January 2008 Volume 86 Number 1

ELSEVIER

Journal of Environmental Management

Journal of

Environmental Management

Aims and Scope

The Journal of Environmental Management is a journal for the publication of peer reviewed, original research for all aspects of management and the managed use of the environment, both natural and man-made. Critical review articles are also welcome; submission of these is strongly encouraged.

As governments and the general public become more keenly aware of the critical issues arising from man's use of his environment, this journal provides a forum for the discussion of environmental problems around the world and for the presentation of management results. It is aimed not only at the environmental manager, but at anyone concerned with the sustainable use of environmental resources.

Research Areas Include, but are not exclusive to:

• resource quality, quantity and sustainability • economics of environmental management • transport and fate of pollutants in the environment • spill prevention and management • remediation of contaminated sites • process modification for pollution prevention • improved energy efficiency • waste treatment and disposal

Papers submitted should address environmental management issues using a range of techniques e.g. case studies, observational and theoretical analyses, the application of science, engineering and technology to questions of environmental concern or mathematical and computer modeling techniques with the aim of informing both the researcher and practitioner.

Note to contributors

Online submission is available for this journal, to submit your article online, go to http://ees.elsevier.com/jema. Instructions on how to submit a manuscript to the journal can be found in the Guide for Authors on the journal homepage at http://www.elsevier.com/locate/jenvman.

Journal of Environmental Management

Editor-in-Chief

A. GILL

ELSEVIER

Orders, claims, and journal enquiries: please contact the Customer Service Department at the Regional Sales Office nearest you:

Orlando: Elsevier, Customer Service Department, 6277 Sea Harbor Drive, Orlando, FL 32887-4800, USA; phone: (877) 8397126 or (800) 6542452 [toll free numbers for US customers]; (+1) (407) 3454020 or (+1) (407) 3454000 [customers outside US]; fax: (+1) (407) 3631354 or (+1) (407) 3639661; e-mail: usjcs@elsevier.com or elspcs@elsevier.com

Amsterdam: Elsevier, Customer Service Department, PO Box 211, 1000 AE Amsterdam, The Netherlands; phone: (+31) (20) 4853757; fax: (+31) (20) 4853432; e-mail: nlinfo-f@elsevier.com

Tokyo: Elsevier, Customer Service Department, 4F Higashi-Azabu, 1-Chome Bldg, 1-9-15 Higashi-Azabu, Minato-ku, Tokyo 106-0044, Japan; phone: (+81) (3) 5561 5037; fax: (+81) (3) 5561 5047; e-mail: jp.info@elsevier.com

Singapore: Elsevier, Customer Service Department, 3 Killiney Road, #08-01 Winsland House I, Singapore 239519; phone: (+65) 63490222; fax: (+65) 67331510; e-mail: asianfo@elsevier.com

USA mailing notice: Journal of Environmental Management (ISSN 0301-4797) is published monthly, with additional copies in January, April, July, and October, by Elsevier Ltd. (The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK). Annual subscription price in the USA US$ 1146 (valid in North, Central and South America), including air speed delivery. Application to mail at periodical postage rate is pending at Rahway, NJ.

USA POSTMASTER: Send address changes to Journal of Environmental Management, Elsevier, 6277 Sea Harbor Drive, Orlando, FL 32887-4800

AIRFREIGHT AND MAILING in the USA by Mercury International Limited, 365, Blair Road, Avenel, NJ 07001.

Printed by the Alden Group, Oxford, UK

Author enquiries
For enquiries relating to the submission of articles (including electronic submission where available) please visit this journal's homepage at http://www.elsevier.com/locate/jenvman. You can track accepted articles at http://www.elsevier.com/trackarticle and set up e-mail alerts to inform you of when an article's status has changed. Also accessible from here is information on copyright, frequently asked questions and more.

Contact details for questions arising after acceptance of an article, especially those relating to proofs, will be provided by the publisher.

Funding body agreements and policies
Elsevier has established agreements and developed policies to allow authors who publish in Elsevier journals to comply with potential manuscript archiving requirements as specified as conditions of their grant awards. To learn more about existing agreements and policies please visit http://www.elsevier.com/fundingbodies

ELSEVIER

Journal of Environmental Management 86 (2008) 1–13

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Biophysical and anthropogenic controls of forest fires in the Deccan Plateau, India

V. Krishna Prasad[a,*], K.V.S. Badarinath[b], Anuradha Eaturu[c]

[a] Agroecosystem Management Program, The Ohio State University, USA

[b] Forestry and Ecology Division, National Remote Sensing Agency, Hyderabad, India

[c] Department of Mathematical Sciences, Kent State University, USA

Received 22 February 2006; received in revised form 31 October 2006; accepted 20 November 2006

Available online 1 February 2007

Abstract

Forest fires constitute one of the most serious environmental problems in several forested regions of India. In the Indian sub-continent, relatively few studies have focused on the assessment of biophysical and anthropogenic controls of forest fires at a landscape scale and the spatial aspects of these relationships. In this study, we used fire count data sets from satellite remote sensing data covering 78 districts over four different states of the Deccan Plateau, India, for assessing the underlying causes of fires. Spatial data for explanatory variables of fires pertaining to topography, vegetation, climate, anthropogenic and accessibility factors have been gathered corresponding with fire presence/absence. A logistic regression model was used to estimate the probability of the presence of fires as a function of the explanatory variables. Results for fire area estimates suggested that, of the total fires covering 47,043 km^2 that occurred during the year 2000 for the entire Indian region, 29.0% occurred in the Deccan Plateau, with Andhra Pradesh having 13.5%, Karnataka 14.7%, Kerala 0.1%, and Tamilnadu 1.15%. Results from the logistic regression suggest that the strongest influences on the fire occurrences were the amount of forest area, biomass densities, rural population density (PD), average precipitation of the warmest quarter, elevation (ELE) and mean annual temperature (MAT). Among these variables, biomass density (BD) and average precipitation of the warmest quarter had the highest significance, followed by others. These results on the best predictors of forest fires can be used both as a strategic planning tool to address broad scale fire risk concerns, and also as a tactical guide to help forest managers to design fire mitigation measures at the district level.

Keywords: Fires; Logistic regression; Biophysical and anthropogenic variables; Deccan Plateau; India

1. Introduction

Fire plays a key role in ecosystem dynamics in many biomes throughout the world (Whelan, 1995; Bond and van Wilgen, 1996). Fire has been frequently documented as a forest clearing tool during slash and burn agriculture (Ramakrishnan, 1992; Tomich and Lewis, 2002; Stolle et al., 2003 and references therein; Fearnside, 2005), tropical forest conversion to pasture (Fujisaka and White, 1998), clearing of savannas (Allen, 1986), clearing of plantations (Ataga et al., 1986), etc. Of the several ecological impacts, forest fires are considered to be a major threat for the loss of biodiversity, including disruption of nutrient cycles (Neary et al., 1996). Small and large fires of varying intensity strongly affect species composition and age distribution of several forest species (Hough and Forbes, 1943). Fire creates a mosaic of burned and unburned forest patches, leaving complex heterogeneous patterns across the landscape. The resulting landscape heterogeneity can further influence successional processes, which in turn may affect the spatial spread of subsequent fires (Turner and Romme, 1994). Fire frequency, cycle, size distribution, intensity, severity, season, and type are all-important in characterizing fire disturbance patterns and thereby landscape heterogeneity (Siegert et al., 2001; Dennis et al., 2005). In addition to alteration of landscapes, vegetation fires are also one of the major causes of greenhouse gas emissions, aerosols and smoke pollution which have an important impact on atmospheric chemistry, visibility and

*Corresponding author. Tel.: +1 330 234 0387; fax: +1 330 263 3686.
E-mail address: krisvkp@yahoo.com (V.K. Prasad).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.017

health (Goldammer, 1990; Levine, 1991; Crutzen and Andreae, 1990; Tomich and Lewis, 2002; Murdiyarso et al., 2004).

At fine spatial and temporal scales (individual trees and forest stands, seconds to hours), the physics of fire spread are well known (Albini, 1976; Rothermel, 1991). The climatic and vegetation constraints on fire occurrence at a variety of scales have also been documented (Gruell, 1985; Swetnam and Betancourt, 1990). For example, during 1997/98 large-scale wildfires were prevalent throughout the tropical forests, particularly in Amazonia and Southeast Asia. The wildfires were triggered partly by the occurrence of extensive droughts, associated with the 1997/98 El Niño events (Siegert et al., 2001; Dennis et al., 2005). Nepstad et al. (1999) assessed the potential for large-scale, drought-induced Amazonian forest burning as a result of ENSO episodes using a regional water balance model and estimated that nearly 270,000 km^2 of forest became vulnerable to fire in the 1998 dry season. Cochrane et al. (1999) inferred that fire incidence rates in the Amazon basin increased during El NinÄ episodes and such fires have the potential for transforming large areas of tropical forests into fragmented scrub forests or savannas. Studies also suggest that in the tropical forest environments of Amazonia and Indonesia, selective logging may lead to an increased susceptibility of forests to fire suggesting the relationship between logging and fire occurrence (Siegert et al., 2001; Dennis et al., 2005 and references therein). vanderWerf et al. (2004) found that fire emissions of greenhouse gases increased across multiple continents in 1997–98 due to El Niño-induced drought including an increase of 60% in Southeast Asia, 30% in Central and South America, and 10% in the boreal forests of North America and Eurasia. In addition to these climatic factors, the assessment of fires as man-made and/or natural hazards involves consideration of a wide range of factors that affect both the occurrence of a single fire event, and the potential effects that the fire may have on human lives and properties, including the ecosystem values (Chuvieco et al., 2003). Also, fire has been used as a tool in farming or land clearing, as a weapon in social conflict, or as a result of accidental or unintended actions in several regions of the world (Goldammer, 1990; Stolle and Tomich, 1999; Tomich and Lewis, 2002; Murdiyarso et al., 2004) including India (Ramakrishnan, 1992; Prasad et al., 2001; Narendran, 2001; Saha, 2002).

The importance of anthropogenic factors in regulating fire events in addition to climate, vegetation and topographic factors, makes fire risk prediction highly challenging (Perry, 1998). A relevant question in this context is then to what extent can regional fire occurrence be related to each of these components. In particular, in the tropical countries, as a result of growing changes in land use, conversion and fragmentation of tropical forests are occurring at a rapid rate and the extent and use of anthropogenic fire are becoming more widespread (Goldammer, 1990; Roberts, 2000). Building predictive models that aid in quantifying fire occurrences is one of the challenging tasks in landscape ecology. In particular, understanding of the interactions between the natural and human environments leading to fire occurrence and subsequent fire regime characteristics is one of the primary concerns in the tropics, which are highly threatened due to rapid population growth and anthropogenic interference (Viegas, 1998).

Also, over the last decades, spatial information technologies have been widely used in forest fire studies. In particular, remote sensing with its multitemporal, multispectral, synoptic and repetitive coverage can provide valuable information on the fire counts, the amount of area burned and the type of ecosystem burned (Morisette et al., 2005). The two primary ways of assessing burned area with remote sensing are (a) active fire detection and (b) post-fire burn detection (Fraser et al., 2000; Eva and Steffen, 2003; vanderWerf et al., 2004). Fires, because of their high temperature, emit thermal radiation with a peak in the middle infrared region, in accordance with Planck's theory of blackbody radiation. Therefore, active fire sensing is often done using middle infrared and also thermal infrared (usually around 3.7–11 mm) information from satellites (Dozier, 1981). For example, Pozo et al. (1997) mapped fire growth by using multitemproal and multispectral images from NOAA-AVHRR data. The technique is mainly based on comparison of brightness temperature differences and AVHRR channels 3 and 4 for successive images, for mapping fires. In contrast, post-fire burn detection is mostly done through measuring changes in surface reflectance before and after the fire mainly from red (0.65–0.70 μm) and near infrared (NIR) (0.7–3.0 μm) information (Eva and Steffen, 2003; vanderWerf, 2004; Smith and Wooster, 2005). Burnt areas have a lower reflectance in the red and near infrared channels than healthy vegetation. Thus, vegetation indices such as the normalized difference vegetation index, computed as NIR−Red/NIR+Red reflectances that exhibit decreased values after vegetation burning, were widely used for post-fire burnt area assessment (Flannigan and Vonder Haar, 1986; Kaufman et al., 1990; Setzer and Pereira, 1991; Justice et al., 1996). Several of the post-fire burn detection studies measure the degree of NDVI change by subtracting pre- and post-fire NDVI (Kasischke et al., 1993). Also, to detect the burnt surfaces, information on surface temperature derived from thermal bands has been used (Eva and Lambin, 2000). After the occurrence of fires, land surface brightness temperature (T_s) rises along with the decrease in the latent heat flux from the surfaces. Landmann (2003) and Smith et al. (2005) showed that, in the case of Savanna fires, surface spectral reflectance increases with increasing fire severity due to the formation of increasing quantities of white mineral ash. Also, their studies showed linear relationships between fire duration and post-fire surface spectral reflectance, with the optimal relationship being a ratio of the 450 and 2034 nm spectral reflectance observations. In addition, Stroppiana et al. (2002), through exploring different spectral methods and band combinations for burnt area mapping, found that minimum near infrared composites contain the most recently burnt pixels and are least affected by artifacts. Their analysis of burnt area spectral signatures showed that the NIR band and the global environmental monitoring index (GEMI) (Pinty and Verstraete (1991) are the most appropriate variables for designing a burnt area algorithm due to their sensitivity to changes induced by the fire on the vegetation cover. Fraser et al. (2000) used a hybrid approach involving both thermal and NDVI data for detecting active fires.

Most of the studies conducted so far in the tropical countries using satellite data for forest fires including risk mapping and modeling have proved the usefulness of remote sensing both by medium resolution satellites (Tanaka et al., 1983; Milne, 1986; Ribed and Lopez, 1995) and the coarse resolution satellite NOAA-AVHRR (Muirhead and Cracknell, 1985; Matson and Holben, 1987; Chuvieco and Martin, 1994). In contrast to several such studies on fire detection and modeling attempted in different parts of the world (Ehrlich et al., 1997; Cochrane and Schulze, 1998; Nepstad et al., 1999; Vazquez and Moreno, 2001; Eva and Lambin, 2000; Stolle and Lambin, 2003; Setiawan et al., 2004), there are relatively few studies in the Indian region that actually analyzed the post-fire data and assessed the underlying causative factors. Considering the huge impact and potential losses caused due to forest fires in the Indian region (IFFN, 2002), studies that characterize fire occurrence/events, their spatial density and temporal evolution, in addition to assessing the causative factors of forest fires, are significant.

The main objective of this study is to understand the major spatial determinants of fires in the Deccan Plateau, India. We used fire count data sets from SPOT satellite data to (a) document the spatial extent of the area burned at the state level in the Deccan Plateau, (b) analyze forest fire events and underlying causes using topographic, vegetation, climatic, and socioeconomic factors and (c) develop a statistical predictive model that best explains the fire occurrence patterns (presence and absence) across diverse geographical and climatic gradients at the district level in the Deccan Plateau. We addressed these objectives using remote sensing and Geographic Information Systems (GIS) as they represent the best available tools to deal with the spatial nature of fires (Morgan et al., 2001). The results from this study will be useful to forest managers and scientific researchers to address some forest fire mitigation strategies and to those seeking to manage fire-prone vegetation types in the study area.

2. Study area

2.1. Deccan Plateau

The Deccan Plateau is a vast plateau in India, encompassing most of Central and Southern India. The term "Deccan" comes from the Sanskrit word dakshina, meaning "the south". It comprises the whole of peninsular India south of the Vindhya range and mainly encompasses the four states of Andhra Pradesh, Karnataka, Kerala and Tamil Nadu, covering nearly 78 districts (Figs. 1 and 2). The Deccan Plateau is bounded by the Western Ghats to the west, the Eastern Ghats to the east, the Nilgiris to the south and the Satpura and Vindhya ranges to the north. The terrain is undulating with elevation (ELE) ranging from 1500 to 2500 ft (450–750 m). At the plateau's margins in the Western Ghats, a steep escarpment drops sharply to the narrow Malabar Coast on the Arabian Sea. To the east the plateau descends more gradually to a broader alluvial plain extending into Andhra Pradesh and Orissa states. Several major rivers, including the Kâveri (Cauvery), Godâvari, Krishna, and Penner, flow across the eastward-tilting plateau before reaching the Bay of Bengal. The Deccan Plateau has a dry season that lasts 6–9 months.

2.2. Forests and people of the study area

Three different ecoregions dominate the Deccan Plateau. They are the South Western hill ranges occupying dense and rich montane rain forests, the Southwestern hill ranges with moist deciduous forests, and the south Deccan Plateau with dry deciduous forests. The tall Western mountain range towards the western part of the Deccan Plateau intercepts the moisture from the southwest monsoon; therefore, the eastern slopes of the Deccan Plateau receive very little rainfall, resulting in totally different vegetation formations. For example, the interior districts of Andhra Pradesh contain tropical dry forests intermingled with pockets of moist mixed deciduous forests. The forest types of the Deccan Plateau mainly fall into six major categories (Champion and Seth, 1968), the tropical wet evergreen forests, south montane wet temperate forests, tropical semi-evergreen forest, tropical moist deciduous forest, tropical dry deciduous forests and tropical thorn forests. The area estimates of the very dense, moderately dense and open forests based on the forest cover assessment from Indian Remote Sensing Satellite (IRS)-1D LISS III data with 23 m resolution at 1:50,000 scale (SFR, 2003), for the study area are given in Table 1. The very dense forest refers to forest canopy density above 70%, the moderately dense forests have 40–70% canopy density and the open forests have 10–40% crown density.

The main reasons for forest degradation in the Deccan Plateau include fire, cattle grazing and fuel wood extraction. Several indigenous people inhabit the states of the Deccan Plateau. Those include the Konda reddis, Koyas, Konda Kapus, Bodo Gadaba, Chenchus, Kammara, Kondh Koya, Lambadas, Malis, Nayaks, Thoti, Yenadis, Yerukulas, Valmiki, Kani's, Gond's, etc. Large-scale livestock grazing is the most common problem in several forest areas of India (Panwar et al., 1993) including the Deccan Plateau (Silori and Mishra, 2001). Although restrictions were imposed on cattle grazing by the local

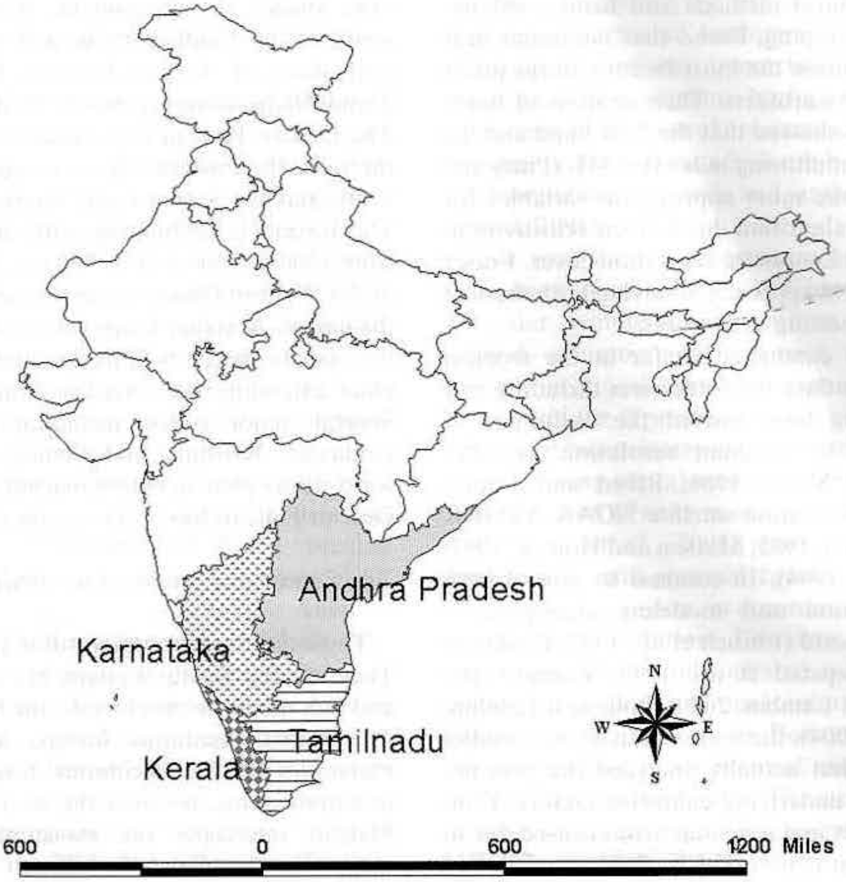


Fig. 1. Indian region depicting the states of the Deccan Plateau.

forest departments, they are not stringent. With the increasing cattle numbers, over-grazing is one of the major factors in forest degradation. For example, in the Mudumalai Wildlife Sanctuary of the Deccan Plateau (Western Ghats, Tamil Nadu) covering an area of $321\,Km^2$, on average, each village family owns 10–15 animals (cattle, water buffalo, goats, sheep, etc) (Silori and Mishra, 2001). Livestock herds enter 5–6 km into the surrounding forest through different entry points and spend about 8–9 h grazing. It is estimated that almost 12,000–15,000 livestock graze in this forest sanctuary every year, resulting in tremendous pressure on the vegetation. Also, Silori and Misra (2001) noted that almost 98% of the total livestock population grazes throughout the year inside the forest corridors and adjoining reserve forests, resulting in huge forest degradation. In addition, for several indigenous people, slash and burn agriculture is one of the major occupations. Indigenous people's rights to the use of the forests they occupy are closely linked to traditions and customs, but are not based on land ownership as legally defined. While the indigenous people's management and use of forest resources may be sound from the ecological point of view, their traditional rights that go back many centuries are not recognized by the forest department of India in several states covering the Deccan Plateau. In many ways, the official policies of the State forest departments directly conflict with the livelihood strategies of local indigenous people. Conflicts arise over forest product collection, encroachment into protected areas and non-compliance with recent environmental laws enacted to protect forests. In colonial times, there was no scarcity of forests for several indigenous people to clear and cultivate (slash and burn agriculture). Villages were small and were moved along to the next cultivation patch. Within an area that the people considered as theirs, and where they had exclusive hunting rights, there was enough land to cultivate and abandon for a required fallow. However, in recent times, a combination of many development forces has gradually made inroads into the living space of the indigenous people, affecting their lifestyle. Mainly due to population growth, spreading urbanization and expanding commercial interests, the available forest spaces as well as the traditional resource use areas of indigenous people are being reduced.

3. Spatial data sets

3.1. SPOT fire data set

In this study, we used the SPOT fire count data set over the Indian region (Fig. 2) for assessing the greenhouse gas emissions from biomass burning. Detailed descriptions of

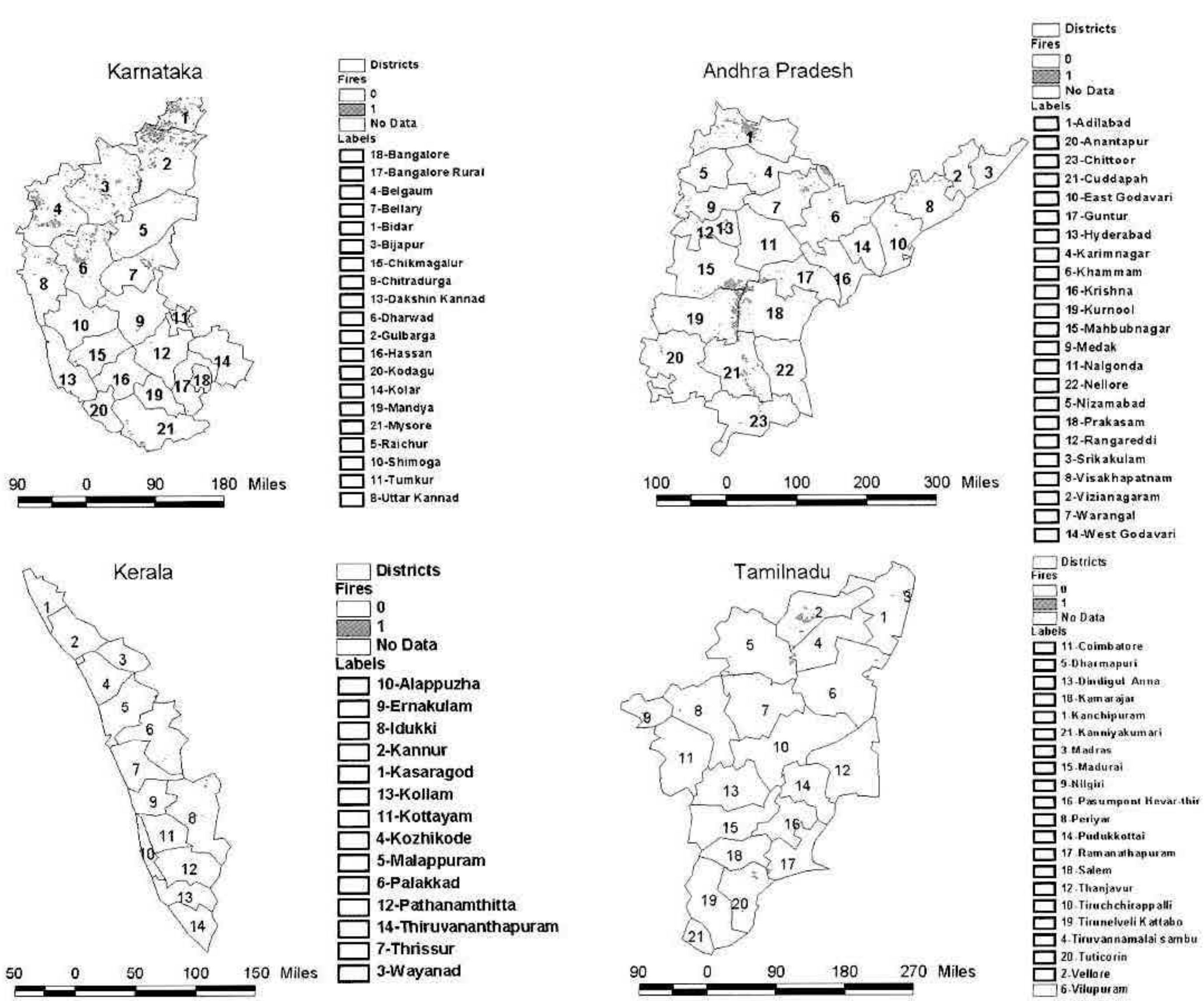


Fig. 2. Four different states with the districts in the Deccan Plateau. Fire counts from SPOT satellite data were overlaid on district cover.

Table 1
Forest area estimates (km^2) for different states of Deccan Plateau

States	Dense	Moderately dense	Open forest	Total	% of country's forest cover
Andhra Pradesh	23	23356	20040	44419	6.55
Karnataka	431	22030	13988	36449	5.37
Kerala	334	9294	5949	15557	2.30
Tamilnadu	2440	9567	10636	22643	3.34

Source: SFR (2003).

these SPOT fire count data sets were provided by Tansey (2002), Stroppiana and Grégoire (2002), Gregoire et al. (2003), Tansey et al. (2004). For developing the SPOT fire data set, a set of 14 months of daily global imagery acquired by the VEGETATION (VGT) instrument on-board the SPOT-4 satellite was assembled as a part of the Millenium Ecosystem Assessment. The products were calibrated, geo-referenced and corrected for atmospheric effects on surface reflectance in four spectral wavelengths (blue, red, near-infrared and short wave infrared), with a pixel size of approximately 1 km, and a multi-temporal registration accuracy better than 0.5 km. The global burnt area maps for the year 2000 are based on a series of regional algorithms that fit specific climatic zones and ecosystem conditions. The final version of the global burnt area maps was completed in December 2002 and made accessible to the user community via the Internet at http://www.gvm.jrc.it/fire/gba2000/index.htm. Extensive details on these fire products were discussed in Gregoire et al. (2003) and Tansey et al. (2004). In this study, we used this data set over the Indian region to characterize the forest fires. For estimating the spatial density of fires at the state and district level, we overlaid the corresponding political and district boundaries of India in Geographic Information Systems (GIS) using ARCGIS (9.x) (ESRI$^{@TM}$) software. We used the Spatial Analyst extension to create individual maps of fire counts at a district level. The data from these maps were then used for extracting the fire statistics. The data has been coded into binary variables, primarily, the districts in which fires were observed are coded as 1 and the others as zero. Corresponding to the fire count data, extensive biophysical and socioeconomic information has been gathered from local sources and spatial methods. The data included topographic, vegetation, climatic, and anthropogenic and accessibility components.

3.1.1. Topographic parameters

For the past several years, fire behavior models have incorporated the interaction of fire spread with fuels, weather and terrain (Albini, 1976). Fire spread was also attributed to fire line intensity, i.e., the rate of energy released by the flaming front (Rothermel, 1991). Other terrain effects on fire intensity and spread were incorporated indirectly through fuel type and moisture (Anderson, 1982). The effect of terrain attributes on forest survival following wildfire has been assessed by Kushla and Ripple (1997) and others. We used five different topographic parameters as causative factors of fires, these included ELE, slope, aspect and compound topographic index. The explanations of these variables are provided in Table 2a.

3.1.2. Vegetation parameters

Vegetation parameters play a vital role in the ignition, spread and dispersal of fires. In this study, for predicting the forest fire probability, we used both the vegetation type parameters as well as biomass data. An area's vegetation must be considered because some vegetation types are more flammable than others, thereby increasing the fire hazard. Fuels represent the organic matter available for fire ignition and combustion (Rothermel, 1991; Albini, 1976; Chuvieco et al., 2003). Thus a spatial distribution of fuels is fundamental to assessing fire hazard across a landscape. In this study, we used two different parameters representing vegetation (1) biomass density (BD) and (2) percent forest area. The explanations of these parameters are provided in Table 2b.

3.1.3. Climatic parameters

Fire occurrence, frequency and intensity are primarily dependent on climate, directly through weather conditions, which allow fires to develop, and indirectly through the supply of a sufficient vegetation fuel load to sustain fire. Climate also plays an essential role in fire ignition and propagation, either through its influence on fuel moisture content (Rothermel, 1991), on the conditions for fire propagation (Albini, 1976) or in providing natural (i.e., lightning) ignition sources. In this study, we used temperature as well as precipitation (summed over different time periods) as predictors of fires at the district level in the study region. Six different subsets of temperature and precipitation parameters have been used (Table 2c).

3.1.4. Anthropogenic factors

The demand for fuel wood by rural people in the study region is closely linked with slash and burn agriculture, in which the biomass is clear felled and then burnt subsequently for land clearing purposes. The tribal dependence on forests for fuel wood as a major source of energy is causing serious deforestation in several parts of the study area. Commercial fuel is beyond the reach of the tribal communities due to their poor socioeconomic conditions. Also, due to the ever-increasing population, fuel wood consumption in some of the districts is increasing rapidly. For example, the average fuel wood consumption is significantly high at about $5.23\,kg^{-1}\,day^{-1}$. Since the majority of the population lives in the rural areas where fuel wood provides most of the energy requirements, we used indices that combined population measures with forest area parameters as causative factors of fire. These included rural population density (PD), illiteracy rates and rural population to forest area ratio (Table 2d). The database for these variables has been created from the locally available census data at the village level for the entire Deccan Plateau from the local government records as well from the Census of India database (Census of India, 2005).

3.1.5. Accessisibility factors

Accessibility of remote forested regions increases the influx of people into the areas, spurring forest disturbance. This may cause accidental fires mainly due to negligence. In this study, we used two different indices, the alpha and gamma indices, to assess the connectivity of roads at the district level. GIS-based analysis using ARCGIS network analyst has been attempted at the individual district level using road layers. In network analysis, a segment of a linear feature is called a link or an edge. The two vertices or nodes at either end of the edge in the network define an edge. The number of edges and vertices in a network is often used to derive statistics reflecting the structural characteristics of a road network (Wong and Lee, 2005). The definitions of the alpha and gamma indices were provided in Table (2e).

Table 2

Sl. no.	Topographic variables	Explanation
(a) Topographic parameters		
1.	Elevation (ELE)	It is an important physiographic factor that is related to wind behavior and hence affects fire-proneness (Rothermel, 1991). Fire travels most rapidly up-slope and least rapidly down-slope. Elevation values for fire pixels have been extracted from GTOPO30 digital elevation model (DEM) with a horizontal grid spacing of 30 arcsec (~1 km).
2.	Slope (SLP)	Is an indicator of rate of change of elevation and steepness of the terrain. Slope affects both the rate and direction of the fire spread. Fires usually move faster uphill than downhill, thus steeper the slope, the faster the fire will move (Rothermel, 1991; Kushla and Ripple, 1997). Values were derived from GTOPO 30 DEM.
3.	Aspect (ASP)	Describes the direction of the maximum rate of change in the elevations between each cell and its neighbors. Southern aspects receive more direct heat from the sun, drying both the soil and the vegetation. Also, fuels are usually drier and less dense on southern slopes than fuels on northern slopes. South-facing slopes will normally be with higher temperatures, stronger winds, lower humidity and lower fuel moistures (Anderson, 1982). Aspect values were derived from GTOPO 30 DEM.
4.	Compound topographic index (CTI)	Compound topographic index (CTI) also known as wetness index is a function of upstream contributing area and the slope of the landscape. The spatial distribution of water on a field is influenced by lateral flow and thus controlled by elevation differences. The topographic wetness index is a compound terrain attribute calculated from specific catchments area of a point (A_s) and the local slope gradient $\tan\beta$ (Beven and Kirby, 1979). Wetness index given as $$\text{Wetness index} = \frac{\ln(A)}{\tan\beta}. \qquad (1)$$ where ln is the natural logarithm, β is the slope angle and 'A' is the upslope area per unit width of contour. In general, the index essentially is a measure of the tendency of water to accumulate at any point on a slope. The topographic wetness index maps indicate zones of high potential soil moisture (high values) and zones that dry up first (low values). Thus higher CTI values are not a good indicator of fires. CTI values have been calculated for individual pixels using GTOPO 30 elevation data sets.
5.	Topographic position index (TPI)	The TPI is calculated from DEM as the difference between a cell elevation value and the average elevation of the neighborhood around that cell. Positive values mean the cell is higher than its surroundings while negative values mean it is lower. TPI values can easily be classified into slope position classes based on how extreme they are and by the slope at each point. Four categories based on TPI and slope have been derived. *Canyon Bottom*: TPI$\leqslant -8$; *Gentle Slope*: $-8<$TPI$\leqslant 8$, Slope $<6°$; *Steep Slope*: $-8<$TPI $\leqslant 8$, Slope $\geqslant 6°$; *Ridgeline*: TPI $\geqslant 8$ (Weiss, 2001; Jennessent, 2006).
(b) Vegetation parameters		
Sl. no.	Vegetation parameters	Explanation
1.	Biomass density (BD)	Higher the quantity of fuel, higher the flammability. As the amount of flammable material in a given area increases, the amount of heat produced by the fire also increases (Albini, 1976). Biomass density values for individual forest types at district level have been aggregated from local forest records and then calculated for individual pixels based on forest type.
2.	% forest area	The amount of area susceptible for fire increases as forest area increases. Fire spread also depends on the continuous availability of biomass material.
(c) Climatic parameters		
Sl. no.	Climatic parameters	Explanation
1.	Mean annual temperature (MAT)	High temperature facilitates rapid drying of biomass, in particular dry grass, dead leaves, tree needles and small trees. Fuels pre-heated by high average temperate burn more rapidly than cold fuels, hence a positive indicator of fires.
2.	Average temperature of the warmest quarter (ATWQ)	As above. Temperature data averaged from March to June.
3.	Mean annual precipitation (MAP)	Negative indicator of forest fires as high moisture levels suppress fires.
4.	Average precipitation of the warmest quarter (APWQ)	Higher values contribute to high moisture in fuels, thus negative indicator of fire spread. Precipitation data averaged from March to June.
(d) Anthropogenic parameters		
Sl. no.	Socioeconomic parameters	Explanation
1.	Population density (PD)	The higher the density, the higher the dependence on surrounding forest resources. Fires in several districts of the study area are caused mainly due to land clearing for agriculture purpose (slash and burn agriculture).

Table 2 (*continued*)

Sl. no.	Topographic variables	Explanation
2.	Illiteracy rates (ILR)	Literacy rates are an indicator of awareness relating to forest sustainability and accidents that may cause forest fires. Low illiteracy rates can have negative implications on environmental concerns, including forest fires.
3.	Rural population to forest area ratio (RPFR)	An indicator of demographic pressure that indicates the possible dependence of rural people on forest resources. Rural population also includes indigenous tribes in the districts of West Godavari, East Godavari, Adilabad, Khammam, etc. This index has been calculated from local district statistics aggregated for local *mandals* (*subcategory of a district*) *and calculated for individual pixels as density values.*
(c) Accessibility criteria		
Sl.No	Accessibility indicators	Explanation
1.	Alpha index	An index of connectivity is based on the number of circuits that a network can support. A circuit is defined as a closed loop along a road network. In a circuit, the beginning node of the loop is also the ending node. The existence of a circuit in a road network implies that there are multiple paths to connect any two vertices in the road network. Alpha index is computed as the ratio of the number of actual circuits to the number of maximum possible circuits. Alpha index thus is an index of road connectivity (Wong and Lee, 2005). The higher the connectivity, the more susceptibility of the forests to fire danger, mainly due to anthropogenic interference.
2.	Gamma index	Gamma index is defined as the ratio of the actual number of edges to the maximum possible number of edges in the network. The index is computed as, $\gamma = e/e_{max}$. Gamma index is most useful in comparing different road networks to differentiate their levels of connectivity. The higher the index value, higher the level of connectivity (Wong and Lee, 2005). The index has been computed at district level using road layer in ARCGIS Network analyst.

4. Statistical analysis

For predicting the presence/absence of fire, site-specific fire data in conjunction with explanatory variables of topographic, vegetation, climatic, anthropogenic and accessibility factors were exported to spreadsheets for statistical analysis. The range of values for the explanatory variables is given in Table 3. We used logistic regression to model the presence/absence of fires in different districts of the Deccan Plateau. Logistic regression is based on transformation of a linear equation that uses a binomial logistic distribution where the probability of the response can be mathematically expressed as a function of several explanatory variables. In the logistic regression analysis, presence/absence of fire was used as the dependent variable and the others as independent explanatory variables. The logistic regression model for more than one independent variable can be expressed as the probability of fire occurrence

$e^z/(1+e^z)$, where z is the linear combination of

$$z = B_0 + B_1X_1 + B_2X_2 + B_3X_3 \ldots B_PX_P.$$

$B_0, B_1 \ldots B_P$ are the coefficients estimated from the data; $x_1, x_2, \ldots, x_P$ are the independent variables and 'e' is the base of the natural logarithm. In the logistic regression, the parameters of the model were estimated using a maximum likelihood method (McCullagh and Nelder, 1989). We evaluated all possible combinations of the explanatory variables in the logistic regression framework. We performed a forward stepwise logistic regression with a P-to-enter of 0.1 and P-to-remove of 0.15 on the predictor variables to select the best subset of independent variables for the model selection (Hosmer and Lemeshow, 1989). The variables left out of the analysis at the last step all have significance values larger than 0.05, so no more explanatory variables were added. The outcome variable was the presence and absence of fires in individual districts of the Deccan Plateau, coded as 1 and 0, respectively. These variables were then included as dependent variables and the rest of the topographic, vegetation, climatic, anthropogenic and accessibility variables, as independent explanatory variables in a logistic regression framework. So as to select the best predictors among the independent variables, we used Akaike's information criteria (AIC); a penalized version of the likelihood function in which the best model fit is given by the lowest value (Akaike, 1973; Burnham and Anderson, 1998; Lancelot et al., 2002; Breck et al., 2003; Zucchini, 2000; Frappier and Eckert, 2003). Significant variables at each step have to reduce the scaled deviance significantly (McCullagh and Nelder, 1989; Buckland et al., 1997). Although initially all explanatory variables were potential predictors, only those variables selected using the above criteria were used in the final predictive model for detecting the presence/absence of fires. Regarding model selection between successive models, the models were ranked by AIC (lower values indicate a better fit) and evaluated with respect to the difference in AIC between a given model and the highest ranked model. We selected the model that has a delta AIC value of zero (Burnham and Anderson, 2002). Nagelkerke's R^2 multiplied by 100 closely approximates the percent variance in the outcome variable that was explained by the model

Table 3
Parameters used in logistic regression analysis for modeling fires in the Deccan Plateau

Sl. no	Factors	Range of values (minimum and maximum)
1.	Elevation (ELE) in meters.	50–860
2.	Slope (SLP) (%)	4.85–8.65
3.	Aspect (ASP) (deg)	131–290
4.	Compound topographic index (CTI)	4.18–7.12
5.	Topographic position index (TPI)	1–4
6	Percent forest cover (PFC)	5.08–81.25
7	Biomass density (BD) (per km^2)	0.2–0.62
8	Mean annual temperature (MAT) (°C)	21.3–28.02
9	Average temperature of the warmest quarter (ATWQ) (°C)	24.2–36.0
10	Mean annual precipitation (MAP) (mm)	699–1479
11	Average precipitation of the warmest quarter (APWQ) (mm)	55–459
12	Rural population density (RPD)	158–1489
13	Illiteracy rate (ILR) (%)	8–54
14	Rural population to forest area ratio (RPFR)	265–84,301
15	Alpha index	0.04–0.09
16	Gamma index	0.03–0.43

because it is adjusted for sample size and transformed so that a value between 0.0 and 1.0 can be achieved (Tabachnik and Fidell, 2001). One of the important criteria for assessing the model fit is the Walds statistic. The Wald statistic is an alternative test that is commonly used to test the significance of individual logistic regression coefficients for each independent variable (that is, to test the null hypothesis in logistic regression that a particular logit (effect) coefficient is zero). For dichotomous independents, the Wald statistic is the squared ratio of the unstandardized logit coefficient to its standard error. In addition, the odds ratio is a way of comparing whether the probability of a certain event is the same for two groups, i.e., in our case fire presence and absence. An odds ratio of 1 implies that the event is equally likely in both groups. An odds ratio greater than one implies that the event is more likely in the first group. An odds ratio less than one implies that the event is less likely in the first group. The results relating to coefficients, standard errors, goodness-of-fit statistics, and odds ratio in addition to classification results for the data set have been reported.

5. Results

For assessing the presence ~~and~~ absence of fires and contributing factors, 78 districts covering four different states in the Deccan Plateau have been analyzed. Of the total fires covering 47,043 km^2 that occurred during the year 2000, 29.0% occurred in the Deccan Plateau, covering the states of Andhra Pradesh (13.5%), Karnataka (14.7%), Kerala (0.1%), and Tamilnadu (1.15%) (Table 4). In the forward model selection, the AIC values continued to decline as stepwise model selection was completed. Although the characteristic increase in AIC values associated with model over fit did not occur in any of the selected models, the decline in AIC values approached an asymptote after the ninth step in the model selection iteration. Thus, the model selected in the ninth step of the forward stepwise procedure represented the model least likely to over fit the training data while retaining the optimal number of factors important for discriminating presence and absence of fires. The final selected model retained six variables out of the 16 explanatory variables used in the study. Thus, the best predictors of fire presence and absence in the Deccan Plateau were found to be forest area, BD, PD, annual precipitation of the warmest quarter (APWQ), ELE and mean annual temperature (MAT). The beta coefficients for the model along with standard error, Wald's statistic, significance values and odds ratio were given in Table 5. The model differed significantly from the constant only model and had a Nagelkerke's R^2 of 0.611. Among the six best predictors, the Wald statistic with higher significance has been found for predictors of BD and APWQ. Thus, for example, the odds of the dependent variable (fires) increases 20-fold when the forest area increases by three times. A classification probability cutoff value of 0.5 yielded the best classification of fires (Table 6). The classification table (Table 6) is a 2×2 table, which tallies correct and incorrect estimates for the full model with the independents as well as the constant. The columns

Table 4
Burnt areas derived from SPOT fire data sets in different states of India for year 2000 and % contribution of the states of Deccan Plateau to total fires in Indian region

State	Fire pixels (km^2)	% Contribution
Andhra Pradesh	6369	13.53
Karnataka	6938	14.74
Kerala	48	0.10
Tamilnadu	542	1.15
Total	13,897	29.5
Total fires in India	47,043	—

Table 5
Individual predictor variables that were retained in the final stepwise logistic regression model associated with fires in Deccan Plateau

Predictors/indicators	β	SE	Wald	df	Significance	Odds ratio
Forest area	3.05	1.7	3.218858	1	0.002	21.11534
Biomass density (BD)	3.8	1.2	10.02778	1	<0.001	44.70118
Population density (PD)	0.034	0.018	3.567901	1	0.003	1.034585
Annual precipitation of the warmest quarter (APWQ)	−0.42	0.15	7.84	1	<0.001	0.657047
Elevation (ELE)	0.4	0.2	4	1	0.01	1.491825
Mean annual temperature (MAT)	0.611	1.06	0.332254	1	0.006	1.842273
Constant	−22.91	9.2	6.201183	1	0.000	0

Table 6
Number of fires as predicted from logistic regression in different districts of Deccan Plateau

Number of fires present	Fires absent	Fires present	% Correct
Fires absent	25	5	83.3
Fires present	5	42	89.4
Overall percentage	—	—	87.0

Classification table with accuracy assessment (the cutoff value is 0.500).

are the two predicted values of the dependent variable, while the rows are the two observed values of the dependent variable. The model correctly classified 89.0% of the fire presence and 83.0% of the fire absence in different districts of the Deccan Plateau yielding an overall correct classification rate of 87.0% (Table 6).

6. Discussion

This study represents the first multi-statewide effort in the Deccan Plateau, India that estimated the causative factors of fire using SPOT satellite data. The results from the logistic regression framework clearly suggested that in the different districts of the Deccan Plateau, the strongest influences on the fire occurrences were forest area, BD, PD, average precipitation of the warmest quarter, ELE and MAT. Among these variables, BD and average precipitation of the warmest quarter (March–June) had the highest significance, followed by the others. Of the different measures of human influence, only PD was significant. As expected, population densities showed positive influence on the fire risk, suggesting that human caused ignitions can be prominent in the Deccan Plateau. Although accessibility factors from roads were also reported by several researchers as one of the major causes of fires (Pfaff, 1999; Chomitz and Gray, 1996), results for these metrics were not significant. ELE as an important predictor of fires in the study area clearly suggests that fires in the Deccan Plateau are highly related to slash and burn agricultural practices that are mostly practiced by the indigenous people in the highly elevated areas. The importance of MAT as a positive predictor in addition to average precipitation during the warmest quarter as a negative predictor of fires clearly suggests the importance of climatic parameters in regulating the fire events. Most importantly, higher temperature would favor drying of biomass material, thus favoring ignition patterns, whereas higher precipitation events suppress the ignitions.

The quantitative results from the above logistic regression framework clearly provide several benefits for risk prioritization in several forested districts of the Deccan Plateau. However, we also acknowledge some of the data gaps and limitations of the modeling approach followed in this study. The quality of the probabilistic fire risk model is dependent on the quality of the data used to create it. The limitations of the current study are with respect to some ground measurements such as moisture content of fuel materials, fuel packing ratios, soil type data, etc. Although SPOT vegetation data sets have considerable advantages in detecting burnt areas mainly due to their geometric and radiometric characteristics, the VGT sensor lacks thermal channels as in NOAA-AVHRR and MODIS (Fraser et al., 2000; Stroppiana et al., 2002; Giglio, 2005). Use of MODIS data sets with thermal channels would have helped in detecting the real-time fire radiative energy as 'fire counts' in the study region, instead of burnt areas. MODIS data thus can be effectively used for detecting the strong infrared thermal emission signal associated with active fires (Kaufman et al., 1998; Wooster et al., 2003; Giglio et al., 2006). Also, MODIS data sets offer a larger dynamic range of radiance values (12-bit quantization) than AVHRR data (10-bit quantization), thereby avoiding or lessening the saturation problem that is most often the problem in fire detection using NOAA-AVHRR (Ichoku et al., 2003). Further, in our analysis, no attempt was made to account for spatial autocorrelation in the data sets and therefore, there is a chance for inflation of the significance values. Nevertheless, our approach combining the fire count data sets with other biophysical and socioeconomic variables in a logistic regression framework provided useful insights into the best predictors of forest fire risk in the study area.

Given the number of potential explanatory variables tested and the limitations associated with them, the model results still make sense, as these predictors were also reported to be causative factors of fire by previous researchers elsewhere (Skole and Tucker, 1993; Pfaff, 1999; Eva and Steffen, 2003; Laake et al., 2004; Fearnside, 2005). The results obtained from this study have implications for forest fire management. In India, forest fires and their management have a long history. The traditional methods of fire protection in the past used networks of fire lines, block lines, and guidelines. For example, the forest department in different parts of the country is using 'controlled fires' of 5–10 m in width in areas prone to fire due to human negligence. All forms of vegetation are pulled down along this fire line. Though these practices were successful when the population pressure was low, they no longer work effectively (Ramnath, 1997). The existing human resources in the forest department at present are disproportionate with respect to forest degradation caused by human interference, more so from forest fires. On an average 500 ha of forests have to be patrolled by one forest guard, and it does not seem to be likely that this human resource will be sufficient to detect and prevent all forest fires. Therefore intensive management to prevent, detect and suppress forest fires is required. To develop the effective fire management strategies, forest managers and environmental scientists need to utilize the full spectrum of biophysical as well as socioeconomic parameters that can aid in fire prediction and implement the necessary mitigation strategies. Assessment of fire potential at any particular scale requires basic information on the biophysical and weather conditions that combine to produce the potential fire environment. The results from this study relating to these factors in addition to population densities can also be used to assess the susceptibility of any vegetation to fires and for determining future fire risks. Further, remote-sensing data with its spatial information, when combined with GIS and statistical models, allow fire managers and personnel to predict 'where and when' forest fires will most likely occur. Local forest departments can use such databases and models to assess the 'potential fire risk' by combining different types of information. Finally, it may also be mentioned that the forest policies which have been adopted to prevent deforestation and shifting cultivation in the study area have had limited success because of their narrow focus (Narendran, 2001). For the most part, the policy measures have not addressed the constraints of the local rural poor, especially in enhancing their livelihood security through meeting the energy as well as food demands. Unless these demands are met, exploitation of forest resources will continue (Saha, 2002). There is a need for a paradigm shift in forest management and conservation involving local indigenous people. Local communities will actively participate in forest fire prevention and control only when they have a stake in forest management and benefits from the forests (Ramakrishnan, 1992). Such practices are urgently needed to protect the dwindling forest resources in different states of the Deccan Plateau, India.

7. Conclusions

Analysis of fire data from SPOT satellite data sets for the year 2000, covering four different states and 78 districts in the Deccan Plateau, India, suggests that 28% of the fires in India occurred in this region. A GIS framework in conjunction with logistic regression analysis was quite useful in characterizing the landscape level controls of fire. The best predictors of fires in the Deccan Plateau were found to be forest area, BD, PD, APWQ, ELE and MAT. These results can be used both as a strategic planning tool to address broad scale fire risk concerns and also as a tactical guide to help managers to design fire mitigation measures at the district level in the study area.

Acknowledgments

Authors gratefully acknowledge the Global Vegetation Monitoring Unit, Joint Research Center, European Commission, France for vegetation and fire data sets. Comments and suggestions received on the draft manuscript from scientists of National Remote Sensing Agency, Hyderabad, India are gratefully acknowledged.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principal. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory, Akedmia Kiado, Budapest, pp. 267–281.

Albini, F.A., 1976. Estimating wildfire behavior and effects. General Technical Report INT-30. US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 92pp.

Allen, T.G., 1986. Land clearing in African savannas. In: Lal, R., Sanchez, P.A., Cummings, R.W. (Eds.), Land Clearing and Development in the Tropics. Balkema, Rotterdam, pp. 69–80.

Anderson, H.E., 1982. Aids to determining fuel models for estimating fire behavior. USDA Forest Service, Intermountain Forest and Range Experiment Station General Technical Report INT-122, Ogden, UT, 22pp.

Ataga, D.O., Onwubuya, I.I., Omoti, U., 1986. In: Lal, R., Sanchez, P.A., Cummings, R.W. (Eds.), Land Clearing and Development in the Tropics. Balkema, Rotterdam, pp. 87–96.

Beven, K., Kirby, M.J., 1979. A physically based variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 303–325.

Bond, W.J., van Wilgen, B.W., 1996. Fire and Plants. Chapman & Hall, London.

Breck, S.W., Wilson, K.R., Andersen, D.C., 2003. Beaver herbivory and its effect on cottonwood trees: influence of flooding along matched regulated and unregulated rivers. River Research and Applications 19, 43–58.

Buckland, S.T., Burnham, K.P., Augustin, N.H., 1997. Model selection: an integral part of inference. Biometrics 53, 603–618.

Burnham, K.P., Anderson, D.R., 1998. Model Selection and Inference: A Practical Information-Theoretic Approach. Springer, New York.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multi Model Inference: A Practical-Theoretic Approach. Springer, New York, 32pp.

Census of India, 2005. 〈http://www.censusindia.net/〉

Champion, H.G., Seth, S.K., 1968. A Revised Survey of the Forest Types of India. Manager of Publications, Delhi, India.

Chomitz, K.M., Gray, D.A., 1996. Roads, land use and deforestation: a spatial model applied to Belize. World Bank Economic Review 10 (3), 487–512.

Chuvieco, E., Martin, M.P., 1994. A simple method for fire growth mapping using AVHRR channel-3 data. International Journal of Remote Sensing 15, 3141–3146.

Chuvieco, E., Allgower, B., Salas, J., 2003. Integration of physical and human factors in fire danger assessment. In: Chuvieco, E. (Ed.), Wildland Fire Danger Estimation and Mapping. The Role of Remote sensing Data. World Scientific, New Jersey, pp. 197–217.

Cochrane, M.A., Schulze, M.D., 1998. Forest fire in the Brazilian Amazon. Conservation Biology 12, 948–950.

Cochrane, M.A., Alencar, A., Schulze, M.D., Souza, C.M., Nepstad, D.C., Lefebvre, P., Davidson, E.A., 1999. Positive feedbacks in the fire dynamic of closed canopy tropical forests. Science 284, 1834–1836.

Crutzen, P.J., Andreae, M.O., 1990. Biomass burning in the tropics: impact on atmospheric chemistry and biogeochemical cycles. Science 250, 1669–1678.

Dennis, R., Mayer, J., Applegate, G.B., Chokkalingam, U., Colfer, C.J.P., Kurniawan, I., Lachowski, H., Maus, P., Permana, R.P., Ruchait, Y., Stolle, F., Suyanto, S., Tomich, T., 2005. Fire, people and pixels: linking remote sensing and social science to understand underlying causes and impacts of fires in Indonesia. Human Ecology 33 (4), 465–504.

Dozier, J., 1981. A method for satellite identification of surface temperature fields of sub pixel resolution. Remote Sensing of Environment 11, 221–229.

Ehrlich, D., Lambin, E., Malingreau, J.-P., 1997. Biomass burning and broad-scale land cover changes in western Africa. Remote Sensing of Environment 61, 201–209.

Eva, H., Lambin, E.F., 2000. Fires and land-cover change in the tropics: a remote sensing analysis at the landscape scale. Journal of Biogeography 27, 765–776.

Eva, H., Steffen, F., 2003. Examining the potential of using remotely sensed fire data to predict areas of rapid forest change in South America. Applied Geography 23 (2–3), 189–204.

Fearnside, P.M., 2005. Deforestation in Brazilian Amazonia: history, rates, and consequences. Conservation Biology 19 (3), 680–688.

Flannigan, M.D., Vonder Haar, T.H., 1986. Forest re monitoring using NOAAsatellite AVHRR. Canadian Journal of Forest Research 16, 975–982.

Frappier, B., Eckert, R.T., 2003. Utilizing the USDA plants database to predict exotic woody plant invasiveness in New Hampshire. Forest Ecology and Management 185, 207–215.

Fraser, R.H., Li, Z., Cihlar, J., 2000. Hotspot and NDVI differencing synergy (HANDS): a new technique for burned area mapping over boreal forest. Remote Sensing of Environment 74, 362–376.

Fujisaka, S., White, D., 1998. Pasture or permanent crops after slash-and-burn cultivation? Land -use choice in three Amazon colonies. Agroforestry systems 42 (1), 45–59.

Giglio, L., 2005. MODIS Collection 4 Active Fire Product User's Guide, Version 2.1. Science Systems and Applications, NASA.

Giglio, L., Csiszar, I., Justice, C.O., 2006. Global distribution and seasonality of active fires as observed with the Terra and Aqua MODIS sensors. Journal of Geophysical Research—Biogeosciences 111, G02016.

Goldammer, J.G. (Ed.), 1990. Fire in the Tropical Biota. Ecosystem Processes and Global Challenges. Ecological Studies, vol. 84. Springer, Berlin.

Gregoire, J.-M., Tansey, K., Silva, J.M.N., 2003. The GBA2000 initiative: developing a global burnt area database from SPOT-VEGETATION imagery. International Journal of Remote Sensing 24 (6), 1369–1376.

Gruell, G.E., 1985. Indian fires in the interior West: a widespread influence. In: Lotan, J.E., Kilgore, B.M., Fischer, W.C., Mutch, R.W., technical coordinators. Proceedings–Symposium and Workshop on Wilderness Fire. USDA Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report, INT-182, pp. 68–74.

Hosmer, D.W., Lemeshow, S., 1989. Applied Logistic Regression. Wiley, New York.

Hough, A.F., Forbes, R.D., 1943. The ecology and silvics of forests in the high Plateaus of Pennsylvania. Ecological Monographs 13 (3), 299–320.

Ichoku, C., Kaufman, Y., Giglio, L., Li, Z., Fraser, R.H., Jin, J.-Z., Park, W.M., 2003. Comparative analysis of daytime fire detection algorithms, using AVHRR data for the 1995 fire season in Canada: perspective for MODIS. International Journal of Remote Sensing 24, 1669–1690.

IFFN, 2002. Fire situation in India, 26th January 2002, pp. 23–27.

Jennessent, 2006. ARCVIEW extensions. 〈http://jennessent.com/arcview/arcview_extensions.htm〉

Justice, C.O., Kendall, J.D., Dowty, P.R., Scholes, R.J., 1996. Satellite remote sensing of fires during the SAFARI campaign using NOAA advanced very high-resolution radiometer data. Journal of Geophysical Research 101, 23851–23863.

Kasischke, E.S., French, N.H.F., Harrell, P., Christensen, N.L., Ustin, S.L., Barry, D., 1993. Monitoring of wildfires in boreal forests using large area AVHRR NDVI composite data. Remote Sensing of Environment 44, 61–71.

Kaufman, Y.J., Tucker, C.J., Fung, I., 1990. Remote sensing of biomass burning in the tropics. Journal of Geophysical Research 95, 9927–9939.

Kaufman, Y.J., Justice, C.O., Flynn, L.P., Kendall, J.D., Prins, E.M., Giglio, L., Ward, D.E., Menzel, W.P., Setzer, A.W., 1998. Potential global fire monitoring for EOS-MODIS. Journal of Geophysical Research 103, 32215–32238.

Kushla, J.D., Ripple, W., 1997. The role of terrain in fire mosaic of a temperate coniferous forest. Forest Ecology and Management 95, 87–107.

Laake, V., Patrick, E., Sánchez-Azofeifa, Arturo, G., 2004. Focus on deforestation: zooming in on hot spots in highly fragmented ecosystems in Costa Rica. Agriculture, Ecosystems & Environment 102 (1), 3–15.

Lancelot, R., Lesnoff, M., McDermott, J.J., 2002. Use of Akaike information criteria for model selection and inference: an application to assess prevention of gastrointestinal parasitism and respiratory mortality of Guinean goats in Kolda, Senegal. Preventative Veterinary Medicine 55 (4), 217–240.

Landmann, T., 2003. Characterizing sub-pixel Landsat ETM+ fire severity on experimental fires in the Kruger National Park, South Africa. South Africa Journal of Science 99, 357–360.

Levine, J.S., 1991. Global Biomass Burning: Atmospheric, Climatic, and Biospheric Implications. MIT Press, Cambridge, MA, 569pp.

Matson, M., Holben, B., 1987. Satellite detection of tropical burning in Brazil. International Journal of Remote Sensing 8, 509–516.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall, London.

Milne, A.K., 1986. The use of remote sensing in mapping and monitoring vegetational change associated with bushfire events in Eastern Australia. Geocarto International 1 (1), 25–34.

Morgan, P., Hardy, C.C., Swetnam, T., Rollins, M.G., Long, L.G., 2001. Mapping fire regimes across time and space: understanding coarse and fine-scale fire patterns. International Journal of Wild land Fire 10, 329–342.

Morisette, J.T., Giglio, L., Csiszar, I., Setzer, A., Schroeder, W., Morton, D., Justice, C.O., 2005. Validation of MODIS active fire detection products derived from two algorithms. Earth Interactions 9, 1–25.

Muirhead, K., Cracknell, A.P., 1985. Straw burning over Great Britain detected by AVHRR. International Journal of Remote Sensing 6 (5), 827–833.

Murdiyarso, D., Lebel, L., Gintings, A.N., Tampubolon, S.M.H., Heil, A., Wasson, M., 2004. Policy responses to complex environmental problems: insights from a science-policy activitiy on transboundary haze from vegetation fires in Southeast Asia. Agriculture, Ecosystems and Environment 104, 47–56.

Narendran, K., 2001. Forest fires. Origins and ecological paradoxes. Resonance, 34–41.

Neary, D.G., Overby, S.T., Gottfried, G.J., Perry, H.M., 1996. Nutrients in fire dominated ecosystems. General Technical Report RM-289. US Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO, pp. 107–117.

Nepstad, D.C., Verissimo, A., Alencar, A., Nobre, C., Lima, E., Lefebvre, P., Schlesinger, P., Potter, C., Mountinho, P., Mendoza, E., Cochrane, M.A., Brooks, V., 1999. Large-scale impoverishment of Amazonian forests by logging and fire. Nature 398, 505–508.

Panwar, H.S., Badola, R., Sinha, B., Silori, C.S., 1993. The wilderness factor: protected areas and people—compatible strategies. Sanctaury Asia 12, 75–85.

Perry, G.L.W., 1998. Current approaches to modeling the spread of wild land fire: a review. Progress in Physical Geography 22 (2), 222–245.

Pfaff, A.S.P., 1999. What drives deforestation in the Brazilian Amazon. Journal of Environmental Economics and Management 37, 26–43.

Pinty, B., Verstraete, M.M., 1991. Extracting information on surface properties from bi-directional reflectance measurements. Journal of Geophysical Research 96, 2865–2874.

Pozo, D., Olmo, F.J., Alados Arboledas, L., 1997. Fire detection and growth monitoring using a multitemporal technique on AVHRR mid-infrared and thermal channels. Remote Sensing of Environment 60 (2), 111–120.

Prasad, K.V., Kant, Y., Gupta, P.K., Sharma, C., Mitra, A.P., Badarinath, K.V.S., 2001. Biomass and combustion characteristics of secondary mixed deciduous forests of India. Atmospheric Environment 35 (18), 3085–3095.

Ramakrishnan, P.S., 1992. Shifting agriculture and sustainable development. MAB Series, UNESCO, Parthenon, 424pp.

Ramnath, M., 1997. Plants, people and management. Natural Resources Forum 21 (4), 257–271.

Ribed, S.P., Lopez, M.A., 1995. Monitoring burnt areas by principle components analysis of multi-temporal TM data. International Journal of Remote Sensing 16 (9), 1577–1587.

Roberts, S.J., 2000. Tropical fire ecology. Progress in Physical Geography 24, 281–288.

Rothermel, R.C., 1991. Predicting behavior and size of crown fires in the Northern Rocky Mountains. General Technical Report INT-438. US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 46pp.

Saha, S., 2002. Anthropogenic fire regime in deciduous forests of central India. Current Science 82, 101–104.

Setiawan, I., Mahmud, A.R., Mansor, S., Mohamed., S., Nuruddin, A.A., 2004. GIS grid-based and multi-criteria analysis for identifying and mapping peat swamp forest fire hazard in Pahang, Malaysia. Disaster Prevention and Management 13 (5), 379–386.

Setzer, A.W., Pereira, M.C., 1991. Amazonian biomass burnings in 1987 and estimate of their tropospheric emissions. Ambio 20, 19–22.

SFR, 2003. State of the Forestry Report. Forest Survey of India, Ministry of Environment and Forests, Dehradun, India.

Siegert, F., Ruecker, G., Hinrichs, A., Hoffmann, A.A., 2001. Increased damage from fires in logged forests during droughts caused by El Nino. Nature 414, 437–440.

Silori, C.S., Mishra, B.K., 2001. Assessment of livestock grazing pressure in and around the elephant corridors in Mudumalai Wildlife Sanctuary, south India. Biodiversity and Conservation 10, 2181–2195.

Skole, D., Tucker, C., 1993. Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. Science 260, 1905–1910.

Smith, A.M.S., Wooster, M.J., 2005. Remote classification of head and backfire types from MODIS fire radiative power observations. International Journal of Wild land Fire 14, 249–254.

Smith, A.M.S., Wooster, M.J., Drake, N., Dipostso, F., Falkowski, M., Hudak, A.T., 2005. Testing the potential of multi-spectral remote sensing for retrospectively estimating fire severity in African Savanna environments. Remote Sensing of Environment 97 (1), 92–115.

Stölle, F., Lambin, E.F., 2003. Interprovincial and inter annual differences in causes of land-use fires in Sumatra, Indonesia. Environmental Conservation 30 (4), 375–387.

Stolle, F., Tomich, T., 1999. The 1997–1998 fire event in Indonesia. Nature and Resources 35 (3), 22–30.

Stolle, F., Chomitz, K.M., Lambin, E.F., Tomich, T.P., 2003. Land use and vegetation fires in Jambi Province, Sumatra, Indonesia. Forest Ecology and Management 179, 277–292.

Stroppiana, D., Grégoire, J.-M., 2002. Using temporal change of the land cover spectral signal to improve burnt area mapping. In: Bruzzone, L., Smith, P. (Eds.), Analysis of Multi-temporal Remote Sensing Images. World Scientific, Singapore, pp. 209–216.

Stroppiana, D., Pinnock, S., Pereira, J.M.C., Gregoire, J.M., 2002. Radiometric analysis of SPOT-VEGETATION images for burnt area detection in Northern Australia. Remote Sensing of Environment 82, 21–37.

Swetnam, T.W., Betancourt, J.L., 1990. Fire-Southern Oscillation relations in the southwestern United States. Science 249, 1017–1021.

Tabachnik, B.G., Fidell, L.S., 2001. Using Multivariate Statistics. 3rd ed. HarperCollins College Publishers, New York.

Tanaka, S., Kimura, H., Suga, Y., 1983. Preparation of a 1:25,000 LANDSAT map for assessment of burnt area on Etajima Island. International Journal of Remote Sensing 4 (1), 17–31.

Tansey, K., 2002. Implementation of regional burnt area algorithms for the GBA2000 initiative. Publication of the European Commission, EUR 20532 EN, 2002.

Tansey, K., Gregoire, J.M., Binaghi, E., Boschetti, L., Brivio, P.A., Ershov, D., Flasse, S., Fraser, R., Graetz, D., Maggi, M., Peduzzi, P., Pereira, J., Silva, J., Sousa, A., Stroppiana, D., 2004. A global inventory of burned areas at 1 km resolution for the year 2000 derived from SPOT vegetation data. Climatic Change 67 (2–3), 345–377.

Tomich, T.P., Lewis, J., 2002. Reducing smoke pollution from tropical forests. ASB Policy brief No. 4.

Turner, M.G., Romme, W.H., 1994. Landscape dynamics in crown fire ecosystems. Landscape Ecology 9, 59–77.

vanderWerf, G.R., Randerson, J.T., Collatz, G.J., Giglio, L., Kasibhatla, P., Arellano, A., Olsen, S., Kasischke, E.S., 2004. Continental-scale partitioning of fire emissions during the 1997 to 2001 El Niño/La Niña period. Science 303, 73–76.

Vazquez, A., Moreno, J.M., 2001. Spatial distribution of forest fires in Sierra de Gredos Central Spain. Forest Ecology and Management 147, 55–65.

Viegas, D.X., 1998. Forest fire propagation. 1998. Philosophical Transactions of the Royal Society of London A 356, 2907–2928.

Weiss, A., 2001. Topographic position and landforms analysis. Poster presentation, ESRI User Conference, San Diego, CA.

Whelan, R.J., 1995. The Ecology of Fire. Cambridge University Press, New York, NY.

Wong, D.W.S., Lee, J., 2005. Statistical Analysis of Geographic Information with Arcview and ArcGIS. Wiley, New York.

Wooster, M.J., Shukiv, B., Oertel, D., 2003. Fire radiative energy for quantitative study of biomass burning: derivation from the BIRD experimental satellite and comparison to MODIS fire products. Remote Sensing of Environment 86 (1), 83–107.

Zucchini, W., 2000. An introduction to model selection. Journal of Mathematical Psychology 44, 41–61.

ELSEVIER

Journal of Environmental Management 86 (2008) 14–26

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Combination of multispectral remote sensing, variable rate technology and environmental modeling for citrus pest management

Qian Du[a], Ni-Bin Chang[b,*], Chenghai Yang[c], Kanth R. Srilakshmi[d]

[a]*Department of Electrical and Computer Engineering, Mississippi State University, MS, USA*
[b]*Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL, USA*
[c]*USDA-ARS, Kika de la Garza Subtropical Agricultural Research Center, Weslaco, TX, USA*
[d]*Department of Environmental Engineering, Texas A&M University-Kingsville, Kingsville, TX, USA*

Received 10 April 2006; received in revised form 7 November 2006; accepted 20 November 2006
Available online 12 January 2007

Abstract

The Lower Rio Grande Valley (LRGV) of south Texas is an agriculturally rich area supporting intensive production of vegetables, fruits, grain sorghum, and cotton. Modern agricultural practices involve the combined use of irrigation with the application of large amounts of agrochemicals to maximize crop yields. Intensive agricultural activities in past decades might have caused potential contamination of soil, surface water, and groundwater due to leaching of pesticides in the vadose zone. In an effort to promote precision farming in citrus production, this paper aims at developing an airborne multispectral technique for identifying tree health problems in a citrus grove that can be combined with variable rate technology (VRT) for required pesticide application and environmental modeling for assessment of pollution prevention. An unsupervised linear unmixing method was applied to classify the image for the grove and quantify the symptom severity for appropriate infection control. The PRZM-3 model was used to estimate environmental impacts that contribute to nonpoint source pollution with and without the use of multispectral remote sensing and VRT. Research findings using site-specific environmental assessment clearly indicate that combination of remote sensing and VRT may result in benefit to the environment by reducing the nonpoint source pollution by 92.15%. Overall, this study demonstrates the potential of precision farming for citrus production in the nexus of industrial ecology and agricultural sustainability.

Keywords: Citrus grove; Pesticide control; Precision farming; Remote sensing; Environmental impact assessment; Industrial ecology; Sustainability

1. Introduction

Agricultural practices are considered the largest contributor to surface water quality degradation in terms of sediment, runoff of nutrients, and leaching of chemicals (Crutchfield et al., 1993). Among the list of environmental damages, agrochemicals (such as pesticides and fertilizers) are suspected to be one of the major contributors to nonpoint source pollution of surface water. The term nonpoint source pollution refers to pollutants that cannot be identified as coming from one discrete location or point. In fact, nonpoint pollution is the primary cause of impairment of fresh water bodies, affecting one-third of the surveyed lake acres, streams, and rivers in the United States (US EPA, 1996). It is also known that nonpoint sources are some of the major sources causing pollution in surface water in Texas (Texas Environmental Profiles, 2005). As the Safe Drinking Water Act and the Clean Water Act play a complementary role in the protection of our water resources in the States, the protection of ground water and the assessment of waters designated for drinking water use are crucial. Because pesticide-contaminated ground water can pose health risks to humans and animals that drink it, a number of government programs have been introduced to directly limit its environmental impact. On the other hand, farmers throughout the world are constantly searching for ways to maximize their returns.

*Corresponding author. Tel.: 407 7547521.
E-mail address: nchang@mail.ucf.edu (N.-B. Chang).

0301-4797/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.019

olatility in the cost of agricultural inputs and the income enerated from farm products leads to instability in the ırm economy. This scenario calls for the introduction of ıodern precision farming technologies to improve crop ield, provide information to enable better in-field manageıent decisions, reduce chemical and fertilizer costs through ıore efficient application, permit more accurate farm ecords, increase profit margin, and reduce environmental ollution (Santhosh et al., 2003). Even though technology as the potential to help alleviate the problem facing future enerations, an integrated approach is needed to promote s use among farmers.

Precision agriculture is a production system that romotes variable management practices within a field ccording to site conditions—such as soil characteristics nd weather conditions—in order to adjust the inputs used nd ultimately achieve optimal output. Precision agriculıre technology is hypothesized to limit the amount of utrient and chemical runoff to the environment for it recisely matches fertilizer and pesticide applications to the eeds of the crop in terms of both quantity and timing. recision farming incorporates several technological tools ıat include variable rate technology (VRT), remote ensing technologies, global positioning systems (GPS) nd geographical information systems (GIS). In general, recision agriculture involves three application processes: athering information inputs such as yield mapping, rocessing the precision information, and prescribing ecommendations for input applications. To collect the ata, farmers could choose a local sensing technique, which ıkes place simultaneously with recommended input pplications. Alternatively, they could use a GPS to collect ıformation related to crop production, including grid soil ımpling, yield monitoring, remote sensing, and crop couting, all of which provide information inputs for ıanagement decisions (Hrubovcak et al., 1999). With such dvantages, precision farming technology is applied in a ariety of agricultural management systems and agriculturl products such as crops, livestock, and forestry, which lso affect the success of environmental management in all spects.

The aim of this study is to demonstrate an integrated ıultidisciplinary approach using an airborne multispectral emote sensing technique, VRT, and environmental impact ssessment to evaluate precision agriculture for citrus roduction. By using a holistic approach, this study articularly attempts to investigate the comparative nvironmental impacts of precision agriculture technology ssociated with variable rate (site specific) and convenonal, single-rate pesticide application. To apply this echnology, site-specific data collected in real time using ocal airborne remote sensing is utilized and the potential nvironmental impact of this VRT technology is assessed ırough simulating a Pesticide Root Zone Model 3 PRZM-3) model for a period of 10 years. Hence, the rimary objectives of this paper are to (1) identify ıe severity of tree health problems with unsupervised classification of airborne remote sensing images; and (2) assess the environmental impact of site-specific chemical application. The study will contribute to the overall goal of improving the environmental sustainability of the citrus industry, both in Texas and elsewhere.

2. Background

2.1. Precision farming, variable rate application, and remote sensing

Managing the land/crop within a field with different levels of input depending upon the yield potential of the crop results in two benefits: (1) the cost of producing the crop in that area can be reduced, and (2) the risk of environmental pollution from pesticides applied at levels greater than those required by the crop can be reduced (Earl et al., 1997). One method of controlling variability within the field is VRT, which allows variable rates of fertiliser application, seeding, chemical application, and tillage throughout a single field. The rate is changed according to a preset map or through information gathered "on the go" by sensors. VRT is most commonly used in conjunction with mapping information such as yield maps or satellite maps, but benefits can still be gained from VRT without such practices. The greatest benefits of VRT can be seen in areas with high variability in soil fertility, weed growth or soil compaction where varying amounts of an input are required throughout a pasture. Instead of applying a single rate of input throughout an entire field, lesser amounts of input can be applied where they are not needed and saved for areas within the field that need greater amounts of input. Savings of over 80% have been reported for some sensor-based selective spray systems (Rob and Troy, 2004). Although VRT can control inputs applied to crops, it cannot control factors such as soil type, weather climate, and topography.

While VRT has become increasingly popular for fertilizer applications in recent years, several complicated issues continue to limit the success of VRT herbicide applications. For example, researchers have not yet developed an appropriate strategy for defining rate scenarios for VRT weed management (Thorp and Tian, 2004). VRT for pesticide application rates depends on variability and complexity of pests such as size, density, and composition. In the case of herbicide application, a rate scenario consists of several parameters, including the number of rate options, the percent by which each rate is reduced from the full rate, the herbicide doses associated with each rate, and the criteria for rate selection. Other criteria, which influence herbicide rate requirements to a lesser degree, include the age of the pest, the level of plant activity at the time of spraying, the pest history, and the desired level of pest control. If defined properly using VRT, however, the ideal rate scenario will provide adequate weed control while maintaining efficient use of herbicide when compared with conventional methods.

The adoption of precision agriculture technologies varies considerably from country to country, and region to region. The adoption rate of variable-rate fertilizer applications and yield monitors is greater than 5% in USA and Canada, and ranges from 1% to 5% in Australia, Brazil, Denmark, United Kingdom, and Germany (Swinton and Lowenberg-Deboer, 2001). In the United Kingdom, however, 15% of the farmers claimed to have used some technologies for precision agriculture (Fountas, 2001). In the United States, the highest rate of adoption was found in corn and soybeans in the Midwest region, while the lowest rate was found along the Southern Seaboard (Daberkow and Mcbride, 2000). By crop type 13.7% of the grains and oil seeds sector used precision technologies, while only 1.6% of the livestock sector had adopted them (Daberkow and Mcbride, 2000). The Northern Great Plains region was ranked second in the US for precision agriculture technology adoption with a rate of 5.8% (Daberkow and Mcbride, 2000). Overall, grid soil sampling is the most commonly used method for variable-rate fertilizer application. Although farms adopting precision agriculture tend to increase remote sensing implementation, remote sensing as well as seed and pesticide variable-rate applications represent less than 1% of the precision agriculture technologies used by all farms (Daberkow and Mcbride, 2000).

Previous implementation of remote sensing for precision agriculture focused on surface temperature measurement and evapotranspiration estimation. The difference between remotely sensed surface temperature and ground-based measurement of air temperature has been established as a method to detect water stress in plants (Jackson et al., 1981). Remote sensing for assessing crop condition is based on the relationship between solar spectral reflectance, temperature of the crop canopy, photosynthesis, and evapotranspiration. To model crop growth and yield one has to combine spectral data with meteorological data, soil data, and other crop parameters (Bauer, 1985). The four main requirements for remote sensing systems for farm management are: frequent coverage, rapid data delivery, fine spatial resolution, and integration with meteorological and agronomic data into expert systems (Jackson, 1984). Remote sensing has shown potential for use in agricultural management for a number of years. For example, powdery mildew has been shown to be detectable with reflectance measurements in the visible portion of the spectrum (Lorenzen and Jensen, 1989). However, the availability of fine spatial resolution, near real-time data has limited its application in the past (Jackson, 1984). New companies that provide aircraft-based imagery to meet the resolution and temporal requirements for agricultural management are now available to provide fine spatial resolution, near real-time data. Sprayer mounted sensors have been found to be useful for the control of herbicide applications (Shearer and Jones, 1991). Remote sensing of crop stress brought on by weeds, diseases, insects, water, frost, and soil temperature was described by Hatfield and Pinter (1993). A vegetation index derived based on vegetation spectral response was proved useful in mapping soil salinity over a sugar cane field (Wiegand et al., 1994). More recently, methods to integrate spectral vegetation indices with temperature have been used to improve remotely-sensed estimates of evapotranspiration (Carlson et al., 1995; Moran et al., 1994). The nitrogen status of crops has also been estimated using remotely sensed data (Blackmer et al., 1995; Filella et al., 1995). Brown and Steckler (1995) developed a method to use digitized color-infrared photographs to classify weeds in a no-till cornfield. The classified data were placed in a GIS, and a decision support system was then used to determine the appropriate herbicide and amount to apply. Penuelas et al. (1995) used reflectance measurements to assess mite effects on apple trees. Previous applications of multispectral remote sensing for farm management can be found in the literature with implications for site-specific cotton management (Barnes et al., 1996) and for the determination of within-field management zones of vegetated fields for application to site-specific farming (Yang and Anderson, 1996). The promise of commercially available, high-resolution satellite imagery will also provide additional sources of remotely sensed data (Fritz, 1996). Retrieval of leaf area index in different vegetation types using a high resolution satellite was established by the inclusion of IKONOS satellite images (Colombo et al., 2003).

2.2. *Environmental impact assessment and pesticide root zone model*

The pesticides which reach the soil or plant material in the target area begin to disappear by degradation or dispersion. Pesticides may volatilize into the air, runoff or leach into surface water and groundwater, be taken up by plants or soil organisms or stay in the soil. The total seasonal losses in runoff for soil-surface applied pesticides average about 2% of the application and rarely exceed 5–10% of the total applied (Leonard, 1990; Schiavon et al., 1995); the fraction removed by leaching is generally less (Taylor and Spencer, 1990; Schiavon et al., 1995). In contrast, volatilization losses of 80–90% have sometimes been measured within a few days after application (Glotfelty et al., 1984; Taylor and Spencer, 1990). Concern about the presence of pesticides in surface water may date back to the 1960s, when residues of chlorinated hydrocarbon insecticides getting into bodies of water were shown to be directly toxic to aquatic organisms (Carson, 1962; Cope, 1965). During the 1970s and 1980s, increasing numbers of pesticides were found in groundwater (US EPA, 1977; Cohen et al., 1984; Leistra and Boesten, 1989; Schiavon et al., 1995), causing great concern, as groundwater is a major source for drinking water in many countries. Worries about the movement of pesticides in the atmosphere arose during the 1970s and 1980s. Transport and redeposition of pesticides may occur over very long distances, as evidenced by the presence of pesticides in ocean fog (Schomburg and Glotfelty, 1991) and arctic snow (Gregor and Gummer, 1989).

In practice, it is extremely difficult and time-consuming to estimate environmental impacts because the impact of pesticides might appear in different environmental media. At a higher level of sophistication, a wide variety of computer models are available that can quantitatively simulate pesticide concentrations in different media. It is important to use a model that has been validated in more than one study, has good user support, requires an amount of data input appropriate for the application, and has a history of producing results acceptable to scientists and regulatory authorities. Considering these various criteria for acceptability, US EPA's Pesticide Root Zone Model (PRZM-3) is an appropriate tool to estimate pesticide concentration in both soil and groundwater (Chang et al., 2006).

In the past, economic and environmental evaluation of alternative pollution-reducing nitrogen management practices has linked farm management with environmental impact (Roderick and Hornbaker, 1999). The ability to detect and map insect damage with remotely sensed imagery implies that methods can be developed to focus pesticide applications in the areas that are most infected, thus decreasing the damage or loss (Barnes et al., 1996). Potentials and limitations of remote sensing data for precision crop management were reviewed by Barnes et al. (1996) and Moran et al. (1997). Although the potential of remote sensing for agriculture has been clearly established, its adoption by farmers remains low. There is still considerable work to be done before the full benefits of remotely sensed data linked with VRT and PRZM-3 can be realized at the present time.

3. Methodology

The Texas citrus industry, an important part of the Lower Rio Grande Valley (LRGV) economy, is heavily dependent on chemical control of pests and diseases. Pest management is essential to fruit yield and economic value, but chemical controls have a high cost, both economically (50% of total production costs) and environmentally. Citrus pest identification depends on human scouting, which is time-consuming and prone to missing of infected trees. Uncertainty regarding tree health results in insurance applications where no problems exist. Multispectral remote sensing can detect trees under stress; there has been no reliable success in determining the source of that stress without human observation. This study, focusing on a citrus farm of LRGV, Texas, emphasizes the synergy between remote sensing, VRT, and environmental impact assessment. Airborne multispectral remote sensing was applied to identify tree health problems, and finally, the PRZM-3 model was simulated to assess the environmental impact of site-specific chemical application for controlling the identified problems.

3.1. Study area

The study location was the South Research Farm of Texas A&M University-Kingsville Citrus Center in Weslaco, Texas. The experimental block is in the form of a square of about 9.2 acres (3.72 ha) and it is part of a large citrus operation of about 250 acres (101.17 ha). Agricultural chemicals are used year-round in this region. Flood irrigation is the most common irrigation method followed in the citrus orchard where simazine is the most intensively used pesticide. Fig. 1 represents the study area showing the citrus groves. The Texas citrus industry is almost totally located in the LRGV, with about 80% of the acreage in Hidalgo County, 15% in Cameron County and 5% in Willacy County. Rio Red grapefruit trees, some of which are heavily infected with Greasy Spot, a common citrus foliar disease in the LRGV caused by the fungus *Mycosphaerella citri Whiteside*, have received wide attention. Greasy Spot reduces tree vigor and thereby fruit size. Infected trees have leaves decaying with yellow and dark spots, and eventually are defoliated. This block may also suffer from some nutrition stresses. Fig. 2 shows the ground pictures of various trees in the citrus grove in the study area.

3.2. Identification of tree health problems with unsupervised classification of airborne images

Multispectral remote sensing data can meet many of the information requirements of site-specific precision farming. Recent advances in precision farming technology (GIS, GPS, and VRT) provide the tools needed to apply information from multispectral images to agricultural management problems. Digital imagery is obtained in distinct areas of the electromagnetic spectrum. Sensors used in vegetation monitoring are typically in the green, red, and near infrared portions of the spectrum (Barnes et al., 1996). As the canopy develops, there is a definite increase in reflectance in the near-infrared (~725–900 nm), as the internal leaf structure of the plant reflects more of the energy in this portion of the spectrum compared to a bare soil. There is also development of a green peak (~550 nm) and decrease in red reflectance (~650–690 nm) due to chlorophyll reflectance and absorption respectively. Thermal imagery (8000–12,000 nm) has also been proven useful in monitoring vegetation, as this imagery can be used to determine surface temperature. Any stress which lessens a crop's transpiration ability will result in a relative increase in the surface temperature of the leaves (Barnes et al., 1996).

Because of the in-field spectral variation due to atmospheric, sensor, and background factors, unsupervised analysis is often used in practical applications where object/material signatures are not required to be known *a priori*. For unsupervised classification, the most famous techniques include K-means clustering and ISODATA (Schowengerdt, 1997). But the resultant classification maps

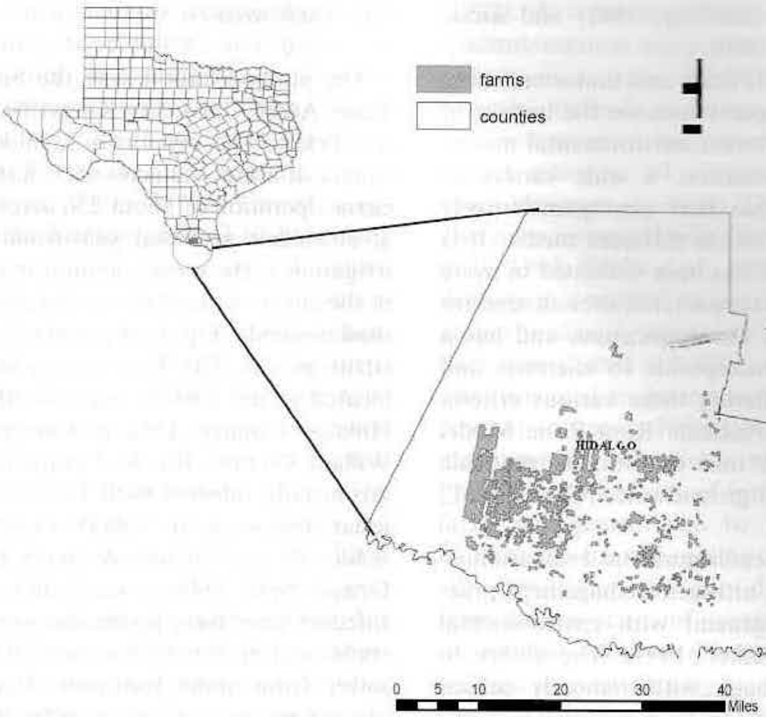


Fig. 1. Map showing the citrus farms in the study area.

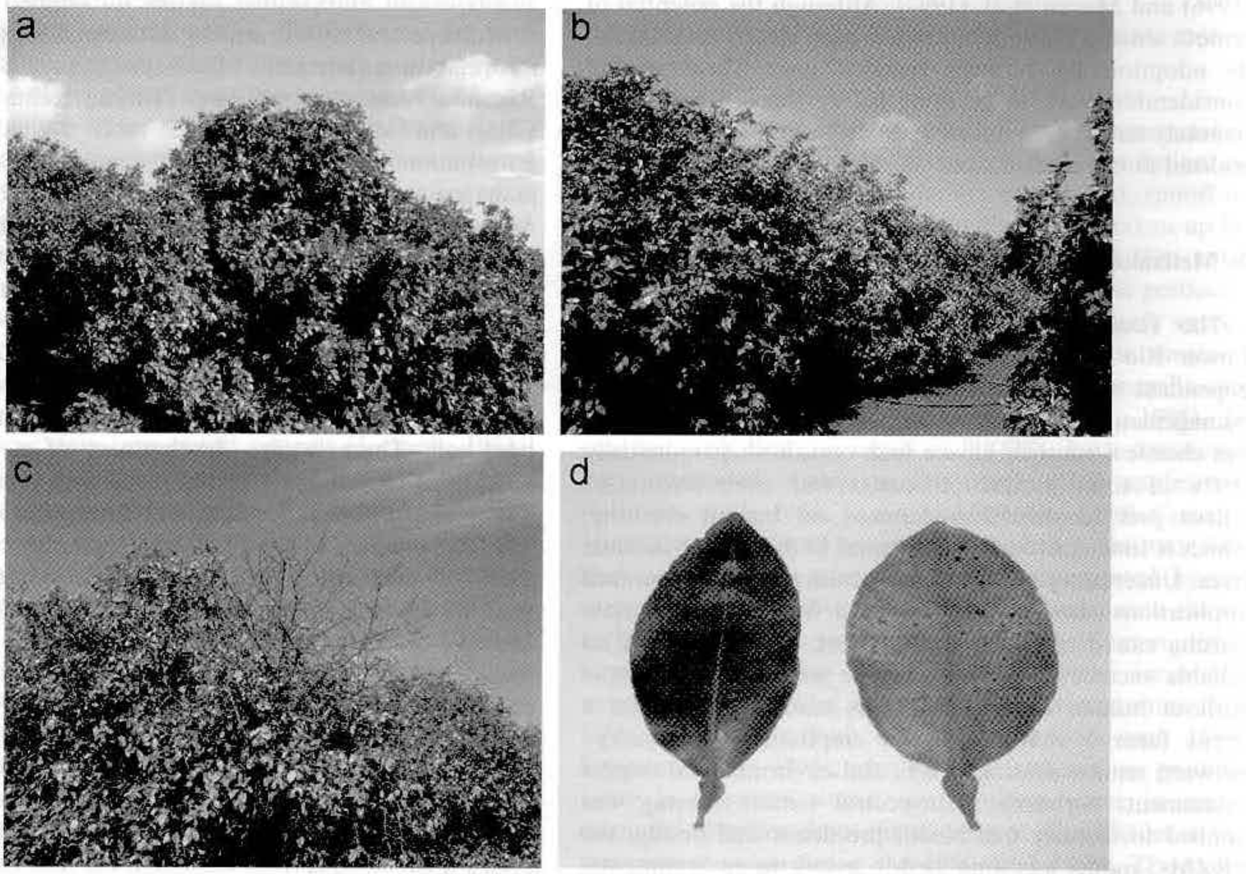


Fig. 2. Ground pictures showing the various trees in citrus farms: (a) vigorous trees, (b) yellowish trees, (c) defoliated trees and (d) healthy leave vs. greasy spot damaged leave.

are binary images. In other words, they employ the "hard" classification, i.e., 0–1 membership assignment. In our research, the classification should quantify the stress severity, which is "soft" classification. Hence, a linear unmixing technique is adopted for this purpose.

Linear unmixing has been widely used for remote sensing image analysis. It models an image pixel vector in a 3D image cube as a linear mixture of a set of finite and distinct objects/materials (i.e., end members) present in an image scene, whose spectral signatures are assumed to be linearly independent. Then classification can be achieved by unmixing each pixel by finding or estimating the respective abundance fractions of the end members resident in the area covered by this pixel (Du et al., 2004). The detailed algorithm can be seen in the Appendix I.

The multispectral imaging system used in this experiment was assembled by the USDA-ARS Kika de la Garza Subtropical Agricultural Research Center at Weslaco, Texas (Yang et al., 2002). The system was composed of three Kodak MegaPlus 1.4i digital charge-coupled device (CCD) cameras with a NIR (845–857 nm) filter, a red (625–635 nm) filter, and a yellow-green (555–565 nm) filter, respectively, and a computer equipped with three image digitizing boards that had the capability of obtaining images with 1024×1024 pixels. The cameras had a built-in analog-to-digital (A/D) converter that produced a digital output signal with 256 gray levels. A Cessna 206 aircraft was used to acquire the imagery at an altitude of 760 m. The ground pixel size achieved was approximately 0.7 m.

The NIR, red, and green band images in the color-infrared composite image were registered to correct the misalignments among them. The co-registered composite image was georeferenced to the Universal Transverse Mercator (UTM), World Geodetic Survey 1984 (WGS-84), Zone 14, coordinate system based on a set of ground control points around the field. For radiometric calibration of the image, four 8 m by 8 m tarpaulins with nominal reflectance values of 4%, 16%, 32% and 48%, respectively, were placed near the field during image acquisition. The georeferenced image was converted to reflectance based on three calibration equations (one for each band) relating reflectance values to the digital count values on the four panels. Image registration and georeferencing were performed using ERDAS IMAGINE (ERDAS, Inc., Atlanta, Georgia, 2004). The longitude-latitude and UTM coordinates of the four corners of the grove are as follows:

- upper left: (97° 57′ 14.25″ W, 26° 08′ 15.68″ N), (604566, 2891353),
- upper right: (97° 57′ 06.80″ W, 26° 08′ 15.88″ N), (604773, 2891360),
- lower left: (97° 57′ 14.10″ W, 26° 08′ 09.88″ N), (604572, 2891174),
- lower right: (97° 57′ 06.61″ W, 26° 08′ 10.05″ N), (604780, 2891181).

Figs. 3a–c show the three bands of the grapefruit block image. As we can see from the image, trees are arranged in a 24×37 matrix (24 rows and 37 columns). The tree at Row 12 (from the top) and Column 18 (from the left) is missing, and this location has a concrete panel (bare soil) for water pump installation. The tree at Row 12 and Column 30 is also missing, and this location is covered by grass. The trees at Row 5 and Column 2, Row 5 and Column 12, Row 10 and Column 5, Row 11 and Column 27 are small and young, so these locations may have more shadow and more exposed grass. This block is surrounded by grassy paths at the northern, southern and western sides. In the southern side, there is a bare soil strip between the trees and the path.

3.3. *Environmental impact assessment and modeling analysis*

The pesticides applied in citrus cultivation have strong potential to transport via different environmental media. Pesticides have been detected in the LRGV water bodies, creating a need to reduce environmental impact. Drift deposits associated with different spraying devices have caused public concern regarding contamination of surface and groundwater systems in local communities. To supplement the uneven and insufficient rainfall, conventional flood irrigation is the primary system used in Texas citrus production while pressurized irrigation systems utilizing pumps to deliver water under low pressure have long been applied. Microsprayer/microsprinkler irrigation systems may result in even higher environmental impact. However, site-specific applications, compared to constant rate application, result in lower levels of residual drift deposits.

The PRZM-3 model was used to assess the effect of the precision-agriculture practices described above on some designated environmental parameters. The model can simulate transformation of pesticides over a long period of time in a wide range of soil, climate, and crop conditions. It is also capable of simulating environmental-related parameters under various management scenarios (Chang et al., 2006). The PRZM-3 model has two major components–hydrology and chemical transport. The hydrologic component for calculating runoff and erosion is based on the Natural Resources Conservation Service curve number technique and the Universal Soil Loss Equation. Evapotranspiration is estimated from pan evaporation data and is classified into evaporation from crop interception, evaporation from soil, and transpiration by the crop (Carlsel et al., 2003). Water movement is simulated by the use of generalized soil parameters, including field capacity, wilting point, and saturation water content. Dissolved, adsorbed, and vapor-phase concentrations in the soil are estimated by simultaneously considering the processes of pesticide uptake by plants, surface runoff, erosion, decay, volatilization, foliar washoff, advection, dispersion, and retardation

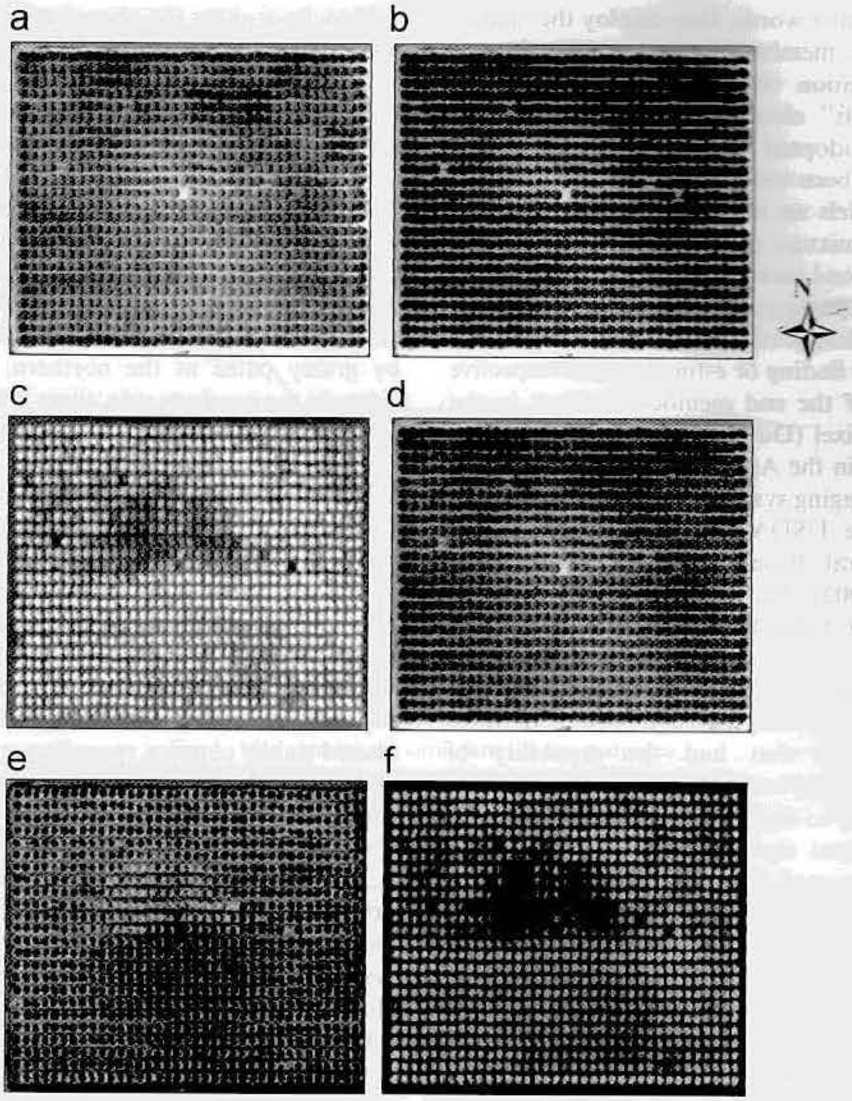


Fig. 3. (a–c) The 3-band multispectral image and (d–f) multispectral image classification using UFCLSLU: (a) Band 1 (yellow–green), (b) Band 2 (red), (c) Band 3 (NIR), (d) Class 1 (bare soil), (e) Class 2 (shadow) and (f) Class 3 (vigorous trees).

in the vadose zone:

$$
\begin{aligned}
&D_{\mathrm{w}}\frac{\partial^2(C_{\mathrm{w}}\theta)}{\partial z^2}+D_{\mathrm{g}}\frac{\partial^2(aC_{\mathrm{g}}K_{\mathrm{H}})}{\partial z^2}-\frac{\partial^2(C_{\mathrm{w}}\theta v)}{\partial z}\\
&-C_{\mathrm{w}}\Big[K_{\mathrm{s}}(\theta+K_{\mathrm{d}}\rho_{\mathrm{b}})+k_{\mathrm{g}}aK_{\mathrm{H}}\\
&+f\theta\varepsilon+\frac{Q}{A_{\mathrm{w}}\Delta z}+\frac{PX_{\mathrm{e}}r_{\mathrm{om}}K_{\mathrm{d}}}{A_{\mathrm{w}}\Delta z}\Big]\\
&+\frac{J_{\mathrm{APP}}}{A\Delta z}+\frac{EP_rM}{\Delta z}-K_{\mathrm{TRN}}C_{\mathrm{w}}\theta+\sum_{\mathrm{K}}K_{\mathrm{TRN}}C_{\mathrm{w}}\theta,\\
&=\frac{\partial^2[C_{\mathrm{w}}(\theta+K_{\mathrm{d}}\rho_{\mathrm{b}}+aK_{\mathrm{H}})]}{\partial t},
\end{aligned}
\tag{1}
$$

where D_{w} is the diffusion-dispersion coefficient for the dissolved phase, assumed constant ($\mathrm{L^2\,T^{-1}}$), t the time (T), D_{g} the molecular diffusivity of the pesticide in the air-filled pore space ($\mathrm{L^2\,T^{-1}}$), a the volumetric air content of the soil ($\mathrm{L^3\,L^{-3}}$), C_{g} the gaseous concentration of the pesticide ($\mathrm{M\,L^{-3}}$), K_{H} the distribution coefficient between the liquid and vapor phases (Henry's constant), z the vertical space dimension (L), Δz the space increment (L), C_{w} the dissolved concentration of the pesticide($\mathrm{M\,L^{-3}}$), θ the volumetric soil-water content ($\mathrm{L^3\,L^{-3}}$), v the velocity of the soil water ($\mathrm{L\,T^{-1}}$), K_{s} the lumped, first-order decay constant for the solid and dissolved phases ($\mathrm{T^{-1}}$), K_{d} the partition coefficient between the dissolved and solid phases ($\mathrm{T^{-1}}$), ρ_{b} the soil bulk density ($\mathrm{M\,L^{-3}}$), J_{APP} the mass gain due to pesticide application at or near the surface ($\mathrm{M\,T^{-1}}$), X_{e} the erosion sediment loss ($\mathrm{M\,T^{-1}}$), p a units conversion factor ($\mathrm{M\,M^{-1}}$), A the cross-sectional area of the soil column, A_{w} the watershed area ($\mathrm{M^2}$), Q the daily runoff volume ($\mathrm{L^3\,T^{-1}}$), P_{r} the daily rainfall depth ($\mathrm{L\,T^{-1}}$), ε an uptake efficiency factor or reflectance coefficient (dimensionless), f the fraction of total water in the zone used for transpiration (T^{-1}), K_{TRN} the transformation rate constant (T^{-1}) and r_{om} the enrichment ratio for organic matter ($M M^{-1}$).The above equation can be solved easily in the PRZM-3 model for surface layers with an assumption of $f\theta = 0$.

To estimate environmental impacts, we compared the agricultural practice consisting of a single pesticide application rate and the site-specific rate based on VRT applied on citrus crops where the stress is more as prescribed by remote sensing multispectral images. The single application rate refers to an unvaried pesticide application rate on crops, regardless of the variability of soil characteristics within the field. Information about single-rate pesticide application in this study was obtained from the citrus farm operations at the citrus center, where this process has been applied for years. For a sound modeling process, simazine pesticide runoff, erosion, and volatilization losses are the primary environmental parameters of interest because these are the primary factors contributing to nonpoint source pollution.

4. Results and discussion

4.1. Remote sensing analysis

The unsupervised fully constrained linear unmixing results (gray-scale images of a citrus field) are shown in Figs. 3d–f where three classes were classified: bare soil (including the concrete panel), shadow, and vigorous trees. The last class is particularly important in the assessment of pest stress. The brightness of a pixel in this gray scale classification map reveals the healthy status of a tree. A bright pixel means a very healthy tree, a gray shaded pixel indicates the tree may have a certain level of stress, and a dark pixel indicates the tree is defoliated and the stress is very severe. Therefore, the pixel gray levels in this class allow the quantification of stress severity. In the unmixing model, three materials (soil, shadow, and trees) were used. The brightness of a pixel for the tree material represents the fractional abundance of the tree material in the pixel (Adams et al., 1995). The abundance value for each pixel ranges from 0 (darkest) to 1 (brightest). Therefore, a bright pixel means a very healthy tree (Adams et al., 1995). This observation can be verified through a ground truthing process.

The Canny edge detection method was applied to Class 3 in Fig. 3 for the detection of tree size. The tree centroids were estimated by finding the local maxima within the crown boundaries. As shown in Fig. 4, the bright spots represent the tree centroids. Those trees whose centroids are not surrounded by boundaries are severely defoliated. Multispectral images used in the analysis can then be translated directly to maps of pesticide application rates. This may provide a basis for environmental modeling analysis.

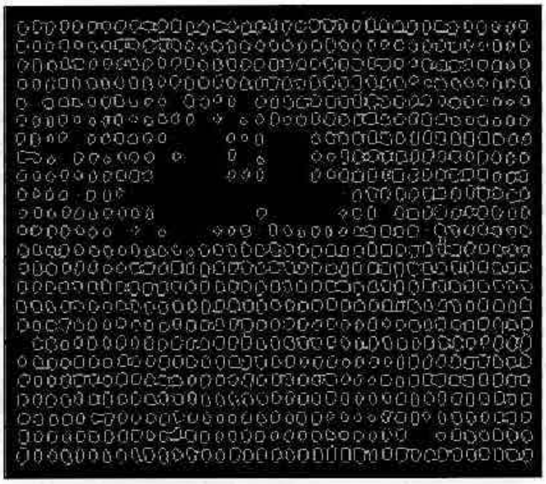


Fig. 4. Detected tree centroid and crown boundary.

4.2. PRZM modeling analysis

4.2.1. Input parameters for the model

Meteorological data (1981–1990) required for the model simulation, such as temperatures, precipitation, evapotranspiration, etc. were obtained from the US EPA's Brownsville weather station located in the Rio Grande Valley. The reason behind using this data set is that these data files are compatible with the PRZM software model and the current version of this data set is not available. The rainfall events as shown in Fig. 5 measured in the LRGV area can be generally classified into three different types, fine rain (i.e., total accumulated rainfall is less than 30 mm and the maximum rainfall intensity is less than 10 mm/h), heavy rain (i.e., total accumulated rainfall is between 30 and 100 mm, and the maximum rainfall intensity is higher than 10 mm/h), and torrential rain (i.e., total accumulated rainfall is higher than 100 mm). The rainfall intensity during 1981–1990 in the LRGV is shown in Fig. 5. According to the meteorological data during the time period between 1981 and 1990, there were 716 rainfall events recorded and all the events were fine rain events. However, variations in annual rainfall create different environmental impacts.

STATSGO is the State Soil Geographic Data Base developed by the United States Department of Agriculture–Natural Resource Conservation Service (USDA-NRCS). USGS STATSGO soil data and GIS were used to map the soil type in the citrus orchard. Fig. 6 shows the map of the orchards and soil type in the LRGV. Approximately 9% of the total land area in the LRGV is covered by sandy clay loam (SCL). Most of the soil in the citrus orchards is SCL. Based on the soil type, different input parameters for the model, including soil permeability (i.e., hydraulic conductivity), were obtained from the STATSGO soil database (USDA NRCS, 2004), and the organic carbon partitioning coefficient was obtained from the PRZM-3 user manual (Carlsel et al., 2003).

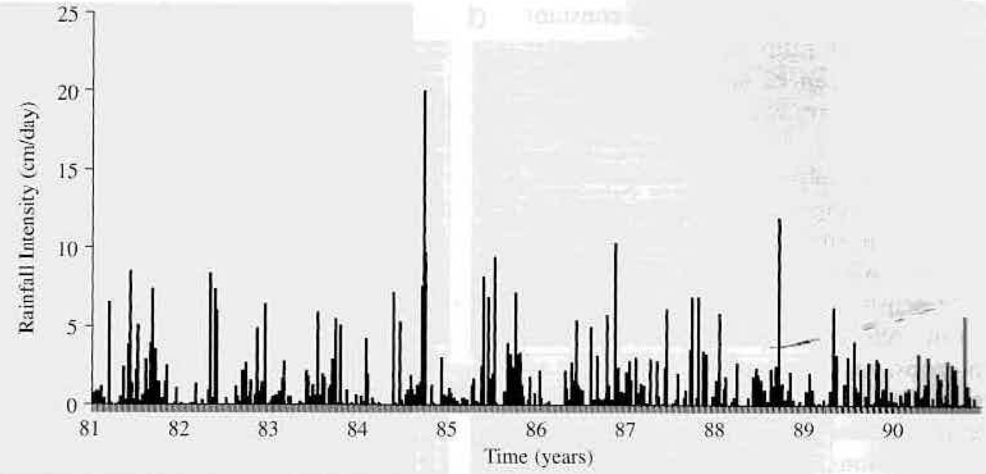


Fig. 5. Rainfall intensity during 1981–1990 in Lower Rio Grande Valley.

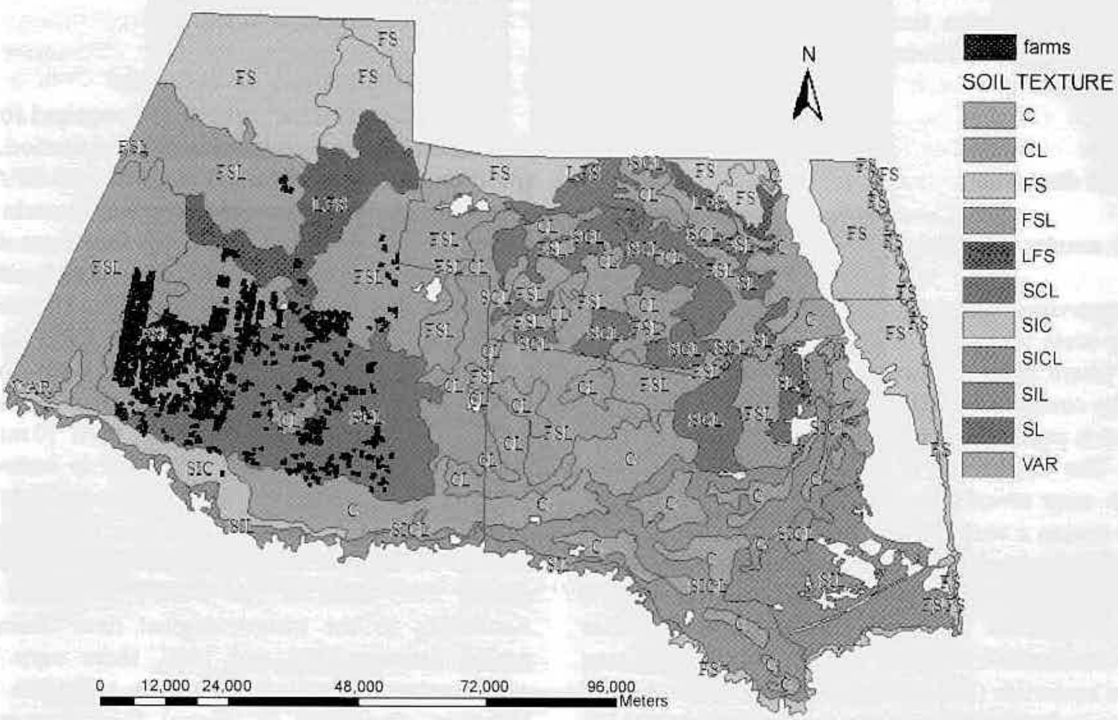


Fig. 6. Map showing the farms and soil texture in the LRGV.

4.2.2. Scenario analysis

In the present study, two scenarios are considered with respect to the exclusion and inclusion of remote sensing and VRT technology. In the former case, the PRZM model is applied to the entire farm area to determine the amount of simazine loading lost in runoff, soil erosion, and in volatilization. In the latter case, the PRZM-3 model is applied to the defoliated trees where the stress is very severe in the farm area to determine the amount of simazine lost in runoff, soil erosion and in volatilization. In both cases, however, the simazine runoff, erosion and volatilization loadings were calculated and compared using the PRZM-3 model. The outcomes are described below. In general, loss from volatilization is insignificant (Extension Toxicology Network, 1996).

4.2.2.1. Scenario 1: model simulation without remote sensing and variable rate application. Simazine runoff values obtained from the PRZM-3 model during the simulation period were averaged and a graph was drawn to show the variation of the simazine loss in the study region (see Fig. 7a). It can be inferred from the graph that only a fraction of a percent of the annual pesticide application was lost in the runoff. The maximum amount of simazine was lost in runoff in the year 1984 because of the high rainfall intensity in that year. Fig. 7b shows the annual simazine erosion

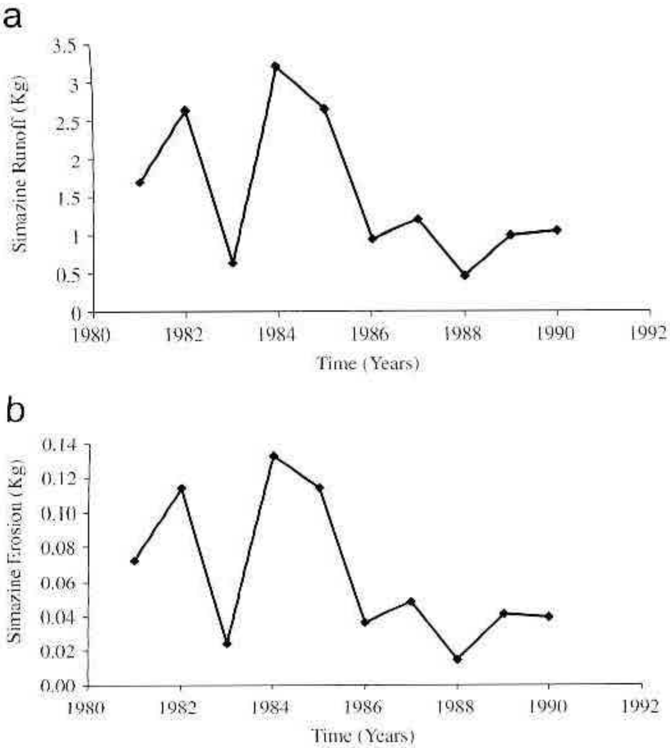


Fig. 7. Scenario 1: (a) annual simazine runoff loading, (b) annual simazine volatilization loading.

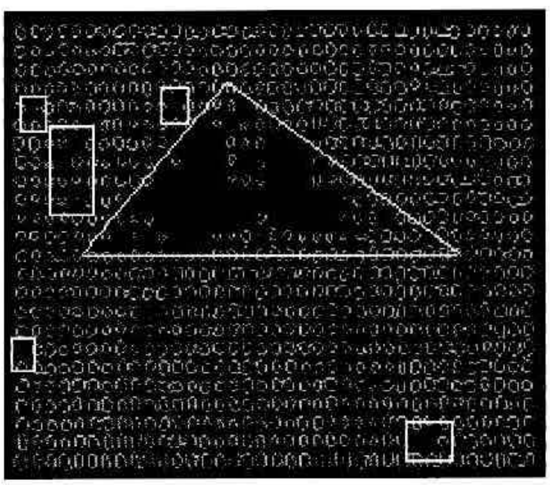


Fig. 8. Multispectral image showing the defoliated trees (severe stress in the farm).

loading values obtained from the PRZM-3 model during the simulation period. It can be observed from the graph that only a fraction of a percent of the annual pesticide application was lost due to erosion. The maximum amount of simazine loss due to erosion occurred during the year 1984 and this is because of the high rainfall intensity which caused the soil containing simazine to erode.

4.2.2.2. Scenario 2: model simulation with remote sensing and variable rate application. Fig. 8 shows the multispectral image of defoliated trees where the stress is very severe in the farm and these are highlighted with white line boundaries. In this case, the area under stress is calculated by dividing the whole block into small cells and thereafter counting the number of cells in the affected region. The total block was divided into 888 cells (which is equal to number of trees in the block), each measuring an area of 0.004193 ha (tree area plus area between two trees). There are a total of 70 cells in the stressed area and the total area of the affected region is approximately 0.293488 ha. The obtained stressed area is given as input to the model to calculate the simazine lost in runoff, erosion, and volatilization.

Simazine runoff loading values obtained from the PRZM-3 model during the simulation period were averaged and a graph was drawn to show the variation of the simazine runoff loss in the study region. The maximum loss occurred during 1984 because of high rainfall during that period (see Fig. 9a). The loss is very small when compared to simazine runoff loss before remote sensing and VRT were used. Fig. 9b shows the annual simazine pesticide erosion values obtained from the PRZM-3 model during the simulation period. It can be observed from the graph that a fraction of a percent of the annual pesticide application was lost due to erosion. The loss is very small when compared to simazine erosion loss in the case without the inclusion of remote sensing and VRT.

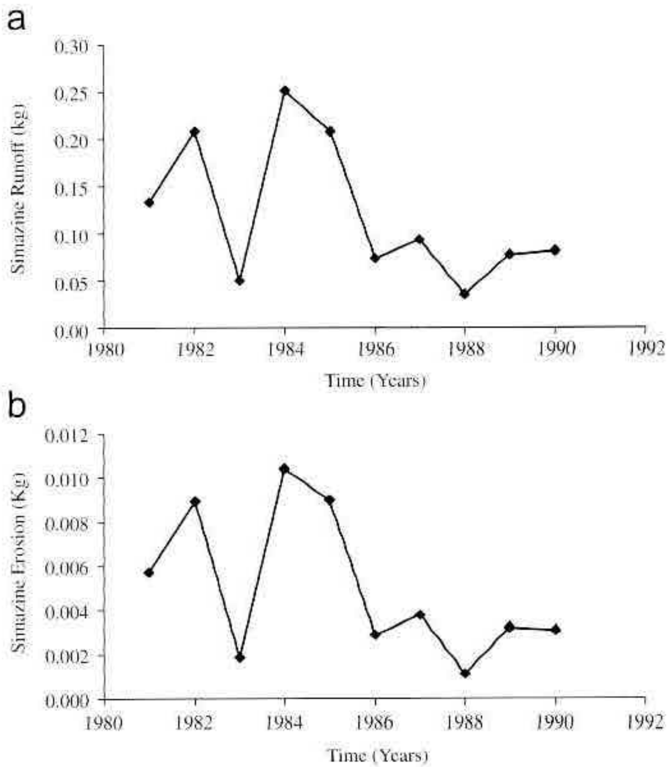


Fig. 9. Scenario 2: (a) annual simazine runoff, (b) annual simazine volatilization.

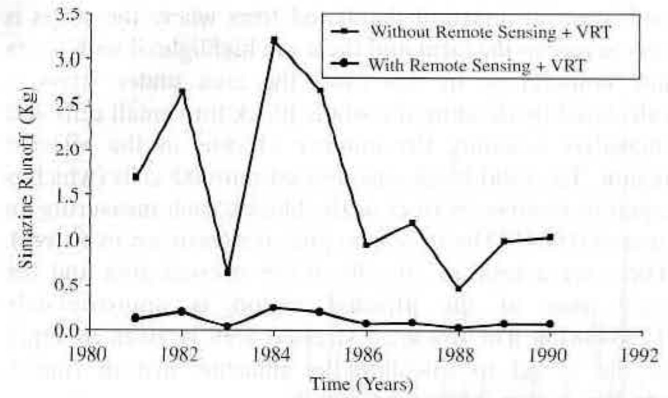


Fig. 10. Comparison of surface runoff between scenarios 1 and 2.

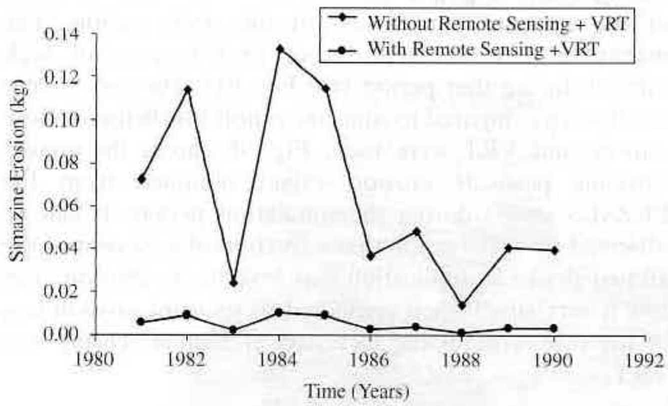


Fig. 11. Comparison of erosion between scenarios 1 and 2.

4.2.3. Comparative study

Environmental impacts associated with two types of agricultural practices (scenarios) deserve further discussion. While the one with a single pesticide application rate on citrus crops is a traditional approach, the site-specific application with precision farming skill where the stress is more as prescribed by remote sensing multispectral images shows a cost-effective and risk-informed advantage. Fig. 10 shows the comparison of simazine surface runoff loss between scenarios 1 and 2 (i.e., before and after the site-specific application). Site-specific application of pesticide resulted in a 92.15% decrease in the runoff value during the simulation period. Fig. 11 shows the comparison of simazine erosion loss between scenarios 1 and 2, and site-specific application of pesticide resulted in a 92.15% decrease in the erosion loss value during the simulation period as well.

5. Conclusion

This analysis demonstrates that precision agriculture is a promising, cost-effective, and environmentally benign practice based on the case study with a citrus grove. The case study in south Texas provides an example of how remotely sensed data and environmental models can be utilized in site-specific agricultural management for pest control and impact assessment. It quantitatively shows the advantage of using an integrated approach combining remote sensing, VRT, and environmental modeling for the characterization of environmental benefit in precision farming. In general, multispectral images were used for anomaly detection; however, anomaly detection does not provide quantitative recommendations that can be directly applied to precision farming. The spectral linear unmixing based approach in the application of remote sensing and site-specific agriculture management was proved useful for quantifying stress severity and detecting early infection. When effectively applied, site-specific pesticide applications can provide environmental benefits by reducing the environmental impacts of surface runoff, erosion, and volatilization. Multispectral images can be translated directly to maps of pesticide application rates. The PRZM-3 modeling outputs indicated a positive decrease in environmental impacts. From an environmental point of view, applying site-specific applications resulted in the maximum benefit to the environment, reducing surface runoff loss and erosion loss by 92.15%. It may become a standard practice in the citrus industry in the future. Future work may also focus on the recognition of the degree of stress/damage by using hyperspectral remote sensing techniques.

Acknowledgment

The authors are grateful to the Texas Higher Education Coordinating Board (THECB) for funding this project.

Appendix I. The algorithm of the linear unmixing classification technique

Let L be the number of spectral bands and r an $L \times 1$ column pixel vector in a multispectral or hyperspectral image. Assume that there are p objects (materials) present in an image scene, which construct an $L \times p$ signature matrix $\mathbf{M} = [\mathbf{m}_1 \mathbf{m}_2 \cdots \mathbf{m}_p]$, where $\mathbf{m}_j$ represents the jth object. Assume that $\boldsymbol{\alpha} = (\alpha_1 \alpha_2 \cdots \alpha_p)^T$ is a $p \times 1$ abundance vector associated with r, where α_j denotes the abundance fraction of the $\mathbf{m}_j$ in r, which is unknown. In the linear mixture model, r is considered as the linear mixture of $m_1, m_2, \ldots, m_p$ as $\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n}$, where $\mathbf{n}$ is included to account for either a measurement or model error. A general approach to solving or estimating the unknown mixing coefficients, $\alpha_1, \alpha_2, \ldots, \alpha_p$, is linear spectral mixture analysis (LSMA), where a fractional abundance image is generated by the LSMA for each of the mixing coefficients. In other words, a fractional abundance image is a gray scale image with gray scales representing abundance fractions of a mixing substance that is specified by a particular object present in each image pixel vector. The mixed pixel r is then classified according to the set of p fractional abundance images generated by the LSMA that correspond to $\alpha_1, \alpha_2, \ldots, \alpha_p$. In addition, if the estimated abundance vector, $\boldsymbol{\alpha}$, can faithfully represent image pixel vector r, the loss of information will be immaterial and have little impact on image analysis. In order to accomplish this goal, two constraints are imposed on $\boldsymbol{\alpha}$: (a) abundance sum-to-one constraint, referred to as the ASC, $\sum_{j=1}^{p} \alpha_j = 1$; and (b) abundance nonnegativity constraint, referred to as the ANC, $\alpha_j \geqslant 0$ for all $1 \leqslant j \leqslant p$. Because there are generally no closed-form solutions to linear mixing problems imposing both constraints, we must rely on numerical algorithms to generate optimal solutions. A fully constrained least squares linear unmixing (FCLSLU) method can be used to find the estimate $\hat{\boldsymbol{\alpha}}$, which minimizes the least square estimation error $\mathrm{LSE} = (\mathbf{r} - \mathbf{M}\hat{\boldsymbol{\alpha}})^T(\mathbf{r} - \mathbf{M}\hat{\boldsymbol{\alpha}})$ and satisfies ASC and ANC simultaneously. It should be noted that there is a data dimensionality limitation for LSMA, i.e., the number of objects/materials to be unmixed cannot be larger than the number of spectral bands. In addition, the FCLSLU requires a complete knowledge of the signature matrix $\mathbf{M}$. In order for it to be applied to a situation where no *a priori* information about $\mathbf{M}$ is available, an unsupervised process, referred to as UFCLSLU, is applied to generate the $\mathbf{M}$ from the data (Du et al., 2004).

References

Adams, J.B., Sabol, D.E., Kapos, V., Filho, R.A., Roberts, D.A., Smith, M.O., Gillespie, A.R., 1995. Classification of multispectral images based on fractions of endmembers: application to land-cover change in the Brazilian Amazon. Remote Sensing of Environment 52, 137–154.

Barnes, E.M., Moran, M.S., Pinter Jr., P.J., Clarke, T.R., 1996. Multispectral remote sensing and site-specific agriculture: examples of current technology and future possibilities. In: Proceedings of the Third International Conference on Precision Agriculture, 23–26 June 1996, Minneapolis, Minnesota, pp. 843–854.

Bauer, M.E., 1985. Spectral inputs to crop identification and condition assessment. Proceedings of the IEEE 73, 1071–1085.

Blackmer, T.M., Schepers, J.S., Meyer, G.E., 1995. Remote sensing to detect nitrogen deficiency in corn. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Proceedings of Site-Specific Management for Agricultural Systems, Minneapolis, MN, 27–30 March 1994. ASA-CSSA-SSSA, Madison, WI, pp. 505–512.

Brown, R.B., Steckler, J.P., 1995. Prescription maps for spatially variable herbicide application in no-till corn. Transactions of ASAE 38, 1659–1666.

Carlson, T.N., Capehart, W.J., Gillies, R.R., 1995. A new look at the simplified method for remote sensing of daily evapotranspiration. Remote Sensing of Environment 54, 161–167.

Carlsel, R.F., Smith, C.N., Mulkey, L.A., Dean, J.D., Jowis, P., 2003. User Manual for the Pesticide Root Zone Model (PRZM) Release I. EPA-600/3-84-109. US EPA, Athens, GA.

Carson, R.L., 1962. Silent Spring. Riverside Press, Cambridge, MA, USA.

Chang, N.B., Srilakshmi, K.R., Parvathinathan, G., 2006. Comparison of models of simazine transport and fate in subsurface environment in a citrus farm. Journal of Environmental Management, in press.

Colombo, R., Bellingeri, D., Fasolini, D., Marino, C.M., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote Sensing of Environment 86 (1), 30, (120–131).

Cohen, S.Z., Creeger, S.M., Carsel, R.F., Enfield, C.G., 1984. Potential for pesticide contamination of groundwater resulting from agricultural uses. In: Kruger, R.F., Seiber, J.N. (Eds.), Treatment and Disposal of Pesticide Wastes. ACS Symposium Series No. 259. American Chemical Society, Washington, DC, pp. 297–325.

Cope, O.B., 1965. Agricultural chemicals and freshwater ecological systems. In: Chichester, C. (Ed.), Research in Pesticides. Academic Press, New York, pp. 115–128.

Crutchfield, S.R., Hansen, L.T., Ribaudo, M.O., 1993. Agricultural and Water Quality Conflicts: Economic Dimensions of the Problem. AIB-676. USDA, Economic Research Service, Washington, DC.

Daberkow, S.G., Mcbride, W.D., 2000. Adoption of precision agriculture technologies by US farmers. In: Proceedings of the Fifth International Conference on Precision Agriculture, Madison, WI, USA, ASA-CSSA-SSSA.

Du, Q., French, V., Skaria, M., Yang C.H., Everitt, J.H., 2004. Citrus pest stress monitoring using airborne hyperspectral imagery. In: Proceedings of 2004 International Geoscience and Remote Sensing Symposium, Anchorage, AK.

ERDAS, Inc., 2004. 2801 Buford Highway, Atlanta, Georgia 30329 USA. Web ⟨www.erdas.com⟩.

Earl, E.C., Wheeler, M.G., Blackmore, J.J., 1997. Precision farming-the management of variability. The Journal of the Institution of Agricultural Engineers 4, 18–23.

Extension Toxicology Network, 1996 ⟨http://extoxnet.orst.edu/⟩, accessed in June, 2004.

Fountas, S., 2001. Farmers' attitude to precision farming. In: Proceedings of the Third European Conference on Precision Agriculture, ENSA, Montpellier, France, pp. 515–519.

Fritz, L.W., 1996. The era of commercial earth observation satellites. Photogrammetric Engineering and Remote Sensing 62, 39–45.

Filella, I., Serrano, L., Serra, J., Penuelas, J., 1995. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. Crop Science 35, 1400–1405.

Glotfelty, D.E., Taylor, A.W., Turner, B.C., Zoller, W.H., 1984. Volatilization of surface-applied pesticides from fallow soil. Journal of Agricultural and Food Chemistry 32, 634–638.

Gregor, D.J., Gummer, W.D., 1989. Evidence of atmospheric transport and deposition of organochlorine pesticides and polychlorinated biphenyls in Canadian arctic snow. Environmental Science and Technology 23, 561–565.

Hatfield, J.L., Pinter Jr., P.J., 1993. Remote sensing for crop protection. Crop Protection 12, 403–414.

Hrubovcak, J., Vasavada, U., Aldy, J.E., 1999. Green technologies for a more sustainable agriculture. Agriculture information bulletin, no. 752, Economic Research Service, US Department of Agriculture, Washington, DC.

Jackson, R.D., 1984. Remote Sensing of vegetation characteristics for farm management. Proceedings of the Society of Photo-Optical Instrumentation Engineers 475, 81–96.

Jackson, R.D., Idso, S.B., Reginato, R.J., Pinter Jr., P.J., 1981. Crop temperature as a crop water stress indicator. Water Resources Research 17, 1133–1138.

Leistra, M., Boesten, J.J.T.I., 1989. Pesticide contamination of groundwater in western Europe. Agriculture Ecosystems and Environent 26, 369–389.

Leonard, R.A., 1990. Movement of pesticides into surface waters. In: Pesticides in the Soil Environment. Book Series No. 2. Soil Science Society of America, Madison, WI, pp. 303–349.

Lorenzen, B., Jensen, A., 1989. Changes in leaf spectral properties induced in barley by cereal powdery mildew. Remote Sensing of Environment 27, 201–209.

Moran, S.M., Clarke, T.R., Inoue, Y., Vidal, A., 1994. Estimating crop water deficit using the relationship between surface-air temperature and spectral vegetation index. Remote Sensing of Environment 49, 246–263.

Moran, M.S., Inoue, Y., Barnes, E.M., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. Remote Sensing of Environment 61, 319–346.

Penuelas, J., Filella, I., Lloret, P., Munoz, F., Vilajeliu, M., 1995. Reflectance assessment of mite effects on apple trees. International Journal of Remote Sensing 16, 2727–2733.

Rob, K., Troy, J., 2004. Precision farming in the northern grains region using variable rate technology (VRT) in cropping land. File no. FS0609, Queensland Government.

Roderick, M.R., Hornbaker, R.H., 1999. Economic and environmental evaluation of alternative pollution-reducing nitrogen management practices in central Illinois. Agriculture, Ecosystems and Environment 75 (1–2), 41–53.

Shearer, S.A., Jones, P.T., 1991. Selective application of post-emergence herbicides using photoelectrics. Transactions of ASAE 34, 1661–1666.

Schowengerdt, R.A., 1997. Remote Sensing Models and Methods for Image Processing, second ed. Academic Press, San Diego, CA, 522pp.

Santhosh, K.S., Laguette, S., Casady, G.M., Seielstad, G.A., 2003. Remote sensing applications for precision agriculture: a learning community approach. Remote Sensing of Environment 88, 157–169.

Schiavon, M., Perrin-Ganier, C., Portal, J.M., 1995. La pollution de l'eau par les produits phytosanitaires: état et origine. Agronomie 15, 157–170.

Schomburg, C.J., Glotfelty, D.E., 1991. Pesticide occurrence and distribution in fog collected near Monterey, California. Environmental Science and Technology 25, 155–160.

Swinton, S.M., Lowenberg-Deboer, J., 2001. Global adoption of precision agriculture technologies: who, when, why? In: Proceedings of the Third European Conference on Precision Agriculture ENSA-Montpellier, pp. 557–562.

Taylor, A.W., Spencer, W.F., 1990. Volatilization and vapor transport processes. In: Pesticides in the Soil Environment. Book Series No. 2. Soil Science Society of America, Madison, WI, pp. 213–269.

Texas Environmental Profiles, 2005. 〈http://www.texasep.org/html/wql/wql_2sfc.html〉.

Thorp, K.R., Tian, L.F., 2004. Performance study of variable-rate herbicide applications based on remote sensing imagery. Biosystems Engineering 88 (1), 35–47.

USDA Natural Resources Conservation Service, National Soil Survey Center, Soil Survey Laboratory. Available online at 〈http://vmhost.cdp.state.ne.us:96〉, (accessed August 2004).

US Environmental Protection Agency, 1977. Waste disposal practices and their effects on groundwater Report to Congress. USEPA, Washington, DC.

US Environmental Protection Agency, 1996. National Water Quality Inventory: 1996 Report to Congress 〈http://www.epa.gov/305b/96report/〉, accessed in October 2004.

Wiegand, C.L., Escobar, D.E., Lingle, S.E., 1994. Detecting growth variation and salt stress in sugarcane using videography. In: Proceedings of the 14th Biennial Workshop on Color Aerial Photography and Videography for Resource Monitoring. American Society for Photogrammetry and Remote Sensing, pp. 185–199.

Yang, C., Anderson, G.L., 1996. Determining within-field management zones for grain sorghum using aerial videography. In: Proceedings of the 26th Symposium on Remote Sensing of Environment, 25–29 March 1996, Vancouver, BC.

Yang, C., Everitt, J.H., Bradford, J.M., 2002. Optimum time lag determination for yield monitoring with remotely sensed imagery. Transactions of the American Society of Agricultural Engineers 45 (6), 1737–1745.

ELSEVIER

Journal of Environmental Management 86 (2008) 27–43

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Comparison of models of simazine transport and fate in the subsurface environment in a citrus farm

Ni-Bin Chang[a,*], K.R. Srilakshmi[b], G. Parvathinathan[b]

[a]Department of Civil and Environmental Engineering, University of Central Florida, Orlando FL 32816, USA
[b]Department of Environmental Engineering, Texas A&M University—Kingsville, Kingsville TX 78363, USA

Received 15 April 2006; received in revised form 7 November 2006; accepted 20 November 2006
Available online 22 January 2007

bstract

Contamination of groundwater by agrochemicals is now widely recognized as an extremely important environmental problem. Iodern agricultural practices involve the combined use of irrigation with the application of large amounts of agrochemicals to maximize ·op yield. Due to flood irrigation and natural runoff, agricultural activities might generate soil, surface water and groundwater ontamination problems and leaching of pesticides. Modeling of the transport and fate of pesticides, such as simazine, may help nderstand the long-term potential risk to the subsurface environment. This paper illustrates a comparative study via the use of three ifferent pesticide transport simulation models and the applicability of those models in determining the groundwater vulnerability to esticides contamination in a citrus orchard located at the Lower Rio Grande Valley (LRGV). The three models used in the study are the esticide root zone model-3 (PRZM-3), the pesticide analytical model (PESTAN) and integrated pesticide transport modeling (IPTM). he concentration values obtained from all three models are in agreement, and they show a decreasing trend from the surface through ie vadose zone. The problem is how to use this information and, specifically, how to combine the testimony of a number of experts into single useful judgment. With the aid of the fuzzy multiattribute decision making method, PRZM-3 is deemed as the most promising one or such precision farming applications.

eywords: Pesticides; Groundwater modeling; Model selection; Fuzzy multicriteria decision making; Environmental impact assessment; Citrus grove

. Introduction

Contamination of groundwater by agrochemicals (pescides and fertilizers) is now widely recognized as an xtremely important environmental problem. Pesticides pplied at or near the soil surface can leach to considerable epths (Loague et al., 1998). The assessment and remediaon of non-point source groundwater contamination can asily pose problems that have significantly greater conomic impact than those that have long been recogized for point sources. The commercial Texas citrus roduction is almost totally located in the Lower Rio Grande Valley (LRGV) along the US–Mexico Border, with bout 80% of the acreage in Hidalgo County, 15% in Cameron County and 5% in Willacy County. Flood irrigation is common in citrus orchards in the LRGV, which has certain disadvantages. Such a method may result in excessive loss of water due to evaporation and percolation into the ground, and leaching of fertilizers and pesticides.

Pesticides are used year-round in the LRGV. More than 100 different pesticides were used on crops throughout the region (US FWS, 1986). The pesticides applied in citrus orchards have strong potential to move via different environmental media (Vredeveld et al., 1983). The most widely used pesticides in citrus orchards are atrazine, aldicarb, carbofuran, dicrotophos, dicamba, glyphosate, simazine and methomyl (Bryant et al., 1993; NCFAP, 1997). Continuous use of pesticides results in damage to the environment and can cause human health problems, which could have a negative impact on agricultural production and reduce agricultural sustainability (Pimentel et al., 1992). Environmental impact is dependent upon a variety

*Corresponding author. Tel.: +1 407 7547521.
E-mail address: nchang@mail.ucf.edu (N.-B. Chang).

301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
oi:10.1016/j.jenvman.2006.11.020

Nomenclature

D_w, D	diffusion–dispersion coefficient for the dissolved phase ($L^2 T^{-1}$)
t	time (T)
D_g	molecular diffusivity of the pesticide in the air-filled pore space ($L^2 T^{-1}$)
a	volumetric air content of the soil ($L^3 L^{-3}$)
C_g	gaseous concentration of the pesticide ($M L^{-3}$)
K_H	Henry's constant
z, x	vertical space dimension (L)
C_w	dissolved concentration of the pesticide ($M L^{-3}$)
θ	volumetric soil-water content ($L^3 L^{-3}$)
v	velocity of the soil water($L T^{-1}$)
K_s, k_1 and λ	first-order decay constant for the solid and dissolved phases (T^{-1})
K_d	partition coefficient between the dissolved and solid phases (T^{-1})
ρ_b	soil bulk density ($M L^{-3}$)
J_{APP}	mass gain due to pesticide application at or near the surface ($M T^{-1}$)
X_e	erosion sediment loss ($M T^{-1}$)
A_w	watershed area (M^2)
Q	daily runoff volume ($L^3 T^{-1}$)
P_r	daily rainfall depth ($L T^{-1}$)
ε	uptake efficiency factor
K_{TRN}	transformation rate constant (T^{-1}) and
r_{om}	enrichment ratio for organic matter ($M M^{-1}$)
q	water flux ($L T^{-1}$)
r_e	pesticide erosion rate(T^{-1})
r_u	pesticide root uptake rate (T^{-1})
M	pesticide loading ($ML^{-3} T^{-1}$)
D_g	diffusion coefficient of the vapor phase pesticide ($L^2 T^{-1}$)
ψ	pressure head (L)
K	saturated hydraulic conductivity ($L T^{-1}$)
k_{rw}	relative permeability
η	effective waste storage capacity (L^{-1})
α, β and γ	empirical parameters
Ψ_a	air entry pressure head value
S_{wr}	residual water phase saturation
S_w	water saturation
R	retardation coefficient

of parameters such as pesticide dose, application technique, and environmental conditions (e.g., weather, soil type, land use, presence of surface water or sensitive biological species, etc.) (Cohen et al., 1995; Reus et al., 2002). Estimating concentrations of pesticide that are transported through soil (root and vadose zones) to groundwater is the main purpose behind this study. At a higher level of sophistication, a wide variety of computer models are available that can quantitatively simulate pesticide leaching and runoff in the aqueous phase. There is a need to pick a model that has been validated in more than one study and has good user support. However, this would require a credible selection of a suite of similar models appropriate for the application, which shows a history of producing sound results acceptable to scientists and regulatory authorities.

Considering these various criteria for acceptability, EPA's pesticide root zone model-3 (PRZM-3) is an appropriate tool to estimate pesticide concentration in both soil and groundwater (Carsel et al., 2003). PRZM-3 has already been used in Fresno County in California to simulate 1,2-dibromo-3-chloropropane (DBCP) pesticide contamination in groundwater (Loague et al., 1998). This model has also been used to estimate the transport and fate of pesticides in potato cultures in the Nicolet River Basin, Canada (Pierre et al., 1996) and the fate and transport of ethoprophos and bentazone in a sandy humic soil (Trevisan et al., 2000). Moreover, PRZM-3 has the capability to simulate pesticide concentration in multiple zones. This allows the model to combine different root zone and vadose zone (i.e., unsaturated zone) characteristics into a single simulation, and the model has the ability to simulate as many as three chemicals simultaneously so that it gives the user the option to observe the concentrations of multiple chemicals without making additional runs. The pesticide analytical model (PESTAN) is also used for estimating the transport of organic solutes through soil to groundwater and has been used by the US Environmental Protection Agency Office of Pesticide Program (Carsel et al., 2003) for initial screening assessments to evaluate the potential for groundwater contamination of currently registered pesticides (Donigian and Rao, 1986). integrated pesticide transport modeling (IPTM) has been used to study aldicarb transport in the subsurface environment (Chu et al., 2000; Chu, 2004) and also was used in Orestimba Creek Basin, California for evaluating the vulnerability of the hydrosystem to diazinon contamination (Chu and Marino, 2004). The risk of unsaturated/saturated transport and transportation of chemical concentrations (RUSTIC) (Dean et al., 1989), groundwater loading effects of agricultural management systems (GLEAMS) (Leonard et al., 1987), leaching estimation and chemistry model (LEACHM) (Wagenet and Huston, 1989), and calculation flow (CALF) (Walker, 1987) models are also scientifically acceptable, but have not been as widely used because in CALF there is no allowance for volatilization and crop growth and the GLEAMS model is appropriate for quantifying runoff potential in simple, field-scale drainage patterns (Cohen et al., 1995).

The applicability of the three models, PRZM-3, PESTAN, and IPTM, for different pesticides is interesting in the management of chemical leaching in citrus orchards. The main objectives of this paper are to compare the application potential of the PRZM-3 model against the

;STAN and IPTM models and to select one of them for ure advanced precision farming study. Multiple criteria :ision making (MCDM) was introduced as a promising ›l for decision analysis in early 1970s. The key ilosophical structure for MCDM lies in the representa-n of several conflicting criteria (Stewart, 1992). Since :n, the number of contributions to the theory has ntinued to grow at a steady rate. Zeleny (1982) shows it multiple criteria include both multiple attributes and iltiple objectives resulting in two theoretical variations, ich can be divided into multiple attribute decision king (MADM) and multiple objective decision making ODM) in decision science. The multiattribute decision-king (MADM) method has been used to deal with the ›blem of ranking and selecting plant locations under iltiple criteria (Hwang and Yoon, 1981; Tompkins and iite, 1984; Spohrer and Kmak, 1984; Pavic and Babic,)1; Rietveld and Ouwersloot, 1992; Prato, 1999; Kuo al., 2002). However, such a model based screening ›cedure is designed based on a structured process for lecting and distilling knowledge from a group of experts means of a series of questionnaires in an uncertain ironment. Fuzzy linguistic models permit the transla-n of verbal expressions into numerical ones, thereby iling quantitatively with imprecision in the expression of importance of each criterion (Zadeh, 1965). This ilysis counts on the integration of fuzzy sets with ditional MADM. The fuzzy MADM (FMADM) thods have been widely used in environmental planning l decision-making processes in order to clarify the nning process, to avoid various distortions, and to nage all the information, uncertainties, and importance the criteria (Teng and Tzeng, 1996; Tran et al., 2002; iou and Tzeng, 2002). It provides a philosophical base ranking the models under uncertainty (Chen and ang, 1993; Carlsson and Fuller, 1996).

Methodologies

Model comparison is not trivial as the size of the model e grows over time. The need for model selection can be n in different disciplinary areas such as air quality nagement, water quality management, and groundwater ility management. For example, assessing pesticide ›acts on water quality can be conducted by a variety nodels. Gualtieri (2002) compared an individual water ility model against the TOXI5 model which is part of WASP5 modeling framework developed by USEPA. h models were applied to an idealized case in order predict steady-state and time-variable concentrations he water-column and in the active sediment layer for r pesticides. On the other hand, ground-level ozone is ned by a chemical reaction between VOCs and oxides nitrogen (NO_x) in the presence of sunlight in many an regions. Liang et al. (2004) compared the predictions two well-known photochemical models, including Mx and CMAQ, to simulate the ozone episodes and examined the sensitivity of both models to pollutant emissions.

Since water quality in groundwater and surface water bodies depends on activities in the surrounding soils and ecosystems, model comparison should show an intrinsic linkage between soil and water quality when considering technical, economic, and risk factors. Fig. 1 shows the schematic representation of the overall work flow to determine the transport and fate of simazine in the subsurface environment of citrus orchards. The overall work is divided into three parts (1) simulation and comparison of three models (PRZM-3, PESTAN and IPTM) and (2) data collection and analysis, and (3) model selection analysis using fuzzy multicriteria decision making (FMCDM). The symbols are summarized in the Nomenclature.

2.1. Pesticide root zone model-3

The PRZM-3 links two modules, including PRZM and VADOFT, to predict pesticide transport and transformation down through the crop root and vadose (unsaturated) zone to the water table. The PRZM-3 incorporates soil temperature simulation, volatilization and vapor phase transport in soils, irrigation simulation, and microbial transformation. The PRZM sub-model is a one-dimensional, finite-difference model that uses a method of characteristics (MOC) algorithm to eliminate numerical dispersion and the VADOFT sub-model is a one-dimensional finite-element code that solves Richards' equation for flow in the unsaturated zone.

The PRZM sub-model has two major components—hydrology and chemical transport. The hydrologic component for calculating runoff and erosion is based on the Soil Conservation Service curve number technique and the Universal Soil Loss Equation (Carsel et al., 2003). Evapotranspiration is estimated from pan evaporation data and is divided among evaporation from crop interception, evaporation from soil, and transpiration by the crop. Water movement is simulated by the use of generalized soil parameters, including field capacity, wilting point, and saturation water content. The PRZM sub-model adopts an empirical water balance procedure describing soil water movement in unsaturated subsurface. Soil-water velocities in vertical unsaturated profiles are estimated in PRMZ-3 simulations based on the following simple relationship:

$$SW^{t+1} = SW^{t} + (v_i - v_{i-1})\Delta t, \tag{1}$$

where SW is the soil-water content, t and $t+1$ denote the beginning and end of the time step, v the velocity, i the soil layer index, and Δt the time step. Dissolved, adsorbed, and vapor-phase concentrations in the soil are estimated by simultaneously considering the processes of pesticide uptake by plants, surface runoff, erosion, decay, volatilization, foliar washoff, advection, dispersion, and retardation. The combined surface zone expression for dissolved,

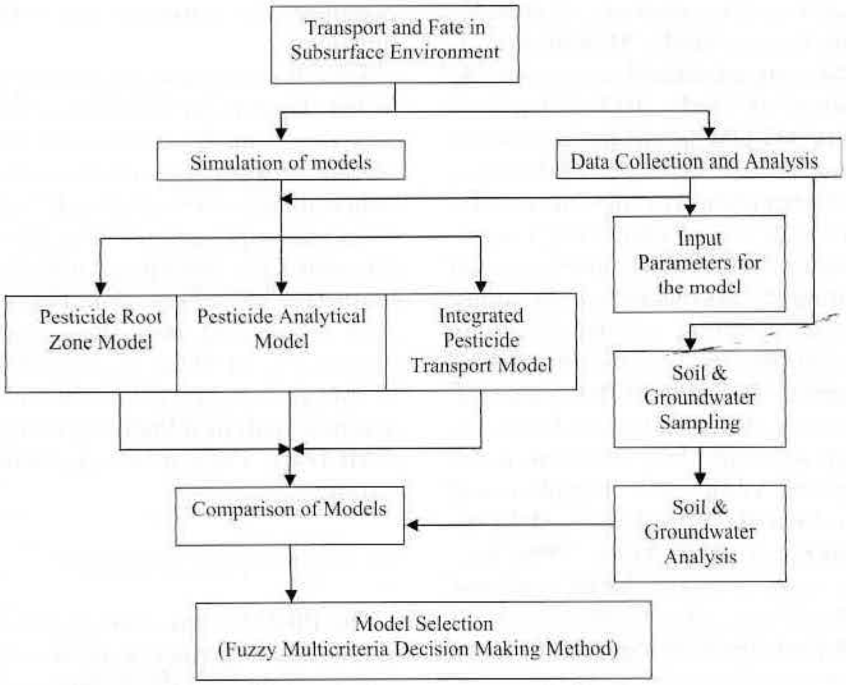


Fig. 1. Schematic representation of the workflow.

adsorbed, and vapor phases is given by

$$D_{\mathrm{w}}\frac{\partial^2(C_{\mathrm{w}}\theta)}{\partial z^2}+D_{\mathrm{g}}\frac{\partial^2(aC_{\mathrm{g}}K_{\mathrm{H}})}{\partial z^2}-\frac{\partial(C_{\mathrm{w}}\theta v)}{\partial z} - C_{\mathrm{w}}\Big[K_{\mathrm{s}}(\theta+K_{\mathrm{d}}\rho_{\mathrm{b}})+k_{\mathrm{g}}aK_{\mathrm{H}}+f\theta\varepsilon + \frac{Q}{A_{\mathrm{w}}\Delta z}+\frac{PX_{\mathrm{e}}r_{\mathrm{om}}K_{\mathrm{d}}}{A_{\mathrm{w}}\Delta z}\Big]+\frac{J_{\mathrm{APP}}}{A\Delta z}+\frac{EP_{\mathrm{r}}M}{\Delta z} - K_{\mathrm{TRN}}C_{\mathrm{w}}\theta+\sum_K K_{\mathrm{TRN}}C_{\mathrm{w}}\theta = \frac{\partial[C_{\mathrm{w}}(\theta+K_{\mathrm{d}}\rho_{\mathrm{b}}+aK_{\mathrm{H}})]}{\partial t}, \tag{2}$$

where D_{w} is the diffusion–dispersion coefficient for the dissolved phase, assumed constant ($\mathrm{L^2\,T^{-1}}$), t the time (T), D_g the molecular diffusivity of the pesticide in the air-filled pore space ($\mathrm{L^2\,T^{-1}}$), a the volumetric air content of the soil ($\mathrm{L^3\,L^{-3}}$), C_{g} the gaseous concentration of the pesticide ($\mathrm{M\,L^{-3}}$), K_{H} the distribution coefficient between the liquid and vapor phases (Henry's constant), z the vertical space dimension (L), Δz the space increment (L), C_{w} the dissolved concentration of the pesticide($\mathrm{M\,L^{-3}}$), θ the volumetric soil-water content ($\mathrm{L^3\,L^{-3}}$), v the velocity of the soil water($\mathrm{L\,T^{-1}}$), K_{s} the lumped, first-order decay constant for the solid and dissolved phases($\mathrm{T^{-1}}$), K_{d} the partition coefficient between the dissolved and solid phases($\mathrm{T^{-1}}$), ρ_{b} the soil bulk density ($\mathrm{M\,L^{-3}}$), J_{APP} the mass gain due to pesticide application at or near the surface ($\mathrm{M\,T^{-1}}$), X_{e} the erosion sediment loss ($\mathrm{M\,T^{-1}}$), p a units conversion factor ($\mathrm{M\,M^{-1}}$), A the cross-sectional area of the soil column, A_{w} is the watershed area ($\mathrm{M^2}$), Q the daily runoff volume ($\mathrm{L^3\,T^{-1}}$), P_{r} the daily rainfall depth ($\mathrm{LT^{-1}}$), ε an uptake efficiency factor or reflectance coefficient (dimensionless) f the fraction of total water in the zone used for transpiration ($\mathrm{T^{-1}}$), K_{TRN} the transformation rate constant ($\mathrm{T^{-1}}$) and r_{om} the enrichment ratio for organic matter ($\mathrm{M\,M^{-1}}$).The above equation can be solved in PRZM for a surface layer with $f\theta = 0$.

The VADOFT sub-model is a finite-element code for simulating one-dimensional, single-phase moisture movement in unconfined, variably saturated porous media and considers solute transport in the vadose zone which is defined as single-porosity media. The second part of the coupled PRZM-VADOFT model can predict the movement of pesticides within and below the plant root zone and assess consequent groundwater contamination. Transport of dissolved contaminants may also be simulated within the same domain. Transport processes accounted for include: hydrodynamic dispersion, advection, linear equilibrium sorption, and first-order decay. The VADOFT sub-model also simulates solute transformations in order to account for parent/daughter relationships. Overall, the VADOFT sub-model consists of flow and solute transport models.

2.1.1. Flow equation

VADOFT considers the problem of variably saturated flow in a soil column in the vadose zone of an unconfined aquifer. Infiltration of water in the vadose zone is given by the Richards' equation,

$$\frac{\partial[Kk_{\mathrm{rw}}(\partial\psi/\partial z)]}{\partial z}=\eta\frac{\partial\psi}{\partial t}, \tag{3}$$

where ψ is the pressure head (L), K the saturated hydraulic conductivity ($\mathrm{L\,T^{-1}}$), k_{rw} the relative permeability, z the vertical coordinate pointing in the downward direction (L), he time (T) and η the effective waste storage capacity $^{-1}$). The above equation can be solved by specifying the ationship between water saturation and pressure head ven by

$$\frac{{}_{\mathrm{w}}S_{\mathrm{wr}}}{-S_{\mathrm{wr}}} = \begin{cases} \dfrac{1}{[1+(a|\Psi-\Psi_{\mathrm{a}}|)\beta]\gamma} & \text{for } \Psi<\Psi_{\mathrm{a}}, \\ 1 & \text{for } \Psi\geqslant\Psi_{\mathrm{a}}, \end{cases} \quad (4)$$

iere α and β are empirical parameters, Ψ_{a} the air entry essure head value, S_{wr} the residual water phase satura-n, S_{w} the water saturation and γ the empirical parameter.

.2. Transport equation

The governing equation for one-dimensional transport a non-conservative solute species in a variably saturated il is given by

$$\left(D\frac{\partial C}{\partial z}\right) - v\frac{\partial C}{\partial z} = \theta R\left(\frac{\partial C}{\partial t}+\lambda C\right), \quad (5)$$

iere D is the apparent dispersion coefficient ($\mathrm{L^2\,T^{-1}}$), C solute concentration ($\mathrm{M\,L^{-3}}$), θ the volumetric water ntent, R the retardation coefficient, and λ the first-order cay constant ($\mathrm{T^{-1}}$).

. Pesticide analytical model

The PESTAN is used for evaluating the transport of ganic solutes through the vadose zone to ground water. ie PESTAN uses an analytical solution to calculate ganic movement based on a linear isotherm, first-order gradation and hydrodynamic dispersion. The model is sed on a closed form analytical solution of the vective–dispersive–reactive transport equation and has en tested under field and laboratory conditions. Input ta includes water solubility, infiltration rate, bulk nsity, sorption constant, degradation rates, saturated iter content, characteristic curve coefficient, saturated draulic conductivity, and dispersion coefficient. The rtical transport of a pollutant dissolved in water through soil is described by the following equation:

$$= D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x} - \frac{\rho C}{\theta}\frac{\partial S}{\partial t} - K_1 C, \quad (6)$$

iere C is the liquid phase pollutant concentration $\mathrm{L^{-3}}$), t the time (T), x the distance along the flow path), D the dispersion coefficient ($\mathrm{L^2\,T^{-1}}$), v the interstitial iter velocity ($\mathrm{L\,T^{-1}}$), ρ the bulk density ($\mathrm{M\,L^{-3}}$), θ the lumetric water content ($\mathrm{L^3\,L^{-3}}$), S the solid phase ncentration ($\mathrm{M\,M^{-1}}$), and k_1 the first-order decay efficient in liquid phase ($\mathrm{T^{-1}}$).

The above partial differential equation can be solved for C t) along with following initial and boundary conditions:

$$x, t=0) = \begin{cases} 0, & x<-x_0 \\ C_0, & -x_0<x<0, \\ 0, & 0<x \end{cases} \quad (7)$$

$$\lim \frac{\partial C}{\partial x} = 0 \quad \text{when} |x| \to \infty, \quad (8)$$

where x_0 is the slug thickness and C_0 the initial concentration.

2.3. Integrated pesticide transport model

The IPTM is suitable for simulating one-dimensional, three-phase (dissolved, adsorbed, and vapor phases) pesticide transport and transformation in the vadose zone, which consists of the surface zone, plant root zone, and deep vadose zone. The model takes into account heterogeneous media, unsteady flow fields, and space-time-dependent physical and biochemical processes related to pesticide environmental fate. The primary pesticide processes simulated in the model include advection, diffusion/dispersion, linear equilibrium sorption, linear equilibrium partitioning between vapor and dissolved phases, first-order decay, plant root uptake, volatilization from the soil surface to the overlying atmosphere, as well as pesticide runoff and erosion in the surface zone.

Assuming linear equilibrium sorption, linear equilibrium liquid vapor partitioning and first-order decay, the second-order partial differential equation governing one-dimensional three-phase pesticide transport in the vadose zone can be expressed as

$$\begin{aligned} &\frac{\partial}{\partial t}[(\theta+\rho K_{\mathrm{d}}+aK_{\mathrm{H}})C_{\mathrm{w}}] \\ &= \frac{\partial}{\partial z}\left[aD_{\mathrm{g}}\frac{\partial}{\partial z}(K_{\mathrm{H}}C_{\mathrm{w}})\right] + \frac{\partial}{\partial z}\left[\theta D_{\mathrm{w}}\frac{\partial C_{\mathrm{w}}}{\partial z}\right] \\ &\quad - \frac{\partial}{\partial z}(qC_{\mathrm{w}}) - (r_{\mathrm{r}}+r_{\mathrm{e}}+r_{\mathrm{u}})C_{\mathrm{w}} \\ &\quad - (\theta+\rho K_{\mathrm{d}}+aK_{\mathrm{H}})K_{\mathrm{s}}C_{\mathrm{w}} + M(z,t), \end{aligned} \quad (9)$$

where C_{w} is the dissolved phase pollutant concentration ($\mathrm{M\,L^{-3}}$), D_{g} the diffusion coefficient of the vapor phase pesticide ($\mathrm{L^2\,T^{-1}}$), D_{w} the dispersion coefficient of the dissolved phase pesticide ($\mathrm{L^2\,T^{-1}}$), k_{s} the first order decay coefficient in liquid phase ($\mathrm{T^{-1}}$), q the water flux ($\mathrm{L\,T^{-1}}$), r_{r} the pesticide runoff ($\mathrm{T^{-1}}$), r_{e} the pesticide erosion rate($\mathrm{T^{-1}}$), r_{u} the pesticide root uptake rate ($\mathrm{T^{-1}}$) and M pesticide loading term ($\mathrm{ML^{-3}\,T^{-1}}$). The above partial differential equation can be solved using the following initial and boundary conditions:

Initial condition:

$$C(z, t_0) = C_0(z),$$

where t_0 denotes the initial time and $C_0(z)$ is the corresponding initial concentration at location z. Boundary conditions are:

(1) For $t>0$ and $z=0$ (assuming $C_{\mathrm{g}}^{\mathrm{a}}=0$):

$$\begin{aligned} \frac{\partial(K_{\mathrm{H}}C_{\mathrm{w}})}{\partial z} &= K_{\mathrm{H}}\frac{\partial C_{\mathrm{w}}}{\partial z} - (aD_{\mathrm{g}}K_{\mathrm{H}}+\theta D_{\mathrm{w}})\frac{\partial C_{\mathrm{w}}}{\partial z} \\ &\quad + \left(q+\frac{D_{\mathrm{a}}K_{\mathrm{H}}}{d}\right)C_{\mathrm{w}} = qC_{\mathrm{w}}\mathrm{in}. \end{aligned} \quad (10)$$

(2) For $t>0$ and $z=L$:

$$-(aD_gK_H+\theta D_w)\frac{\partial C_w}{\partial z}+qC_w=qC_w\text{out}. \tag{11}$$

Fig. 2 represents the schematic diagram of the root and vadose zones in the soil subsurface. The above mentioned pesticide simulation models were applied to a simplified system that spreads over a vadose zone with 4 m depth to compare the performance of the models (Carsel et al., 2003). Table 1 summarizes characteristics and capabilities of the models evaluated for this study. All three models can calculate contaminant concentrations in leachate that ha[s] infiltrated down to the water table from the vadose zone.

3. Case study

3.1. Study location

The study location was the South Research Farm o[f] Texas A&M University—Kingsville Citrus Center i[n] Weslaco, Texas. The experimental block was about 9.[?] acres, and it is part of a large citrus operation of about 25[0] acres. Agricultural chemicals are used regularly in thes[e]

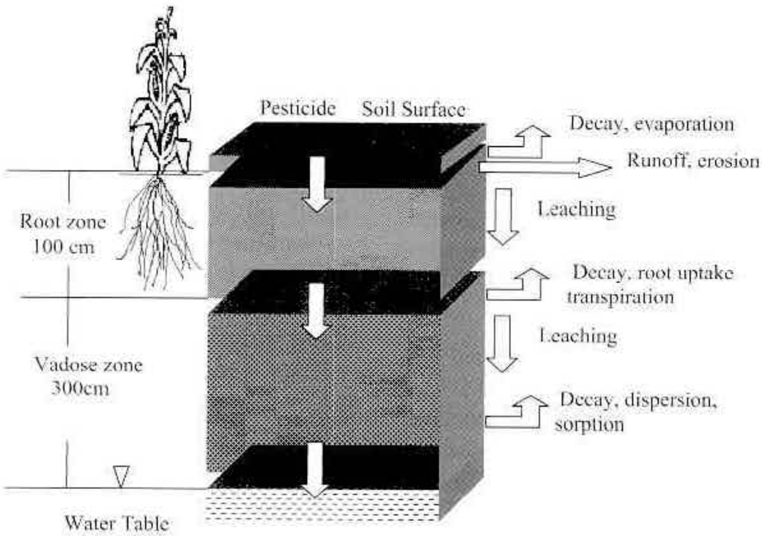


Fig. 2. Schematic representation of root and vadose zones in the subsurface soil.

Table 1
Characteristics and capabilities of the three models

		PRZM 3	PESTAN	IPTM
Type	Analytical		■	
	Hybrid (analytical and numerical)			■
	Numerical	■		
Transport and fate process considered	Finite source	■	■	■
	Volatilization	■		■
	Hydrodynamic dispersion	■	■	■
	Sorption	■	■	■
	Non-equilibrium partitioning			
	First-order decay	■	■	■
	Biodegradation	■	■	■
	Layered soils	■		■
	Root zone uptake	■	■	■
	Runoff	■		■
	Erosion	■		■
	Saturated zone included			
Other	Monte Carlo analysis	■		
	Water balance calculations	■		■
	Large output data set transfer compatibility	■		
	Simulation of transport of nitrogen	■		

:hards. Flooding is the most common irrigation method ng used, especially during the 1980s. Simazine and phosate are intensively used pesticides in the citrus :hards. Fig. 3 shows the map of groundwater wells with ter level from the surface. It can be observed from Fig. 3 it ground water is shallow (43.2 ft) near the citrus farm. 3. 4 shows the flow gradient of groundwater in the study gion. Flood irrigation in the farm can cause leaching of sticides into the shallow Gulf coast aquifer. Hence, there a need to conduct proper studies for assessing the tential of ground water contamination in this region. e fate of pesticides in soils is determined by a variety of

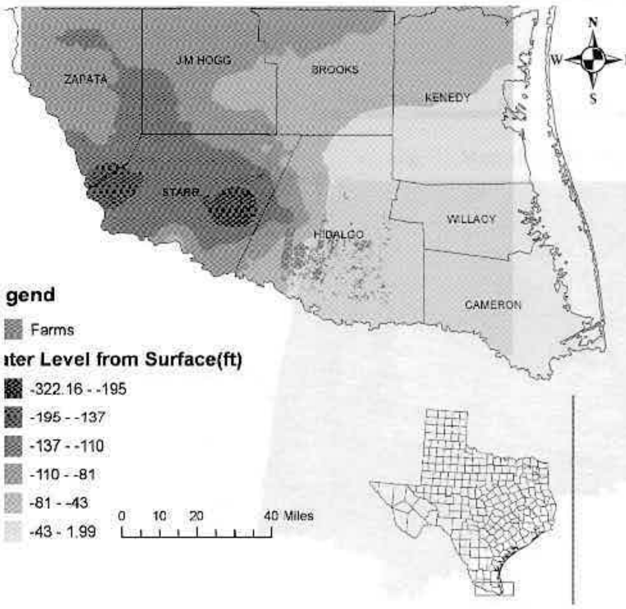


Fig. 3. Map showing the water level in the wells from the surface.

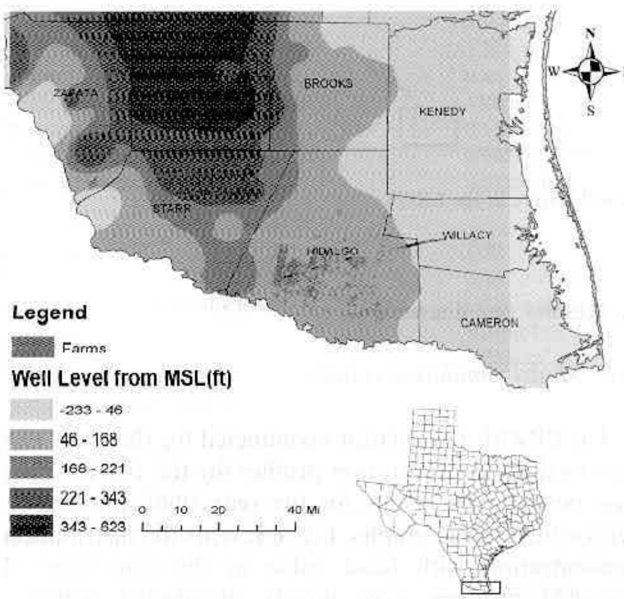


Fig. 4. Map showing the flow gradient of groundwater.

physico-chemical soil processes and crop processes (Brown et al., 1995).

3.2. *Input parameters for the models*

3.2.1. *Pesticide application rate*

The historical record of simazine pesticide application rates was obtained from the citrus farm operations at the citrus center and is shown in Table 2. The pesticide application was implemented during the 1980s in the study area and may have caused contamination of soil, surface water, and groundwater (Fig. 5).

3.2.2. *Meteorological data*

Meteorological data (1981–1990) required for the model simulation, such as temperature, precipitation, evapotranspiration, were obtained from the EPA's Brownsville weather station located in the LRGV. The reason behind using this data set is that these data files are compatible with the PRZM-3 software model and the current version of this data set is not available. The rainfall events measured in the LRGV area can be generally classified into three different types, fine rain (i.e., total accumulated rainfall is less than 30 mm and the maximum of rainfall intensity is less than 10 mm/h), heavy rain (i.e., total accumulated rainfall is between 30 and 100 mm, and the maximum of rainfall intensity is higher than 10 mm/h), and torrential rain (i.e., total accumulated rainfall is higher than 100 mm). The rainfall intensity during 1981–1990 in the LRGV is shown in Fig. 6. According to the meteorological data during the time period between 1981 and 1990, there were 716 rainfall events recorded and all the events were fine rain events. However, variations in annual rainfall create different environmental impacts. Storms come in normally in late spring and early fall in every year.

3.2.3. *Soil properties*

USGS STASTGO soil data and GIS were used to map the soil type in the citrus orchard. Fig. 7. shows the map of the orchards and soil type in the LRGV. Approximately

Table 2
Application history of simazine in citrus farm

Total application (kg/ha)	Application year
15	1981
15	1982
15	1983
16	1984
14	1985
15	1986
14	1987
15	1988
18	1989
23	1990

Note: Application rates are rounded to nearest whole numbers

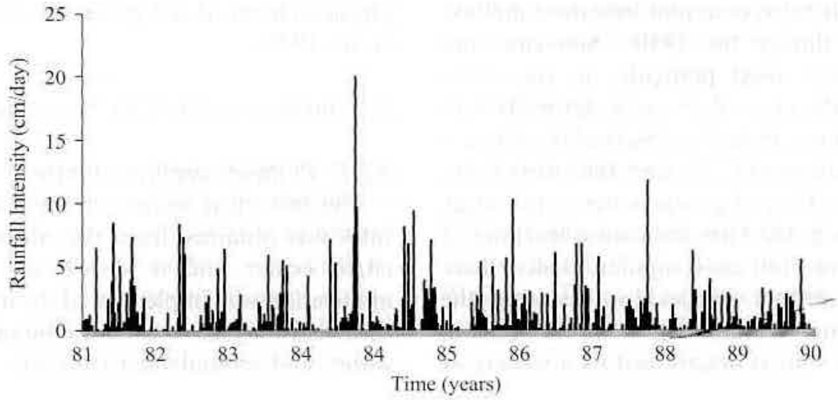


Fig. 5. Rainfall intensity during 1981–1990 in Lower Rio Grande Valley.

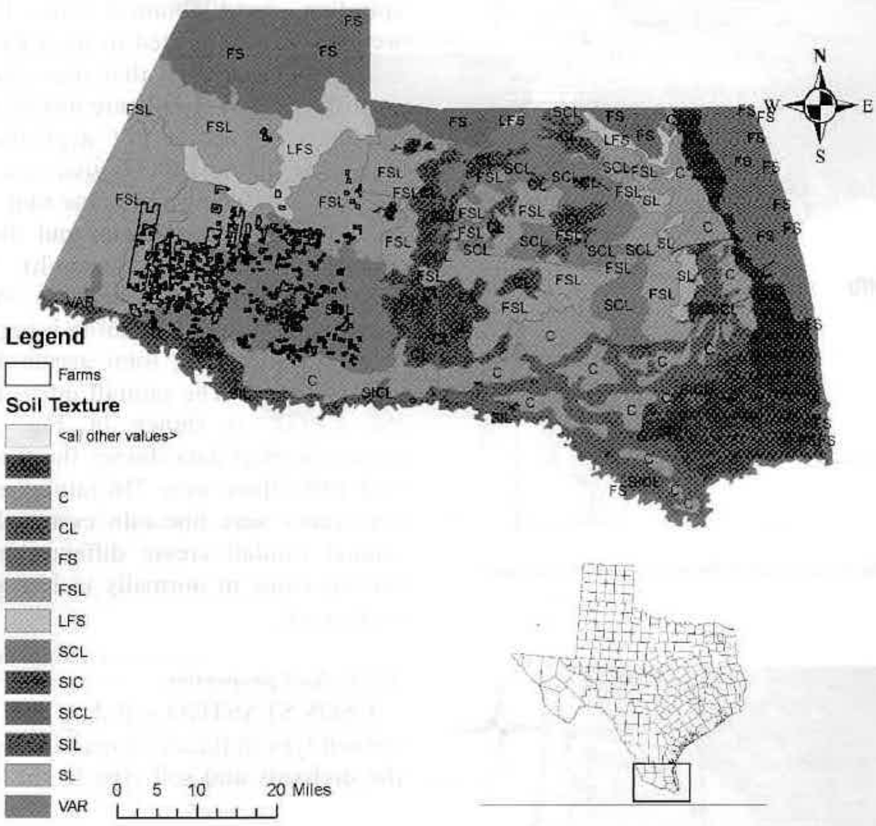


Fig. 6. Map showing the farms and soil texture in the LRGV.

9% of the total land area in the LRGV is covered by sandy clay loam (SCL). Most of the soil texture in the citrus orchards is SCL. Based on the soil type, different input parameters for the models PRZM and VADOFT, such as hydraulic conductivity and organic carbon partitioning coefficient, were obtained from the STASTGO soil database (USDA NRCS, 1996) and PRZM 3 user manual (Carsel et al., 2003). Table 3 summarizes the list of all parameters that were used in the models. Some of them were determined based on the values in the literature.

4. Results and discussion

4.1. Model simulation results

The PRZM-3 simulations conducted for the study region provided the concentration profiles for the 10-year simulation period. The results for the year 1990 simulation are represented in the graphs. Fig. 8 reveals the distribution of concentration with head value in the root zone. The simazine residues were mostly distributed within the shallow depth (from 0 to 40 cm below the ground surface)

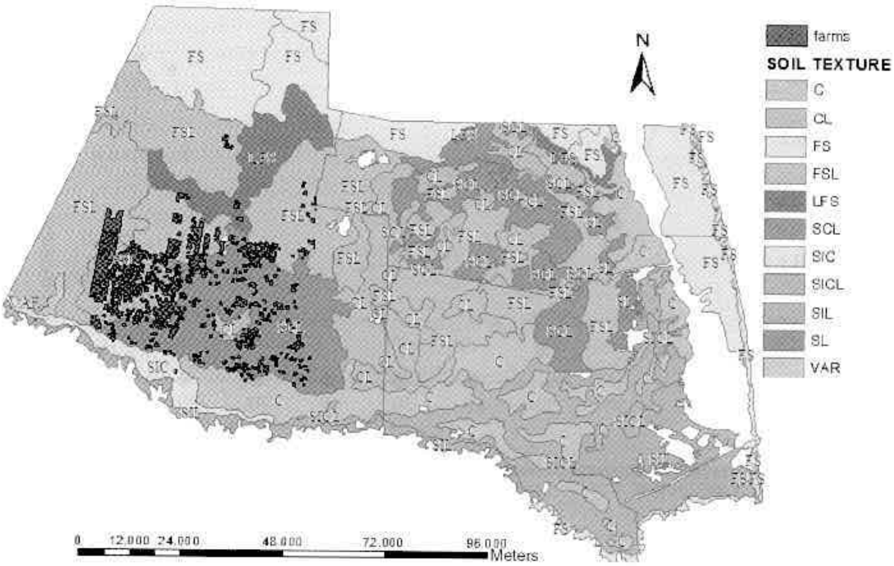


Fig. 7. Map showing the farms and soil texture in the LRGV.

ıble 3
ːst of all parameters that were used in three models

ırameter	Value	Reference
ːop: citrus		
aximum canopy coverage	60%	Texas A&M, Citrus Center
umber of different crops	1	
ate of emergence	October 1	Texas A&M, Citrus Center
ate of maturity	April 1	Texas A&M, Citrus Center
ate of crop harvest	April 15	Texas A&M, Citrus Center
SLEC factor	0.10	
il properties		
oot zone		
ıtal soil depth (cored)	100 cm	
›. of horizons	2	
ılk density	1.51 g/m^3	Carsel et al. (2003)
ırosity	0.45	
ganic carbon partitioning coefficient	130 cm^3/g	USDA-ARS, GLEAMS
ilting point	0.15 cm^3/cm^3	
stribution coefficient	0.405 cm^3/g	IPTM user manual
ːst order decay rate	0.01155/day	USDA-ARS, GLEAMS
dose zone		
lk density	1.59	Carsel et al. (2003)
rosity	0.43	
turated hydraulic conductivity	0.314 m/day	IPTM user manual
ıter solubility	6.2 mg/L@ 22 °C	USDA-ARS, GLEAMS
stribution coefficient	0.205 cm^3/g	IPTM user manual
sticide properties: simazine		
›. of applications	10	
erage field half-life	28–60 days	Extension Toxicology Network
ffusion coefficient in air	0.43 m^2/day	IPTM user manual
nrys law constant	1.3E−08	Carsel et al. (2003)
g K_{ow}	1.94	Extension Toxicology Network
ղuid phase decay rate	0.01155/day	USDA-ARS, GLEAMS
id phase decay rate	0.01155/day	USDA-ARS, GLEAMS
ffusion coefficient in water	0.000043 m^2/day	Ref. Jury et al. (1983)

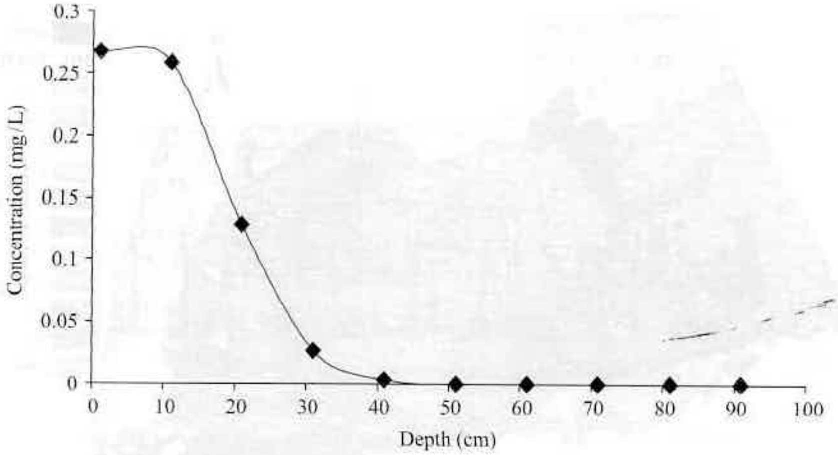


Fig. 8. PRZM 3 simulation results: dissolved concentration of simazine in the root zone in 1990.

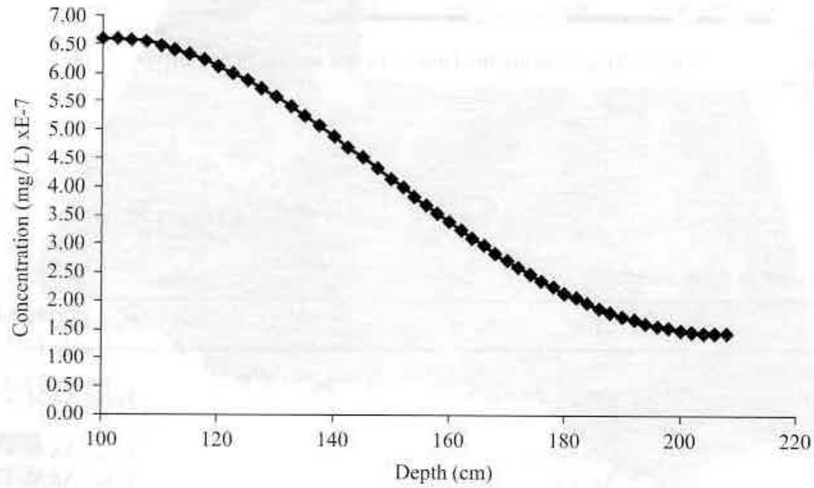


Fig. 9. PRZM-3 simulation results: dissolved concentration of simazine in the vadose zone in 1990.

with a maximum concentration of 0.2673 mg/L located 1 cm below the ground surface, and reached 2.01E−07 mg/L at the interface of the root and vadose zones (100 cm below the ground level). Fig. 9 shows the concentration profile of simazine in the vadose zone, 100–400 cm (100 cm root zone + 300 cm vadose zone) below the ground level. The variation of concentration of simazine residues in the vadose zone exhibited the same pattern as in the root zone. Obviously, the simazine concentration decreases with depth in the vadose zone and the peak concentrations (6.6E−07 mg/L) were located at the interface of the root and vadose zones. Fig. 10 shows the overall distribution of simazine in both the root and vadose zones. It can be inferred from the concentration profiles that there is no considerable simazine impact in both root and vadose zones, and the concentration values are well below the regulatory MCL standard provided by US EPA.

Fig. 11 shows the simulated concentration profiles of simazine in the root zone using the PESTAN model. The simazine residues were mostly distributed within the root zone with a maximum concentration of 0.24 mg/L located 1 cm below the ground surface, and reached 2.41E−07 mg/L at the interface of the root and vadose zones. Fig. 12 shows the distribution of simazine concentration in the root and vadose zones obtained from the PESTAN model. The concentration profile exhibits the same pattern as in the root zone, that is, the concentration decreases with increase in depth of the vadose zone.

Fig. 13 shows the concentration profile of simazine in the root zone using the IPTM model. Very small amounts of simazine residues were distributed in the root zone (from 0 to 100 cm below the ground surface) with a maximum concentration of 0.02636 mg/L located just below the ground surface and reached 4.231E−05 mg/L at the interface of the root and vadose zones. Fig. 14 shows the simazine concentration in the root and vadose zones. The concentration of simazine decreases with an increase in the depth of the vadose zone, reaching a value close to zero. Simazine concentrations from the IPTM model are very low when compared to PRZM. This is due to the fact that the pesticide loss due to decay, runoff volatilization, and

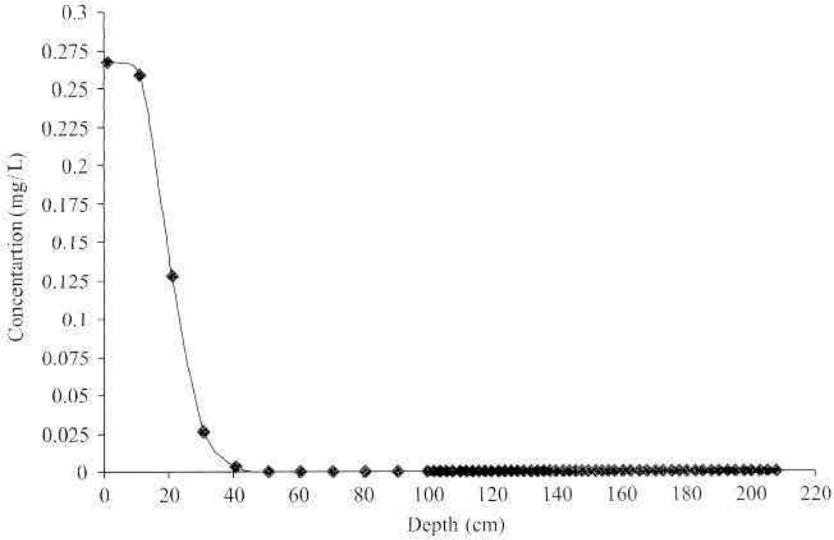


Fig. 10. PRZM 3 simulation results: concentration profile of simazine in the root and vadose zone in 1990.

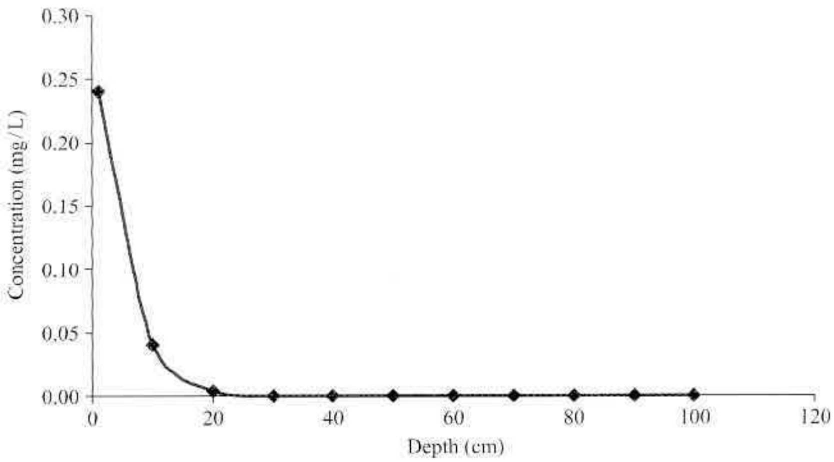


Fig. 11. PESTAN simulation results: dissolved concentration of simazine in the root zone in 1990.

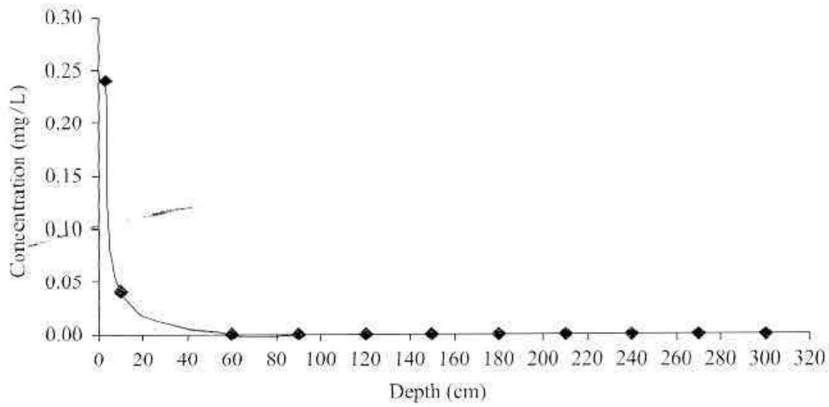


Fig. 12. PESTAN simulation results: dissolved concentration of simazine in the root and vadose zones in 1990.

osion calculated by the IPTM model during the year 1990 is comparatively higher than the PRZM values. Concentration values in the vadose zone obtained from the PESTAN and IPTM models are well below the MCL values (4 μg/L). The trends of the simazine concentration profiles in the root and vadose zones in all the three models

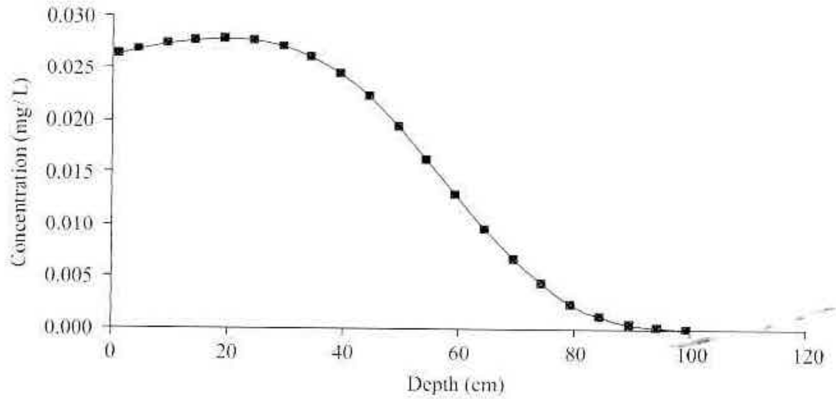


Fig. 13. IPTM simulation results: dissolved concentration of simazine in the root zone in 1990.

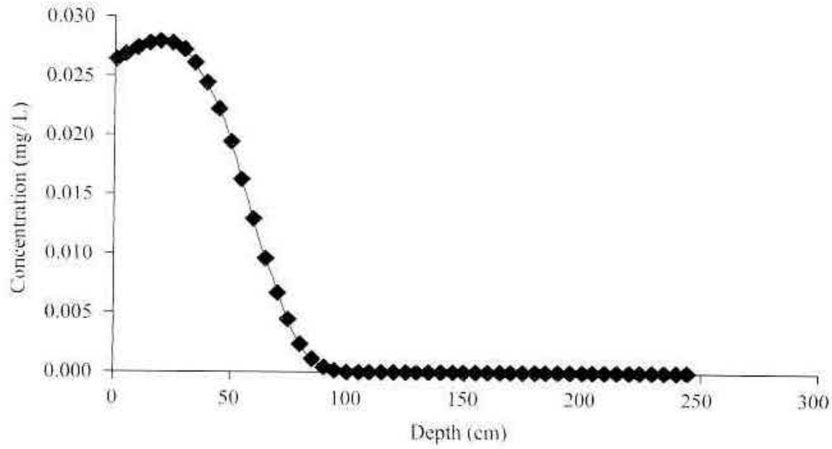


Fig. 14. IPTM simulation results: dissolved concentration of simazine in the root and vadose zones in 1990.

are similar, which implies that the concentration decreases as the depth increases. The model also predicted that most of the pesticide applied is degraded in the soil before reaching the water table, while some is lost in runoff, erosion and volatilization.

4.2. Model selection criteria

Computer simulation models are valuable tools for assessment of the behavior of chemicals in the environment. Yet models never completely reflect reality: they always simplify various physical, chemical and biological processes in the environment or in environmental compartments. Therefore, the results obtained by computer models should be interpreted considering the simplifications of the model (Cohen et al., 1995). Selection of models depends on some criteria like model validation and model usage, the availability of input data for the model, user friendliness, number of studies in which the model is evaluated (Cohen et al., 1995) and track record of successful applications (Stephen et al., 1999).

Following the method developed by Chang and Chen (1994) and Chen (2001) the criteria that were selected for evaluating the merits of the three models in this study are: (1) validation and calibration (VC), (2) model availability and user support (MA), (3) availability of input data (ID), (4) user friendliness (UF), (5) model usage (MU), (6) predictive results (PR), and (7) independent replication (IR).

The decision objective was to select the most appropriate one from three different alternative models. The alternatives were the models {PRZM-3, PESTAN, IPTM}. The decision criteria were defined as {VC, MA, ID, UF, MU, PR, IR}. A weighting set {very low, low, medium low, medium, medium high, high, and very high} was assumed to evaluate the importance of each criteria and a rating set {very poor, poor, medium poor, fair, medium good, good, and very good} was assumed to evaluate the three alternatives (models) based on the criteria. The ratings of each alternative (models) and the weight of each criteria are described by linguistic variables which can be expressed in triangular fuzzy numbers (Chen, 2001), as shown in Tables 4 and 5.

ble 4
eights to evaluate importance of each criteria

ry low	(0, 0, 0.1)
w	(0, 0.1, 0.3)
edium low	(0.1, 0.3, 0.5)
edium	(0.3, 0.5, 0.7)
edium high	(0.5, 0.7, 0.9)
gh	(0.7, 0.9, 1)
ry high	(0.9, 1, 1)

ble 5
tings to evaluate the three alternatives

ry poor	(0, 0, 1)
or	(0, 1, 3)
edium poor	(1, 3, 5)
ir	(3, 5, 7)
edium good	(5, 7, 9)
od	(7, 9, 10)
ry good	(9, 10, 10)

ble 6a
portance weights of criteria

	D1	D2	D3
	(0.9, 1, 1)	(0.9, 1, 1)	(0.7, 0.9, 1)
A	(0.9, 1, 1)	(0.7, 0.9, 1)	(0.7, 0.9, 1)
A	(0.9, 1, 1)	(0.5, 0.7, 0.9)	(0.5, 0.7, 0.9)
	(0.3, 0.5, 0.7)	(0.7, 0.9, 1)	(0.9, 1, 1)
U	(0.7, 0.9, 1)	(0.7, 0.9, 1)	(0.7, 0.9, 1)
	(0.7, 0.9, 1)	(0.7, 0.9, 1)	(0.5, 0.7, 0.9)
	(0.3, 0.5, 0.7)	(0.5, 0.7, 0.9)	(0.7, 0.9, 1)

te: (1) validation and calibration (VC), (2) model availability and user pport (MA), (3) availability of input data (ID), (4) user friendliness F), (5) model usage (MU), (6) predictive results (PR) and (7) lependent replication (IR).

ble 6b
zzy weights of the criteria

iteria	fuzzy weights
	(0.83, 0.97, 1.00)
A	(0.77, 0.93, 1.00)
	(0.63, 0.80, 0.93)
	(0.63, 0.80, 0.90)
U	(0.70, 0.90, 1.00)
	(0.63, 0.83, 0.97)
	(0.50, 0.70, 0.87)

te: (1) validation and calibration (VC), (2) model availability and user pport (MA), (3) availability of input data (ID), (4) user friendliness F), (5) model usage (MU), (6) predictive results (PR) and (7) lependent replication (IR).

Using values in Table 5 the importance of each criteria ven by decision makers D1, D2, D3 is shown in Tables 6a d b which show the fuzzy weights of the criteria calculated as

$$X_{ij} = \frac{1}{K}\left[X_{ij}^{1}(+)X_{ij}^{2}(+)\cdots(+)X_{ij}^{K}\right], \qquad (12)$$

where X_{ij}^{K} is the rating of the Kth decision maker.

The decision makers (D1, D2, and D3) use the linguistic rating variables (Table 5) to evaluate the rating of alternatives with respect to each criterion as shown in Table 7. With the aid of Table 7, a fuzzy decision matrix can be constructed as shown in Table 8.

With the aid of a fuzzy decision matrix, a fuzzy normalized decision matrix was constructed as shown in Table 9. The normalized fuzzy decision matrix was constructed using the following equation:

$$r_{ij} = \left(\frac{a_{ij}}{c_j^*}, \frac{b_{ij}}{c_j^*}, \frac{c_{ij}}{c_j^*}\right), \qquad (13)$$

$$c_j^* = \max c_{ij},$$

where a, b, and c represent the triangular fuzzy numbers of the decision matrix, $i = 1, 2, \ldots, n$ (alternatives with respect to criterion) and $j = 1, 2, \ldots, m$ (weights of criterion).

The normalized method mentioned above preserves the property that the ranges of normalized fuzzy numbers

Table 7
Ratings of three alternatives by decision makers under all criteria

Criteria	Candidate	Decision makers		
		D1	D2	D3
VC	PRZM-3	(9, 10, 10)	(9, 10, 10)	(7, 9, 10)
	PESTAN	(3, 5, 7)	(3, 5, 7)	(3, 5, 7)
	IPTM	(0, 1, 3)	(3, 5, 7)	(1, 3, 5)
MA	PRZM-3	(7, 9, 10)	(7, 9, 10)	(9, 10, 10)
	PESTAN	(7, 9, 10)	(7, 9, 10)	(9, 10, 10)
	IPTM	(3, 5, 7)	(3, 5, 7)	(1, 3, 5)
ID	PRZM-3	(9, 10, 10)	(7, 9, 10)	(5, 7, 9)
	PESTAN	(0, 0, 1)	(7, 9, 10)	(5, 7, 9)
	IPTM	(0, 0, 1)	(5, 7, 9)	(3, 5, 7)
UF	PRZM-3	(5, 7, 9)	(5, 7, 9)	(5, 7, 9)
	PESTAN	(5, 7, 9)	(5, 7, 9)	(7, 9, 10)
	IPTM	(5, 7, 9)	(5, 7, 9)	(7, 9, 10)
MU	PRZM-3	(9, 10, 10)	(9, 10, 10)	(7, 9, 10)
	PESTAN	(3, 5, 7)	(3, 5, 7)	(5, 7, 9)
	IPTM	(0, 1, 3)	(0, 0, 1)	(1, 3, 5)
PR	PRZM-3	(3, 5, 7)	(5, 7, 9)	(0, 0, 1)
	PESTAN	(3, 5, 7)	(5, 7, 9)	(0, 0, 1)
	IPTM	(1, 3, 5)	(3, 5, 7)	(0, 0, 1)
IR	PRZM-3	(5, 7, 9)	(7, 9, 10)	(0, 0, 1)
	PESTAN	(3, 5, 7)	(5, 7, 9)	(0, 0, 1)
	IPTM	(0, 1, 3)	(0, 1, 3)	(0, 0, 1)

Note: (1) validation and calibration (VC), (2) model availability and user support (MA), (3) availability of input data (ID), (4) user friendliness (UF), (5) model usage (MU), (6) predictive results (PR) and (7) independent replication (IR).

Table 8
Fuzzy decision matrix

	VC	MA	IDA	UF	MU	PR	IR
PRZM-3	(8.3, 9.6, 10)	(7.7, 9.3, 10)	(7, 8.6, 9.6)	(5, 7, 9)	(8.3, 9.6, 10)	(2.6, 4, 5.6)	(4, 5.3, 6.6)
PESTAN	(3, 5, 7)	(7.7, 9.3, 10)	(4, 5.3, 6.6)	5.6, 7.6, 9.3)	3.6, 5.6, 7.6)	(2.6, 4, 5.6)	(2.6, 4, 5.6)
IPTM	(1.3, 3, 5)	(2.3, 4.3, 6.3)	(2.7, 4, 5.6)	(5.6, 7.6, 9.3)	(0.3, 1.3, 3)	(1.3, 2.6, 4.3)	(0, 0.67, 2.3)

Note: (1) validation and calibration (VC), (2) model availability and user support (MA), (3) availability of input data (ID), (4) user friendliness (UF), (5) model usage (MU), (6) predictive results (PR) and (7) independent replication (IR).

Table 9
Fuzzy normalized decision matrix

	VC	MA	ID	UF	MU	PR	IR
PRZM-3	(0.83, 0.97, 1)	(0.8, 0.93, 1)	(0.7, 0.9, 1)	(0.5, 0.8, 0.96)	(0.83, 0.97, 1)	(0.4, 0.7, 1)	(0.6, 0.8,1)
PESTAN	(0.3, 0.5, 1)	(0.8, 0.93, 1)	(0.4, 0.5, 0.7)	(0.6, 0.8, 1)	(0.37, 0.57, 0.77)	(0.4, 0.7, 1)	(0.4, 0.6, 0.85)
IPTM	(0.13, 0.3, 1)	(0.2, 0.43, 1)	(0.3, 0.4, 0.6)	(0.6, 0.8, 1)	(0.03, 0.13, 0.3)	(0.23, 0.4, 1)	(0, 0.1, 0.35)

Note: (1) validation and calibration (VC), (2) model availability and user support (MA), (3) availability of input data (ID), (4) user friendliness (UF), (5) model usage (MU), (6) predictive results (PR) and (7) independent replication (IR).

belong to [0, 1]. Considering the different importance of each criterion, the final fuzzy evaluation values of each alternative were calculated using the following equation:

$$P_i = \sum_{j=1}^{n} r_{ij}(\bullet)w_j \quad i = 1, 2, \ldots, m, \tag{14}$$

where P_i is the final fuzzy evaluation value of ith alternative and w_j are the fuzzy weights of the criteria. The final fuzzy evaluation values of the three alternatives are:

$$\text{PRZM-3(A1)} = (3.26 \quad 5.14 \quad 6.62),$$
$$\text{PESTAN(A2)} = (2.24 \quad 3.97 \quad 5.70),$$
$$\text{IPTM(A3)} = (1.02 \quad 2.26 \quad 3.91).$$

After the calculation of the final fuzzy evaluation values for each alternative, a pair-wise comparison of the preference relationships between the alternatives was established and a fuzzy preference relation matrix was constructed as $E = [e_{ij}]$.

The e_{ij} is defined as

$$e_{ij} = S1/S2, \tag{15}$$

where

$$S1 = \int_{x>0} \mu(x)\mathrm{d}x \text{ and } S2 = \int_{x<0} \mu(x)\mathrm{d}x \quad (S1 + S2 = 1), \tag{16}$$

where μ is the membership function defined as

$$\mu = \begin{cases} 0 & x < n1, \\ (x-n1)/(n2-n1) & n1 <= x <= n2, \\ (x-n3)/(n2-n3) & n2 <= x <= n3, \\ 0 & x > n3. \end{cases} \tag{17}$$

The fuzzy preference relation matrix is given by

$$E = \begin{vmatrix} 0.50 & 0.76 & 0.98 \\ 0.24 & 0.50 & 0.87 \\ 0.02 & 0.13 & 0.50 \end{vmatrix}.$$

The fuzzy preference relation matrix represents the degree of preference of each pair of alternatives. For example the value in the 1st row, 3rd column is 0.98 and it represents the degree of preference of PRZM-3 over IPTM. Using the fuzzy preference matrix a fuzzy strict preference relation matrix can be constructed using the following equation:

$$E^{s} = [e_{ij}^{s}], \tag{18}$$

where

$$e_{ij}^{s} \begin{array}{ll} = e_{ij} - e_{ji} & (\text{when } e_{ij} > e_{ji}), \\ = 0 & (\text{otherwise}). \end{array}$$

The fuzzy strict matrix is

$$E^{s} = \begin{vmatrix} 0 & 0.52 & 0.96 \\ 0 & 0 & 0.74 \\ 0 & 0 & 0 \end{vmatrix}.$$

For example, the value in the 1st row, 2nd column of the strict matrix is 0.52, which means 0.52 is the strict dominance of PRZM-3 over PESTAN. Then, the non-dominated degree of each alternative was determined using the fuzzy strict preference relation matrix as

$$\mu^{\mathrm{ND}}(A_i) = \min\{1 - e_{ij}^{s}\} = 1 - \max e_{ij}^{s}(\forall j \neq i). \tag{19}$$

The non-dominated degree of each alternative is there-re

μ^{ND}(PRZM-3) = 1,

μ^{ND}(PESTAN) = 0.48,

μ^{ND}(IPTM) = 0.04.

μ^{ND} is the non-dominated degree of each alternative. A ·ge value of μ^{ND} indicates that the alternate has a higher ın-dominated degree than the others. Then, μ^{ND} values e used to rank the alternatives. If the alternative has a gh μ^{ND} value, then top rank is given to that alternative. ıe ranking order of the three models is {PRZM-> {PESTAN} > {IPTM}. Therefore, PRZM-3 is the best ıtion for advanced precision farming study.

Simulation analysis

I. Simazine runoff flux

Simazine runoff flux values obtained from the PRZM ɔdel during the simulation period (1981–1990) were eraged and a graph was drawn to show the variation of the simazine loss in the study region (Fig. 15). Therefore, only a fraction of a percent of the annual pesticides application was lost in the runoff.

5.2. Soil decay of simazine

Fig. 16 shows the amount of simazine pesticide that decayed in the soil during the simulation period (1981–1990). It can be inferred from the graph that most of the applied simazine pesticide decayed in the soil. Only a fraction of the annual pesticide application was leached into the soil, because the half-life of simazine in soil ranges from 28 to 60 days (Extension Toxicology Network, 1996).

5.3. Soil pesticide volatilization

Fig. 17 shows the amount of simazine volatilized from the soil during the simulation period (1981–1990). It can be inferred from the graph that a very small fraction of the applied pesticide volatilized from the soil. Simazine is subject to decomposition by ultraviolet radiation, but this effect is small under normal field conditions. Thus loss

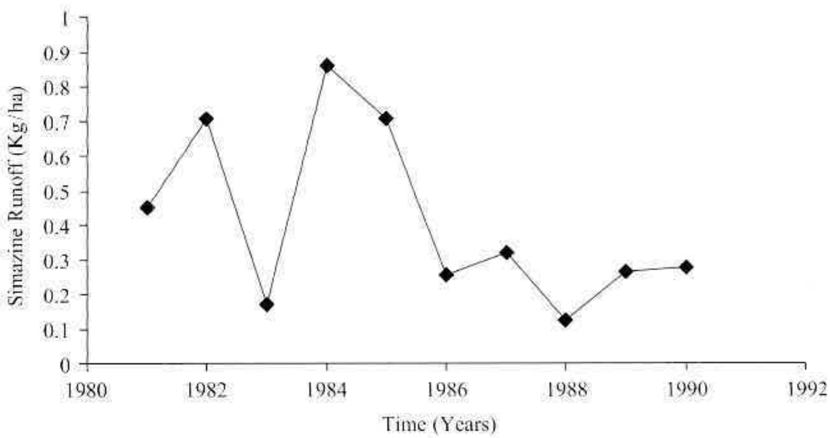


Fig. 15. Annual simazine runoff flux in the study region.

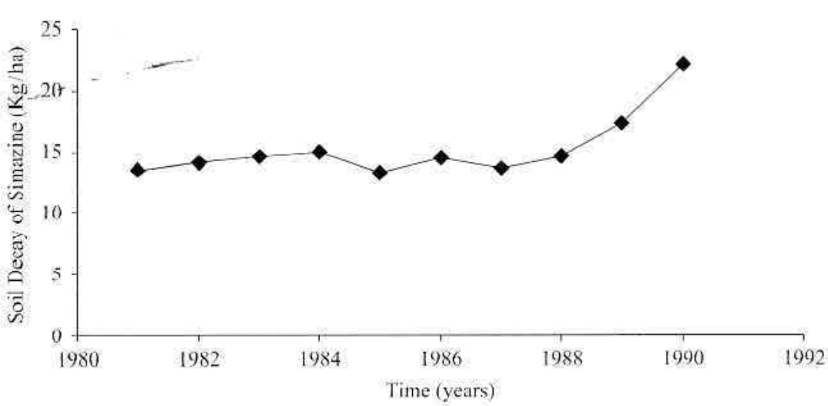


Fig. 16. Annual soil decay of simazine in the study region.

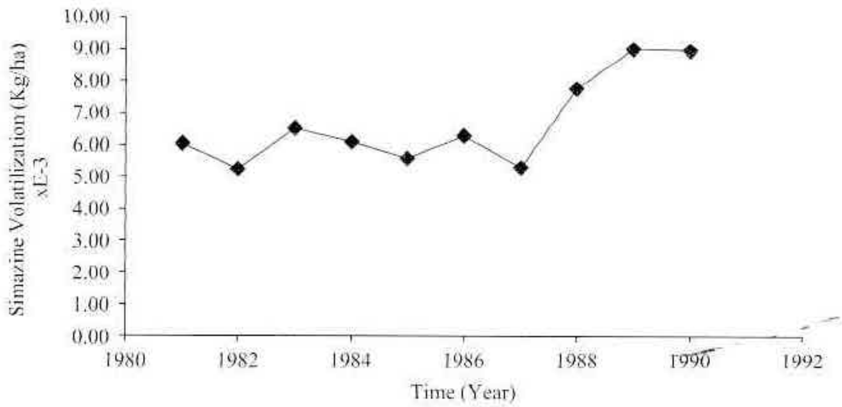


Fig. 17. Annual simazine volatilization in the study region.

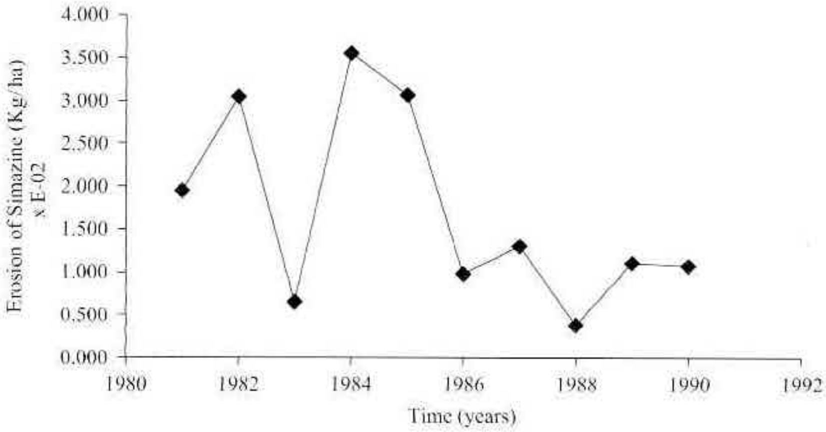


Fig. 18. Annual simazine erosion flux in the study region.

from volatilization is insignificant (Extension Toxicology Network, 1996).

5.4. Pesticide erosion

Fig. 18 shows the annual simazine pesticide erosion flux values obtained from the PRZM-3 model during the simulation period (1981–1990). Chemical runoff and erosion from agricultural management systems are of importance. Based on the graph, only a fraction of a percent of the annual pesticide application was lost in the erosion.

6. Conclusion

This analysis presents a unique comparative study in which simazine pesticide concentrations calculated from the PRZM-3, PESTAN and IPTM models were compared. Concentration values obtained from PRZM-3 are in agreement with the concentration values obtained from the other two models. All of the concentration values followed a decreasing trend from the surface through the vadose zone; and as the depth increased the concentration of simazine decreased. Thus the overall performances of the three models are similar in terms of the decay trend. With the aid of the FFADM method, it can be concluded that PRZM-3 is the suitable option for applications in advanced precision farming studies when compared to the other two models. Such applications are discussed in a subsequent paper (Du et al., 2006).

Acknowledgments

The authors are grateful to the Texas Higher Education Coordinating Board (THECB) for funding this project.

References

Brown, C.D., Hodgkinson, R.A., Rose, D.A., Syers, J.K., Wilcockson, S.J., 1995. Movement of pesticides to surface waters from a heavy clay soil. Pesticide Science 43, 31–140.

Bryant, K.J., Lace, R.D., Robinson, J.R.C., Norman Jr., J.W., Sparks Jr., A.N., Bremer, J.E., 1993. Report, T. W. R. I. T., Ed. no. 157, 16pp.

Carlsson, C., Fuller, R., 1996. Fuzzy multiple criteria decision making: recent developments. Fuzzy Sets and Systems 78, 139–153.

Carsel, R.F., Imhoff, J.C., Hummel, P.R., Cheplick, J.M., Donigian, A.S., 2003. PRZM-3, a Model for Predicting Pesticide and Nitrogen Fate in

the Crop Root and Unsaturated Soil Zones: Users Manual for Release 3.12.USEPA.
ıng, P.L., Chen, Y.C., 1994. Fuzzy multi-criteria decision making method for technology transfer strategy selection in biotechnology. Fuzzy Sets and Systems 63, 131–139.
ın, C.T., 2001. A fuzzy approach to select the location of the distribution center. Fuzzy Sets and Systems 118, 65–73.
ın, S.J., Hwang, C.L., 1993. Fuzzy Multiple Attribute Decision-making Methods and Applications. Lecture Notes in Economics and Mathematical Systems, vol. 375. Springer, Heidelberg, Germany.
ou, H.K., Tzeng, G.H., 2002. Fuzzy multiple criteria decision making approach for industrial green engineering. Environmental Management 30 (6), 816–830.
ı, X.F., 2004. Integrated Pesticide Transport Model (IPTM). Version 1.00b. User's manual 〈http://www.gvsu.edu/wri/envbio/chu/iptm.htm〉, accessed in October 2004.
ı, X.F., Marino, M.A., 2004. Semi-discrete pesticide transport modeling and application. Journal of Hydrology 285, 19–40.
ı, X.F., Basagaoglu, H., Marion, M.A., Volker, R.E., 2000. Aldicarb transport in the subsurface environment: Comparison of models. Journal of Environmental Engineering, ASCE 126 (2), 121–129.
ıen, S.Z., Wauchope, R.D., Klein, A.W., Eadsforth, C.V., Graney, R., 1995. Offsite transport of pesticides in water: mathematical models of pesticide leaching and runoff. Pure and Applied Chemistry 67 (12), 2109–2148.
Q., Chang, N.B., Srilakshmi Kanth, R., 2006. Combination of multispectral remote sensing, variable rate technology and environmental modeling for citrus pest management. Journal of Environmental Management, in press, doi:10.1016/j.jenvman.2006.11.019.
.n, J.D., Huyakorn, P.S., Donigian, A.S., Voss, K.A., Schanz, R.W., Meeks, Y.T., Carsel, R.F., 1989. Risk of Unsaturated/Saturated Transport and Transformation of Chemical Concentrations (RUSTIC), Theory and Code Verification, vol. I. US Environmental Protection Agency, Athens, GA EPA-600/3-89/048a.
ıigian Jr., A.S., Rao, P.S.C., 1986. In: Hern, S.C., Melancon, S.M. (Eds.), Vadose Zone Modeling of Organic Pollutants. Lewis Publishers, Chelsea, MI (Chapter 1).
ension Toxicology Network, 1996. Pesticide Information Profiles (Simazine). Available at 〈http://extoxnet.orst.edu/pips/simazine.htm〉.
ıltieri, C., 2002. Comparison between Water Quality Models for Toxics. In: Rizzoli, A.E., Jakeman, A.J. (Eds.), Proceedings CD of the 2002 Integrated Assessment and Decision Support Conference. Lugano Switzerland, 24–27 June 2002, International Environmental Modelling and Software Society, Lugano, Switzerland.
ıng, C.L., Yoon, K., 1981. Multiple Attributes Decision Making Methods and Applications. Springer, Berlin.
, W.A., Spencer, W.F., Farmer, W.J., 1983. Behavior assessment model for trace organics in soil: I. Model description. Journal of Environmental Quality 12 (4), 558–564.
), R.J., Chi, S.C., Kao, S.S., 2002. A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network. Computers in Industry 47, 199–214.
nard, R.A., Knisel, W.G., Still, D.A., 1987. ~~GL~~EAMS: groundwater loading effects of agricultural management systems. Trans. ASAE 30, 1403–1418.
ıg, J.Y., Martien, P.T., Soong, S.T., Tanrikulu, S., 2004. A photochemical model comparison study: CAMx and CMAQ performance in central California. In: Proceedings CD of 13th Conference on the Applications of Air Pollution Meteorolgy with the Air and Waste Management Association, San Francisco, CA, 23–25 August 2004.
guc, K., Anh Nguyen, D.L., Davis, S.N., Abrams, R.H., 1998. A case study simulation of DBCP groundwater contamination in Fresno County, California 1. Leaching through the unsaturated subsurface. Journal of Contaminant Hydrology 29, 109–136.
National Center for Food and Agricultural Policy (NCFAP), 1997. Pesticide use database 〈http://pestdata.ncsu.edu/ncfap/search.cfm〉, accessed in October 2004.
Pavic, I., Babic, Z., 1991. The use of the PROMETHEE method in the location choice of a production. International Journal Production Economics 23, 165–174.
Pierre, Y.C., Charlotte, B., Allen, C., 1996. Fate and impact of pesticides applied to potato cultures: the Nicolet River Basin. Ecotoxicology and Environmental Safety 33, 175–185.
Pimentel, D., Acquay, H., Biltonen, M., Rice, P., Silva, M., Nelson, J., Lipner, V., Giordano, S., Horowitz, A.D., Amore, M., 1992. Environmental and human costs of pesticide use. Bioscience 42, 750–760.
Prato, T., 1999. Multiple attributes decision analysis for ecosystem management. Ecological Economics 30, 207–222.
Reus, J., Leendertse, P., Bockstaller, C., Fomsgaard, I., Gutsche, V., Lewis, K., Nilsson, C., Pussemier, L., Trevisan, M., Van der Werf, H., Alfarroba, F., Blümel, S., Isart, J., McGrath, Seppälä, D., 2002. Comparison and evaluation of eight pesticide environmental risk indicators developed in Europe and recommendations for future use. Agriculture, Ecosystems and Environment 90, 177–187.
Rietveld, P., Ouwersloot, H., 1992. Ordinal data in multicriteria decision making, a stochastic dominance approach to sitting nuclear power plants. European Journal Operational Research 56, 249–262.
Spohrer, G.A., Kmak, T.R., 1984. Qualitative analysis used in evaluating alternative plant location scenarios. Industrial Engineering, 52–56.
Stephen, D.D., Steven, P., Porter, M.J., Anderson, G., Smith, W.G., Weingart, D., 1999. Applicability of a Simulation Model to Determine the Environmental Fate of Atrazine in the Canajoharie Creek Watershed. New York NYS Water Resources Institute, Cornell University Rice Hall, Ithaca, NY 14853 〈http://www.cfe.cornell.edu/wri/〉, accessed in October 2004.
Stewart, T.J., 1992. A critical survey on the status of multiple criteria decision making theory and practice. OMEGA—International Journal of Management Science 20, 569–586.
Teng, J.Y., Tzeng, G.H., 1996. Fuzzy multicriteria ranking of urban transportation investment alternatives. Transportion Planning Technology 20 (1), 15–31.
Tompkins, A.J., White, A., 1984. Facilities Planning. Wiley, New York, USA.
Tran, L.T., Knight, C.G., O'nell, R.V.L., Smith, E.R., Ritters, K.H., Wickham, J., 2002. Fuzzy decision analysis for integrated environmental vulnerability assessment of the Mid-Atlantic region. Environmental Management 29 (6), 845–859.
Trevisan, M., Errera, G., Goerlitz, G., Remy, B., Sweeney, P., 2000. Modelling ethoprophos and bentazone fate in a sandy humic soil with primary pesticide fate model PRZM-2. Agricultural Water Management 44, 317–335.
US Fish and Wildlife Service (US FWS), 1986. Preliminary Survey of Contaminant Issues of Concern on National Wildlife Refuges. Division of Refuge Management, Washington, DC, p. 162.
Vredeveld, G., Bullard, R., Sells, M., Sims, S., West, J., 1983. Energy comparison in three cases of pesticide versus bio-control pest management. Agriculture, Ecosystems and Environment 9, 51–56.
Wagenet, R.J., Huston, J.L., 1989. LEACHM: Leaching Estimation and Chemistry Model. Version 2, vol. 2. Continuum Water Resources Inst., Cornell University, Ithaca, NY.
Walker, A., 1987. CALF: calculation flow model: version 1. Weed Research 27, 143–152.
Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8, 338–353.
Zeleny, L.A., 1982. Multiple Criteria Decision Making. McGraw-Hill, New York.

Journal of Environmental Management 86 (2008) 44–62

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Characteristics of coal mine ventilation air flows

Shi Su*, Hongwei Chen[1], Philip Teakle, Sheng Xue

CSIRO Exploration and Mining, P.O. Box 883, Kenmore, Qld 4069, Australia

Received 3 January 2006; received in revised form 13 November 2006; accepted 21 November 2006
Available online 18 January 2007

Abstract

Coal mine methane (CMM) is not only a greenhouse gas but also a wasted energy resource if not utilised. Underground coal mining is by far the most important source of fugitive methane emissions, and ~70% of all coal mining related methane is emitted to the atmosphere through mine ventilation air. Therefore, research and development on mine methane mitigation and utilisation now focuses on methane emitted from underground coal mines, in particular ventilation air methane (VAM) capture and utilisation. To date, most work has focused on the oxidation of very low concentration methane. These processes may be classified based on their combustion kinetic mechanisms into thermal oxidation and catalytic oxidation. VAM mitigation/utilisation technologies are generally divided into two basic categories: ancillary uses and principal uses. However, it is possible that the characteristics of ventilation air flows, for example the variations in methane concentration and the presence of certain compounds, which have not been reported so far, could make some potential VAM mitigation and utilisation technologies unfeasible if they cannot cope with the characteristics of mine site ventilation air flows. Therefore, it is important to understand the characteristics of mine ventilation air flows. Moreover, dust, hydrogen sulphide, sulphur dioxide, and other possible compounds emitted through mine ventilation air into the atmosphere are also pollutants. Therefore, this paper presents mine-site experimental results on the characteristics of mine ventilation air flows, including methane concentration and its variations, dust loadings, particle size, mineral matter of the dust, and other compounds in the ventilation air flows. The paper also discusses possible correlations between ventilation air characteristics and underground mining activities.

Keywords: Coal mine methane; Greenhouse gas; Waste energy; Dust emission; Hydrogen sulphide; Sulphur dioxide; Characteristic parameters

1. Introduction

Coal mine methane (CMM) is not only a greenhouse gas but also a wasted energy resource if not utilised. In 2000, it was estimated that the world total of methane emitted from mine ventilation air was over 237 Mt CO_2-e, and that ventilation air methane (VAM) derived power projects could theoretically create 3GW$_e$ of net usable capacity (US EPA, 2003). Underground mining is by far the most important source of fugitive methane emissions, and ~70% of all coal mining related methane is emitted through underground mine ventilation air, rather than as more highly concentrated drainage gas in advance of or immediately after mining (Moore et al., 1998; Sloss, 2005). Therefore, research and development on mine methane mitigation and utilisation now focuses on methane emitted from underground coal mines, especially VAM capture and utilisation, because (1) it represents most of the methane emission from coal mines; and (2) it is most difficult to capture and use as the air volume is large and the methane resource is dilute and variable in concentration and flow rate.

To date, most work has focused on the oxidation of very low concentration methane. These processes may be classified based on their combustion kinetic mechanisms into thermal oxidation and catalytic oxidation. VAM mitigation/utilisation technologies are generally divided into two basic categories: ancillary uses and principal uses, which have been reviewed by Su et al. (2005). No matter what technologies have been/are being developed for the ancillary and principal uses of VAM, an important issue related to the implementation of these technologies at mine sites has been ignored, namely the characteristics of the mine ventilation air flows. The characteristics, including

*Corresponding author. Tel.: +61 7 3327 4679; fax: +61 7 3327 4455.
E-mail address: shi.su@csiro.au (S. Su).
[1]Current address: North China Electric Power University, China.

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.11.025

ust loading, particle size, mineral matters of the dust, ariations in methane concentration and flow rate, and ther compounds (H_2S and SO_2), could make a potential 'AM mitigation and utilisation technology unfeasible if it annot cope with the characteristics of mine site ventilation ir flows. Therefore, it is important to understand the haracteristics of mine ventilation air flows. Moreover, ust, H_2S, SO_2, and other possible compounds emitted hrough mine ventilation air into the atmosphere are also ollutants.

To gather this information and measure the levels of hese emissions, an isokinetic particle sampling system vas specifically designed and constructed, which met oal mine intrinsic safety requirements and suited the nine ventilation air shafts. Field sampling trials at our mines were carried out, and dust and gas samples vere analysed using a variety of techniques. This aper presents the mine-site experimental results n characteristics of mine ventilation air flows, including nethane concentration and its variations, dust loadings, article size/distribution, mineral matter of the dust, nd other compounds in the ventilation air flows. 'he paper also discusses possible correlations between entilation air characteristics and underground mining ctivities.

2. Experiments

2.1. Isokinetic sampling system and sampling procedure

To develop an isokinetic dust sampling system suited to mine ventilation air shafts, it is necessary to understand mine ventilation air shafts and major isokinetic sampling principles before designing and fabricating the isokinetic sampling train, in particular one which can meet coal mine intrinsic safety requirements. Since there remained many unknowns at the outset, the design had to be readily adaptable to the various flow speeds, port sizes and dust loadings at each site.

The data collection was done in collaboration with four mines located in eastern Australia, designated mines A, B, C and D. At each mine the vertical ventilation air shaft (upcast shaft) was topped with a 90° elbow which redirected the ventilation air flow horizontally. The flow was split into two or three branches, each with its own fan and evasee. Following preliminary mine site visits it was concluded that the most suitable sampling locations were near the upcast shaft collars because in each case the shafts were fitted with several sampling ports at a convenient height, and the ports provided access to the entire ventilation air flow before it was branched.

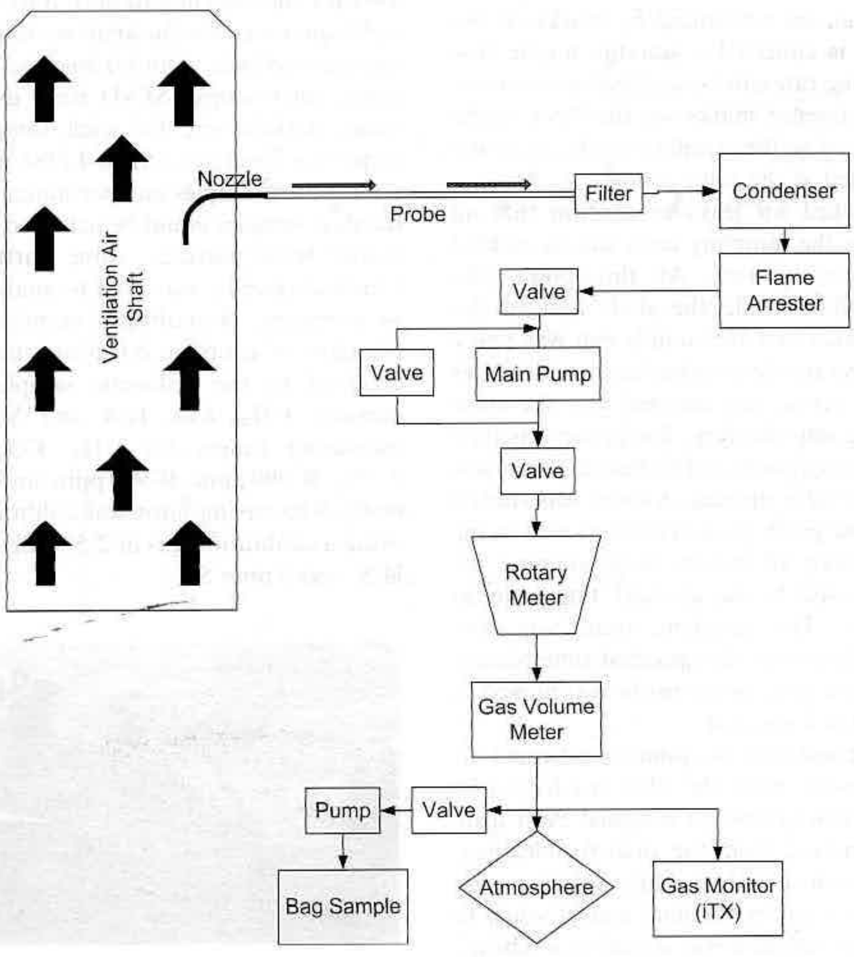


Fig. 1. Schematic design of the Isokinetic sampling train adapted in this study.

A sampling train with an in-stack collector is superior to out-stack collection as the latter is susceptible to dust deposition inside the sampling tube. The nozzle and probe of an out-stack collector need to be thoroughly washed with a solution to remove any particulate matter stuck to the tube and nozzle according to AS 4323.2-1995 (Standards Australia, 1995a), and the washed particles should be counted in the calculation of dust loading for each sample as discussed later. However, since the sampling ports on the ventilation air shafts of most of the mines were smaller than the particulate filter holder, out-stack collection was required. The relative humidity of the ventilation air in most of these mines was about 70–100%, however probe heating to prevent moisture deposition was not necessary because the ventilation air temperature was less than the ambient air temperature. Therefore, the sampling train selected for the mine site trials corresponded to Type D in AS 4323.2-1995 (Standards Australia, 1995a). It was a general-purpose configuration, which provided maximum flexibility for sampling at different mines and was capable of out-stack sampling (Refer Fig. 1).

The dust isokinetic sampling procedure was based on AS 4323.1-1995 (Standards Australia, 1995b) and US EPA Method 5. Before the sampling can begin, a velocity profile for the shaft needs to be determined. This is achieved by using a type S Pitot tube (US EPA Method 5) at designated points in the shaft which are referenced by marks on the probe. Once this data is collected a suitable nozzle that gives the desired sampling rate can be selected based on the velocity profile. The reference marks on the Pitot probe also need to be transferred to the sampling probe to ensure that samples are collected at the same points.

The system was checked for leaks to confirm that all seals were working after the sampling train was assembled and before every sampling run. At this point, the temperature and humidity inside the shaft and of the atmosphere were measured and the sample run was ready to begin. With the correct nozzle attached and a clean filter in the filter holder the probe was inserted into the shaft with the nozzle pointing into the flow. The probe was then secured at its first sampling point and the main pump was turned on and the main valve opened. A timer was started to keep track of when the probe needed to be moved to the next sampling point. Once all systems were running the rotary meter was adjusted to the desired flow rate to achieve isokinetic flow. The sampling train was now running and collected data over the specified time period, normally 1 h. At each time interval the probe was moved to the next point, checked and secured.

At the end of the specified time the pump was turned off and the main valve closed. First the filter needed to be carefully removed and placed into its assigned Petri dish. Then the probe was removed from the shaft for cleaning. The probe was then washed with a dilute solution of a cleaning agent to remove any particulate matter stuck to the tube and nozzle. This solution was stored in a labelled container for filtration at the laboratory. Once the probe was thoroughly washed it was dried by passing dry nitrogen through it. Once these steps were completed the system was ready for the next sampling run. It should be pointed out that the dust loading was determined from dusts both deposited on the filter paper and washed from inside the tube and nozzle.

It was not possible to select sampling plane positions which met the ideal sampling criteria specified in Australian Standard 4323.1-1995 (Standards Australia, 1995b). Hence, the non-ideal sampling plane position, as defined in the Standard, was attempted. However, this still required three sampling traverses and six access holes to perform the isokinetic sampling process. Unfortunately, the number of available access holes at mine sites was often fewer because some were either occupied for permanent monitoring equipment, located too close to the surrounding structures, or the plugs rusted in place. As a result, the isokinetic sampling process was not usually fully compliant with the Standard at the mine sites. Therefore, best practise for the mine site sampling was that dust samples were taken from 10 sampling points of an accessible radius.

2.2. *Analytic techniques*

Besides the above-described isokinetic sampling train used for the determination of dust loadings, some analytic techniques used for the analysis of dust and gas samples are summarised here. Optical microscopy and scanning electronic microscopy (SEM) were used to determine maximum particle size for each sample. SEM with energy dispersive spectrum (SEM-EDS) was used to analyse 24 typical dust samples (six per mine) so an understanding of the dust samples could be achieved in terms of the mineral matter (coal particles, stone particles, etc.). Micro Gas Chromatography was used to analyse the gas bag samples as necessary. A multi-gas monitor (iTX), based on the principle of catalytic diffusion and electrochemistry, was attached to the isokinetic sampling train and used to measure CH_4, CO, H_2S and SO_2 concentrations. Its measuring ranges for CH_4, CO, H_2S and SO_2 were 0–5%, 0–999 ppm, 0–999 ppm and 0.2–99.9 ppm, respectively. The gas monitor was calibrated before each run by using a calibration gas of 2.5% CH_4, 100 ppm CO, 25 ppm H_2S, and 5 ppm SO_2.

Fig. 2. Isokinetic sampling trails at mine B.

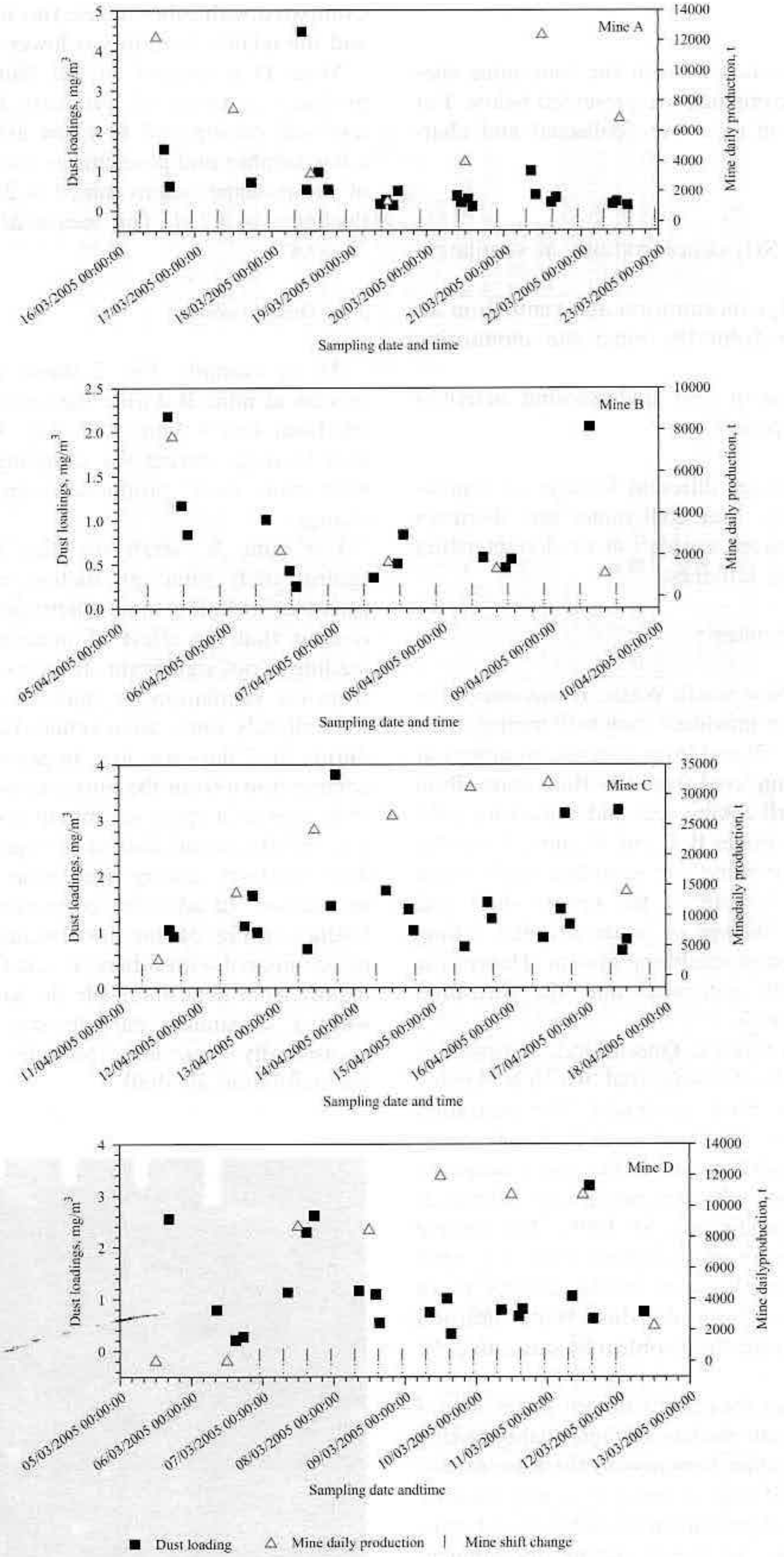


Fig. 3. Dust loadings at four mines.

3. Results

Most of the results obtained from the four mine sites during the field testing campaign are presented below. For each mine, the following data were collected and characterised:

- dust loadings,
- CH_4, CO, H_2S and SO_2 concentrations in ventilation air,
- retrieved data of CH_4 concentration and ventilation air flow rate each hour from the mine site monitoring system,
- mine layout information and underground activities during the sampling period.

This paper then addresses different features of ventilation air flows among the four coal mines and discusses possible correlations between ventilation air characteristics and underground mining activities.

3.1. Information on coal mines

Mine A is located in New South Wales. It was one of the first mines in Australia to introduce long wall mining. This mine operates one longwall and three continuous miners at depths of 550 m extracting coal from the Bulli seam. Bulli coal is prime quality hard coking coal and is used for coke making. Compared with mines B, C and D, mine A was the most humid. During the mine site sampling trials, team members went to the bottom of the upcast shaft and observed a continuous shower of water droplets falling from the shaft like a spray scrubbing system. Hence, the ventilation air was fully saturated and the measured relative humidity was 100%.

Mine B is located in Central Queensland. It produces low to medium volatile hard coking coal, which is blended with coal from a nearby mine for export. The maximum depth of cover is ~400 m. The coal seam thickness varies between 2.2 and 2.7 m and dips ~5° to the east. Compared with mine A, the ventilation air was not always saturated, however, the relative humidity was 85–100%. The relative humidity was 100% during site sampling from 4–9 April 2005. A lack of water in the seam results directly in an increased concentration of respirable dust. Water infusion has been trialed to overcome this problem by saturating the seam with water.

Mine C operates within the central Bowen Basin, one of the world's most important sources of high-quality coking and thermal coals. The seam thickness in the lease area is 2–3.5 m. The majority of coal at mine C is extracted by longwall mining, with ~9 percent of the total run-of-mine coal produced by continuous miners during development activities. The mine has two longwall units that are operated alternately to minimise downtime and ensure seamless production and reliability. Each longwall block is 250 m wide and the average length is more than 2 km. Compared with other mines, this mine is not a gassy mine, and the relative humidity is lower at 74.5–83.5%.

Mine D is located in the Hunter Valley. This mine produces a range of products from a highly volatile semi-soft coking coal to a low ash thermal coal, all with a low sulphur and phosphorus content. The average depth of the principal seams mined is 200 m, and their average thickness is 3.2 m. The measured relative humidity was 73–99.8%.

3.2. Dust loadings

As an example, Fig. 2 shows the mine site sampling process at mine B during the second sampling trial at this site from 4 to 9 April 2005. Fig. 3 presents the results of dust loadings during the sampling period for each mine with mine daily production, and indicates mine shift changes.

For mine A, when the dust loadings are examined against daily mine production rates and underground activities, including mine maintenance and stone dusting, it is clear that the effect of mine production on the dust loading is not significant. It seems that the dust emissions from the ventilation air shaft on 19 and 20 March 2005 were slightly lower than others because mine production during the 2 days was low. In general, the volume of water coming down from the upper section of the ventilation air shaft acts as a spray scrubbing system for the ventilation air, and this could lead to no significant difference in the dust loadings during the mine production and non-production. In addition, corresponding to sample A5 (as labelled in Fig. 5), the dust loading is very high, 4.47 mg/m^3, compared with others. It was found that the mass was mainly from deposits inside the sampling tube and nozzle with a maximum particle size of ~0.5 mm. Hence, occasionally some large particles can be carried out of the ventilation air shaft.

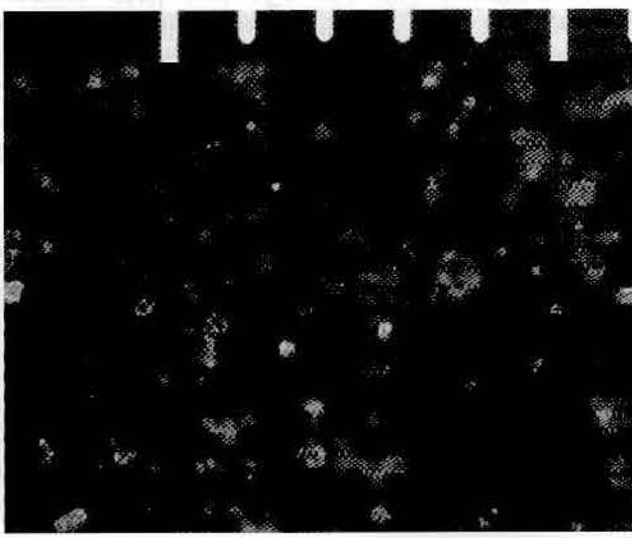


Fig. 4. Optical image (100 ×) of sample A21 (mine A).

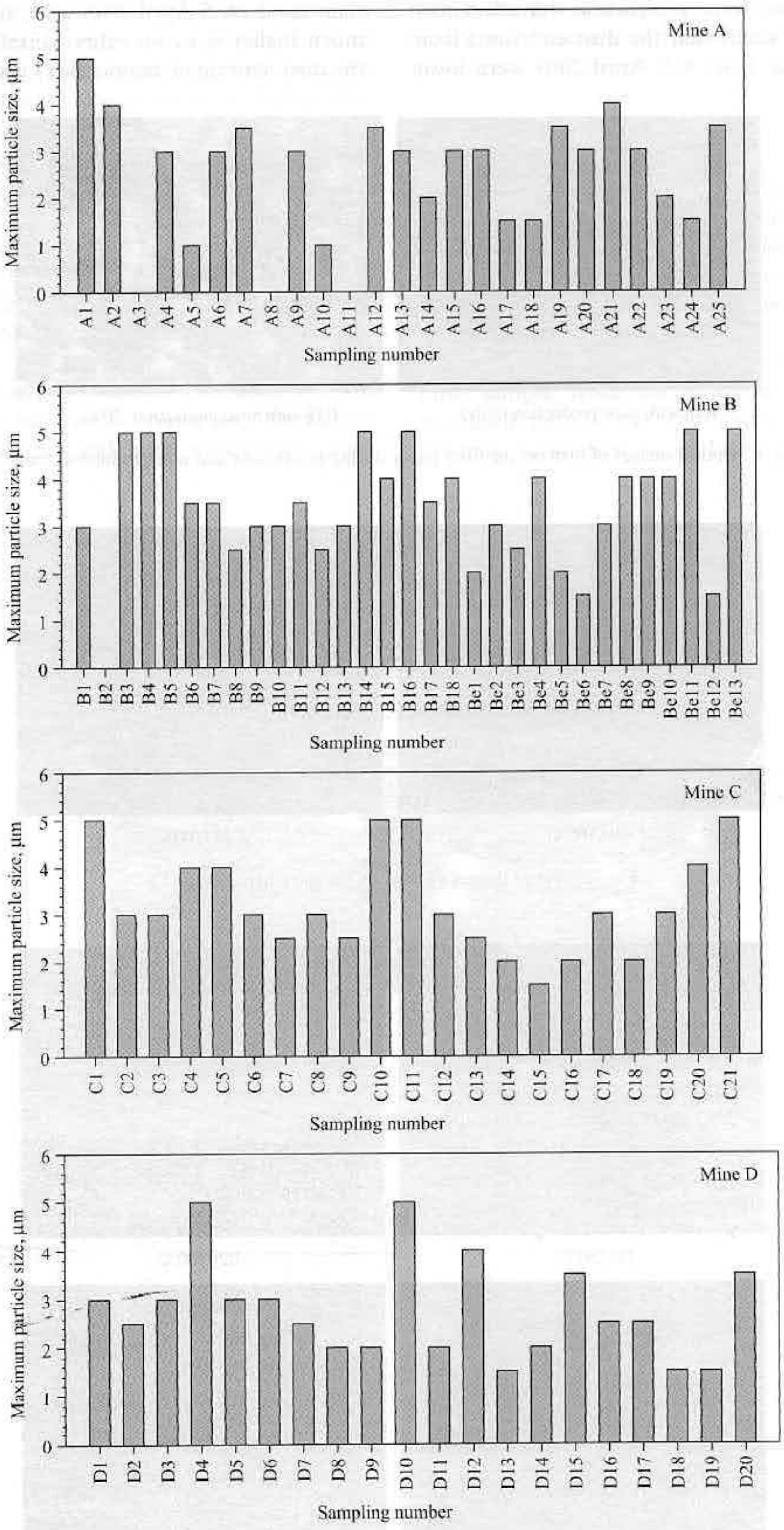
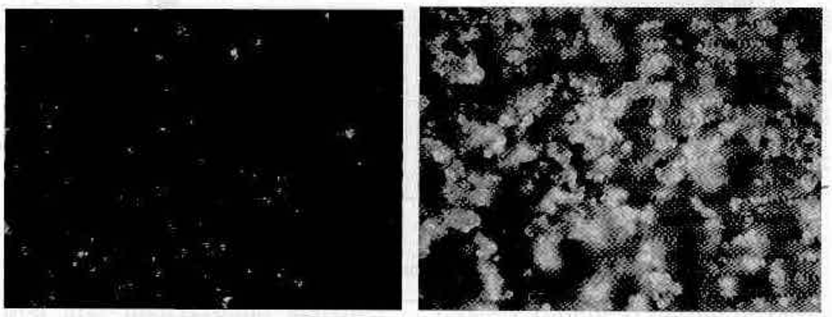


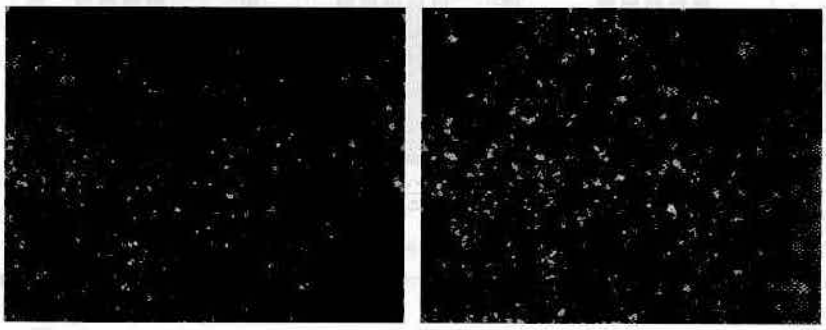
Fig. 5. Maximum particle size of each sample at the four mines.

For mine B, however, mine production did affect dust loading. In general, it seems that the dust emissions from the ventilation air shaft from 6–9 April 2005 were lower than those on 5 April 2005 when the mine production was much higher than on other sampling days. In particular, the dust loading of sample Be13 was a bit higher compared



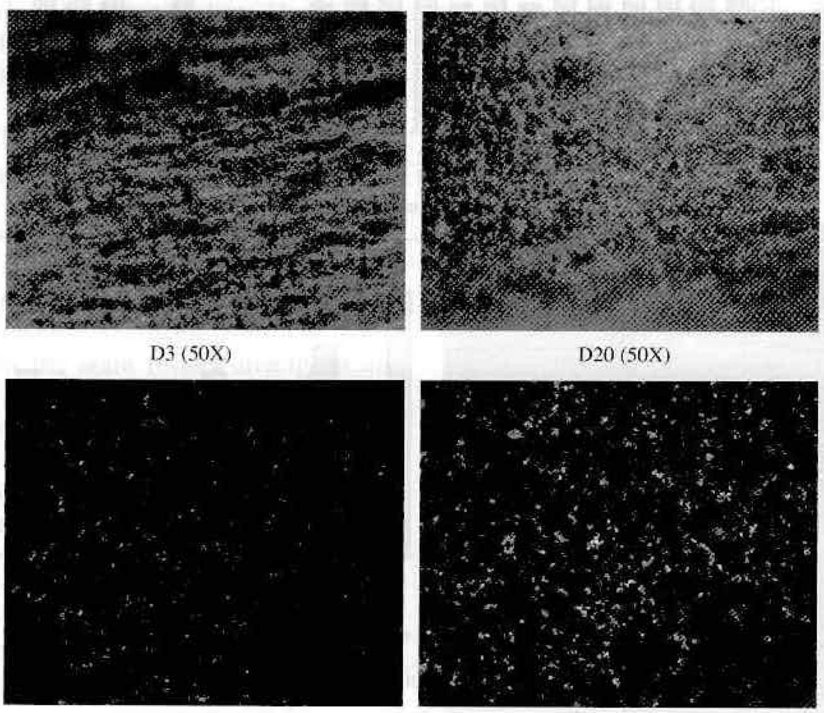
B18 with mine production (50X) B18 with mine production (50X)

Fig. 6. Optical images of dust on the filter paper during production and non-production (mine B).



C1 (100X) C18 (100X)

Fig. 7. Optical images of dust on the filter paper (mine C).



D3 (50X) D20 (50X)

D13 (50X) D16 (50X)

Fig. 8. Optical images of dust on the filter paper (mine D).

with other samples taken from 5–8 April 2005. During the collection of sample Be13, there was no production but maintenance work was being conducted, on occasion therefore, maintenance work could result in higher dust emissions.

For mine C, the dust loadings measured from 11–17 April 2005 and mine site production and maintenance data indicate a trend: higher mine production results in higher dust emissions when the three highest dust loading points are ignored from the figure. The dust loadings on 11 and 17 April 2005 were lower due to lower production rates compared with other production days.

Based on the dust loadings measured at mine D from 5–12 March 2005, it is hard to determine any correlation between the dust emissions and mine production rates. While there was no production on 6 March 2005, there could have been stone dusting that generated airborne particles.

3.3. Maximum particle size

As mentioned above, the maximum particle size for each sample was determined by using optical microscopy. Fig. 4 shows an optical micrograph of sample A21 as an example of how the particle size was measured. The particle sizes of some samples were also determined by the SEM images, shown in Figs. 9–15.

Fig. 5 summarises the maximum particle size for each dust sample from the four mines. In summary, the maximum particle size was 5 μm for the dust emitted from the mine ventilation air in this study.

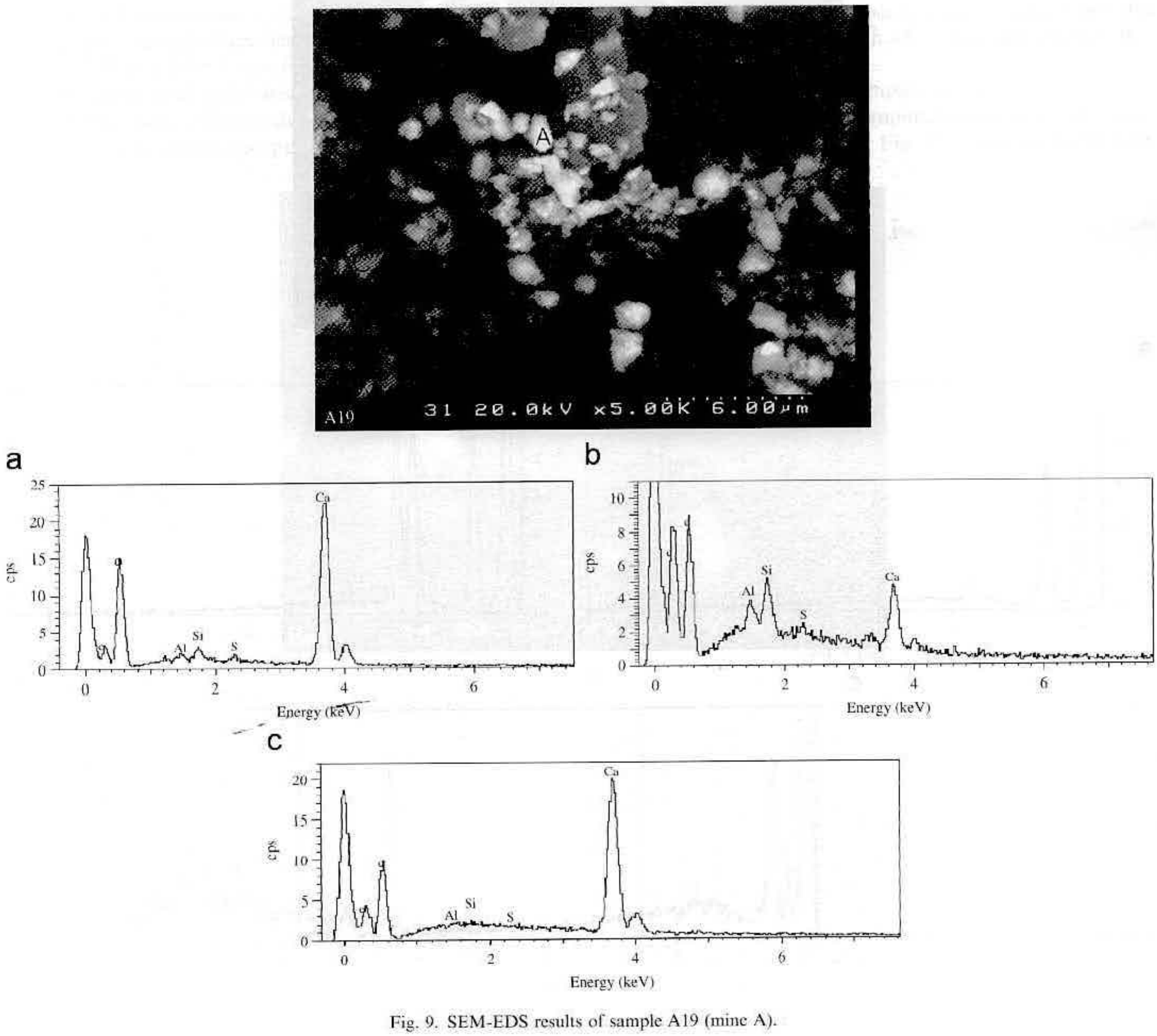


Fig. 9. SEM-EDS results of sample A19 (mine A).

3.4. Visual properties and mineral matter

An effort was made to determine physical visual differences between dust samples by using an optical microscope, and to understand the mineral matter of selected typical dust samples by using the SEM-EDS.

3.4.1. Visual properties

As outlined above, mine A is the most humid mine of the four mines tested. The high humidity caused some water to collect on the filter holder, so the dust was not deposited evenly on the membrane filter paper like the samples collected from the other mines. This made it very hard to get a reasonable visual assessment of the dust samples for mine A.

For mine B, there are obvious colour differences between the production and non-production samples. The colour of the dust samples collected during the mine production is darker than those collected during non-production. Fig. 6 shows the difference as an example from optical microscope images. Sample B18 was collected during the mine production on 15 November 2004 (Monday), and sample B13 during non-production on 14 November 2004 (Sunday), when stone dusting was carried out. It is apparent that the black dust samples contain more coal particles than the grey/white dust samples.

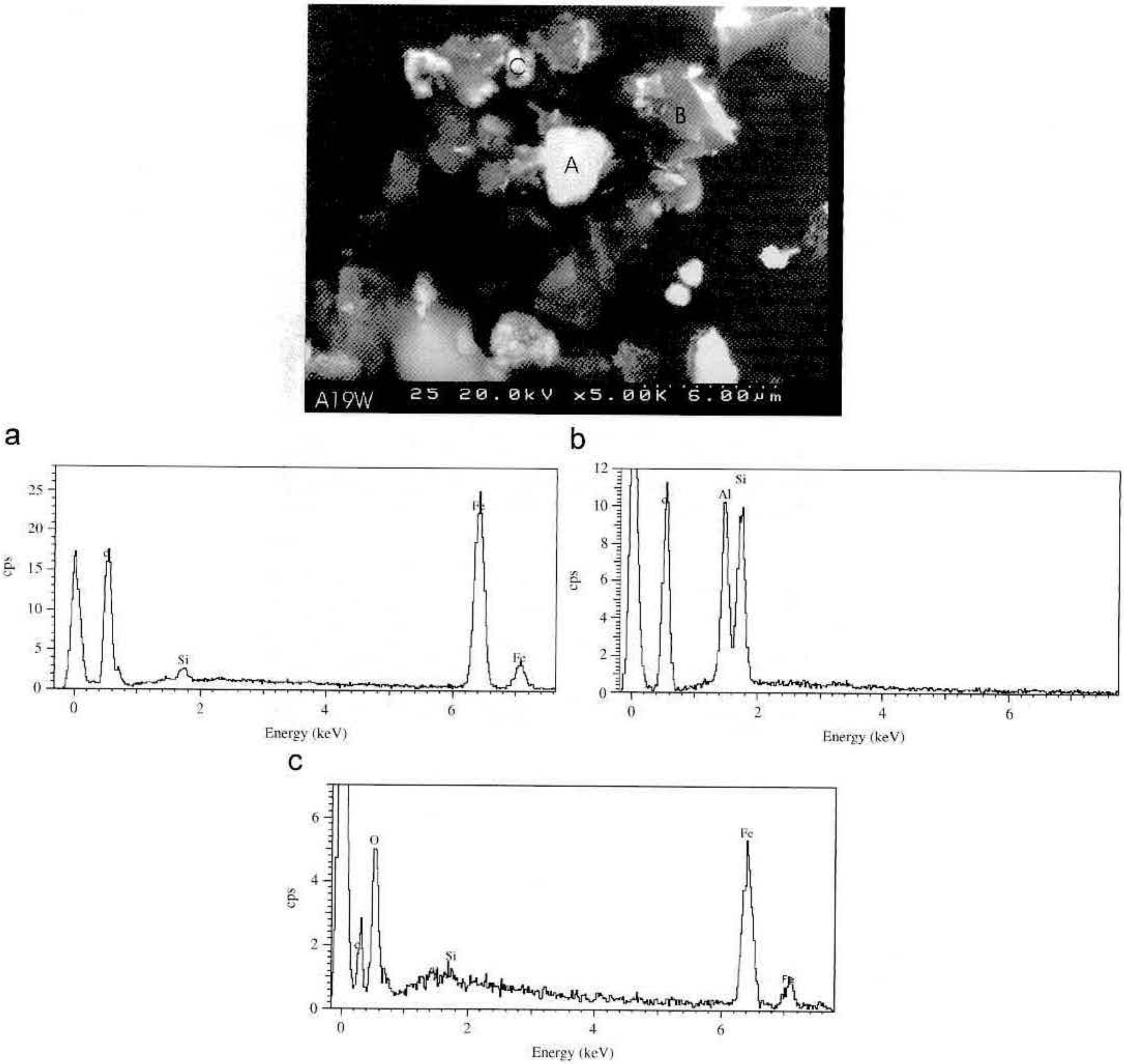


Fig. 10. SEM-EDS results of sample A19w (mine A).

For mine C, the mine produced coal every day during the sampling period with the lowest production of 2596 tonnes on 11 April 2005. At this mine two longwall units are operated alternately to minimise downtime and ensure seamless production and reliability. According to the visual assessment of 21 dust samples, they all appear to be brownish-black dust. Fig. 7 shows two examples of optical microscope images of dust. Sample C1 was collected on 11 April 2005 with the lowest production, and sample C18 was collected on 16 April 2005 with the highest production. There appears to be minimal physical difference between the two samples. This may be due to the relatively low humidity at mine C of between 74.5 and 83.5%.

Based on the visual assessment of 20 dust samples from mine D, colours appear to differ between mine production and no production. Fig. 8 shows four dust examples from optical microscope images. Sample D3 was collected on 6 March 2005 (Sunday), and D20 on 12 March 2005 (Saturday). There was no production when collecting these two samples though there was no production in early or late of 12 March 2005. From Fig. 8 it can be seen that there was little black coal in the dust samples, and it appears the two samples were collected during stone dusting. Clearly, during the mine production process more coal particles are contained in the dust samples such as D13 and D16 shown in Fig. 8.

3.4.2. Mineral matter

For mine A, six dust samples, including two samples of washed dust from the sampling probe and tube, were examined. Fig. 9 shows the SEM-EDS examination results of the dust deposit on the filter paper. From this figure, it was determined that particle B is coal with some ash, and particles A & C are calcium oxides. Fig. 10 shows the SEM-EDS examination results of the dust collected from the sampling probe and nozzle for the sample A19, and from this figure, particle A is iron oxide, and particle B is clay (Si–Al–O), and particle C is coal containing iron oxide.

For mine B, six dust samples, including one sample of washed dust from the sampling probe and tube, were examined by the SEM-EDS. Fig. 11 shows the SEM-EDS examination results of the dust collected from the sampling probe and nozzle for the sample Be2w, and from this figure, particle A is coal with some ash, and particle B is quartz (SiO_2).

For mine C, seven dust samples, including three samples of washed dust from the sampling probe and tube, were examined by the SEM-EDS. Fig. 12 shows the SEM-EDS

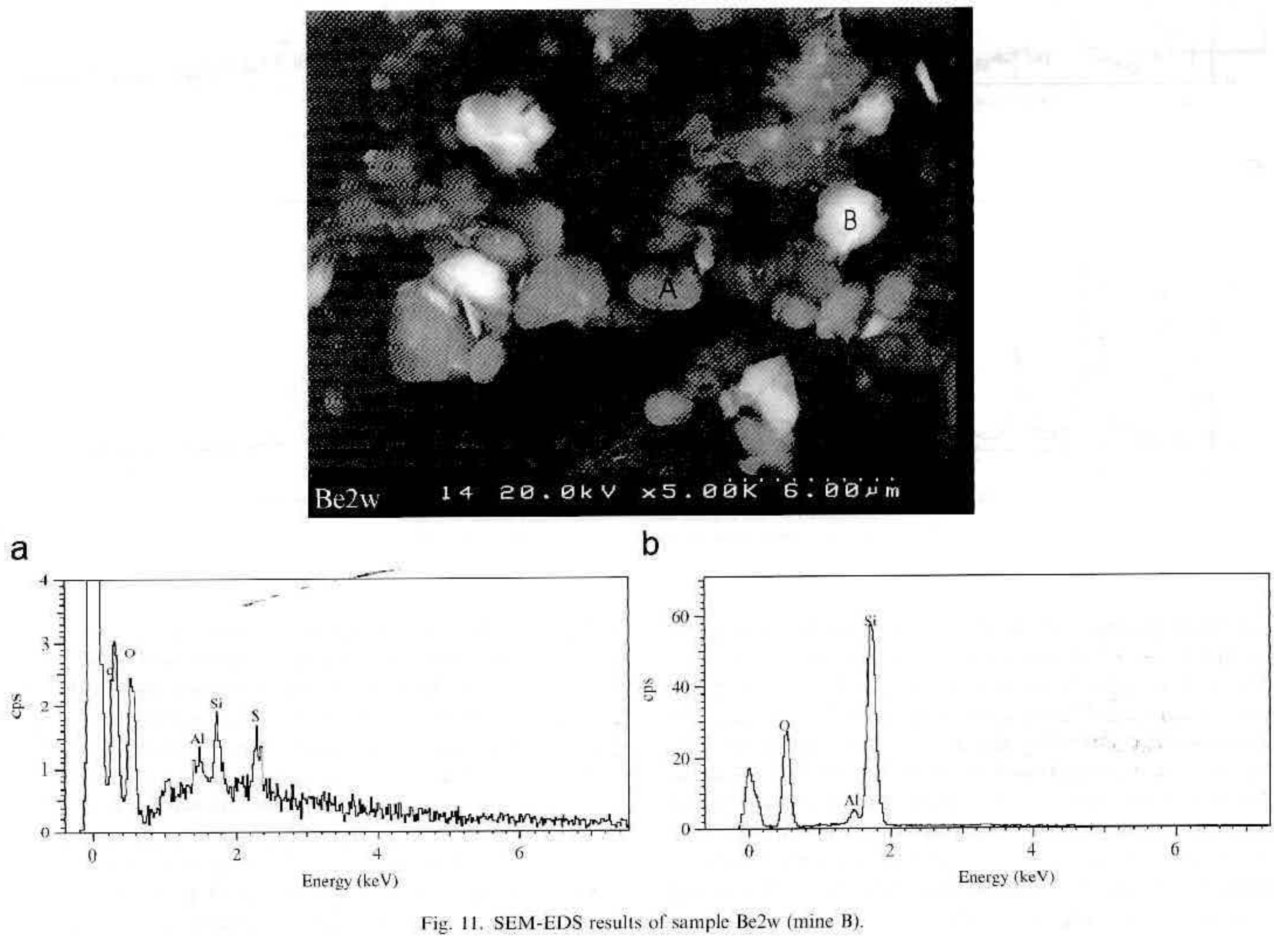


Fig. 11. SEM-EDS results of sample Be2w (mine B).

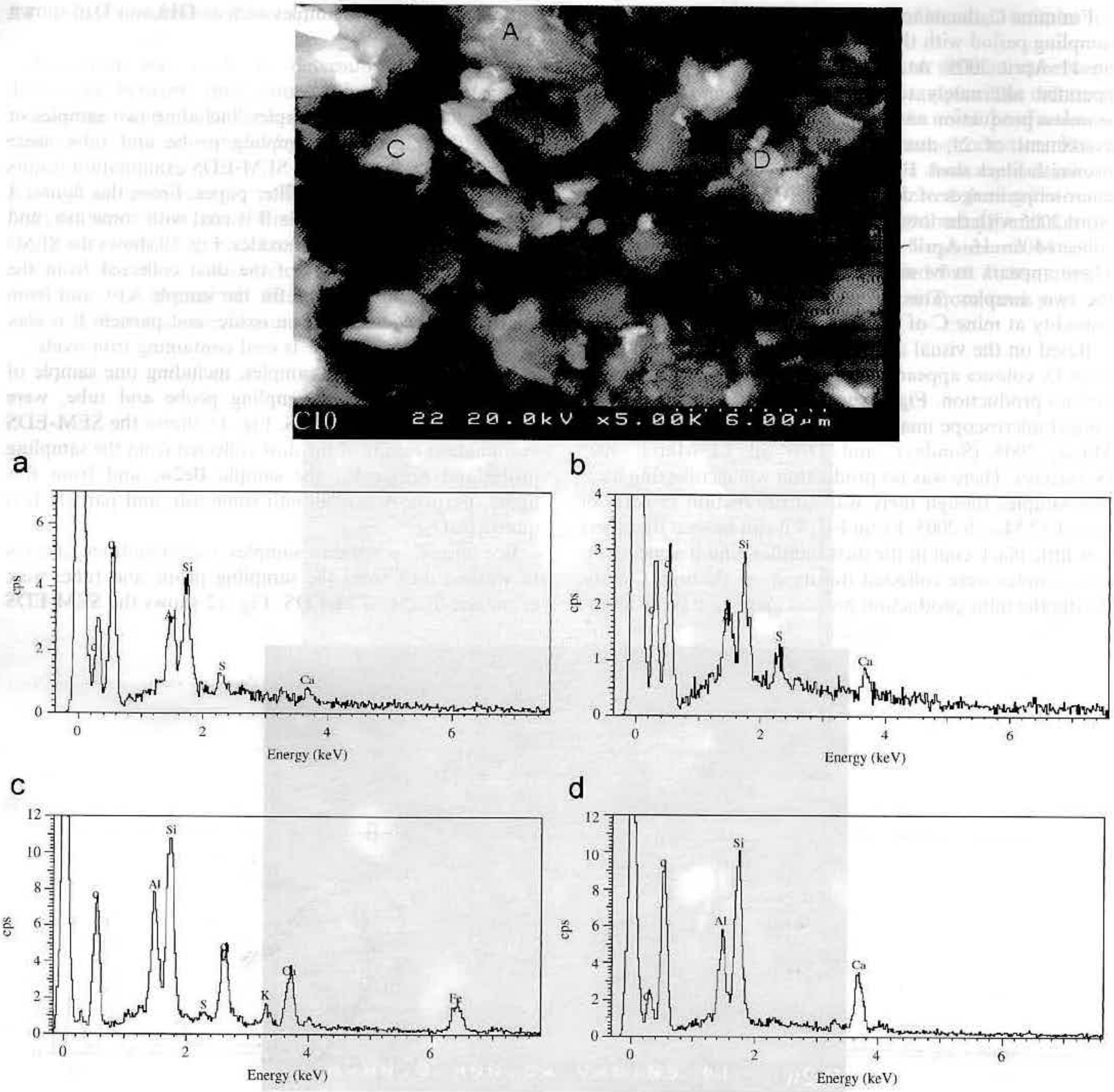


Fig. 12. SEM-EDS results of the sample C10 (mine C).

examination results of the dust deposit on the filter paper (sample C10). From this figure, it can be seen that particles A and B are coal containing some ash. It is noticed that some S is retained in these two coal particles. Particle C is stone (Si–Al–Ca–Fe–O), and also some Cl and K occur in this particle. Particle D is clay (Al–Si–Ca–O). Fig. 13 shows the SEM-EDS examination results of the dust collected from the sampling probe and nozzle for the sample C10w. As shown in this Figure, particle A is coal with a little ash contained inside, and particle B is illite (Al–Si–K–O), and particle C is clay (Al–Si–Na–O).

For mine D, six dust samples, including one sample of washed dust from the sampling probe and tube, were examined by the SEM-EDS. Fig. 14 shows the SEM-EDS examination results of the dust deposit on the filter paper (sample D4), which was collected on 6 March 2005, when there was no mine production. As mentioned earlier, few coal particles are contained in this sample because of no production, and also perhaps due to stone dusting. From Fig. 14, it could be summarised that particle A is calcium oxide, and particle B is Al–Si–Ca–O–K–Na–Cl with a small amount of carbon, which is complicated in terms of

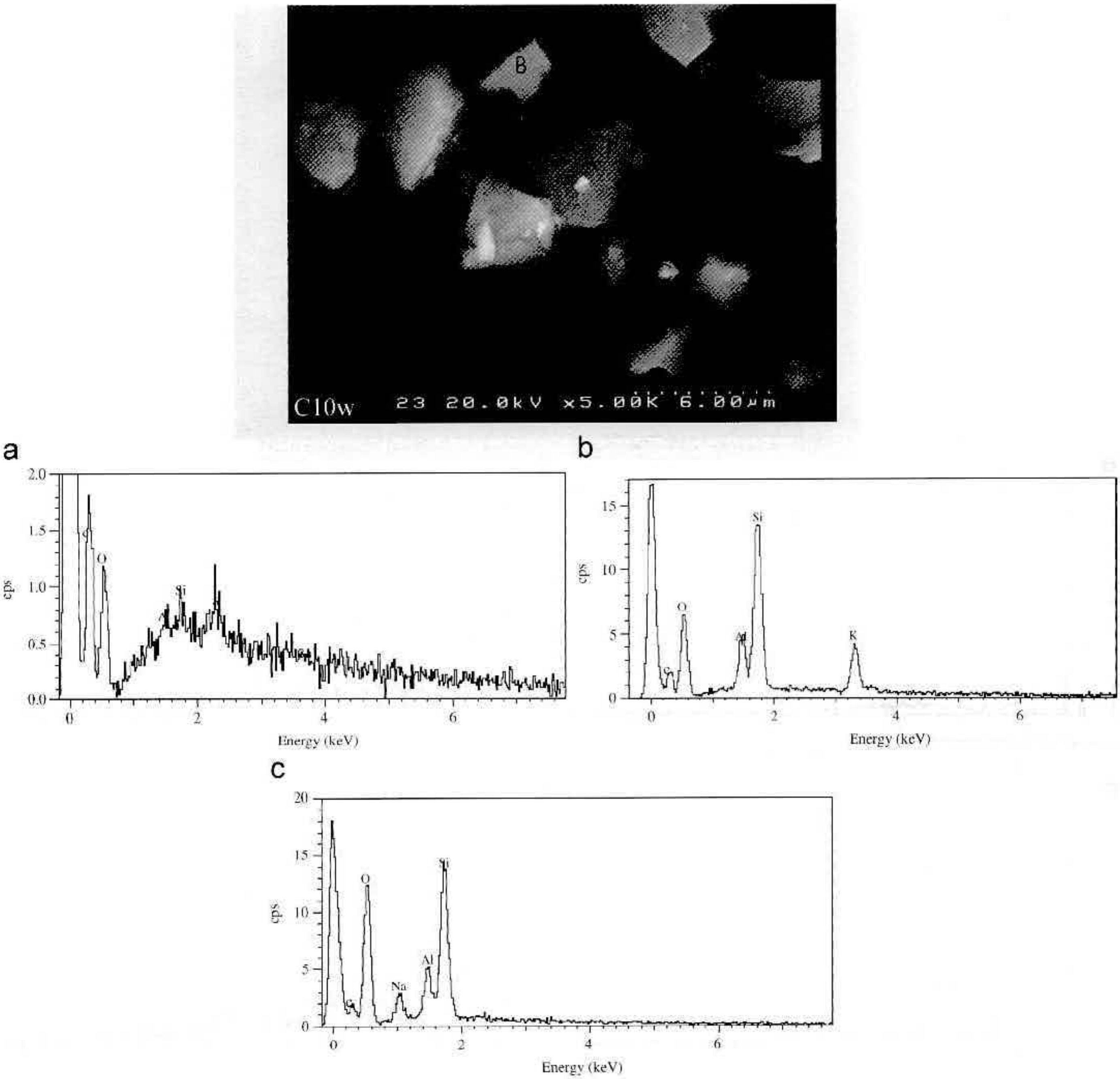


Fig. 13. SEM-EDS results of the sample C10w (mine C).

mineral matter. Particle C is Al–Si–Ca–K–O and particle D is coal with some ash (Si–Al–Ca–O). Fig. 15 shows the SEM-EDS examination results of the dust from sample D13. This sample was collected during mine production, and a lot of coal particles are contained in the sample. As shown in this figure, particle A is quartz (SiO_2). Particle B is coal with high ash (Ca–S–O) content containing some sulphur. Particle C is coal with high ash (Ca–O) content.

In summary, based on the examination of typical dust samples from the four mines, collected dust samples consisted of coal particles and stone particles. When there was no mine production, the dust samples contained fewer coal particles. When the dust sample was collected during mine production, it contained many coal particles.

3.5. Gas compositions

In this study CH_4, CO, H_2S and SO_2 gas concentrations were measured. In addition to measured data on the gas compositions of the ventilation air, mine site data on methane concentration and ventilation air flow rates was retrieved hourly from each mine over a period of ~3 months. In Australia, for mine safety every underground coal mine has an on-line gas monitoring system, mostly

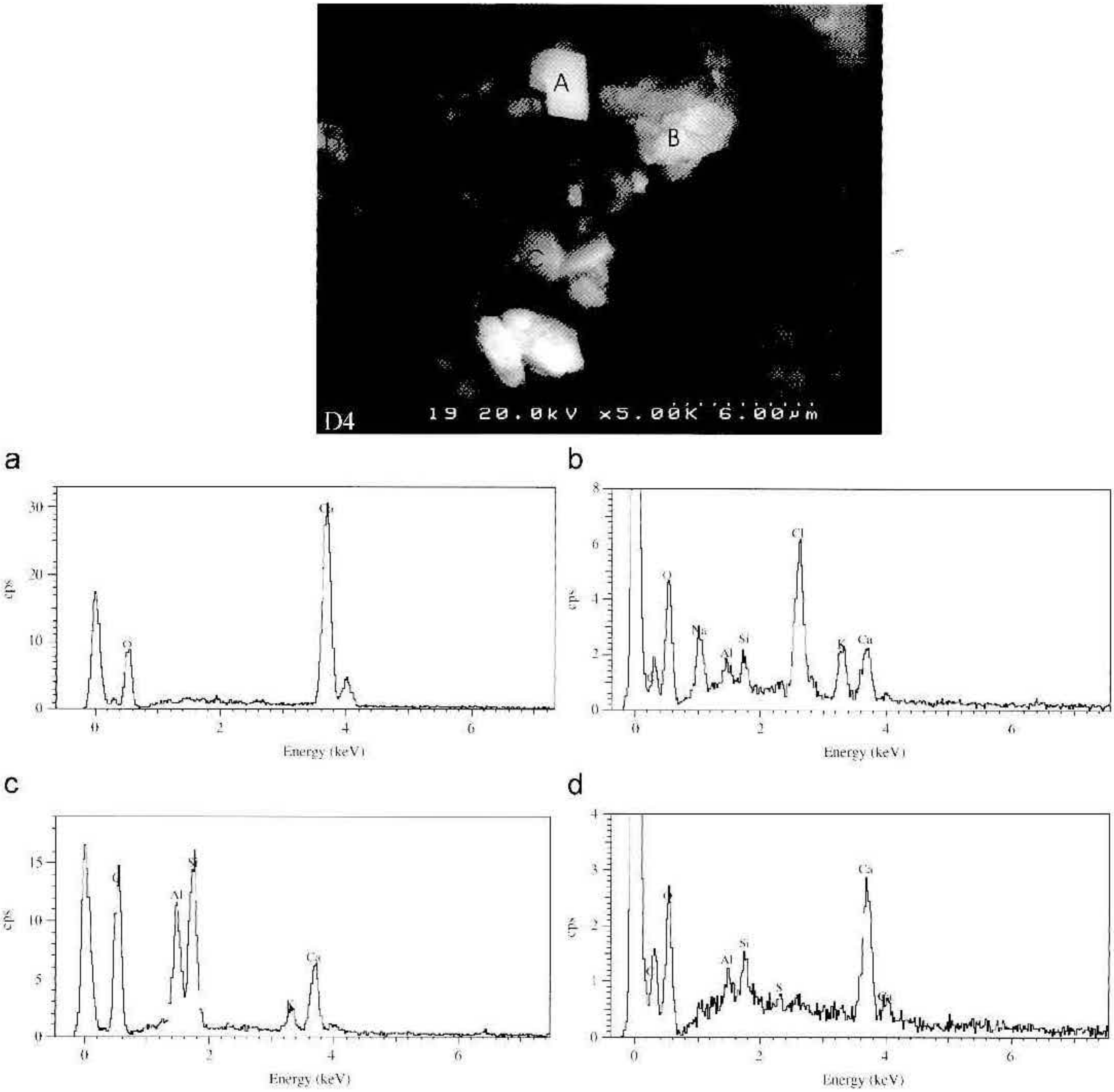


Fig. 14. SEM-EDS results of sample D4.

using Gas Chromatography to analyse sampling gases from tube bundles which take gas at different positions underground including some sampling points near the bottom of mine ventilation air shafts.

Fig. 16 presents and compares methane concentration in the mine ventilation air, retrieved from mine A and measured by the multi-gas monitor. From Fig. 16 the measured methane concentration of the ventilation air is higher than that retrieved from the mine data system, and this is consistent with our investigation carried out in early 2003 as well. In fact, at that time, the mine ventilation engineer reported that the methane concentration of the ventilation air retrieved from the mine data system needs to be increased by 0.1–0.3% on the basis of manual measurements. Hence, in this case, the methane concentration obtained by the gas monitor almost agreed with the data retrieved from the mine once the manually measured data from this mine were corrected.

The concentrations of H_2S and SO_2 were less than the minimum detectable level of the multi-gas monitor (1 ppm). Some CO "spikes" (transient concentration peaks lasting a few seconds) were measured during sampling of A19 and

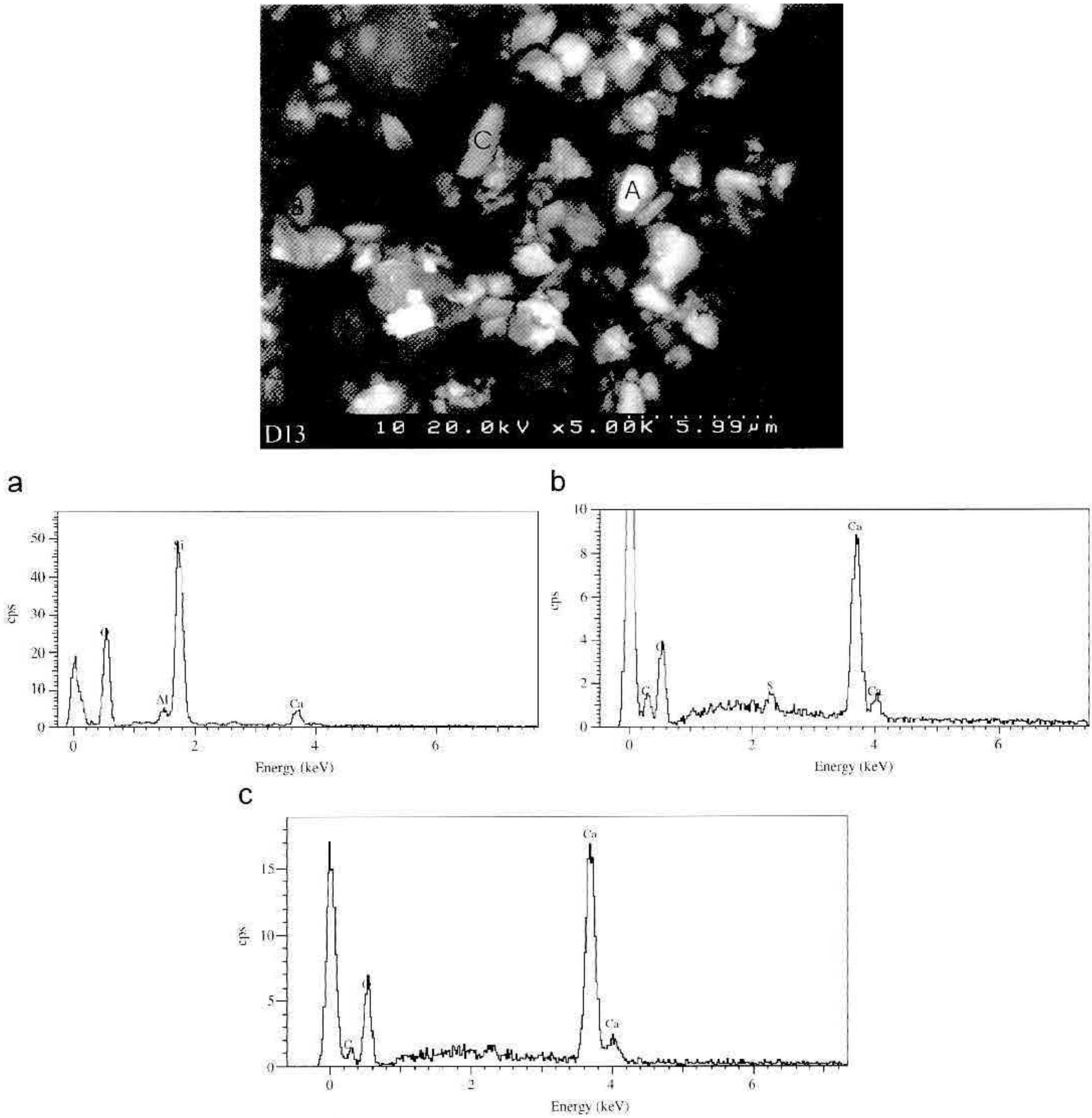


Fig. 15. SEM-EDS results of sample D13.

A25. For sample 19 the reason for this is unknown, and for sample 25 it could be from vehicular activity during the shift change.

For mine B, again, H_2S and SO_2 were not detectable by the multi-gas monitor. No CO spikes were observed during sampling from 5 to 9 April 2005. Fig. 17 presents and compares methane concentration in the mine ventilation air, obtained from the mine and measured by the multi-gas monitor. From Fig. 17 it can be seen that the ventilation air flow was significantly reduced from 22 March 2005. This is because the mine shut down the No. 2 fan due to reduced ventilation requirements. Also, it is clear that our measured methane concentration of the ventilation air correlates closely with the data retrieved from the mine data system.

For mine C, H_2S and SO_2 were <1 ppm. No CO spikes were measured during sampling from 11–16 April 2005. Fig. 18 presents and compares methane concentration in

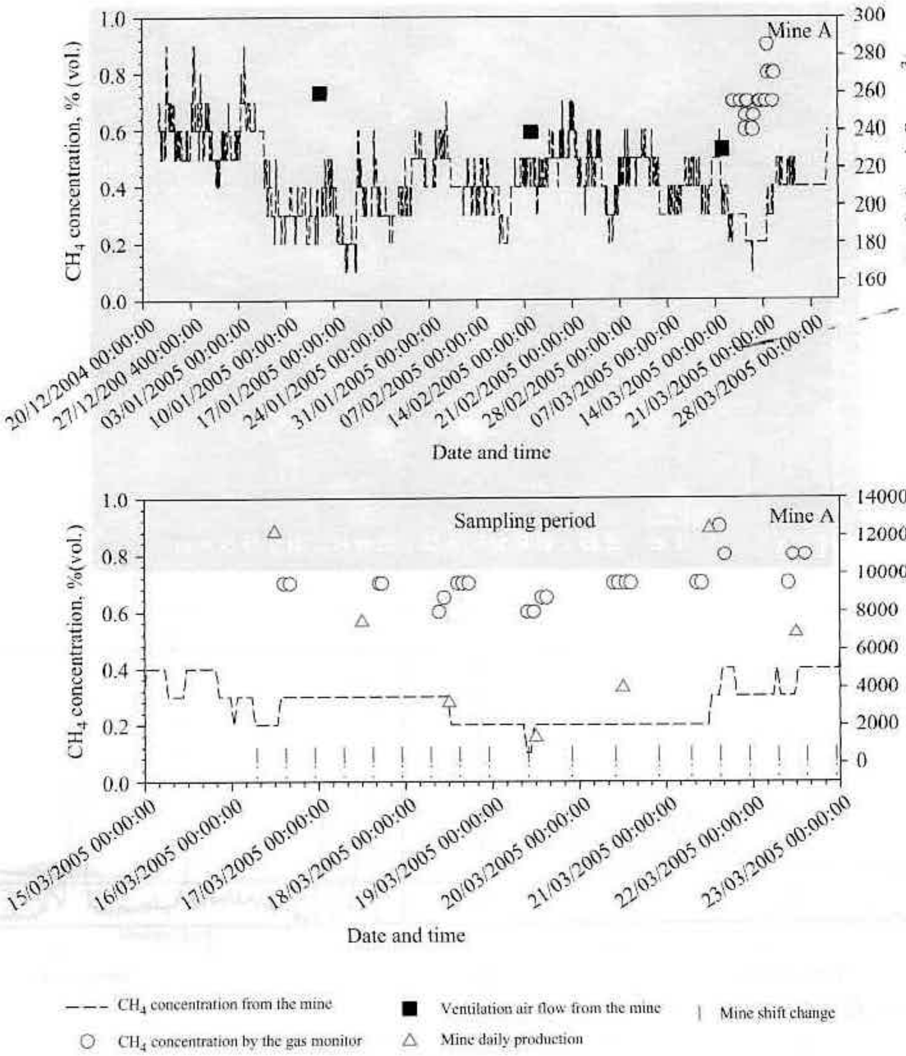


Fig. 16. Methane concentration in mine A ventilation air.

the mine ventilation air, obtained from the mine and measured by the multi-gas monitor. The methane concentration of ventilation air is very low for this mine compared with the other three mines. The average methane concentration in the ventilation air was 0.09% from 17 February to 18 May 2005 based on the data retrieved from mine C, and it was too low for VAM mitigation/utilisation to be presently feasible. In addition, on 17 April 2005 during the collection of sample C21, almost constant CO concentrations of 1–2 ppm were recorded for about 20 min by the gas monitor, but the mine data shows levels near 0 ppm at the same time. This discrepancy has not been explained.

For mine D, H_2S and SO_2 were not detectable by the multi-gas monitor, i.e. <1 ppm. CO spikes as high as 28 ppm were observed for a few seconds during collection of sample D6 on 7 March 2005. This was around the change of shift (14:30–15:00), and may have been caused by diesel equipment being started. Fig. 19 presents and compares methane concentrations in the mine ventilation air, obtained from the mine and measured by the multi-gas monitor. It is clear that the methane concentrations were higher for this mine than the other three mines investigated in this paper. As with mines A and B, the methane concentration by the gas monitor was compared to the data retrieved from the mine data system. On 12 March 2005, the measurements by the multi-gas monitor of methane concentrations in the ventilation air before and after using the ventilation air shaft fans showed consistent readings of 1%. Based on the data retrieved from mine D, the average methane concentration in the ventilation air was 1.01% from 6 February to 30 April 2005. The methane in the ventilation air from this mine should be relatively easy to mitigate/utilise.

It is interesting to note that, compared with the other three mines, the effect of mine production on the methane concentration in the ventilation air was significant. For example, when there was no production on 5 and 6 March 2005, the average methane concentration was 0.5%, which is much lower than the ~1% during the mine production.

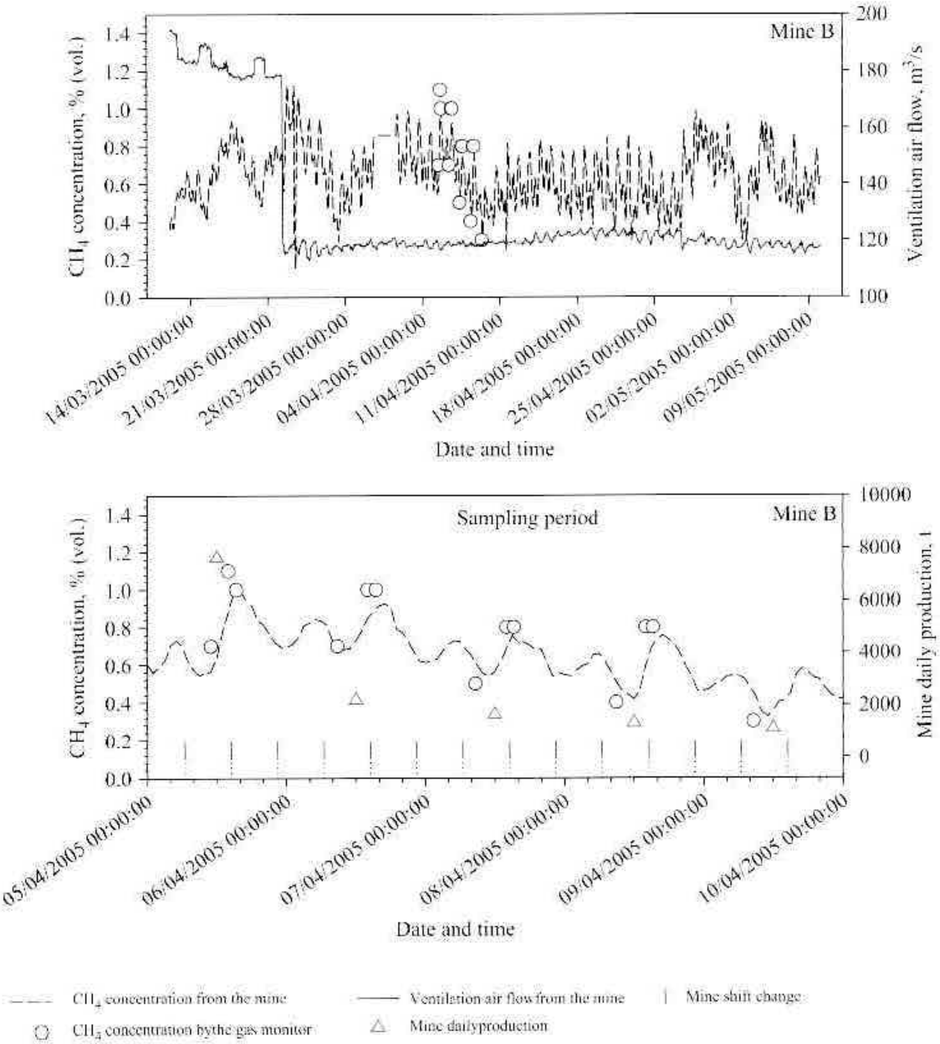


Fig. 17. Methane concentration in mine B ventilation air.

. Conclusions

This paper presented mine-site experimental results on haracteristics of mine ventilation air flows, including ethane concentration and its variations, dust loadings, article size, mineral matter of the dust, and other ompounds in the ventilation air flows. The major haracteristic parameters are summarised in Table 1, and teresting findings are given below.

Mine A: This mine was the most humid of the four iines. The ventilation air was saturated (100% relative umidity). Results showed little correlation between mine roduction and the dust loading. This could be due to the iany water droplets observed circulating in the upcast haft acting as a spray scrubbing system for the ventilation ir. The maximum measured dust particle size in the entilation air outlet was 5 µm. SEM-EDS results showed hat the dust samples consisted mostly of coal particles and one particles.

H_2S and SO_2 concentrations were less than 1 ppm. Some O spikes were captured during sampling which could have been from vehicular activity during the shift change. It is interesting to note that this mine was the first in the world to commercially use VAM as an ancillary fuel source. However, it is not used anymore because the amount of particulate matter in the air required constant replacement of the air filters, resulting in high costs.

Mine B: This mine had measured relative humidities of 85–100%. Dust samples collected during mine production were visibly darker than those collected during non-production. Analysis showed that the dark dust samples contained more coal particles than the grey/white dust samples collected during non-production. Therefore, the mine production rate affected the dust composition in this mine. Also the data indicated that the higher the mine production rate, the higher the dust loading in this mine.

H_2S and SO_2 concentrations were less than 1 ppm. No CO spikes were observed, in contrast with mine A. The data collected by the mine methane monitoring system were consistent with our measurements.

Mine C: This mine is not a gassy mine, and the relative humidity was low at 74.5–83.5%. The mine produced coal

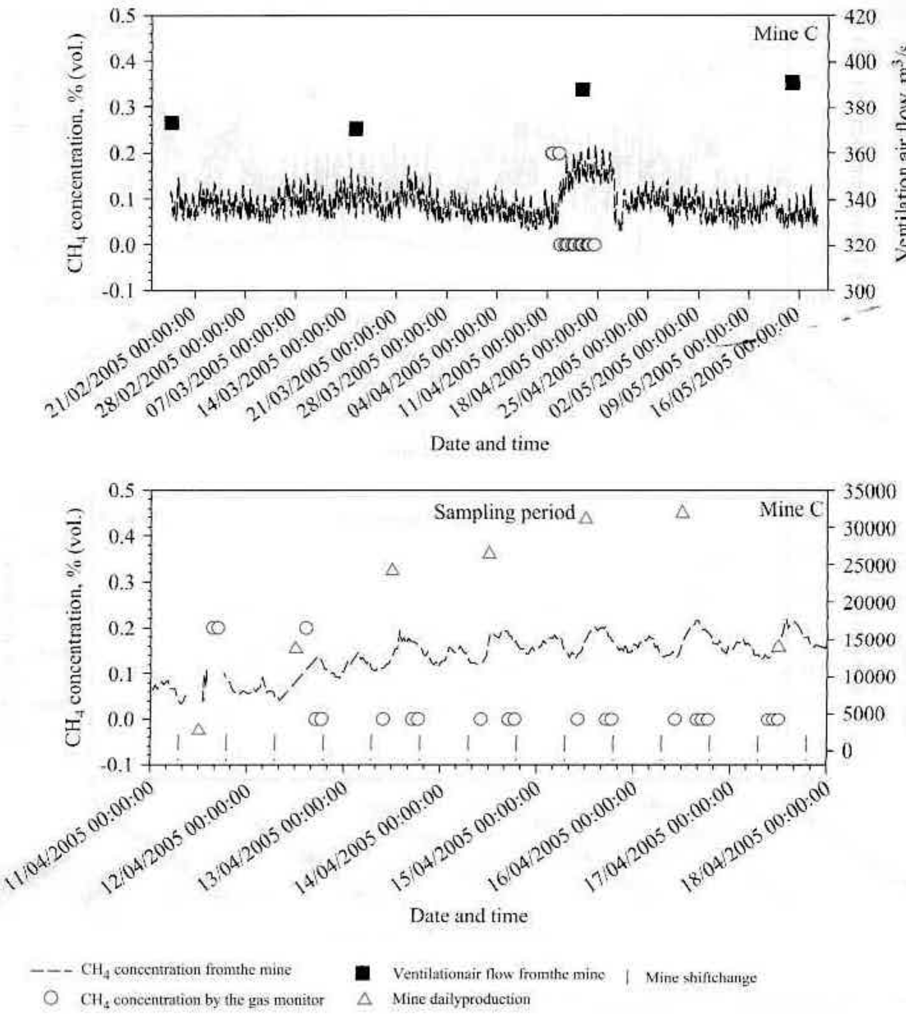


Fig. 18. Methane concentration in mine C ventilation air.

every day during the sampling period. There was a small correlation between coal production and dust loading. It is interesting to note that, based on the SEM-EDS results, some sulphur was retained in the coal particles. This may imply a technical issue that when such mine ventilation air is directly fed into a catalytic oxidation system without filtration, the sulphur could poison the catalysts.

H_2S and SO_2 concentrations were less than 1 ppm. No CO spikes were captured during sampling. The methane concentration of ventilation air was very low for this mine compared with the other three mines. The average methane concentration in the ventilation air was only 0.09%, hence any existing and developing VAM mitigation/utilisation technologies would not be suitable for this mine.

Mine D: The relative humidity of the ventilation air at this mine was 73–99.8%. There was little correlation between the dust emissions and mine production rate, potentially because of stone dusting during non-production periods. This is supported by SEM-EDS examination results which showed that the dust collected when there was no mine production contained few coal particles.

H_2S and SO_2 concentrations were less than 1 ppm. CO spikes as high as 28 ppm were captured during shift changes, and may have been caused by diesel equipment being started. The methane concentration of the ventilation air was relatively high, 1.01% on average. This means that the methane in ventilation air from this mine is relatively easy to mitigate and utilise. It is very interesting to note that, compared with the other three mines, the effect of mine production on the methane concentration of the ventilation air was significant. For example, when there was no production on 5 and 6 March 2005, the average methane concentration halved to 0.5%.

General remarks: In general, mines A, B and D are gassy mines. All have drainage gas systems (at least post-drainage gas practise, i.e. degasification after coal mining, at the mine sites). Mine C is not a gassy mine with an average methane concentration of 0.09% in the ventilation air. In conclusion, for all the mines investigated in this paper the dust samples consisted of coal particles and stone particles. When there was no mine production, the dust samples contained few coal particles. When the dust

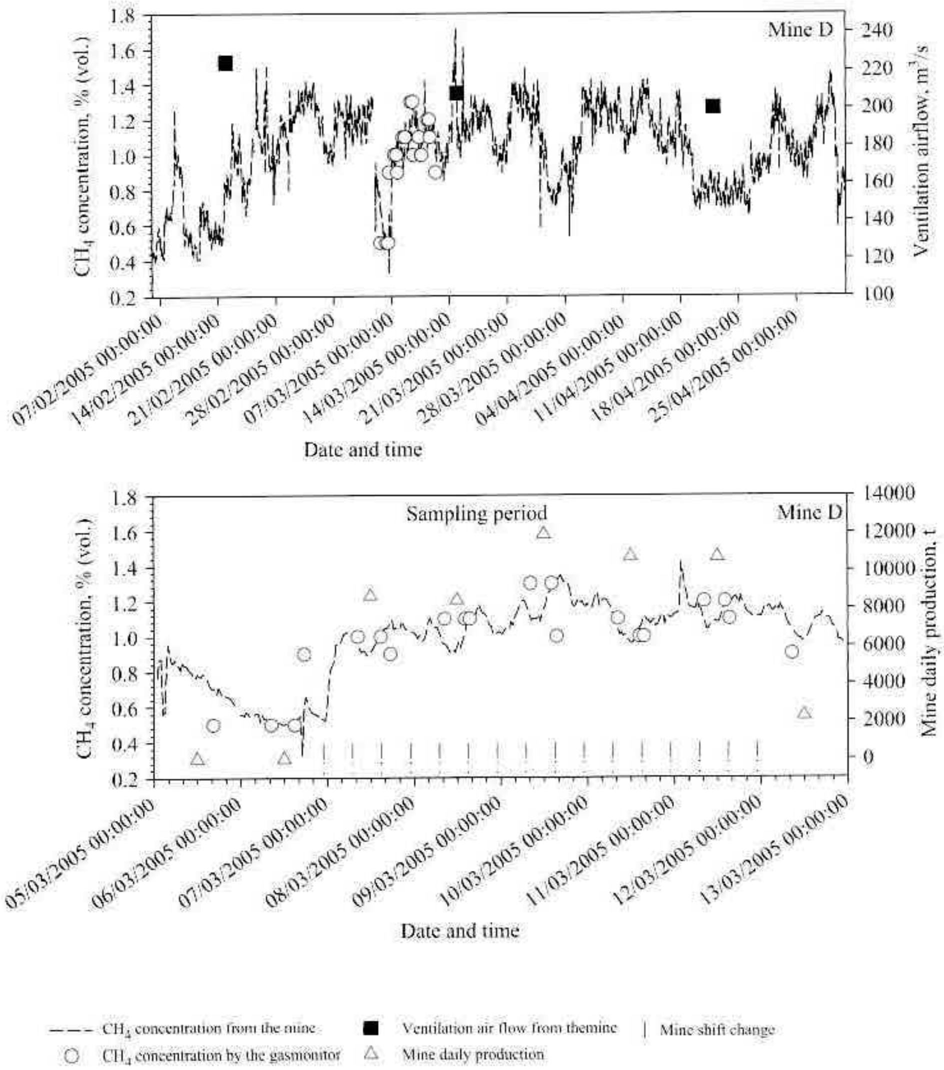


Fig. 19. Methane concentration in mine D ventilation air.

Table 1
Major characteristic parameters of ventilation air flows at four Australian coal mines

Mine	Sampling time	Dust loading (mg/m^3)	Maximum particle size (μm)	Gas compositions (average)				Average VA flow rate (m^3/s)	Relative humidity (%)
				CH_4 (%)	CO (ppm)	H_2S (ppm)	SO_2 (ppm)		
A	March 2005	0.13–4.47	5.0	0.75	<1	<1	<1	250	100
B	November 2004 and April 2005	0.25–3.87	5.0	0.65	<1	<1	<1	200[a]	85–100
C	April 2005	0.67–3.82	5.0	0.09	<1	<1	<1	385	74–84
D	March 2005	0.21–3.19	5.0	1.01	<1	<1	<1	210	73–100

[a]Average VA flow rate is about 120 m^3/s when the No.2 fan was off during the mine site sampling in April 2005.

samples were collected during mine production, they contained many coal particles.

Moreover, the data presented in this paper provide basic information for selection, assessment, and development of ventilation air cleaning technology and VAM mitigation and utilisation technology. Variations in methane concentration and ventilation air flow rate need to be considered carefully when choosing appropriate technology.

Acknowledgements

We wish to acknowledge ACARP for funding the project of characterisation and cleaning of mine ventilation air flows. We would also like to thank Mr. Dwight van Diem for his mine site sampling work, and mine site personnel for their supervision and assistance at the mine sites. Special thanks are due to project monitor, Dr. John Rich, for his suggestions and support of the project, and Mr. Jim Craigen for his coordination of the project, and Dr. John Carras for his valuable comments.

References

Moore, S., Freund, P., Riemer, P., Smith, A., 1998. Abatement of methane emissions. IEA Greenhouse Gas R&D Programme, June 1998.

Sloss, L., 2005. Coalbed methane emissions—capture and utilisation. IEA Clean Coal Centre, CCC/104, November 2005.

Standards Australia, 1995a. AS 4323.2-1995 Stationary source emissions—determination of total particulate matter—isokinetic manual sampling – gravimetric method.

Standards Australia, 1995b. AS 4323.1-1995 Stationary source emissions—selection of sampling positions, 1995.

Su, S., Beath, A., Guo, H., Mallett, C., 2005. An assessment of mine methane mitigation and utilisation technologies. Progress in Energy and Combustion Science 31 (2), 123–170.

US EPA, 2003. Assessment of the worldwide market potential for oxidising coal mine ventilation air methane. United States Environmental Protection Agency, EPA 430-R-03-002, July 2003

US EPA Method 5. Determination of particulate matter emission from stationary sources.

ELSEVIER

Journal of Environmental Management 86 (2008) 63–79

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Environmental protection, the economy, and jobs: National and regional analyses☆

Roger H. Bezdek[a,*], Robert M. Wendling[a], Paula DiPerna[b]

[a] *Management Information Services, Inc., 2716 Colt Run Road, Oakton, VA 22124, Washington, DC, USA*
[b] *Jobs and the Environment Initiative, New York, USA*

Received 1 March 2005; received in revised form 19 November 2006; accepted 23 November 2006
Available online 17 January 2007

Abstract

The relationship between environmental protection (EP), the economy, and jobs has been an issue of harsh contention for decades. Does EP harm the economy and destroy jobs or facilitate economic growth and create jobs? We address this issue by summarizing the results of the Jobs and the Environment Initiative, research funded by nonprofit foundations to quantify the relationship between EP, the economy, and jobs. We estimate the size of the US environmental industry and the numbers of environment-related jobs at the national level and in the states of Florida, Michigan, Minnesota, North Carolina, Ohio, and Wisconsin. This is the first time that such comprehensive, detailed estimates have been developed.

Our major finding is that, contrary to conventional wisdom, EP, economic growth, and jobs creation are complementary and compatible: Investments in EP create jobs and displace jobs, but the net effect on employment is positive.

Second, environment protection has grown rapidly to become a major sales-generating, job-creating industry—$300 billion/year and 5 million jobs in 2003.

Third, most of the 5 million jobs created are standard jobs for accountants, engineers, computer analysts, clerks, factory workers, etc., and the classic environmental job (environmental engineer, ecologist, etc.) constitutes only a small portion of the jobs created. Most of the persons employed in the jobs created may not even realize that they owe their livelihood to protecting the environment.

Fourth, at the state level, the relationship between environmental policies and economic/job growth is positive, not negative. States can have strong economies and simultaneously protect the environment.

Finally, environmental jobs are concentrated in manufacturing and professional, information, scientific, and technical services, and are thus disproportionately the types of jobs all states seek to attract.

Keywords: Jobs; Environmental protection; Economy; Industry; Manufacturing; Skills; State; Occupations

1. Introduction: the issue

The relationship between environmental protection (EP), the economy, and jobs has been an issue of harsh contention for decades. Analysts and policymakers of all points of view seem to agree that a strong relationship exists between EP and jobs; the debate is over the sign of the correlation coefficient. Does EP tend to harm the economy and destroy jobs or to facilitate economic growth and create jobs? If the latter is the case, can the positive affects be quantified and estimated at a meaningful level of detail?

Here, we address this issue by summarizing the initial results of the Jobs and the Environment Initiative, a research effort funded by nonprofit foundations designed to quantify the relationship between EP, the economy, and jobs.[1] We estimate the size of the US environmental

☆This paper finds that, contrary to the conventional wisdom, environmental protection has evolved into a major US industry, that most of the 5 million jobs created are for occupations not related to the environment, and that detailed economic and employment impacts can be estimated for individual states.

*Corresponding author. Tel.: +1 7036 204120; fax: +1 7035 699733.
E-mail address: rbezdek@misi-net.com (R.H. Bezdek).

[1]The research summarized here was supported by the Rockefeller Foundation, the Merck Fund, and the Beldon Fund.

0301-4797/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.028

industry in 2003 and the numbers of environment-related jobs created at the national level and in the states of Florida, Michigan, Minnesota, North Carolina, Ohio, and Wisconsin.

2. Previous studies

Numerous studies have been undertaken over the past two decades to estimate the economic and employment effects of EP. These can be grouped into three types: (i) theoretical analyses and cases studies, (ii) econometric simulations of policy alternatives, and (iii) empirical estimates derived using historical data. Below we review some of the major studies in each category.

2.1. Theoretical analyses and cases studies

In 1992, Meyer analyzed the impact of environmental legislation on differential interstate rates of economic performance and tested the hypothesis that pursuit of environmental quality hinders economic growth and job creation (Meyer, 1992). He ranked the 50 states on the basis of the stringency of their environmental laws and then compared the environmental rankings with measures of economic growth and job creation between 1973 and 1989. He found no evidence to support a negative relationship between environmental regulation and economic performance, and his results showed the opposite. Meyer found that the states with the most ambitious environmental programs had the highest levels of economic growth and job creation over the period.

In 1993 and 1995, Bezdek examined the available empirical evidence and found that, contrary to conventional wisdom, strict environmental regulations do not damage US industry, reduce international competitiveness, or cost thousands of jobs, and he found that strict environmental standards may even foster economic development (Bezdek, 1993). He concluded that recent major empirical studies reject the hypothesis that there is a negative relationship between EP and economic and job growth.

During the 1990s, Goodstein conducted several studies examining the relationship between EP and employment (Goodstein, 1994). He examined the impact of existing regulations on overall employment rates, shutdowns and layoffs, regulation-induced capital flight, estimates of the costs of environmental regulation, and specific industries and case studies. He found that little empirical evidence exists that environmental regulation destroys jobs and that ex ante estimates of the costs of compliance have been much higher than actual costs. He also showed the conditions under which EP might lead to increased employment.

In 1995, Templet hypothesized that the economy is dependent on the environment to provide resources and accept wastes, and that a healthy environment should make for a better economy (Templet, 1995). He cited empirical evidence from a number of studies substantiating that finding and showing that states with lower pollution levels and better environmental policies generally have more jobs, better socioeconomic conditions, and are more attractive to new business. He conducted a case study of Louisiana case that found that jobs increased while pollution declines. He concluded that there is little evidence that progressive environmental policies are detrimental to a state's economy and that there is substantial evidence that the converse is true.

In 1995, Repetto reported that EP requirements have not contributed to job loss or reduced international competitiveness for US companies. He also found that firms with superior environmental performance are no less profitable than others in the same industry (Repetto, 1995). He concluded that appropriate US environmental policies could protect the environment with far greater economic efficiency.

In 1997, Berman analyzed the regulation of air pollution in manufacturing plants in Los Angeles. He examined employment growth in the Los Angeles region in plants subject to these regulations, and compared growth at these plants to employment growth at similar plants in Texas and Louisiana. He found that, while the Los Angeles regulations imposed costs on regulated plants, they had little effect on employment and that there were no large job losses due to these regulations. His major finding was that the most severe episode of increased air quality regulation of manufacturing industries did not have a large effect on manufacturing employment.

In 1998, Morgenstern, Pizer, and Shih examined the possibility that workers could be adversely affected in heavily regulated industries, which has led to claims of a "jobs vs. environment" tradeoff (Morgenstern et al., 1998). They explored how increased environmental stringency can influence the industry-level demand for labor and developed estimates for four heavily polluting industries (pulp and paper, plastics, petroleum refining, and iron and steel). Their results indicated that increases in environmental spending do not cause job loss. Their model showed that the overall demand effect is mitigated by employment increases associated with new environmental spending.

In 1999, Arnold, Forrest, and Dujack examined claims about the costs of environment regulations by reviewing the available research (Arnold et al., 1999). They found that, while the claims about damage to the economy can mostly be attributed to misinformed advocates or exaggeration, the majority of the fault lies in a lack of accurate communication of economists' findings about the effect of environmental regulation to the general public. Worst-case economic impact scenarios for a regulation—such as potential increases in unemployment and plant closures—are reported not as low probabilities, but as serious threats. They concluded that the view that environmental regulation seriously harms the US economy is not supported by the data.

In 1999, Bliese reviewed dozens of well-designed studies that tested the assertion that EP harms the economy (Bliese, 1999). The results of these studies indicate that EP normally has no negative impact on the economy overall, and often has a positive effect. He noted that the studies only searched for economic impacts of environmental policies—and found none; they did not estimate environmental or public health benefits. He concluded that the "environment vs. the economy trade-off" is a myth, even in narrowly economic terms.

In 1999, Yapijakis found that widespread fears of job losses from EP are unfounded and that, when job creation aspects of pollution control policies are factored in, EP has increased net employment in the US (Yapijakis, 1999) Further, actual layoffs due to regulation have been extremely small. EP raises employment levels and provides some recession-proof stimulus to aggregate demand. Government data reveal that few manufacturing plants are shut down as a result of environmental or safety regulations.

In 2000, Renner found that creating an environmentally sustainable economy has already generated an estimated 14 million jobs worldwide (Renner, 2000). He reported that many new opportunities for job creation are emerging, ranging from recycling and remanufacturing of goods, to greater energy and materials efficiency and the development of renewable energy. Jobs are more likely to be at risk where environmental standards are low. He concluded that investing in the environment, in renewable energy, and energy efficiency will generate more jobs than investing in extractive industries and fossil fuels.

2.2. Econometric simulations of policy alternatives

In 1989, Arvind Teotia and his associates estimated the macroeconomic impacts of the use of clean diesel engine technology in light trucks to comply with corporate average fuel economy (CAFÉ) standards. They assumed that the new engines would capture 15 percent of the light truck market and estimated that by 2022, between 70,000 and 110,000 jobs would be created (Teotia et al., 1999).

In 1989, Bezdek and Wendling simulated the impact of the two major acid rain control bills that were then being considered in the US Congress (Bezdek and Wendling, 1989). They found that between 100,000 and 195,000 net jobs would be created, depending on which bill was enacted. Economic and job impacts were estimated for each state and employment requirements by occupation at the national level were also estimated.

In 1990, Jorgensen and Wilcoxen estimated the impact of environmental regulations on the US economy by simulating the growth of the economy between 1974 and 1985 with and without these regulations (Jorgenson and Wilcoxen, 1990). They concluded that the effect of these regulations was that the economy grew 0.2%/year more slowly than it would have otherwise, and that by the early 1990s GNP was about 2.5% less. In 1993, they extended the analysis to assess the impacts of the Clean Air Act Amendments of 1990 and concluded that the net impact of the Amendments would further reduce the rate of growth of GNP (Jorgenson et al., 1993). They did assess the benefits of EP or estimate employment impacts.

In 1990, a Motor Vehicles Manufacturers Association study of the potential impact of increased CAFE standards predicted that tighter CAFE standards would result in the loss of between 159,000 and 315,000 jobs in the motor vehicle industry (Motor Vehicle Manufactures Association, 1990). Secondary effects from consumer fuel savings were not estimated and the report did not consider that fuel savings by consumers would result in additional spending on other products and higher employment in the affected industries.

In 1992, Geller, DeCicco, and Laitner estimated the impact of a "high efficiency" scenario for the energy-using sectors of the economy and found that a national strategy of investment in environmentally benign energy sources and energy efficiency would create one million net new jobs in the United States within 10 years (Geller et al., 1992). They also found that by increasing the fuel efficiency of passenger cars from 28 mpg in 1990 to 40 mpg in 2000 and 50 mpg in 2010, 244,000 additional jobs would be created by 2010.

1993, MISI simulated the impact that the environmental provisions of the North American Free Trade Agreement would have on US exports and jobs (Management Information Services, Inc., 1993). The study estimated that by 2000, NAFTA would generate \$3.8 billion in US environmental export sales to Mexico and create nearly 70,000 jobs in the US Jobs estimates were disaggregated by state.

During the 1990s. Laitner, DeCicco, Elliott, Geller, Goldberg, Morris, and Nadel examined energy consumption patterns in the four-state Midwest region of Illinois, Indiana, Michigan, and Ohio and the states of New York, New Jersey, and Pennsylvania, and projected energy consumption through 2010 assuming business-as-usual policies and trends (Laitner et al., 1994). They then developed an energy efficiency scenario assuming more aggressive implementation of energy efficiency measures and analyzed the potential economic benefits of the scenario. They found that by 2010, the energy efficiency scenario would create 132,000 net new jobs in Midwest region and 164,000 net new jobs in New York, New Jersey, and Pennsylvania.

In 1999, Bernow, Cory, Dougherty, Duckworth, Kartha, and Ruth examined the impact of implementing a set of integrated policies designed to bring the US in compliance with the Kyoto Protocol (Bernow et al., 1999). They found that the US could reduce its carbon emissions to its Kyoto target and that the prescribed policies would produce net economic savings. Specifically, they estimated that by 2010 almost 900,000 net new jobs would be created, relative to the baseline.

A 2001, a Friedman study for the Union of Concerned Scientists analyzed the effects of increasing CAFE standards to 40 mpg by 2012 and to 55 mpg by 2020 (Friedman et al., 2001). UCS estimated that employment, wages, and income would increase over the 10–20-year horizon of the study. By 2010, the analysis projected a net increase of over 40,000 jobs; by 2020, the study projected an increase of 104,000 jobs. In a 2002 update of this study, UCS highlighted the potential job gains by industry and state resulting from increased CAFE standards, and estimated that 183,000 new jobs would be generated (Union of Concerned Scientists, 2002).

In 2002, research staff at the University of Illinois analyzed the Midwest's Clean Energy Development Plan, which advocated energy efficient technologies and development of renewable energy resources, especially wind power and biomass energy (Regional Economics Applications Laboratory, 2002). They estimated that implementing the plan would create more than 200,000 new jobs across the 10-state Midwest region by 2020.

In 2002 and 2004, Barret and Heorner assessed the impact of a set of policies designed to provide steady increases in energy efficiency and reductions in carbon emissions, while improving overall economic efficiency (Barrett and Hoerner, 2002). They analyzed the macroeconomic impact of these policies and estimated that an additional 660,000 net jobs would be created in 2010 and 1.4 million in 2020. This would increase employment in the service sector and reduce the rate of decline in employment in manufacturing.

In 2004, the New Apollo Initiative proposed an economic development plan for the US based on diversifying energy sources, making the US less dependent on foreign oil, investing in industries of the future, promoting construction of energy efficient buildings, and investing in cities and communities (New Energy for America, 2004). It estimated that a $30 billion investment per year for 10 years would add more than 3.3 million jobs to the economy and stimulate $1.4 trillion in new GDP.

In 2004, UCS analyzed the effects of implementing a national renewable electricity standard (RES) that would require electric utilities to supply a set percentage of their electricity from renewable sources (Union of Concerned Scientists, 2004). It found that a under a national RES of 20% by 2020, the US would increase its total renewable power capacity by nearly 11 times over present levels and would create more than 355,000 new jobs.

In 2004, Levinson and Taylor examined the effect of environmental regulations on trade flows by developing an economic model to demonstrate how unobserved heterogeneity, endogeneity, and aggregation issues bias measurements of the relationship between regulatory costs and trade (Leninson and Taylor, 2004). They applied an estimating equation derived from the model to data on US regulations and net trade flows among the US, Canada, and Mexico for 130 manufacturing industries from 1977 to 1986. Their results indicated that industries whose abatement costs increased most experienced the largest increases in net imports. For the 20 industries hardest hit by regulation, the change in net imports they ascribed to the increase in regulatory costs amounted to more than half of the total increase in trade volume over the period.

In 2005, Bezdek and Wendling estimated the economic impacts on the US of enhanced CAFE standards and found that such changes would have positive economic effects and create 300,000 jobs, although the costs in terms of vehicle characteristics and prices and limited consumer choice could be significant (Bezdek and Wendling, 2005). There would be widespread job displacement and job impacts were disaggregated by industry, state, and occupation.

2.3. *Empirical estimates of actual environmental employment*

For two decades, Environmental Business International has been publishing estimates of the size of the US environmental industry with times series data beginning in 1970. The data are disaggregated by Services (analytical services, wastewater treatment, solid waste, hazardous waste, remediation, and consulting and engineering), Equipment (water and chemicals, instruments and information, air pollution control, waste management, and process and prevention), and Resources (water utilities, resource recovery, and clean energy systems and power). EBI estimates that the size of the US environmental industry has increased from $18 billion in 1970 to $227 billion in 2003. Corresponding employment or jobs numbers are not published.[2]

MISI has been estimating the economic and jobs impact of the environmental industry for two decades.[3] Using an econometric input–output (I–O) model to estimate the direct and indirect impact of the industry, MISI estimates that EP has increased from $39 billion in sales (2003 dollars) and 700,000 jobs in 1970 to $300 billion (2003

[2] Environmental Business International, Inc., San Diego, California, www.ebiusa.com

[3] Roger H. Bezdek. "The Environmental Protection Industry and Environmental Jobs in the U.S.A." In Leal Filho and Kate Crowley, eds., *Environmental Careers, Environmental Employment, and Environmental Training: International Approaches and Contexts.* Frankfurt am Main: Peter Lang Publishers, 2001, pp. 161–179; "State of the Industry: Jobs and Sales Created by Environmental Protection." *New England's Environment*. Vol. 1, No. 8 (August 1999), pp. 12–16; "The Net Impact of Environmental Protection on Jobs and the Economy." Chapter 7 in Bunyan Bryant, editor., *Environmental Justice: Issues, Polices, and Solutions*, Washington, DC: Island Press, 1995, pp. 86–105; "The Economy, Jobs, and the Environment." *Proceedings of GEMI '95: Environment and Sustainable Development*. Arlington, Virginia, March 1995, pp. 65–79; "Environment and Economy: What's the Bottom Line?" *Environment*, vol. 35(7) (September 1993), pp. 7–32. Roger H. Bezdek and Robert M. Wendling. "Environmental Market Opportunities." Chapter 9 in T.F.P. Sullivan, editor, *The Greening of American Business*. Rockville, Maryland: GII Press, 1992, pp. 196–224; Management Information Services, Inc., *Jobs and Economic Opportunities in the U.S. Created by Environmental Protection*. Periodic reports, 1986–2004.

dollars) and 5 million jobs in 2003—see the discussion in Section 5.

In 2001, MISI analyzed the environmental industry and jobs in six Midwestern states: Illinois, Iowa, Michigan, Minnesota, Ohio, and Wisconsin.[4] It found that in 1998, environment-related employment in these states totaled 893,000 widely distributed among sectors, industries, jobs, and skills. Jobs estimates were disaggregated among each of the six states.

In 1999, the US International Trade Administration (ITA) estimated the world market for environmental products and services and the size of the US market, including estimates at the state and metropolitan statistical area (MSA) levels.[5] ITA estimated that the 1999 US environmental market totaled $189 billion, almost 38% of the global $499 billion market. In meeting the demands of those markets, the US environmental industry was estimated to have generated $196 billion of revenues and over 1.4 million jobs. The ITA US employment estimates were disaggregated by state and by selected MSAs.

The Census MA200 survey has been one of the more respected sources for information on the US environmental industry.[6] This report was not available for a number of years after 1994, but was revived for 1999. The MA200 results are not consistent with previous reports, but they presented a snapshot of major portions of the environmental industry by detailed North American Industry Classification System (NAICS) industry and by state. However, the survey's biggest weakness is that it only covers the mining (NAICS 21), manufacturing (NAICS 31-33), and electric power generation industries (NAICS 22111). Thus, while the survey estimates are of sufficient quality, they lack comprehensiveness and describe only a fraction of the environment-related business activities in the US Pollution abatement costs were disaggregated by capital expenditures and operating costs, but employment estimates were not included.

3. Methodology

The economic and employment effects of EP expenditures were estimated using the Management Information Services, Inc. model, database, and information system. A simplified version of the MISI model as applied in this study is shown in Fig. 1.

The first step involves translation of environmental expenditures into per unit output requirements from every industry in the economy. Second, the direct output requirements of every industry affected by the expenditures are estimated, and they reflect the production and technology requirements implied by the environmental spending. These direct requirements show, proportionately, how much an industry must purchase from every other industry to produce one unit of output. Direct requirements, however, give rise to subsequent rounds of indirect requirements. The sum of the direct plus the indirect requirements represents the total output requirements from an industry necessary to produce one unit of output.

Economic I–O techniques allow the computation of the direct and the indirect production requirements. Direct industry output requirements are converted into total output requirements from every industry by means of the I–O inverse equations. These equations show not only the direct requirements, but also the second, third, fourth, *n*th round indirect industry and service sector requirements resulting from environmental expenditures. Next, the total output requirements from each industry are used to compute sales volumes, profits, and value added for each industry. Then, using data on manhours, labor requirements, and productivity, and employment requirements, the number of jobs created within each industry are estimated.

The next step requires the conversion of total employment requirements by industry into job requirements for specific occupations and skills. To accomplish this, MISI utilizes data on the occupational composition of the labor force within each industry and estimates job requirements for 700 occupations encompassing the entire US labor force. This permits estimation of the impact of environmental expenditures on jobs for specific occupations.

Utilizing the modeling approach outlined above, the MISI model allows estimation of the effects on employment, personal income, corporate sales and profits, and government tax revenues in the US Estimates can then be developed for detailed industries and occupations.

The final step in the analysis required assessing the economic impacts on individual states, which were estimated using the MISI regional model, which allows the flexibility of specifying multi-state, state, or county levels of detail. Because of the comprehensive nature of the modeling system, these regional impacts are consistent with impacts at the national level.

4. What constitutes an environmental job?

4.1. Ambiguities and questions

As discussed below, we estimate that EP created nearly 5 million jobs in the US in 2003, and these were distributed widely throughout all states and regions within the US But how many of these are "environmental jobs" or "green jobs?" More specifically, what constitutes an "environmental job?" While a definitive analysis of this important

[4] Management Information Services, Inc. *Survey of Jobs and the Environment Issues in Six Midwestern States: Identifying Policy Challenges and Opportunities*. Report prepared for the Joyce Foundation, Chicago, IL, July 2001.

[5] US Department of Commerce, ITA, Office of Environmental Technologies Industries. *Environmental Industry of the United States*, a USDOC/ITA web-accessible briefing generated by Environmental Business International, Inc. for 1999.

[6] US Department of Commerce, Bureau of the Census. *Pollution Abatement Cost and Expenditures: 1999*. MA200(99), November 2002.

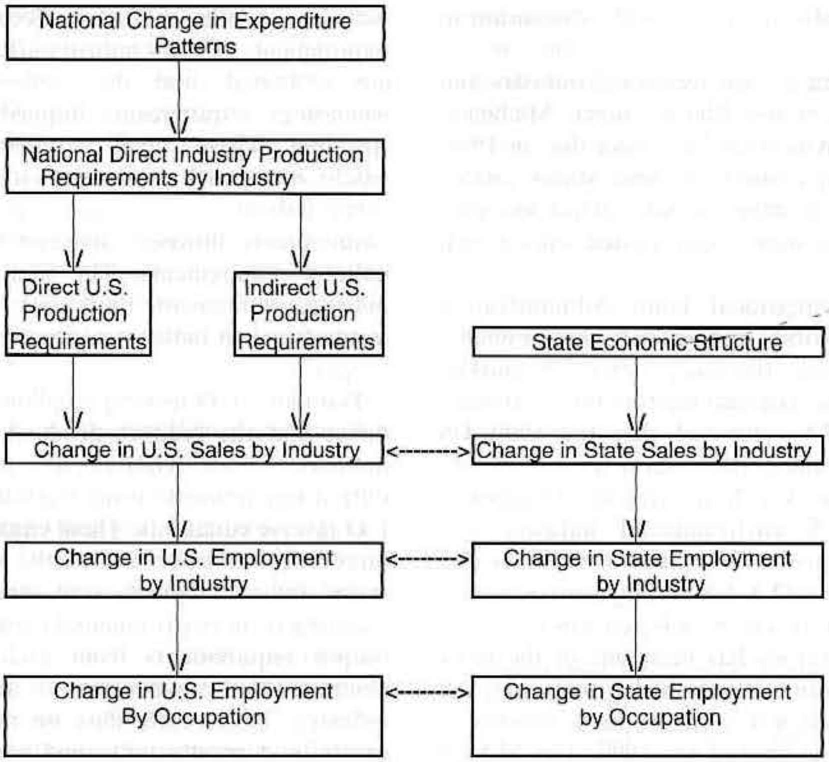


Fig. 1. Use of the MISI model to estimate the economic, employment, and occupational impacts of environmental protection. *Source*: Management Information Services, Inc., 2006.

topic is outside the scope of this report, our review of the literature indicates that there is no rigorous, well-accepted definition of an environmental job. Rather, the definitions used are often loose and contradictory.

Clearly, an ecologist or an environmental engineer would constitute an environmental job, as would an employee of the federal or a state EP agency. However, there are ambiguities. For example, most people would agree that the positions in a firm that assembles and installs solar thermal collectors would be considered environmental jobs. But what about the jobs involved in producing those solar panels, especially if the factory involved used coal-based energy, one of the most controversial fossil fuels in terms of emissions? Here, these manufacturing jobs are included as jobs created indirectly by environmental expenditures.

Most analysts would consider jobs in a recycling plant to be environmental jobs. But what if the recycling plant itself produces air pollution? What about a firm in North Carolina that produces emissions control equipment for power plants in Alabama? It seems clear that the jobs in the North Carolina company should be considered green or environmental jobs, even though the user of the equipment in Alabama may cause pollution in North Carolina. What about environmental engineers and environmental controls specialists working in a coal-fired power plant? What about the workers who produce environmental control equipment for the plant?

There are many firms in the US that produce products for the automotive industry. Should those that produce components for fuel-efficient vehicles be considered part of the environmental industry, but not those that produce components for gas guzzlers? If so, is there any way to accurately distinguish between these? Should all factories producing catalytic converters be considered environmental jobs, even when some of these converters are used on low miles-per-gallon vehicles?

These relevant questions have, in fact, been generated by shifts in environmental policy itself. The early stages of the environmental movement in the 1970s and 1980s focused primarily on "end-of-the pipe" solutions: The remedies focused on cleaning or minimizing air, water, or solid waste pollutants after they had been produced. However, EP has evolved to include entire processes, so, rather than cleaning up at the end of the pipe, the entire manufacturing and servicing processes are being designed to minimize the production of pollutants. Therefore, it is possible that efficient processes designed to produce relatively little waste output could actually result in a decrease in the number of "environmental" jobs if these are defined strictly as "end of the pipe" jobs. Energy efficiency could ultimately result in less need for electric power and could result in the shutting down of a coal-fired electric power plant. While some may view such a shutdown as an environmental plus, many environmental jobs in that power plant involving pollution abatement would be in this case lost.

While solid waste abatement is a major area of environmental concern, does this imply that all persons engaged in trash collection are performing environmental jobs? What part of the tourism industry constitutes "ecotourism," and are all jobs associated with ecotourism green jobs? Are forms of alternative energy green industries, with all jobs counting as environmental jobs?

There is also the issue of how to account for indirect job creation and how broadly or narrowly to define an indirect environmental job. For example, what of ancillary jobs created across the street from a factory producing solar collectors, such as those in a fast food restaurant, dry cleaner, etc. whose customers are primarily the workers at the renewable energy factory. Are these latter jobs also considered to be "indirect" green jobs or environmental jobs? We include such indirect jobs here.

4.2. Definitions and concepts used here

Here, we consider that jobs can be considered to be "green" relative to the way the job was performed previously, i.e., in a production process, a change in technology that reduces waste emissions or energy consumption makes the jobs in that process "greener" than before. Based on extensive research and literature review, we determine that environmental jobs are best understood when viewed in a continuum, with jobs that generate environmental degradation or extraction at one end; a range of greener jobs involving clean production measures and technologies to reduce environmental impacts in the center, and the other end of the spectrum where jobs have a positive environmental impact (see Fig. 1). Using this concept, we define environmental industries and green jobs as those which, as a result of environmental pressures and concerns, have produced the development of products, processes, and services, which specifically target the reduction of environmental impact. Environment-related jobs include those created both directly and indirectly by EP expenditures (Fig. 2).

There exists relatively little rigorous research addressing the practical relationship between EP and job creation. Even some research in this area sponsored by environmental organizations is off the mark, in that it has tended to emphasize jobs creation in classically green activities, such as environmental lawyers or workers in recycling plants. However, while these jobs count as jobs related to the environment, we found that classic environmental jobs constitute only a small portion of the jobs created by EP. The vast majority of the jobs created by EP are standard jobs for accountants, engineers, computer analysts, clerks, factory workers, truck drivers, mechanics, etc. In fact, most of the persons employed in these jobs may not even realize that they owe their livelihood to protecting the environment. For example, as illustrated in Fig. 3, in the US in 2003, we estimate that EP created: More jobs for secretaries (97,900) than for environmental scientists (50,700); more jobs for management analysts (82,600) than for environmental engineers (45,200); more jobs for bookkeepers (71,600) than for hazardous materials workers (33,300); more jobs for janitors (56,400) than for environmental science technicians (25,000); more jobs for computer systems analysts (30,000) than for chemical engineers (8200); and more jobs for truck drivers (25,200) than for biological technicians (12,100).

More generally, arguments stressing the economic benefits and job creation resulting from EP and clean energy initiatives are not currently being made in a rigorous manner, which disaggregates these benefits to a level of detail that is meaningful to policymakers. The level

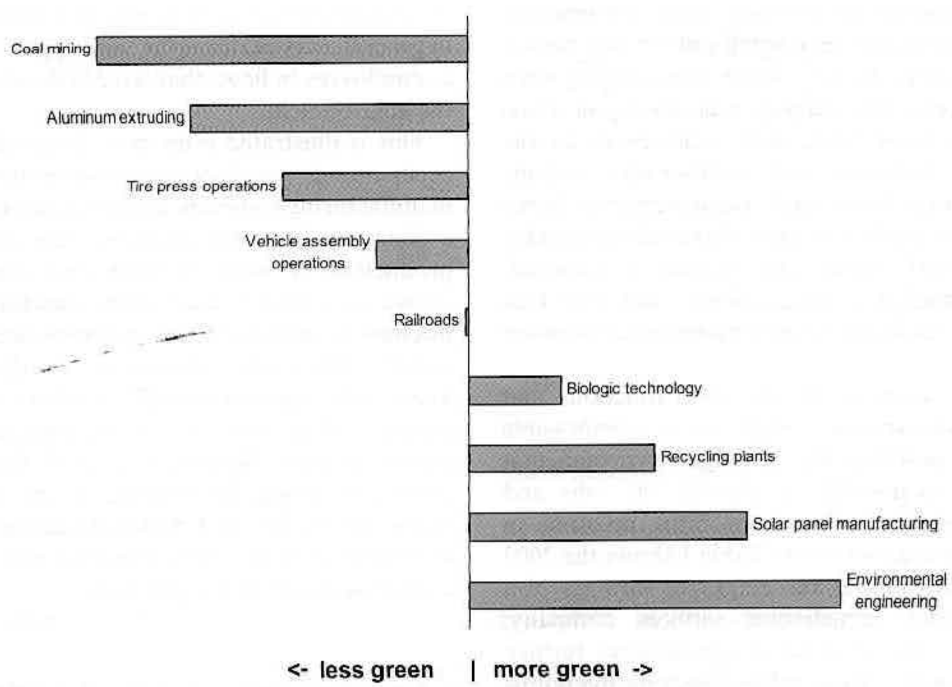


Fig. 2. The environmental job spectrum. *Source*: Management Information Services, Inc., 2006.

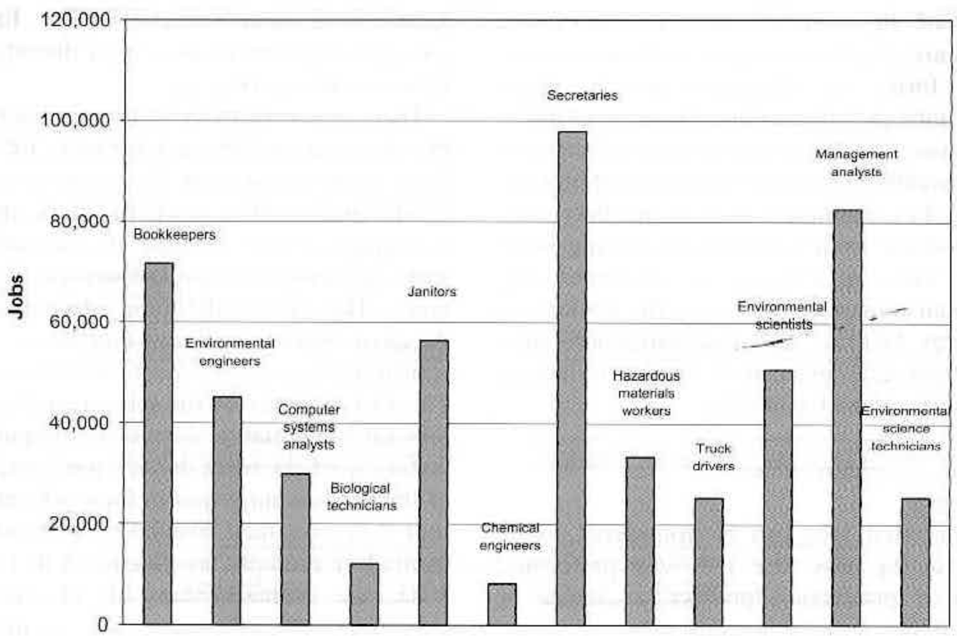


Fig. 3. Selected US jobs created in 2003 by environmental expenditures. *Source*: Management Information Services, Inc., 2006.

of detail required is at the sector, industry, state, city, and county level, and the jobs created have to be identified by industry, category, skill, and specific occupation at the state and local level. The findings summarized here provide data at such levels of detail.

4.3. Jobs distribution in typical environmental companies

There are thousands of environmental companies located throughout the US and they generate jobs for nearly 5 million workers in virtually every community. These firms range from the very small one or two person "mom and pop" shops to very large firms employment thousands of workers; they employ workers at all levels of skills, from the most basic and rudimentary to the very high skilled technical and professional; include environmental service firms and manufacturing firms; include those whose market is local, those whose market is state and regional, those who market is national, and those whose market is international, and they face the same problems, challenges, and opportunities as other companies

Given the wide diversity in the size, function, and technologies of environmental companies, it is impossible to estimate the job profile of the "average" environmental firm. However, it is possible to identify the jobs and earnings profiles of typical types of firms involved in environment-related areas of work. Table 1 shows the 2003 occupational job distribution and employee earnings of a typical environmental remediation services company; Table 2 shows the same data for a typical wind turbine manufacturing company. These tables illustrate the points made above.

First, firms working in the environmental and related areas employ a wide range of workers at all educational and skills levels and at widely differing earnings levels.

Second, even in environmental companies, most of the employees are not classified as "environmental specialists." For example, in the environmental remediation services firm profiled in Table 1, most of the workers are in occupations such as laborers, clerks, bookkeepers, accountants, maintenance workers, cost estimators, etc. All of these employees owe their jobs and livelihoods to EP, but, in general, they perform the same types of activities at work as employees in firms that have little or nothing to do with the environment.

This is illustrated even more forcefully in Table 2. The occupational job distribution of a typical wind turbine manufacturing company differs relatively little from that of a company that manufactures other products. Thus, the production of wind turbines and components requires engine assemblers, machinists, machine tool operators, mechanical and industrial engineers, welders, tool and die makers, mechanics, managers, purchasing agents, etc. These are "environmental" workers only because the company they work for is manufacturing a renewable energy product. Importantly, with the current national angst concerning the erosion of the US manufacturing sector and the loss of US manufacturing jobs, it is relevant to note that many environmental and renewable energy technologies are growing rapidly.[7]

[7]For example, windpower is the most rapidly growing source of electrical power in the world.

Table 1
Typical employee profile of a 100-person Environmental Remediation Services Company, 2003

Occupation	Employees	Earnings
Hazardous materials removal workers	22	$36,204
Septic tank servicers and sewer pipe cleaners	8	30,419
Construction laborers	7	32,382
First-line supervisors/managers of construction/extraction	5	50,673
Truck drivers, heavy and tractor-trailer	5	33,044
General and operations managers	3	86,258
Laborers and freight, stock, and material movers	2	21,620
Truck drivers, light or delivery services	2	27,437
Office clerks	2	23,384
Refuse and recyclable material collectors	2	26,796
Insulation workers	2	32,256
Secretaries (except legal, medical, and executive)	2	25,998
Bookkeeping, accounting, and auditing clerks	2	31,217
Plumbers, pipefitters, and steamfitters	1	41,202
Executive secretaries and administrative assistants	1	36,729
Maintenance and repair workers	1	30,849
Environmental engineering technicians	1	36,939
Operating engineers and other const. equip. operators	1	40,520
First-line supervisors/managers of office/administrative	1	47,576
Chief executives	1	116,435
Construction managers	1	73,994
Cleaners of vehicles and equipment	1	21,704
Cost estimators	1	56,753
Janitors and cleaners	1	25,746
Environmental engineers	1	69,930
Industrial truck and tractor operators	1	27,741
Carpenters	1	38,588
Construction and maintenance painters	1	33,296
Accountants and auditors	1	53,865
Dispatchers (except police, fire, and ambulance)	1	29,537
Water and liquid waste treatment plant and system operators	1	31,049
First-line supervisors/managers of transportation operators	1	46,914
Sales representatives, wholesale and manufacturing	1	42,683
Customer service representatives	1	30,366
First-line supervisors/managers of mechanics and repairers	1	49,088
Environmental scientists and specialists	1	62,003
Receptionists and Information clerks	1	22,775
Environmental science and protection technicians	1	44,867
Other employees	12	47,422
Employee total	100	$39,621

Source: Management Information Services, Inc., 2006.

Table 2
Typical Employee Profile of a 250-person Wind Turbine Manufacturing Company, 2003

Occupation	Employees	Earnings
Engine and other machine assemblers	31	$33,359
Machinists	27	37,191
Team assemblers	16	27,668
Computer-controlled machine tool operators	12	37,254
Mechanical engineers	10	65,772
First-line supervisors/managers of production/operating	10	54,705
Inspectors, testers, sorters, samplers, and weighers	8	37,202
Lathe and turning machine tool setters/operators/tenders	6	36,729
Drilling and boring machine tool setters/operators/tenders	4	36,509
Welders, cutters, solderers, and brazers	4	36,530
Laborers and freight, stock, and material movers	4	28,466
Maintenance and repair workers	4	41,318
Tool and die makers	4	40,047
Grinding/lapping/polishing/buffing machine tool operators	4	31,899
Multiple machine tool setters/operators/tenders	4	37,517
Industrial engineers	3	64,659
Industrial machinery mechanics	3	42,315
Engineering managers	3	99,404
Shipping, receiving, and traffic clerks	3	29,516
General and operations managers	3	110,702
Industrial production managers	3	85,512
Industrial truck and tractor operators	3	31,416
Purchasing agents	3	51,702
Cutting/punching/press machine setters/operators/tenders	3	28,907
Production, planning, and expediting clerks	3	41,601
Milling and planing machine setters/operators/tenders	3	37,380
Mechanical drafters	2	44,090
Customer service representatives	2	36,036
Bookkeeping, accounting, and auditing clerks	2	32,760
Office clerks, general	2	27,227
Sales representatives, wholesale and manufacturing	2	50,757
Janitors and cleaners	2	28,476
Sales engineers	2	66,591
Accountants and auditors	2	54,873
Tool grinders, filers, and sharpeners	2	40,520
Executive secretaries and administrative assistants	2	39,638
Mechanical engineering technicians	2	46,767
Electricians	2	45,570
Other employees	48	45,969
Employee total	250	$42,726

Source: Management Information Services, Inc., 2006.

5. Findings at the national level

We found that, contrary to general public perception and public policy understanding, since the late 1960s, protection of the environment has grown rapidly to become a major sales-generating, profit-making, job-creating industry. Expenditures in the US for EP have grown (in constant 2003 dollars) from $39 billion per year

Table 3
Environmental protection expenditures and jobs in the us economy, 1970–2003

	Expenditures (billions of 2003 dollars)	Jobs (thousands)
1970	\$39	704
1975	77	1352
1980	121	2117
1985	158	2838
1990	204	3517
1995	235	4255
2003	\$301	4974

Source: Management Information Services, Inc., 2006.

in 1970 to \$301 billion per year by 2003—increasing more rapidly than GDP over the same period—see Table 3. If "EP" were a corporation, it would rank higher than the top of the Fortune 500, for our estimate of 2003 EP expenditures (\$301 billion) ranks it higher than the sales of \$259 billion for Wal-Mart, the largest corporation in the US In 2003, EP generated five million jobs distributed widely throughout the nation.

Many companies, whether they realize it or not, owe their profits—and in some cases their existence—to EP expenditures.[8] Many workers, whether they realize it or not, would be unemployed were it not for these expenditures: In 2003, EP created nearly five million jobs distributed widely throughout the nation. To put this into perspective, the size of environment-related employment is over ten times larger than employment in the US pharmaceuticals industry, nearly six times larger than the apparel industry, almost three times larger than the chemical industry, nearly half the employment in hospitals, and almost one-third the size of the entire construction industry.

We estimate that in 2003 protecting the environment generated \$301 billion in total industry sales, \$20 billion in corporate profits, 4.97 million jobs, and \$45 billion in Federal, state, and local government tax revenues.[9] Clearly, providing the goods and services required for EP has become a major US industry with significant effects on the national economy and labor market and on those of individual states.[10]

6. Findings at the state level

As part of the research initiative we have thus far estimated and assessed the environmental industry and jobs in six states: Florida, Michigan, Minnesota, North Carolina, Ohio, and Wisconsin.[11] Our findings are summarized in Tables 4 and 5.

6.1. Aggregate and sectoral findings

Table 4 summarizes the parameters of the environmental industries in each state. The size of the industry in each state differs considerably, from \$5.4 billion in Wisconsin to \$15.4 billion in Florida, generally corresponding to the differences in state GDP. However, the industry share of state GDP differs from a high of 3.9% in Michigan to a low of 2.6% in Minnesota. Similarly, environment-related employment ranges from 220,000 in Florida to 92,000 in Minnesota—again reflecting mainly the differences in the sizes of the state labor forces. Environmental employment ranges from a high of 4.9% of total employment in Ohio to 2.9% in North Carolina.

The shares of each state of the total US environmental industry and environment-related jobs also differ substantially, depending largely on the size of state GDP and labor force. Nevertheless, there are some important differences among the states. For example, while the number of environment-related jobs is about the same in both Michigan and Florida and each state has about 4.4% of the national total, the population of Florida is nearly twice that of Michigan—Florida represents about six percent of the US population while Michigan comprises 3.4%. That is, per capita, the size of the environmental industry in Michigan is nearly twice that of the industry in Florida.

Table 5 shows the industry sector distribution of total employment and of environmental employment in each of the six states. It and Table 4 illustrate that environment-related jobs are distributed among all sectors, but are heavily concentrated in several. Significant portions of the environmental jobs in each state are in the public administration sector which, given the public nature of EP, is to be expected. However, most of the environmental jobs in the states are in the private sector, and focusing on these reveals that they are heavily concentrated in several sectors.

[8] In this paper, "expenditures" refers to all public and private spending in the environmental sector (EP spending) and is used interchangeably with "sales."

[9] The national estimates have been developed by MISI beginning in 1986 using the model and database summarized in Section 2, and have been updated periodically over the past two decades. The six states discussed here were selected for detailed analysis at the request of the funders of the work. The overall project goal is to eventually conduct similar analyses for as many states as possible and, at present, estimates are being developed for three more states—Arizona, California, and Connecticut—to provide better geographic coverage. In addition, analyses for states such as New York, Oregon, and Washington that have traditionally been viewed as environmentally aggressive can help determine if environmental job growth has been more rapid in these states. Findings will be posted on the MISI web site when available: www.misi-net.com

[10] As discussed, all estimates of the size of the environmental industry rely critically on the exact definition of the industry. Since there is no official definition, estimates of the size of the environmental industry differ according to the source. In MISI's case, the definition of the industry includes human and environmental sustainability principles, and MISI's estimates thus include a broader range of environmental activities in the economy than some other definitions that have been developed.

[11] The detailed findings for each state are available on the MISI web site: www.misi-net.com

Table 4
Summary of the environmental industries in six states in 2003

	Environmental industry (billions) ($)	Environmental jobs	Environmental industry as a percent of		State environmental industry as a percent of		Private sector environmental jobs	
			State GDP (%)	State jobs (%)	Total US environmental industry (%)	Total US environmental jobs (%)	Manufacturing (%)	Professional, scientific, technical (%)
Florida	15.4	220,000	3.1	3.0	5.0	4.4	7	22
Michigan	12.9	217,000	3.9	4.9	4.3	4.4	29	29
Minnesota	5.1	92,000	2.6	3.5	1.7	1.8	21	23
North Carolina	9.1	112,000	3.1	2.9	3.0	2.9	24	20
Ohio	12.2	176,000	3.2	3.3	4.1	3.5	29	25
Wisconsin	5.4	97,000	2.9	3.5	1.8	2.0	31	16

Source: Management Information Services, Inc., 2006.

Of particular note is that the private sector environmental industry is more manufacturing intensive than other average private sector activity in the states. As shown in Fig. 4, in Florida, 7.4% of private sector jobs in the environmental industry is in manufacturing, compared to 6.2% in manufacturing among all private sector jobs in the state; in Michigan, 29% of private sector jobs in the environmental industry is in manufacturing, compared to 17% in manufacturing among all private sector jobs; in Minnesota, the comparable shares are 21% and 15%; in North Carolina, the comparable shares are 24% and 19%; in Ohio, the comparable shares are 29% and 18%; in Wisconsin, the comparable shares are 31% and 21%.

The jobs concentration is even more pronounced with respect to employment in the professional, scientific, and technical services sector. As shown in Fig. 5, in Florida, 22% of private sector environmental jobs is in professional, scientific, and technical services, compared to 6% of all private sector jobs in the state; in Michigan, 29% of private sector environmental jobs is in professional, scientific, and technical services, compared to 8% of all private sector jobs in the state; in Minnesota, the comparable shares are 23% and 5%; in North Carolina, the comparable shares are 20% and 5%; in Ohio, the comparable shares are 25% and 7%; in Wisconsin, the comparable shares are 16% and 4%.

Conversely, there are relatively few private sector environmental jobs in other parts of the states' economies, including retail trade, finance and insurance, health care and social services, and transportation and warehousing.

The concentration of environmental jobs within certain industrial sectors is instructive and interesting. While accounting for only abut 3–5% of total employment in each state, the industry sector composition of environmental employment is highly skewed in favor of certain sectors including manufacturing. This indicates that investments in the environment will provide a greater than proportionate assist to the states' manufacturing sectors. All of these states are seeking to modernize and expand their high-tech industrial and manufacturing bases. Table 5 and Fig. 4 indicate that the environmental industry can aid in this objective.

Similarly, environmental investments generate, proportionately, 3–4 times as many jobs in professional, scientific, and technical services as the state averages. Jobs in this sector include the high-skilled, high-wage, technical, and professional jobs that all states seek to attract and retain. Table 5 and Fig. 5 indicate that investments in EP can be of considerable assistance here.

6.2. *Environmental jobs by occupation*

We disaggregated environmental employment in each state by specific occupations and skills. The results for Florida and Michigan are representative of those for the six states, and this information for selected occupations is given in Tables 6 and 7. These tables illustrate that environmental jobs are widely distributed among all occupations and skill levels and, while the number of jobs created in different occupations differs substantially, employment in virtually all occupations is generated by environmental spending.

As noted in Section 4, the vast majority of the jobs created by EP are standard jobs for accountants, engineers, computer analysts, clerks, factory workers, truck drivers, mechanics, etc., and most of the persons employed in these jobs may not even realize that they owe their livelihood to protecting the environment. This is further illustrated in Tables 6 and 7, which list the jobs created by EP in Florida and Michigan in 2003 within selected occupations. For example, Table 6 shows that EP generated in Florida: More jobs for sheet metal workers (821) than for geoscientists (241); more jobs for office clerks (4968) than for environmental engineers (2545); more jobs for executive secretaries (2432) than for landscape architects (313); more jobs for janitors (1827) than for natural science managers (207); more jobs for electricians (708) than for chemists (242); more jobs for truck drivers (2870) than for septic tank servicers (2181); more jobs for financial managers (684) than for conservation scientists (371); more jobs for

Table 5
Environmental-related jobs in each state, by industry

Industry	Florida employment		Michigan employment		Minnesota employment		N. Carolina employment		Ohio employment		Wisconsin employment	
	Total	Environmental	Total	Environmental	Total	Environmental	Total	Environmental	Total	Environmental	Total	Environmental
Agriculture, forestry, fishing and hunting	2300	192	3515	216	800	86	3700	120	1564	129	2500	208
Mining	4900	459	5226	627	5200	515	4000	293	10,505	678	1300	145
Utilities	26,800	4973	24,136	6914	12,000	2902	14,000	2114	26,109	5949	11,600	2782
Construction	445,900	9966	173,244	8633	125,200	4497	211,800	4732	212,409	7061	123,500	4295
Manufacturing	388,800	9849	659,736	38,895	344,300	11,974	604,300	14,013	805,716	28,149	506,500	17,400
Wholesale trade	313,200	3692	178,545	4021	127,800	2151	163,600	1827	243,493	3634	113,000	1752
Retail trade	920,400	5833	503,576	351	301,700	1778	432,500	2582	591,557	322	319,000	1962
Transportation and warehousing	202,100	1300	90,412	544	80,100	507	110,700	632	130,002	516	94,600	555
Information	171,800	4278	86,397	170	62,600	1751	75,600	1797	103,334	148	49,700	1382
Finance and insurance	330,900	1962	168,065	202	138,100	1062	143,700	855	248,897	209	129,800	861
Real estate and rental and leasing	153,400	1680	61,676	278	37,900	527	47,800	577	66,212	248	27,900	416
Professional, scientific, and technical services	384,400	28,606	195,553	39,432	118,200	12,922	146,300	11,616	221,765	24,657	89,000	9341
Management of companies and enterprises	65,600	1032	152,641	2188	59,000	1385	61,200	971	134,502	1848	37,600	861
Administrative/support/waste management/ remediation services	807,500	41,971	294,857	25,287	117,300	7622	213,700	10,001	319,058	17,242	118,200	7586
Educational services	108,400	3198	70,286	2537	48,400	1676	61,600	1753	97,489	3186	46,100	1807
Health care and social assistance	777,200	4364	516,974	1269	318,300	2099	366,600	1848	678,618	1205	320,500	2330
Arts, entertainment, and recreation	157,200	1030	53,009	449	36,900	247	44,000	240	58,265	399	35,500	229
Accommodation and food services	651,300	5286	327,545	188	196,200	1525	291,000	1837	410,303	187	209,500	1641
Other services	317,800	3107	175,892	2676	118,900	1330	162,400	1335	229,701	2465	131,300	1310
Public administration	1,055,500	86,723	670,515	81,624	402,400	35,545	644,600	52,865	801,500	77,877	411,800	40,337
State total	7,285,400	219,500	4,411,800	216,500	2,651,300	92,100	3,803,100	112,007	5,390,999	176,109	2,778,900	97,200

Source: Management Information Services, Inc., 2006.

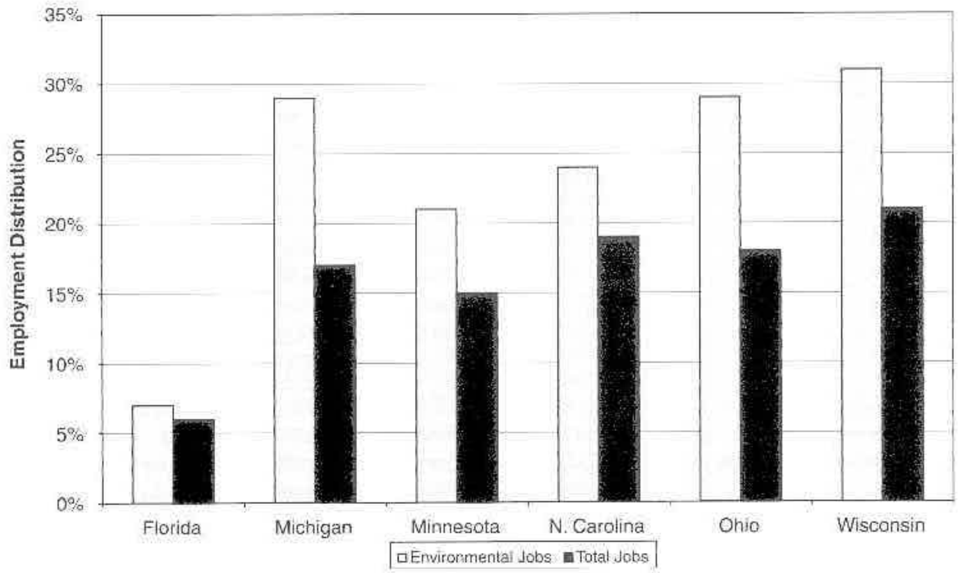


Fig. 4. Private sector manufacturing jobs. *Source*: Management Information Services, Inc., 2006.

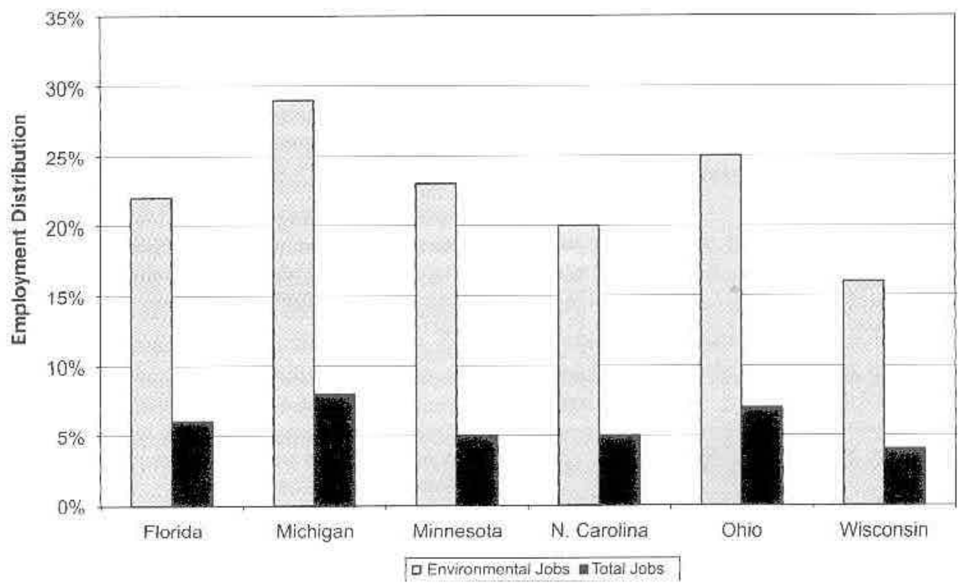


Fig. 5. Private sector professional, scientific and technical jobs. *Source*: Management Information Services, Inc., 2006.

management analysts (2049) than for environmental engineering technicians (1289); and more jobs for computer software engineers (1839) than for hazardous material removal workers (1267).

Table 7 shows similar findings for Michigan. Thus, many workers in Florida and Michigan are dependent on EP for their employment, although they often would have no way of recognizing that connection unless it is brought to their attention.

The importance of environmental spending for jobs in some occupations is much greater than in others. For some occupations, such as environmental scientists and specialists, environmental engineers, hazardous materials workers, water and liquid waste treatment plant operators, environmental science protection technicians, refuse and recyclable material collectors, and environmental engineering technicians, virtually all of the demand in both states is created by EP activities. This is hardly surprising, for most of these jobs are clearly identifiable as "environmental" jobs.

However, in many occupations not traditionally identified as environment-related, a greater than proportionate share of the jobs is also generated by EP. On average, environment-related employment in Florida comprises only 3% of total employment and in Michigan comprises 4.9%, in 2003 EP expenditures generated jobs for a greater than proportionate share—as much as 10% or more—of many professional occupations in the two states, including chemists, civil engineers, computer software engineers, electronics engineers, geoscientists, landscape architects, medical scientists, natural sciences managers, surveyors,

Table 6
Environmental jobs generated in florida in 2003, by selected occupations

Occupation	Jobs
Accountants and auditors	1272
Bookkeeping and accounting clerks	2092
Cashiers	3591
Chemists	242
Computer software engineers	1873
Conservation scientists	371
Customer service representatives	2334
Electricians	708
Electronics engineers	781
Environmental engineers	2545
Environmental engineering technicians	1289
Environmental scientists and specialists	5659
Executive secretaries and administrative assistants	2432
Financial managers	684
Forest and conservation workers	199
Geoscientists	241
Graphic designers	296
Hazardous material removal workers	1267
Inspectors, testers, and sorters	323
Janitors and cleaners	1827
Landscape architects	313
Mechanical engineers	250
Management analysts	2049
Marketing managers	454
Medical scientists, except epidemiologists	255
Natural science managers	207
Office clerks	4949
Pest control workers	1161
Security guards	1614
Septic tank servicers and sewer pipe cleaners	2141
Sheet metal workers	821
Stock clerks	2587
Training and development specialists	431
Truck drivers	2870
Water and liquid waste treatment plant operators	5484
Welders and Solderers	328

Source: Management Information Services, Inc., 2006.

Table 7
Environmental jobs generated in Michigan in 2003, by selected occupations

Occupation	Jobs
Accountants and auditors	1780
Chemical engineers	197
Computer and information systems managers	535
Construction laborers	880
Customer service representative	2425
Electricians	1079
Engine and other machine assemblers	186
Environmental engineers	1382
Environmental scientists and specialists	1523
Employment, recruitment, and placement specialists	525
Financial analysts	353
Forest and conservation technicians	190
Forging machine setters, operators, and tenders	204
Geoscientists, except hydrologists and geographers	272
Hazardous material removal workers	1210
Human resource managers	297
Industrial engineers	739
Industrial machinery mechanics	464
Inspectors, testers, and sorters	1161
Janitors and cleaners	3040
Landscaping and grounds workers	1101
Machinists	966
Management analysts	1134
Marketing managers	311
Mechanical engineering technicians	307
Medical scientists, except epidemiologists	225
Office clerks	4118
Packers and packagers	952
Receptionists and information clerks	1512
Refuse and recyclable material collectors	5454
Sales representatives, technical and scientific products	563
Secretaries	2522
Security guards	1115
Septic tank services and sewer pipe cleaners	702
Tool and die makers	524
Truck drivers, heavy and tractor trailer	2176
Water and liquid waste treatment plant operators	5130
Word processors and typists	523

Source: Management Information Services, Inc., 2006.

urban and regional planners, chemical engineers, and engineering managers.

For many other occupations, also not traditionally identified as environment-related, a greater than proportionate share of the jobs is also generated by EP. On average, environment-related employment in Florida comprises only 3% of total employment and in Michigan 4.9%, in 2003 EP generated jobs for as much as 10% or more of many highly skilled, technical occupations in the two states, including architectural and civil drafters, chemical technicians, civil engineering technicians, electrical and electronics engineering technicians, electrical and electronics equipment assemblers, electrical and electronics drafters, fiberglass laminators and fabricators, forest and conservation technicians, heating, air conditioning, and refrigeration mechanics and installers, industrial engineering technicians, surveying and mapping technicians, chemical plant and system operators, electrical and electronics repairers, engine and other machine assemblers, surveying and mapping technicians, and network systems and data communications analysts.

The above findings are significant for they indicate that EP creates jobs in greater than proportionate share in two categories that Florida and Michigan—and other states—are eager to attract: (i) college-educated professional workers, many with advanced degrees, and (ii) highly skilled, technical workers, with advanced training and technical expertise, many of them in the manufacturing sector. EP thus generates jobs that are disproportionately for highly skilled, well-paid, technical and professional workers, who in turn underpin and provide foundation for entrepreneurship and economic growth.

Our work thus demonstrates that EP can form an important part of a strategy for states based on attracting and retaining professional, scientific, technical, high-skilled, well paying jobs, including manufacturing jobs. While a successful strategy must have other components as ell, rarely has any state recognized the economic and jobs benefits that could flow from specifically encouraging the development of environmental and environment-related industries as an economic development initiative. Indeed, usually the opposite is the case: Most states usually tend to view EP as economically negative.[12]

Another important finding derived here is the significance of the environmental industry compared to other sectors of the state economies. For example, the tourism industry generates about 540,000 jobs in the Florida, and this state well recognizes the key role that tourism plays in the state economy. Here, we estimate that environment-related jobs in Florida total 220,000—jobs that tend to be more highly skilled and better paying than those in the tourism sector. This fact is not widely known or appreciated by state policy-makers.

Comparison to other estimates of environmental spending

Aside from the estimates presented here, the only other comprehensive, consistent time series of estimates of US environmental expenditures over the past four decades are those developed by Environmental Business, International (EBI).[13] The MISI and EBI data series are not strictly comparable. For example, MISI estimates environment-related spending using the expenditure concept and disaggregates spending by media (air, water, land, etc.) and other categories such as R&D, energy-related environmental programs, and so forth. EBI focuses on revenues to business and classifies spending into services (analytical, hazardous waste, consulting & engineering, etc.) equipment (air pollution control, waste management, instruments & information, etc.), and resources (water utilities, resource recovery, and clean energy & power).

A comparison of the MISI and EBI estimates[14] is given in Table 8, which shows that:

During the 1970s, the EBI estimates of environmental expenditures were significantly higher than the MISI estimates.

Table 8
Comparison of estimates of the growth of environmental expenditures in the US (expenditures in billions of 2003 dollars)

	MISI		EBI[a]	
	Expenditures (millions/$)	Growth (%)	Expenditures (millions)	Growth (%)
1970	39		73	
1975	77	97	100	37
1980	121	57	125	25
1985	158	31	148	18
1990	204	29	201	36
1995	235	15	210	5
2000	273	16	221	5
2003	301	10	227	3
2010	357	19	268	18
2015	398	11	NA[b]	
2020	439	10	NA	

Source: Management Information Services, Inc. and Environmental Business International, Inc., 2006.

[a]EBI expenditures in current dollars were converted by MISI to 2003 dollars.

[b]NA, not available. EBI did not forecast expenditures beyond 2010.

- From 1980 through 1995, the MISI and EBI estimates were roughly comparable
- By 2000, the MISI estimates were larger than the EBI estimates.
- The percentage growth rates in expenditures over the past four decades were roughly comparable, and both data sets show the rate of increase in environmental spending decreasing after 1970.
- The forecast rate of growth of environmental spending through 2010 by both EBI and MISI are nearly identical, although from different bases.
- EBI shows the rate of growth of expenditures to be between three and 5% from 1995 to 2003, whereas MISI shows the rates of growth to be considerably higher, although declining.
- MISI forecasts that from 2010 to 2020, environmental expenditures will increase by 23%, whereas EBI presents no forecasts beyond 2010.

8. Conclusions and suggestions for further research

8.1. Findings at the national level

Our first major finding is that EP, economic growth, and jobs creation can be complementary and compatible: Investments in EP can create jobs, not destroy them.[15]

[12]These policies differ considerably among the states, and some states have belatedly begun to recognize the economic benefits of environmental protection. For example, Florida has initiated a major Everglades restoration program and has prohibited offshore drilling, Michigan has implemented a hydrogen program, Arizona has aggressively promoted solar and wind, and Washington is initiating an ambitious biomass program.

[13]The EBI data are available for purchase at www.ebiusa.com. In 1990, EPA published estimates of environmental costs (US Environmental Protection Agency, Office of Policy, Planning, and Evaluation. *Environmental Investments: The Cost of a Clean Environment*. EPA-230-11-90-083, November 1990) and Pace University published estimates of the environmental costs of electricity (Richard Ottinger et al., *Environmental Costs of Electricity*. New York: Oceana publications, 1990). However, no time series data are available for these data and the estimates are not comparable to the MISI estimates.

[14]EBI present its estimates in current dollars. For comparison here, MISI converted the EBI current dollar estimates to constant 2003 dollar estimates using the GDP deflator series.

[15]While environmental protection both creates and displaces jobs, we have found the net jobs effect to be strongly positive, although jobs impacts will vary from case to case. Further, even when the net jobs effect is strongly positive, it must be recognized that significant job displacement may be occurring. For example, in analyzing the likely economic and jobs effects of enhanced CAFE standards, we estimated that by 2020 347,000 net jobs would be created. However, this estimate was the combination of

This finding is important because it differs from what many legislators and policy-makers currently believe.

Second, contrary to general public perception and public policy understanding, since the late 1960s protection of the environment has grown rapidly to become a major sales-generating, profit-making, job-creating industry—$300 billion/year. and 5 million jobs in 2003. The size and the job creating potential of the environmental industry is something that few are aware of.

Third, the vast majority of the five million jobs created by EP are standard jobs for accountants, engineers, computer analysts, clerks, factory workers, truck drivers, etc., and the classic environmental job (environmental engineer, ecologist, conservation technician, etc.) constitutes only a small portion of the jobs created. In fact, most of the persons employed in the jobs created may not even realize that they owe their livelihood to protecting the environment.

This finding is important for, even recognizing that EP is good for the economy and is creating 5 million jobs, the first impression is likely that these are jobs for environment specialists, ecologists, environmental regulators, etc. We found that jobs for all occupations and skills are generated, and this should be of interest to organized labor, trade and professional associations, and policy-makers.

8.2. *Findings at the state level*

Our first major finding at the state level, derived from detailed analyses of the environmental industry and jobs in Florida, Michigan, Minnesota, North Carolina, Ohio, and Wisconsin, is that the overall relationship between state environmental policies and economic/job growth is positive, not negative. States can and do have strong economies and simultaneously protect the environment, and states with the strongest environmental records also have the best job opportunities and climate for long-term economic development.

This is a key finding. In our analysis of the six states we found that all of them assume that there is a negative relationship between protecting the environment and economic and job growth. Thus, the states' policies relating to EP and economic/job development focus on "reforming," "streamlining," and "rationalizing" environmental rules and regulations (a euphemism for weakening them), "simplifying" and "accelerating" environmental permitting, and otherwise sacrificing the environment to economic growth and job creation. Hopefully, the research reported here will begin to change these state attitudes and polices.

(*footnote continued*)
total gross job creation of 433,000 jobs and the displacement of 86,000 jobs. That is, while nearly 350,000 *net* jobs would be created, nearly 90,000 jobs would still be lost. This has obvious policy implications. See Roger H. Bezdek and Robert M. Wendling, "Fuel Efficiency and the Economy," *American Scientist*, op. cit., and Roger H. Bezdek and Robert M. Wendling, "Potential Long-term Impacts of Changes in US Vehicle Fuel Efficiency Standards," *Energy Policy*, op. cit.

Second, environmental jobs in each of the states are concentrated within a number of sectors, including manufacturing and professional, information, and scientific, and technical services, and this is significant because the states are seeking to modernize and expand their high-tech industrial and manufacturing bases. Thus not only is the relationship between EP and jobs positive, but the types of jobs created are disproportionately scientific, professional, technical, high-skilled, manufacturing, and high-wage jobs—the very types of jobs that all states are attempting to retrain and attract. These types of jobs are a prerequisite for a prosperous, middle class society able to support state and local governments with tax revenues—which states already recognize. Of particular note, in the six states studied thus far data show that investments in the environment will provide a greater than proportionate assist to the manufacturing sector.

Finally, EP can form an important part of states' economic development strategies, and there is no inherent institutional impediment in any state to using existing economic assistance policies and incentives to facilitate development of environmental industries and jobs. This is a key policy finding because, at present, none of the states we examined appreciates this potential: (i) no state has integrated environmental industry and job development into its general strategic or economic development plan; (ii) state environmental departments and agencies have little or no focus on environmental employment or job development; (iii) state labor and workforce departments and agencies have little or no focus on environmental industries or jobs.

Each state is home to diverse environmental companies, many global leaders in their field,[16] but their strong role in employment generation is largely overlooked in economic development initiatives and policy. Altering states' perceptions and policies here is essential.

8.3. *Suggestions for further research*

Our work has identified several areas requiring further research. First, a more rigorous and generally accepted definition of what constitutes an "environment-related job" is required. Environmental advocates have tended to identify the more glamorous types of jobs, such as ecologist, wildlife biologist, conservation specialist, solar energy researcher, etc., but we found that the overwhelming majority of environment-related jobs are for the standard occupations, skills, and professions. Nevertheless, the numbers and types of jobs—both in general and in specific industries and firms—are in need of much additional research.

[16] As part of this research project, we identified and assessed a representative sample of environmental firms in each state selected for heterogeneity with respect to size, geographic location, and services and products provided. These findings are available on the MISI web site: www.misi-net.com

Second, the empirical work reported here needs to be expanded. While we have analyzed the environmental industries in six states, it remains to be determined how representative our findings are for the rest of the US At least as important, our analyses of each state were not comprehensive, and much more detailed assessment of several individual states is required. Such an assessment would look in detail below the state level to specific geographic regions and industries and conduct in-depth analyses of specific environmental firms.

Finally, it would be useful to have international perspective. We found that in the US environment-related activities account for 3–5% of national and state GDP and jobs. Using generally consistent concepts and definitions, it would be interesting to determine how these estimates compare to estimates of environmental industries and jobs in other developed nations. International comparative analyses of detailed results at the sector, industrial, and occupational level would be especially useful.

References

Arnold, F.S., Forest, A.S., Dujack, S.R., 1999. Environmental protection: is it bad for the economy? Report prepared for the US Environmental Protection Agency.

Barrett, J.P., Hoerner, J.A., 2002. Clean Energy and Jobs: A Comprehensive Approach to Climate Change and Energy Policy. Economic Policy Institute, Washington, DC (Clean Energy and Jobs: A Comprehensive Approach to Climate Change and Energy Policy, Redefining Progress, Oakland, CA, 2004).

Bernow, S., Dougherty, W., Duckworth, M., Kartha, S., Lazams, M., Ruth, M., 1999. America's Global Warming Solutions. Tellus Institute and Stockholm Environment Institute, Boston, MA.

Bezdek, R.H., 1993. Environment and Economy: What's the Bottom Line? Environment 35(7), 7–32 (The Economy, Jobs, and the Environment. In: Proceedings of the GEMI '95: Environment and Sustainable Development. Arlington, VA. March 1995, pp. 65–79).

Bezdek, R.H., Wendling, R.M., 1989. Acid rain abatement: costs and benefits. International Journal of Management Science 17 (3), 251–261.

Bezdek, R.H., Wendling, R.M., 2005. Potential long-term impacts of changes in US vehicle fuel efficiency standards. Energy Policy 33(3), 407–419 (Fuel efficiency and the economy. American Scientist, March).

Bliese, J.R., 1999. The Great "Environment Versus Economy" Myth. Brownstone Policy Institute, New York.

Friedman, D., et al., 2001. Drilling in Detroit: Tapping Automaker Ingenuity to Build Safe and Efficient Automobiles. Union of Concerned Scientists, UCS Publications, Cambridge, MA.

Geller, H., DeCicco, J., Laitner, S., 1992. Energy Efficiency and job creation: the employment and income benefits from investing in energy conservation technologies. Report Number ED922, American Council for an Energy Efficient Economy, Washington, DC, November.

Goodstein, E.B., 1994. Jobs and the Environment: The Myth of a National Trade-Off. Economic Policy Institute, Washington, DC (Jobs or the Environment? No Trade-off," Challenge (January–February 1995), pp. 41–45 (The Trade-Off Myth: Fact and Fiction About Jobs and the Environment. Island Press, New York, 1999; Eban B. Goodstein, Hart Hodges, Polluted Data: Overestimating the Costs of Environmental Regulation. The American Prospect, November/December 1997).

Jorgenson, D., Wilcoxen, P., 1990. Environmental regulation and US economic growth. RAND Journal of Economics 21 (2), 153–167.

Jorgenson, D., Goettle, R., Gaynor, D., Wilcoxen, P., Slesnick, D., 1993. The Clean Air Act and the US Economy: Final Report of Results and Findings. Environmental Economics Report Inventory, August 27.

Laitner, S., DeCicco, J., Elliott, N., Geller, H., Goldberg, M., Morris, R., Nadel, S., 1994. Energy Efficiency as an Investment in Ohio's Economic Future. American Council for An Energy-Efficient Economy, Washington, DC, November (Energy Efficiency and Economic Development in the Midwest. American Council for an Energy-Efficient Economy, April 1995; Energy Efficiency and Economic Development in New York, New Jersey, and Pennsylvania. American Council for an Energy-Efficient Economy, February 1997).

Leninson, A., Taylor, M.S., 2004. Unmasking the pollution haven effect. National Bureau of Economic Research Working Paper No. W10629, July.

Management Information Services, Inc., 1993. Potential economic and employment Impact on the US economy of increased exports of environmental and energy efficiency technologies under NAFTA. Report prepared for the White House.

Meyer, S.S., 1992. Environmentalism and Economic Prosperity: Testing the Environmental Impact Hypothesis. Massachusetts Institute of Technology Project on Environmental Policies and Policy, Cambridge, MA.

Morgenstern, R.D., Pizer, W.A., Ahih, J.S., 1998. Jobs Versus Environment: Is there a Trade-off? Resources for the Future, Washington, DC.

Motor Vehicle Manufactures Association, 1990. US employment effect of higher fuel economy standards. Unpublished Paper, January 30 (The MVMA is now known as the American Automobile Manufacturers Association).

New Energy for America, 2004. The Apollo Jobs Report, Apollo Alliance.

Regional Economics Applications Laboratory, 2002. Job Jolt: The Economic Impacts of Repowering the Midwest. University of Illinois, Chicago.

Renner, M., 2000. Working for the environment: a growing source of jobs. Worldwatch Paper 152, Worldwatch Institute, Washington, DC.

Repetto, R., 1995. Jobs, Competitiveness, and Environmental Regulations: What are the Real Issues? World Resources Institute.

Templet, P.H., 1995. The Positive Relationship Between Jobs, Environment, and Economy. Spectrum of the Institute of Electrical and Electronics Engineers.

Teotia, A., et al., 1999. CAFE compliance by light trucks: economic impacts of clean diesel engines. Energy Policy 27, 889–900.

Union of Concerned Scientists, 2002. Fuel Economy as an Engine for Job Growth. Cambridge, MA.

Union of Concerned Scientists, 2004. A 20 Percent National Renewable Electricity Standard Will Create Jobs and Save Consumers Money. Cambridge, MA.

Yapijakis, C., 1999. The Myth of 'Jobs Versus the Environment. Environmental Research Laboratory, Cooper Union School of Engineering, New York.

ELSEVIER

Journal of Environmental Management 86 (2008) 80–87

Journal of Environmental Management

www.elsevier.com/locate/jenvman

A multivariate statistical analysis of surface water chemistry data—The Ankobra Basin, Ghana

Sandow M. Yidana[a,*], Duke Ophori[a], Bruce Banoeng-Yakubo[b]

[a]*Department of Earth and Environmental Studies, Montclair State University, Montclair, NJ 07043, USA*
[b]*Department of Geology, University of Ghana, Box LG 58, Legon, Ghana*

Received 29 October 2006; received in revised form 19 November 2006; accepted 23 November 2006
Available online 16 January 2007

Abstract

R-mode hierarchical cluster and principal component analysis (PCA) were simultaneously applied to surface water hydrochemical data from three different locations, Ankwaso, Dominase and Prestea, along the Ankobra Basin, Ghana, to extract principal factors corresponding to the different sources of variation in the hydrochemistry, with the objective of defining the main controls on the hydrochemistry at the basin scale. Using the Kaiser criterion, principal components (PC) were extracted from the data and rotated using varimax normalization, for each location. The varimax rotation ensured that variation in the data was maximized for easy interpretation of the results. The analysis reduced 30, 33 and 33 data points, respectively, for Ankwaso, Dominase and Prestea to four, three and four PC representing the sources of variation in the hydrochemistry at the three different locations. Though the PC analysis proved to be more robust at unveiling the sources of variation in the hydrochemistry than the *R*-mode hierarchical cluster analysis (HCA), the combined use of both techniques resulted in more reliable interpretations of the hydrochemistry. On the basis of these analyses, the hydrochemistry of the basin is controlled largely by the weathering of minerals (silicates, carbonates, gypsum and apatite) from the underlying meta-sediments of the Birimian and Tarkwaian Systems, and the decay of organic matter from the heavily forested regions. Concentrations of the major chemical parameters are within naturally acceptable limits and do not pose threats to the local ecology and humans. There is no strong evidence of high anthropogenic impacts on the major anions and cations used for this research, though there are variations at the different locations studied. The hydrochemistry at Ankwaso is principally controlled by the weathering of silicate minerals, whereas those of Dominase and Prestea are, respectively, influenced by precipitation and domestic wastewaters, and the decay of organic matter.

Keywords: Ankobra; Dendrogram; Hierarchical cluster analysis; Principal component analysis; Varimax rotation; Weathering

1. Introduction

In the current world economic paradigms, sustainable socioeconomic development of every community depends much on the sustainability of the available water resources. Water of adequate quantity and quality is required to meet growing household, industrial and agricultural needs. Surface water quality is a very sensitive issue, which transcends national boundaries. It is influenced by many factors, including atmospheric chemistry, the underlying geology, the vegetation (or organic matter decay), and anthropogenic agents. The solubility of minerals in water places an upper limit on the maximum amounts of certain species of chemicals in natural waters. Some minerals like carbonates and evaporites dissolve quickly and change the composition of water faster, while other minerals like silicates dissolve more slowly and have less conspicuous effects on the composition of water. Temperature also plays a vital role in controlling the chemical and biological composition of a freshwater body. Previous studies (Frape et al., 1984; Garrels and McKenzie, 1967; Hem, 1989; Hartman et al., 2005) have revealed that the chemistry of natural waters can often be traced to the reaction of these waters with sediments or rocks through which they flow. Based on catchments studies in the USA, Walling (1980) observed differences in the weathering mechanisms of different rocks. Walling (1980) concluded that total dissolved solids in the

*Corresponding author. Tel.: +1 973 655 3457; fax: +1 973 655 4072.
E-mail address: yidanas1@mail.montclair.edu (S.M. Yidana).

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.11.023

water from limestones, volcanics and sand and gravel is almost independent of the amount of runoff. These differences in behavior accrue from the differences in the solubilities of the minerals present in these rocks. Since ultrabasic rocks are rich in pyroxenes and olivine, the predominant element expected from these rocks in freshwater is Mg^{2+}. Similarly, Ca^{2+} is the dominant cationic contribution from calcareous soils, and when Ca^{2+} and Mg^{2+} are present in about the same concentrations in a water body, they are probably derived from dolomite ($CaMg(CO_3)_2$). Atmospheric chemistry is the main source of Cl^- in freshwater bodies.

Organic matter decay can contribute phosphates, bicarbonate, nitrate, ammonia and dissolved solids to the chemistry of surface waters. Organic matter decomposition also has the tendency to reduce the dissolved oxygen (DO) content of natural water. Factors such as the residence time of water, the temperature and the presence of other ions play important roles in determining how much organic matter decays. Anthropogenic forces have the immense tendency to accelerate natural processes that affect water quality. Surface mining, mineral leaching and the uncontrolled use of toxics in industry can adversely affect the suitability of surface water resources for many purposes. Where chemical fertilizers are being used in agriculture, nitrate and ammonia levels in surface waters have the tendency to rise to unacceptable levels. Many of the factors that influence water quality vary on spatial and temporal scales. A complete assessment of the role of these factors would require the consideration of both scales in the analysis.

This paper applies cluster and factor analysis to assess the main controls on the chemistry of surface water resources from the Ankobra Basin at three different locations along its course. It makes use of the strength inherent in the joint application of two multivariate variable classification tools, to establish the source of variance in the hydrochemistry of the basin at the three different locations.

2. The Ankobra Basin

The Ankobra Basin (Fig. 1) is part of the southwestern river system, which makes up 22% of the total runoff in Ghana. It drains a total area of $8272\,km^2$ into the Atlantic Ocean in the south. Annual rainfall ranges between 1500 and 2150 mm. This much precipitation helps sustain a thick forest cover within the area. Rocks of the Birimian and Tarkwaian systems, which host a huge percentage of the country's mineral resource, underlie the Ankobra Basin. These include crystalline igneous and metamorphic rocks of Precambrian age making up rocks of the Birimian system, overlain unconformably by rocks of the Tarkwaian system. The Tarkwaian system consists of slightly metamorphosed, shallow water sedimentary strata, chiefly sandstones, quartzites, shale and conglomerate, believed to be the weathered derivatives of the Birimian rocks. The Birimian System is composed of great thicknesses of isoclinally folded, metamorphosed sediments intercalated with metamorphosed tuff and lava (Dapaah-Siakwan and Gyau-Boakye, 2000). The tuff and lava are predominant in

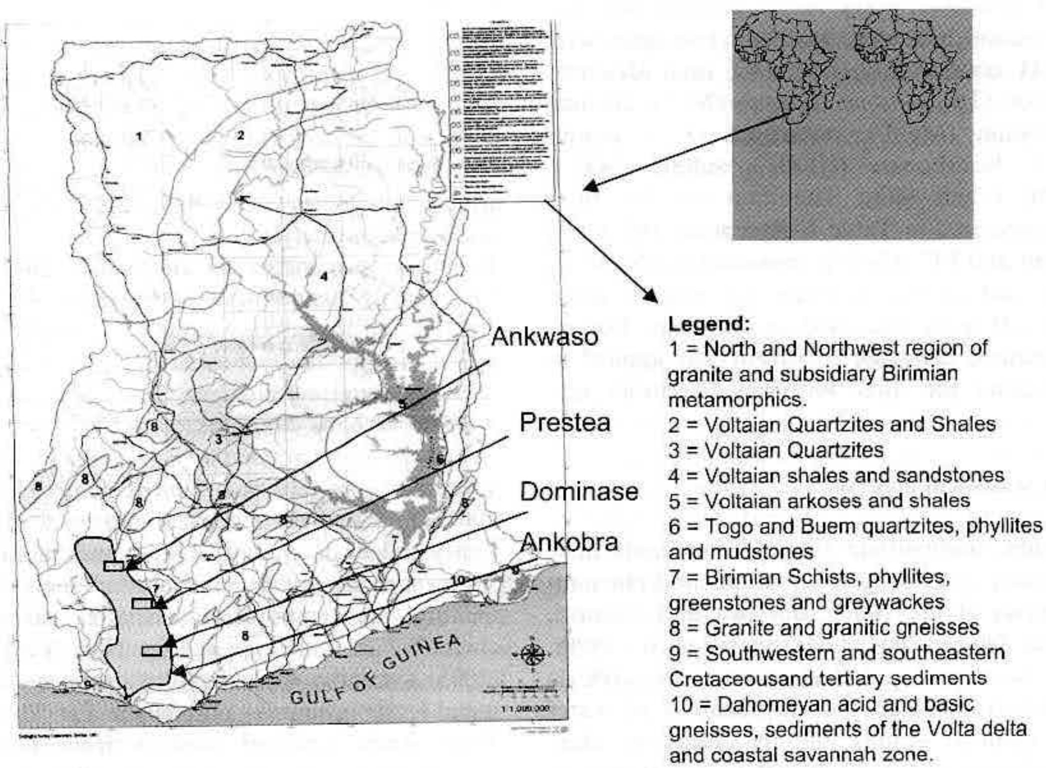


Fig. 1. A map of Ghana showing the three locations along the Ankobra Basin.

the upper part of the System whilst the sediments are mainly in the lower parts. The entire sequence is intruded by batholithic masses of granites and gneisses. In the wake of this intrusion, the argillaceous sediments of the sequence were metamorphosed to schist and phyllite, intercalated with greywacke.

Ankwaso, Dominase and Prestea, the three locations sampled for this research, are all underlain by the Birimian schists, phyllites, greenstones and greywackes (Fig. 1), which are highly manganiferous (Dapaah-Siakwan and Gyau-Boakye, 2000). These rock units are highly folded in the region. They contain feldspars, quartz, carbonates, sulfates and several heavy metals. The three regions were selected to identify any impacts of anthropogenic factors such as surface mining activities on the concentrations of the major cations and anions of the water in the Ankobra Basin at the three different locations.

3. Chemical data

The chemical analysis that forms the basis of this paper was compiled by workers of the Water Resources Research Institute, of the Center for Scientific and Industrial Research, Ghana. The purpose was to present data on all the surface water basins (see Fig. 1) in Ghana in a readily available format for use. Observation stations were located at three different places in Prestea, Dominase and Ankwaso, for the period ranging between 1989 and 1992. There are gaps in the data due to the drying up of some streams following prolonged drought, and the lack of reagents and logistics to analyze some other water quality indicators. For instance, heavy metal analysis was not done. For the period indicated, grab water samples were analyzed for pH, conductivity (EC), DO, total alkalinity (TA), total solids (TS), cations (sodium (Na^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), potassium (K^+), anions (chloride (Cl^-), bicarbonate (HCO_3^-), sulfate (SO_4^{2-}), phosphate (PO_4^{3-}) and SiO_2. The data for the three locations is summarized in Table 1. Except for pH which is a pure number and EC, which is measured in μS/cm, all the parameters are quoted in parts per million units. Alkalinity and pH were measured in the field. For all chemical parameters, the same treatment was applied to samples taken from the three locations to ensure uniformity.

Table 1
Summary statistics of the concentrations of the major chemical parameters used for this study

	N	Minimum	Maximum	Mean	Std. deviation
Ankwaso					
pH	30	6.4	7.6	7.2	0.3
EC	30	70.0	182.0	139.9	29.5
Na^+	28	6.1	48.0	14.4	8.3
K^+	30	2.70	25.9	6.4	4.6
Ca^{2+}	31	8.0	22.4	12.7	3.8
Mg^{2+}	30	0.0	18.5	5.0	3.5
Cl^-	30	9.0	59.0	17.0	10.1
HCO_3^-	30	26.8	956.0	122.4	240.4
PO_4^{3-}	30	0.0	0.96.0	0.1	0.2
TS	30	65.0	575.0	259.6	166.4
TA	33	20.0	80.0	51.1	15.5
SiO_2	30	2.0	65.0	18.8	13.5
SO_4^{2-}	30	0.0	45.0	5.4	9.9
DO	33	3.3	8.1	6.0	1.2
Dominase					
pH	30	6.3	7.6	7.0	3648.0
EC	33	27.0	116.0	73.3	20.3
Na^+	35	2.9	22.4	9.0	4.4
K^+	35	0.9	120.0	11.8	28.0
Ca^{2+}	33	3.0	109.0	15.5	25.7
Mg^{2+}	32	1.0	15.6	4.1	2.9
Cl^-	32	6.0	27.9	12.4	5.4
HCO_3^-	33	12.0	56.0	25.3	12.8
PO_4^{3-}	30	0.0	0.7	0.1	0.2
TS	30	100.0	400.0	221.5	103.4
TA	30	12.0	60.0	27.7	14.1
SiO_2	33	1.0	28.0	10.3	6.9
SO_4^{2-}	30	1.0	540.0	41.3	114.2
DO	30	5.1	8.0	6.6	0.8
Prestea					
pH	30	6.2	7.7	7.1	0.4
EC	30	33.0	348.0	85.6	52.4
Na^+	33	0.0	43.0	8.3	7.4
K^+	30	0.9	7.4	2.5	1.3
Ca^{2+}	30	3.0	21.0	7.2	3.4
Mg^{2+}	30	0.0	27.7	4.2	4.8
Cl^-	30	7.0	20.0	11.0	3.1
HCO_3^-	30	20.7	73.2	34.7	13.3
PO_4^{2-}	30	0.0	0.5	0.1	0.1
TS	30	42.0	1055.0	204.9	254.4
TA	35	14.0	60.0	29.3	11.2
SiO_2	30	0.0	43.0	14.5	10.8
SO_4^{2-}	33	0.0	15.0	3.6	4.2
DO	33	5.4	9.8	7.0	1.2

N = total number of data points.

4. Multivariate statistical analysis

In recent times, multivariate statistical methods have been applied widely to investigate environmental phenomena (e.g. Anazawa et al., 2003; Anazawa and Ohmori, 2005; Güler and Thyne, 2004; Laaksoharju et al., 1999). The combined use of principal component analysis (PCA) and cluster analysis enabled the classification of water samples into distinct groups on the basis of their hydrochemical characteristics. Multivariate statistical tools have been successfully used to study and classify different sediment types (Huisman and Kiden, 1998; Tebens et al., 2001), and hydrogeochemical processes (Cameron, 1996; Duffy and Brandes, 2001; Gupta and Subramanian, 1998). Momen et al. (1996) used cluster analysis and PCA to identify the temporal and spatial variation of water chemistry in Lake George in New York. Tariq et al. (2005) similarly applied multivariate techniques to trace metal levels in tannery effluents in Pershawar in Pakistan. Their study involved samples from tannery effluents, groundwater and soils, and with the aid of multivariate tools, they were able to correlate important chemical species of the three media and established significant relationships. Using factor analysis, Zeng and Rasmussen (2005) attributed the variations in quality of water from Lake Lanier in Georgia to anoxia associated with lake stratification. This paper dwells on the strength of multivariate techniques to characterize the hydrochemical variations along the Ankobra Basin. It employs the combined use of cluster and factor analysis to assess the spatial variation of surface water chemistry.

4.1. Cluster analysis

Cluster analysis groups a system of variables into clusters on the basis of similarities (or dissimilarities) such that each cluster represents a specific process in the system. In this study, the hierarchical cluster analysis (HCA) was applied to the raw data for each of the three different locations, using SYSTAT 11 (Systat Software Incorporated, 2004). HCA is a powerful tool for analyzing water chemistry data, and formulating geochemical models (Meng and Maynard, 2001). Several similarity/dissimilarity measures are available from the SYSTAT 11 package. A classification scheme using the Euclidean distance for similarity measures and the Ward's method for linkage produces the most distinctive classification where each member within a group is more similar to its fellow members than to any member outside of the group (Güler et al., 2002).

4.2. Factor analysis

Factor analysis is a multivariate analytical technique, which derives a subset of uncorrelated variables called factors that explain the variance observed in the original dataset (Anazawa and Ohmori, 2005; Brown, 1998). Factor analysis is used to uncover the latent structure of a set of variables. In technical terms, common factor analysis represents the common variance of variables, excluding unique variance, and is thus a correlation-focused approach seeking to reproduce the intercorrelation among the variables. On the other hand, components (from PCA) reflect both common and unique variance of the variables and may be seen as a variance-focused approach that reproduces both the total variable variance with all components as well as the correlations. PCA is far more commonly used than principal factor analysis (PFA). However, it is common to use "factors" interchangeably with "components" in multivariate analysis.

Factor analysis can be performed on any kind of scientific data to establish a pattern of variation among variables or reduce large data sets into factors for easy handling and interpretation. The total number of factors generated from a typical factor analysis indicates the total number of possible sources of variation in the data. Factors are ranked in order of merit. The first factor or component has the highest eigenvector sum and represents the most important source of variation in the data. The last factor is the least important process contributing to the chemical variation. Factor loadings on the factor loadings tables are interpreted as correlation coefficients between the variables and the factors.

In this research, PCA was applied to chemical data from the Ankobra Basin to extract the principal factors corresponding to the different sources of variation in the data. Here, PCA was selected for the reasons stated above. In order to maximize the variation among the variables under each factor, the factor axes were subsequently varimax rotated.

5. Results and discussions

Normal and random distribution are required for optimality in all multivariate statistical analysis. The Shapiro–Wilk (Shapiro and Wilk, 1965) test was performed on each dataset to determine normal distribution. Except for Na^+, K^+, HCO_3^- and PO_4^{3-} for Ankwaso, and Na^+, and K^+ for Dominase and Prestea, which were initially not normally distributed due to the presence of obvious outliers in the data, the rest of the parameters are normally distributed at $p<0.05$. When the outliers were replaced by mean values, data for the above named parameters assumed normal distribution and were well suited for optimal multivariate analysis.

5.1. Cluster analysis

Results from the *R*-mode HCA for each region were reported in the form of dendrograms. On the basis of the connecting distances between parameters and their positions on the dendrograms, distinctive clusters of the variables were defined for each of the three locations along the Basin. Though this procedure is subjective, the distinction between clusters in this analysis is quite clear from the dendrograms.

5.1.1. Ankwaso region

Three main groups are visible from the cluster analysis of data from the Ankwaso region. Fig. 2 presents the dendrogram from the HCA. Group 1 comprises PO_4^{3-}, K^+, pH, DO, Mg^{2+}, and SO_4^{2-}, and represents the dissolution of sulfide bearing rocks and the subsequent oxidation of the sulfide to sulfate. This is an oxidation process which also affects the pH of the medium. Group 2 is made up of SiO_2, Na^+, Ca^{2+} and Cl^- and represents the weathering of albitic and calcic feldspar and the contribution of precipitation to water chemistry. Group 3 indicates that EC in the water has a dominant contribution from the HCO_3^- in the water. It comprises TS, EC, HCO_3^- and TA. Group 3 represents the decay of organic matter, which enriches the water with HCO_3^- ions.

5.1.2. Dominase

The HCA for data from the Dominase area reveal two main groups of clusters of significance (Fig. 3). Group 1 is

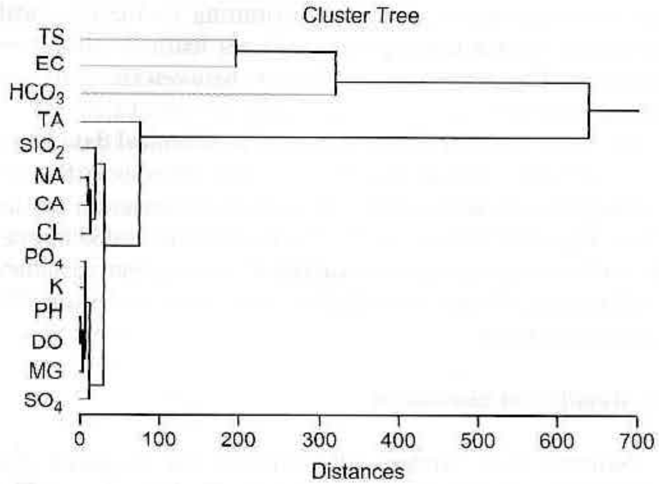


Fig. 2. The dendrogram showing the clustering of chemical parameters of Ankwaso water.

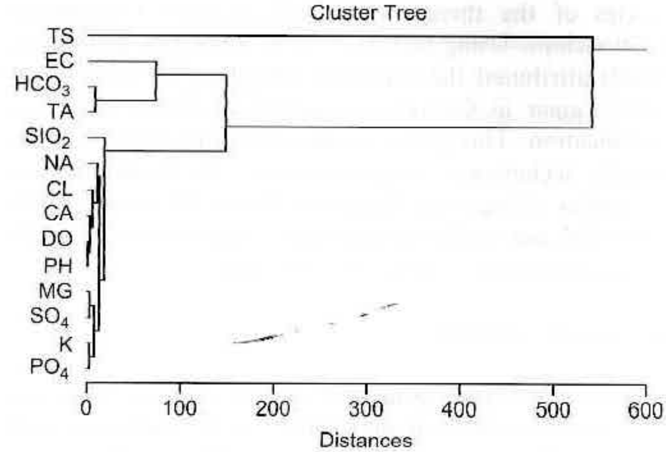


Fig. 4. A dendrogram developed from the HCA on the data from the Prestea region.

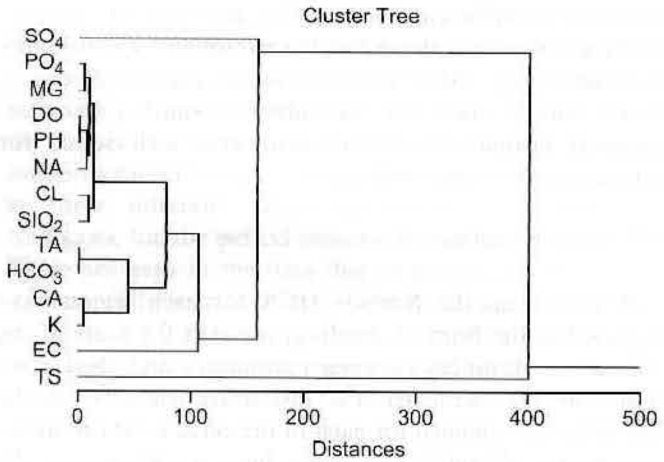


Fig. 3. The dendrogram developed from HCA on data from the Dominase area.

made up of PO_4^{3-}, Mg^{2+}, DO, pH, Na^+, Cl^-, and SiO_2 which are derived from silicate mineral weathering and precipitation. Group 2 consists of TA, HCO_3^-, Ca^{2+} and K, and represents the weathering of calcite in the organic sediments.

5.1.3. *Prestea*

Three main groups are visible from the results of the HCA. The result of the cluster analysis is illustrated in Fig. 4. Group 1 comprises PO_4^{3-}, K^+, SO_4^{2-}, and Mg^{2+} and reflects the contribution of organic and artificial fertilizers from agricultural activities in the area. pH, DO, Ca^{2+}, Cl^-, Na^+ and SiO_2 make up group 2 which represents the joint effects of precipitation and the weathering of calcic and albitic feldspars from the underlying geology. Group 3 is made up of TA, HCO_3^-, EC and TS and reflects the decay of organic matter from plant debris in the area.

5.2. *Factor analysis*

The Kaiser criterion (Kaiser, 1960) was applied to determine the total number of factors for each dataset in this analysis. Under this criterion, only factors with eigenvalues greater than or equal to 1 will be accepted as possible sources of variance in the data, with the highest priority ascribed to the factor that has the highest eigenvector sum. The rationale for choosing 1 is that a factor must have a variance at least as large as that of a single standardized original variable to be acceptable. When the seemingly complete factor model was developed using this criterion, the first four factors that account for at least 80% of the variance were selected for the next analysis. This was to ward off factors which by virtue of their low loadings do not constitute unique sources of variance in the hydrochemistry and could therefore be dispensed with.

5.2.1. *Ankwaso region*

Four principal components (PC) were extracted and rotated using the varimax normalization (Kaiser, 1960). An initial run using the Kaiser criterion (Kaiser, 1960) resulted in five principal components. However, it was observed that the fifth factor would not constitute a unique source of variance in the hydrochemistry since it had only one loading greater than 0.50. It was therefore dropped and four factors were chosen for varimax rotation. The results (Table 2) show that the four PC account for more than 86% percent of the variance, which is quite good and can be relied upon to identify the main sources of variation in the hydrochemistry. PC 1 has high loadings ($>|0.50|$) for pH, EC, K^+, PO_4^{3-}, SiO_2 and DO, and accounts for 31.6% of the total variance in the hydrochemistry in the area. Potassium and SiO_2 both show positive loadings under PC1 and are derived from the weathering of K-feldspars from the underlying geology, a process which is accompanied by a rise in pH (Eq. (6)). PC 2, which accounts for 26.3% of the total variance, contains high loadings for EC,

Table 2
Rotation PCA loading matrix (VARIMAX, Gamma = 1.0000) for the Ankwaso subregion

	Component 1	2	3	4
pH	**0.7**	−0.2	−0.4	**−0.6**
EC	**0.5**	**0.6**	**0.6**	−0.1
Na^+	0.0	**−0.9**	−0.0	−0.4
K^+	**0.9**	0.2	−0.0	0.2
Ca^{2+}	−0.4	−0.1	**0.9**	−0.2
Mg^{2+}	0.1	0.0	−0.2	**0.9**
Cl^-	−0.0	0.1	−0.2	**0.9**
HCO_3^-	−0.2	**0.7**	−0.1	−0.0
PO_4^{3-}	**−0.6**	**0.6**	−0.1	−0.2
TS	−0.2	−0.5	**0.6**	−0.2
TA	0.1	**0.9**	−0.3	0.2
SiO_2	**0.7**	−0.4	−0.3	−0.4
SO_4^{2-}	−0.1	−0.2	**0.9**	−0.1
DO	**0.9**	−0.0	−0.4	0.0

Total variance explained

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.1	31.6	31.6	4.1	31.6	31.6
2	3.4	26.3	57.9	3.4	26.3	57.9
3	1.9	14.5	72.4	1.9	14.5	72.4
4	1.8	13.9	86.3	1.8	13.9	86.3

The bold values indicate absolute component loadings higher than 0.5, which are considered significant contributors to the variance in the hydrochemistry.

Na^+, HCO_3^-, PO_4^{3-}, TS and TA, and represents the contribution of agricultural activities and organic matter. PC 3 has high loadings for EC, Ca^{2+}, TS and SO_4^{2-}, representing 14.5% of the total variance in the hydrochemistry at this location. PC 3 represents the weathering of gypsum from the underlying sedimentary material. PC 4, which accounts for 13.9% of the total variance in the hydrochemistry, shows high loadings for pH, Mg^{2+} and Cl^-. PC4 represents the contribution of domestic wastes, which contain $MgCl_2$.

5.2.2. *Dominase*

A trial analysis was performed using the Kaiser criterion (Kaiser, 1960), to determine the maximum number of PC. After the first PCA, PO_4^{3-} and SO_4^{2-} did not load significantly under any PC. They were consequently dropped since they did not contribute to the variation in the hydrochemistry. Three PC with significant loadings of at least three variables were selected for the varimax rotation. This was to ensure that the resulting model represented necessary sources of variation in the hydrochemistry at this location. This resulted in three PCs, accounting for 80% of the total variation in the hydrochemistry. Table 3 summarizes the results of the PCA. PC 1 contains high loadings for pH, EC, Na^+, K^+, Mg^{2+} and Cl^-, accounting for 43% of the total variance. It represents the contribution of precipitation and domestic wastewaters, which are rich in Na^+, Mg^{2+} and Cl^-. PC 2 represents 21% of the total variation in the hydrochemistry and has high loadings for K^+, Ca^{2+} and HCO_3^-, derived from the weathering of carbonate minerals from the underlying geology.

5.2.3. *Prestea*

Four principal components, which account for 82% of the total variance, were extracted for varimax rotation. Table 4 presents the rotated factor matrix. PC 1 contains high loadings for EC, Na^+, Cl^-, HCO_3^-, TA, and DO, and represents 30.8% of the total variance in the hydrochemistry. PC 1 represents the decay of organic matter, an oxidation process which requires the presence of DO. This explains why DO and HCO_3^- are positively correlated with PC 1. PC 2 contains high loadings for K^+, Mg^{2+}, Cl^-, TS and DO, accounting for 22.4% of the total variance in the hydrochemistry at this location along the basin. PC 2 represents the weathering of apatite and gypsum from the sediments. PC 3 has high loadings for Na^+, Ca^{2+}, PO_4^{3-} and SO_4^{2-}, representing 18.6% of the total variance. PC 4 records high loadings for pH and SiO_2. Under PC 1, the Na^+ registers a loading of 0.5, which is the same as that of the Cl^-, 0.5. PC 1 represents the contribution of precipitation to the hydrochemistry. The Mg^{2+} and Cl^- are inversely correlated with PC 2, whereas K^+ and TS are positively correlated. PC 2 represents wastewater from communities and agricultural activities in the

Table 3
Rotation PCA loading matrix for chemical parameters—Dominase

	Component		
	1	2	3
pH	**0.7**	0.2	**0.6**
EC	**0.8**	0.1	0.1
Na^+	**0.7**	0.1	−0.3
K^+	**0.6**	**0.5**	−0.2
Ca^{2+}	0.0	**1.0**	−0.0
Mg^{2+}	**0.9**	0.0	−0.1
Cl^-	**0.8**	0.1	−0.1
HCO_3^-	0.2	**0.9**	−0.2
TS	0.2	0.3	**−0.7**
SiO_2	−0.1	−0.1	**0.9**

Total variance explained

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.3	43.2	43.2	4.3	43.2	43.2
2	2.1	20.8	64.1	2.1	20.8	64.1
3	1.4	13.9	78.0	1.4	13.9	78.0

The bold values indicate absolute component loadings higher than 0.5, which are considered significant contributors to the variance in the hydrochemistry.

Table 4
Rotation PCA loading matrix for chemical parameters from Prestea

	Component			
	1	2	3	4
pH	0.2	0.2	−0.1	**0.9**
EC	**0.9**	0.1	−0.2	0.1
Na^+	**0.5**	−0.3	**−0.7**	−0.4
K^+	0.2	**0.8**	0.1	−0.0
Ca^{2+}	−0.0	0.2	**0.9**	−0.1
Mg^{2+}	0.4	**−0.7**	−0.1	−0.1
Cl^-	**0.5**	**−0.6**	0.4	0.1
HCO_3^-	**0.9**	0.1	−0.2	−0.0
PO_4^{3-}	−0.0	−0.2	**0.8**	−0.4
SiO_2	−0.2	−0.4	−0.2	**0.8**
SO_4^{2-}	−0.4	−0.2	**0.5**	−0.4
TS	0.1	**0.9**	−0.0	0.0
TA	**0.9**	0.1	−0.0	−0.0
DO	**0.6**	**0.7**	−0.2	0.1

Total variance explained

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.3	30.8	30.8	4.3	30.8	30.8
2	3.1	22.5	53.2	3.1	22.4	53.2
3	2.6	18.6	71.8	2.6	18.6	71.8
4	1.4	10.1	81.9	1.4	10.1	81.9

The bold values indicate absolute component loadings higher than 0.5, which are considered significant contributors to the variance in the hydrochemistry.

neighborhood. Household wastewaters contain high concentrations of Mg^{2+}, Cl^- and K^+ from salts and soaps. PC 3 represents the weathering of apatite and gypsum from the underlying sediments and rocks. PC 4, which has high positive loadings for pH and SiO_2 represents the breakdown of clay minerals, typically, kaolinite.

6. Conclusions

The above analysis demonstrates the use of multivariate statistical techniques to study the source/genesis of chemical parameters in surface water systems. Though the two multivariate techniques provide powerful means of studying the source genesis of the hydrochemistry of the Basin, the HCA gives broader and less definitive classifications than the PCA since the clustering and interpretations based on the dendrograms are largely subjective. The PCA, with the varimax rotation, is much more definitive and provides much more insight into the processes controlling the hydrochemistry at the different locations along the Basin. The application of both the *R*-mode HCA and PCA with varimax rotation, however, has been more effective than the use of either one of them alone. The hydrochemistry of the basin is controlled largely by the weathering of minerals (silicates, carbonates, gypsum and apatite) from the underlying meta-sediments of the Birimian and Tarkwaian Systems, and the decay of organic matter from the heavily forested regions. There is no strong evidence of high anthropogenic impacts on the hydrochemistry of the basin, though there are variations at the different locations studied. The hydrochemistry at Ankwaso is principally controlled by the weathering of silicate minerals, whereas those of Dominase and Prestea are, respectively, influenced by precipitation and domestic wastewaters, and the decay of organic matter.

References

Anazawa, K., Ohmori, H., 2005. The hydrochemistry of surface waters in Andesitic Volcanic area, Norikura volcano, central Japan. Chemosphere 59, 605–615.

Anazawa, K., Ohmori, H., Tomiyasu, T., Sakamoto, H., 2003. Hydrochemistry at a volcanic summit area, Norikura, central Japan. Geochimica et Cosmochimica Acta 67 (18S), 17.

Brown, C.E., 1998. Applied Multivariate Statistics in Geohydrology and Related Sciences. Springer, New York.

Cameron, E.M., 1996. The hydrochemistry of the Fraser River, British Columbia: seasonal variation in major and minor components. Journal of Hydrology 182, 209–215.

Dapaah-Siakwan, S., Gyau-Boakye, P., 2000. Hydrogeologic framework and Borehole yields in Ghana. Hydrogeology Journal 8, 405–416.

Duffy, C.J., Brandes, D., 2001. Dimension reduction and source identification for multispecies groundwater contamination. Journal of Contaminant Hydrology 48, 151–165.

Frape, S.K., Fritz, P., McNutt, R.H., 1984. Water–rock interaction and the chemistry of groundwaters from the Canadian Shield. Geochimica et Cosmochimica Acta 48, 1617–1627.

Garrels, R.M., McKenzie, F.T., 1967. Origin of the chemical compositions of some springs and lakes. In: Stumm, W. (Ed.), Equilibrium Concepts in Natural Waters. American Cancer Society, Washington, DC, pp. 222–242.

Güler, C., Thyne, G.D., 2004. Hydrologic and geologic factors controlling surface and Groundwater chemistry in Indian wells – Owens Valley area, southeastern California, USA. Journal of Hydrology 285, 177–198.

Güler, C., Thyne, G.D., McCray, J.E., Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydro Hydrogeology Journal 10, 455–474.

Gupta, L.P., Subramanian, V., 1998. Geochemical factors controlling the chemical nature of water and sediments in the Gomte River, India. Envronmental Geology 36, 102–108.

Hartman, J., Berna, Z., Stuben, D., Henze, N., 2005. A statistical procedure for the analysis of seismotechtonically induced hydrochemical signals: a case study from the Eastern Carpathians, Romania. Tectonophysics 405, 77–98.

Hem, J.D., 1989. Study and interpretation of the chemical characteristics of natural water. US Geological Survey Water-Supply Paper, p. 2254.

Huisman, D.J., Kiden, P., 1998. A geochemical record of late Cenozoic sedimentation history in southern Netherlands. Geologie en Mijnbouw 76, 277–292.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educational and Psychological Measurement 20, 141–151.

Laaksoharju, M., Gurban, I., Skarman, C., Skarman, E., 1999. Multivariate mixing and mass balance (M3) calculations, a new tool for decoding hydrogeochemical information. Applied Geochemistry 14, 861–871.

Meng, S.X., Maynard, J.B., 2001. Use of statistical analysis to formulate conceptual models of geochemical behavior: water chemical data from Butucatu aquifer in Sao Paulo State, Brazil. Journal of Hydrology 250, 78–97.

Momen, B., Eichler, L.W., Boylen, C.W., Zehr, J.P., 1996. Application of multivariate statistics in detecting temporal and spatial patterns of water chemistry in Lake George, New York. Ecological Modelling 91, 183–192.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality. Biometrika 52 (3), 591–599.

Systat Software Incorporated, 2004. SYSTAT 11 for Windows, Point Richmond, CA, USA.

Tariq, R.S., Shah, M.H., Shaheen, N., Khalique, K., Manzoor, S., Jaffar, M., 2005. Multivariate analysis of trace metal levels in tannery effluents in relation to soil and water: a case study from Peshawar, Pakistan. Journal of Environmental Mangement 79, 20–29.

Tebens, L., Veldkamp, A., Kroonenberg, S.B., 2001. The impact of climate change on the bulk and clay geochemistry of fluvial residual channel infillings: the Late Weichselian and Early Holocene River Meuse sediments (The Netherlands). Journal of Quaternary Science 13, 345–356.

Walling, D.E., 1980. Water in catchment ecosystems. In: Gower, A.M. (Ed.), Geochemistry, Groundwater and Pollution. A.A. Balkema, Rotterdam.

Zeng, X., Rasmussen, C.D., 2005. Multivariate statistical characterization of water quality in Lake Lanier, Georgia, USA. Journal of Environmental Quality 34, 1980–1991.

ELSEVIER

Journal of Environmental Management 86 (2008) 88–103

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Environmental strategy and performance in small firms: A resource-based perspective

J. Alberto Aragón-Correa[a,b,*], Nuria Hurtado-Torres[a], Sanjay Sharma[c], Víctor J. García-Morales[a]

[a] University of Granada, Spain
[b] Rotterdam-Erasmus University, The Netherlands
[c] Wilfrid Laurier University, Canada

Received 24 January 2006; received in revised form 22 November 2006; accepted 24 November 2006
Available online 18 January 2007

Abstract

In spite of the widespread recognition of the important roles that small and medium sized enterprises (SMEs) play in most economies, limited research has focused on their impacts on the natural environment and the strategies such enterprises adopt to reduce these impacts. It is usually assumed that SMEs lack the resources to implement proactive environmental strategies that go beyond minimum regulatory compliance. In this study of 108 SMEs in the automotive repair sector in Southern Spain, we found that SMEs undertake a range of environmental strategies from reactive regulatory compliance to proactive pollution prevention and environmental leadership. These strategies are associated with three organizational capabilities: shared vision, stakeholder management, and strategic proactivity, hypothesized based on the unique strategic characteristics of SMEs—shorter lines of communication and closer interaction within the SMEs, the presence of a founder's vision, flexibility in managing external relationships, and an entrepreneurial orientation. We also found that firms with the most proactive practices exhibited a significantly positive financial performance.

Keywords: Corporate environmental strategy; Natural environment; Organizational capabilities; Small and medium-sized enterprises (SMEs)

1. Introduction

Global environmental problems such as climate change that require urgent solutions have increased societal awareness about the impact of business operations on the natural environment. Scholars have expressed concern about the difficulty of achieving real environmental improvements if the current social paradigms and normative frameworks that guide business decision-making remain unaltered (e.g. Newton and Harte, 1997). Another group of scholars have highlighted that managers tend to frame the natural environment more as a strategic or subjective issue than as a normative or ethical one (Aragón-Correa et al., 2004; Banerjee, 2001; Cordano and Frieze, 2000; Sharma, 2000), and therefore advocate a strategic approach to promote change from within the organization (e.g. Clemens, 2001). While we recognize the complementarity and importance of both perspectives, in this study we adopt a strategic focus in an under-researched context: small and medium-sized enterprises (SMEs).

Research on organizations and the natural environment from the resource-based view of the firm (Barney, 1991; Rumelt, 1984; Wernerfelt, 1984) has shown that proactive corporate environmental strategies that go beyond regulatory compliance have a positive effect on corporate financial performance when mediated by valuable organizational capabilities (e.g. Christmann, 2000; Hart, 1995; Marcus and Geffen, 1998; Russo and Fouts, 1997; Sharma

*Corresponding author. School of Economics and Business, Campus Cartuja, s.n., Granada 18071, Spain.
Tel.: +34 958 24 37 05; fax: +34 958 24 62 22.
E-mail addresses: jaragon@ugr.es (J.A. Aragón-Correa), nhurtado@ugr.es (N. Hurtado-Torres), ssharma@wlu.ca (S. Sharma), victorj@ugr.es (V.J. García-Morales).

0301-4797/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.022

and Vredenburg, 1998; Wagner, 2005). Studies have also usually found that firm size has a significant effect on the degree of proactiveness, with larger organizations being more likely to adopt proactive environmental practices (e.g. Aragón-Correa, 1998; Buysse and Verbeke, 2003; Russo and Fouts, 1997; Sharma, 2000).

The findings that show that firm size affects the proactiveness of environmental strategies have perhaps led to an assumption that SMEs' lack of resources prevents them from implementing proactive strategies and that such initiatives may reduce their profitability (e.g. Russo and Fouts, 1997; Rutherfoord et al., 2000; Schaper, 2002). It may therefore lead to the assumption that the 'natural resource-based view of the firm' (Hart, 1995) may not be a useful perspective for SMEs. However, the above studies only include populations of large companies in their samples. Although some cases suggest that environmental policy in lesser-developed countries should develop to enable small-scale and artisan firms to form entities that are of a sufficiently large scale to allow adequate environmental protection (Tarras-Wahlberg, 2002): this option is not necessarily useful for competitive markets in developed countries. Therefore, it cannot be concluded from the available empirical evidence that SMEs are not likely to adopt proactive environmental strategies or even that they may not possess valuable organizational capabilities that enable them to generate such strategies.

SMEs have also not been the focus of environmental strategy research based on arguments about their presumed lack of interest in going beyond regulatory compliance (Greening and Gray, 1994; Russo and Fouts, 1997; Sharma and Vredenburg, 1998), the low degree of public interest in SMEs (Scott, 1990), and the difficulty of obtaining data from SMEs (Aragón-Correa, 1998; Rutherfoord et al., 2000). However, SMEs produce around 70% of the total global pollution (Smith and Kemp, 1998), 60% of the total carbon emissions (Marshall, 1998), and the sum total of SMEs' environmental impacts outweighs the combined environmental impact of large firms (Hillary, 2000). Therefore, it is imperative that such assumptions about the lower importance of studying environmental strategies of SMEs be based on empirical data rather than on conjecture. There is also broad agreement in the strategic management literature that the strategic distinctiveness of SMEs makes research focused in this context necessary (e.g. Barney et al., 2001; Dean et al., 1998; Way, 2002). Thus, strategic differences between big and small firms, the scope of SMEs' impacts on the global economy and on the natural environment, and the absence of previous analysis, all suggest the importance of giving detailed attention to the issue of the strategic behaviour of SMEs in their interface with the natural environment.

Accordingly, in this study we develop an organizational size-dependent perspective that links the unique strategic characteristics of SMEs—shorter lines of communication and closer interaction within the SMEs, the presence of a founder's vision, flexibility in managing external relationships, and an entrepreneurial orientation—to the organizational capabilities that they are likely to develop and deploy to generate proactive environmental strategies. Our organizational size-dependent natural resource-based view for SMEs addresses two questions: (1) Based on the unique strategic characteristics of SMEs, which are the organizational capabilities likely to be associated with their proactive environmental strategies? and (2) What effect do such proactive environmental strategies have on the financial and competitive performance of these SMEs? We contribute to both: the natural resource-based view by showing that SMEs can adopt proactive environmental strategies based on specific capabilities associated with their unique strategic characteristics, and the resource-based view by identifying the type of organizational capabilities that are likely to help SMEs develop their competitive strategies in general.

2. Strategies of SMEs

2.1. Strategic management literature and SMEs

Despite the widespread recognition of the significant role that SMEs play in most economies, the competitive strategies of SMEs have attracted limited research effort as compared to the focus on large firms. This may be partly because early studies using the PIMS database (e.g. Strategic Planning Institute, 1977) reinforced the idea that large firms possessed numerous advantages over smaller firms, and espoused the virtue of growth-oriented strategies. SMEs were often marginalized as a residual class of firms that failed to become big, that frequently used old fashioned managerial approaches, that occupied secondary labour markets and niches (Scranton, 1999), and/or that were less likely to use strategic analysis and planning practices (e.g. Shuman et al., 1985).

Later studies showed that certain advantages might also accrue to smaller firms (e.g. Woo, 1987) and accordingly research began to focus on demonstrating the effectiveness of different strategies and practices for SMEs' performance (e.g. Risseeuw and Masurel, 1994). It was argued that SMEs could gain advantage via focused (e.g. Brown, 1995; Porter, 1980) or flexible strategies for producing specialty products/services for niche markets (Lescure, 1999) as the only viable competitive options given their lack of resources (Lee et al., 1999).

2.2. Strategic characteristics of SMEs

Although the dominant perspective on SMEs has emphasized their lack of resources (e.g. D'Amboise and Muldowney, 1988; Eden et al., 1997), recent literature has proposed that SMEs possess certain characteristics such as internally generated funds, a simple capital structure, and an entrepreneurial orientation of the founders/managers

(Rangone, 1999; Yu, 2001), that can contribute to competitive advantage. It has also been argued that SMEs possess the flexibility to respond to changes in the general business environment, innovativeness to respond with agility to competitors' moves, and closer interaction amongst organizational employees. Each of these characteristics is discussed below.

Flexibility is probably the most widely cited among SME characteristics (e.g. Chen and Hambrick, 1995; Fiegenbaum and Karnani, 1991; Yu, 2001). Flexibility allows SMEs to pay greater attention to managing external relationships upon which they are dependant for critical resources to survive, more so than larger firms with larger internal resources. These include interfirm relationships (especially where SMEs are sub-contractors or parts/services suppliers to large companies), personal relationships that provide them with new market opportunities (Hendry et al., 1995; Conner and Prahalad, 1996), and relationships with government agencies for identifying and attracting subsidies and technical assistance for small enterprises (Darnall, 2002).

Entrepreneurial orientation and innovativeness have also been cited as important characteristics of small firms (e.g. Hitt et al., 1991; Woo, 1987). However, given that SMEs have limited human resources, the competitive advantage from these characteristics depends mainly on the managers-founder's vision and her ability to extend her views to the rest of the employees (Merz and Sauber, 1995; Miller et al., 1988) and motivate their opportunity-seeking and problem solving behaviour.

Finally, small firms may also be distinguished from large ones in terms of the closer interaction among departments, shorter lines of communication, better personal links, more unified culture and stronger identity. While these features permit easier communications and the generation of shared understanding (Kogut and Zander, 1996), they may also create problems due to the greater intensity of personal interactions if shared values and vision are not sufficiently strong (Lawrence and Lorsch, 1969).

In summary, while the literature on SME strategies has traditionally focused on efficacy problems because of small firms' potential lack of recourses, conflicting evidence for this argument has been found (Chen and Hambrick, 1995). We argue that the unique characteristics of SMEs' enable them to develop and deploy certain organizational capabilities which may be one of the main reasons for these paradoxical findings. Accordingly, the following section develops arguments, from a resource-based view of the firm, to link the three characteristics discussed above—shorter lines of communication and closer interaction within the SMEs, the presence of a founder's vision, the flexibility in managing external relationships, and entrepreneurial orientation—with the generation of proactive environmental strategies, and the relationship of such strategies with SMEs' performance.

3. A size-dependent perspective on SMEs' environmental strategy

3.1. Corporate environmental strategies

In spite of differences in nomenclature, typologies of corporate environmental strategy place firms' environmental strategies along a continuum ranging from reactive strategies that merely aim to meet legal requirements and implement pollution controls, to more proactive strategies that include voluntary eco-efficient practices for reducing energy and waste and pollution prevention practices that require innovations in processes, products and operations to reduce energy and material use at the source, to environmental leadership strategies where products, processes, and even business models are re-designed to minimize the ecological footprint along the entire product life cycle (Aragón-Correa, 1998; Buysse and Verbeke, 2003; Hart, 1995; Roome, 1992; Sharma, 2000; Sharma and Vredenburg, 1998).

A proactive environmental strategy requires changes in routines and operations and has been identified as an organizational competence (e.g. Christmann, 2000; Hart, 1995) because it requires the complex coordination of several human and technical skills and heterogeneous resources (Amit and Schoemaker, 1993) in order to reduce environmental impacts and simultaneously maintain or increase firm competitiveness. Extant research based on samples of only larger firms has shown that organizations with a larger size are more likely to undertake the most proactive environmental strategies (Aragón-Correa, 1998; Russo and Fouts, 1997; Sharma, 2000). Scholars have consequently argued that because proactive environmental strategies require accumulation of, and complex interaction among, skills and resources such as physical assets, technologies, and people (Ramus and Steger, 2000; Russo and Fouts, 1997; Sharma, 2000; Shrivastava, 1995), SMEs' limited resources might prevent them from adopting such practices (e.g. Greening and Gray, 1994; Russo and Fouts, 1997). Although a curvilinear relationship between firm size and environmental strategy may be possible based on potential synergies between complementary assets (Christmann, 2000), the lack of flexibility of large firms, and the limited resources of very small firms,[1] to our knowledge there is no evidence to support this relationship. The extant literature supports a linear relationship considering that big firms have greater access to the resources required for the implementation of the proactive environmental strategies.

This generalized impression of SMEs has persisted although systematic research on SMEs' environmental strategies has been absent from extant literature. Supporting this assumption, descriptive studies of SMEs have often highlighted their poor level of environmental commitment, describing them as mainly interested in controlling emis-

[1]We thank our reviewers for suggesting this idea as potential research for the future.

ons of pollution to comply with environmental regula- ons (e.g. Rutherfoord et al., 2000; Schaper, 2002; /illiamson and Lynch-Wood, 2001).

However, a few descriptive studies in several countries ɔntradict this assumption and have shown that SMEs ıay successfully implement environmental strategies con- stent with the advanced environmental practices of big rms (e.g. Bianchi and Noci, 1998; Carlson-Skalak, 2000; [illary, 2000) including innovations that prevent pollution t the source rather than pollution control at the end-of- ıe-pipe. Therefore, we argue that, contrary to the current ominant thinking about SME environmental strategies as eing mainly reactive, they are also likely to exhibit roactive strategies based on certain organizational cap- bilities related to the unique strategic characteristics of MEs.

2. Organizational capabilities for SMEs' proactive ıvironmental strategies

Drawing on the natural resource-based view of the firm ɛ.g. Hart, 1995) and the unique characteristics of SMEs iscussed above, we present arguments and develop ypotheses in this section for the capabilities that will be ssociated with SMEs' proactive environmental strategies ased on a fit with their unique characteristics discussed in ıe strategic management literature. These capabilities are l) shared vision, which is related to the owner-founder's ision and the close interaction and communication etween the owner–founder and the organizational mem- ers; (2) strategic proactivity which is related to SME's ıtrepreneurial orientation and innovativeness; and (3) :akeholder management, which is related to SMEs exibility in managing their inter-organizational and xternal relationships. These three capabilities have also een analysed in the natural resource-based view literature, 'hich is predominantly based on large firm samples (e.g. .ragón-Correa and Sharma, 2003; Christmann, 2000; Iarcus and Geffen, 1998; Russo and Fouts, 1997; Sharma nd Vredenburg, 1998). Our contribution in this study is to evelop a theoretical basis to identify a subset of the several ıpabilities discussed in the literature that are relevant and ıore likely for SMEs for developing proactive environ- ıental strategies.

.2.1. Shared vision

The organizational capability of shared vision exists 'hen an organization's members collectively have similar alues and beliefs about its objectives and mission (Oswald t al., 1994). The capability of shared vision does not mply mean that employees know their managers' bjectives; rather, shared vision entails a shared feeling ıat the firm's objectives are important and appropriate nd that all of its members may contribute to defining ıem. Goal clarity and shared responsibility for organiza- onal objectives are two basic characteristics of shared ision and positively affect organizational learning and employee creativity at the interface between business and the natural environment (Ramus and Steger, 2000).

Hart (1995) proposed that firms having a demonstrated capability of shared vision would be able to accumulate the skills necessary for developing a proactive environmental strategy earlier than firms without such a capability because these strategies depend "upon tacit skill development through employee involvement" (Hart, 1995, p. 999). Similarly, research on corporate environmental change argues for the importance of employee support in the SME context (Barret and Murphy, 1996; Ruiz-Quintanilla et al., 1996; Wehrmeyer and Parker, 1996) although the empirical evidence is mainly based on the large firms' context (e.g. Andersson and Bateman, 2000; Ramus and Steger, 2000).

On the positive side, as compared to large firms, SMEs may be burdened with less bureaucracy and fewer restrictions and provide more opportunities for direct communication and shared experiences amongst organizational members (O'Gorman and Doran, 1999). On the negative side, SME managers have been shown to have difficulty developing clear objectives and communicating with subordinates owing to lack of resources and unprofessional management (Smeltzer and Fann, 1989; Way, 2002), and lack interest in analysis offered by employees during the process of setting precise objectives and in reaching consensus among managers and employees (Merz and Sauber, 1995). Therefore, size alone does not make a shared vision capability intrinsic to an SME. Rather, we argue that only the SMEs that are able to exploit close interactions between the owner-founder's vision and employees into a capability of a shared vision for a sustainable business are more likely to implement a proactive environmental strategy. For instance, Raymond et al. (1998) showed that organizational support is positively associated with the implementation of business process reengineering in SMEs. This leads us to argue that the importance of a shared vision capability for a proactive environmental strategy will be at least as high in SMEs as in larger organizations. Therefore:

Hypothesis 1. *A capability of shared vision will be positively associated with the development of proactive environmental strategies by SMEs.*

3.2.2. Stakeholder management

Stakeholder pressures have often been cited as factors contributing to the adoption of proactive environmental practices by firms (e.g. Céspedes-Lorente et al., 2003; Cordano et al., 2004; Henriques and Sadorsky, 1999; Sharma and Henriques, 2005; Wheeler et al., 2003). This literature has mostly offered empirical evidence in the large firm context for the importance of this capability for generating proactive environmental strategies.

Sharma and Vredenburg (1998, p. 735) specifically defined this capability as "the ability to establish trust-based collaborative relationships with a wide variety of stakeholders, especially those with non-economic goals".

Henriques and Sadorsky (1999) showed that environmentally proactive firms usually view all their stakeholders as important and actively manage their environmental concerns and, for instance, Sharma and Henriques (2005) linked specific stakeholder pressures to specific sustainability practices by facilities in the Canadian forest products industry.

Stakeholders such as environmental non-governmental organizations (NGOs) usually target larger firms (and not necessarily the worst polluters), because these firms are often the most likely to respond in order to avoid damage to their reputations (Bianchi and Noci, 1998; Greve, 1989). Small firms may enjoy a degree of anonymity and therefore might be able to avoid undertaking some environmental practices, if they are so inclined (Dean et al., 2000). However, SMEs that are interested in proactive environmental practices need to pay careful attention to their stakeholders' interests. Although their small size provides them with flexibility in responding to changes in the general business environment, small firms are also seriously challenged by unfavourable and hostile environments (Merz and Sauber, 1995).

SMEs have limited internal resources to survive hostility by external forces and are less likely than large firms to have access to media or publicity. Therefore, SMEs need an organizational ability to be sensitive to the preferences of, and collaborative with, relevant external groups to garner external resources for adopting technologies, processes, and systems required for proactive environmental practices. Understanding and managing societal concerns via engagement in trust-based relationships can help expand SME's resources for undertaking proactive environmental practices through environmental coalitions and alliances (Rondinelli and London, 2003), governmental technological help and grants (Darnall, 2002), participation in green networks (Lehmann et al., 2005), and free consulting (Bianchi and Noci, 1998). Similarly, Flannery and May (2000, p. 646) found that, in a sample of 139 SMEs in the US metal-finishing industry, managers' decisions and intentions concerning the treatment of hazardous wastewater were influenced positively by "their assessment of support from important others", that is external stakeholders. McEvily and Marcus (2005) showed how SMEs engaged their suppliers to develop capabilities for pollution prevention. Therefore:

Hypothesis 2. *A capability of stakeholder management will be positively associated with the development of proactive environmental strategies by SMEs.*

3.2.3. Strategic proactivity

Strategic proactivity is a firm's ability to initiate changes in its strategic policies regarding its entrepreneurial, engineering, and administrative activities, rather than reacting to events (Aragón-Correa, 1998). The concept draws on the prospector orientation described in Miles and Snow's (1978) typology and involves taking initiative to shape the general business environment to one's own advantage (Chen and Hambrick, 1995). Proactiveness has been proposed as a key dimension (with innovation and risk taking) of an entrepreneurial orientation (Covin and Slevin, 1990; Lumpkin and Dess, 1996; Miller, 1987) and Stopford and Baden-Fuller (1994) showed that these dimensions help firms to gain new capabilities.

Based on data from 105 large firms in different Spanish economic sectors, Aragón-Correa (1998) showed that strategic proactivity encouraged adoption of proactive natural environmental strategies. Although previous empirical research regarding the influence of strategic proactivity on environmental approaches has been mainly based on large companies, some evidence shows the importance of strategic proactivity for small firms. For example, Dean et al. (1998) showed that environmental regulation of specific activities or sectors tends to generally discourage SMEs' presence but attracts specific kinds of proactive SMEs. This finding is especially interesting considering that small firms often show a greater propensity for action than their larger rivals (Chen and Hambrick, 1995). Moreover, entrepreneurial orientation and innovativeness are important characteristics of small firms (e.g. Hitt et al., 1991; Woo, 1987) seeking focused niche strategies in competitive markets dominated by large firms with deep pockets (Lescure, 1999; Lee et al., 1999). Therefore:

Hypothesis 3. *A capability of strategic proactivity will be positively associated with the development of proactive environmental strategies by SMEs.*

3.3. The influence of proactive environmental strategy on SME performance

Although there is mixed evidence regarding the influence of proactive environmental strategies on the financial performance of the firms (e.g. Bansal, 2005; Christmann, 2000; Margolis and Walsh, 2003), the majority of the studies have found a positive relationship. This positive relationship in the context of large firms (e.g. Klassen and McLaughlin, 1996; Russo and Fouts, 1997) has been explained as a result of mutual influence between proactive environmental strategies and valuable competitive capabilities (Christmann, 2000; Hart, 1995; Majumdar and Marcus, 2001; Russo and Fouts, 1997; Sharma and Vredenburg, 1998). The most proactive strategies focused on business redefinition and innovation in products and processes to prevent pollution and waste at the source enable an organization to align itself with changes in its general business environment (Aragón-Correa and Sharma, 2003), and have been shown to be associated with lower costs, improved reputation, and generation of new organizational capabilities (Christmann, 2000; Hart, 1995; Sharma and Vredenburg, 1998). The adoption of mid-range environmental strategies focused on eco-efficiency to reduce energy and waste have been found to reduce nvironmental impacts and simultaneously provide firms ith competitive advantage through reduction of costs and ddition of net value (Lehni, 2000; WBCSD, 2001).

Although researchers have extensively examined the ıfluence of environmental practices on large firms' erformance, studies of small firms' environmental approaches have often used varying dependent variables. For xample, Flannery and May (2000) investigated the ıfluences on shaping environmentally ethical decision ıtentions, and Dean and colleagues (1998, 2000) estimated ıe effect of environmental regulations on the formation of mall US manufacturing establishments. The descriptive tudies on SME environmental practices generally assume ıat legislative pressures are the only way to generate dvanced environmental practices among small firms ecause environmental activities do not have any positive nplications for SMEs' performance (e.g. Rutherfoord et l., 2000). However, Miles et al. (1999) emphasized that the elationship between financial performance and the adopion of environmental standards for SMEs needed further nalysis. They stated that "it is reasonable to expect that ompetent management and a more proactive stance on nvironmental issues will be rewarded in the small business ector as well" (Miles et al., 1999, p. 120).

Since organizational capabilities have been empirically ound to mediate the link between environmental strategy nd financial performance in the context of large firms (e.g. Christmann, 2000; Russo and Fouts, 1997; Sharma and Vredenburg, 1998), we argue that small firms will reap imilar benefits. However, in this study, we did not examine hese mediation effects and leave it for future research. We ropose that there will also be a direct relationship between roactive environmental strategy and financial performance based on the arguments presented above.

Iypothesis 4. *Proactive environmental strategies will be ositively associated with financial performance of SMEs.*

. Research method

.1. Sample and procedures

We used data from automotive garages, mainly truck nd car repair and maintenance facilities in Southern Spain, to examine our hypotheses. The automobile ndustry has a major impact on the natural environment e.g. Orsato et al., 2002) at each stage of its life cycle—from roduction, utilization and operation, repair and maintenance, to final disposal. The garages focus on repair and naintenance of automotive products which generates ignificant environmental impacts such as noise, high levels of CO_2 generation in tests, high levels of consumption of nergy and water, use of contaminants and toxic materials uch as chemicals and paint, and dangerous wastes. Consequently, this industry is receiving growing attention rom legislators.

We initially interviewed six managers and key members of an industry association of the automotive garage sector, three consultants, and two academics interested in the area. Data on SMEs' environmental practices, performance, and their organizational capabilities are not available from published sources. Therefore, based on the interviews and extant literature, we developed a questionnaire to measure our constructs.

The population for this study consisted of all the automotive garages (210 firms) located in the region of Southern Spain (including the provinces of Malaga, Granada, and Almeria). All these firms were designated SMEs according to official standards[2] (European Commission, 1999). By controlling for industry and geographical effects we limited our potential to draw broadly generalizable conclusions but this allowed us to control for extraneous confounding influences and focus on our variables of interest. Our secondary research suggests that the majority of the European automotive garages face a fairly homogeneous context in competitive and legal terms. Automotive garages also have many similarities with small firms in other sectors including intensive competition, complex regulation, and lack of resources—especially financial—as compared to larger companies.

Studies based on data gathered from SMEs usually tend to be inconclusive because of very low response rates (Merrit, 1998) or the difficulty that the respondents may have in interpreting answers to questions (Smith and Kemp, 1998). Therefore, we decided to administer the structured questionnaire personally via interviews with each informant. Additionally, because the vast majority of the CEOs were native Spanish speakers, the questionnaire was written in Spanish to avoid problems in interpretation. All the items are provided in the Appendix after translation into English. The personal survey administration took significantly greater time and effort compared to mail surveys but allowed us to ensure the appropriate identity of the respondents and their understanding of the questions. These factors contributed to greater data accuracy and reliability.

As is usual in strategic and environmental research (e.g. Banerjee, 2001; Cordano and Frieze, 2000; Christmann, 2000; Flannery and May, 2000; Sharma, 2000), data were collected from the general managers because they are the most knowledgeable about their organization as a whole and usually play a crucial role in designing environmental strategies, especially in the SME context. To reduce possible social desirability bias, we: (1) promised that our analyses would be aggregated and no organization would be identified individually; (2) randomly compared

[2]When the research was conducted, the European Commission's definition of SME was based on companies with fewer than 250 employees with either an annual turnover not exceeding €40 million or an annual balance sheet total not exceeding €27 million (adjustments are made in these values regularly), and those that were independent with less than 25% of the capital or voting rights owned by one enterprise or jointly by several enterprises.

self-evaluation of managers with the evaluation of environmental practices of their firms by competitors[3]; and (3) included questions about specific actions and strategies rather than about general ethical claims (Banerjee, 2001).

The general managers of 149 firms agreed to participate in this research and we were able to interview 126 of these managers. Fourteen interviews were not used for this study because of missing values, leaving a final sample of 108 interviews representing 51.42% of the contacted population. The sampled firms had an average size of six employees. We did not find significant differences between the descriptive characteristics of firms that were included in the study in terms of location, activities, and size versus the overall population.

Although a single informant in each firm is often used in strategic management research due to the difficulty of obtaining multiple informants in larger surveys, we recognize the potential for mono-method bias in our data. Therefore we increased the confidence in our data by: (1) undertaking a factor analysis which showed the absence of a single general factor to account for most of the covariance in our variables, indicating the absence of common method variance problems for our data (Podsakoff and Organ, 1986); (2) comparing self-evaluation of managers with competitors' perceptions of their firms which allowed us to compare managers' responses with external sources of well-informed evaluators (see footnote 2 for details); and (3) personally identifying qualified respondents and administering the questionnaires. Previous literature has shown that the views of a single but well-qualified informant may better capture a firm's approach as compared to the views of several respondents in the case of small organizations where relevant decisions are often highly centralized (Chandler and Hanks, 1993; Lyon et al., 2000).

4.2. Measures

4.2.1. Environmental strategy

Due to limited or non-availability of publicly available environmental performance data, a firm's environmental strategy has been usually measured in terms of self-perceptions of managers (e.g. Aragón-Correa, 1998; Christmann, 2000; Flannery and May, 2000; Sharma, 2000; Sharma and Vredenburg, 1998). In our case, this was the only feasible approach since there are no publicly available data on environmental practices of garages in Spain.

[3]A random sample of 46 managers from our final respondent firms was asked about the most visible environmental practices of one or several competitors (usually nearby garages), resulting in a sample of 70 externally evaluated firms. Averaging evaluations created external evaluations of specific firms when several of them were available. A correlation of 0.71 between rate of external evaluations developed by competitors and self-evaluations developed by managers for sampled practices increases our confidence in the data.

We used two groups of items to measure the proactiveness of the environmental strategy of each garage. The items were grouped based on theoretically well-differentiated categories ranging from the most innovative prevention practices to the simplest eco-efficient practices. Although both groups of items include proactive practices they are significantly different in terms of their complexity and objectives, ranging from major changes in product and process design to saving energy and reducing waste. These have been analysed differently in extant literature (e.g. Hart, 1995; Marcus and Anderson, 2006).

The first group of practices was measured using 14 items drawn from Aragón-Correa's (1998) measure for environmental strategies designed to cover the whole range of environmental practices that a firm might adopt including product and process innovations for pollution prevention. Table A1 in the Appendix shows the results of standardized varimax rotation of these items resulting in two significant factors (eigenvalues > 1). Responses were on a five-point scale ranging from 1 for "we have not addressed this issue at all" to 5 for "we are the leaders on this practice in our sector". Unless otherwise noted, all subsequent scales were created by using the same five-point response format.

The second group of practices included nine items to evaluate the importance that managers gave to their firm's implementation of various eco-efficient practices. These environmental practices have been suggested as the first steps toward proactive environmental practices, and/or as specifically applicable to SMEs (e.g. Schmidheiny, 1992; Smith and Kemp, 1998). Exploratory principal components analysis with varimax rotation showed that these items formed two factors with eigenvalues > 1 (Please refer to Table A2 in the Appendix).

4.2.2. Stakeholder management

Following previous environmental research (Buysse and Verbeke, 2003; Cordano and Frieze, 2000; Flannery and May, 2000), we used Ajzen and Fishbein's (1980) technique as a guide to measure the capability of stakeholder management. First, we listed nine categories of stakeholders (such as local communities, shareholders, the media, environmentalists, and customers), and asked the general managers to rate (from 1 to 5) the level of attention they paid to each category in managing their enterprises. Second, we asked the informants to evaluate (from 1 to 5) their perception of the importance of each stakeholder for the environmental performance of the firm. Similar to Buysse and Verbeke's (2003) methodology, we weighted the managers' attention to each stakeholder by the perceived attention of each stakeholder to environmental impacts of the firm. The scores ranged from 1 to 25, with a higher score indicating a higher capability of stakeholder management. Exploratory principal components analysis with varimax rotation of those nine items showed that they formed three factors with eigenvalues > 1 (Please refer to Table A3 in the Appendix).

4.2.3. *Shared vision*

We used three items (shown in the Appendix) based on previous literature on shared ideas about organizational objectives and the ways in which employees influence those objectives (e.g. Jehn, 1995; Oswald et al., 1994). Each item was measured on a five-point Likert response scale (1="strongly disagree", 5="strongly agree"). Exploratory principal components analysis with varimax rotation of those three items showed that they formed only one factor with eigenvalue > 1 (Cronbach's alpha = 0.791). A high average score was indicative of a high degree of shared vision in a garage.

4.2.4. *Strategic proactivity*

We used three bipolar items from Aragón-Correa's (1998) validated scale measuring proactivity in a firm's generic business strategy (refer to the Appendix). The scale draws on Miles and Snow's (1978) well-known typology. We asked each respondent to position his or her firm on a scale of 1–5 that was constructed so that high values matched a "prospector" strategy and low values, a "defender" strategy. We inverted this pattern for the second question to avoid skewing the answers. Exploratory principal components analysis with varimax rotation of those three items showed that they formed only one factor with eigenvalue > 1 (alpha = 0.770). A high score was indicative of high development of strategic proactivity in a garage.

4.2.5. *Firm performance*

Strategy and environment scholars have used subjective perceptions of managers (e.g. Judge and Douglas, 1998; Sharma and Vredenburg, 1998) and objective data (e.g. Russo and Fouts, 1997) to measure firm performance. We included questions tapping both types of assessment in our interviews, but the garage managers were more open to offering their perceptions rather than to offering precise quantitative data (only 61 offered quantitative data). Therefore, we tested the model using a perceptual measure of financial performance in which each respondent rated his or her organization's performance relative to that of other firms in the garage industry using two items (see Appendix). These two items were drawn from Judge and Douglas (1998). These authors provide an extensive discussion and details about the potential of this process of measurement and similar processes in previous strategic literature (e.g. Miller and Friesen, 1984; Powell, 1995). The final measure was an average of the two items (alpha = 0.69), and a high score was indicative of a high level of performance in the garage as compared to the garage industry. Where possible, we calculated the correlation between the objective and subjective data. These were high and statistically significant (0.62, $p < 0.01$).

4.2.6. *Control variables*

All our sampled firms were SMEs according to institutional definitions. However, the breadth of this category suggested the need to use organization size to control for potential differences. The size indicators initially used were annual turnover and number of employees. Both indicators were highly and significantly correlated in this sample and we chose to use the number of employees. The logarithm of an organization's number of employees was used to measure size. We also added a dummy variable to control potential external influences based on whether or not the garage was associated with an automotive dealer or was independent.

5. Analysis and results

5.1. *Analysis*

In a preliminary analysis, the two variables measuring proactive environmental practices were first subjected to a cluster analysis using a non-hierarchical optimising procedure to determine whether different groups of similar firms could be grouped on the basis of their environmental strategies (Hair et al., 1998).

Buysse and Verbeke's (2003) classification of environmental strategies suggested a three-cluster solution which showed in our data a relatively small group of firms ($n = 26$) characterized by high levels of innovative and preventive environmental practices and of eco-efficient practices. A second group ($n = 45$) showed a high level of eco-efficient practices and an intermediate level of innovative and preventive practices. Finally, a group of firms ($n = 37$) exhibited the lowest levels of both eco-efficient and innovative practices. We identified these groups as exhibiting "leadership," "pollution prevention," and "reactive" environmental strategies, respectively, based on Buysse and Verbeke's (2003) terminology. The first group exhibits the greatest proactiveness, the second group exhibits some proactive practices focusing on simpler eco-efficiency solutions, and the third group does not exhibit any proactive practices. A four-cluster solution would divide the last group into two but not provide additional information.

Although the cluster analysis was developed using standardized values as recommended to avoid any influences of scales, Table 1 shows average scores using raw values for the different variables in each emerging group to avoid problems in interpreting standardized values. The "leadership" group have the highest values for all the analysed variables, the "pollution prevention" group shows intermediate values, and the 'reactor' group has the lowest values for the organizational capabilities and environmental performance. Test statistics from one-way analyses of variance (ANOVA *F*s) were all significant except for shared vision, showing the robustness of the solution (Hair et al., 1998). Cluster analyses were repeated on randomly selected sub-samples, indicating that the results can be considered independent of sample characteristics (Henriques and Sadorsky, 1999).

Table 1
Clustering of sampled firms based on their environmental strategies

	Group 1: "Leader"	Group 2: "Pollution prevention"	Group 3: "Reactive"	ANOVA F
N	26	45	37	
Innovative preventive practices	4.04	2.76	1.76	196.021**
Eco-efficient practices	4.66	4.45	3.45	58.780**
Shared vision	4.73	4.47	4.38	1.728
Stakeholder management	13.49	12.23	10.51	6.881**
Strategic proactivity	3.11	1.97	1.91	7.312**
Performance	3.34	2.89	2.56	9.302**

$^{**}p<0.01$.
$^{***}p<0.001$.

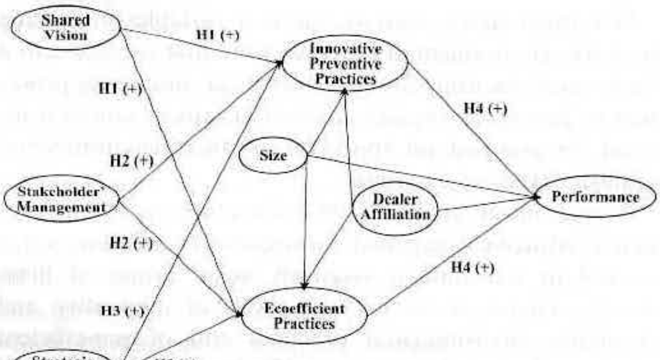


Fig. 1. Hypothesized model.

The results show that the patterns of environmental strategies for the sampled SMEs are congruent with theoretical typologies ranging from reactive ("reactor") to the most proactive strategies ("pollution prevention" and "leader"). Our classification is also consistent with previous classifications for large firms (e.g. Buysse and Verbeke, 2003). The values in Table 1 for the rest of the variables suggest support for Hypotheses 1–4. Next, we used a structural equation model for a definitive test of our hypotheses (Fig. 1).

5.2. Results

We examined our hypotheses using structural equation modelling (LISREL 8.3) with the covariance matrix as input and the assumption of no error. This provided a conservative test of the model (Hair et al., 1998) and was estimated using the weighted least squares (WLS)[4] method. We used a recursive non-saturated model, taking shared vision (ξ_1), stakeholder management (ξ_2), strategic proactivity (ξ_3), size (ξ_4), and dealer affiliation (ξ_5) as the exogenous latent variables. Size and dealer affiliation wer control variables, innovative preventive practices (η_1) and eco-efficient practices (η_2) were the first-grade endogenou latent variables, and performance (η_3) was the second grade endogenous latent variable. Such analysis allows fo modelling based on both latent variables and manifes variables, a property well suited for the hypothesized model where most of the represented constructs are abstractions of unobservable phenomena.

Fig. 1 shows our proposed model and the hypotheses Table 2 presents the means, standard deviations, reliability coefficients, and correlations among the variables. Prior to conducting our analyses, we followed the procedure outlined by Bono and Judge (2003) for examining the measurement properties of our three multidimensiona variables to prevent interpretational problems inherent in simultaneous estimation of measurement and structura models.

As suggested in extant literature (e.g. Cordano and Frieze, 2000; Flannery and May, 2000), the dimensions of stakeholder management emerging from the exploratory factorial analysis were highly interrelated in our data. A second-order confirmatory factor analysis (loading items on the three dimensions on a single transformationa factor) demonstrated a reasonable fit for the data ($\chi^2 = 18.95$; $df = 20$; RMSEA = 0.001; CFI = 0.99, IFI = 0.99) with standardized loadings ranging from 0.97 to 0.40 (Fig. 2). Whereas this model is not a perfect fit for the data, because of cross-loading among some of the items, alternative models were unambiguously rejected. For instance, our model is a better fit than a single-factor model in which items are loaded directly on the final factor ($\chi^2 = 137.56$; $df = 27$; RMSEA = 0.195; CFI = 0.74; IFI = 0.74). Hence, the three factors were considered indicators of a single measure, which we labelled "stakeholder management." Subsequent analyses were conducted using a single factor for stakeholder management. The final measure was a weighted average of the three factors using the standardized loadings obtained from the second-order factor analysis, and a high score was indicative of a high capability for stakeholder management.

[4]The sample sized was 108. Sample size, as in any other statistical method, provides a basis for the estimation of sampling error……. While there is no correct sample size, recommendations are for a size ranging between 100 to 200 (Hair et al., 1998, p. 637).

able 2
leans, standard deviations, and correlations

	Mean	s.d.	1	2	3	4	5	6	7	8
Shared vision	4.49	0.75	**0.791**							
Stakeholder management	11.93	3.36	0.076	**0.693**						
Strategic proactivity	2.24	1.42	−0.065	0.257**	**0.770**					
Innovative preventive practices	2.73	0.96	0.167	0.324***	0.293**	**0.778**				
Eco-efficient practices	4.21	0.94	0.180	0.067	0.110	0.417***	**0.705**			
Performance	2.90	0.76	0.266**	0.087	−0.003	0.450***	0.430***	**0.691**		
Size	1.17	1.05	−0.079	0.035	0.427***	0.190*	−0.106	−0.002	—	
Dealer affiliation	0.51	0.50	−0.031	−0.070	0.216*	0.202*	0.005	0.099	0.280**	—

cale reliabilities (Cronbach's alpha) are on the diagonal in boldface.
*$p<0.05$.
**$p<0.01$.
***$p<0.001$.

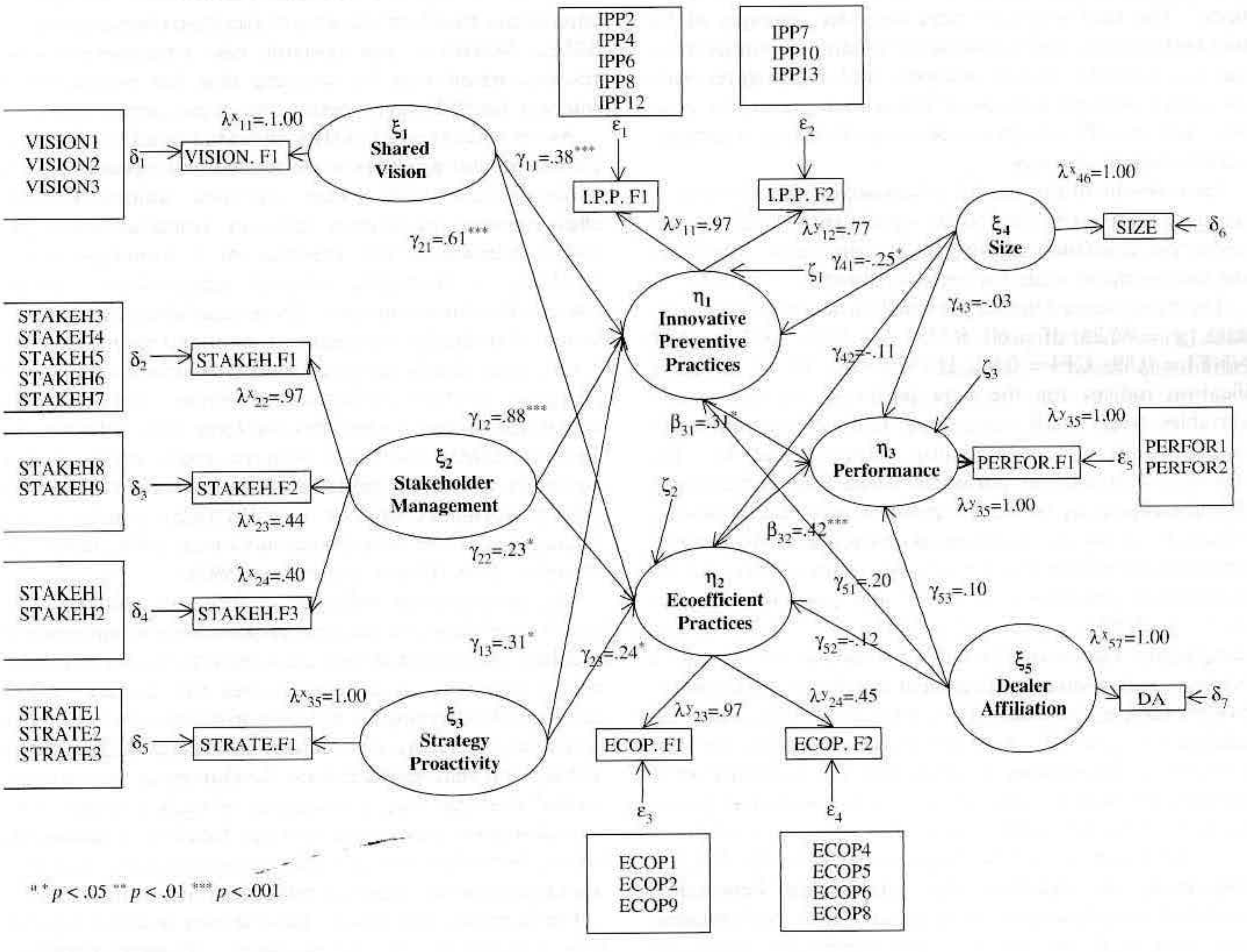


Fig. 2. Results of processes for estimation of measurements and model (*$p<0.05$, **$p<0.01$, ***$p<0.001$).

Similar analysis was conducted for the measures of the nnovative-preventive practices and the eco-efficient practices. Examination of residuals and modification indices of he second-order confirmatory factor analyses (loading the tems on the two dimensions on a single transformational factor for each variable) suggested to avoid the final inclusion of specific items (IPP1, IPP3, IPP5, IPP9, IPP11 and IPP14, and ECOP3, and ECOP7) to reinforce the fit of each measure. The final model (Fig. 2) demonstrated a reasonable fit for the data for innovative-preventive practices ($\chi^2 = 27.95$; df = 19; RMSEA = 0.066; CFI = 0.95; IFI = 0.95) with standardized loadings ranging from 0.97 to 0.77, and the eco-efficient practices ($\chi^2 = 33.40$; df = 14; RMSEA = 0.113; CFI = 0.87; IFI = 0.87) with standardized loadings ranging from 0.97 to 0.45. Whereas these models were not a perfect fit for the data, because of cross-loading among some of the items, alternative models were unambiguously rejected. For instance, these models were a better fit than a single-factor model, in which items are loaded directly on the final factor of the innovative-preventive practices ($\chi^2 = 39.10$; df = 20; RMSEA = 0.094; CFI = 0.89; IFI = 0.90) or eco-efficient practices ($\chi^2 = 33.40$; df = 14; RMSEA = 0.113; CFI = 0.87; IFI = 0.87). Hence, the two factors were considered indicators of a single measure, which we labelled "innovative-preventive practices" and "eco-efficient practices." The final measures were weighted averages of the two factors using the standardized loadings obtained from the second-order factor analysis, and high scores were indicative of high degrees of innovative-preventive practices and eco-efficient practices, respectively, in a garage's environmental strategy.

Final results of this model, which are displayed in Fig. 2, revealed that each final item significantly loads on its respective construct showing high convergent validity of the measurement scale for each construct.

The hypothesized model provided an acceptable fit to the data ($\chi^2 = 86.25$; df = 60; RMSEA = 0.08; AGFI = 0.89; NNFI = 0.89; CFI = 0.92; IFI = 0.91). All of the modification indices for the beta pathways between major variables were small, suggesting that adding additional paths would not significantly improve the fit. The statistically significant parameters (see Fig. 2) predicting that the capability of shared vision would have a positive influence on the development of proactive environmental strategies by influencing the innovative-preventive and the eco-efficient practices ($\gamma_{11} = 0.38$ and $\gamma_{21} = 0.61$, respectively; $p<0.001$) support Hypothesis 1. A positive and statistically significant parameter estimate for the path between stakeholder management capability and innovative-preventive practices ($\gamma_{12} = 0.88$, $p<0.001$), and eco-efficiency ($\gamma_{22} = 0.23$, $p<0.05$) provides support for Hypothesis 2. Hypothesis 3, regarding the influence of a strategic proactivity capability on environmental proactiveness, was also supported by the parameter estimates ($\gamma_{13} = 0.31$ and $\gamma_{23} = 0.24$, respectively; $p<0.05$). Finally, Hypothesis 4, regarding the relationship between a proactive environmental strategy and firm performance, was supported by positive and significant parameter estimates for the path between innovative-preventive strategies and performance ($\beta_{31} = 0.31$, $p<0.01$) and for the path between eco-efficiency and performance ($\beta_{32} = 0.42$, $p<0.01$).

The influence of control variables was tested for each of the three dependent variables in our model. The control variable for size showed a positive and statistically significant ($p<0.01$) association with innovative-preventive practices. The influence of size on eco-efficient practices and performance was not significant for the sampled firms. The analysis of the influence of dealer affiliation was not significant for our sample. We also fit several nested models to the data, each one incorporating different assumptions about parameters (Bollen and Long, 1993). These alternative models indicated that our proposed model was parsimonious and a good fit. Detailed information about alternative models is available upon request.

6. Discussion, conclusions, and future research

Our contribution is to show, contrary to conventional wisdom in the extant literature, that even SMEs can adopt proactive environmental practices and that these practices can lead to superior financial performance via specific capabilities based on the unique strategic characteristics of SMEs. Moreover, our research also contributes to the resource-based view by showing that this perspective is relevant for SMEs' competitive strategies generally.

As we had expected, SMEs' potential to adopt proactive environmental practices is associated with specific organizational capabilities based on their unique strategic characteristics of shorter lines of communication and closer interaction, the presence of a founder's vision, flexibility in managing external relationships, and an entrepreneurial orientation. These capabilities were shared vision, stakeholder management, and strategic proactivity.

Our data clearly clustered SMEs' environmental strategies into reactive, pollution prevention, and leadership categories, similar to findings for large firms (Buysse and Verbeke, 2003) and firms with proactive environmental strategies showed a positive and significant relationship with performance for our sampled firms, similar to the findings for larger firms (Aragón-Correa, 1998; Judge and Douglas, 1998; Russo and Fouts, 1997).

The simultaneous influence of firm size and organizational capabilities on innovative-preventive environmental practices suggests that size, a common proxy for organizational resources, is a relevant but not a deterministic condition for developing the most proactive environmental strategies. Therefore, our findings contradict the traditional assumption that SMEs cannot develop proactive environmental strategies owing to scarcity of slack resources. Our results support the natural resource-based view perspective (Hart, 1995) that indicates that organizational capabilities are critical for strategies of both large firms and SMEs.

Furthermore, our results have shown that the capabilities promoting the development of environmentally proactive approaches in the sampled SMEs coincide with those found for large companies in previous research (e.g. Marcus and Geffen, 1998; Ramus and Steger, 2000; Sharma and Vredenburg, 1998). Although researchers concerned with organizational size have stated that what applies to large firms may not apply to small ones, recent studies (e.g. Flannery and May, 2000) and our own results show similar relationships for large and small firms.

Traditional arguments in the small firm context may have to be restated from the perspective of the resource-based view, that is, large and small firms both require organizational capabilities for competitive strategies. However, large and small firms may follow different paths and generate different sets of capabilities based on different sets of characteristics. Therefore, although some features of these three capabilities for SMEs may be similar to those in large organizations, the foundations for those capabilities may be significantly different for big and small-size organizations. While larger firms may deploy resources such as sophisticated systems for knowledge management, public relations campaigns, or significant investments in R&D to develop these three capabilities, small organizations may build these capabilities based on their unique strategic characteristics discussed above.

We also showed a positive and significant relationship between innovative-preventive environmental practices and eco-efficient practices and firm performance for the sampled SMEs. The positive relationship between the most proactive environmental practices and firm performance is also consistent with those obtained for larger firms (e.g. Judge and Douglas, 1998; Russo and Fouts, 1997), and it shows that a proactiveness in environmental strategy may be an appropriate alternative for both small and large firms.

It is especially interesting that our results show, in a SME context, a positive and significant relationship between firm performance and eco-efficient practices analysed as a systematic pattern of simple but multiple, consistent and co-ordinated practices that simultaneously reduce environmental impacts and organizational costs. This is relevant for practitioners and policy makers because eco-efficient practices have been often suggested as first steps for adoption by small firms and others starting on the path of environmental change. Viewed from a resource-based perspective, this result may help to differentiate between an isolated practice (e.g. switching off lights when not necessary) and a systematic pattern of practices demanding certain organizational capabilities (rather than resources or critical size). Partial analysis of our data indicates that certain eco-efficient practices might indirectly prevent the generation of organizational capabilities that are positively related to firm performance if implemented in an isolated way by deterring innovation.

The main limitation of our study arises from our modelling of environmental strategy as exclusively a function of internal capabilities. External conditions are also relevant to the development of environmental initiatives (Aragón-Correa and Sharma, 2003; Clemens, 2001; Newton and Harte, 1997). Future research needs to complement the internal influences with the role of social and normative paradigms in the SMEs' environmental behaviour. We attempt in our study to minimize external influences via homogeneity of the context of the sampled organizations. Finally, we caution that our results do not allow definitive statement about the direction of the causality in the analysed relationships and may have limited generalizability due to the business and geographical peculiarities of our sample.

Overall, our study confirms that SMEs have an important role to play in reducing the negative impacts of business on the natural environment. Considering the absolute impact of the SME sector on global economies, more research attention should be paid to the environmental practices of SMEs. Although previous research has shown that larger firms are often environmentally more proactive than smaller ones, paradoxically many of the capabilities needed to develop proactive environmental approaches may be fostered by certain strategic characteristics of small firms. Future longitudinal work could confirm the direction of causality between capabilities, environmental performance, and financial performance. Research comparing large and small firms can show that SMEs may be at a resource disadvantage but not a capability disadvantage when it comes to environmental advances and this would have important implications for policy makers and practitioners.

Acknowledgments

We would like to thank all the managers and the two anonymous reviewers who contributed their time and ideas to this study. Project SEC 2003-07755 (European Union and Spanish Ministry of Education) and Foundation BBVA partially supported this research.

ppendix A

The results of standardized varimax rotation for innovative preventive practices are shown in Table A1. Exploratory incipal components analysis results are shown in Table A2 and A3.

❖ **Performance**

Please, rate your firm overall performance on each of the following objectives relative to others in their industry (1, much orse to 5, much better) ($\alpha = 0.691$):

1. Return on investment.
2. Earnings growth.

Please, could you provide the approximate return on investment for your firm for the last year.

❖ **Shared vision** ($\alpha = 0.76$):

1. The objectives of this organization are very well-known to everybody working here.
2. Everybody working in this garage influences the way to work and the objectives of the firm.

3. Everybody in this organization freely contributes his/her points of view about how to run it smoothly.

❖ **Strategic proactivity** ($\alpha = 0.77$):

For each item, respondents positioned their firms on a 1–5 scale anchored by the responses given here.

1. The field within which the firm currently conducts our business is:

Narrow (related areas with prospect of change)	1 2 3 4 5	Broad (diversified and continuing to develop)

2. The main focus of concern in relation to the garage's technological process is:

Having cost-efficient technologies	1 2 3 4 5	Having flexible and innovative technologies

3. Planning in this garage is:

Tremendously rigorous and predetermined	1 2 3 4 5	Tremendously open, impossible to complete before acting

Table A1
Factor loadings of exploratory principal components analysis for innovative preventive practices[a]

Item	Factor 1: IPPF1	Factor 2: IPPF2
1. Sponsorship of natural environmental events (IPP1)		0.419
2. Use of natural environmental arguments in marketing (IPP2)	0.526	
3. Natural environmental aspects in administrative work (IPP3)		0.476
4. Periodic natural environmental audits (IPP4)	0.795	
5. Systematic programme for water recycling (IPP5)	—	—
6. Program of residue recycling (IPP6)	0.435	
7. Purchasing criteria including ecological requirements (IPP7)		0.710
8. Natural environmental seminars for executives (IPP8)	0.591	
9. Environmental training for employees (IPP9)	0.805	
10. Use of ISO certifications on quality and/or environmental aspects (IPP10)		0.637
11. Insurance planning to cover potential environmental risks (IPP11)		0.524
12. Procedures' manual including precise instructions on environmental operations in the garage (IPP12)	0.546	
13. Filters and controls on emissions and discharges (IPP13)		0.628
14. Use of "Life cycle analysis" (IPP14)		0.665
Eigenvalues	4.330	1.322
Cronbach alpha	0.754	0.733
Percentage of variance explained	20.21	20.16

[a]"Varimax" rotation was performed.

Table A2
Factor loadings of exploratory principal components analysis results of eco-efficient practices[a]

Items	Factor 1: ECOF1	Factor 2: ECOF2
1. We always switch off those lights and machines which are not necessary (ECOP1)	0.862	
2. All the water taps are always perfectly closed when they are not in use (ECOP2)	0.723	
3. When possible, we used recycled water to save water (ECOP3)	—	
4. We systematically separate dangerous wastes of the rest (ECOP4)		0.650
5. We systematically separate all the different kind of wastes (ECOP5)		0.658
6. We are activate to participate in buying and selling in waste markets (ECOP6)		0.538
7. We always avoid wasting the chemical products that we use in the garage (ECOP7)	0.629	
8. We store boxes and papers to use again or recycle them (ECOP8)		0.790
9. We always try to avoid high level of noises with potential to generate fines (ECOP9)	0.523	
Eigenvalues	2.954	1.273
Cronbach alpha	0.720	0.610
Percentage of variance explained	24.47	22.50

[a]"Varimax" rotation was performed.

Table A3
Factor loading of exploratory principal components analysis results of stakeholder management[a]

Item	Factors		
	Factor 1: STAKHF1	Factor 2: STAKHF2	Factor 3: STAKHF3
1. Competitors (STK1)			0.876
2. Leaders in the sector (STK2)			0.843
3. Customers (STK3)	0.725		
4. Suppliers (STK4)	0.766		
5. Stockholders/owners (STK5)	0.672		
6. Friends and relatives (STK6)	0.613		
7. Employees (STK7)	0.780		
8. Unions (STK8)		0.726	
9. Environmental activists (STK9)		0.879	
Eigenvalues	2.779	1.521	1.261
Cronbach alpha	0.765	0.515	0.695
Percentage of variance explained	28.605	17.514	15.738

[a]"Varimax" rotation was performed.

References

Ajzen, I., Fishbein, M., 1980. Understanding attitudes and predicting social behavior. Prentice-Hall, Englewood Cliffs, NJ.

Amit, R., Schoemaker, P.J., 1993. Strategic assets and organizational rent. Strategic Management Journal 14, 33–45.

Andersson, L.M., Bateman, T.S., 2000. Individual environmental initiative: championing natural environmental issues in U.S. business organizations. Academy of Management Journal 43, 548–570.

Aragón-Correa, J.A., 1998. Strategic proactivity and firm approach to the natural environment. Academy of Management Journal 41, 556–567.

Aragón-Correa, J.A., Sharma, S., 2003. A contingent resource-based view of proactive corporate environmental strategy. Academy of Management Review 28, 71–88.

Aragón-Correa, J.A., Matías-Reche, F., Senise-Barrio, M.E., 2004. Managerial discretion and corporate commitment to the natural environment. Journal of Business Research 57 (9), 964–975.

Banerjee, S.B., 2001. Managerial perceptions of corporate environmentalism: interpretations from industry and strategic implications for organizations. Journal of Management Studies 38, 489–513.

Bansal, T., 2005. Evolving sustainability. A longitudinal study of corporate sustainable development. Strategic Management Journal 26.

Barney, J.B., 1991. Firm resources and sustained competitive advantage. Journal of Management 17, 99–120.

Barney, J., Wright, M., Ketchen, D.J., 2001. The resource-based view of the firm: ten years after 1991. Journal of Management 27, 625–641.

Barret, S., Murphy, D., 1996. Managing corporate environmental policy: a process of complex change. In: Wehrmeyer, W. (Ed.), Greening People. Greenleaf, Sheffield, England, pp. 75–98.

Bianchi, R., Noci, G., 1998. Greening SMEs' competitiveness. Small Business Economics 11, 269–281.

Bollen, K.A., Long, J.S., 1993. Testing Structural Equation Models. Sage, Newbury Park, CA.

Bono, J.E., Judge, T.A., 2003. Self-concordance at work: toward understanding the motivational effects of transformational leaders. Academy of Management Journal 46, 554–571.

Brown, R., 1995. Marketing for Small Firms. Holt, Rinehart and Winston, London.

Buysse, K., Verbeke, A., 2003. Proactive environmental strategies: a stakeholder management perspective. Strategic Management Journal 24, 453–470.

Carlson-Skalak, S., 2000. E Media's Global zero: design for environment in a small firm. Interfaces 30, 66–83.

Céspedes-Lorente, J., Burgos-Jiménez, J., de Álvarezil, M.J., 2003. Stakeholders' environmental influence. An empirical analysis in the Spanish hotel industry. Scandinavian Journal of Management 19, 333–359.

Chandler, G.N., Hanks, S.H., 1993. Measuring the performance of emerging business: a validation study. Journal of Business Venturing 8, 391–408.

Chen, M., Hambrick, D.C., 1995. Speed, stealth, and selective attack: how small firms differ from large firms in competitive behavior. Academy of Management Journal 38, 453–482.

Christmann, P., 2000. Effects of 'best practices' of environmental management on cost advantage: the role of complementary assets. Academy of Management Journal 43, 663–680.

Clemens, B., 2001. Three phases of environmental strategies. Journal of Environmental Management 62, 221–231.

Conner, R.K., Prahalad, C.K., 1996. A resource-based theory of the firm: knowledge versus opportunism. Organizational Science 7, 477–501.

Cordano, M., Frieze, I.H., 2000. Pollution reduction preferences of U.S. environmental managers: applying Ajzen's theory of planned behavior. Academy of Management Journal 43, 627–641.

Cordano, M., Frieze, I.H., Ellis, K.M., 2004. Entangled affiliations and attitudes: an analysis of the influences on environmental policy stakeholders' behavioral intentions. Journal of Business Ethics 49, 27–40.

Covin, J.G., Slevin, D.P., 1990. New venture strategic posture, structure and performance: an industry life cycle analysis. Journal of Business Venturing 5, 123–135.

D'Amboise, G., Muldowney, M., 1988. Management theory for small business: attempts and requirements. Academy of Management Review 13, 226–240.

Darnall, N., 2002. Motivations for participating in a US voluntary environmental initiative: the multi-state working group and EPA's EMS pilot program. In: Sharma, S., Starik, M. (Eds.), Research in Corporate Sustainability: The Evolving Theory and Practice of Organizations in the Natural Environment. Edward Elgar Publishing, Inc., Northampton, MA, pp. 123–154.

Dean, T.J., Brown, R.L., Bamford, C.E., 1998. Differences in large and small firm responses to environmental context: strategic implications from a comparative analysis of business formations. Strategic Management Journal 19, 709–728.

Dean, T.J., Brown, R.L., Stango, V., 2000. Environmental regulation as a barrier to the formation of small manufacturing establishments: a longitudinal examination. Journal of Environmental Economics and Management 40, 56–75.

Eden, L., Levitas, E., Martinez, R.J., 1997. The production, transfer and spillover of technology: comparing large and small multinationals as technology producers. Small Business Economics 9, 53–66.

European Commission, 1999. The European Observatory for SMEs Fifth Annual Report 1997. European Network for SME Research, Brussels.

Fiegenbaum, A., Karnani, A., 1991. Output flexibility: a competitive advantage for small firms. Strategic Management Journal 12, 101–114.

Flannery, B.L., May, D.R., 2000. Environmental ethical decision making in the U.S. metal-finishing industry. Academy of Management Journal 43, 642–662.

Greening, D.W., Gray, B., 1994. Testing a model of organizational response to social and political issues. Academy of Management Journal 37, 467–498.

Greve, M.S., 1989. Environmentalism and bounty hunting. Public Interest fall, 15–19.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis, fifth ed. Prentice-Hall, Englewood Cliffs, NJ.

Hart, S.L., 1995. A natural-resource-based view of the firm. Academy of Management Review 20, 874–907.

Hendry, C., Arthur, M.B., Jones, A.M., 1995. Strategy Through People: Adaptation and Learning in the Small-Medium Enterprise. Routledge, London.

Henriques, I., Sadorsky, P., 1999. The relationship between environmental commitment and managerial perceptions of stakeholder importance. Academy of Management Journal 42, 87–99.

Hillary, R., 2000. Small and Medium-Sized Enterprises and the Environment. Greenleaf, Sheffield.

Hitt, M.A., Hoskisson, R.E., Harrison, J.S., 1991. Strategic competitiveness in the 1990s: challenges and opportunities for U.S. executives. Academy of Management Executive 5, 7–22.

Jehn, K.A., 1995. A multimethod examination of the benefits and detriments of intragroup conflict. Administrative Science Quarterly 40, 256–282.

Judge, W.Q., Douglas, T.J., 1998. Performance implications of incorporating natural environmental issues into the strategic planning process: an empirical assessment. Journal of Management Studies 35, 241–262.

Klassen, R.D., McLaughlin, C.P., 1996. The impact of environmental management on firm performance. Management Science 42, 1199–1214.

Kogut, B., Zander, U., 1996. What firms do? Coordination, identity, and learning. Organization Science 7, 502–519.

Lawrence, P.R., Lorsch, J.W., 1969. Organization and Environment; Managing Differentiation and Integration. R.D. Irwin, Homewood, IL.

Lee, K.S., Lim, G.H., Tan, S.J., 1999. Dealing with resource disadvantage: generic strategies for SMEs. Small Business Economics 12, 299–311.

Lehmann, M., Christensen, P., Larsen, J.M., 2005. Self-regulation and new institutions: the case of Green Network in Denmark. In: Sharma, S., Aragón-Correa, J.A. (Eds.), New Perspectives in Research on Corporate Sustainability: Corporate Environmental Strategy and Competitive Advantage. Edward Elgar Publishing, pp. 96–113.

Lehni, M., 2000. Eco-efficiency: Creating More Value with Less Impact. World Business Center for Sustainable Development, Geneva.

Lescure, M., 1999. Small- and medium-size industrial enterprises in France. In: Odaka, K., Sawai, M. (Eds.), Small Firms, Large Concerns. Oxford University Press, Oxford.

Lumpkin, G.T., Dess, G.G., 1996. Clarifying the entrepreneurial orientation construct and linking it to performance. Academy of Management Review 21, 135–172.

Lyon, D.W., Lumpkin, G.T., Dess, G.G., 2000. Enhancing entrepreneurial orientation research: operationalizing and measuring a key strategic decision making process. Journal of Management 26, 1055–1085.

Majumdar, S.K., Marcus, A.A., 2001. Rules versus discretion: the productivity consequences of flexible regulation. Academy of Management Journal 44, 170–179.

Marcus, A., Anderson, M., 2006. A dynamic capability and the acquisition of competencies in supply chain and environmental management. Journal of Management Studies 43, 19–46.

Marcus, A.A., Geffen, D., 1998. The dialectics of competency acquisition: pollution prevention in electric generation. Strategic Management Journal 19, 1145–1168.

Margolis, J.D., Walsh, J.P., 2003. Misery loves companies: rethinking social initiatives by business. Administrative Science Quarterly 48, 268–305.

Marshall, 1998. Economic Instruments and the Business Use of Energy. Stationery Office, London.

McEvily, W., Marcus, A., 2005. Embeddedness and the acquisition of competitive capabilities. Strategic Management Journal 26, 1033–1055.

Merrit, J.Q., 1998. EM into SME won't go: attitudes, awareness and practices in the London Borough of Croydon. Business Strategy and the Environment 7, 90–100.

Merz, G.R., Sauber, M., 1995. Profiles of managerial activities in small firms. Strategic Management Journal 16, 551–564.

Miles, M.P., Munilla, L.S., McClurg, T., 1999. The impact of ISO 14000 environmental management standards on small and medium sized enterprises. Journal of Quality Management 4, 11–122.

Miles, R., Snow, C., 1978. Organizational Strategy, Structure and Process. McGraw Hill, New York.

Miller, D., 1987. The structural and environmental correlates of business strategy. Strategic Management Journal 8, 55–76.

Miller, D., Friesen, P.H., 1984. Organizations: A Quantum View. Prentice-Hall, Englewood Cliffs, NJ.

Miller, D., Droge, C., Toulouse, J.M., 1988. Strategic process and content as mediators between organizational context and structure. Academy of Management Journal 31 (3), 544–569.

Newton, T., Harte, G., 1997. Green business: Technicist kitsch? Journal of Management Studies 34, 75–98.

O'Gorman, C., Doran, R., 1999. Mission statements in small and medium-sized businesses. Journal of Small Business Management 37, 59–66.

Orsato, R.J., den Hond, F., Clegg, S.R., 2002. The political ecology of automobile recycling in Europe. Organization Studies 23, 639–665.

Oswald, S.L., Mossholder, K.W., Harris, S.G., 1994. Vision salience and strategic involvement: implications for psychological attachment to organization and job. Strategic Management Journal 15, 477–489.

Podsakoff, P.M., Organ, D.W., 1986. Self reports in organizational research: problems and prospects. Journal of Management 12, 531–544.

Porter, M.E., 1980. Competitive Strategy. The Free Press, New York.

Powell, T., 1995. Total quality management as competitive advantage: a review and empirical study. Strategic Management Journal 16, 15–37.

Ramus, C.A., Steger, U., 2000. The roles of supervisory support behaviors and environmental policy in employee 'ecoinitiatives' at leading-edge European companies. Academy of Management Journal 43, 605–626.

Rangone, A., 1999. A resource-based approach to strategy analysis in small-medium sized enterprises. Small Business Economics 12, 233–248.

Raymond, L., Bergeron, F., Rivard, S., 1998. Determinants of business process reengineering success in small and large enterprises: an empirical study in the Canadian context. Journal of Small Business Management 36, 72–85.

Risseeuw, P., Masurel, E., 1994. The role of planning in small firms: empirical evidence from a service industry. Small Business Economics 7.

Rondinelli, D.A., London, T., 2003. How corporations and environmental groups cooperate: assessing cross-sectoral alliances and collaborations. Academy of Management Executive 17, 61–76.

Roome, N., 1992. Developing environmental management strategies. Business Strategy and the Environment 1, 11–24.

Ruiz-Quintanilla, S.A., Bunge, J., Freeman-Gallant, A., Cohen-Rosenthal, E., 1996. Employee participation in pollution reduction: a socio-technical perspective. Business Strategy and the Environment 5, 137–144.

Rumelt, R., 1984. Toward a strategic theory of the firm. In: Lamb, R. (Ed.), Competitive Strategic Management. Prentice-Hall, Englewood Cliffs, NJ, pp. 556–570.

Russo, M.V., Fouts, P.A., 1997. A resource-based perspective on corporate environmental performance and profitability. Academy of Management Journal 40, 534–559.

Rutherfoord, R., Blackburn, R.A., Spence, L.J., 2000. Environmental management and the small firm: an international comparison. International Journal of Entrepreneurial Behaviour and Research 6, 310–325.

Schaper, M., 2002. Small firms and environmental management: predictors of green purchasing in Western Australian pharmacies. International Small Business Journal 20, 235–251.

Schmidheiny, S., 1992. Changing Course: A Global Business Perspective on Development and the Environment. MIT Press, Cambridge, MA.

Scott, W.R., 1990. Ideology and the New Social Movements. Unwin Hyman, London.

Scranton, P., 1999. Moving outside manufacturing: research perspectives on small business in twentieth-century America. In: Odaka, K., Sawai, M. (Eds.), Small Firms, Large Concerns. Oxford University Press, Oxford.

Sharma, S., 2000. Managerial interpretations and organizational context as predictors of corporate choice of environmental strategy. Academy of Management Journal 43, 681–697.

Sharma, S., Henriques, I., 2005. Stakeholder influences on sustainability practices in the Canadian forest products industry. Strategic Management Journal 26, 159–180.

Sharma, S., Vredenburg, H., 1998. Proactive corporate environmental strategy and the development of competitively valuable organizational capabilities. Strategic Management Journal 19, 729–753.

Shrivastava, P., 1995. Environmental technologies and competitive advantage. Strategic Management Journal 16, 183–200.

Shuman, J.C., Shaw, J.J., Sussman, G., 1985. Strategic planning in smaller rapid growth companies. Long Range Planning 18, 48–53.

Smeltzer, L.R., Fann, G.L., 1989. Comparison of managerial communication patterns in small, entrepreneurial organizations and large, mature organizations. Group and Organization Studies 14, 198–215.

Smith, M.A., Kemp, R., 1998. Small Firms and the Environment 1998: A Grounded Report. Groundwork, Birmingham.

Stopford, J.M., Baden-Fuller, C., 1994. Creating corporate entrepreneurship. Strategic Management Journal 15, 521–536.

Strategic Planning Institute, 1977. Selected Findings from the PIMS Program. Strategic Planning Institute, Cambridge, MA.

Tarras-Wahlberg, N.H., 2002. Environmental management of small-scale and artisanal mining: the Portovelo–Zaruma goldmining area, southern Ecuador. Journal of Environmental Management 65, 165–179.

Wagner, M., 2005. How to reconcile environmental and economic performance to improve corporate sustainability: corporate environmental strategies in the European paper industry. Journal of Environmental Management 76, 105–118.

Way, S.A., 2002. High performance work systems and intermediate indicators of firm performance within the US small business sector. Journal of Management 28, 765–785.

WBCSD, 2001. The Business Case for Sustainable Development: Making a Difference Toward the Johannesburg Summit 2002 and Beyond. World Business Center for Sustainable Development, Geneva.

Wehrmeyer, W., Parker, K.T., 1996. Identification and relevance of environmental corporate cultures as part of a coherent environmental policy. In: Wehrmeyer, W. (Ed.), Greening People. Greenleaf, Sheffield, England, pp. 163–184.

Wernerfelt, B., 1984. A resource-based view of the firm. Strategic Management Journal 5, 171–1804.

Wheeler, D., Colbert, B., Freeman, R.E., 2003. Focusing on value: reconciling corporate social responsibility, sustainability and a stakeholder approach in a network world. Journal of General Management 28, 1–28.

Williamson, D., Lynch-Wood, G., 2001. A new paradigm for SME environmental practice. TQM Magazine 13, 424–432.

Woo, C.Y., 1987. Path analysis of the relationship between market share, business-level conduct and risk. Strategic Management Journal 8, 149–168.

Yu, T.F., 2001. Toward a capabilities perspective of the small firm. International Journal of Management Reviews 3, 185–197.

Journal of Environmental Management 86 (2008) 104–113

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Assisting Australian indigenous resource management and sustainable utilization of species through the use of GIS and environmental modeling techniques

Julian Gorman[a], Diane Pearson[b,*], Peter Whitehead[c,1]

[a]*School for Environmental Research, Institute of Advanced Studies, Charles Darwin University, Darwin, Northern Territory 0909, Australia*
[b]*Tropical Spatial Science Group, School of Science and Primary Industries, Faculty of Education Health and Science, Charles Darwin University, Darwin, Northern Territory 0909, Australia*
[c]*Key Centre for Tropical Wildlife Management, Charles Darwin University, Darwin, Northern Territory 0909, Australia*

Received 3 November 2005; received in revised form 3 September 2006; accepted 25 November 2006
Available online 15 February 2007

Abstract

Information on distribution and relative abundance of species is integral to sustainable management, especially if they are to be harvested for subsistence or commerce. In northern Australia, natural landscapes are vast, centers of population few, access is difficult, and Aboriginal resource centers and communities have limited funds and infrastructure. Consequently defining distribution and relative abundance by comprehensive ground survey is difficult and expensive. This highlights the need for simple, cheap, automated methodologies to predict the distribution of species in use, or having potential for use, in commercial enterprise. The technique applied here uses a Geographic Information System (GIS) to make predictions of probability of occurrence using an inductive modeling technique based on Bayes' theorem. The study area is in the Maningrida region, central Arnhem Land, in the Northern Territory, Australia. The species examined, *Cycas arnhemica* and *Brachychiton diversifolius*, are currently being 'wild harvested' in commercial trials, involving sale of decorative plants and use as carving wood, respectively. This study involved limited and relatively simple ground surveys requiring approximately 7 days of effort for each species. The overall model performance was evaluated using Cohen's kappa statistics. The predictive ability of the model for *C. arnhemica* was classified as *moderate* and for *B. diversifolius* as *fair*. The difference in model performance can be attributed to the pattern of distribution of these species. *C. arnhemica* tends to occur in a clumped distribution due to relatively short distance dispersal of its large seeds and vegetative growth from long-lived rhizomes, while *B. diversifolius* seeds are smaller and more widely dispersed across the landscape. The output from analysis predicts trends in species distribution that are consistent with independent on-site sampling for each species and therefore should prove useful in gauging the extent of resource availability. However, some caution needs to be applied as the models tend to over predict presence which is a function of distribution patterns and of other variables operating in the landscape such as fire histories which were not included in the model due to limited availability of data.

Keywords: Sustainable utilization; Predictive modeling; Bayes' Theorem; Indigenous; Commercial

1. Introduction

The Northern Territory covers approximately one sixth of the Australian continent and has a population of about 211,000 people (ABS, 2001) at a density of one person every 6.4 km^2. In 2001, 28% of the Northern Territory population was identified as being of indigenous origin (Kinfu and Taylor, 2002) of whom 60.3% were recorded as living in rural localities. Approximately half (53%) of the

*Corresponding author. Tel.: +61 8 89466046; fax: +61 8 89467088.
E-mail addresses: julian.gorman@cdu.edu.au (J. Gorman), diane.pearson@cdu.edu.au (D. Pearson), peter.whitehead@cdu.edu.au, Peter.Whitehead@nt.gov.au (P. Whitehead).

[1]Present Address: School for Environmental Research, Institute of Advanced Studies, Charles Darwin University, Darwin, Northern Territory, 0909, Australia and Department of Natural Resources Environment and The Arts, Northern Territory Government, Darwin, Northern Territory, Australia.

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.033

and in the Northern Territory is under Aboriginal Title or under claim and most of this land is being managed by the Aboriginal custodians, often for subsistence use (Altman, 1987). Aboriginal people have detailed knowledge of native flora and fauna and evidence suggests that their management practice is important in sustaining contemporary patterns of biodiversity (Yibarbuk et al., 2001).

Indigenous people in remote areas of northern Australia have limited opportunity to engage in mainstream employment and there are many challenges in developing viable resource-based industries. They confront the tyranny of distance, limited infrastructure, lack of capital and difficulties in raising capital on land held under inalienable communal title. A Senate committee inquiring into commercial use of native wildlife concluded that activities involving wild plants and animals present important commercial opportunity for Aboriginal people (SRRATRC, 1998). The reasons given for this conclusion were that: there are few alternatives on lands owned by Aboriginal people; wildlife use and management makes good use of existing skills and interest; communities already manage a substantial customary economy based in part on wildlife and that involvement with wildlife fosters customary connections to land that many are eager to maintain (Martin, 1995).

Across the Top End of the Northern Territory over 30 Aboriginal land and sea ranger groups are involved in natural resource management activities using a combination of customary skills and formal science. These ranger groups are increasingly seeking opportunities to become more actively involved in commercial use of wildlife.

In order to make informed judgments regarding sustainability of use and to fix conditions for commercial use, the Traditional Aboriginal Owners and the Northern Territory regulating authority, the Parks and Wildlife Service, need to understand the distribution and abundance of the species to be harvested. Collecting this information in remote areas for which there is little existing environmental data is both time consuming and expensive. Limited economic returns from such small-scale ventures cannot justify the expense of the comprehensive resource inventories that may be needed to support large harvests. A methodology is needed that is congruent with the low levels of utilization most often involved, easy to understand and use, spatially explicit, has a visual output, and is not too data intensive.

Many methods have been developed for predicting habitat suitability and species distributions. The approaches for modeling habitat suitability can be grouped into seven categories (Guisan and Zimmerman, 2000). These include multiple regression and generalized forms (e.g. Generalized Linear Models (GLMs) (Guisan et al., 1998, 1999) and Generalized Additive Models (GAMs) (Yee, 1991); classification techniques (e.g. classification and regression trees (CART) (Franklin, 1998; Franklin et al., 2000; Yee, 1991); environmental envelopes (e.g. BIOCLIM (Busby, 1991), HABITAT (Walker and Cocks, 1991) DOMAIN (Carpenter et al., 1993); ordination techniques (e.g. canonical correspondence analysis (CCA) (Hill, 1991); Bayesian modeling (Aspinall, 1992); artificial neural networks (ANN) (Fitzgerald and Lees, 1992); and others including combinations, e.g. discriminant function analysis (DFA) (Frank, 1988; Lowell, 1991), ecological niche-factor analysis ENFA (Hirzel, 2001), MONOMAX (a monotonic maximum-likelihood function) (Bayes and Mackey, 1991; Mackey, 1993) and SIMPLE (a spatial and inductive modeling tool) (Walker and Moore, 1988). Since there are already a variety of papers which provide overviews of these approaches (e.g. Guisan and Zimmerman, 2000; Hirzel et al., 2001; Wintle et al., 2005), it is not the intention of this paper to also review the suite of methodologies available.

The approach used in this study to analyze habitat associations and predict distribution was Bayesian probability modeling in a GIS environment (Aspinall, 1992) which works by analyzing spatial patterns in the data, i.e. it predicts distribution within one data set by combining a number of data sets based on Bayes theorem. This particular methodology can be justified because it offered a conceptually simple approach that used basic spatial data and generated spatially explicit outputs in the form of maps. Bayesian statistics have been widely used in other discipline areas for decision-making under conditions of uncertainty (*e.g.* Agterberg, 1989; Lusted, 1968) and have been recognized for some time as having good potential for analysis of spatial data (Cressie, 1991). A Bayesian approach has also been used in a variety of applications to successfully model species distributions (Aspinall, 1992; Aspinall and Veitch, 1993; Aspinall et al., 1998; Brzeziecki et al., 1993; Skidmore, 1989) and has recently been applied to map vegetation in the form of a weed species in northern Australia (Ferdinands, 2006). A comparison of this approach to other habitat modeling procedures can be seen in (Guisan and Zimmerman, 2000). The predicted distribution that results from the model generated by the Bayesian approach (Aspinall, 1992) is expressed as the probability of occurrence that can be treated as a measure of habitat suitability. This probability of occurrence is generated by comparing the attributes of the data set to be modeled with attributes of predictor data sets using a type of inductive learning.

This approach was applied to two plant species currently being harvested from wild populations in Maningrida (central Arnhem Land). These were *Cycas arnhemica*, which is being harvested for collectors and the landscape gardening market, and *Brachychiton diversifolius*, which is commonly harvested by Aboriginal artists as a carving wood (Griffiths et al., 2003).

To meet the requirements of management authorities an appropriate modeling technique must provide reasonable robust reliable predictions despite the available environmental data being relatively coarse for much of the landscape. Information deficits are unlikely to be overcome quickly because priorities for mapping environmental variables are determined by demand for such things as horticultural land and exploitable water resources, rather than to assess availability of individual species found in less productive, often agriculturally marginal land.

There is a demonstrable need to improve capacity to make broad-scale resource assessments that are sufficiently reliable to inform management planning. This study examines the utility of one readily available and easily implemented application of Geographic Information Systems and associated statistical methods for describing resource distribution.

2. Background and study site

2.1. Species

Cycads belong to an ancient family of slow growing, diecious plants, dominant during most of the Mesozoic period, but which are now restricted to tropical and subtropical regions (Whiting, 1963). Eleven species are recognized in the Northern Territory and all are listed under Appendix 2 of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Listing imposes management obligations on Australia which are met through provision of the *Environmental Protection and Biodiversity Conservation Act* 1999. The Northern Territory Cycad Management Plan (PWCNT, 1997) allows for experimental harvest of *C. arnhemica* (and other cycad species), limited to 500 plants of each species per annually. A local indigenous organization, the Bawinanga Aboriginal Corporation has acquired a permit from the Parks and Wildlife Service to harvest adult stems of *C. arnhemica* for commerce (Griffiths et al., 2005; Gorman et al., 2006). Estimating the extent of the cycad population will have an important influence on future management plans and in particular setting of harvest limits.

B. diversifolius is a semi-deciduous tree that grows to 7–15 m in height. It is found in open forest and woodland on a wide variety of well-drained sites extending to sparse savanna woodland in drier regions. Aboriginal people use the seed for food, the inner bark for medicinal purposes and rope making, and the wood for fire sticks and spear making (Brock, 1988) as well as for carving and paintings for sale. As the demand for carvings is growing (Griffiths et al., 2003) improved information on the species distribution and abundance in areas of use is increasingly important.

2.2. Study site

The study area is the Maningrida region of central Arnhem Land, 500 km east of the NT's capital city Darwin (Fig. 1). The regional landscape is dominated by the Arnhem Land sandstone escarpment to the south, extensive wetland systems nearer the coast, and gently undulating lowland plains (Griffiths et al., 2000). The dominant vegetation is *Eucalyptus* savanna, interspersed

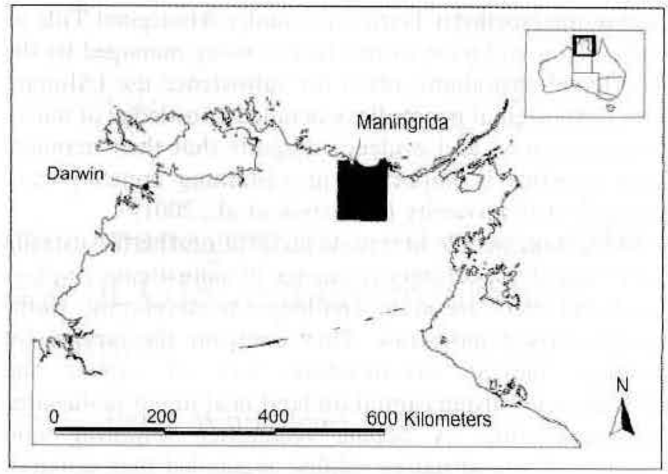


Fig. 1. Map of Maningrida study site.

elsewhere with floodplain, wetlands and monsoon rainforest patches that are most common in the coastal/sub coastal belt (Griffiths et al., 2003). Elevation ranges up to 60 m.a.s.l., mean annual rainfall is 1245 mm of which 92.6% falls between December and April inclusive (Maningrida Airport, Bureau of Meteorology). Maningrida township has a population that may reach more than 2000 people in the wet season when access to other areas is difficult, and up to 800 people reside in 32 outstations or homeland communities serviced from Maningrida (Hall, 2002).

3. Methods

As mentioned above there are many approaches to habitat suitability modeling, each with their own set of advantages and disadvantages. The type of survey data that is available plays a major role in determining the type of modeling method that can be applied (Wintle et al., 2005). A Bayesian approach was taken in this study, rather than applying some of the other habitat modeling procedures. This was because many of the other approaches that meet the criteria of being simple to use and easy to interpret spatially, like BIOCLIM (Nix, 1986), DOMAIN (Carpenter et al., 1993), ENFA (Hirzel, 2001) and GARP (Stockwell and Peters, 1999), are presence only methods, whereas the Bayesian approach can also work with presence–absence data. Presence–absence models have been noted as being superior to presence-only models (Wintle et al., 2005). The method in this case is implemented within a GIS running ArcView 3.3 (ESRI, 2002) and uses an extension developed by Aspinall (1999) which is based on the Bayesian Modeling approach described in Aspinall (1992). Additionally, some Aboriginal rangers in the area are familiar with ArcView 3.3 software and the approach could be equally applied in other versions of GISs by reproducing the code in the relevant programming language. The model works by calculating conditional probabilities from the relative frequency of association between a response variable (e.g. presence or absence of a species) and attributes of a variety of predictor data sets (e.g. environmental data).

3.1. Field observations

The objective of this study was to define, and test a methodology that could provide a description of the distribution of a plant population at a reasonable cost.

Presence/absence data were collected from quadrats (5 × 4 m for *C. arnhemica* and 10 × 5 m for *B. diversifolius*) which were positioned at 5 km intervals along all accessible roads and hunting tracks in each of the study sites (Fig. 2). To try and minimize the influence of the roads/hunting tracks the quadrats were positioned 50 m into the bush in a random direction without regard for the presence of the target species. The difference in quadrat size reflects the demographics of the species with *C. arnhemica* growing either vegetatively or with limited dispersal and therefore tending to occur in clumps while *B. diversifolius* tending to be more widely dispersed generally at lower densities. The number of sample/quadrats for *C. arnhemica* was 126 (64 presence and 62 absence) and for *B. diversifolius* was 224 (123 presence and 101 absence)—the difference in number of data points being due to difference in size of study areas and accessibility. The coordinates of the middle of each quadrat was taken using a Garmin GPS 12 XL. The accuracy of this model of GPS is recorded in the manual as having error of up to 15 m. However, a satellite averaging function, which reduces the effects of selective availability upon position error, would have provided a more accurate saved position.

3.2. Environmental/predictor data sets

As discussed above, spatial data available for the Maningrida area is limited and most often only available at a relatively coarse environmental scale. Data were selected from available mapped environmental themes to reflect the factors thought most likely to influence species distribution, based on both a knowledge of the species ecology from consultation with experts, and indigenous knowledge of the species. A mixture of raster and broad scale vector data was used. Spatial data used included satellite imagery, maps of habitat and soil types, plus coverage reflecting distance to water. Despite the important role that climate plays in the natural distribution of species it was not treated as a significant environmental variable in this particular study. This was because climate can be considered to be the same across the study site, as data is only captured at one recording station in the vicinity. Instead of climate, location relative to water is thought to play the most important role in determining species occurrence (particularly in the case of *B. diversifolius*). Higher densities of *B. diversifolius* have been recorded in coastal areas and adjacent to riverine areas, this is thought to be associated with soil fertility (Jennifer Koenig-Price, Unpublished data). As well as having little climatic variability, the study sites also have very little variation in elevation. Again data for this area is not available at a scale where differences in elevation (and its derivatives of slope and aspect) would have much influence on the output, so this data was not included in the model.

A soil data set was extracted from a vector soil coverage for the Northern Territory that was derived from the Geology of Australia digital data set. This was the only soil data set that was available to the researchers at the time of analysis. Soil classes for the Northern Territory consist of 321 different categories—6 of which were found in the area examined for *B. diversifolius* and 5 in the area examined for *C. arnhemica*. Habitat types were taken from Griffiths et al. (2000) which identified and characterized 15 distinct types, i.e. dry rainforest, *Eucalyptus* open forest on laterised plateau, *Eucalyptus* open forest on sand sheets, *Eucalyptus* open woodland on clay plains, *Eucalyptus* woodland on

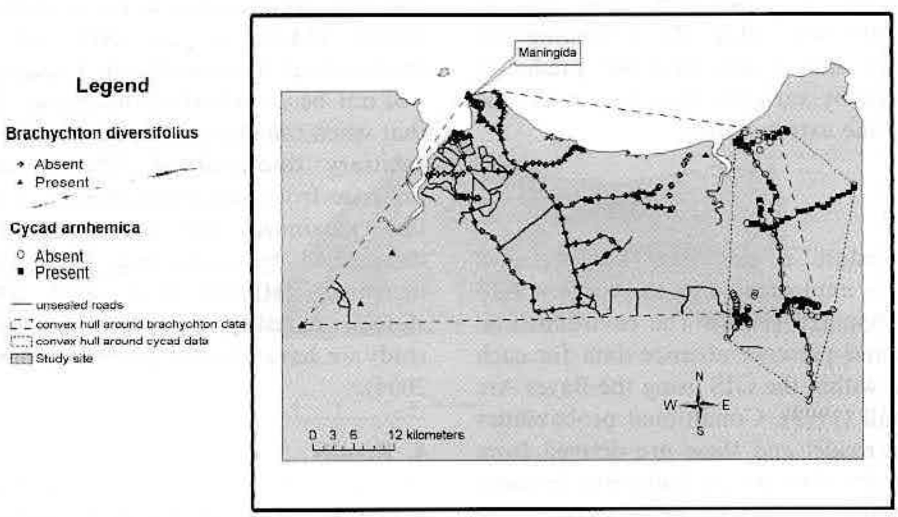


Fig. 2. Illustration of study site and distribution of sample data collected using roads and hunting tracks.

clay foot slopes, floodplain complex, grassland on hills, mangroves, *Melaleuca cajaputi* forest, open chenier ridges, saline wetlands, stream complex, sandstone complex, swamps, and wet rainforest (on springs). Of these habitat types 12 where found to occur in the area examined for *B. diversifolius* and 6 in the area examined for *C. arnhemica*. Both the soil and habitat data were treated as distinct categorical data sets.

The satellite imagery consisted of band 3 of a Landsat TM image at 30 m resolution collected in the dry season in 1999. The raw unclassified data were used to reflect the fine-scale variability in vegetation present across the landscape which is not captured by broad categorical soil class and habitat data. This unclassified band was used to represent a surface of continuous variation. The use of continuous data of this type has been shown to be useful in quantifying the structure present in heterogeneous northern Australian landscapes (Pearson, 2002). This continuous data set was thought to offset the limitations of just using the categorical soil and habitat data by providing additional finer-scale information within the boundaries of the soil and habitat classes.

Topographic data in the form of water bodies generated by Geoscience Australia at the scale 1:250,000 was used to generate a surface of distance to permanent water bodies using the GIS buffer function. Distance to water was broken down into 8 categories $\leqslant 50$ m, 51–200 m, 102–500 m, 501–1000 m, 1001–2000 m, 2001–3000 m, and 3001–5000 m. All data were projected to UTM Australian Map Grid Zone 53 and a datum of GDA94. All analysis was done in a raster data structure after conversion of vector data.

To reduce the error in the model that may arise in application to large areas not being covered by the field survey, a convex hull was placed round the survey points for each species and this was used to reduce the area over which modeling analysis was done. In mathematical terms a convex hull is the minimal vector which contains all the points sampled. The extent of the area over which *C. arnhemica* was modeled was 546.3 km^2 and for *B. diversifolius* it was 1915.2 km^2 (Fig. 2). Confining the model within a convex hull means that the predictive output is not influenced by variables that lie outside the area of interest during the extrapolation.

3.3. The model

The modeling procedure is described at length in Aspinall (1992) and was implemented in an ArcView GIS extension outlined in Aspinall (1999). The environmental data described above and presence/absence data for each species were combined within the GIS using the Bayes Arc View extension Aspinall (1999). Conditional probabilities are the input into the model and these are derived from what Aspinall (1992) describes as 'an inductive learning process in which attributes of data set to be modeled are compared with attributes of a variety of predictor data sets'. With the given combination of predictor data sets a probability value is generated for each grid cell within the study site, allowing probability surfaces to be mapped for each species.

Model outputs were mapped and then grouped into three categories: probability of occurrence 70–100% (high probability of occurrence), 30–70% (intermediate probability of occurrence) and 0–30% (low probability of occurrence) for ease of presentation.

3.4. Validation

Validation of the predictive maps involved independent additional ground sampling. This involved surveying 271 and 235 quadrats (50 × 50 m) within each of the *C. arnhemica* and *B. diversifolius* study sites, respectively, and recorded presence or absence of the species.

We used Arc View 3.3 to place a 200 m buffer around the roads and hunting tracks and generate random points in each of the three probability classes. At each study site a number of random points were chosen for each of the high, medium or low probability categories (in total 22 for *C. arnhemica* and 16 for *B. diversifolius*) and these were used as start points from which a further 12 50 × 50 m quadrats were sampled every 100 m, the 100 m mark being the center of the smaller quadrat. The direction of the quadrats from these starting points was randomly chosen (from degrees written on paper and chosen in a hat) with possible directions being between 0° and 180° from the road to ensure transects did not cross over the road. Subsequently, the individual 50 × 50 m quadrats fell in the same direction as the larger quadrat. This method was used to survey away from the road/track while still minimizing the time spent collecting the validation data.

3.5. Testing the model

Cohen's kappa statistics were used to test overall model performance and provide a simple, effective, standardized and appropriate statistic for evaluating presence–absence models (Manel et al., 2001). A description of this methodology is available in Fielding and Bell (1997) so will not be described in this paper. It has been suggested that when calculating Kappa statistics, instead of using an arbitrary threshold of 50% to distinguish simulated presence from simulated absence, a probability threshold that maximizes the model's performance should be determined by evaluating k at successive probability increments (Huntley et al., 1995, 2004). In using Kappa statistics to test the accuracy of the models generated in this study we have adopted a similar approach (Huntley et al., 2004).

4. Results

The coverage of *C. arnhemica* and *B. diversifolius* predicted from the Bayesian models using the presence/

absence field data are illustrated in Figs. 3 and 4, respectively.

Landis and Koch (1977) suggest the following agreement measures for categorical data using Kappa statistics. A Kappa statistic below 0.00 equates to a *poor* strength of agreement; between 0.00 and 0.20 equates to *slight* agreement; between 0.21 and 0.40 equates to *fair* agreement; between 0.41 and 0.60 equates to *moderate* agreement; between 0.61 and 0.80 equates to *substantial* agreement; and between 0.81 and 1.00 equates to almost perfect agreement.

The evaluation of the models using Kappa statistics can be seen in Table 1. The validation for the model generated for *C. arnhemica* gave a Kappa statistical peak at 0.56 with a probability threshold of 50%. This has a classification of *moderate* (Landis and Koch, 1977) and indicates that this model is most reliable for mapping distribution of this species at 50% probability of occurrence or above.

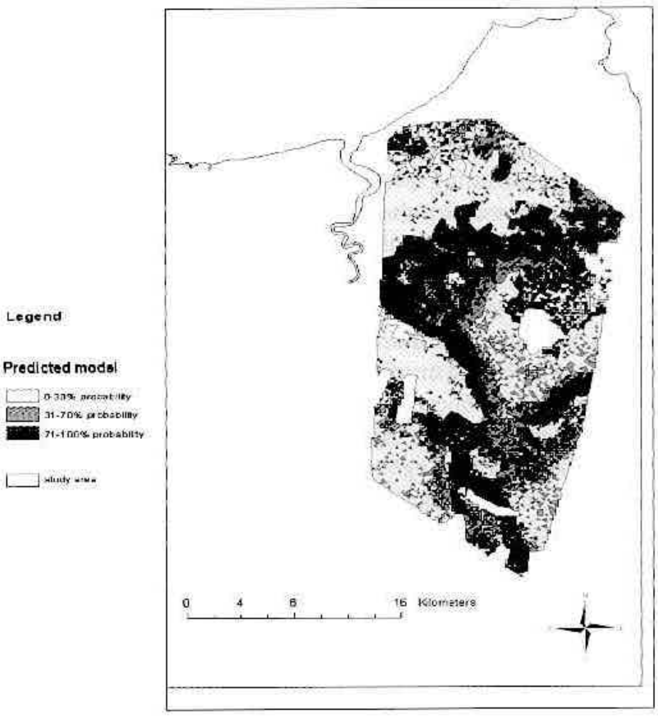


Fig. 3. Map showing predicted probability of occurrence of *Cycas arnhemica* using Bayesian modeling.

The Kappa statistic for the model produced for *B. diversifolius* peaked at 0.36 with a probability threshold of 90%. This has a classification of *fair* (Landis and Koch, 1977) and indicates that this model is most reliable using the predicted output that falls between 91% and 100% probability of occurrence. Below this threshold the model for *B. diversifolius* appears to over estimate the presence of this species.

The coverage of *C. arnhemica* and *B. diversifolius* using the probability threshold associated with the Kappa statistical peaks are illustrated in Figs. 5 and 6 respectively.

5. Discussion

The methodologies described above, involving GIS and the Bayes' Theorem, have produced predictions of species

Table 1
Cohen's kappa statistics using different probability thresholds for *C. arnhemica* and *B. diversifolius*

Probability threshold (%)	Kappa statistic	
	C. arnhemica	*B. diversifolius*
30	0.47	0.24
50	0.56	0.30
55	0.56	0.28
60	0.54	0.31
70	0.47	0.32
75	0.47	0.32
80	0.47	0.33
85	0.36	0.35
90	0.24	0.36
95	0.19	0.36

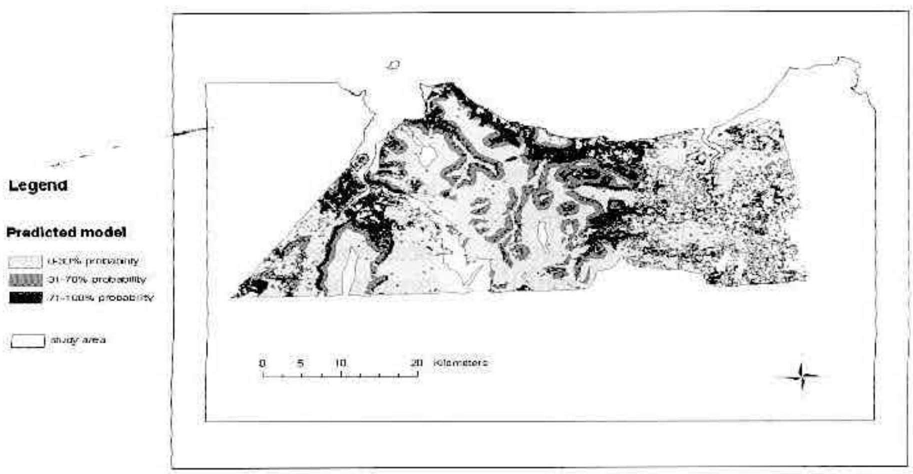


Fig. 4. Map showing predicted probability of occurrence of *Bracychiton diversifolius* using Bayesian modeling.

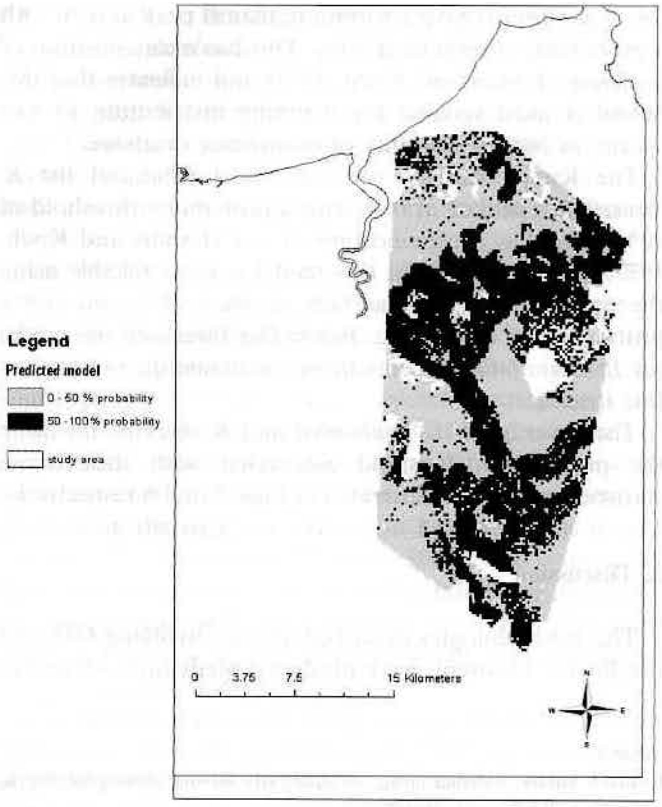


Fig. 5. Map of most reliable probability distribution for *Cycas arnhemica* based on Kappa statistic.

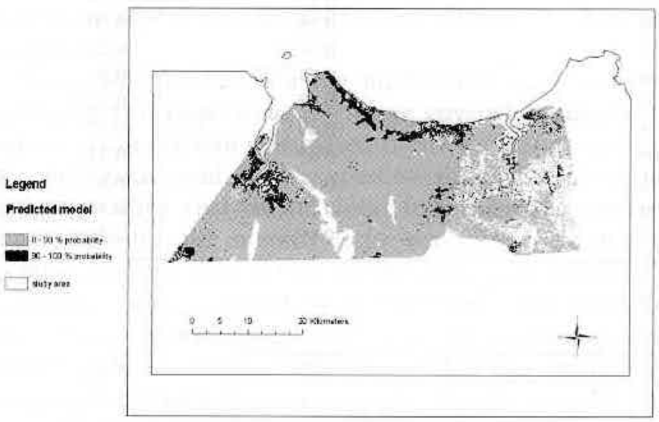


Fig. 6. Map of most reliable probability distribution for *Bracychiton diversifolius* based on Kappa statistic.

distribution that match well with what could be expected from the known species distribution patterns for both *C. arnhemica* and *B. diversifolius*. The Kappa statistics evaluate the models as performing *moderate* and *fair* for *C. arnhemica* and *B. diversifolius* respectively. The fact that the models were found to peak at different probability thresholds for each species suggests the model outputs should be presented and used for management purposes differently for each species. For example, the Kappa statistic indicates that the predicted distribution of *C. arnhemica* is most accurate when mapped at a probability threshold of 50% and above, whereas the predicted distribution of *B. diversifolius* can only reliably be mapped at a probability threshold of 90% and above.

Although the variables used to generate the predictive models were the same for both species the relative importance of the predictor data sets was different. In the case of *B. diversifolius* distance to water was the most important variable in the model generated, whereas for *C. arnhemica* the most important variable was the unclassified satellite imagery. This confirms what we would expect from expert knowledge of these species and their distribution patterns which occur relative to water for *B. diversifolius* and relative to the fine scale variability in the landscape for *C. arnhemica*.

The predictive ability and reliability of the models can be explained by the way these species reproduce and disperse through the landscape. Cycads have been characterized as slow growing and reproduction can occur from seed or vegetative sprouting (Watkinson and Powell, 1997). Dispersal agents and mechanisms of *C. arnhemica* are poorly understood but dispersal is limited, possibly due to the fact that the seed are highly toxic. Previous studies on similar species of cycads have reported dispersal as being generally less than 1 m in *Cycas armstrongii* (Watkinson and Powell, 1997) and up to 24 m in *Macrozamia riedlei* (Burbidge and Whelan, 1982). These low dispersal mechanisms results in cycad species occurring most commonly in clumped agglomerations throughout the landscape. This predictive capability of the Bayesian approach works well with this clumped pattern of occurrence and this is indicated by the *moderate* classification at a probability threshold of 50%. Additionally, the Bushfire Council of the Northern Territory have actively been promoting cooler fires in central Arnhem Land area and this strategy may have benefited the recruitment and vegetative growth of *C. arnhemica* resulting in an increase of occurrence.

The seed of *B. diversifolius* is attractive to both birds and arboreal mammals and as a result it is well dispersed. At favorable sites this species can be found in high density but it is more often found in low density and scattered across the landscape. This pattern can be seen quite clearly in Fig. 4 where the predicted occurrence is mainly in Category 1 (low probability of occurrence) with much less area in Category 2 and 3 (intermediate and high probability of occurrence). The *fair* classification for the prediction of *B. diversifolius* with a probability threshold at 90% indicates that the Bayesian modeling has over predicted the occurrence of this species and that the most accurate prediction of occurrence is between 91% and 100%. Possibly a greater number of sample points is needed to increase the accuracy in this instance.

There are a number of factors that may influence the accuracy of predictive models and these should be acknowledged. Available data was a considerable limiting factor in this study. The choice of predictor data sets was drawn from what could be freely accessed at the time of the study. It is probable that the inclusion of other predictor variables in the analysis could have improved the predictive accuracy of the models. In particular, information such as fire history could be important in influencing the distribution of these species as extensive areas of the landscape in this region are burnt on a regular basis which influences the patterns of vegetation present in the landscape (Whitehead et al., 2003; Yibarbuk et al., 2001). However, data on fire history in the Maningrida region was not available for this analysis. The scale of accuracy of the spatial data used to predict occurrence based on species presence and absence was also fairly course and this level of accuracy will be reflected in the modeled output. The pattern that was chosen to validate the modeled predictions is classic pseudo replication in that only a small subset of the validation points were randomly selected and the others were clustered around these random points. This methodology can be supported as a trade off in meeting one of the main objectives of this spatial analysis study, which was to keep field costs to a minimum. The level of access in remote landscapes of the Northern Territory will always inhibit the completeness and randomness of data collection and there will often be a trade off between the amount and thoroughness of fieldwork and the expected accuracy of the end result. In collecting data in close proximity to roads it is possible to incorporate a bias due to the environmental characteristics of where roads are positioned (soils, hydrology, slope, etc). Additionally, an ease of access may accentuated this bias because of higher harvest rates occurring along road sides. To minimize these influences in this study we collected the verification data at random points that were up to 500 m away from roads/hunting tracks.

Applying GIS, and an inductive modeling procedure like Bayes' Theorem, can be seen to provide a simple methodology for predicting species distribution which involves limited data collection and verification, a visual output which may facilitate the incorporation of local knowledge, and an acceptable level of accuracy. It appears that the way species are distributed throughout the landscape, clumped or dispersed, may influence the accuracy of predictions and the size of quadrats chosen to measure occurrence should take this into account. A big advantage of this methodology is that it is easily replicable and this is an important component in measuring how a resource changes over time.

In remote locations and with a limited budget and resource availability this methodology is capable of producing estimates of distribution that we ascertain would be acceptable for initial assessments of resource availability. In combination with linked assessments of variation in density stratified by the measures of habitat suitability provided here, proponents of commercial harvest will be able to place their proposals in context. Namely to:

(i) compare the size of the proposed harvest with resource availability in the regional landscape,
(ii) develop and provide details of spatially explicit designs for harvest management, including specification of harvested and protected sites,
(iii) develop designs for monitoring impact of harvests that are congruent with the scale of harvest relative to resource availability,
(iv) develop proposals for ongoing refinement of understanding of resource availability and dynamics, should that appear necessary or desirable.

Regional management authorities in the Northern Territory indicate that application of this approach will be acceptable to support initiation of trial harvests, where the proposed scale of harvest is demonstrably minor in terms of both spatial extent and the quantity of resource to be taken. The need to refine estimates of distribution and abundance would be determined as an element of the trials.

6. Conclusion

The results of this study show that GIS and environmental modeling techniques are useful for predicting the distribution of species that may be of interest to Australian indigenous groups wanting to engage in sustainable utilization of wildlife and enterprise development. As with any tool that predicts distribution of a potential commercial commodity it is important that caution is applied to ensure over prediction does not lead to over utilization that could affect the long-term viability of the species. However, the overall aim of having produced such a predictive model is to ensure that regulating bodies such as the Parks and Wildlife Service have a good starting point for understanding the extent of the resource and can monitor impacts. Without such simple modeling procedures it would be expensive and difficult to quantify the extent of available resources over large and inaccessible areas, such as those found in central Arnhem Land.

Acknowledgments

We would like to thank Charles Darwin University for funding this project, the Djelk Rangers and Bawinanga Aboriginal Corporation for providing assistance with the field work, numerous Traditional Owners in the Maningrida region for allowing access to land to conduct the surveys, Richard Aspinall for allowing use of his Bayesian modeling extension for ArcView and Jennifer Koenig-Price and Tony Griffiths for their input to the design of the project knowledge of the target species.

References

ABS, 2001. Census of Population and Housing. Australian Bureau of Statistics, Canberra.

Agterberg, F.P., 1989. Systematic approach to dealing with uncertainty of geoscience information in mineral exploration. In: 21st APCOM Symposium; March 1989, Las Vegas.

Altman, J.C., 1987. Hunter-Gatherers Today: an Aboriginal Economy in North Australia. Angus and Robertson, Canberra.

Aspinall, R., 1992. An inductive modelling procedure based on Bayes' theorem for analysis of pattern in spatial data. International Journal of Geographical Information System 6, 105–121.

Aspinall, R., 1999. An Introduction to Bayesian Modeling, Workshop Handbook. Darwin, pp. 1–79.

Aspinall, R., Veitch, N., 1993. Habitat mapping from satellite imagery and wildlife survey using a Bayesian modeling procedure in a GIS. Photogrammetic Engineering Remote Sensing 59, 537–543.

Aspinall, R.J., Burton, G., Landenburger, L., 1998. Mapping and modeling wildlife species distribution for biodiversity management. In: Environmental Systems Research Unit User Conference; Darwin.

Bayes, A.J., Mackey, B.G., 1991. Algorithms for monotonic functions and their application to ecological studies in vegetation science. Ecological Modelling 56, 135–159.

Brock, J., 1988. Top End Native Plants: a Comprehensive Guide to the Trees and the Shrubs of the Top End of the Northern Territory. Reed New Holland, Sydney.

Brzeziecki, B., Kienast, F., Wildli, O., 1993. A simulated map of the potential natural forest vegetation of Switzerland. Journal of Vegetation Science 4, 499–508.

Burbidge, A., Whelan, R., 1982. Seed dispersal in a cycad, *Macrozamia riedlei*. Australian Journal of Ecology 7, 63–67.

Busby, J.R., 1991. BIOCLIM—a bioclimate analysis and prediction system. In: Margules, C.R., Austin, M.P. (Eds.), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne Chapter 10.

Carpenter, G., Gillison, A.N., Winter, J., 1993. DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals. Biodiversity Conservation (2), 667–680.

Cressie, N., 1991. Statistics for Spatial Data. Wiley-Interscience, New York.

Ferdinands, K., 2006. Assessing the threat posed by an invasive African grass *Urochloa mutica* Forssk to biodiversity conservation in the Mary River wetlands, Northern Territory. [PhD]. Darwin, Charles Darwin University.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24 (1), 38–49.

Fitzgerald, R.W., Lees, B.G., 1992. The application of neural networks to the floristic classification of remote sensing and GIS data in complex terrain. In: American Society of Photogrammetry and Remote Sensing (Eds.), XVII Congress American Society of Photogrammetry and Remote Sensing, Washington, DC.

Frank, T.D., 1988. Mapping dominant vegetation communities in the Colorado rock mountain front range with landsat thematic mapper and digital terrain data. Photogrammetric Engineering Remote Sensing 54, 1727–1734.

Franklin, J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain derived variables. Journal of Vegetation Science 9, 733–748.

Franklin, J., McCullough, P., Gray, C., 2000. Terrain variables used for predictive mapping of vegetation communities in Southern California. In: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis: Principles and Applications. Wiley, New York.

Gorman, J.T., Griffiths, A.D., Whitehead, P.J., 2006. An analysis of the use of plant products for commerce in remote Aboriginal communities of Northern Australia. Journal of Economic Botany 60 (4), 362–373.

Griffiths, A.D., Bowman, D., Cowie, I., Fensham, R., 2000. Vegetation of the Maningrida Region. Technical Report for Bawinanga Aboriginal Corporation. Darwin, Charles Darwin University.

Griffiths, A.D., Philips, A., Godjuwa, C., 2003. Harvest of *Bombax ceiba* for the Aboriginal arts industry, central Arnhem Land, Australia. Biological Conservation 113 (2), 295–305.

Griffiths, A.D., Schult, H.J., Gorman, J., 2005. Wild harvest of *Cycas arnhemica*: impact on survival, recruitment and growth in Arnhem Land, northern Australia. Australian Journal of Botany 53, 1–9.

Guisan, A., Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135 (2–1), 147–186.

Guisan, A., Theurillat, J.P., Kienast, F., 1998. Predicting the potential distribution of plant species in an alpine environment. Journal of Vegetation Science 9, 65–74.

Guisan, A., Weiss, S.B., Weiss, A.D., 1999. GLM versus CCA spatial modeling of plant species distribution. Plant Ecology 143, 107–122.

Hall, R., 2002. Use of forests in the Blyth and Liverpool River catchments. Arnhem Land NT. Darwin, Department of Agriculture Fisheries and Forestry.

Hill, M.O., 1991. Patterns of species distribution in Britain elucidated by canonical correspondence analysis. Journal of Biogeography 18, 247–255.

Hirzel, A.H., 2001. When GIS comes to life. Linking Landscape and Population Ecology for Large Population Management and Modelling: the Case of Ibex (*Capra Ibex*) in Switzerland [PhD]. Lausanne, The University of Lausanne.

Hirzel, A.H., Helfer, V., Metral, F., 2001. Assessing habitat suitability models with a virtual species. Ecological Modelling 145, 111–121.

Huntley, B., Berry, P.M., Cramer, W.P., McDonals, A.P., 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. Journal of Biogeography 22, 967–1001.

Huntley, B., Green, R.E., Collingham, Y.C., Hill, J.K., Willis, S.G., Bartlein, P.J., Cramer, W., Hagemeijer, W.J.M., Thomas, C.J., 2004. The performance of models relating species geographical distributions to climate is independent of trophic level. Ecology Letters 7, 417–426.

Kinfu, Y., Taylor, J., 2002. Estimating the Components of Indigenous Population Change: 1996–2001. Australian National University, Canberra.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Lowell, K., 1991. Utilizing discriminant function analysis with a geographical information system to model ecological succession spatially. International Journal of Geographical Information Systems 5, 175–191.

Lusted, L.B., 1968. Introduction to Medical Decision Making. Illinois, Springfield.

Mackey, B.G., 1993. Predicting the potential distribution of rain-forest structural characteristics. Journal of Vegetation Science 4, 43–54.

Manel, S., Ceri Williams, H., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Appllied Ecology 38, 921–931.

Martin, D.F., 1995. Money, Business and Culture: Issues for Aboriginal Economic Policy. Australian National University, Canberra.

Nix, H.A., 1986. A biogeographic analysis of Australian elapid snakes. In: Londmore, R. (Ed.), Atlas of Elapid Snakes of Australia. Australian Government Publishing Service, Canberra, pp. 4–15.

Pearson, D.M., 2002. The application of local measures of spatial autocorrelation for describing pattern in north Australian landscapes. Journal of Environmental Management 64, 85–95.

PWCNT (Parks and Wildlife Commission of the Northern Territory), 1997. A Management Program for Cycads in the Northern Territory of Australia. Government Printer of the Northern Territory, Darwin, pp. 1–29.

Skidmore, A.K., 1989. An expert system classifies eucalypt forest types using Landsat thematic mapper data and a digital terrain model. Photogrammetric Engineering and Remote Sensing 55, 1449–1464.

SRRATRC, 1998. Commercial Utilisation of Australian Native Wildlife. Report of the Senate Rural and Regional Affairs and Transport References Committee. Commonwealth of Australia, Canberra, June 1998.

Stockwell, D., Peters, D., 1999. The GARP modeling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Systems 13, 143–158.

Walker, P.A., Cocks, K.D., 1991. HABITAT: a procedure for modeling a disjoint environment envelope for a plant or animal species. Global Ecology and Biogeography 1, 108–118.

Walker, P.A., Moore, D.M., 1988. SIMPLE. An inductive modeling and mapping tool for spatially-oriented data. International Journal of Geographical Information Systems 2, 347–363.

Watkinson, A., Powell, J.I., 1997. The life history and population structure of *Cycas armstrongii* in monsoonal northern Australia. Oecologia 111, 341–349.

Whitehead, P.J., Bowman, D.M.J.S., Preece, N., Fraser, F., Cooke, P., 2003. Customary use of fire by indigenous peoples in northern Australia: its contemporary role in savanna management. International Journal of Wildland Fire 12, 415–425.

Whiting, M.G., 1963. Toxicity of cycads. Economic Botany 17, 271–302.

Wintle, B.A., Elith, J., Potts, J.M., 2005. Fauna habitat modeling: a review and case study in the lower hunter central coast region of NSW. Austral Ecology 30, 719–738.

Yee, T.W., 1991. Generalised additive models in plant ecology. Journal of Vegetation Science 2, 587–602.

Yibarbuk, D., Whitehead, P.J., Russell-Smith, J., Jackson, D., Godjuwa, C., Cooke, P., Choquenot, D., Bowman, D.M.J.S., 2001. Fire ecology and aboriginal land management in central Arnhem Land, northern Australia: a tradition of ecosystem management. Journal of Biogeography 28, 325–344.

ELSEVIER

Journal of Environmental Management 86 (2008) 114–120

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Dairy washwater treatment using a horizontal flow biofilm system

M. Rodgers*, D. de Paor, E. Clifford

Department of Civil Engineering, National University of Ireland, Galway, Ireland

Received 7 February 2006; received in revised form 16 November 2006; accepted 26 November 2006
Available online 22 January 2007

Abstract

In Ireland, dairy farmyard washwater commonly comprises farmyard run-off and dairy parlour washings. Land-spreading is the most widely used method for treating this wastewater. However, this method can be labour intensive and can cause, in some cases, the degradation of surface and ground waters, mainly due to nitrogen contamination.

In this study, a horizontal flow biofilm reactor (HFBR) with step-feed was constructed and tested in the laboratory, to remove organic carbon and nitrogen from a agricultural strength synthetic washwater (SWW). The HFBR had an average top plan surface area (TPSA) of $0.1002\,m^2$ and consisted of a stack of 45 polystyrene horizontal sheets—15 sheets embedded with 25 mm deep frustums above 30 sheets with 10 mm deep frustums. The frustums acted as miniature reservoirs. The sheets were alternately offset to allow the wastewater to flow horizontally along each sheet and vertically from sheet to sheet down through the reactor. Biofilms developed on the sheets and treated the wastewater.

During the 212-d study, the total hydraulic loading rate based on the TPSA of the sheets was $35\,l\,m^{-2}\,d^{-1}$. SWW was pumped for 10 min each hour, in a step feed arrangement at a rate of $23.33\,l\,m^{-2}\,d^{-1}$ on to the top sheet during Phases 1 and 2, and $11.67\,l\,m^{-2}\,d^{-1}$ onto Sheet 16 during Phase 1 (days 1–92) and onto Sheet 30 during Phase 2 (days 93–212). The substrate loading rate during Phases 1 and 2 was 94.8 g total chemical oxygen demand (COD) $m^{-2}\,d^{-1}$ and 10.5 g total nitrogen (TN) $m^{-2}\,d^{-1}$, based on the TPSA.

At steady state in Phase 2, the unit achieved excellent carbon removal of 99.7% 5-day biochemical oxygen demand (BOD_5) and 96.7% total COD, equivalent to TPSA removal rates of 67.5 g $BOD_5\,m^{-2}\,d^{-1}$ and 91.7 g COD $m^{-2}\,d^{-1}$. The nitrogen removal percentages were 98.3% total ammonium-nitrogen (NH_4-N_t) and 72.8% TN, which equated to TPSA removal rates of 4.8 g NH_4-$N_t\,m^{-2}\,d^{-1}$ and 7.6 g TN $m^{-2}\,d^{-1}$.

No sloughing of solids or clogging of media occurred during the study. The unit was simple to construct and operate, with little maintenance.

Keywords: Horizontal flow biofilm reactor; Step-feed; Organic carbon and nitrogen removal; Dairy washings; Agricultural wastewater

1. Introduction

With the introduction of the Nitrates Directive 91/676/EEC, which limits the amount of nitrogen (N) that can be applied to land to 170 kg $ha^{-1}\,year^{-1}$ (European Economic Community, 1991), increased attention has been given to the treatment of agricultural wastewaters in Ireland (Department of Agriculture and Food, 2005).

Dairy farmyard dirty water or washwater, which is not properly treated or managed, can be a major cause of point source pollution (Dunne et al., 2004). This can lead to a reduction in dissolved oxygen (DO) levels, and subsequent damage to aquatic organisms in receiving waters (Klees and Silverstein, 1992). Generally, dairy washwater is stored before being land-spread using spray irrigation. This method can achieve complete biodegradable chemical oxygen demand (COD) and phosphorus (P) removal on suitable well-drained soils (DeBusk et al., 1995). However, it may only achieve low N removal rates (Harter et al., 2002; Rodgers et al., 2003), which can increase the potential of eutrophication (Rodgers et al., 2005). Spray irrigation can be labour intensive and unattractive to farmers.

A process that replaces land-spreading and reduces the concentrations of COD and N so that washwaters can be

*Corresponding author. Tel.: +353 (0)91 492219; fax: +353 (0)91 494507.
E-mail address: michael.rodgers@nuigalway.ie (M. Rodgers).

0301-4797/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.027

safely discharged into soil or can be re-used for selected washing in the milking parlour, is preferable to the spray irrigation method.

Constructed wetlands and reed beds have been investigated as a possible solution (Sun et al., 1998; Knight et al., 2000; Pant and Reddy, 2003). They can have initial high capital costs. Alternative methods for the treatment of high strength wastewaters include a vertically moving biofilm system (Rodgers and Zhan, 2003), stratified sand filters (Rodgers et al., 2005) and an alternating pumped sequencing batch biofilm reactor (APSBBR) (Zhan et al., 2006).

The aim of this study was to develop a new horizontal flow biofilm reactor (HFBR) system, with step-feed, that has the potential to remove organic carbon and N from dairy washwaters, is robust, inexpensive, and easy to construct, maintain and operate.

2. Materials and methods

2.1. Design and construction of the horizontal flow biofilm reactor (HFBR)

The outer frame of the HFBR (Fig. 1), which was constructed from timber sheets, was 350 mm wide, 350 mm deep and 1900 mm high. The reactor comprised 45 horizontal polystyrene sheets (Terram DC), stacked one above the other and placed on 340 mm × 350 mm × 10 mm high shelves, made of lightweight, rigid polypropylene panels, which were supported by the frame. The system was kept in a temperature-controlled room at 11 °C.

The reactor comprised two sets of polystyrene sheets. The top set had 15 sheets, each of which measured 315 mm × 285 mm in plan with embedded 25 mm deep frustums, and was located above a set of 30 polystyrene sheets, each 340 mm × 310 mm in plan with 10 mm deep frustums. The frustums acted as miniature reservoirs. The sheets were placed horizontally and alternately offset so that the wastewater could flow along the sheets before falling to the sheet underneath and so on down through the reactor (Fig. 2). Three of the four sides of each sheet were turned up, acting as walls to direct the flow toward the open end of the sheet, where the bottoms of a row of frustums were cut off, to allow the wastewater drop from one sheet to the sheet below it.

The plan area of each 25 mm deep frustum sheet was $0.0898\,m^2$ and that of each 10 mm deep frustum sheet was $0.1054\,m^2$. The total surface area (TSA) of the sheets was $4.509\,m^2$ $(\{15 \times 0.0898\,m^2\} + \{30 \times 0.1054\,m^2\})$. The top plan surface area (TPSA) averaged $0.1002\,m^2$ (TSA/45). The frustums increased the available biofilm plan surface area and provided for solids accumulation. It was assumed that the majority of the COD removal would take place in the top sheets, and as a result, the 25 mm deep frustum sheets were used to accommodate the expected thick biofilm. It was thought the 10 mm deep frustum sheets would provide sufficient space for the thinner

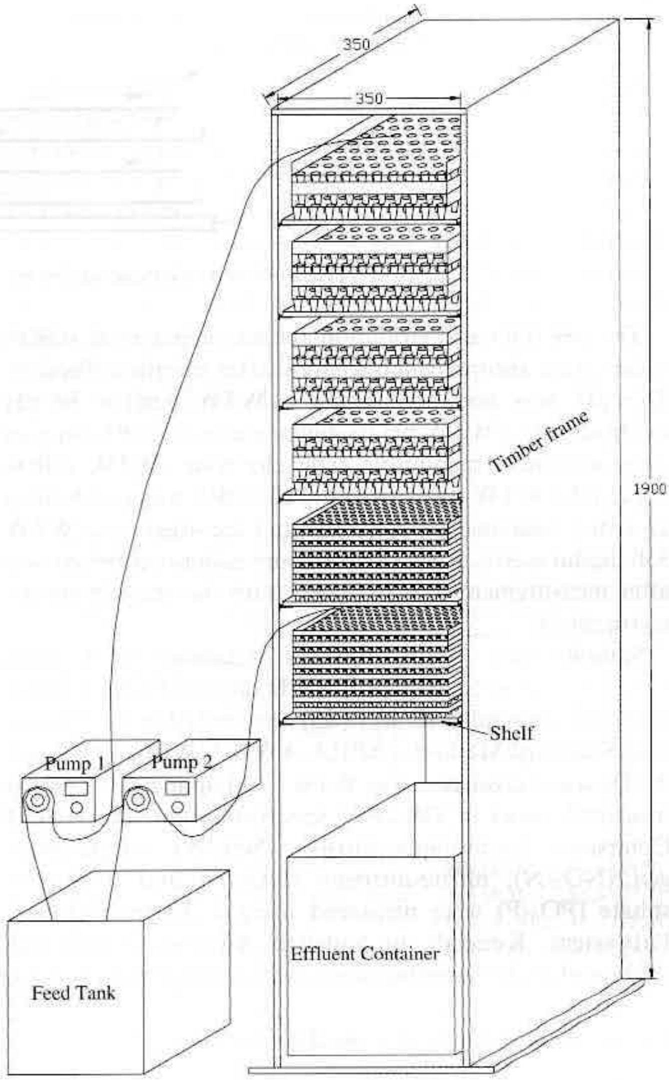


Fig. 1. Schematic of the agricultural horizontal flow biofilm reactor (HFBR). Dimensions in mm.

biofilm associated with nitrification. To minimise the degradation of total COD in the influent, the influent tank was cleaned and the agricultural strength synthetic washwater (SWW) was prepared daily, and a small circulation pump was placed at the bottom of the influent tank to gently mix the feed wastewater.

The following activities were also carried out daily: cleaning of the pump tubing, influent flow checks and emptying of the final effluent tank.

2.2. Sampling and analysis

Throughout the study, samples were taken from the influent feed tank, from individual sheets where the wastewater dropped from one sheet onto the next, and of the effluent from the last polystyrene sheet. On each test day, 10 samples were taken from the unit and tested for a range of parameters, to provide a profile of the processes occurring within the unit.

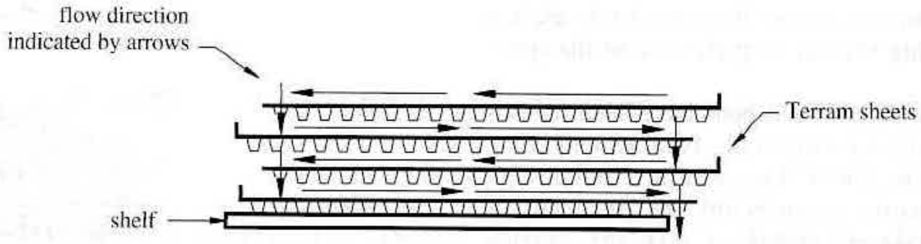


Fig. 2. Vertical schematic section through HFBR showing the direction of flow.

The pH, DO and oxidation reduction potential (ORP) values were recorded immediately after sample collection. The pH was measured using a WTW SenTix 50 pH electrode and a WTW pH 91 digital metre, the DO using an electrochemical membrane type electrode (WTW cellOx 325) and a WTW oxi 330 m, and the ORP using a Dolmen 23 redox combination electrode and recorded on a WTW 330 digital metre. All electrodes were calibrated before and after measurement in accordance with the manufacturers' instructions.

Samples were filtered through Whatman GF/C glass fibre filters (pore size 1.2 μm). COD, filtered COD (COD_f), and total suspended solids (TSS) were tested in accordance with Standard Methods (APHA-AWWA-WEF, 1995), and BOD_5 was measured using WTW OxiTop metres. TN was measured using a DR/2010 spectrophotometer (HACH Company). Ammonium-nitrogen (NH_4-N), nitrate-nitrogen (NO_3-N), nitrite-nitrogen (NO_2-N) and orthophosphate (PO_4-P) were measured using a Thermo Clinical Labsystem, Konelab 20 Nutrient Analyser, which was calibrated to the manufacturer's instructions before testing.

2.3. Hydraulic and organic loading rates

The influent was pumped onto the unit at two points, with a flow ratio of 2:1, for 10 min every hour using two Masterflex L/S Economy pumps fitted with Tygon Laboratory L/S 14 silicone tubing. An electronic timer controlled this intermittent loading. The influent SWW had similar constituents, but higher concentrations (Table 1), to that used by Ødegaard and Rusten (1980). The SWW, which had average total concentrations of 2709 mg COD l^{-1}, 1933 mg BOD l^{-1}, 333 mg TSS l^{-1}, 299 mg TN l^{-1} and 138 mg NH_4-N l^{-1}, was designed to have similar concentrations to dirty water from a 160 cow dairy farm, which had an average annual COD concentrations of 3061 mg l^{-1} and BOD of 1700 mg l^{-1} (Ryan, 1990) and to dairy parlour wastewater samples from a farm in Co. Cork, Ireland, which had average COD concentrations of 2920 mg l^{-1} and BOD of 2210 mg l^{-1} (Rodgers et al., 2003).

There were two loading phases. In the preliminary phase, Phase 1 (days 1–92), during the 10 min dosing period: 97 ml SWW was pumped onto Sheet 1, giving TPSA loading rates of 23.33 l m^{-2} d^{-1} and 63.2 g COD m^{-2} d^{-1}; and 48.5 ml SWW onto Sheet 16, giving TPSA loading rates of 11.67 l m^{-2} d^{-1} and 31.6 g COD m^{-2} d^{-1}. Initial results indicated that nitrification of the top feed could be

Table 1
Wastewater constituents and concentrations

Constituents	Concentration ($mg\,l^{-1}$)
Glucose	1600
Yeast	240
Dried milk powder	960
Urea	240
NH_4Cl	480
$Na_2PO_4 \cdot 12H_2O$	800
$KHCO_3$	400
$NaHCO_3$	1040
$MgSO_4 \cdot 7H_2O$	400
$FeSO_4 \cdot 7H_2O$	16
$MnSO_4 \cdot 1H_2O$	16
$CaCl_2 \cdot H_2O$	24

increased by changing the step-feed location from Sheet 16 to a lower sheet. The new step-feed location in Phase 2 operated from days 93 to 212, with the same loading rates as Phase 1 but with the location of the step-feed (11.67 l m^{-2} d^{-1}) changed from Sheet 16 to Sheet 30. Steady state behaviour for Phase 2 was reached 10 days after changing the step feed location (day 103).

Seeding of the unit took place on day 1 using biosolids from a nitrifying wastewater treatment plant, at Moycullen, Co. Galway. The application of biosolids to the unit occurred four times, at 2 h intervals. Each application volume consisted of 1 l of biosolids poured on to Sheet 1 and a half-litre poured on to Sheet 16.

2.4. Biofilm characteristics

Biofilm solids concentrations were measured twice during the study, once mid-way through the Phase 2 steady-state period on day 139 and again just before the end of the study on day 205. The volatile solids (VS) and total solids (TS) of the biofilm samples were tested in accordance with Standard Methods (APHA-AWWA-WEF, 1995).

3. Phase 2 results and discussion

3.1. COD and BOD_5 removal

During steady state, the average total influent COD of 2709.2 mg l^{-1} was reduced by 96.7% to an average total

COD in the effluent of $89.7\,mg\,l^{-1}$. The average influent COD_f was $2229.1\,mg\,l^{-1}$ and was reduced by 96.6% to an average effluent value of $74.9\,mg\,l^{-1}$. Most COD_f (91%) removal took place in the top 7 sheets of the unit (Fig. 3), where the TPSA and average TSA removal rates were 68.6 and $9.8\,g\,COD_f\,m^{-2}\,d^{-1}$, respectively.

The increase in COD_f at Sheet 30 resulted from the introduction of the step-feed SWW combining with the effluent from Sheet 29 to give a concentration considerably lower than the concentration of the step-feed wastewater. The unit quickly (from Sheets 30 to 38) reduced the COD_f to the final effluent values of $74.9\,mg\,l^{-1}$. The overall TPSA removal rate was $75.4\,g\ COD_f\,m^{-2}\,d^{-1}$and the TSA removal rate was $1.7\,g\ COD_f\,m^{-2}\,d^{-1}$.

The average influent BOD_5 of $1933.3\,mg\,l^{-1}$ was reduced by 99.7% to an average effluent BOD_5 of $5.4\,mg\,l^{-1}$. With a similar wastewater, Rodgers et al. (2005) achieved 99% BOD removal using a 900 mm high stratified sand filter and TSPA hydraulic loading rate of $20\,l\,m^{-2}\,d^{-1}$, which was lower than the $35\,l\,m^{-2}\,d^{-1}$ used in this HFBR study. Schaafsma et al. (2000) recorded 96.9% BOD_5 removal from dairy farm wastewater with a BOD_5 concentration of $1914\,mg\,l^{-1}$, in wetland cells that had a capacity of $166\,m^3$. The HFBR had an average TPSA removal rate of $67.5\,g\ BOD_5\,m^{-2}\,d^{-1}$ and a TSA removal rate of $1.5\,g\ BOD_5\,m^{-2}\,d^{-1}$. This compares with Cronk (1996), whose review on wetlands found that BOD_5 removal rates varied from 3.7 to $24.3\,g\,m^{-2}\,d^{-1}$ and Geary and Moore (1999) who reported a $5.6\,g\ BOD_5\,m^{-2}\,d^{-1}$ removal rate from dairy parlour waters using a treatment wetland. Due to the stacking arrangement and flow contact area, the HFBR can have a smaller footprint than these systems.

3.2. Nutrient removal

The average total NH_4-N of $138.4\,mg\,l^{-1}$ in the influent was reduced by 98.3% to an average total NH_4-N of $2.4\,mg\,l^{-1}$ in the effluent (Table 2). The average filtered influent NH_4-N (NH_4-N_f) of $117.5\,mg\,l^{-1}$ was reduced by 98.1% to an average NH_4-N_f in the effluent of $2.2\,mg\,l^{-1}$ (Fig. 4). At Sheet 7, most of the carbon had been removed, and the majority of nitrification occurred between Sheets 7 and 15, where NH_4-N decreased and NO_3-N levels increased. Further nitrification occurred between Sheets 15 and 29, with an average of $18.5\,mg\,l^{-1}$ NH_4-N_f on Sheet 29, before the step-feed. There was an increase at Sheet 30

Table 2
Average sheet effluent concentrations of NH_4-N_f, TN_f, NO_3-N_f and PO_4-P_f during Phase 2 (days 103–212), standard deviation shown in ( )

Sheet number	Average sheet effluent concentrations ($mg\,l^{-1}$)			
	NH4-N_f	TN_f	NO_3-N_f	PO_4-P_f
0	117.5 (5.0)	235.3 (35.7)	0.2 (0.7)	67.4 (25.5)
1	132.8 (24.8)	188.2 (34.9)	0.3 (2.2)	60.0 (20.7)
7	115.9 (23.4)	155.5 (28.5)	3.2 (2.3)	36.5 (12.9)
11	98.6 (17.9)	134.5 (24.3)	10.1 (2.0)	39.4 (13.1)
15	71.8 (19.2)	125.0 (23.4)	30.3 (14.1)	42.9 (15.8)
24	50.3 (14.8)	118.4 (22.2)	53.9 (20.7)	39.9 (13.3)
29	18.5 (19.4)	100.9 (22.5)	70.9 (23.8)	37.3 (11.9)
30	39.0 (19.5)	96.6 (26.0)	36.4 (8.9)	55.5 (24.3)
38	11.9 (6.83)	68.3 (20.6)	40.7 (11.4)	44.2 (16.3)
45	2.2 (2.5)	60.7 (16.1)	57.9 (9.9)	40.3 (14.9)
% Removal	98.1	74.2	—	40.2

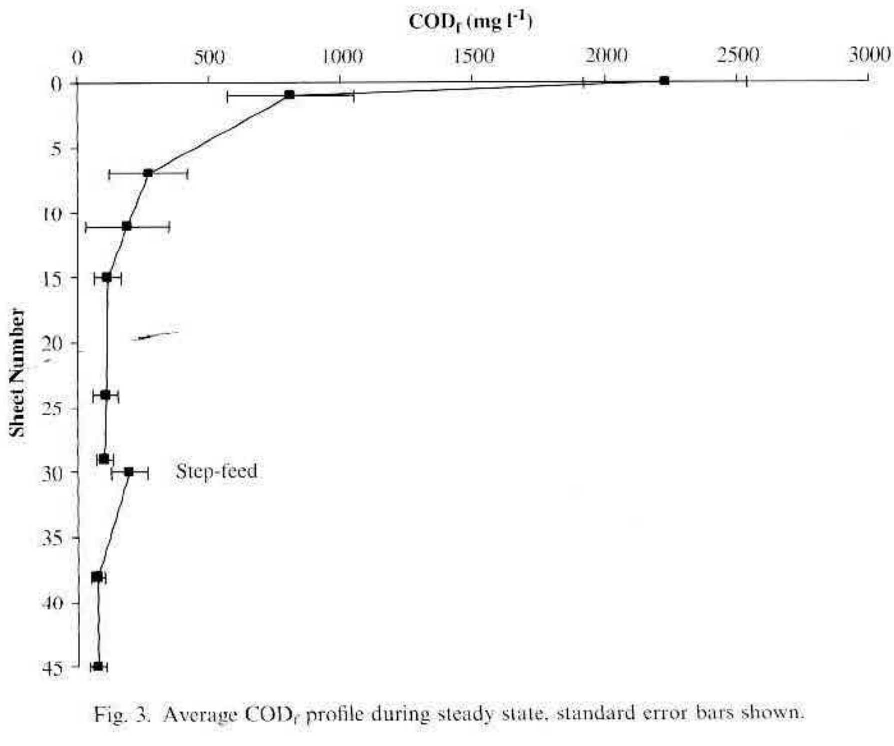


Fig. 3. Average COD_f profile during steady state, standard error bars shown.

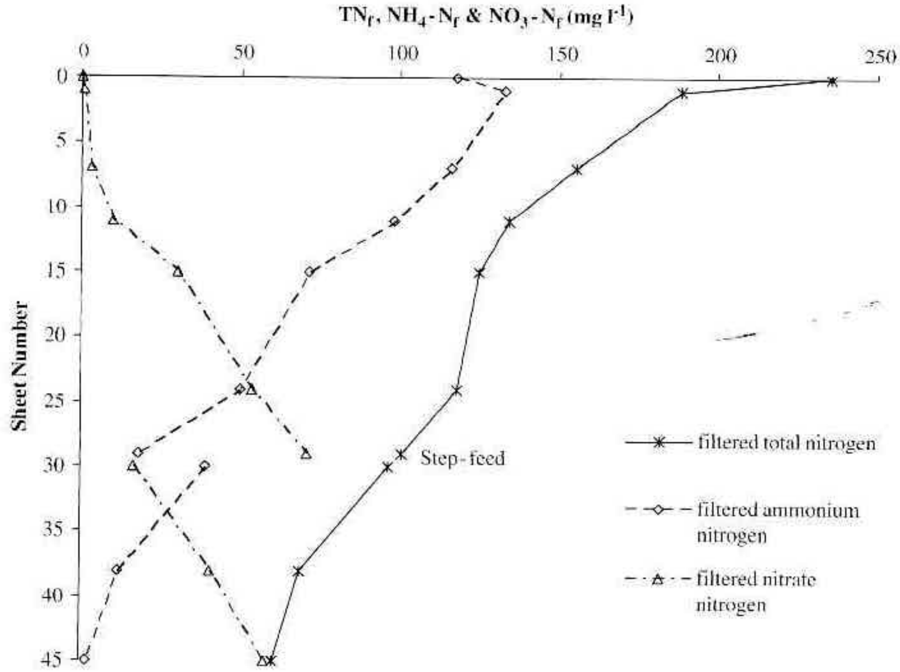


Fig. 4. Average TN_f, NH_4-N_f and NO_3-N_f profiles during steady state.

due to the step-feed. Further nitrification occurred in the bottom 7 sheets of the unit. The TPSA NH_4-N_f removal rate was 4.0 g NH_4-N m^{-2} d^{-1} and from Sheets 7 to 29 the removal rate was 3.4 g NH_4-N m^{-2} d^{-1}. With a similar wastewater and a loading rate of 20 l m^{-2} d^{-1}, Rodgers et al. (2005) achieved 88% removal in NH_4-N using a stratified sand filter.

NO_3-N increased from Sheets 7 to 29, due to nitrification (Table 2). Denitrification occurred after the step-feed at Sheet 30 (Fig. 4). NO_3-N increased to an average of 69.7 mg l^{-1} in the total effluent and 57.9 mg l^{-1} in the filtered effluent. Healy et al. (2005) found that 60 mg l^{-1} NO_3-N could be reduced by 97% when it flowed horizontally through a tank of wood-chippings and sand.

Filtered NO_2-N (NO_2-N_f) increased from 0.7 mg l^{-1} in the influent to 16.4 mg l^{-1} on Sheet 29. At the step-feed, the NO_2-N_f was 11.0 mg l^{-1} and decreased to 1.3 mg l^{-1} in the effluent. The concentration of total NO_2-N was 2.1 mg l^{-1} in the effluent.

Overall, the TN in the unfiltered samples was reduced by 72.8% from 299.0 mg l^{-1} in the influent to 81.3 mg l^{-1} in the effluent. The TN in the filtered influent (TN_f), with an average of 235.3 mg l^{-1} was reduced by 74.2% to 60.7 mg l^{-1} (Table 2). Tanner et al. (1995) observed that TN removal was 48% when a planted wetland was used to treat dairy farm wastewater which had a nitrogen loading rate of 2.7 g TN m^{-2} d^{-1}. The TSPA loading rate for the HFBR was 10.5 g TN m^{-2} d^{-1}.

The total PO_4-P was reduced by 39.0%, from 77.2 mg l^{-1} in the influent to 47.1 mg l^{-1} in the effluent. The filtered influent PO_4-P (PO_4-P_f) with an average of 67.4 mg l^{-1} was reduced by 40.2% to 40.3 mg l^{-1} in the effluent.

3.3. DO, ORP and pH

Throughout the study pH, DO and ORP values were recorded. For carbonaceous removal, the optimal performance occurs near a neutral pH (Metcalf and Eddy, 2003). During steady state, the pH increased from 7 to 8.5 in the top 7 sheets where carbon removal was taking place. Nitrification took place between Sheets 7 and 29, where the pH decreased from 8.5 to 7.6. Nitrification reactions depress pH and optimal nitrification rates occur between pH 7 and 8 (Metcalf and Eddy, 2003). The pH dropped to 7.4 on Sheet 30, where the step-feed was introduced, and subsequently increased to 8.2 in the effluent.

Similar to the pH profile, the DO and ORP readings indicated the location within the reactor where different processes were occurring. During carbon removal on the top sheets of the unit, the DO was close to zero and the ORP had an average negative value of $-126.4 E_H$ mV. From Sheets 7 to 29 where nitrification occurred, the ORP varied from 208.5 to 226.8 E_H mV and the DO from 5.05 to 5.85 mg l^{-1}. At Sheet 30 which was the location of the step-feed and where denitrification occurred, the DO decreased to 1.5 mg l^{-1} and the ORP had negative readings (Fig. 5). Both parameters subsequently increased in the bottom sheets of the unit to 7.6 mg DO l^{-1} and 360.7 E_H mV ORP in the effluent.

3.4. Steady-state TSS and biofilm measurements

The overall reduction in TSS was 86.2%, from 333.3 mg l^{-1} in the influent to 46.0 mg l^{-1} in the effluent. The average TS and VS concentrations of the biofilm.

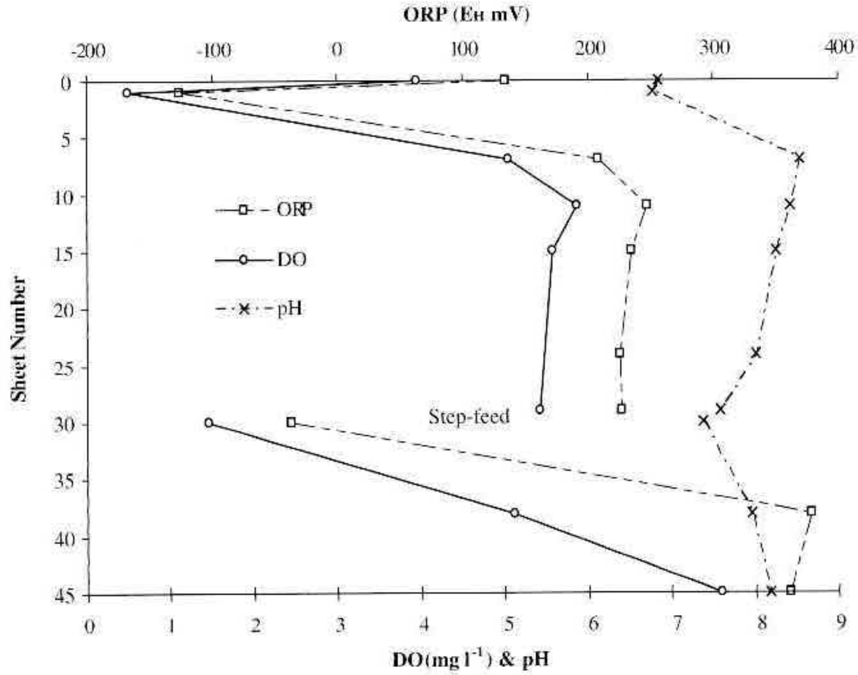


Fig. 5. Average pH, DO and ORP profiles during steady state in Phase 2.

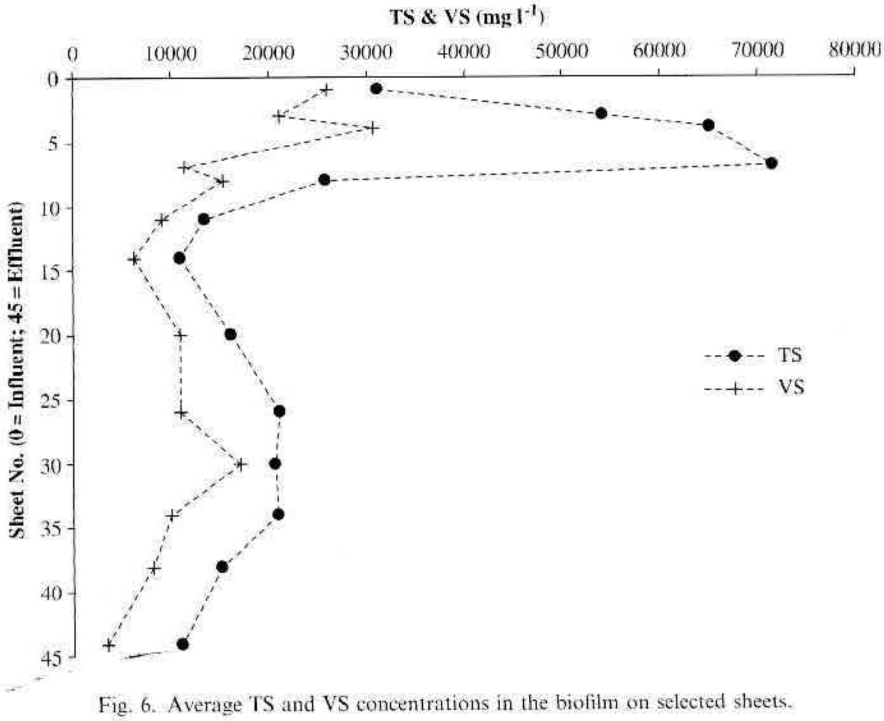


Fig. 6. Average TS and VS concentrations in the biofilm on selected sheets.

measured on the two biofilm sampling days, followed a similar profile, down through the unit (Fig. 6). The highest solids concentrations were recorded on Sheets 1–7, where most carbon removal occurred and where TS and VS average concentrations were 49,554 and 20,911 mg l^{-1}, respectively. Lower concentrations were recorded in the nitrification regions of the reactor. The average TS and VS concentration values from Sheets 8 to 29 were 17,514 and 10,532 mg l^{-1}, respectively. After the step-feed at Sheet 30, the biofilm concentration increased with an average TS concentration of 20,809 mg l^{-1} and VS of 13,573 mg l^{-1} between Sheets 30 and 36. It subsequently decreased in the remaining sheets to average TS values of 13,131 mg l^{-1} and VS of 5777 mg l^{-1} on Sheets 37–45. No leaching of solids occurred during the study; however, clarification may be necessary in field conditions.

4. Conclusions

During the study with TPSA loading rates of $35\,l\,m^{-2}\,d^{-1}$, $94.5\,g\ COD\,m^{-2}\,d^{-1}$ and $10.5\,g\ TN\,m^{-2}\,d^{-1}$, the 45-sheet, step-feed HFBR unit performed excellently in removing BOD_5 (99.7%), total COD (96.6%), COD_f (96.7%), total NH_4-N (98.3%), NH_4-N_f (98.1%), TN (72.8%), TN_f (74.2%), and TSS (86.2%). PO_4-P removal percentages were 39.0% in the unfiltered and 40.2% in the filtered samples. The results compare favourably with other proposed systems for dealing with high strength dairy wastewater, such as wetlands, reed-beds and stratified sand filters. The use of stacked horizontal sheets reduces the footprint required in comparison with many other systems.

Based on the laboratory results, it is estimated that a 100-cow dairy farm, would require a treatment unit measuring $12\,m \times 12\,m$ in plan and 45 sheets in depth, based on a wastewater flow of $50\,l\,cow^{-1}\,day^{-1}$ to give a similar performance. It is appreciated that laboratory and field conditions are different. The system would be preceded by a settlement tank, which could be an existing holding tank, and followed by a clarifier. The HFBR could be manufactured in units that would treat a certain volume of washwater per day, allowing farm owners to buy the number of units they required. Since there is little hydrostatic pressure associated with the system, a simple light plastic tank formed in a mould could be used to contain the HFBR sheet stacks. The sheets could be specifically manufactured for the HFBR system with 3 sides turned up, and simply stacked in alternate directions on one another. It would also be possible to assemble a tower of sheet stacks with a suitable wastewater distribution system to reduce the footprint further.

Using the HFBR system, the current method of land-spreading might be eliminated or greatly reduced. The treated effluent could be used to wash farmyards or it could be passed through soil to further reduce the N and P before entering a water body. The unit was easy to construct and maintain, and there was no sloughing of solids or clogging of the reactor.

Acknowledgements

The authors gratefully acknowledge the grant from Enterprise Ireland (Commercialisation Fund) for this project. Sincere thanks are given to Ms. Mary O'Brien and Mr. Liwen Xiao, Department of Civil Engineering, NUI, Galway.

References

APHA-AWWA-WEF, 1995. Standard Methods for Examination of Water and Wastewater, 19th ed. American Public Health Association, Washington.

Cronk, J.K., 1996. Constructed wetlands to treat wastewater from dairy and swine operations: a review. Agriculture, Ecosystems and Environment 58, 97–114.

DeBusk, T.A., Peterson, J.E., Reday, K.R., 1995. Use of aquatic and terrestrial plants for removing phosphorus from dairy wastewaters. Ecological Engineering 5, 371–390.

Department of Agriculture and Food, 2005. ⟨www.agriculture.gov.ie/publicat/factsheet/Nov05.pdf⟩.

Dunne, E.J., Culleton, N., O'Donovan, G., Harrington, R., Olsen, A.E., 2004. An integrated constructed wetland to treat contaminants and nutrients from dairy farmyard dirty water. Ecological Engineering 24, 221–234.

EEC, 1991. The Protection of Waters Against Pollution Caused by Nitrates from Agricultural Sources. EEC Directive 91/676/EEC. Official Journal No. L375, Brussels, Belgium.

Geary, P.M., Moore, J.A., 1999. Suitability of a treatment wetland for dairy wastewaters. Water Science and Technology 40 (3), 179–185.

Harter, T., Davis, H., Mathews, M.C., Meyer, R.D., 2002. Shallow groundwater quality on dairy farms with irrigated forage crops. Journal of Contaminant Hydrology 55, 287–315.

Healy, M.G., Rodgers, M., Mulqueen, J., 2005. Denitrification of a nitrate-rich wastewater using various wood-based media materials. Journal of Environmental Science and Health Part A 41 (5).

Klees, R., Silverstein, J., 1992. Improved biological nitrification using recirculation in rotating biological contactors. Water Science and Technology 26 (3–4), 545–553.

Knight, R.L., Payne Jr., W.E., Borer, R.E., Clarke Jr., R.A., Pries, J.H., 2000. Constructed wetlands for livestock wastewater management. Ecological Engineering 15, 41–55.

Metcalf, Eddy, 2003. Wastewater Engineering, Treatment and Reuse, forth ed. McGraw-Hill, New York.

Odegaard, H., Rusten, R., 1980. Nitrogen removal in rotating biological contactors without the use of an external carbon source. In: Proceedings of the First National Symposium/Workshop on Rotating Biological Contactor Technology, vol. 2, Champion, pp. 1301–317.

Pant, H.K., Reddy, K.R., 2003. Potential internal loading of phosphorus in a wetland constructed in agricultural land. Water Research 37, 965–972.

Rodgers, M., Gibbons, P., Mulqueen, J., 2003. Nitrate leaching on a sandy loam soil under different dairy wastewater applications. In: Proceedings of the IWA Seventh International Specialised Conference on Diffuse Pollution and Basin Management, vol. 7, Dublin, 7A Groundwater, pp. 14–19.

Rodgers, M., Zhan, X.M., 2003. Biological nitrogen removal using a vertically moving biofilm system. Bioresource Technology 93 (3), 313–319.

Rodgers, M., Healy, M.G., Mulqueen, J., 2005. Organic carbon removal and nitrification of high strength wastewaters using stratified sand filters. Water Research 39, 3279–3286.

Ryan, M., 1990. Properties of Different Grades of Soiled Water and Strategies for Safe Disposal. Environmental Impact of Landspreading of Wastes, Wexford, Ireland, pp. 28–43.

Schaafsma, J.A., Baldwin, A.H., Streb, C.A., 2000. An evaluation of a constructed wetland to treat wastewater from a dairy farm in Maryland, USA. Ecological Engineering 14, 199–206.

Sun, G., Gray, K.R., Biddlestone, A.J., 1998. Treatment of agricultural wastewater in downflow reed beds: experimental trials and mathematical models. Journal of Agricultural Engineering Research 69, 63–71.

Tanner, C.C., Clayton, J.S., Upsdell, M.P., 1995. Effect of loading rate and planting on treatment of dairy farm wastewaters in constructed wetlands—II. Removal of nitrogen and phosphorus. Water Research 29, 27–34.

Zhan, X.M., Rodgers, M., O' Reilly, E., 2006. Biofilm growth and characteristics in an alternating pumped sequencing batch biofilm reactor (APSBBR). Water Research 40, 1–2.

Journal of Environmental Management 86 (2008) 121–131

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Kinetic of carbonaceous substrate in an upflow anaerobic sludge sludge blanket (UASB) reactor treating 2,4 dichlorophenol (2,4 DCP)

Delia Teresa Sponza*, Ayşen Uluköy

Dokuz Eylül University, Engineering Faculty, Environmental Engineering Department, Buca Kaynaklar Campus Tınaztepe 35160, İzmir

Received 27 December 2005; received in revised form 23 November 2006; accepted 28 November 2006
Available online 24 January 2007

Abstract

The performance of an upflow anaerobic sludge blanket (UASB) reactor treating 2,4 dichlorophenol (2,4 DCP) was evaluated at different hydraulic retention times (HRTs) using synthetic wastewater in order to obtain the growth substrate (glucose-COD) and 2,4 DCP removal kinetics. Treatment efficiencies of the UASB reactor were investigated at different hydraulic retention times (2–20 h) corresponding to a food to mass (F/M) ratio of 1.2–1.92 g-COD g^{-1} VSS day^{-1}. A total of 65–83% COD removal efficiencies were obtained at HRTs of 2–20 h. In all, 83% and 99% 2,4 DCP removals were achieved at the same HRTs in the UASB reactor. Conventional Monod, Grau Second-order and Modified Stover–Kincannon models were applied to determine the substrate removal kinetics of the UASB reactor. The experimental data obtained from the kinetic models showed that the Monod kinetic model is more appropriate for correlating the substrate removals compared to the other models for the UASB reactor. The maximum specific substrate utilization rate (k) (mg-COD mg^{-1} SS day^{-1}), half-velocity concentration (K_s) (mg COD l^{-1}), growth yield coefficient (Y) (mg mg^{-1}) and bacterial decay coefficient (b) (day^{-1}) were 0.954 mg-COD mg^{-1} SS day^{-1}, 560.29 mg-COD l^{-1}, 0.78 mg-SS g^{-1}-COD, 0.093 day^{-1} in the Conventional Monod kinetic model. The second-order kinetic coefficient (k_2) was calculated as 0.26 day^{-1} in the Grau reaction kinetic model. The maximum COD removal rate constant (U_{max}) and saturation value (K_B) were calculated as 7.502 mg COD l^{-1} day^{-1} and 34.56 mg l^{-1} day^{-1} in the Modified Stover–Kincannon Model. The (k)(mg-2,4 DCP mg^{-1} SS day^{-1}), (K_s) (mg 2,4 DCP l^{-1}), (Y) (mg SS mg^{-1} 2,4 DCP) and (k_d) (day^{-1}) were 0.0041 mg-2,4 DCP mg^{-1} SS day^{-1}, 2.06 mg-COD l^{-1}, 0.0017 mg-SS mg^{-1} 2,4 DCP and 3.1×10^{-5} day^{-1} in the Conventional Monod kinetic model for 2,4 DCP degradation. The second-order kinetic coefficient (k_2) was calculated as 0.30 day^{-1} in the Grau reaction kinetic model. The maximum 2,4 DCP removal rate constant (U_{max}) and saturation value (K_B) were calculated as 0.01 mg COD l^{-1} day^{-1} and 9.8×10^{-3} mg l^{-1} day^{-1} in the Modified Stover–Kincannon model.

Keywords: 2,4 dichlorophenol (2,4 DCP); UASB; Monod; Grau second order; Modified Stover–Kincannon model

1. Introduction

Chlorinated phenols have been placed in the list of priority pollutants of the United States Environmental Protection Agency due to their toxicity, carcinogenicity and persistence in the environment (Notice of the first priority list of hazardous substances that will be the subject of toxicological profiles, 1987). The treatability of chlorophenols has been studied in both anaerobic (Boyd et al., 1983; Boyd and Shelton, 1984; Ning et al., 1997; Kim et al., 1998; Nitayapat and Watson-Craik, 2002) and aerobic reactors (Puhakka and Jarvinen, 1992; Armmenante et al., 1999). 2,4 DCP is one non-volatile toxic organic compound that is used in large quantities in industries for production of wood preservatives, pesticides, herbicides, and plastic materials (Ning et al., 1997; Krumme and Boyd, 1988). Furthermore 2,4 DCP is released as an anaerobic degradation product (Ning et al., 1997).

Mathematical models are used in anaerobic processes to determine the importance of relationships between the design data and experimental results. They are also used to control and predict the plant design performance for optimizing the plant design. Gu and Knaebel studied the degradation kinetics of pentachlorophenol (PCP) (Gu et al., 1995; Perkins et al., 1994). Ning et al. (1997)

*Corresponding author. Tel.: +90 232 4127119; fax: +90 232 4531143.
E-mail address: delya.sponza@deu.edu.tr (D.T. Sponza).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.030

Nomenclature

U	specific substrate utilization rate ($mg\,mg^{-1}\,day^{-1}$)
X	concentration of microorganisms ($mg\,l^{-1}$)
k	maximum specific substrate utilization rate (mg-COD mg^{-1} SS.day^{-1})
Q	influent flow rate ($l\,day^{-1}$)
a	constant in Grau order kinetic model (dimensionless)
S_e	effluent COD or 2,4 DCP concentrations ($mg\,l^{-1}$)
b	constant in Grau order kinetic model (dimensionless)
S_o	influent COD or 2,4 DCP concentrations ($mg\,l^{-1}$)
μ	specific growth rate ($mg\,SS\,mg^{-1}\,COD\,day^{-1}$)
μ_{max}	maximum specific growth rate ($mg\,SS\,mg^{-1}\,COD\,day^{-1}$)
V	reactor volume (l)
SRT (Θ_c)	sludge retention time (day)
Y	growth yield coefficient (g SS g-COD^{-1})
k_d	decay coefficient (day^{-1})
X_e, X	effluent microbial concentration ($mg\,l^{-1}$)
K_s	half velocity coefficient (mg-COD l^{-1})
HRT(Θ_H)(Θ_h)	hydraulic retention time (h)
E	COD, 2,4 DCP removal efficienciencies (%)
k_2	second-order substrate removal rate constant ($l\,day^{-1}$)
U_{max}, R_{max}	maximum substrate removal rate constant (mg-COD mg^{-1} SS day^{-1})
K_B	saturation value constant ($mg\,l^{-1}\,day^{-1}$)

investigated the degradation of 2,4 DCP using a modified Haldane inhibition kinetic model to estimate the performance of anaerobic sludge granules in an anaerobic digester (Ning et al., 1997). Beitrame et al. showed that the 2,4 DCP was degraded according to the Monod kinetic model (Ning et al., 1997; Wik et al., 2000; Kesavan and Law, 2005; Marcos et al., 2004). Monod-type kinetic models have been widely used to describe the process kinetics of anaerobic digesters (Beitrame et al., 1982). Although there has been some success in applying Monod-type kinetics to the anaerobic process, some research workers found it difficult to apply them to their systems. For instance, Grady et al. (1972) have shown that the effluent substrate concentration, expressed as COD, was not dependent on the substrate concentration entering the reactor. In the equation proposed by Stover–Kincannon (Beitrame et al., 1982; Weiland and Manur, 1997; Yu et al., 1998; Hammer and Sidney, 2002), the specific growth rate was considered as a function of the growth limiting nutrient, by using an empirical constant, which was related to microbial concentrations. Chen and Hashimoto (1980) developed a kinetic model for substrate utilization and methane production and suggested that this kinetic model would be more suitable than the Monod-type kinetic models to predict the performance of a digester. Grau found that the anaerobic reactors could be modeled using his kinetic model (Sanchez et al., 2001; Grau et al., 1975).

Determination of kinetic constants of a UASB reactor is useful to be able to describe and predict the performance of the system. Recent literature surveys showed that the kinetics of upflow anaerobic sludge blanket (UASB) reactors fed with 2,4 DCP wastewater have not been investigated. Therefore in this study the treatment performance of a UASB reactor treating 2,4 DCP was evaluated at decreasing hydraulic retention times and the kinetic constants of the reactor were evaluated according to the experimental results. In order to determine the most suitable biokinetic model for the UASB reactor treating 2,4 DCP, some kinetic models, such as the Conventional Monod, Second-order kinetic, and Modified Stover–Kincannon models were applied to the experimental data. The models and the kinetic constants were interpreted.

2. Materials and methods

2.1. Reactor model

In this study, a laboratory scale reactor was designed. A continuously fed stainless steel UASB was used for the experimentation. The schematic diagram of the UASB reactor is given in Fig. 1 and it was operated at 37 °C. The UASB reactor employed in this study had 2.5 l of effective volume with an internal diameter of 6 cm and a height of 100 cm.

2.2. Seed

Partially granulated anaerobic sludge was used as seed in the UASB reactor and was taken from the methanogenic reactor of the Pakmaya Yeast Baker Factory in İzmir.

2.3. Composition of synthetic wastewater

The anaerobic UASB reactor was inoculated with settled partially granulated anaerobic sludge and fed with synthetic wastewater containing glucose (3000 $mg\,l^{1}$) and 10 $mg\,l^{1}$ 2,4 DCP for 88 days. On day 88, the OLRs were increased by decreasing the HRTs. The feed solution of the UASB reactor consisted of glucose as the primary carbon source and 2,4 DCP dissolved in Vanderbilt mineral medium (Speece, 1996). Anaerobic conditions were maintained by adding 0.5 $mg\,l^{1}$ of sodium thioglycollate into the feed. The desired alkalinity and neutral pH were obtained by addition of 6000 $mg\,l^{-1}$ $NaHCO_3$ into the feed media.

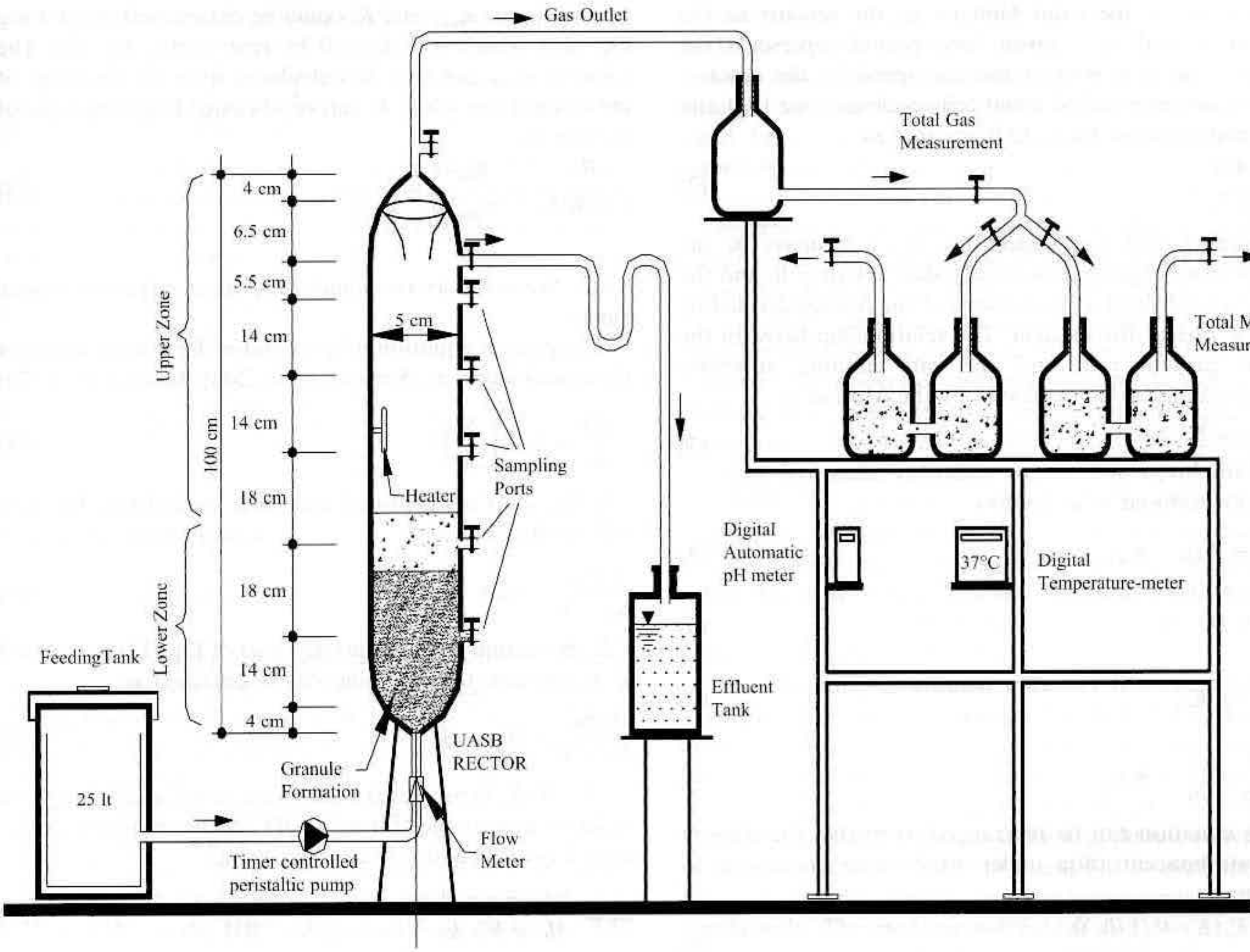


Fig. 1. Schematic configuration of UASB reactor.

2.4. *Operating conditions*

The lab scale UASB reactor was operated with glucose and 2,4 DCP during a 96-day period in order to investigate the process kinetics at different hydraulic retention times. HRT was decreased from 20 to 2 h by increasing the organic loading rates (OLR) from 6 to 34 g COD l^{-1} day^{-1} in the UASB reactor for 96 days of operation. The VSS concentrations in the UASB reactor varied between 5 and 17 g l^{-1}. The corresponding 2,4 DCP loading rates varied between 1.2×10^{-2} and 0.232 g l^{-1} day^{-1}. The F/M ratios were between 1.2 and 1.92 g COD g VSS^{-1} day^{-1} through the operation of the UASB reactor.

2.5. *Analytical procedures*

Suspended solids in granulated sludge and in activated sludge were regularly measured 2–3 times per week with a filtration technique using membrane filters with a pore size of 0.45 µm. Soluble COD in influent and effluent samples was measured 3 times per week using the closed Reflux colorimetric method (APHA, 1975). Cumulative gas and methane gas were measured daily by the liquid displacement method (Razo-Flores et al., 1997; Beydilli et al., 1998). Volatile fatty acids (VFA) were measured 3 times per week by a two-stage titration method developed by Anderson and Yang (1992). The temperature was measured daily automatically with a heater held in the UASB reactor. The pH of the effluent was measured daily using a pH meter. Samples were centrifuged at 14 000 rpm for 25 min. 2,4 DCP was analyzed 2–3 times per week with a spectrophotometer at 500 nm wavelength (APHA, 1975).

2.6. *Anaerobic kinetic models*

2.6.1. *Conventional monod kinetic*

For a UASB reactor without biomass recycle, the rate of change of biomass in the system can be expressed as (Ning et al., 1997; Wik et al., 2000; Kesavan and Law, 2005; Marcos et al., 2004):

$$\frac{dX}{dt} = \frac{Q}{V}X_0 - \frac{Q}{V}X_E + \mu X - K_d X. \tag{1}$$

The ratio of the total biomass in the reactor to the biomass wasted in a given time period represents the average time that microorganisms spend in the reactor. This parameter is called mean cell-residence time (θ_C) and is calculated below for UASB reactors as

$$\theta_C = \frac{VX}{QX_E}. \tag{2}$$

Assuming that the concentration of biomass in the influent can be ignored, at steady-state $dX/dt = 0$, and the HRT(θ_H) is defined as the volume of the reactor divided by the flow rate of the influent. The relationship between the specific growth rate and the rate limiting substrate concentration can be expressed by the Monod as

$$\mu = \frac{\mu_{max} S}{K_s + S}. \tag{3}$$

Eq. (1) reduced to as follows:

$$\frac{Q}{V} X_E = X(\mu - K_d), \tag{4}$$

$$\frac{QX_E}{VX} + K_d = \mu, \tag{5}$$

$$\mu = \frac{1}{\theta_C} + K_d, \tag{6}$$

$$\frac{\mu_{max} S}{K_s + S} = \frac{1}{\theta_C} + K_d. \tag{7}$$

This equation can be rearranged to predict the effluent substrate concentration under steady-state conditions as follows:

$$S = \frac{K_s(K_d + (1/\theta_C))}{\mu_{max} - K_d - (1/\theta_C)}. \tag{8}$$

The rate of change in substrate concentration in the system could be expressed as

$$-\frac{dS}{dt} = \frac{Q}{V} S_0 - \frac{Q}{V} S - \frac{\mu X}{Y}. \tag{9}$$

Under steady-state conditions, the rate of change in substrate concentration ($-dS/dt$) is negligible and by a similar technique to that used for the substrate concentration, the above equation can be reduced to Eq. (10) by substituting Eq. (6)

$$\frac{(S_0 - S)}{\theta_H} = \frac{X}{Y}\left(\frac{1}{\theta_C} + K_d\right). \tag{10}$$

The above equation can then be rearranged to predict the effluent biomass concentration under steady-state conditions as follows:

$$X = \frac{\theta_C Y(S_0 - S)}{\theta_H(1 + K_d\theta_C)}. \tag{11}$$

The kinetic parameters Y and K_d can be obtained by rearranging Eq. (10) as shown below:

$$\frac{S_0 - S}{\theta_H X} = \frac{1}{Y}\left(\frac{1}{\theta_C}\right) + \frac{1}{Y} K_d. \tag{12}$$

By plotting Eq. (12), the values of Y and K_d can be calculated from the slope and intercept of the best fit line. The values of μ_{max} and K_s could be determined by plotting Eq. (13), which was derived by rearranging Eq. (7). The value of μ_{max} can then be calculated from the intercept of the straight line while K_s can be obtained from the slope of the line as

$$\frac{\theta_C}{1 + \theta_C K_d} = \frac{K_s}{\mu_{max}}\frac{1}{S} + \frac{1}{\mu_{max}}. \tag{13}$$

2.6.2. Second-order Grau multicomponent substrate removal model

The general equation of a second-order kinetic model is illustrated below as (Sanchez et al., 2001; Grau et al., 1975)

$$-\frac{dS}{dt} = k_s X\left(\frac{S}{S_0}\right)^2. \tag{14}$$

If Eq. (14) is integrated and then linearilized, Eq. (15) will be obtained as

$$\frac{S_0\theta_H}{S_0 - S} = \theta_H + \frac{S_0}{k_s X}. \tag{15}$$

If the second term of the right part of Eq. (15) is accepted as a constant, the equation will be obtained as

$$\frac{S_0\theta_H}{S_0 - S} = b\theta_H + a. \tag{16}$$

$(S_0 - S)/S_0$ expresses the substrate removal efficiency. In order to predict the effluent COD and biomass concentrations Eqs. (17) and (18) can be used as

$$S_e = \frac{S_0(1 + a \times \theta_c)}{\theta_C \times b \times k_2}, \tag{17}$$

$$X = \frac{b\theta_C(S_0 - S_e)}{\theta_H(1 + a\theta_C)}. \tag{18}$$

2.6.3. Stover–Kincannon model

In this model the substrate utilization rate is expressed as a function of the organic loading rate using monomolecular kinetics for UASB reactors. A special feature of the Modified Stover/Kincannon model is the utilization of the concept of total organic loading rate as the major parameter to describe the kinetics of an anaerobic reactor in terms of organic matter removal and methane production (Yu et al., 1998b). The removal of organic substrate in the UASB process can be determined on the basis of the substrate removal rate as a function of the substrate concentration.

Equations of the modified Stover/Kincannon model are as follows (Beitrame et al., 1982; Weiland and Manur, 1997; Yu et al., 1998a; Hammer and Sidney, 2002):

$$\frac{dS}{dt} = \frac{R_{max}(QS_0/V)}{K_B + (QS_0/V)}, \tag{19}$$

where dS/dt is defined as

$$\frac{dS}{dt} = \frac{Q}{V}(S_0 - S). \tag{20}$$

Eq. (21) is obtained from the linearization of Eq. (20) as follows:

$$\frac{V}{Q(S_0 - S)} = \frac{K_B}{R_{\max}}\frac{V}{QS_0} + \frac{1}{R_{\max}}a. \tag{21}$$

This expression can be solved for either the effluent substrate concentration (Eq. (22)) or the required volume of the UASB reactor (Eq. (23)) by substituting the kinetic constans $U_{\max}$ and K_B:

$$S_e = S_0 - \frac{U_{\max} \times S_0}{K_B + (QS_0/V)}, \tag{22}$$

$$V = \frac{Q \times S_0}{(U_{\max} \times S_0/S_0 - S_e) - K_B}. \tag{23}$$

3. Results and discussion

3.1. Reactor performance

In order to obtain the kinetic coefficients three different kinetic models were used for the UASB reactor based on degradation of COD and 2,4 DCP at six different HRTs through 96 days of operation. The results obtained for steady-state conditions during the reactor operation at six different HRTs are summarized in Table 1. A total of 83% COD removal efficiency was obtained at a HRT of 20 h. The COD removal efficiencies decreased to 26% at a HRT of 2 h. The 2,4 DCP removal efficiencies decreased from 99% to 83% as the HRTs decreased from 20 to 2 h. The methane percentages were 73.8% and 50.8% for HRTs varying between 5 and 20 h. As the HRT decreased from 20 to 2 h, the pH and B.Alk levels decreased to 6.79 and 2000 mg $CaCO_3$ l^{-1} from 7.41 and 3700 mg l^{-1} while the VFA concentrations increased to 1900 from 605 mg l^{-1}, respectively. Although the pH remained in the optimal working range for anaerobic conditions (6.5–7.5) the VFA concentrations increased to 1900 mg CH_3COOH l^{-1} through the operation at a HRT of 2 h. The optimal pH could be explained by the neutralization of hydrogen anions being released from the volatile fatty acid together with the carbonates dissociated from the carbonic acid with the bicarbonate alkalinity inside the UASB reactor (Razo-Flores et al., 1997; Beydilli et al., 1998). The VFA/B.Alk. ratios increased to 0.39 from 0.19 as the HRTs decreased from 20 to 3 h, indicating the reactor stability. It was found that this ratio was higher than 0.4 for HRTs of 2 days showing the moderate instability of the UASB reactor (Behling et al., 1997). In order to obtain the UASB reactor performance the reactor system was operated at six different hydraulic retention times. The optimum HRT and SRT for maximum COD and 2,4 DCP removal efficiencies and methane percentage were 20 h and 646 days for the UASB reactor, respectively. The optimum HRTs varied between 5 and 20 h for optimum COD removal ($E = 64$–82%), 2,4 DCP removal ($E = 90$–99%) and methane percentages (50–73%).

High alkalinity in effluent indicates that the alkalinity supplied to the reactor is not used at high HRTs. In other words the system is under steady-state conditions and there is no need for the additional alkalinity. The VFA concentrations in the effluent increased after a HRT of 5 h indicating that the methane bacteria do not convert the VFA to methane. At HRTs of 2 and 3 h the methanogenic archae lost their activity and methanogenesis decreased significantly. The methane percentage also decreased from 50% to 30% after a HRT of 5 h. The VFA/B.Alk ratios also indicate the reactor stability. These results showed that the UASB reactor was operated with high 2,4 DCP, COD removals and methane gas percentages for HRTs varying between 5 and 20 h.

Table 1
Reactor performance under steady-state conditions at six different HRTs

Parameter	HRT (h)					
	20 ($n = 4$, mean values)	10 ($n = 4$, mean values)	8 ($n = 3$, mean values)	5 ($n = 4$, mean values)	3 ($n = 4$, mean values)	2 ($n = 4$, mean values)
Operation period (days)	0–12	13–27	28–45	46–60	61–82	83–96
F/M ratio (mg COD mg^{-1} VSS day^{-1})	1.2	1.35	1.52	1.78	1.82	1.92
Sludge retention time (days)	646	234	198	96	59	38
pH	7.41±0.04	7.38±0.13	7.30±0.02	6.95±0.01	6.82±0.04	6.79±0.02
VFA (mg CH_3COOH l^{-1})	605±32	740±5	800±44	1200±20	1600±90	1900±±97
B.Alk (mg $CaCO_3$ l^{-1})	3700±30	3403±135	3000±136	2765±4.9	2298±100	2000±90
VFA/B.Alk ratio	0.19±0.02	0.20±0.01	0.24±0.03	0.30±0.01	0.39±0.05	0.47±0.10
Effluent COD (mg l^{-1})	514±42	954±23	1126±24	1292±96	1598±89	2000±21
Effluent 2,4 DCP (mg l^{-1})	0.1	0.3	0.5	1	1.5	1.7
2,4 DCP removal efficiencies (%)	99	97	95	90	85	83
COD removal efficiencies (%)	82.95±1.96	70.14.5±0..5	69.8±1.2	64.32±2.1	55.49±1.4	25.78±2.5
Total gas production rate (ml day^{-1})	976±98	1400±78	4300±789	5230±120	5980±450	6050±234
Methane gas production rate (ml day^{-1})	497±98	769±98	1200±342	1583±95	1800±257	1200±164
Methane percentage (%)	73.8±1.8	70.3±2.2	56.7±1.2	50.8±3.18	48.2±1.6	45.7±2.08
Methane yield (l methane gCOD^{-1} added)	0.318±0.060	0.248±0.015	0.404±0.114	0.238±0.052	0.182±0.070	0.104±0.013

3.2. Conventional Monod kinetic model

To determine the yield (Y) and decay constant (k_d), the inverse of sludge retention time (1/SRT, day^{-1}) was plotted against the specific substrate utilization rate (U, mg COD g^{-1}SS day^{-1}). Y was determined from the slope as 0.78 mg SSg^{-1} COD and b was determined from the intercept of the y-axis as 0.093 day^{-1}(see Table 2).

Similarly, K_s and k were determined by plotting the inverse of the effluent substrate concentration ($1/S_e$) against the inverse of the substrate utilization rate ($1/U$). The half-saturation concentration K_s was calculated from the intercept of the x-axis as 560.69 mg COD l^{-1}, and the maximum substrate utilization rate (k) was calculated from the intercept of the y-axis as 0.954 day^{-1}. μ_{max} was calculated by multiplying Y by k to be 0.74 mg SS mg^{-1} COD day^{-1}.

Table 2
Kinetic parameters of UASB reactor treating COD as primary substrate

Kinetic models	Kinetic parameters	Values	Regression equation and coefficients (R^2)
Conventional Monod	Y (mg SS mg COD^{-1})	0.78	$y = 0.78x+0.093$ 0.982
	k_d (day^{-1})	0.093	0.982
	μ (day^{-1})	$Y \times k_d$	
	k (day^{-1})	0.954	$y = 560x+0.954$ 0.943
	μ_{max} (day^{-1})	0.213	0.943
	K_S (mg l^{-1})	560	0.943
Grau second order	a (dimensionless)	0.0291	$y = 0.011x+0.02$ 0.942
	b (dimensionless)	0.0113	0.942
	k_2 (day^{-1})	0.26	0.942
Modified Stover Kincannon	K_B (mg (l day)$^{-1}$)	34.567	$y = 34.56x+7.501$ 0.991
	U_{max} (mg COD(l day)$^{-1}$)	7.501	0.991

The K_s and k values were 2.06 mg 2,4 DCP l^{-1} and 0.0041 day^{-1} through the degradation of 2,4 DCP in the UASB reactor while μ_{max} was 0.42 day^{-1}. Y was determined as 0.0017 mg 2,4 DCP mg^{-1} SS and k_d was determined as 3.1×10^{-5} day^{-1}(see Table 3).

These results demonstrate that the growth yield coefficient was higher and the death rate constant was lower. This can be explained by the long hydraulic and sludge retention times having a positive effect on microbiological growth. A high K_s value indicates the affinity of methanogens to COD. From these results the design equations can be written as follows:

The effluent COD concentrations can be predicted using Eq. (8) based on Monod kinetics:

$$S = \frac{560(0.09 + 1/\theta_C)}{0.213 - 0.093(1/\theta_C)}.$$

The microorganism concentration in the effluent can be predicted using Eq. (11) for COD removal based on Monod kinetics:

$$X = \frac{\theta_C 0.7(S_0 - S)}{\theta_H(1 + 0.09\theta_C)}.$$

The effluent 2,4 DCP and the microorganism concentrations can be predicted using Eqs. (10) and (11) as

$$S = \frac{2.06(3.1 \times 10^{-3} + 1/\theta_C)}{0.41 - 3.1 \times 10^{-5} - 1/\theta_C}$$

$$X = \frac{\theta_C \times 0.0017(S_0 - S)}{\theta_H(1 + 3.1 \times 10^{-5}\theta_C)}.$$

Table 3
Kinetic parameters of UASB reactor treating 2,4 DCP

Kinetic models	Kinetic parameters	Values	Regression equation and coefficients (R^2)
Conventional Monod	Y (mg SS mg COD^{-1})	0.0017	$y = 0.017x + 3E{-}05$ 0.951
	k_d (day^{-1})	3.1×10^{-5}	0.951
	μ (day^{-1})	Yk_d 6.97×10^{-6}	
	k (day^{-1})	0.0041	$y = 0.0041x + 2.06$ *0.902*
	μ_{max} (day^{-1})	0.41	
	K_S (mg l^{-1})	2.02	0.902
Grau second order	a (dimensionless)	0.541	$y = 0.9829x + 0.5413$ 0.991
	b (dimensionless)	0.98	0.991
	k_2 (day^{-1})	0.30	0.991
Modified Stover Kincannon	K_B (mg (l day)$^{-1}$)	9.8×10^{-3}	$y = 0.9829x + 53.11$ 0.992
	U_{max} (mg COD(l day)$^{-1}$)	0.01	0.992

3.3. Grau second-order model application

From the plots between HRT/E and HRT, a and b constant values were found with high correlation coefficient ($R^2 = 0.94$) from the intercept and the slope of the best fit line. The Second-order substrate removal rate constant k_2 ($a = S_0/(k_2X)$), and the a and b constants were calculated as 0.26 day^{-1}, 0.0291 and 0.0113 (dimensionless), respectively through COD degradation (see Table 2). The Second-order substrate removal rate constant k_2, and the a and b constants were calculated as 0.30 day^{-1}, 0.541 and 0.98 (dimensionless), respectively through anaerobic treatment of 2,4 DCP in the UASB reactor (see Table 3). From these results the design equations can be written as follows:

The effluent COD and biomass concentrations can be predicted using Eqs. (17) and (18) as

$$S_e = \frac{S_0(1 + 0.0291\theta_C)}{\theta_C x 0.0113 \times 0.26},$$

$$X = \frac{0.0113\theta_C(S_0 - S)}{\theta_H(1 + 0.0291\theta_C)}.$$

The effluent 2,4 DCP and biomass concentrations can be predicted using Eqs. (17) and (18) as

$$S_e = \frac{S_0(1 + 0.541 \times \theta_C)}{\theta_C \times 0.98 \times 0.30},$$

$$X = \frac{0.98\theta_C(S_0 - S_e)}{\theta_H(1 + 0.540\theta_C)}.$$

3.4. Modified Stover–Kincannon model application

The removal of substrate in the anaerobic UASB reactor can be determined by the substrate removal rate as a function of the substrate concentration. The straight line portion of the intercept $1/U_{max}$ and the slope of K_B/U_{max} from the graph plotted between (K_B/U_{max}) and U_{max} and K_B were found to be 7.501 and 34.56501 mg l^{-1} day^{-1}, respectively, from the experimental data given in Table 2 with a high regression coefficient ($R^2 = 0.991$). The maximum 2,4 DCP removal rate constant (U_{max}) and saturation value (K_B) were calculated as 0.01 mg COD l^{-1} day^{-1} and 9.8×10^{-3} mg l^{-1} day^{-1} from the Modified Stover–Kincannon model (see Table 3).

The effluent COD and the volume of the UASB reactor can be predicted from Eqs. (19) and (20) as

$$S_e = S_0 - \frac{7.501 \times S_0}{34.56 + (QS_0/V)},$$

$$V = \frac{Q \times S_0}{(7.501 \times S_0/S_0 - S_e) - 34.56}.$$

The effluent 2,4 DCP and the volume of the UASB reactor can be predicted from the equations given below:

$$S_e = S_0 - \frac{0.01 \times S_0}{9.8 \times 10^{-3} + (QS_0/V)},$$

$$V = \frac{Q \times S_0}{(0.01 \times S_0/S_0 - S_e) - 9.8 \times 10^{-3}}.$$

Fig. 2a and b indicate the relationship between the observed effluent COD and 2,4 DCP concentrations from the experiments carried out under six different HRTs and predicted COD and 2,4 DCP effluent concentrations calculated by using Eq. (8) based on Monod kinetics. Fig. 3a and b indicate the relationship between the predicted and observed effluent COD and 2,4 DCP concentrations and predicted effluent COD and 2,4 DCP concentrations were calculated using Eq. (19) on the basis of the Grau second-order kinetics. Fig. 4a and b indicate the relationship between the observed effluent COD and 2,4 DCP concentrations from the experiments carried out under six different HRTs and predicted COD and 2,4 DCP effluent concentrations calculated by using Eq. (19) based on Modified Stover–Kincannon kinetics. There is a linear

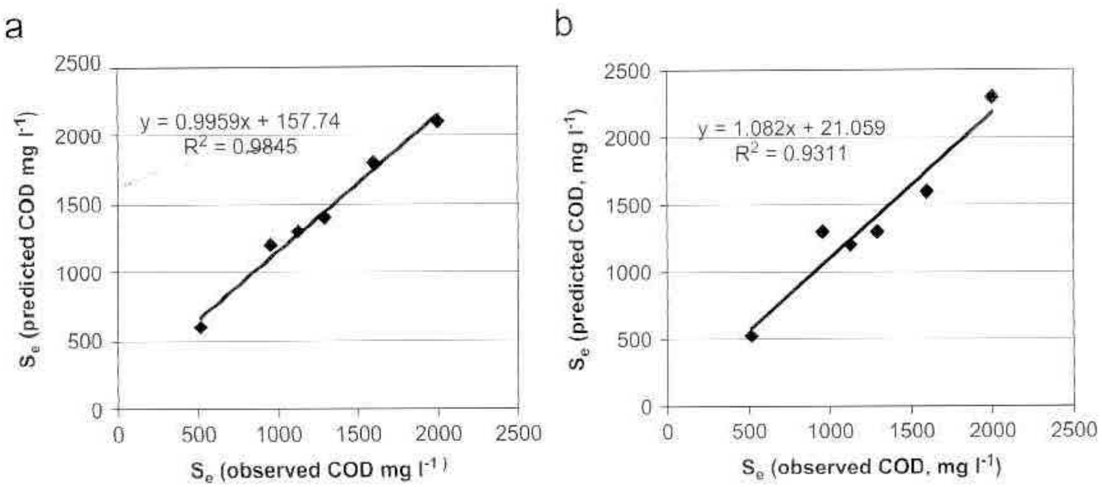


Fig. 2. The relationships between observed and predicted COD concentrations (a) and 2,4 DCP concentrations (b) for Monod kinetic model.

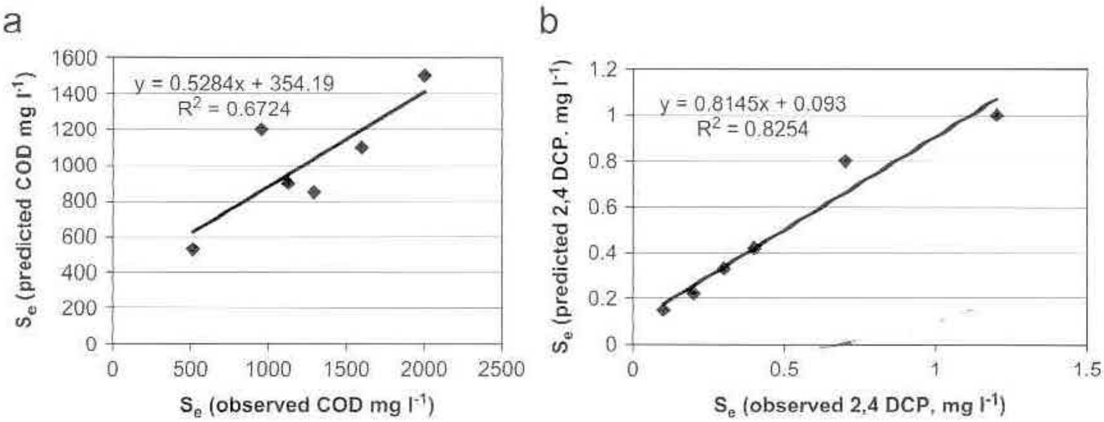


Fig. 3. The relationships between observed and predicted COD concentrations (a) and 2,4 DCP concentrations (b) for Grau second-order kinetic model.

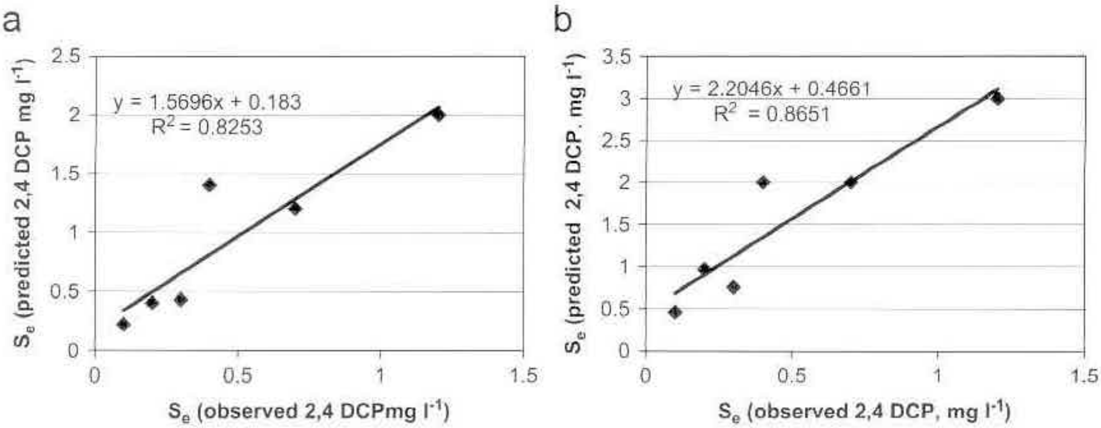


Fig. 4. The relationships between observed and predicted COD concentrations (a) and 2,4 DCP concentrations (b) for Modified Stover Kincannon kinetic model.

relationship between the observed and predicted COD and 2,4 DCP concentrations with a high R^2 (regression coefficient). From these figures it can be seen that the regression coefficients between the predicted and observed COD and 2,4 DCP concentrations are higher for Monod kinetics ($R^2 = 0.98$ and 0.95, respectively), compared to the regressions between observed and predicted values for Modified Stover–Kincannon kinetics ($R^2 = 0.82$ and 0.86, respectively) and second-order Grau kinetics ($R^2 = 0.67$ and 0.82, respectively).

3.5. *Evaluation of the kinetic models*

The kinetic data showed that the Monod substrate removal kinetic model was a more appropriate model compared to the Grau and Modified Stover–Kincannon models for predicting the performance of the lab scale UASB reactor treating 2,4 DCP with glucose-COD as co-substrate.

The yield value is higher compared to the decay constant and K_s is lower compared to the initial COD concentration of 3000 mg l^{-1} in Monod kinetics. The maximum substrate utilization rate (k) in Monod kinetics is lower than the second-order substrate utilization rate in Grau kinetics through 2,4 DCP degradation. However, the k value in Monod kinetics is significantly higher than the k_2 value in Grau kinetics for COD degradation.

The predicted and observed COD concentrations indicate that there is a high regression coefficient between these parameters in Monod kinetics compared to the other two models (see Figs. 2a, 3a and 4a). The plot of the best fit line between predicted and observed 2,4 DCP concentrations showed that the 2,4 DCP was removed according to Monod kinetics (see Figs. 2b, 3b and 4b). The regression coefficient between predicted and observed values for COD and 2,4 DCP were high ($R^2 = 0.98$, 0.95, respectively) in Monod kinetics compared to the Stover–Kincannon and Grau second-order kinetic models. The kinetic coefficients calculated from the Monod kinetics are more significant.

The maximum second-order removal constant for 2,4 DCP was significantly lower than that observed in Modified Stover–Kincannon kinetics. In Monod kinetics the yield values are lower through degradation of 2,4 DCP since the SS used for calculating the yield contains both

COD and 2,4 DCP removing anaerobic methanogens. In this study the K_s values are lower (2.02 mg l^{-1})than those from another study treating 2,4 DCP ($K_s = 6.49$ mg l^{-1}, initial 2,4 DCP was 12 mg l^{-1}) under anaerobic conditions indicating the consumption of 10 mgl^{-1} 2,4 DCP by anaerobic archae (Majumder and Gupta, 2007). The lower K_s value indicates the affinity of anaerobic microorganisms for the substrate. Therefore, it can be said that no 2,4 DCP accumulation was observed in the UASB reactor. The μ_{max} value is significantly higher than that μ values for 2,4 DCP in Monod kinetics (see Table 3).

Since the effluent COD and 2,4 DCP concentrations varied with HRT, these two parameters are the most important variables in the UASB reactor. Therefore the observed effluent COD and 2,4 DCP concentrations and predicted effluent 2,4 DCP and COD concentrations for every HRT were compared for the three kinetic models (see Figs. 5 and 6). The HRT versus observed and predicted effluent COD and 2,4 DCP concentrations showed that the predicted effluent COD and 2,4 DCP concentrations are closer to the observed values when the calculated kinetic constants were placed in the Monod kinetic model ($R^2 = 0.98$, 0.95 for COD, and 2,4 DCP, respectively). When the UASB reactor was operated between HRTs of 5 and 20 h the COD and 2,4 DCP removal efficiencies were 64–82% and 90–99% , respectively. The COD and 2,4 DCP in the UASB reactor could be removed with high efficiencies according to Monod kinetics for long-term steady-state operations at HRTs varying between 5 and 20 h. The k values obtained in this study for Monod kinetics are much higher than the data obtained by Sponza (2001a, b), Ünal (1999) and Sponza and Atalay (2005). The k_2 value found in this study is much lower than the k_2 values calculated by Büyükkamacı and Filibeli (2002) (10.82 day^{-1}) (and Öztürk et al. (1998) (1.65 and 13.6 day^{-1}) in which they used molasses and glucose as carbon sources without any toxicant in a hybrid and UASB reactor, respectively.

The kinetic constants obtained from the Modified Stover–Kincannon kinetic model are much lower than the U_{max} and K_B values obtained by Büyükkamacı and Filibeli (83.3 and 186 g l^{-1} day^{-1}, respectively) (Sponza and Atalay, 2005) and Yu et al. (83.3 and 85.5 g l^{-1} day^{-1}; Öztürk et al., 1998) in a hybrid and anaerobic filter reactor

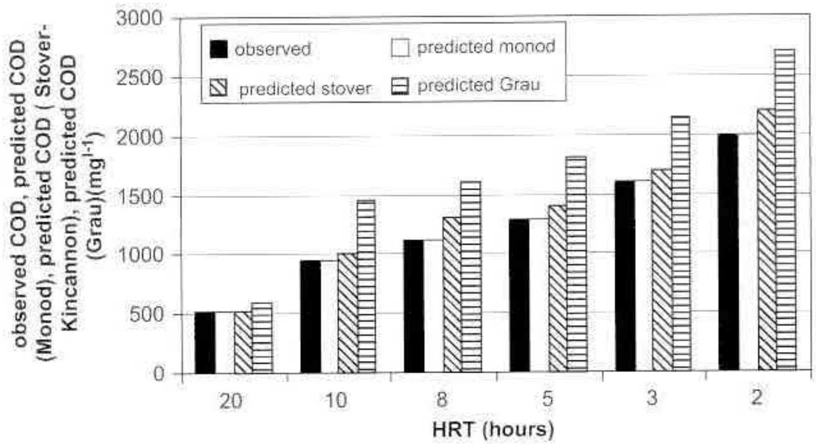


Fig. 5. Comparison of predicted effluent COD concentrations according to models.

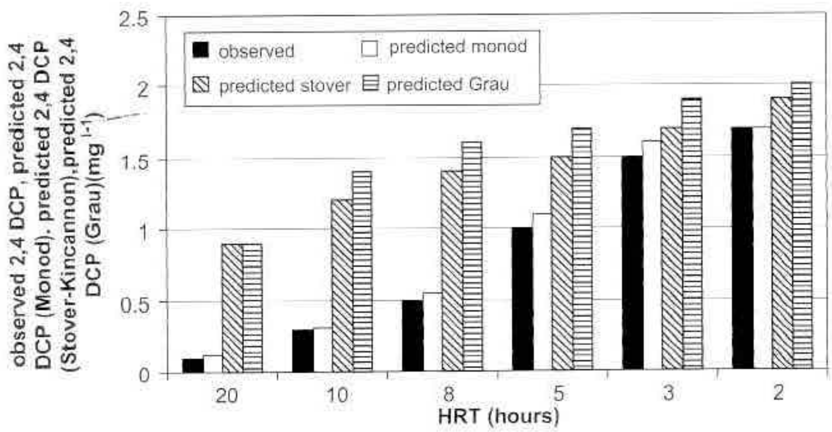


Fig. 6. Comparison of predicted effluent 2,4 DCP concentrations according to models.

treating only molasses and soybean, respectively. This could be attributed to the 2,4 DCP loading rate as high as $0.116\,mg\text{-}2,4\ DCP\,l^{-1}\,day^{-1}$ applied to the UASB reactor in this study. However the aforementioned kinetic constants are significantly lower than those obtained by Sponza and Işık treating a simulated textile wastewater ($8.5\,g\,l^{-1}\,day^{-1}$) (Işık and Sponza, 2005).

4. Conclusion

The results of this study showed that synthetic wastewater containing 2,4 DCP can be effectively treated at HRTs between 5 and 20 h in a UASB reactor. The COD removal efficiencies decreased from 83% to 65% when the HRT was decreased from 20 to 2 h while 83% and 99% 2,4 DCP removals were achieved, respectively. The COD and methane percentage decreased significantly at a HRT of 2 h. The 2,4 DCP removal efficiencies were not significantly affected at HRTs of 2 and 3 h. For maximum COD and 2,4 DCP removal efficiencies and methane gas production the optimum HRTs should be between 5 and 20 days.

The Y value constant was significantly higher than the K_d value indicating that the 2,4 DCP degrading anaerobic microorganisms were active and the death of these was at a low level based on Monod kinetics. The μ value was lower than the μ max value based on Monod kinetics indicating that the maximum specific growth rate of anaerobic methanogens is higher than the specific growth of methanogens during COD removal in a UASB. The k value based on Monod kinetics was higher than that based on Grau second-order reaction kinetics. The K_s values for COD and 2,4 DCP degradations for Monod kinetics were lower indicating no accumulation of substrate in the UASB reactor. The low K_s values indicate the affinity of methanogens for the substrate.

There was good concurrence between observed and predicted concentrations both for COD and 2,4 DCP for Monod kinetics, meaning that Monod kinetic constants can be used to design a UASB reactor for 2,4 DCP and COD removals from wastewaters containing chlorinated organics.

The results of kinetic studies based on the anaerobic removal of 2,4 DCP and COD in a lab scale UASB reactor can be used to predict the treatment performance of a full-scale UASB reactor using the Monod kinetic model. Furthermore, the kinetic constants can also be used for optimizing the UASB design. It was shown that the Grau and Stover–Kincannon models are not very appropriate for predicting the performance of a UASB reactor for degradation of 2,4 DCP.

Acknowledgments

This study was executed as a part of the research activities of the Environmental Microbiology Laboratory of Environmental Engineering Department of projects numbered Fen 021, Fen 051 and 03.KB.Fen 017, which were partially funded by the Dokuz Eylül University Research Foundation. The authors would like to thank this body for the financial support provided.

References

Anderson, G.K., Yang, G., 1992. Determination of biocarbonate and total volatile fatty acid concentration in anaerobic digertser using a simple titration. Water Environment Research 64, 53–59.

APHA, 1975. Standard Methods for the Examination of Water and Wastewater. APHA, Washington, DC.

Armmenante, P.M., Kafkewitz, D., Lewandowski, G.A., Jou, C.J., 1999. Anaerobic–aerobic treatment of halogenated phenolic compounds. Water Research 33, 681–692.

Behling, E., Diaz, A., Colina, G., Herrera, M., Gutierrez, E., Chacin, E., Fernandez, E., Forster, C.F., 1997. Domestic wastewater treatment using a UASB reactor. Bioresource Technology 61, 239–245.

Beitrame, P., Beitrame, P.L., Corniti, P., Pitea, D., 1982. Kinetics of biodegradation of mixture containing 2,4-DCP in a continuous stirred reactor. Water Research 16 (4), 429–433.

Beydilli, M.I., Pavlostathis, S.G., Tincher, W.C., 1998. Dechlorination and toxicity screening of selected reactive azo dyes under methanogenic conditions. Water Science and Technology 38, 225–232.

Boyd, S.A., Shelton, D.R., 1984. Anaerobic biodegradation of chlorophenols in fresh and acclimated sludge. Applied and Environmental Microbiology 47, 272–277.

Boyd, S.A., Shelton, D.R., Berry, D., Tiedje, J.M., 1983. Anaerobic biodegradation of phenolic compunds in digested sludge. Applied and Environmental Microbiology 46, 50–54.

Büyükkamacı, N., Filibeli, A., 2002. Determination of kinetic constant of an anaerobic hybrid reactor. Process Biochemistry 38, 73–79.

Chen, Y.R., Hashimoto, A.G., 1980. Substrate utilisation kinetic model for biological treatment processes. Biotechnology and Bioengineering 22, 2081–2095.

Grady, C.P.L., Harlow, L.J., Riesing, R.R., 1972. Effects of growth rate and influent substrate concentration on effluent quality from chemostats containing bacteria in pure and mixed culture. Biotechnology and Bioengineering 14, 391–410.

Grau, P., Dohanyas, M., Chudoba, J., 1975. Kinetic of multicomponant substrate removal by activated sludge. Water Research 9, 337–342.

Gu, Y.X., Knaebal, D.B., Korus, R.A., Crawford, R.L., 1995. 2,4 Dichlorophenoxyacetic acid (2,4-D) detection using 2,4—D.a-Ketoglutarate Dioxygenase. Environmental Science and Technology 29, 1622–1627.

Hammer, A., Sidney, Q., 2002. A comparison of mesofilic and thermofilic anaerobic upflow filters treating paper pulp liquors. Process Biochemistry 38 (1), 73–79.

Işık, M., Sponza, D.T., 2005. Kinetics in an upflow anaerobic sludge blanket reactor decolorising simulated textile wastewater. Process Biochemistry 40, 1189–1198.

Kesavan, P., Law, V.T., 2005. Practical identifiability of parameters in Monod Kinetics and statistical analysis of residuals. Biochemical Engineering Journal 24, 95–104.

Kim, I.S., Tabak, H.H., Young, J.C., 1998. Modelling of the fate and effect of chlorinated phenols in anaerobic treatment processes. Water Science and Technology 36 (6–7), 287–294.

Krumme, M.L., Boyd, S.A., 1988. Reductive dechlorination of chlorinated phenols in anaerobic upflow reactors. Water Research 22, 171–177.

Majumder, P.S., Gupta, S.K., 2007. Removal of chlorophenols in sequential anaerobic/aerobic reactors. Bioresource Technology 98 (1), 118–129.

Marcos, N.I., Guay, M., Dochain, D., 2004. Output feedback adaptive extremum seeking control of a continuous stirred tank bioreactor with Monod 's kinetic. Journal of Process Control 14 (7), 807–818.

Ning, Z., Kennedy, K.J., Fernandes, L., 1997. Anaerobic degradation kinetics of 2,4 dichlorophenol (DCP) with linear sorption. Water Science and Technology 35 (2–3), 67–75.

Nitayapat, N., Watson-Craik, I.A., 2002. The influence of experimental conditions on the assessment of the toxicity of 2,4 dichlorophenol to anaerobic degradation bacteria. Water Science and Technology 45 (10), 61–63.

Notice of the first priority list of hazardous substances that will be the subject of toxicological profiles, 1987. Federal Register 52, p. 1286.

Öztürk, I., Altinbas, M., Arikan, O., Demir, A., 1998. Anaerobic UASBR treatment of young landfill leachate. In: First International Workshop on Environmental Quality and Environmental Engineering in the Middle East Region, Konya, Turkey.

Perkins, P.S., Komisor, S.J., Puhakka, J.A., Ferguson, J.A., 1994. Effects off electron donors and inhibitors on reductive dechlorination of 2,4,6-trichlorophenol. Water Research 28, 2101–2107.

Puhakka, J.A., Jarvinen, K., 1992. Aerobic fluidized-bed treatment of polychlorinated phenolic wood preservative constituents. Water Research 26, 765–770.

Razo-Flores, E., Luijten, M., Donlon, B.A., Lettinga, G., Field, J.A., 1997. Biodegradation of selected azo dye under methanogenic conditions. Water Science and Technology 36 (6–7), 65–72.

Sanchez, E., Borja, R., Weilant, P., Travieso, L., 2001. Effect of substrate concentration and temperature on the anaerobic digestion of piggery waste in a tropical climate. Process Biochemistry 37, 483–489.

Speece, R.E., 1996. Anaerobic Biotechnology for Industrial Wastewaters. Archae Press, Nashville, TN.

Sponza, D.T., 2001a. Performance of upflow anaerobic sludge blanket (UASB) reactor treating wastewaters containing carbon tetrachloride. World Journal of Microbiology & Biotechnology 17, 839–847.

Sponza, D.T., 2001b. Anaerobic granule formation and tetrachloroethylene (TCE) removal in an upflow anaerobic sludge blanket (UASB) reactor. Enzyme and Microbial Technology 29, 417–427.

Sponza, D.T., Atalay, H., 2005. Treatment of trichlorotoluene in an anaerobic/aerobic sequential reactor system. Process Biochemistry 40, 69–82.

Ünal N. 1999. Factors affecting the treatment efficiency of medium strength wastewaters in UASB reactor. In: Environmental Sciences. Dokuz Eylül University, Izmir, pp. 1–98.

Weiland, S., Manur, O., 1997. Rotating biological contactors for wastewater treatment in an equatorial climate. Water Science and Technology 35 (8), 177–184.

Wik, T., Göransan, E., Breitholtz, C., 2000. Low model order approximations of continuously stirred reactors with Monod kinetics. Biochemical Engineering Journal 30 (200), 16–25.

Yu, H.K., Wilson, F., Tay, J.H., 1998a. Kinetic analysis of an anaerobic filter treating soybean wastewater. Water Research 32 (11), 203–214.

Yu, H., Wilson, F., Tay, J., 1998b. Kinetic analysis of an anaerobic filter treating soybean wastewater. Water Research 32 (11), 3341–3352.

ELSEVIER

Journal of Environmental Management 86 (2008) 132–138

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Dust deposition in a sub-tropical opencast coalmine area, India

Sudhir Kumar Pandey, B.D. Tripathi*, Virendra Kumar Mishra

Pollution Ecology Research Laboratory, Centre of Advanced Study in Botany, Banaras Hindu University, Varanasi 221005, India

Received 2 March 2005; received in revised form 21 November 2006; accepted 28 November 2006
Available online 26 January 2007

Abstract

This paper provides baseline information about the total annual dust fall, and its constituents and seasonal variation, from a sub-tropical opencast coalmine area in Bina, India. Dust samples were collected monthly for 2 years (June 2002–May 2004) from five sampling sites in the region and analyzed in the laboratory for water-soluble and -insoluble matter. Water-insoluble components constituted the major fraction of the total annual dust fall. Two-way ANOVA indicated significant variations in dust fall at different sites, over the months and in their interactions. The dust deposition rate was highest during summer (March–June), followed by winter (November–February) and lowest in the rainy season (July–October). Maximum dust fall was observed near the coal handling plant (at site 2) followed by the receiving pit of the coal handling plant (site 3), near the main sub-station (site 4), Jawahar colony (site 1) and Gharasari village (site 5). An inverse and significant relation was observed between dust fall and precipitation. Our studies have shown that the main residential areas are experiencing higher levels of dust fall which makes them unsuitable for living. We suggest that residential areas should be moved farther away from the mining area in the opposite direction of prevalent winds.

Keywords: Dust fall; Seasonal variation; Sub-tropical; Coalmine; Water-soluble and -insoluble matter

1. Introduction

The economies of China, India and other nations in Asia are now largely coal driven (Elliot et al., 1999) and coal-derived energy will continue to dominate for years to come. Consequently, the number of coalmines is increasing day by day. However, the emphasis is more towards opencast coal mining to achieve bulk production and increased productivity. In India, opencast mining today constitutes nearly 70% of the total coal production. Opencast coal mining using large-scale mechanization results in the release of huge quantities of dust and gases, which are changing the environmental impacts of coalmines, and adversely affecting human health (Dhar, 1994). Given our new understanding of the significance of deposited dust in air quality, it is one of the main causes of complaints about air pollution (Vallack and Shillito, 1998).

The effects of dust clouds and deposition are both visible and tangible in communities around industrial activities or construction sites (Hall et al., 1993; Fuglsang, 2002). Dust emission is the foremost problem of opencast coalmines and results from blasting and drilling operations, transportation of coal and overburden on haul roads, from coal handling plant operation, loading of overburden and coal by shovel dumpers, from crushing, conveying and handling of overburden by draglines, running of other vehicular traffic on the unpaved road and from vehicular emission. Dust fall rate and its chemical constituents are required measurements in the quantitative as well as qualitative study of dust pollution in a region (Harrison, 1986). The distinction between monthly and annual means is important because fugitive dust emissions are often episodic due to plant/site operational failure, varying meteorological conditions or both (Vallack and Shillito, 1998). Numerous workers have reported on various aspects of dust pollution i.e. Brooks and Schwar (1987), Irvine et al. (1989), Rybicka (1989), Tripathi et al. (1996), Adams (1997), Wanquan et al. (2004) and Reynolds et al. (2005). A good example of dust deposition modeling comes from Gao et al. (2003).

Ghose (1989) reported monitoring of air borne dust and its abatement measures whereas Ghose and Sinha (1990) gave information on an air pollution control plan for opencast coalmines. Ghose and Banerjee (1995) highlighted

*Corresponding author. Tel.: +91 542 2368419; fax: +91 542 2368174.
E-mail address: bdtripathi1948@yahoo.co.in (B.D. Tripathi).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.032

the air pollution status in opencast coalmines. Banerjee et al. (1996) discussed the air pollution problem and abatement measures in Indian opencast coalmines. Ghose and Banerjee (1997) conducted a study on the physico-chemical properties of air borne dust in coal washeries of India. Ghose and Majee (2000a, b) assessed the impact of opencast coal mining on the air environment. Ghose and Majee (2001) suggested abatement measures for air pollution caused by opencast coal mining. Ghose (2002) and Chaulya (2004) discussed the air pollution problem in Indian opencast coalmines. However, still there is a paucity of data regarding seasonal variation of dust fall and its chemical constituents in opencast coalmines.

This paper reports on the total quantity of dust fall and its constituents, combustible matter, water-soluble matter, water-insoluble matter, ash and tarry substances, chloride and calcium and their seasonal variation in a sub-tropical Indian opencast coalmine. The present study was conducted in the Bina opencast project, one of the biggest coalmine projects of Northern Coal Fields, Limited (NCL), India, in Sonebhadra district.

2. Details of the study area

2.1. Location

Bina opencast coalmine lies in the Marrack block on the south and Kakari OCP in the north situated between latitudes 24°08′12″ and 24°09′52″ and longitudes 82°44′25″ and 82°45′41″ (Fig. 1). The Bina coalmine is located on a hilly terrain forming a plateau on the west and southwest, whereas towards the east and northeast the area is gently undulating. The elevation in the mining area varies from 275 to 400 m above the mean sea level.

2.2. Meteorology

The climate of the area is characteristically monsoonal with three distinctive seasons, i.e. summer (March–June), rainy (July–October) and winter (November–February) with discernible variation in the temperature and rainfall. This area receives on average approximately 1000 mm of rainfall annually and the maximum and minimum temperatures are 48 and 4 °C. The average annual relative humidity was 63% during the investigation period. The first half of the summer season was associated with strong hot dry westerly winds and high temperature (up to 48 °C), whereas the second half was humid. The mean monthly temperature ranged from 24 to 36 °C in summer. The rainy season accounted for almost 90% of the annual rainfall during which the relative humidity ranged from 79% to 90%. In dry months, the relative humidity ranged from 36% to 85% and 10% to 82% in winter and summer, respectively. Winter was characterized by 10–25 °C day temperature and during the night sometimes the temperature dropped below 4 °C.

The basic meteorological parameters determining the horizontal transport and dispersion of air pollutants are the mean wind speed and the wind direction (Ziomas et al., 1995). Fig. 2 shows seasonal and annual wind rose plots for

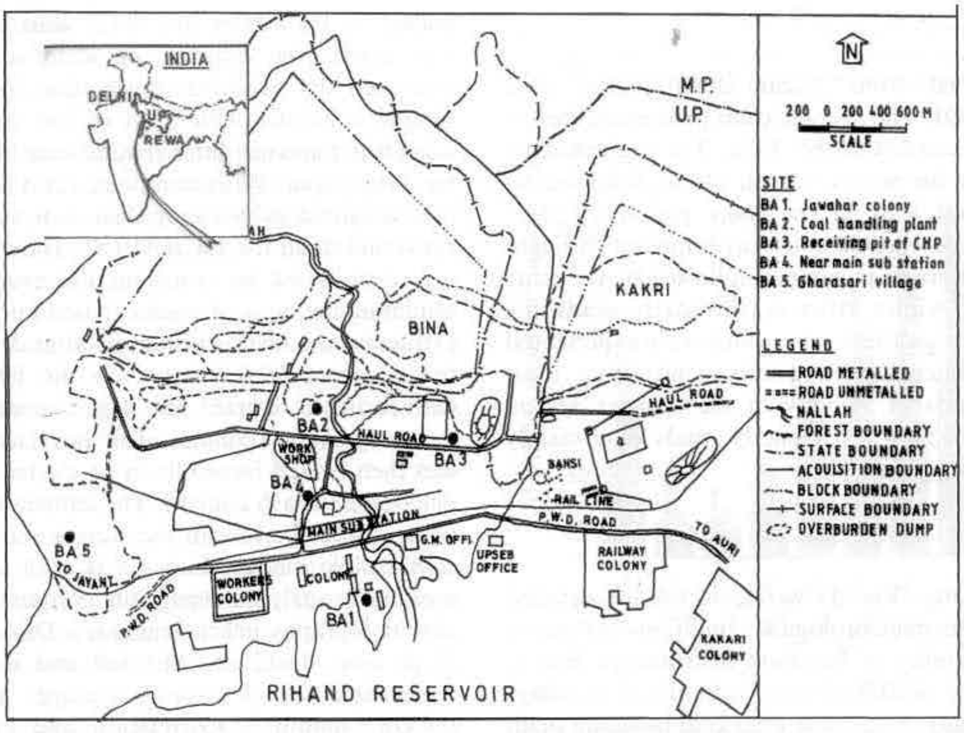


Fig. 1. Location map of Bina opencast coalmine area showing sampling sites.

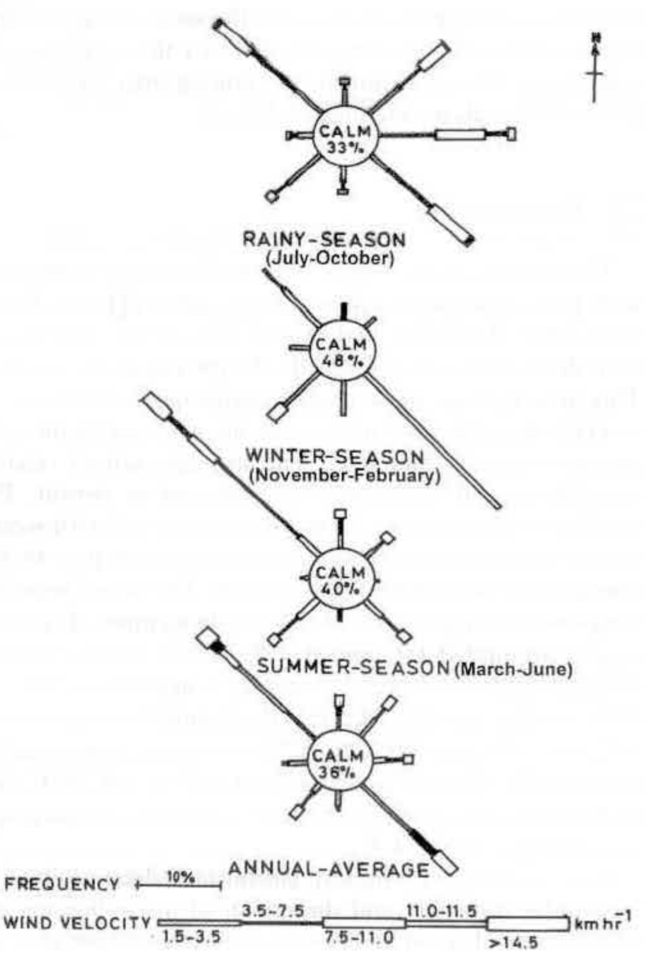


Fig. 2. Wind rose plots of Bina coalmine area developed from 2-year climatological data (years 2002–2004).

the area developed from 2-year climatological data provided by CMPDI. The average calm percentage (below $1.5\,km\,h^{-1}$) was recorded to be 36%. The highest calm percentage was in the winter season (48%) followed by summer (40%) and least in the rainy period at 33%. Northwesterly winds were the most predominant throughout the year, but more so in summer followed by the rainy period and least in winter. After northwesterly winds, the most common were southeasterly, which were experienced more often in winter. Northwesterly winds were most inconsistent in terms of velocity in the summer season ranging from 1 to $15\,km\,h^{-1}$. Easterly winds were mainly confined to the rainy season.

2.3. Sampling sites

Five sampling sites (Fig. 1) were selected for detailed study based on micrometeorological conditions, nature of activities and feasibility of handling instruments. Site 1, Jawahar colony, is a residential area, without tall buildings or trees in the vicinity. Site 2, the Bina coal handling plant site, is situated near the maintenance building of the coal handling plant and is affected by the dust generated by the coal handling plant operations. Site 3, the receiving pit of the coal handling plant, is mostly affected by coal unloading activities. Site 4, near the main sub-station, is mainly affected by transport activities. Site 5, Gharasari village, is residential surrounded by little forest.

3. Materials and methods

Dust fall collection traps were cylindrical glass containers filled with distilled water. These containers were mounted on iron tripods at a height of 10 m above the ground at all of the study sites (Mark and Hall, 1994) and exposed to the atmosphere for 1 month. These devices, placed at sites free from nearby obstructions, measured the fallout rate of coarse particulates i.e. generally above 10 μm in size. Because the period of collection was 1 month, there was a chance of collected substances undergoing chemical changes. In order to prevent this 0.02 N copper sulfate solution was added to the collection bottle. Dust fall measuring devices used in the present study are useful for identifying nuisance sources or monitoring potential sources of dust emissions (Vallack and Shillito, 1998; Schwar, 1998). Five samples from five sub-sites were collected at monthly intervals from every site. Hence, 600 samples were collected during the study period (June 2002–May 2004), and analyzed in the laboratory. Fig. 3 illustrates the general protocol for the analysis of collected dust. Residual water in the container was filtered and the residue after drying and weighing was chemically analyzed (Stern, 1976) for water-insoluble and -soluble matter. The accuracy of these determinations was approximately $\pm 3\%$. After the total dust fall determination was made, the amount of water-soluble matter was measured by washing down the inner sides of the sample container with a jet of hot distilled water. The wash water and insoluble residue were filtered through ashless filter paper (Whatmann No. 42). Filtrate was collected in a weighted evaporating dish, was subsequently evaporated and dried for 1 h at 110 °C. The mass of this residue was determined by weighing the evaporating dish and obtaining the mass of residue based on the difference. For extracting the tarry substances, after drying and weighing the water-insoluble matter on the filter paper acetone extraction was carried out. The acetone residue weight is the tarry matter content. The extracted insoluble matter was then heated forcefully in an electric oven at 800 °C to determine the ash content. The leftover matter was the ash content and the loss in the weight was equivalent to the combustible matter. Calcium (Ca^{2+}) and chloride (Cl^-) ions were analyzed using ion chromatography. The ion chromatography instrument was a Dionex DX-120, AS14, 25-μl loop. Data on dust fall and its constituents are represented as a 2-year average and expressed in $ton\,km^{-2}\,month^{-1}$. Correlation and exponential regression analysis was conducted using Microsoft-Excel and a

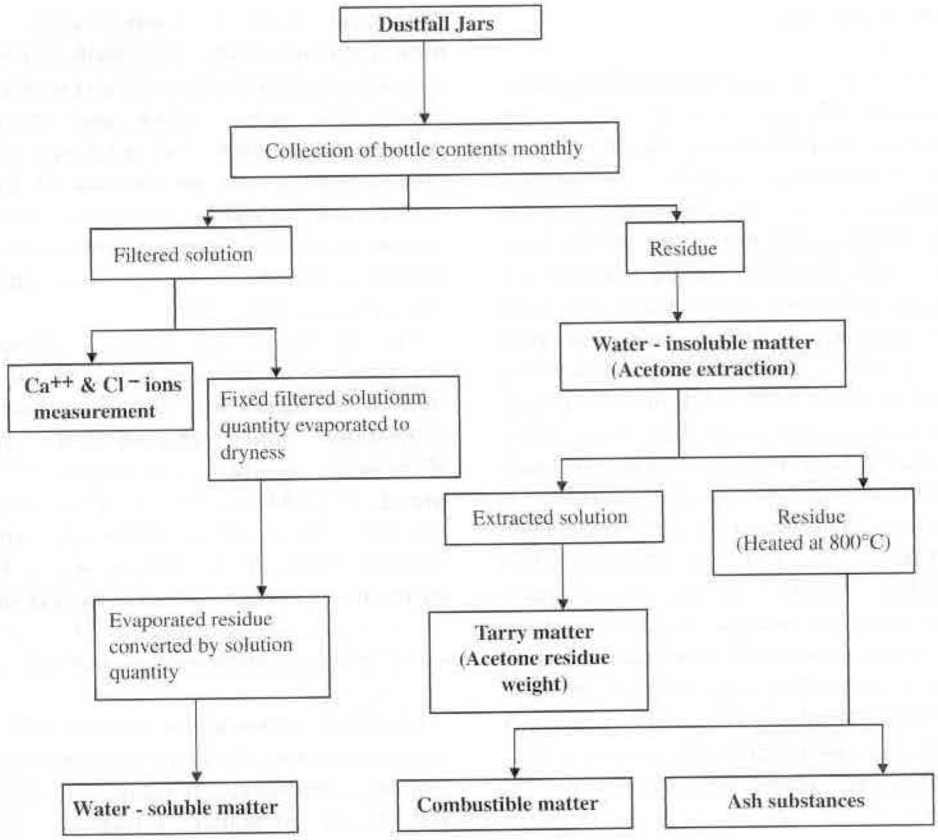


Fig. 3. General protocol of deposited dust analysis.

two-way full factorial analysis of variance (ANOVA) was conducted in SPSS 10.

4. Results and discussion

4.1. Seasonal variations in dust fall

Fig. 4 summarizes the average monthly variation in the dust fall during June 2002–May 2004. ANOVA showed significant temporal variation in the concentration of dust fall over the months ($F_{11,\ 120} = 117{,}861.56$). Maximum dust deposits were encountered in the summer season and ranged from 32.8 ± 1 to 278.9 ± 2.9 ton km^{-2} month^{-1}, when dusty winds and low humidity are a common attribute in tropical areas. The dust levels rose further by surface erosion and dust resuspension due to prolonged high winds (upto 15 km h^{-1}) and thermal turbulence caused by high temperature. The lowest values occurred during the rainy season and ranged from 16.2 ± 1.2 to 111.3 ± 3.2 ton km^{-2} month^{-1}, because heavy rain washed out the dust during these months. The capture of aerosol particles by falling rain takes place with Brownian and turbulent shear diffusion, inertial impaction, diffusiophoresis, thermophoresis and electric charge effects (Chate and Pranesha, 2004). The winter season demonstrated intermediate values with a range of 64.2 ± 3.6–226.3 ± 5.8 ton km^{-2} month^{-1}); this is attributed to climatic inversions and constantly changing wind directions (Lyons and Scott, 1990).

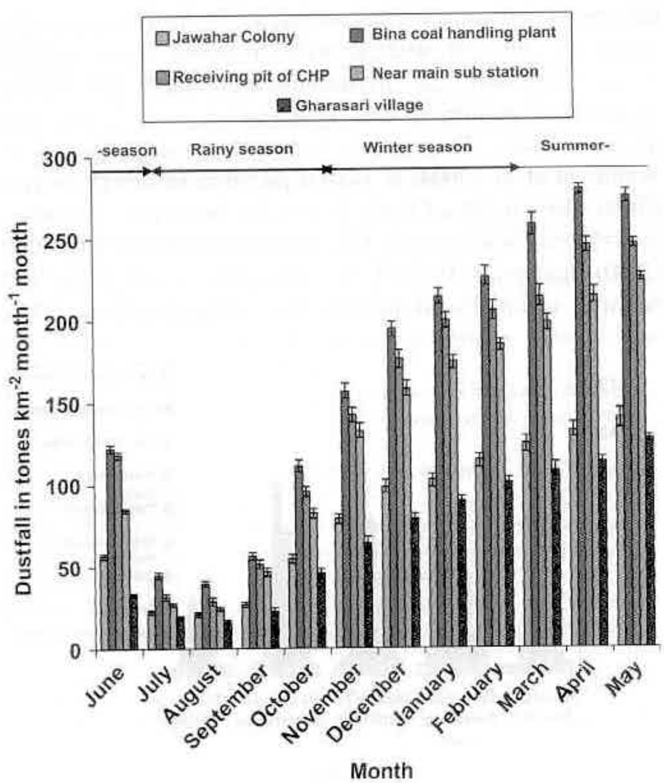


Fig. 4. Monthly variation (2-year average, 2002–2004) in dust fall deposition in Bina opencast project.

4.2. Spatial variations in dust fall

After the ANOVA test, it was also observed that spatial variations were significant ($F_{4,\ 120} = 116{,}397.40$) over the sites. The amount of dust received among the sites (Fig. 4) was highest (278.98 ± 2.98 ton km^{-2} month^{-1} in April) at site 2, the coal handling plant, due to operations like crushing, conveying, loading and unloading of the coal. The most prevalent wind direction during summer i.e. northwest (Fig. 2) elevated the dust load at site 2. Site 3, the receiving pit, also showed a higher value of dust (246.1 ± 4.5 ton km^{-2} month^{-1} in May) due to unloading of coal. The dust load at site 4, near the main sub-station, mainly consisted of resuspended road dust from heavy vehicles. Site 1, Jawahar colony, exhibited relatively lower values of dust deposits (139.8 ± 5.9 ton km^{-2} month^{-1} in May), since the area is mainly residential, but residents still experience high pollution levels. The least amount of dust recorded (32.8 ± 1 ton km^{-2} month^{-1} in June) was at site 5, Gharasari village, as expected because the area is distal from mining activities and covered by moderate forest. It lies southwest of the main mining area (Fig. 1) so small amounts of dust are brought through by winds (Fig. 2). A two-way ANOVA demonstrated significant variations ($F_{44,\ 120} = 132.62$) also in the site × month interactions.

4.3. Annual average of total dust fall

Fig. 5 summarizes the annual average value of total dust fall and its individual constituents at all the sites. The averaged values in different seasons were acquired from monthly values. Annually the maximum total dust fall (96.2 ± 3.6 ton km^{-2} month^{-1}) was observed at site 2, the coal handling plant. This rate is higher than that reported by Wanquan et al. (2004) in Gansu province (a desert area) of China where it ranged from 18.23 to 69.34 ton km^{-2} month^{-1}, in northern China (central Takelamagan desert) by Xuan et al. (2000) (maximum level of 1.5 ton ha year^{-1}) and in the sub-Saharan region by Middleton (1997) (ranging from 3.5 to 200 ton km^{-2} year^{-1}). Another study regarding dust fall measurements in India is by Shah and Shah (1964) in which they have studied the dust fall and sootfall in the Ahmedabad region (an urban area) and reported a range of 9–31 ton km^{-2} month^{-1} of total dust fall. Dust deposition rates depend mainly on the rate of dust supply from the sources, rainfall and air turbulence, and at a given location, depend chiefly on climatic conditions in the source area and climatic conditions in the deposition area (Recheis and Kihil, 1995; Ziomas et al., 1995).

Fig. 6 shows the relation between dust fall and precipitation at different sites. An exponential regression showed a significant regression pattern between dust deposition and precipitation ($y = 98.603e^{-0.0047x}$, $R^2 = -0.76$ at site 1, $y = 201.89e^{-0.0048x}$, $R^2 = -0.78$ at site 2, $y = 184.03e^{-0.0054x}$, $R^2 = -0.83$ at site 3, $y = 163.23e^{-0.0056x}$, $R^2 = -0.85$ at site 4 and $y = 81.776e^{-0.0048x}$, $R^2 = -0.73$ at site 5) However, there was no relation between temperature and dust deposition.

4.4. Chemical constituents of dust fall

Chemical components of dust fall also exhibited the highest values at site 2 (the coal handling plant) due to coal handling operations, loading and unloading of the coal and heavy vehicular transport. The water-insoluble matter was comparatively higher (range: 21.5 ± 1.4–69.4 ± 1.9 ton km^{-2} month^{-1}) than the water-soluble matter (range: 9.6 ± 1.1–26.3 ± 1.6 ton km^{-2} month^{-1}) at all of the study sites (Fig. 5). The ash content of dust was highest in winter months due to burning of low-grade coal by local laborers. It was moderately high in summer, because of dusty winds having the favorable direction. The combustible fraction of water-insoluble matter also showed higher values, which is the result of frequent running of heavy vehicles and movement of heavy earth-moving machines. Tarry matter deposition was the lowest among the water-insoluble matter. It was in the greatest amount in winter at all the study sites, and relatively lower in the summer months, because of the high frequency of ground

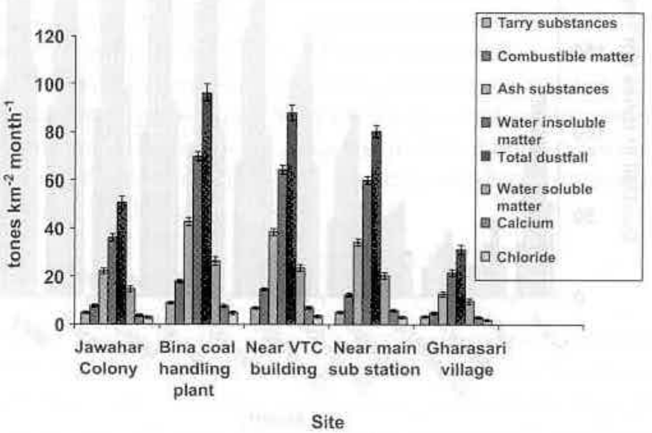


Fig. 5. Annual variation (2-year average, 2002–2004) in the dust fall and its constituents in Bina opencast coalmine area.

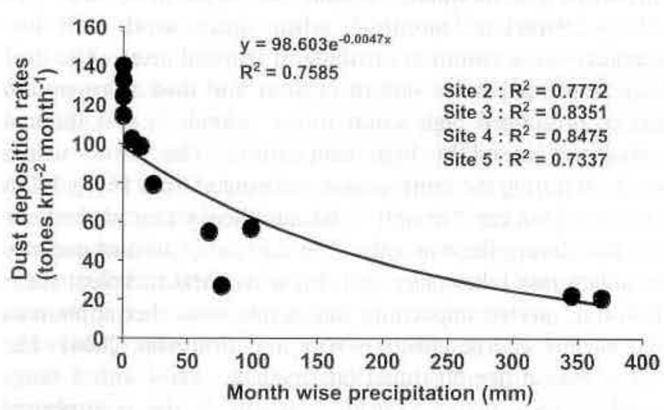


Fig. 6. Exponential regression plot of dust deposition versus precipitation in Bina opencast project at all the study sites.

inversions in winter and high wind velocity in summer. The analyzed water-soluble components (Ca^{2+} and Cl^-) were found to have low concentrations. No significant fluctuations were observed with respect to months and sites (Tripathi et al., 1991). However, they were observed to be the lowest at Gharasari village and the highest at site 2, the coal handling plant.

5. Concluding remarks

Our studies have shown that site 2, the Bina coal handling plant, was the most polluted site followed by site 3, the receiving pit of the coal handling plant, site 4, near the main sub-station, site 1, Jawahar colony, and finally site 5, Gharasari village. There are no pollution standards in India or abroad for dust fall. It is important to recognize that the whole area suffers from high levels of pollution, when we compare it to studies published in other regions, as previously mentioned in Section 4. Site 1, the main residential area of the coalmine, had a higher dust fall rate in comparison to site 5. Exposure to elevated concentrations of air pollutants causes adverse human health effects (Hall, 1996), therefore the local inhabitants are living in unhealthy conditions that could result in health problems (Moorcraft and Laxen, 1990). The pollution control measures used by the mining authorities are inadequate, and urgent action is required to remediate the pollution problem. The residential areas around coalmines should be shifted farther away in the opposite direction of prevalent winds.

Acknowledgments

The authors are thankful to the University Grants Commission, New Delhi, for financial assistance to first author (S.K.P) in the form of Senior Research Fellowship and the mining authorities for providing necessary facilities during the study. Special thanks to Dr. Brian Ridley, National Centre for Atmospheric Research, Boulder, USA, for editing the English. The authors also acknowledge the constructive criticism of the reviewers in improving the final version of the manuscript.

References

Adams, S.J., 1997. Dust deposition and measurement: a modified approach. Environmental Technology 18, 345–350.

Banerjee, S.K., Dhar, R.K., Ghose, M.K., 1996. Air pollution due to coal washery projects and its abatement measures. Environmental Management 20, 235–240.

Brooks, K., Schwar, M.J.R., 1987. Dust deposition and the soiling of glossy surfaces. Environmental Pollution 43, 129–141.

Chate, D.M., Pranesha, T.S., 2004. Field studies of scavenging of aerosols by rain events. Journal of Aerosol Science 35, 695–706.

Chaulya, S.K., 2004. Assessment and management of air quality for an opencast coal mining area. Journal of Environmental Management 70, 1–14.

Dhar, B.B., 1994. Changing environment scenario in mining industry. Journal of Mines, Metals and Fuels, 309–314.

Elliot, S., Blake, D.R., Rowland, F.S., 1999. New directions: rapid industrialization in developing countries: the challenge to earth system, research for the new millennium. Atmospheric Environment 33 (4), 683–684.

Fuglsang, K., 2002. An automatic sampler for measurement of dust deposition rates around fugitive sources. Journal of Air and Waste Management Association 52, 789–795.

Gao, Y., Fan, S.M., Sarmiento, J.L., 2003. Aeolian iron input to the ocean through precipitation scavenging: a modeling perspective and its implication for natural iron fertilization in the ocean. Journal of Geophysical research 108 (D7), 4221.

Ghose, M.K., 1989. Pollution due to air borne dust particles in coal mining, its monitoring and abatement measures. Mine Technology 10, 91–95.

Ghose, M.K., 2002. Air pollution due to opencast coal mining and the characteristics of air-borne dust—an Indian scenario. International Journal of Environmental studies 59 (2), 211–228.

Ghose, M.K., Banerjee, S.K., 1995. Status of air pollution caused by coal washery project in India. Environmental Monitoring and Assessment 38, 97–105.

Ghose, M.K., Banerjee, S.K., 1997. Physico-chemical characteristics of air-borne dust emitted by coal washery in India. Energy Environment Monitor 13, 11–16.

Ghose, M.K., Majee, S.R., 2000a. Assessment of dust generation due to opencast coal mining—an Indian case study. Environmental Monitoring and Assessment 61, 255–263.

Ghose, M.K., Majee, S.R., 2000b. Assessment of the impact on the air environment due to opencast coal mining—an Indian case study. Atmospheric Environment 34, 2791–2796.

Ghose, M.K., Majee, S.R., 2001. Air pollution caused by opencast mining and its abatement measures in India. Journal of Environmental Management 63, 193–202.

Ghose, M.K., Sinha, D.K., 1990. Air pollution control plan in coal mining areas. Indian Journal of Environmental Protection 10, 752–756.

Hall, D.J., Upton, S.L., Marsland, G.W., 1993. Improvements in dust gauge design. In: Couling, S. (Ed.), Measurements of Airborne Pollutants. Butterworth-Heinemann, London.

Hall, J.V., 1996. Assessing health effects of air pollution. Atmospheric Environment 30, 743–746.

Harrison, R.M., 1986. Analysis of particulate pollutants. In: Harrison, R.M., Perry, R. (Eds.), Handbook of Air Pollution Analysis. Chapman & Hall, London.

Irvine, K.N., Murray, S.D., Drake, J.J., Vermette, S.J., 1989. Spatial and temporal variability of dry dustfall and associated trace elements: Hamilton, Canada. Environmental Technology Letters 10, 527–540.

Lyons, T.J., Scott, W.D., 1990. Principles of Air Pollution Meteorology. Belhaven Press, London.

Mark, D., Hall, D.J., 1994. Recent developments in airborne dust monitoring. Clean Air 23, 193–217.

Middleton, N., 1997. Desert dust. In: Thomas, D.S.G. (Ed.), Arid Zone Geomorhology: Process, Form and Change in Drylands. Wiley, Chichester, UK, pp. 413–436.

Moorcraft, J.S., Laxen, D.P.H., 1990. Assessment of dust nuisance. Environmental Health, 215–217.

Recheis, M.C., Kihil, R., 1995. Dust deposition in Southern Nevada and California, 1984–1989: relations to climate, source area and lithology. Journal of Geophysical Research 100, 8893–8918.

Reynolds, L., Jones, T.P., Berube, Wise, H., Richards, R., 2005. Toxicity of airborne dust generated by opencast coal mining. Geoscience 67 (2), 141.

Rybicka, E.H., 1989. Metals and their chemical and mineralogical forms in industrial pollutants of the atmosphere. Environmental Technology Letters 10, 921–928.

Schwar, M.J.R., 1998. Nuisance dust deposition and soiling rate measurements. Environmental Technology 19, 223–229.

Shah, R.K., Shah, C.B., 1964. Dustfall and sootfall study in Ahmedabad area. Current Science 22, 679.

Stern, A.C., 1976. Air Pollution: Measuring, Monitoring and Surveillance of Air Pollution. Academic Press, New York.

Tripathi, B.D., Tripathi, A., Mishra, K., 1991. Atmospheric dustfall deposits in Varanasi city. Atmospheric Environment Part B 25 (1), 109–112.

Tripathi, B.D., Chaturvedi, S.S., Tripathi, R.D., 1996. Seasonal variation in ambient air concentration of nitrate and sulphate aerosols in a tropical city, Varanasi. Atmospheric Environment 30 (15), 2773–2778.

Vallack, H.W., Shillito, D.E., 1998. Suggested guidelines for deposited ambient dust. Atmospheric Environment 32 (16), 2737–2744.

Wanquan, T., Honglang, X., Jianjun, Q., Zheng, X., Gensheng, Y., Tao, W., Xiaoyou, Z., 2004. Measurements of dust deposition in Gansu Province, China, 1986–2000. Geomorphology 57, 41–51.

Xuan, J., Liu, G.L., Du, K., 2000. Dust emission inventory in northern China. Atmospheric Environment 3, 1767–1776.

Ziomas, I.C., Melas, D., Zerefos, Ch.S., Bais, A.F., 1995. Forecasting peak pollutant levels from meteorological variables. Atmospheric Environment 29, 3703–3711.

ELSEVIER

Journal of Environmental Management 86 (2008) 139–147

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Effect of reservoir flushing on downstream river water quality

Se Woong Chung[a,*], Ick Hwan Ko[b], Yu Kyung Kim[a]

[a]Department of Environmental Engineering, Chungbuk National University, 12 Gaeshin-Dong, Heungduk-Gu, Cheongju, 361-763, Republic of Korea
[b]Korea Institute of Water and Environment, Korea Water Resources Corporation, 462-1 Jeonmin-Dong, Yusung-Gu, Daejeon, 305-735, Republic of Korea

Received 26 November 2005; received in revised form 22 November 2006; accepted 28 November 2006
Available online 22 January 2007

Abstract

The effect of short-term reservoir flushing on downstream water quality in the Geum River, Korea was studied using field experiments and computer simulations. The reservoir release was increased from 30 to 200 m^3/s within 6 h for the purpose of this experiment. The flushing discharge decreased the concentrations of soluble nitrogen and phosphorus species considerably, but the experimental results revealed a negative impact on organic forms of nutrients and biochemical oxygen demand (BOD). A dynamic river water quality model was applied to simulate the river hydraulics and water quality variations during the event. The model showed very good performance in predicting the travel time of flushing flow and the variations of dissolved forms of nitrogen and phosphorus constituents. However, it revealed a limited capability in simulating organic forms of nutrients and BOD because it does not consider the re-suspension mechanism of these constituents from sediment during the wave front passage.

Keywords: Reservoir flushing; Water quality control; Unsteady water quality model; KORIV1-WIN; Ammonia nitrogen

1. Introduction

The river water quality below Daecheong Dam, which is located in the Geum River of Korea, deteriorates every drought season from December to February due to loading from polluted urban passing tributaries. The most significant water quality problem is the large increase in ammonia nitrogen (NH_3-N) concentration which reaches as high as 2.0–5.0 mg/L and exceeds the regulated water quality standard (0.5 mg/L) for drinking water during that period (Chung and Kim, 2004). The occurrence of high NH_3-N concentrations continues until it arrives at a drinking water intake site, although it is located 68 and 50 km downstream from the confluences of the tributaries, because the natural flow from the watershed dramatically decreases and water temperature drops during winter.

Since the self-purification capacity of the Geum River is significantly influenced by the amount of upstream reservoir release, it is important and necessary to have integrated operations of the river and reservoir systems, taking into account not only water quantity but also downstream water quality and ecosystem health (Ko et al., 2003, 2004). Some previous studies (Tanaka et al., 2004; Green, 1998; Malatre and Gosse, 1995; Barillier et al., 1993; Martin et al., 1986) reported that flushing discharge could improve the hydrological and ecological regimes as well as water quality below a reservoir. Thus a flushing discharge from Daecheong dam has been suggested as an operational alternative that may improve the downstream water quality and mitigate the ammonia nitrogen problem during drought season. However, the impact of reservoir flushing on its downstream water quality may be diverse depending upon the specific environments of the targeted river, water quality constituents considered, and different reservoir operation scenarios. Furthermore, there have been very limited studies and science-based information about the effect of reservoir flushing on the downstream water quality in most of the rivers in Korea.

Field experiments can provide useful data to evaluate the effectiveness of flushing discharge, but they are costly and difficult to implement under various flushing scenarios. Meanwhile, unsteady river water quality modeling techniques can be effectively used to support the dam and river

*Corresponding author. Tel.: +82 43 261 3370; fax: +82 43 272 3370.
E-mail address: schung@chungbuk.ac.kr (S.W. Chung).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.11.031

system operations through providing useful information on the relationship between dam outflow and downstream water quality variations (Bedford et al., 1982; Martin, 1986; Martin and McCutcheon, 1999; Martin and Wool, 2002). It is expected that a better water quality assessment could be made in the Geum River below Daecheong Dam if relations between reservoir flushing rate and river water quality are adequately determined. Thus the objectives of this study were: (1) to investigate the effect of short-term reservoir flushing on the downstream water quality variations through field experiments in the Geum River; and (2) to apply and evaluate an unsteady river water quality model that was developed for supporting reservoir operations by providing simulated information on the downstream hydraulics and water quality changes under various flushing scenarios.

2. Materials and methods

2.1. Descriptions of site and water quality problem

Daecheong dam is a multipurpose dam that has been operated for water supply, hydropower generation, and flood control since 1981. It supplies 922,000 m^3 of municipal water per day to about 2 million residences in Daejeon, Cheongju, and surrounding cities (Fig. 1). Since there is no selective withdrawal facility in the dam, the water release for hydropower generation is withdrawn at the middle layer (EL. 52.0 m) of the reservoir where the penstock inlet is located. During the experiment, the sampling of reservoir release was conducted at the actual outflow in the re-regulation reservoir. As presented in Table 1 the water quality of the reservoir release is very good, but a significant water quality degradation was noticed in the downstream river reach after the confluence of the Gap and Miho tributaries which receive large-scale wastewater treatment discharges from the urban areas in the vicinity including Daejeon (700,000 m^3/d) and Cheongju (250,000 m^3/d). Currently, the regulated outflow water quality standards for the wastewater treatment plants are 20 mg/L for 5-day biochemical oxygen demand (BOD), 60 mg/L for total nitrogen, and 8 mg/L for total phosphorus. The occurrence of high ammonia concentrations in the tributaries during the winter season was found to be induced by inhibition of nitrification processes due to low water temperature.

In particular, a seasonal occurrence of high NH_3-N concentrations has hampered chemical treatment processes of a water treatment plant (S-WTP in the Fig. 1) whose river water intake is 78 km downstream from the dam. Fig. 2 shows the trend of daily ammonia nitrogen concentrations measured at the intake tank of the S-WTP for selected years. The plant operators have experienced great difficulties with the chemical treatment process because more chlorine should be applied to treat the ammonia for breakpoint chlorination, which may cause creation of carcinogenic byproducts such as Tri-halomethane and Halo-acids. It was suggested that some degree of water quality deterioration can be attenuated by optimal scheduling of reservoir release from Daecheong dam if the relationship between outflow and downstream water quality is adequately identified.

The dam was not designed to supply minimum base flow for downstream water quality. The reservoir release occurs only through the penstock for hydropower generation during peak-time demand, and spillway discharge for flood control. The released water is typically stored in the re-regulation reservoir so that it can be constantly supplied to the downstream. The monthly operation schedule is determined by the Reservoir Operation Center of Korea Water Resources Corporation based on the current reservoir water level and projected inflow amount. The amounts of typical reservoir release for the past 24 years are presented in Table 2. The amount of reservoir release is

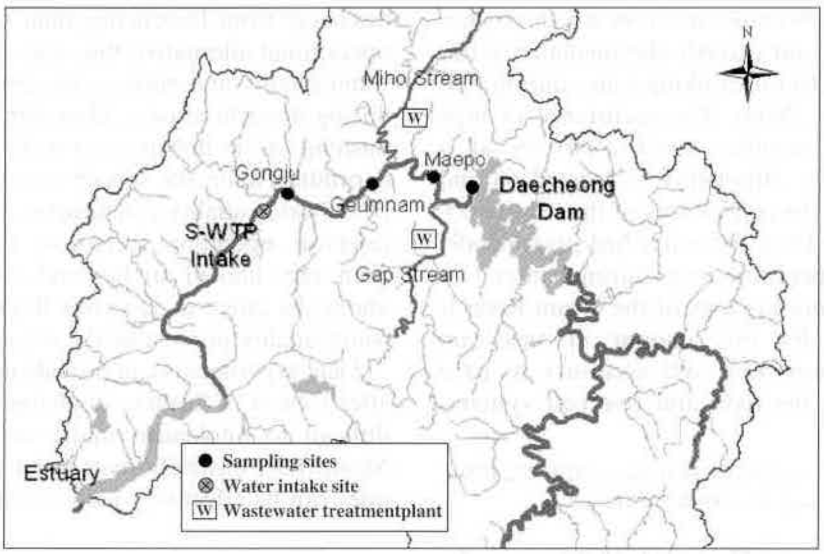


Fig. 1. Study area map and location of sampling sites.

Table 1
The flow and water quality data observed during the flushing event for reservoir release and tributary streams

Site	Mean flow (m^3/s)	Value	Temp. (°C)	BOD_5 (mg/L)	Org-N (mg/L)	NH_3-N (mg/L)	NO_3-N (mg/L)	Org-P (mg/L)	PO_4-P (mg/L)
Reservoir release	200.0	Mean	14.27	0.76	0.718	0.050	0.951	0.171	0.006
		Max.	14.85	0.92	1.332	0.060	0.979	0.175	0.010
		Min.	13.64	0.52	0.399	0.040	0.934	0.165	0.005
		Stdev.	0.49	0.15	0.327	0.008	0.015	0.003	0.002
Tributary Gap stream	11.7	Mean	8.80	2.80	0.221	6.636	3.675	0.270	0.673
		Max.	10.70	3.88	1.059	7.120	3.741	0.337	0.788
		Min.	7.30	2.02	0.105	5.890	3.600	0.163	0.626
		Stdev.	1.48	0.79	0.508	0.459	0.067	0.066	0.066
Tributary Miho stream	13.6	Mean	7.73	3.64	1.850	2.205	2.010	0.283	0.275
		Max.	9.10	4.09	2.014	2.390	2.085	0.407	0.321
		Min.	6.40	3.07	1.733	1.920	1.972	0.233	0.149
		Stdev.	0.96	0.34	0.098	0.156	0.041	0.064	0.065

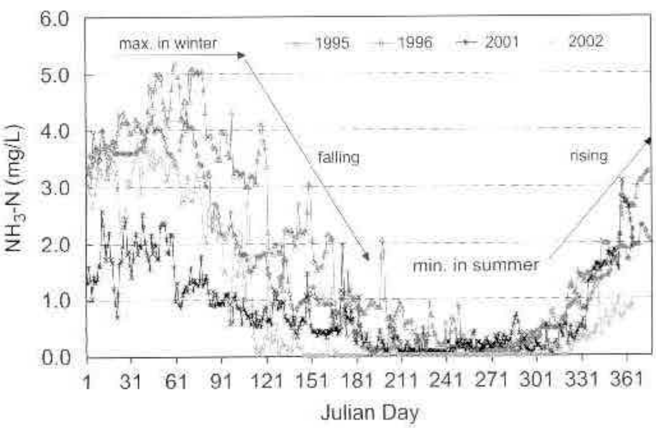


Fig. 2. Seasonal variations of NH_3-N measured at S-WTP intake site in the Geum River.

relatively small during January and February when the ammonia nitrogen concentration is highest in the downstream (Fig. 2). This is the major reason why we did this study and suggested that reservoir flushing could be employed to mitigate downstream water quality problems.

2.2. Field experiments

The field experiment for assessing the impact of short-term reservoir flushing on the river water quality below the dam was conducted on November 22, 2003. The dam outflow was increased from 30 to 200 m^3/s within 6 h for the purpose of the experiment. The reservoir water was withdrawn at the middle layer of the reservoir, EL. 52.0 m where the penstock inlet is located. During the passage of the flushing flow, water samples were taken every hour or two hours at the outlet of the re-regulation reservoir, Maepo and Geumnam gauge stations (Fig. 1). The water quality variables including pH, water temperature, and DO were measured on site using a portable water analysis system (YSI 6600). Other constituents including biological oxygen demand (BOD_5), organic nitrogen (Org-N), ammonia nitrogen (NH_3-N), nitrate nitrogen (NO_3-N), organic phosphorus (Org-P), and soluble reactive phosphate (PO_4-P) were measured through laboratory analyses based on the standard methods (APHA, 1998).

The water quality concentrations measured at the dam outflow and tributaries including the Gap and Miho streams during the flushing event are presented in Table 1. The water quality concentrations observed at the tributaries were much higher than those of the reservoir discharge (see Table 3).

2.3. Unsteady river water quality model

An unsteady river water quality model KORIV1-WIN was used as a tool for evaluating the impact of reservoir flushing on the downstream water quality (Chung, 2004). KORIV1-WIN is a user-friendly one-dimensional hydrodynamic and water quality model that was based on the version 2.0 of CE-QUAL-RIV1 (Environmental Laboratory, 1995). The model can simulate hydraulics of unsteady river flows and the interactions of 16 water quality variables including water temperature, nitrogen species, phosphorus species, dissolved oxygen, carbonaceous oxygen demand, algae, iron, manganese, coliform bacteria and two arbitrary constituents. Although the water quality algorithms for the pollutants simulated by the model are relatively comprehensive, in comparison with other available river water quality models, there are a number of limitations. In particular, the model does not include sediment transport processes such as scour and deposition and their impact on water quality.

3. Results and discussion

3.1. Hydrodynamic results

The hydrographs and peak times of reservoir flushing release, and observed flow at the downstream Maepo and Gongju gauging stations are presented and compared with

Table 2
Statistic summary of monthly water release of the reservoir for last 24 years

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Mean	37.1	38.2	46.1	58.3	78.6	102.0	192.7	196.5	161.0	60.8	46.3	46.5
Max.	75.3	79.3	108.5	142.4	152.6	309.0	714.2	644.4	412.9	141.7	102.0	89.3
Min.	12.9	14.1	13.7	20.1	13.1	29.5	10.6	20.9	20.6	12.7	10.0	11.9

Table 3
The observed travel times of flushing discharge at selected gauging stations

Time	Description	Gauging station		
		Maepo (9.77 km)[a]	Geumnam (28.85 km)	Gongju (45.68 km)
T_a	Time when flushing front arrived	11/22 07:00	11/22 12:00	11/22 14:00
T_p	Time when peak flow occurred	11/22 11:00	11/22 18:00	11/22 20:00
T_e	Time when flushing tail passed	11/22 22:00	11/23 05:00	11/23 10:00

[a]Distance from Dam.

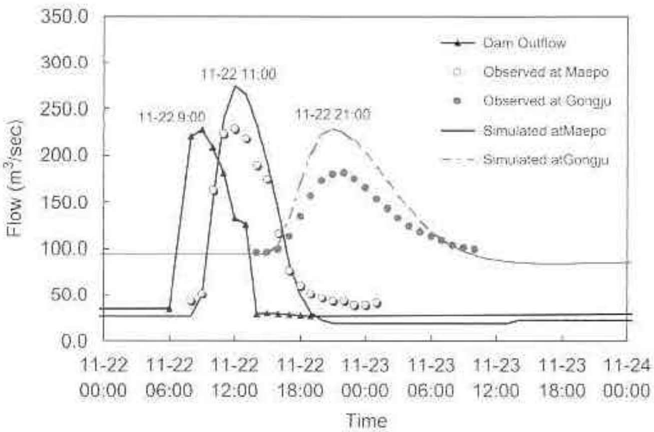


Fig. 3. Comparisons of observed and simulated hydrographs at selected gauging stations.

simulated results in Fig. 3. The travel times from the dam to Maepo and Gongju were approximately 2 and 21 h, respectively (Table 3). Apparently the hydrodynamic model showed very good performance in predicting the wave propagation such as dampening and spreading out of the hydrograph, and travel time of flushing flow. However, it is noticed that the magnitudes of peak flow and total amounts of flushing flow that arrived at both control points were somewhat over estimated by the model. The reason for such deviations is that some portion of the flushing flow was lost in the channel by detention and infiltration due to dry riverbed conditions.

3.2. *Effects of flushing on the river water quality and model performance*

The observed and simulated water quality variations during the experiment are shown in Figs. 4–6. The two dashed lines show the arrival time of the wave front and the end time of the flushing event, respectively. The solid line shows the time when the peak flow occurred. The model performance was evaluated using absolute mean error (AME), correlation coefficient (r), and coefficient of determination (r^2) between the observed and simulated values; these are presented in Table 4.

The effects of short-term flushing from Daecheong dam on the downstream water quality were quite different depending on the specific constituent. The wave front passage of flushing flow caused a sharp increase in BOD and a slight decrease in DO concentrations in the river, possibly due to the re-suspension of organic matterfrom the river bottom sediments (Fig. 4). Barillier et al. (1993) also found similar results from their experimental reservoir water release in the upper Seine River where the water front caused an increase in particulate organic matter concentrations and biomass of the organisms. The water temperature was increased from 10 °C to 13–14 °C because the reservoir release water was withdrawn from the middle layer of the reservoir.

The simulated results are compared with observed data with ±20% error bars to take account of the uncertainty of diverse input variables. Apparently, the model showed satisfactory performance in water temperature and DO simulations. However it revealed poor performance in simulating the dynamic changes in BOD concentrations during the event because the re-suspension of organic matter was not adequately accounted for in the model. The current version of the model is not able to simulate the re-suspension of organic materials by turbulent scour and its impact on water quality as most available river water quality models cannot.

The concentration changes of nitrogen species, Org-N, NH_3-N, NO_3-N, and T-N are presented in Fig. 5. The results revealed that all nitrogen concentrations except

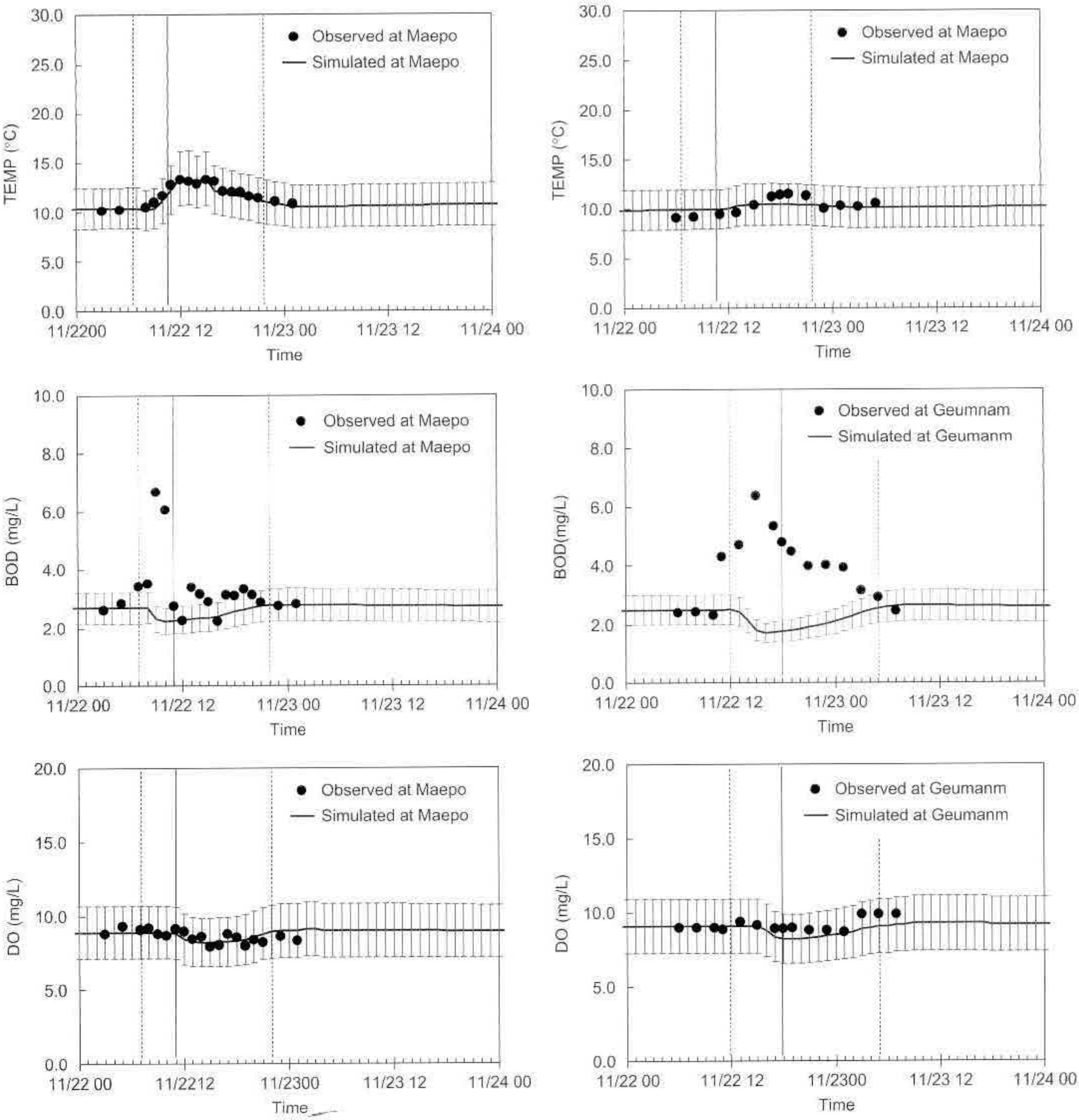


Fig. 4. Comparisons of observed and simulated water temperature, BOD, and DO at Maepo and Geumnam.

organic nitrogen are fairly sensitive to the reservoir flushing flow. The level of NH_3-N concentration was most sensitive to the flushing event and considerably decreased from 1.8 and 2.0 mg/L at Maepo and Geumnam respectively before the flushing to a value less than 0.5 mg/L during the passage of the flushing flow. However, the dilution effectiveness continued only for the time when the reservoir water was being passed.

The water quality model showed satisfactory performance in capturing the effect of the flushing event on the dilutions of NH_3-N, NO_3-N, and T-N concentrations and its duration of effectiveness. However, a large deviation was found between observed and simulated concentrations of Org-N. As observed for BOD, the Org-N suddenly increased when the wave front of flushing flow arrived at Maepo station due to the disturbance of the river bottom

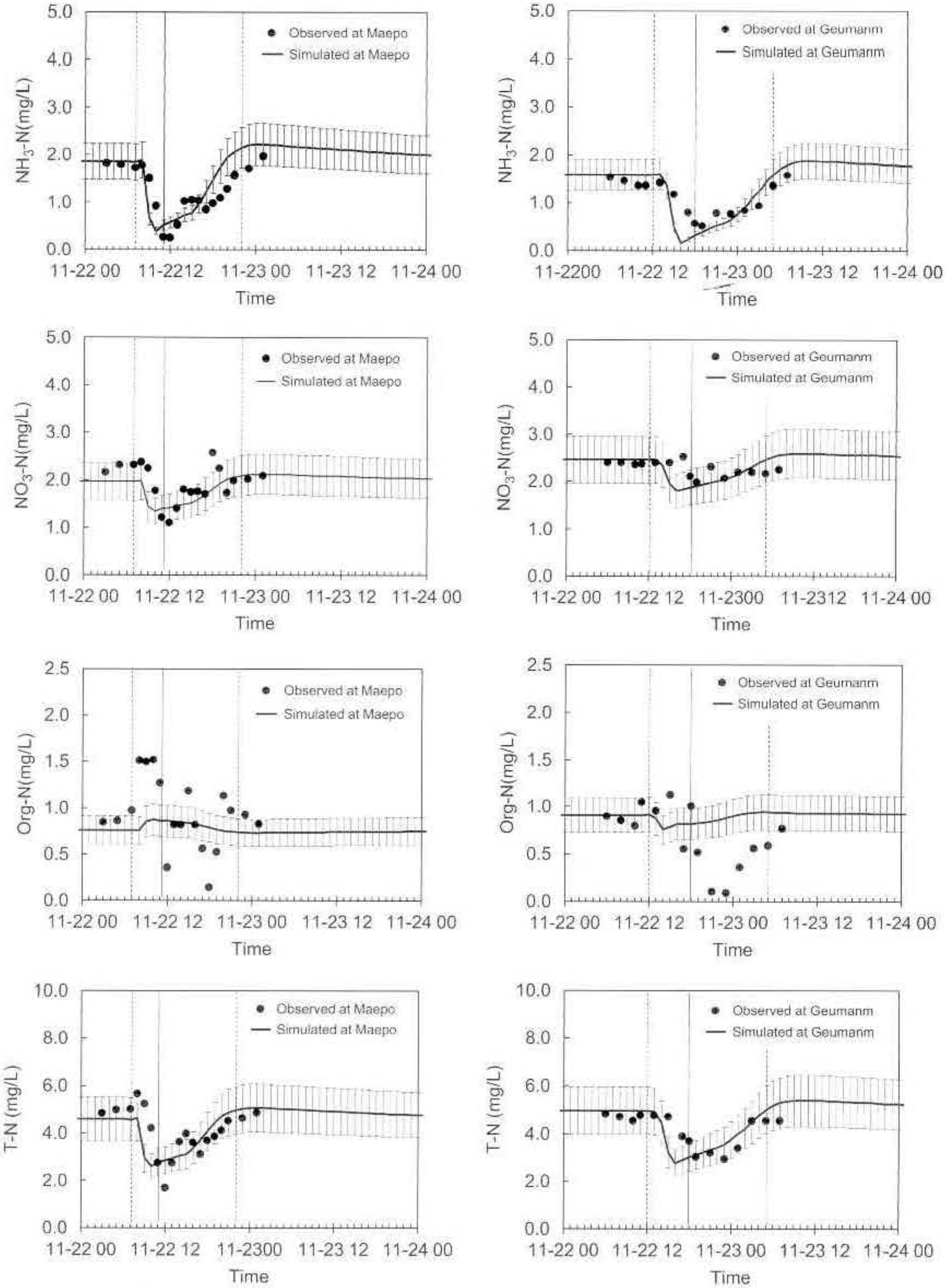


Fig. 5. Comparisons of observed and simulated nitrogen species at Maepo and Geumnam.

sediments. This caused a great deal of fluctuationin Org-N concentrations during the flushing event. The degree of fluctuation was decreased at Geumnam, and even a notable reduction in Org-N concentration was observed after the peak flow passed.

The observed and simulated concentration variations in phosphorus constituents including Org-P, PO_4-P, and T-P during the flushing event are presented in Fig. 6. It can be seen that the dissolved form of phosphorus, PO_4-P, was much more sensitive and was significantly decreased by the

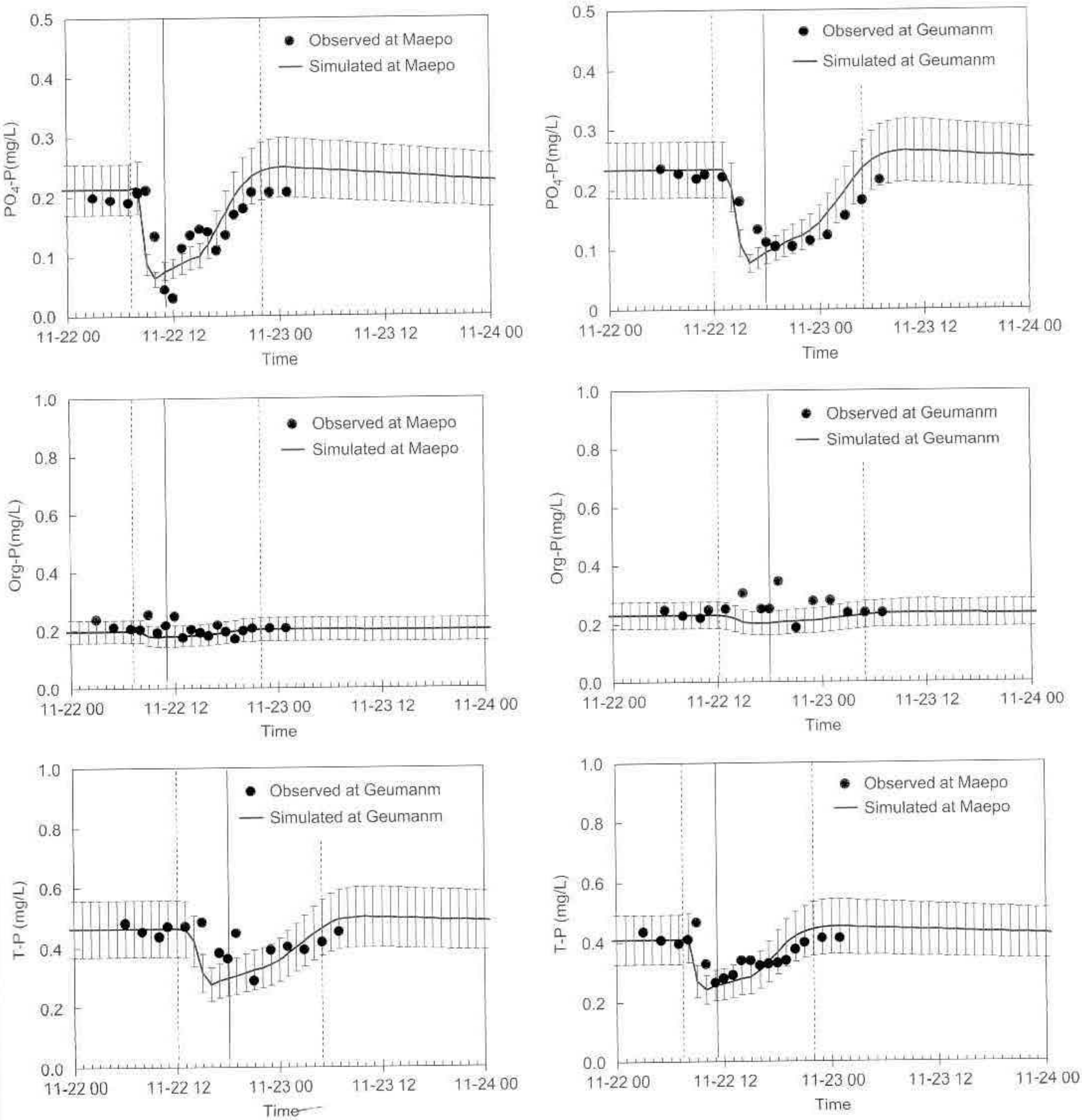


Fig. 6. Comparisons of observed and simulated phosphorus species at Maepo and Geumnam.

flushing flow. The PO_4-P concentrations decreased from 0.2 mg/L to less than 0.05 mg/L at Maepo, and from 0.24 to 0.1 mg/L at Geumnam. Unlike the large fluctuations of Org-N concentrations, the variation of Org-P concentrations during the flushing passage was relatively small possibly because the major phosphorus component is in dissolved form rather than particulate form in the river system. The unsteady river water quality model simulated the dilution effect on soluble reactive phosphate quite well. And the observed concentrations of all phosphorus species were mostly located within the $\pm 20\%$ error bars of the simulated values.

3.3. *Effect of alternative flushing rates on ammonia nitrogen concentrations*

Although the unsteady model revealed a limited capability in replicating the re-suspension of sediment

Table 4
Statistics summary showing the evaluation results of model performance

Site	Index	Temp	BOD	DO	Org-N	NH_3-N	NO_3-N	Org-P	PO_4-P
Maepo	AME[a]	0.32	0.79	0.32	0.29	0.33	0.28	0.02	0.04
	r[b]	0.94	−0.38	0.49	0.17	0.79	0.79	−0.10	0.73
	r^2[c]	0.89	0.14	0.24	0.03	0.63	0.63	0.01	0.54
Geumnam	AME	0.52	1.53	0.41	0.27	0.23	0.20	0.03	0.03
	r	0.91	−0.81	0.55	0.02	0.91	0.28	−0.48	0.87
	r^2	0.83	0.66	0.30	0.00	0.82	0.08	0.23	0.76

[a]Absolute mean error (AME) $= \sum_{i=1}^{n} |y_{si} - y_{oi}|$, where y_s and y_o are simulated and observed values, respectively.
[b]Pearson correlation coefficient.
[c]Coefficient of determination.

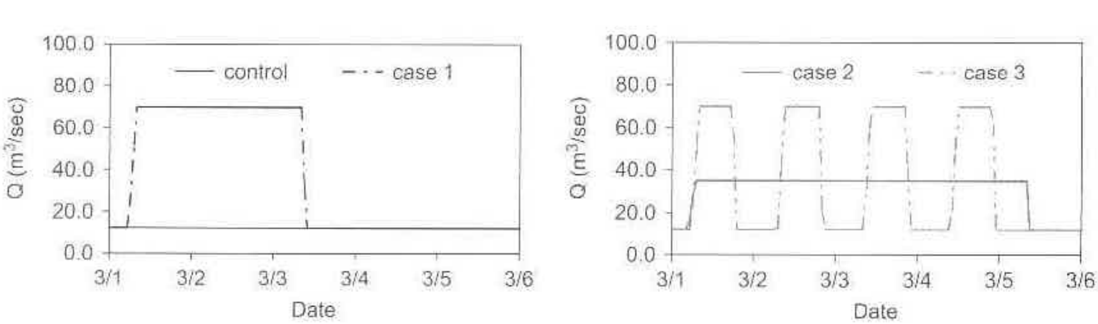


Fig. 7. Alternative dam flushing flow cases to assess their effect on NH_3-N concentrations.

during the passage of the wave front, it showed relatively good performance in simulating the water quality variations of dissolved forms of nitrogen and phosphorus over the period of the flushing event. Thus the model was applied to assess the effect of various alternative flushing rates on the mitigation of ammonia nitrogen concentrations using the field data obtained in March, 2002. Fig. 7 shows the alternative cases of flushing flow discharge from the Daecheong dam. The control was a real condition with no flushing in which the amount of reservoir release was $12\,m^3/s$. The other three alternative cases consisted of different discharge intervals and flushing rates, while the total amounts of release were all the same (12 million m^3). Under cases 1 and 2, the flushing discharges were supplied for 48 and 96 h with constant amounts of 69.4 and $34.7\,m^3/s$, respectively. An intermittent flushing discharge of $69.4\,m^3/s$ with 12 h intervals was assumed in case 3. The initial NH_3-N concentration was 3.5 mg/L in the river 78 km from where the S-WTP intake site is located. The boundary NH_3-N concentrations for reservoir release, Gap stream, and Miho stream were 0.86, 8.08, and 3.35 mg/L, respectively.

Fig. 8 shows the time series of simulated NH_3-N concentrations 78 km downstream from the dam in response to the four different flushing alternatives. All cases showed increasing concentrations because it took more than 10 days for the NH_3-N to reach steady state conditions. Without supplying the flushing discharge

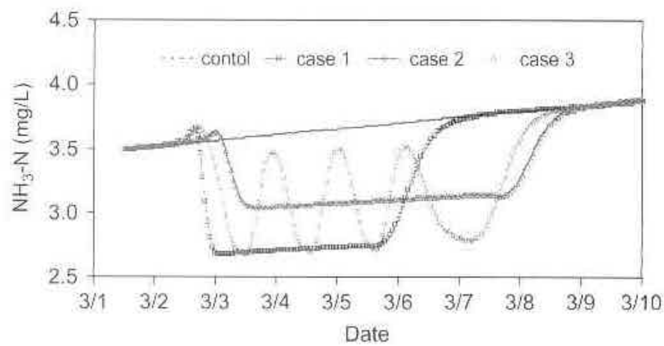


Fig. 8. The dynamic variations of NH_3-N concentrations in response to alternative flushing flow cases.

(control), the NH_3-N concentrations increased up to 3.8 mg/L on March 9, 2002. For cases 1 and 2 the NH_3-N concentrations were maintained at 2.7 mg/L for 68 h and 3.0 mg/L for 106 h, respectively and recovered smoothly up to the level of the control. The intermittent flushing discharge (case3) caused sinusoidal variations in which the NH_3-N concentrations were between those for the control and case 1.

In addition, the amount of reservoir discharge necessary to keep the NH_3-N less than 2.0 mg/L, which is the level expected by the plant operators for proper water treatment, at intake the site was determined under the assumption of steady state boundary conditions. The simulated NH_3-N

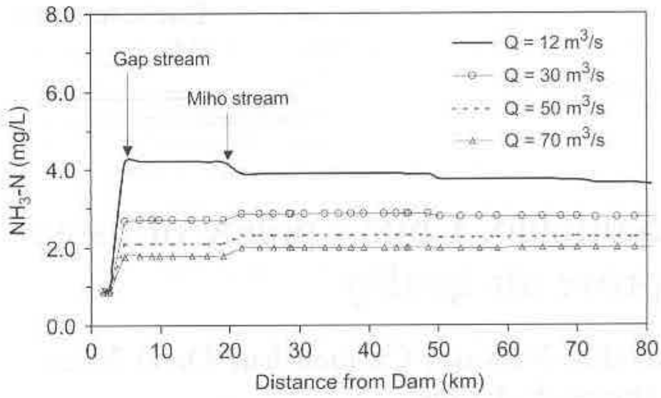


Fig. 9. Downstream NH_3-N concentrations for different reservoir releases under steady-state conditions.

concentrations with different amounts of reservoir release along the distance from the dam are presented in Fig. 9. It is apparently required to supply more than $50\,m^3/s$ to maintain the NH_3-N at less than 2.0 mg/L.

4. Conclusions

The effects of short-term reservoir flushing discharge from Daecheong dam on the downstream water quality during drought season were found to be quite different depending on the type of constituent. It was very effective in reducing the concentrations of dissolved forms of nitrogen (NH_3-N and NO_3-N) and phosphorus (PO_4-P) species, while it had an adverse impact on organic matter and BOD. In particular, the flushing flow caused an increase in BOD concentration in the river, possibly due to the re-suspension of organic matter from sediments. The unsteady river water quality model, KORIV1-WIN, showed very stable performance in simulating the wave propagation and travel time of the reservoir release under various flushing alternatives. It also satisfactorily captured the effect of reservoir flushing on the reductions of dissolved forms of nitrogen and phosphorus concentrations. However, the model needs some improvement to incorporate the re-suspension mechanism of organic matter from bottom sediment during the wave front passage for better assessing the flushing impact on the Geum River.

In conclusion, the suggested flushing using the surplus reservoir water can be an effective management technique for reducing nitrogen concentrations downstream at a drinking water intake during a short duration period in the dry winter months. Although the available amount of flushing water is totally dependent on the hydrological conditions, it is currently sufficient because the water demand is about 40% less than the water supply capacity of the reservoir.

Acknowledgments

This research was supported by a grant (code 1-6-2) from Sustainable Water Resources Research Center of 21st Century Frontier Research Program.

References

APHA, AWWA and WEF, 1998. Standard Methods for the Examination of Water and Wastewater, 20th ed. American Public Health Association, Washington D.C., USA, pp. 4-103–4-126.

Barillier, A., Garnier, J., Coste, M., 1993. Experimental reservoir water release: impact on the water quality on a river 60 km downstream (upper seine river, France). Water Research 27 (4), 635–643.

Bedford, K.W., Sykes, R.M., Libicki, C., 1982. A Dynamic Water Quality Model for Stormwater Assessment. US Army Corps of Engineers, Waterway Experiment Station, MS, USA.

Chung, S.W., 2004. Application of an unsteady river water quality model for the analysis of reservoir flushing effect on downstream water quality. Journal of KWWA 37 (10), 857–868 (in Korean).

Chung, S.W., Kim, J.H., 2004. Development of water quality models for supporting NH_3-N control in a dam regulated river. In: Proceedings of the Fourth IWA World Water Congress and Exhibition, 19–24 September, 2004, Marrakech, Morocco, p. 223.

Environmental Laboratory, 1995. CE-QUAL-RIV1: A Dynamic, One-Dimensional (Longitudinal) Water Quality Model for Streams User's Manual, US Army Corps of Engineers, Waterway Experiment Station, MS, USA.

Green, W.R., 1998. Relations between reservoir flushing rate and water quality. 〈http://aslo.org/phd/dialog/1998January-13.html〉.

Ko, I.H., Chung, S.W., Hwang, M.H., Kim, W., 2003. Decision support tools for integrated water resources management in Korea. In: Proceedings of APHW2003: First International Conference on Hydrology and Water Resources in Asia Pacific Region 13–15 March, 2003, Kyoto, Japan, pp. 385–389.

Ko, I.H., Chung, S.W., Maeng, S.J., 2004. A new technical framework for integrated river system operations in Korea. In: Proceedings of the Fourth IWA World Water Congress and Exhibition, 19–24 September, 2004, Marrakech, Morocco, pp. 203–204.

Malatre, K., Gosse, Ph., 1995. Is it possible to influence water temperature and quality in the river Seine upstream of Paris in summer by managing the upstream reservoir? Water Science and Technology 31 (8), 67–77.

Martin, J.L., 1986. Water quality study of proposed regulation dam downstream of Wolf Creek Dam, Cumberland River, Kentucky, Miscellaneous Paper EL-86-4, US Army Corps of Engineers, Waterway Experiment Station, MS, USA.

Martin, J.L., McCutcheon, S.C., 1999. Hydrodynamics and Transport for Water Quality Modeling. CRC Press, Inc., Boca Raton, FL.

Martin, J.L., Wool, T., 2002. A dynamic one-dimensional model of hydrodynamics and water quality EPD-RIV1 User's Manual. Georgia Environmental Protection Division, Atlanta, Georgia, USA.

Martin, J.L., Curtis, L.T., Nestler, J.M., 1986. Effects of flow alterations on trout habitat in the Cumberland River below Wolf Creek Dam, Kentucky, Miscellaneous Paper EL-86-11, US Army Corps of Engineers, Waterway Experiment Station, MS, USA.

Tanaka, N., Osugi, T., Nanami, Y., Okano, M., 2004. Methods of environmental restoration for downstream of dams. in: Proceedings of the Environmental Considerations for Sustainable Dam Projects, ICOLD 72nd Annual Meeting, May 16–22, 2004, Seoul, Korea, p. 45.

Journal of Environmental Management 86 (2008) 148–157

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Analyzing the cost effectiveness of Santiago, Chile's policy of using urban forests to improve air quality

Francisco J. Escobedo[a,*], John E. Wagner[b], David J. Nowak[c], Carmen Luz De la Maza[d], Manuel Rodriguez[d], Daniel E. Crane[c]

[a] University of Florida, IFAS-School of Forest Resources and Conservation, Building 164, Mowry Road, P.O. Box 110806, Gainesville, FL 32611-0806, USA
[b] SUNY-ESF 304 Bray Hall, Syracuse, NY 13210, USA
[c] USDA Forest Service Northeastern Research Station 5 Moon Library, Syracuse, NY 13210, USA
[d] Universidad de Chile Facultad de Ciencias Forestales, Manejo de Recursos Forestales, Santa Rosa 11315, Santiago, Chile

Received 3 March 2003; received in revised form 25 November 2006; accepted 28 November 2006
Available online 1 February 2007

Abstract

Santiago, Chile has the distinction of having among the worst urban air pollution problems in Latin America. As part of an atmospheric pollution reduction plan, the Santiago Regional Metropolitan government defined an environmental policy goal of using urban forests to remove particulate matter less than 10 μm (PM_{10}) in the *Gran Santiago* area. We used cost effectiveness, or the process of establishing costs and selecting least cost alternatives for obtaining a defined policy goal of PM_{10} removal, to analyze this policy goal. For this study, we quantified PM_{10} removal by Santiago's urban forests based on socioeconomic strata and using field and real-time pollution and climate data via a dry deposition urban forest effects model. Municipal urban forest management costs were estimated using management cost surveys and Chilean Ministry of Planning and Cooperation documents. Results indicate that managing municipal urban forests (trees, shrubs, and grass whose management is under the jurisdiction of Santiago's 36 municipalities) to remove PM_{10} was a cost-effective policy for abating PM_{10} based on criteria set by the World Bank. In addition, we compared the cost effectiveness of managing municipal urban forests and street trees to other control policies (e.g. alternative fuels) to abate PM_{10} in Santiago and determined that municipal urban forest management efficiency was similar to these other air quality improvement measures.

Keywords: Cost-effective analysis; Urban forest management; Air pollution abatement; Street trees; Ecosystem services

1. Introduction

Urban forests (trees, shrubs, and grass) provide many ecosystem services that benefit human well-being (Beckett et al., 1998; McPherson et al., 1999; Nowak et al., 2002; Scott et al., 1998; Ulrich, 1986; WRI, 2001; Yang et al., 2005). Some of these services include improved human health, community empowerment, climate modification, recreational benefits, wildlife habitat, wood, food, and aesthetics (Dwyer et al., 2003; Gutiérrez, 2000; Intendencia Región Metropolitana, 1987; Murray, 1996a, b; Ulrich, 1986; World Bank, 1994). Several Latin American cities, among them Santiago, Chile; Mexico City, Mexico; and São Paolo, Brazil, are integrating trees and other vegetation as part of urban environmental improvement programs, policies, and measures.

Santiago, Chile has the distinction of having among the worst urban air pollution problems in Latin America despite having a steady improvement in air quality over the last 10 years (SESMA, 2000; World Bank, 1994, 1997). The city is located 450–900 m above sea level in a basin surrounded by 2000 m tall mountain ranges. These geographic conditions contribute to thermal inversions and restricted air flow through the basin that aggravate air

*Corresponding author. Tel.: +1 352 378 2159; fax: +1 352 376 4536.
E-mail addresses: fescobed@ufl.edu (F.J. Escobedo), jewagner@esf.edu (J.E. Wagner), dnowak@fs.fed.us (D.J. Nowak), cdlamaza@uchile.cl (C.L. De la Maza), mrodrigu@uchile.cl (M. Rodriguez), dcrane@fs.fed.us (D.E. Crane).

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.11.029

quality problems. The *Gran Santiago* airshed comprises a study area of 967 km^2 located in the northernmost section of a basin referred to as the *Valle Central* and contains well over 5 million residents—nearly 40% of Chile's population. The *Gran Santiago* Metropolitan Area is Chile's administrative, cultural, and industrial center that encompasses residential areas, industrial and commercial districts, transportation networks, agricultural lands, *espinal* shrub-lands, and Andean piedmont. Santiago's semi-arid climate and urbanized environment also poses a major limitation for the establishment of trees in Santiago (Escobedo, 2004; Escobedo et al., 2006). As a result, the current urban forest cover has to be attributed to active management by its human inhabitants (Escobedo, 2004).

Santiago has an existing *Plan de Prevención y Descontaminación Atmosférica* (PPDA; Atmospheric Prevention and Decontamination Plan) that is part of the Chilean Environmental Commission's *Ley de Bases Generales del Medio Ambiente* (Law of General Environmental Baselines) that defines a policy goal of using street trees and green areas to reduce the emissions of particulate matter less than 10 μm (PM_{10}) in the *Gran Santiago* metropolitan region (CONAMA, 1997, 2001; CONAMA-RM, 2002).[1] Street trees are trees within the right of way or easement of any major or minor thoroughfare. Green areas refer to parks, plazas, large medians, squares or any vegetated public or open access area administered by the municipality or other public entity. Laws and ordinances have established the administrative infrastructure for the management of Santiago's street trees and green areas under the jurisdiction of Santiago's 36 *comuna's*, departments of *Aseo y Ornato* (Waste management and landscaping) (Ceballos Ibarra, 1997; Escobedo, 2004; Escobedo et al., 2006). A *comuna* is an autonomous municipality with its own mayor, council, budget and department of *Aseo y Ornato*. As there is no legal definition of an urban forest in either of these laws or ordinances, for convenience we will describe a municipal urban forest (MUF) as trees, shrubs, and grass (i.e. street trees and green areas) whose management is under the jurisdiction of Santiago's 36 *comunas* and an urban forest as all trees, shrubs, and grass within the *Gran Santiago* metropolitan region.

The 36 *comuna*s are currently allocating part of their municipal budgets to manage their MUFs. However, we found no published analysis examining whether MUF management in Latin America is in fact a cost-effective policy for reducing PM_{10}. McPherson et al. (1998) reported that planting residential shade trees for air quality benefits is not cost effective in California, USA. However, Nowak et al. (1998) rebut this conclusion based on the limited scale of analysis and methods used in McPherson et al. (1998) study. The focus of this study was on municipal urban forests because private expenditures on tree maintenance were not readily available given the time frame and budget of this research. Thus, the research question investigated by this study is whether managing Santiago's MUFs are cost effective in reducing PM_{10} concentrations.

[1]The PPDA policy does not mention any other pollutants, hence only PM_{10} will be analyzed.

2. Materials and methods

2.1. Policy analysis model

To analyze the effectiveness of urban forest management for air quality improvement, any analysis model must compare urban forest management with other public investment alternatives. By using this type of approach as a component of the urban forest-decision making framework; economic, social, political, and environmental factors can be weighed against one another and in doing so assist the decision maker in selecting the best alternative. The cost-effective analysis policy model used in this study is a specific type of approach in which the goal of a policy is defined, the threshold costs of obtaining that goal are established, and then the most efficient alternatives are selected (Field, 1997; Larson et al., 1999). As opposed to a cost-benefit analysis, a cost-effective analysis by its more limited frame of reference permits an analyst to compare and advocate policies by quantifying costs in monetary units and effects in units of functions or services (Dunn, 1981; Poister, 1978; Portney, 2000). In doing so the analyst determines *how* the resources should be used and not *whether* they should be used to meet a policy objective in a technologically efficient manner (Larson et al., 1999; Poister, 1978).

Estimating the cost effectiveness of Santiago's policy to use MUFs to remove PM_{10} will require developing a quantitative relationship between the urban forest's ability to remove PM_{10} and its management costs. Determining a direct relationship between MUF management and air quality can be difficult (Brimblecombe, 2001; Krupnick and Portney, 1991). However, a vegetation cover-atmosphere process can be used as a link between MUF cover and air quality. The amount of pollution that vegetation cover removes per unit time is a function of dry deposition velocity (V_d meters per second (m/s)) or the rate at which vegetation cover "removes" a pollutant from the atmosphere given an ambient pollutant concentration (C grams per cubic meter (g/m^3)). By calculating the dry deposition velocity of MUFs and determining ambient pollutant concentration, pollutant flux (F) or removal can be calculated ($F = V_dC$ ($g/m^2/s$)) (Davidson and Wu, 1990; Fowler, 2002; Lovett, 1994; McPherson et al., 1998; Nowak et al., 2002; Scott et al., 1998). Therefore, given that the existing MUF cover in Santiago is the result of purposeful management, by quantifying Santiago's MUFs structure and modeling its ability to remove PM_{10} combined with the management costs of maintaining that cover we will be able to estimate the costs of abating PM_{10}. If this cost estimate is less than a threshold described by the World Bank (1994), then managing Santiago's MUFs are cost effective in reducing PM_{10} concentrations.

Table 1
Santiago demographics

Socioeconomic strata	Area (km^2)	Average annual per capita income (US$2000)	Population (2000)[a]	Population density (pop/km^2)
High	164.9	10 000	773 633	4692
Medium	370.3	4000	1 924 767	5198
Low	431.9	1250	2 823 864	6538
Total	967.1		5 522 264	5710

Source: ICCOM-Novaction (2004) and Instituto Nacional de Estadistica-Chile statistics.
[a]Includes both rural and urban inhabitants within the *comunas*.

Santiago's 36 *comunas* are self-governing municipalities with their own mayor, council, municipal budgets, and MUF management programs. Their demographic and socioeconomic characteristics are different. Consequently, they were divided into three socioeconomic strata based on ICCOM-Novaction (2004) classifications (Escobedo et al., 2006). *Comunas* with 25% of their households in the highest three classifications were defined as the high socioeconomic stratum. *Comunas* with 50% of their household in the middle two classifications were defined as the medium socioeconomic stratum. *Comunas* with 25% of their household in the lowest two classifications were defined as the low socioeconomic stratum (see Table 1 and Fig. 1) (Escobedo, 2004).

2.2. Quantifying urban forest structure

To quantify Santiago's urban forest structure 200, 0.04 ha random, permanent, circular plots were distributed among the three socioeconomic strata proportional to tree cover area: 74, 62, and 64 plots were allocated to the high, medium, and low socioeconomic strata, respectively following standard UFORE methods (Escobedo, 2004; Escobedo et al., 2006). This resulted in a sampling intensity of less than 1% of each stratum's urban forest cover. The plots centers were located by applying a random number generator of x and y coordinates per stratum using a geographic information system (GIS: ARCVIEW 3.2 with spatial analyst extension) and 1:10 000 black and white, digital ortho-photographs across public and private property within the study area.[2] When plot access permission was not given or the plot was inaccessible (approximately 5% of all plots), the plot was relocated in the immediate area within the same land use and general surface cover characteristics. Specifically, the next parcel in a clockwise direction was selected until access was possible and marked on ortho-photograph. The plot was relocated in the same relative position on the parcel as the original plot.

The urban forest field data were collected during January and February 2002. The data recorded from each plot included land use, percent grass, and other ground cover. Shrubs were identified to the species level and measured for height, percent of shrub mass volume occupied by leaves, and percent of total shrub area occupied by the shrub mass. Trees whose stem center was located within the plot and had a minimum diameter at breast height of 2.54 cm, had the following information recorded: species, number of stems, height, height to base of live crown, crown widths along a north–south axis and an east–west axis, percent dieback, percent foliage density, and indication if the tree was located on a street or green area and hence managed by the municipality or other public entity (Nowak et al., 2003; Escobedo et al., 2006). Tree, shrub, and grass cover were quantified independently thereby accounting for spatial overlap.

These data were incorporated into the Urban Forest Effects (UFORE) model to quantify urban forest structure (e.g. leaf area, leaf cover, leaf area index, evergreen leaf composition, and leaf biomass) (Nowak and Crane, 2000). The UFORE model was developed by the USDA Forest Service to quantify urban forest structure and function and aid in urban forest management. In general, the urban forests in the high socioeconomic stratum were in better condition than those in the medium and low socioeconomic strata. The medium and low socioeconomic strata were characterized by relatively larger, older, isolated trees in generally poor condition (Escobedo, 2004).

The model estimated tree density, leaf area index, leaf biomass and other parameters (Escobedo et al., 2006). For this study however, the urban forest structure parameter of interest is leaf area (m^2)[3]. Table 2 gives the MUF cover by socioeconomic strata. The low and medium socioeconomic stratum's MUF cover is greater than the high socioeconomic stratum's even with the medium and low socioeconomic stratum's urban forests in poorer condition than those in the high socioeconomic stratum. This difference is because the low and medium socioeconomic strata encompassed nearly 80% of the study area (Tables 1 and 2).

[2]Because no ortho-photographs were available for 2 of the 36 *comunas*, only 34 *comunas* were used for this analysis. The 2 *comunas* were in the low socioeconomic stratum.

[3]Because leaf area (m^2) can easily be measured, leaf area index, tree density, and leaf type which are important parameters in pollution deposition, are not discussed in the analysis because they are already incorporated into the model (see Escobedo (2004) for discussion of the role of these parameters in pollution removal).

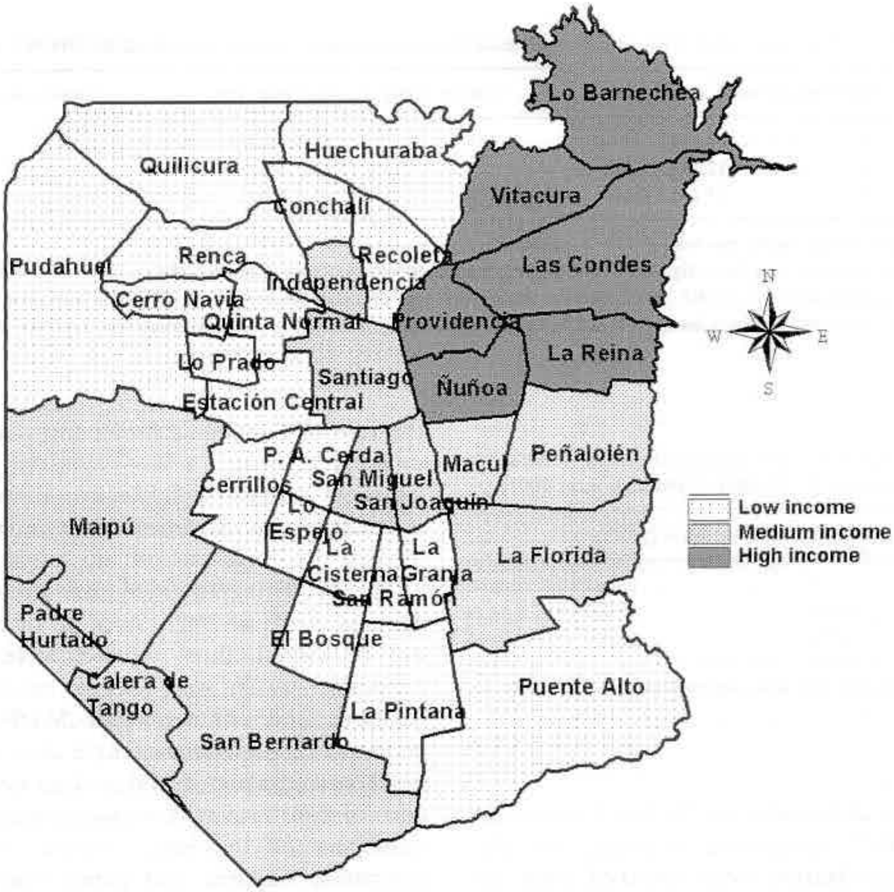


Fig. 1. *The Gran Santiago's* 36 *comunas* and three different socioeconomic strata.

Table 2
Municipal urban forest covers by socioeconomic strata

Socioeconomic strata	Tree cover[a] (m^2)	Shrub cover[b] (m^2)	Grass cover[b] (m^2)	Total MUF cover[c] (m^2)
High	12 517 620	10 487 640	26 388 000	49 393 260
Medium	17 414 804	14 515 760	55 992 400	87 922 964
Low	32 830 110	13 820 800	40 076 608	86 727 518

[a]UFORE calculated cover based on actual field measurements.
[b]The proportion of municipal shrub and grass cover was not measured in the field. Using professional judgment, municipal shrub and grass cover was assumed to be 40% of total shrub and grass cover as calculated by UFORE based on field measurements.
[c]MUF is municipal urban forest (trees, shrubs, and grass whose management is under the jurisdiction of Santiago's 36 municipalities) cover.

2.3. Modeling PM_{10} removal rates

Annual PM_{10} removal rates for the period July 2000–June 2001 were calculated by UFORE based on the MUF structure, hourly weather data, and hourly ambient PM_{10} concentrations.[4] The weather data were obtained from the *La Platina* weather station in the *comuna* of *La Pintana*. The MACAM-2 monitoring network (SESMA, 2000) was used to obtain hourly pollutant concentration data stratified by socioeconomic stratum. Missing hourly PM_{10} concentration data were estimated using the monthly average for the specific hour and particulate matter resuspension accounted for (Nowak and Crane, 2000). Hourly ambient PM_{10} concentrations for the high socio-

[4]During periods of precipitation, pollution is removed via wet deposition and dry deposition is not occurring; therefore, pollution removal by urban forest cover was set to zero during periods of precipitation. The model assumes a 50% resuspension rate of PM_{10} back to the atmosphere based on Zinke (1967).

Table 3a
Annual PM_{10} removal rates for municipal urban trees, shrubs, and grass by socioeconomic strata as calculated by UFORE (July 2000–June 2001)

Socioeconomic strata	Tree removal rates ($g/m^2/yr$)[a]		Shrub removal rates ($g/m^2/yr$)		Grass removal rates[c] ($g/m^2/yr$)	
High	7.5	(2.9–11.7)[b]	8.5	(3.3–13.3)[b]	1.3	(1.2–4.7)
Medium	7.4	(2.3–11.5)	5.7	(2.2–8.8)	1.7	(0.9–4.6)
Low	8.0	(3.1–12.4)	5.8	(2.3–9.1)	1.8	(1.2–5.0)

[a] $g/m^2/yr$ denotes grams per square meter per year.
[b] Ranges are based on reported low and high deposition velocities from the literature (Nowak et al., 2002).
[c] The grass removal rates calculated by UFORE, are based on the lowest, tree dry deposition velocity from the literature (Nowak et al., 2002). The ranges are the low and high tree removal rates divided by 2.5 based on research by Shreffler (1978) for grass SO_2 deposition velocities.

Table 3b
Total annual PM_{10} removal rates for municipal urban forests by socioeconomic strata as calculated by UFORE (July 2000–June 2001)

Socioeconomic strata	Total removal rates ($g/m^2/yr$)	
High	17.3	(7.5–29.7)[a]
Medium	14.8	(5.4–24.9)
Low	15.6	(6.6–26.5)

[a] Ranges are based on the low and high deposition velocities given in Table 3a.

economic stratum were obtained from the *Las Condes* and *Providencia* MACAM-2 monitoring stations; for the medium socioeconomic stratum were obtained from *La Florida*, *La Paz*, and *Parque O'Higgins* MACAM-2 monitoring stations; and for the low socioeconomic stratum were obtained from the *Pudahuel*, *Cerrillos*, and *El Bosque* MACAM-2 monitoring stations. The mean annual ambient PM_{10} concentrations for the study period were 59.1, 78.9, and 84.4 micrograms per cubic meter ($\mu g/m^3$) for the high, medium, and low socioeconomic stratum, respectively. As a point of comparison, in 1995 the cities of Santiago, São Paolo, Bogotá, and Mexico City had average PM_{10} levels of 109, 105, 70, and 87 $\mu g/m^3$, respectively (World Bank, 1997).

The annual PM_{10} removal rates for MUFs by socioeconomic strata are shown in Tables 3a and 3b. Unfortunately, there are no grass PM_{10} deposition velocities reported in the literature. However, the estimates of grass removal rates, as reported by the UFORE model are based on tree dry deposition velocities. Shreffler (1978) states that "observations and predictions indicate the deposition velocity over a forest will be 2–3 times as great as over grass." Therefore, the grass removal ranges are the low and high tree removal rates divided by 2.5.

Annual PM_{10} removal rates for MUF in the high socioeconomic stratum were greater than for MUFs in the medium and low socioeconomic strata. There was however variability among tree, shrub, and grass removal rates for the three strata. Differences in cover, density, leaf area index, composition, and pollution concentrations among the strata also accounted for differences in pollution removal. Escobedo (2004) discusses the role of Santiago's urban forest structure in pollution removal.

2.4. Estimating municipal urban forest management costs

The MUF management cost data were collected from January to April 2002. Since all of the 36 *comunas* could not be visited, three representative *comunas* per socioeconomic stratum were selected based on existing working relations and contacts with MUF managers of those *comunas*. These *comunas* were also representative of the MUF management, social, and economic characteristics of each of their respective socioeconomic stratum. The MUF managers of the nine *comunas* were interviewed to determine budgets and expenditures and management and maintenance activities of MUF. Expenditures included annual variable and fixed investment in the management of MUFs as reported by the managers; such as the direct and indirect costs of capitol, labor, and operation activities such as administration, personnel, equipment, tree maintenance activities (e.g. pruning, planting, transplants), shrub and turf maintenance, watering, fertilization, infrastructure improvement, hazard tree damages, and sidewalk construction and repair (see Escobedo et al. (2006) for detailed list and discussion of the cost items included in this analysis).

During the interview, the managers filled out a self-administered questionnaire with the interviewer (Poister, 1978). The questionnaire was left with the manager to permit the acquisition of additional accounting information. However, most questions were answered during the interview. A final visit was scheduled to complete the questionnaire. Total municipal budgets were determined using data from nine separate *Chilean Ministerio de Planificación y Cooperación* (Chilean Ministry of Planning and Cooperation) documents (MPC, 2000).

Table 4 gives the average percent of the sampled *comuna's* annual budget allocated to MUF management by socioeconomic stratum. Due to different cost accounting methods, inconsistent definitions of costs, and differing bureaucratic levels reporting expenditures, the accuracy of cost estimates cannot be determined. For example, concurrent interviews with one *comuna's* central administrative office (*Secretaría de Planificación Comunal*) and the

Table 4
Socioeconomic strata's 2000 budget allocated to municipal urban forests management

Socioeconomic strata	Municipal urban forest management expenditure (%)		Municipal urban forest management expenditure (US$/m²)[c]	
High	3.6[a]	(1.4–4.2)[b]	0.19	(0.08–0.23)[d]
Medium	3.8	(1.4–4.2)	0.12	(0.04–0.13)
Low	3.0	(1.4–4.2)	0.12	(0.06–0.17)

[a]The average percent of the total *comuna*'s budget allocated to street trees and green areas.
[b]Low and high ranges represent the lowest and highest percentages reported on the survey.
[c]A 2000 average monthly "Reference Exchange Rate" of 550 Ch$ = 1 US$ was used (Banco Central de Chile, 2005).
[d]Based on the lowest and highest percentages reported on the survey.

MUF management department (*Aseo y Ornato*) resulted in different line item expenditures and thus different reported expenditures for MUF management activities. To address this problem, ranges based on the low and high budget expenditures were also defined (Table 4).

2.5. Cost-effective analysis

The World Bank (1994) conducted an economic analysis of environmental problems in Chile. In their analysis of the benefits to health in Santiago from PM_{10} reduction policies and measures, their results indicated that controls reducing PM_{10} emission at a cost below US$ (1994) 18 000/ton/PM_{10} "should be considered worthwhile and a reasonable threshold value for evaluating air pollution controls". Adjusting this value for inflation to the year 2000 results in a PM_{10} control threshold value of 25 000 US$/ton/$PM_{10}$ (Banco Central de Chile, 2002). Therefore, MUF management costs of less than 25 000 US$/ton/$PM_{10}$ will be considered cost effective.

Calculating each socioeconomic stratum's MUF management costs per ton of PM_{10} removed (US$/ton/$PM_{10}$) is a three-step process. First, each stratum's budget allocated to MUF management in US$ is estimated using

$$F_s = \sum_{i=1}^{S} B_{is}\beta_s, \tag{1}$$

where F_s is the budget allocated to MUF management (US$) for socioeconomic stratum s (i.e. high, medium, and low) B_{is} the ith *comuna's* total budget in socioeconomic stratum s, S the number of *comunas* in each socioeconomic stratum, and β_s the percent of the socioeconomic stratum's budget allocated to MUF management (Table 4). Second, the MUF management cost per square meter of municipal tree, shrub, and grass cover by socioeconomic stratum is estimated using

$$C_s = \frac{F_s}{TC_s + SC_s + GC_s} \quad \text{for all } s, \tag{2}$$

where C_s is the annual MUF management cost per square meter (US$/m²) of municipal tree shrub and grass cover by socioeconomic stratum s; TC_s, SC_s, and GC_s are the municipal tree, shrub, and grass cover in square meters by socioeconomic stratum, s, respectively (Table 2). Finally, each socioeconomic stratum's MUF management cost per ton of PM_{10} removed is estimated using Eq. (3):

$$A_s = \left(\frac{C_s}{TR_s + SR_s + GR_s}\right)\theta \quad \text{for all } s, \tag{3}$$

where A_s is the MUF management costs per ton of PM_{10} removed (US$/ton/$PM_{10}$) for socioeconomic stratum s; TR_s, SR_s, and GR_s are the annual PM_{10} removal rates by municipal trees shrubs and grass by socioeconomic stratum, s, respectively (Table 3a); and θ converts grams to metric tons.

Table 5
Cost per ton of PM_{10} removed by municipal urban forest's by socioeconomic strata

Socioeconomic strata	Municipal urban forest (US$/ton/$PM_{10}$)		
High	11 185	(4350–13 050)[a]	(6515–26 150)[b]
Medium	8147	(3002–9005)	(4843–22 330)
Low	7861	(3669–11006)	(4628–18 581)

[a]Ranges based on low and high urban forestry budget allocations given in Table 4.
[b]Ranges based on low and high deposition velocities given in Table 3b.

3. Results

Table 5 shows the cost per ton of PM_{10} removed by MUFs for each socioeconomic stratum. The low socioeconomic stratum's MUFs were the most cost effective at 7861 US$/ton/$PM_{10}$ and the high socioeconomic stratum's MUFs were the least cost effective at 11 185 US$/ton/$PM_{10}$. This difference was due primarily to the MUF cover (Table 2) used in calculating the MUF management cost per square meter (Eq. (2)) and the stratum's PM_{10} concentration used to estimate annual PM_{10} removal rates used in Eq. (3). The medium and low socioeconomic stratum's MUF cover was approximately 1.8 times larger than the high socioeconomic stratum's MUF cover

(Table 2). This caused the MUF management cost per square meter for the medium and low socioeconomic stratum's to be less than that of the high socioeconomic stratum even though the high socioeconomic stratum allocates a larger percent of their municipal budget to MUF management expenditures (Table 4). Finally, national and regional government work and tree planting programs and subsidies might be lowering overall costs in the lower socioeconomic stratum (Escobedo, 2004).

The MUFs management costs per ton of PM_{10} removed in each socioeconomic stratum were below the \$25 000 threshold set by the World Bank. Due to the inaccuracies in the MUF expenditure information summarized in Table 4, we also calculated the ranges of MUF management cost per ton of PM_{10} removed (Table 5). Again, the MUF management cost per ton of PM_{10} removed in each socioeconomic stratum was below the \$25 000 threshold set by the World Bank. In addition, we calculated the MUF management cost ranges per ton of PM_{10} removed based on the low and high annual PM_{10} removal rates given in Tables 3a and 3b. Using the lowest PM_{10} removal rates, the high socioeconomic stratum's MUF management cost per ton of PM_{10} removed by MUFs was greater than the World Bank threshold indicating that MUF management might not be cost effective in this socioeconomic stratum if the removal rate was at the lowest end of its expected range.

Table 6
Cost per ton of PM_{10} removed by street trees by socioeconomic strata

Socioeconomic strata	Street trees (US\$/ton/PM_{10})		
High	7125	(2441–13 636)[a]	(4567–18 426)[b]
Medium	9889	(3209–17 909)	(6364–25 235)
Low	8100	(3352–18 711)	(5226–20 903)

[a]Ranges based on low and high street tree budget allocations (Escobedo, 2004).

[b]Ranges based on low and high deposition velocities given in Table 3a.

Escobedo et al. (2006) also summarized management expenditure and cover information for street trees. Using this information and Eqs. (1)–(3), we examined if managing for municipally owned street trees was cost effective in reducing PM_{10} concentrations. As shown by Table 6, the management of street trees was cost effective in removing PM_{10}. The only exception was in the case of the medium socioeconomic stratum based on the low annual PM_{10} removal rates indicating that street tree management might not be cost effective in this socioeconomic stratum if the removal rate was at the lowest end of its expected range.

A variety of Chilean PM_{10} control devices measures and policies' based on studies conducted by Eskeland (1997), O'Ryan (1993) and the World Bank (1994) were compared against MUF management costs to determine if MUF management's cost efficiency was similar to these other air quality improvement measures and technologies (Fig. 2). MUF management costs in all of Santiago's socioeconomic strata were within the costs of these measures. Only the regulation of light duty gas vehicle emission standards had costs greater than the threshold value of 25 000 US\$/ton/PM_{10}.

4. Discussion

The procedure developed for this analysis presents an innovative, simple, straightforward approach to examine the effectiveness of managing MUFs for air quality improvement within the confines of existing policies. The

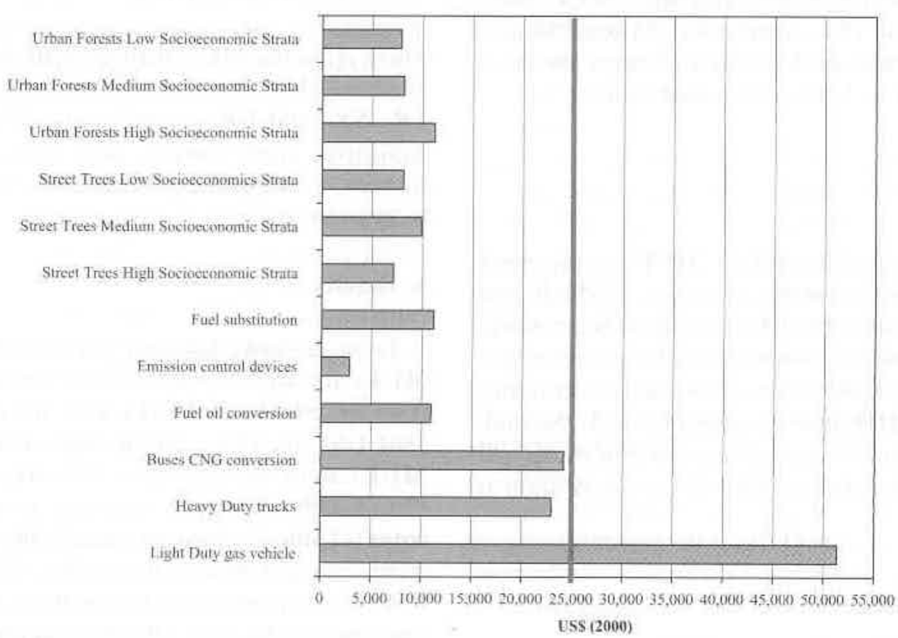


Fig. 2. Cost effectiveness of several PM_{10} abatement policies in Santiago, Chile. *Source*: World Bank (1994) and O'Ryan (1993). CNG, compressed natural gas; heavy duty trucks, regulation of heavy-duty truck emissions; light duty gas vehicles, regulation of light-duty vehicle emissions.

analysis indicates that managing MUFs and street trees are a cost-effective approach for abating PM_{10} in Santiago, Chile according to the PPDA. There are, however, two main caveats. First, the conclusions are based on UFORE estimates of MUF structure and annual PM_{10} removal rates (Tables 2, 3a, and 3b). The UFORE model has been used in previous studies to estimate urban tree and shrub effects on air quality (Nowak et al., 2002; Yang et al., 2005). However, the effect of grass on PM_{10} removal has not been well studied. Consequently, the PM_{10} removal rates and ranges for grasses, while based on the best available information, are more subjective than those for trees and shrubs. Even so, the estimated PM_{10} removal rates for MUFs are likely conservative as urban trees have other effects (e.g. reducing air temperatures, building energy use, and other air pollutants) not accounted for in this analysis. Second, the conclusions are based on the estimates of MUF management costs (Table 4). Given the nature of the cost data, the cost estimates probably overestimate the actual MUF management costs (Escobedo, 2004). Thus, the cost-effective estimates in Tables 5 and 6 are conservative. We have attempted to address both these issues by including a range analysis of both the removal rates and the cost data. Given these caveats, the conclusions of this study are tenable.

The results from this study indicate that in the case of Santiago, Chile urban forests are a cost-effective air quality improvement policy. That said, even if urban forests were not cost effective, urban trees can provide additional environmental benefits, for example, in their potential to sequester carbon and modify climate at no additional management costs (Escobedo, 2004). Urban vegetation also has additional environmental costs. Trees and shrubs can emit biogenic, volatile organic compounds (e.g. isoprenes, monoterpenes, and other organic compounds) that in combination with nitrogen oxides and under certain climate conditions, can contribute to ozone formation (Chameides et al., 1988). Urban vegetation also produces pollen which can aggravate allergies and emits carbon dioxide through maintenance activities and decomposition (Escobedo, 2004). Accounting for these additional environmental and economic benefits and costs was beyond the scope of this analysis.

5. Conclusion

Previous experiences from other parts of the world indicate that as low-cost options for air quality improvement are implemented, the costs of further reduction will increase (Hall, 1995; Maynard, 2001). Once these current technologies and policies have been implemented and exhausted, then the burden will fall on individual's behaviors and other more diffuse sources, thereby complicating air quality improvement programs (Krupnick and Portney, 1991).

As Chile integrates citizen participation in its environmental policies (e.g. the management of privately controlled urban forests), the opportunity presents itself for applying the pollution removal function of urban forests to encourage the political integration of its citizenry and local governments in the improvement of environmental quality. The metrics and methods from this study could provide *comunas* flexibility in satisfying environmental ordinances in a cost-effective manner. For example, one possible approach to incorporate urban forest cover within an air quality control program would be to develop a system of tradable permits based on each *comuna's* urban forest PM_{10} abatement potential (Main et al., 2000). Remote sensing protocols for determining urban forest cover could be used to enforce attainment of urban forest cover goals. Pollution removal rates for urban forests, as calculated by UFORE in this study, could be used to determine PM_{10} reduction effects. According to the World Bank (1994), sector arrangements could even be implemented to counteract the negative effects of urbanized *comunas* trading permits with that of peri-urban *comunas* and in doing so account for the discrepancy in pollution emission dynamics in urbanized central areas being treated as equivalent to pollution dynamics in outer non-urbanized areas.

Finally, many of the controversies involved with nonexistent policies, valuation of benefits, time-dependent assumptions, and the complexities of atmospheric physics and chemistry and individual homeowner behavior were circumvented in this analysis. Variables and the methodology used were adjusted for the socioeconomic, environmental, and policy realities of a major Latin American city. Removal rates as quantified by UFORE were based on actual field measurements and real-time meteorological and pollution data. The procedure applied and the results from this study indicate that MUF and street tree management are cost effective in abating PM_{10} within the context of the PPDA and Chile's existing environmental and economic policies. It is hoped that this same procedure can be applied to examine the cost effectiveness of managing urban forest to improve air quality in other cities.

Acknowledgments

We thank the Chilean *Fondo de Fomento al Desarrollo Científico y Tecnológico* FONDEF D00I1078 and the USDA Forest Service Northeastern Research Station for funding the project. At the University of Chile we would like to especially thank Jaime Hernández, Valeska Pizarro, as well as Maria Teresa Serna and the field crew; Pablo Gonzales, Victoria, Cynnamon, Nico, and Pepe for assistance in field measurements and the cost-management survey. We also thank Pablo Ulricksen from CENMA for assembling pollution and boundary height data and the various urban forest managers who participated in the survey. At the USDA Forest Service we thank Wayne Zipperer and Jack Stevens for assistance and review of drafts. At SUNY-ESF we thank Pablo Donoso and Bill Bentley for their advice.

References

Banco Central de Chile, 2002. Consumer Price Indices. Retrieved May 1, 2005: ⟨www.bcentral.cl/esp/infoeconomica/seriesindicadores/⟩.

Banco Central de Chile, 2005. Reference Exchange Rates. Retrieved March 1, 2005: ⟨www.bcentral.cl/esp/infoeconomica/seriesindicadores/⟩.

Beckett, K.P., Freer-Smith, P.H., Taylor, G., 1998. Urban woodlands: their role in reducing the effects of particulate pollution. Environmental Pollution 9, 347–360.

Brimblecombe, P., 2001. Urban air pollution. In: Bristlecombe, P., Maynard, R. (Eds.), The Urban Atmosphere and Its Effects. Air Pollution Reviews, vol. 1. Imperial College Press, London, pp. 1–18.

Ceballos Ibarra, W., 1997. Enverdecimiento Urbano en Chile. In: Krishnamurthy, L., Nacimiento, J. (Eds.), Áreas Verdes Urbanas en Latino América y el Caribe. Banco Interamericano de Desarrollo, Washington, DC, pp. 231–251.

Chameides, W.L., Lindsay, R.W., Richardson, J., Kiang, C.S., 1988. The role of biogenic hydrocarbons in urban photochemical smog: Atlanta as a case study. Science 241, 1473–1475.

CONAMA, 1997. Plan de Prevención y Descontaminación Atmosférica. Comisión Nacional del Medio Ambiente, Santiago, Chile.

CONAMA, 2001. Anteproyecto de revisión, reformulación y actualización del Plan de Prevención y Descontaminación Atmosférica para la región metropolitana. Comisión Nacional del Medio Ambiente, Santiago, Chile.

CONAMA-RM, 2002. Áreas verdes en el Gran Santiago. Gobierno de Chile, Comisión Nacional del Medio Ambiente, Región Metropolitana, Santiago, Chile.

Davidson, C.I., Wu, Y., 1990. Dry deposition of particles and vapors. In: Lindberg, S.E., Page, A.L., Norton, S.A. (Eds.), Acidic Precipitation: Sources, Deposition, and Canopy Interactions. Springer, New York, pp. 103–209.

Dunn, W., 1981. Public Policy Analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Dwyer, J.F., Nowak, D.J., Noble, M.H., 2003. Sustaining urban forests. Journal of Arboriculture 29 (1), 49–55.

Escobedo, F.J., 2004. A cost-effective analysis of urban forest management's role in improving air quality in Santiago, Chile. Ph.D. Dissertation. State University of New York—College of Environmental Science and Forestry, Syracuse.

Escobedo, F., Nowak, D.J., Wagner, J., De la Maza, C.L., Rodriguez, M., Crane, D.E., Hernández, J., 2006. The socioeconomics and management of Santiago de Chile's public urban forests. Urban Forestry and Urban Greening 4, 105–114.

Eskeland, G.S., 1997. Air pollution requires multi-pollutant analysis: the case of Santiago, Chile. American Journal of Agricultural Economics 79 (5), 1636–1641.

Field, B.C., 1997. Environmental Economics: An Introduction, second ed. McGraw Hill, New York.

Fowler, D., 2002. Pollutant deposition and uptake by vegetation. In: Bell, J.N.B., Treshow, M. (Eds.), Air Pollution and Plant Life, second ed. Wiley, New York, pp. 43–67.

Gutiérrez, J., 2000. Silvicultura Urbana: Salvación para Santiago. Chile Forestal 281, 11–18.

Hall, J.V., 1995. Air quality in developing countries. Contemporary Economic Policy XIII (April).

ICCOM-Novaction, 2004. Datos Estadísticos. Hogares e ingresos por grupo socioeconómico: Solo urbano. Retrieved January 10, 2004: ⟨www.iccom.cl/html/info_estadistica/f_inf_estadistica.html⟩.

Intendencia Región Metropolitana, 1987. Seminario de Arborización Urbana. Area Metropolitana, Intendencia Metropolitana, Santiago de Chile.

Krupnick, A.J., Portney, P.L., 1991. Controlling urban air pollution: a benefit-cost assessment. Science 252, 522–527.

Larson, B.A., Avaliani, S., Golub, A., Rosen, S., Shaposhnikov, D., Strukova, E., Vincent, J., Wolff, S., 1999. The economics of air pollution health risks in Russia: a case study of Volgograd. World Development 27 (10), 1803–1819.

Lovett, G.M., 1994. Atmospheric deposition of nutrients and pollutants in North America: an ecological perspective. Ecological Applications 4, 629–650.

Main, M., Canessa, M., Pollicardo, J., Stein, A., 2000. Análisis de Los Procesos de Participación Ciudadana en la Elaboración de Planes de Prevención y Descontaminación y Normas de Calidad Ambiental y de Emisión. Fundación Casa de la Paz. Fondo para el Estudio de las Políticas Públicas, Chile.

Maynard, R.L., 2001. Particulate air pollution. In: Bristlecombe, P., Maynard, R. (Eds.), The Urban Atmosphere and its Effects. Air Pollution Reviews, vol. 1. Imperial College Press, London, pp. 163–194.

McPherson, E.G., Scott, K.I., Simpson, J.R., 1998. Estimating cost effectiveness of residential yard trees for improving air quality in Sacramento, California using existing models. Atmospheric Environment 32, 75–84.

McPherson, E.G., Simpson, J.R., Pepper, P.J., Xiao, Q., 1999. Benefit-cost analysis of Modesto's municipal urban forest. Journal of Arboriculture 25 (5), 235–248.

MPC, 2000. Documento(s) de Información Comunal. Región Metropolitana, Provincia de Santiago, Comuna(s) de: Vitacura, La Reina, Providencia, Santiago, La Florida, San Bernardo, Pudahuel, Renca y La Pintana. Ministerio de Planificación y Cooperación. División de Planificación Regional, Gobierno de Chile.

Murray, S., 1996a. Managing forest influences in urban and peri-urban areas. Unasylva 47 (185), 38–44.

Murray, S., 1996b. Urban and Peri-Urban Forestry in Quito, Ecuador: A Case-study. Forestry Department, Food and Agriculture Organization, Rome.

Nowak, D.J., Crane, D.E., 2000. The Urban Forest Effects (UFORE) Model: quantifying urban forest structure and functions. In: Hansen, M., Burk, T. (Eds.), Integrated Tools for Natural Resources Inventories in the 21st Century: Proceedings of the IUFRO Conference. US Department of Agriculture, Forest Service, North Central Research Station, General Technical Report NC-212, pp. 714–720.

Nowak, D., Cardelino, C.A., Rao, S.T., Taha, H., 1998. Discussion: estimating cost effectiveness of residential yard trees for improving air quality in Sacramento, California using existing models. Atmospheric Environment 32 (14/15), 2709–2711.

Nowak, D.J., Crane, D.E., Stevens, J.C., Ibarra, M., 2002. Brooklyn's Urban Forest. US Department of Agriculture, Forest Service, Northeastern Research Station. General Technical Report NE-290.

Nowak, D.J., Crane, D.E., Stevens, J.C., Hoehn, R.E., 2003. The Urban Forest Effects (UFORE) Model: Field Data Collection Manual. US Department of Agriculture, Forest Service, Northeastern Research Station. Available (February 2006) at: ⟨www.fs.fed.us/ne/syracuse/Tools/downloads/UFORE_Manual.pdf⟩.

O'Ryan R.E., 1993. Cost effective policies to improve urban air quality in developing countries: case study for Santiago, Chile. Ph.D. Dissertation, University of California, Berkeley.

Poister, T.H., 1978. Public Program Analysis: Applied Research Methods. University Park Press, Baltimore.

Portney, P.R., 2000. The contingent valuation debate: why economists should care. In: Stavins, R. (Ed.), Economics of the Environment: Selected Readings, fourth ed. WW Norton and Co., New York, pp. 253–267.

Scott, K.I., McPherson, E.G., Simpson, J.R., 1998. Air pollutant uptake by Sacramento's urban forest. Journal of Arboriculture 24 (4), 224–233.

SESMA, 2000. Red de Monitoreo de Calidad del Aire del SESMA. Servicio de Salud Metropolitano del Ambiente. Retrieved November 25, 2000: ⟨http://www.sesma.cl/ind_con/graf_ses2.htm⟩.

Shreffler, J.H., 1978. Factors affecting dry deposition of SO_2 on forests and grasslands. Atmospheric Environment 12, 1497–1503.

Ulrich, R.S., 1986. Human responses to vegetation and landscapes. Landscape and Urban Planning 13, 29–44.

World Bank, 1994. Chile—managing environmental problems: economic analysis of selected issues. Report No. 13061-CH. The World Bank, Washington, DC, pp. viii, x–xi, 39–41, 50–59, 83–85, 96–98.

World Bank, 1997.Vehicular air pollution: an overview. World Bank Technical Paper No. 373. The World Bank, Washington, DC.

WRI, 2001. Urban ecosystems. In: World Resources 2000–2001: People and Ecosystems: The Fraying Web of Life. UNDP, World Resources Institute, Washington, DC, pp. 141–145.

Yang, J., McBride, J., Zhou, J., Sun, Z., 2005. The urban forest in Beijing and its role in air pollution reduction. Urban Forestry and Urban Greening 3, 65–78.

Zinke, P.J., 1967. Forest interception studies in the United States. In: Sopper, W.E., Lull, H.W. (Eds.), Forest Hydrology. Pergamon Press, Oxford, pp. 137–161.

ELSEVIER

Journal of Environmental Management 86 (2008) 158–170

Journal of Environmental Management

www.elsevier.com/locate/jenvman

A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas river watershed – Brazil

Philippe Maillard[a,*], Nádia Antônia Pinheiro Santos[b]

[a]Departamento de Cartografia, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Belo Horizonte, Minas Gerais, 31270-901, Brazil
[b]Instituto Mineiro de Gestão das Águas (IGAM), Rua Santa Catarina, 1354, Belo Horizonte, Minas Gerais, 30170-081, Brazil

Received 27 September 2005; received in revised form 27 November 2006; accepted 1 December 2006
Available online 20 February 2007

Abstract

In this article, a methodology for evaluating the effect of land use/land cover on the quality of nearby stream water in a semiarid environment is described and tested on a large watershed in Southeastern Brazil. The approach aims at identifying the width of the riparian area having the strongest effect on different water quality parameters. The land use/land cover data were generated from remotely sensed data while water quality point data were supplied by a government agency. Testing was conducted for both the rainy and dry seasons in an effort to understand the direct effect of surface runoff. The approach combines cartographic modelling using a geographical information system (GIS) and statistics to establish the strength of the relationship between water quality, land use and the distance from the stream. Results suggest a strong relationship between land use/land cover and turbidity, nitrogen and fecal coliforms. They also suggest that each of these parameters has a unique behavior when distance from the stream is considered. Finally, although it was expected that the models would apply better during the wet season, some parameters had the opposite behavior and displayed a better fit during the dry season.

Keywords: Water quality; Non-point pollution; Statistical modelling; Remote sensing

1. Introduction

A watershed is a natural spatial entity for land use planning and environmental management. Given a long time scale (geological), almost everything within the watershed will be deposited in the stream that drains it. Even on a scale of a few hours (a precipitation event), surface runoff can wash unconsolidated sediments downstream and, with it, nutrients and chemicals that will directly affect the quality of the receiving waters. As such, surface runoff is a major source of non-point pollution and is primarily responsible for the relationship between land use/land cover (LULC) and water quality (WQ).

In a natural densely vegetated watershed, most rainfall is absorbed by the land surface and vegetation and released over a long period of time and there is relatively little direct surface runoff with few nutrients being "lost" to the stream (Karr and Schlosser, 1978; McCulloch and Robinson, 1993). The removal of the natural vegetation and its replacement by agricultural or pastoral land promotes surface runoff and sediment transport (Bruijnzeel, 1990; McCulloch and Robinson, 1993). The conversion of natural vegetation to agriculture and pasture is intense in Brazil. This is especially true for the *cerrado* (wooded savana), the second largest biome in Brazil having a rate of "conversion" exceeding that of the Amazon forest (Ratter et al., 1997).

The impact of this conversion to agriculture, pasture and other uses such as eucalyptus plantations, open mining and urbanization can be considered severe in the Velhas river watershed (Euclydes and Ferreira, 2002). It had drastic

*Corresponding author. Tel.: +55 31 3499 5427; fax: +55 31 3499 5410.
E-mail addresses: philippe@cart.igc.ufmg.br (P. Maillard), psnadia@yahoo.com.br (N.A. Pinheiro Santos).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.009

consequences on the quality of water and has caused the disappearance of most fish species downstream of Belo Horizonte (the capital of the state of Minas Gerais) (Polignano et al., 2001). Many point pollution sources are the cause of the problem and local authorities are taking measures such as the implementation of sewage water treatment plants and stricter control of polluting industries but comparatively little is done with regard to non-point source pollution. One initiative has taken the form of a pilot project for the restoration of riparian vegetation of tributaries of the São Francisco river by EMATER (2005), and the Velhas river is one such tributary. The exact effectiveness of riparian vegetation to reduce diffuse pollution and sediment transport still needs to be further studied. The combination of the level of degradation of the Velhas river and the initiative to restore riparian vegetation has triggered a lot of interest and motivated the present study.

2. Research objective

The objective of this study is to develop and test a simplified methodology to determine the effect of LULC on WQ on a large scale watershed in a semiarid environment. The work is based on the analysis of the Velhas river watershed in Minas Gerais—Brazil. Since it is known that the effect of LULC is greater closer to the stream in the so-called riparian zone, the approach incorporates this spatial relation through analytical cartography (in a GIS environment) by creating buffer zones corresponding to riparian zones (RZ) for each sampling point of WQ. It is hoped that this approach will help determine the effect of distance on different WQ parameters. The methodology is also bi-seasonal: wet and dry. The dry season data are used as a means of control since it is expected that the relation between LULC and WQ should decrease when no precipitation has been recorded for some time and there is no surface runoff.

The approach adopted is a direct consequence of the difficulty and cost of acquiring data over such an extensive area. As in most of Brazil, topographical maps of the Velhas river watershed are only available on a scale of 1:100 000 and have not been updated for over twenty-five years. On such a coarse scale, slope analysis would be strongly biased, especially if only the riparian zones are considered. It is hoped that developing a simpler approach that still provides reliable results would promote more studies on the effect of non-point pollution sources on the WQ of Brazilian streams. The approach is also based on the assumption that each WQ sampling point can be treated as independent from the others when processed appropriately. This is also an effort to disregard point (e.g. sewers, industries) and linear (e.g. roads) pollution sources based on the fact that rural areas are much more extensive than urban ones. Therefore, only a few water sampling points are directly affected by point pollution sources and do not strongly bias the majority of the sampling points located in rural areas.

3. Background

3.1. Water quality and land use

Land use plays a complex multi-faceted role in the hydrological cycle (Karr and Schlosser, 1978; Olsson and Pilesjo, 2002). The vegetation intercepts, and re-evaporates (evapotranspiration) part of the precipitation and has an effect on other hydrological parameters such as percolation and surface runoff (Tong and Chen, 2002). Cleared land and some agricultural practices can promote overland flow and erosion (Nisbet, 2001). The construction of impervious surfaces also causes erosion and prevents replenishing of the water table (Cornish, 2001). All of these effects eventually influence the water quality of the nearby streams and rivers. It has been well reported that the use of agricultural fertilizers can result in an increase of levels of nitrate and phosphorus (Karr and Schlosser, 1978; Foster et al., 1989; Meybeck et al., 1989; Mattikalli and Richards, 1996; Basnyat et al., 1999). Similarly, stock grazing and dairies may increase the presence of fecal bacteria in the water (Moore et al., 1989), provoke erosion problems and increase the turbidity of stream waters (Stout et al., 2000; Evans, 2005).

3.2. GIS and remote sensing in water quality studies

The use of GIS technology for integrating LULC data, WQ data and even hydrological models is increasingly popular in the scientific community and numerous articles on WQ issues can be quoted to that effect. In many of these projects, remote sensing is used for mapping and monitoring LULC. Mattikalli and Richards (1996) combined land use data from maps and remotely sensed images with an "export coefficient model" and a GIS to evaluate nitrogen loss and compared them with measured values in the River Glen watershed (UK). Although their estimates generally agree with real measured values, significant differences were reported. Fisher et al. (2000) used a GIS as a means of spatially grouping WQ stations and correlated their results with agricultural activities (poultry, dairy and beef production) and conservation practices. In their discussion they showed that levels of nitrogen, phosphorus, turbidity and fecal coliform bacteria could be directly associated with the localization of certain agricultural activities.

Basnyat et al. (1999) went further and proposed the integration of GIS, ecological modelling and remote sensing as an effective decision support tool for land management. They computed two multiple regression models based on the proportion of different LULC classes to explain nitrate (NO_3) levels in a medium-size watershed ($138\,km^2$) which was divided into eight different basins where WQ data were collected over a two-year period. While the frst model had a relatively low determination coefficient (r^2) of 0.186 because it integrated whole basins, the second model only considered a model-based riparian zone and showed a strong r^2 of 0.959. Many aspects of the current study were inspired by Basnyat et al. and, for consistency reasons, some terms (e.g. LULC, contribution zone) were also borrowed from their study.

In another study (Wang, 2001), twenty-two river catchments near headwaters were used as independent processes to which were associated the proportions of land use and their relative position to water treatment plants. He used Pearson's correlation and multiple regression to reveal the relationship between biological indicators (fish and macro-invertebrates) and land use. Multiple regression results were conclusive for the fish bio-indicator with urban and wooded areas as strong predictors. Gardi (2001) used a GIS and a crop simulation model (CropSyst) to characterize the relationship between cropping practices, land use, soils and nitrate loss as well as concentrations of herbicides in the water. Although he was not able to establish the relationship between nitrate loss and cropping practices, he demonstrated that the practice of row crops had a negative effect on concentrations of herbicides and pesticides in the Centonara watershed (Italy). In an integrated study Tong and Chen (2002) found a significant correlation between land use and a series of WQ variables on a regional scale (the State of Ohio). They then used these correlations to calibrate a hydrological model (BASINS) which was used in combination with a GIS to simulate runoff and levels of various WQ parameters on a local watershed scale.

3.3. Riparian vegetation

All the studies quoted above (except the study by Basnyat et al. (1999)) used the whole watershed as an input. Although using the whole of the watershed processes to explain the hydrological and ecological condition of the stream is valid, the riparian zone has a disproportionate influence (Petersen, 1992; Schuft et al., 1999; Paula Lima (de) and Brito Zakia, 2001). In particular, riparian forests have the following effect on streams, they:

- stabilize river banks through their intertwining root system,
- serve as a buffer zone and filter between the higher watershed and the aquatic ecosystem, absorbing nutrients and other chemicals from surface runoff,
- decrease and filter the sediment load of surface runoff,
- help maintain the thermal stability of the water by intercepting part of the solar radiation and,
- supply organic matter for aquatic organisms.

As an example of the importance of riparian forests, Sparovek et al. (2002) showed that from an economic point of view, to maintain soil loss at a "sustainable" level (blocking 80% of the current sediment yield value), in a Southeastern Brazilian watershed, a minimum width of 52 meters of forest should be maintained on each side of the river (current legislation requires only 30 m). In an attempt to take the importance of riparian zones into account, Schuft et al. (1999) have proposed a methodology for extracting a series of landscape metrics to characterize riparian stream networks to accommodate the dichotomy of WQ point data and spatial LULC information. Their approach is based on defining sampling areas around each observation site. First, concentric circles of equal area ($1\,km^2$) are drawn, then linear buffer zones are defined on each side of the stream section within the rings. The sampling areas are defined by the intersection of the buffer zones and the concentric circles. A similar, yet different sampling approach has been adopted in the present article and is described in Section 4.3. The idea of isolating the land use in riparian zones is an effort to help determine the width of riparian forest that should be preserved (in a legislative sense) to reduce non point pollution sources in agricultural areas.

4. Material and method

4.1. The Velhas river watershed

With 27,867 square kilometers, the Velhas river watershed is one of the most important in the state of Minas Gerais (Fig. 1). Situated in the central part of the state it also hosts the capital Belo Horizonte and the greater metropolitan area with over three million people. Having suffered repeated degradation for the past 400 years, it is the subject of various restoration projects. Point and non-point pollution sources are unevenly distributed in the watershed. The upper watershed concentrates very important urban areas and equally important mining operations, but agriculture and grazing are also found. The land use in the middle and lower Velhas river is mainly dedicated to agriculture and grazing. Although eucalyptus plantations are found everywhere, the majority of the activities are concentrated in the middle section. While the headwater region is characterized by a rugged topography (average altitude: 1500 m) and receives about 2000 mm of rainfall each year, the river mouth is mostly flat (at an altitude of 500 m) and receives only about 1100 mm of rain.

4.2. Data and data generation

In this study, data were obtained from four different sources: (1) Landsat images, (2) digital topographic maps, (3) in situ field surveys and (4) water quality data. The first three sources were used to produce LULC digital maps. The LULC map was generated from a mosaic of five Landsat 7 ETM+ images. Digital maps and in situ field data were used both as ground truth and to complement the classified mosaic with elements that could not be extracted from the images. The five Landsat images were geometrically rectified and registered to a UTM cartographic projection. Table 1 shows the orbit/scene reference of each image along with its date of acquisition and root

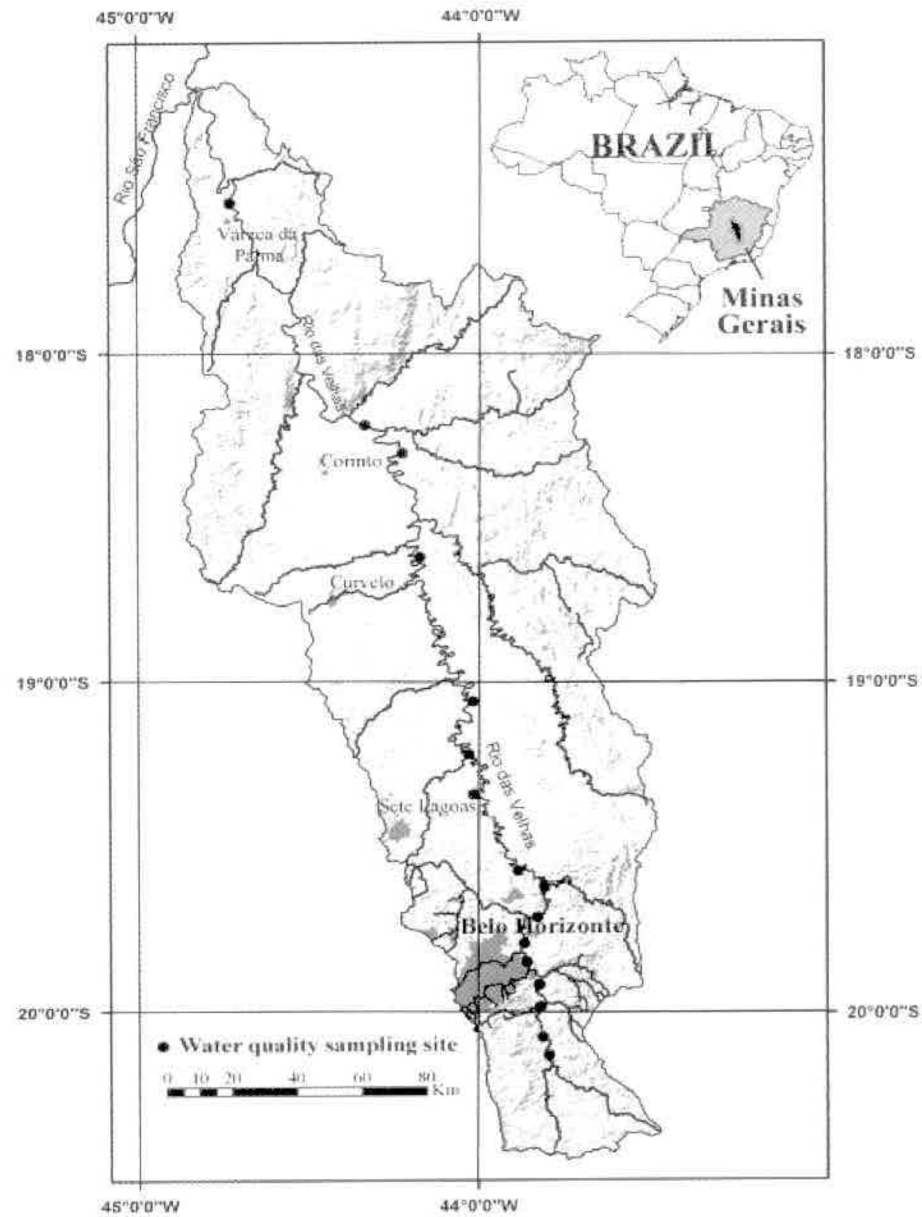


Fig. 1. The Velhas river watershed.

mean square (RMS) registration error. The scenes were first radiometrically corrected using the "improved dark-object subtraction technique"[1] to correct for atmospheric differences (Chavez Jr., 1988). The results from this correction were not completely satisfactory and further *ad hoc* radiometric adjustments were necessary to reduce spectral response differences between the images. These additional corrections were small linear histogram adjustments. The mosaic was constructed by juxtaposing the five images while always leaving the image with the least atmospheric effect on top for overlap. The resulting mosaic

Table 1
Listing of the orbit/scene reference (WRS) for the five Landsat 7 ETM+ scenes with their respective date of acquisition, RMS registration error and standard deviation (σ)

WRS orbit/scene	Date of acquisition	RMS error (m)	σ (m)
218/72	22 October 2002	24.30	13.21
218/73	06 October 2002	49.59	24.96
218/74	06 October 2002	44.27	24.57
219/72	13 October 2002	66.29	26.15
219/73	13 October 2002	16.56	12.00

[1]The dark object subtraction technique consists of subtracting a constant value from all the pixels in a band of an image. The constant is determined by identifying an object that should either absorb all electromagnetic energy (*e.g.* a clear water lake) or not receive any direct light (*e.g.* a very steep slope). If such an object has a value of 20, then 20 is subtracted from all the pixels in that band. The process is repeated for all the other bands. In the improved version, the constants for all the bands are not independent but computed from a look-up table and an estimate of the quality of the atmosphere.

was sufficiently uniform to classify as a single data source using training data acquired in situ and, in some exceptional cases, from maps. The thermal and panchromatic bands were excluded from the classification process. Urban areas were first extracted from topographic maps and then visually updated using the Landsat mosaic. The mosaic was over 99% cloud free. Clouds were treated as "no data" and removed from the analysis. The watershed limits were used as a mask to clip onto the classified mosaic.

Classification was performed using maximum likelihood from the Purdue/NASA MultiSpec software package that is available at no cost from the MultiSpec website (http://dynamo.ecn.purdue.edu/biehl/MultiSpec/). The MultiSpec classification scheme makes it possible to classify any combination of the training areas, testing areas and/or the entire image and reports the results with a wide range of statistics. Biehl and Landgrebe (2002) give a complete description of the MultiSpec package. The LULC map accuracy was tested using only in situ data which were collected using a near-random sampling scheme. A total of 250 random sampling points were defined within the watershed to guide field work. Using a global positioning system (GPS) the sampling sites were then visited (by the authors) on the basis of the "nearest accessible site" within a one kilometer radius of the randomly generated coordinates. Each site generated a bundle of pixels (from 9 to 25 depending on the homogeneity of the terrain) that were used either as training (about 20%) or as test areas. Although sub-optimal, this approach had the benefit of expediency and avoiding land owners' authorization problems. The hydrographic network was extracted from the topographic maps and manually adjusted where needed (and possible). The LULC classification system used was inspired by the Anderson system (Anderson et al., 1976) and adapted for Brazil according to Sokolonski's "Sistema de Classificação de Uso Atual da Terra" (Sokolonski, 1999). The classified mosaic was merged with the adjusted hydrographic network, the updated urban areas and the "no data" zones to produce a single LULC map. Overall accuracy was of 89.8% (kappa $\kappa = 0.883$). Table 2 shows the two-level class system used in this study and gives the producer's and user's accuracies for each LULC class along with sample size. For statistical modelling purposes the classification system was simplified to the following classes: riparian forest, forest, savanna, planted forest, agro-pastoral and barren (the class *urban* was extracted from maps and/or interpreted from the imagery and was not tested for accuracy).

The WQ data were obtained from the *Projeto Águas de Minas* of the *Instituto Mineiro de Gestão das Águas* (IGAM) with 29 water collecting stations for which 13 physical, chemical and biological parameters have been measured on a quarterly basis since 1992. Based on the known relationship between the parameters and LULC, five of these were retained for the statistical analysis: turbidity, nitrate, nitrite, phosphorus, and fecal coliforms. The DELPHI water quality index (Bollmann and Marques, 2000) was also used even though it incorporates some parameters that are more related to point pollution than non-point. The index was computed by IGAM using the following formula:

$$WQI = \prod_{i=1}^{9} q_i^{w_i},$$

where q is the parameter and w is the parameter weight as defined in Table 3.

Of the 29 stations, only 16 were directly from the Velhas river (and not from a tributary). Only data from 2002 were used as this was the year corresponding to the

Table 2
The LULC classification system adapted from Anderson et al.(1976) and Sokolanski (1999) along with the sample size and producer's and user's accuracy values from the image classification

Level I	Level II	Samples	Accuracy (%)	
			Producer's	User's
1 Urban		+		
2 Agro-pastoral	21 Agriculture	*		
	22 Pastoral	382	81.7	76.5
3 Vegetation	31 Primary forest	−		
	32 Secondary forest	667	97.5	95.6
	33 Wooded savanna	269	77.0	73.9
	34 Grassland savanna	67	61.2	100
	35 Dry deciduous forest	272	74.6	90.2
	36 Riparian forest	237	76.8	71.1
Planted forest	Eucalyptus plantation	441	74.4	81.2
Barren	Exposed soil	197	95.4	82.5
	Open soil	+		
	Rock outcrop	857	99.4	97.6
Water		124	99.2	100

Note that the samples were based on bundles of 9 (3 × 3) to 25 (5 × 5) pixels
+, interpreted; *, interpreted and/or merged with pastoral; −, none observed.

LULC data. As a reference, Table 4 shows total precipitation for the month of January and July for the year 2002. It should be noted however that we have no knowledge of the exact day in the month when the water samples were collected. Table 5 shows the five WQ parameters and the WQ index for each of the 16 Velhas river stations and for both the wet and dry seasons.

4.3. Determining contributing areas

Since all the WQ data are from the same watershed and most are located on the same stream (the Velhas river), they are statistically highly correlated and, to some extent, all the upstream points contribute to the measurements of any observation point. This is statistically undesirable and would produce strongly biased results. To solve this problem, the water quality data from the sampling point upstream were subtracted from each point. For the approach to be consistent, the first sampling point near the watershed source had to be eliminated (no sampling point upstream) and only the sampling points directly on the Velhas river could be used. Of course, this resulted in producing some negative values for the water quality parameters. For each remaining point (15 in all), a series of *riparian zones* (RZ) of various widths were defined. To do so, a sub-watershed was defined that contributes exclusively to each water sampling point. Then each sub-watershed is successively overlaid on riparian buffer zones of varying width to produce the *exclusive contribution zones* (ECZs). Buffers were created for a maximum width of 510 m and all were multiples of 30 m to match the LANDSAT image resolution. This approach produced 17 buffer zones which were then reduced to five using a process of elimination. The first two buffers were eliminated (30 and 60 m) for being too close to the stream and possibly affected by positioning errors. Ten of the remaining 15 buffers were eliminated for being uninformative (linearly distributed) based on analyzing the plots of coefficients of determination (see Section 5). Fig. 2 illustrates the process of creating the RZs. Since the topography of riparian zones does not vary as much as in the whole watershed, this procedure also had the effect of reducing the impact of relief variations; an important hydrologic variable for watershed modelling that is not considered in the present study. Although incorporating the relief would probably improve the models, the only available medium scale topographic maps (1:100.000 and a few 1:50.000) could not yield enough precision within the restricted riparian zones.

4.4. Statistical modelling

With only 15 sampling points, a thorough statistical analysis is difficult. Still, considering the high cost of collecting water samples systematically and performing

Table 3
Parameters and parameter weights used in the calculation of the DELPHI water quality index

Parameter	Parameter weight
Dissolved oxygen (%)	0.17
Fecal coliform (NMP/100 ml)	0.15
PH	0.12
Biochemical oxygen demand (mg/L)	0.10
Nitrate (mg/LNO_3)	0.10
Phosphorus (mg/LPO_4)	0.10
Temperature variation (°C)	0.10
Turbidity (UNT)	0.08
Total solid (mg/L)	0.08

Table 4
Total precipitation (mm) for the months of January and July of the year 2002 for the Velhas river watershed

Month	High	Mid	Low
January	306.8	398.4	226.9
July	8.0	6.3	0.0

The data are presented for the high-, mid- and low-watershed.

Table 5
Water quality parameters measured in the 16 Velhas river stations for the wet (top) and dry (bottom) season of 2002

WQ	Velhas river station: WET season: January 2002															
Parameter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Turbidity (NTU)	94.8	603.0	531.0	505.0	436.0	415.0	1088.0	915.0	738.0	627.0	644.0	923.0	726.0	690.0	366.0	434.0
Nitrate (mg/L NO_3)	0.09	0.04	0.04	0.22	0.24	0.25	0.28	0.46	0.66	0.34	0.57	0.67	0.43	0.52	0.48	0.38
Nitrate (mg/L NO_2)	0.005	0.004	0.005	0.011	0.007	0.010	0.105	0.121	0.000	0.051	0.014	0.000	0.008	0.007	0.005	0.006
P (mg/L PO_4)	0.02	0.22	0.01	0.01	0.03	0.14	0.05	0.04	0.31	0.01	0.05	0.03	0.08	0.05	0.09	0.21
Fecal col. (NMP/100 L)	—	—	—	50.0	50.0	90.0	160.0	160.0	50.0	7.0	30.0	17.0	2.2	2.8	0.7	1.7
WQI	77.3	57.4	61.7	41.3	41.9	38.1	30.1	28.9	30.0	38.6	37.5	39.8	46.5	44.3	49.9	44.7
DRY season: July 2002																
Turbidity (NTU)	6.5	8.17	9.29	12.2	9.83	54.9	71.0	55.3	82.4	35.3	6.88	20.7	20.5	11.6	7.72	3.83
Nitrate (mg/L NO_3)	0.16	0.24	0.29	0.38	0.41	0.27	0.02	0.02	0.03	0.18	1.76	2.90	1.83	1.82	1.87	1.02
Nitrite (mg/L NO_2)	0.003	0.022	0.023	0.048	0.046	0.027	0.006	0.006	0.005	0.022	0.223	0.245	0.015	0.019	0.032	0.008
P (mg/L PO_4)	0.02	0.04	0.05	0.08	0.06	1.50	0.95	1.05	0.96	0.34	0.19	0.14	0.12	0.11	0.04	0.04
Fecal col. (NMP/100 L)	0.00	7.0	3.0	7.0	8.0	160.0	160.0	160.0	160.0	13.0	0.13	0.001	0.05	0.03	0.01	0.02
WQI	76.1	62.6	65.2	60.3	61.3	19.6	15.7	14.6	16.3	36.2	59.7	66.0	59.5	55.9	79.5	77.4

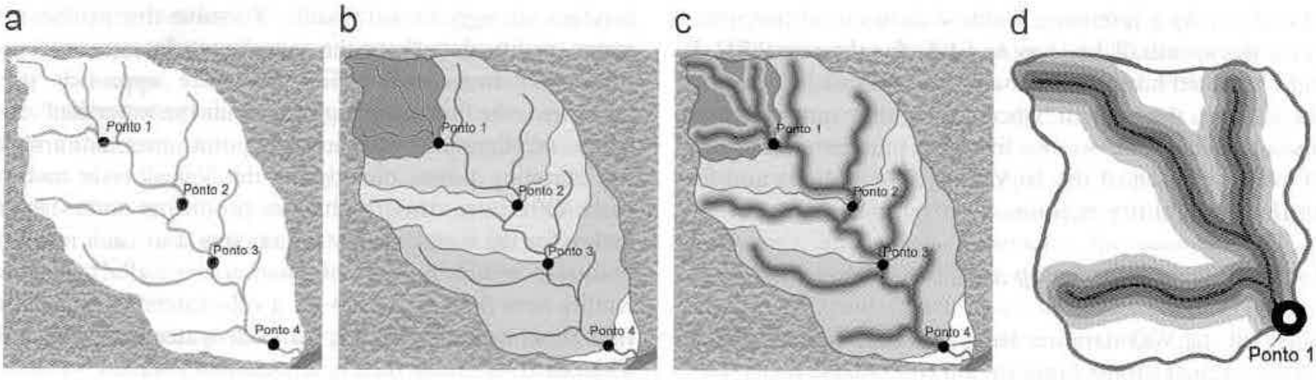


Fig. 2. Illustration of how the riparian zones are defined: (a) the water quality sampling points, (b) definition of the exclusive contribution zone for each sub-watershed, (c) overlaying the riparian buffer zones of different width, and (d) illustration of the riparian zones for point 1.

chemical analyses, the practice of statistical modelling with few samples has been considered acceptable and many other studies have adopted a similar approach. For example, Mattikalli and Richards (1996) and Fisher et al. (2000) used 17 and 10 water sampling sites respectively while Basnyat et al. (1999) and Wang (2001) used 8 and 22 small catchment watersheds respectively for their analysis. To compensate for this weakness and strengthen the results, the analysis was carried out for two different months: January (wet season) and July (dry season). The argument to support this approach was the following: during the wet season, surface runoff is more often active and LULC should have a stronger influence upon WQ than during the dry season. There are however some problems with this statement which will be considered in due time. One such problem is that farmers rarely use fertilizers and nutrients during the wet season because much of them will be washed away by the rain. Another problem is that the soil is more exposed during the dry season and therefore more susceptible to erosion.

Multiple regression using the least square rule has been chosen for the present study. Multiple regression produces linear models in the form $Y = b_0 + b_1X_1 + b_2X_2 + \dots b_nX_n$ where Y is the explained variable, $X_{1,2\dots n}$ are the independent or predictor variables, $b_{1,2\dots n}$ are the coefficients and n is the number of variables. It was preferred to correlation because LULC is made up of a mosaic of different classes in all the exclusive contribution zones of the WQ sampling sites. Taken one at a time, the LULC classes would not have a significant impact on the WQ. This is also consistent with the fact that together these classes make up 100% of the contribution zones (except for a few clouds classified as "no data"). All the models were calculated using proportions of each LULC and not actual areas in units of measure. Only models with coefficients of determination (r^2) superior to 0.6 were retained. To further test the validity of the models, a level of significance $p \geqslant 0.9$ has been considered. Finally, an analysis of variance was performed on the betas ($b_{1,2,\dots,n}$) to test if they are equal to zero ($H0$: $b_{1,2,\dots,n} = 0$; $H1$: $b_{1,2,\dots,n} \neq 0$). A level of 90% (0.1) was also used for this inference test. These precautions also made sure that the results were valid despite the small sample used. Scatterplots of the residuals were also produced to make sure these were randomly distributed.

5. Results and discussion

Tables 6 and 7 show the coefficients of determination of the various models. The tables were formatted to include information on the models that were rejected on the basis of their low coefficients of determination (italic) or on not being able to reject the null hypothesis ($H0$) of the analysis of variance stating that the betas are equal to zero at 90% (values with asterisk). This was the case for the following WQ parameters: phosphorus (P) and water quality index (WQI). Valid models are in bold type. These include nitrite for the wet season and turbidity and fecal coliforms for the dry season.

Models with a plus sign (+) mean that the $H0$ of the analysis of variance could be rejected at 85%. Nitrite and nitrate during the dry season are such models and will be discussed. Although WQI has some valid models ($H0$ rejected at 85%), it was not retained for discussion, because it is an index and not a measured value, and its interpretation would, therefore, be too subjective. Fig. 3 shows a series of plots of the coefficients of determination and how they vary with the buffer width (RZ) for the WQ parameters retained for discussion. The results for these parameters are discussed separately in the following subsections.

5.1. Turbidity

High turbidity is normally associated with the wet season when surface runoff transports sediments from the soil and carries them to the stream. During the wet season the high waters are also much more turbulent and do not allow the sediments to settle on the river bed. Suspended sediments can therefore travel long distances and all the tributaries also contribute to the turbidity. This is also the case here and turbidity is much higher during the wet season (see Table 5).

For modelling purposes, turbidity values were log transformed to obey a more linear behavior. The January

Table 6
Coefficients of determination (r^2) of the wet season multiple regression models for the five water quality parameters and for the different riparian zones (RZ) including the whole exclusive contribution zone (ECZ, last column)

Water quality Parameters	WET season: January 2002 RZs					ECZ
	90 m	150 m	210 m	300 m	510 m	
Turbidity	*0.194*	*0.198*	*0.194*	*0.198*	*0.214*	*0.222*
Nitrate	*0.513*	*0.504*	*0.482*	*0.457*	*0.431*	*0.507*
Nitrite	**0.823**	**0.823**	**0.820**	**0.815**	**0.803**	**0.822**
P	**0.690+**	**0.647***	**0.647***	*0.580*	*0.543*	*0.543*
Fecal col. (× 1000)	**0.755***	**0.735***	**0.725***	**0.709***	**0.691***	**0.789***
WQI	*0.453*	*0.461*	*0.459*	*0.464*	*0.487*	*0.502*

Bold: $r^2 \geqslant 0.6$ betas≠0 (90%); +: $r^2 \geqslant 0.6$ betas≠0 (85%)
*: $r^2 \geqslant 0.6$ betas = 0; italic: $r^2 < 0.6$.

Table 7
Coefficients of determintation (r^2) of the dry season multiple regression models for the five water quality parameters and for the different riparian zones (RZ) including the whole exclusive contribution zone (ECA, last column)

Water quality Parameters	DRY season: January 2002 RZs					ECZ
	90 m	150 m	210 m	300 m	510 m	
Turbidity	**0.874**	**0.906**	**0.918**	**0.925**	**0.895**	**0.831**
Nitrate	**0.726+**	**0.752**	**0.779**	**0.809**	**0.836**	**0.708**
Nitrite	**0.713+**	**0.729+**	**0.744**	**0.700**	**0.772**	*0.621*
P	*0.507*	*0.515*	*0.515*	*0.493*	*0.441*	*0.343*
Fecal col. (× 1000)	**0.850**	**0.849**	**0.834**	**0.801**	**0.726**	**0.621***
WQI	**0.677***	**0.673***	**0.664***	**0.649***	**0.616***	**0.474**

Bold: $r^2 \geqslant 0.6$ betas≠0 (90%); $r^2 \geqslant 0.6$ betas≠0 (85%).
*: $r^2 \geqslant 0.6$ betas = 0; italic: $r^2 < 0.6$.

(wet) model was not conclusive with r^2 values below 0.3 for the RZ buffers and below 0.4 for the ECZ. This is understandable for the wet season since suspended sediment intake before the ECZ are generally high and may come from a long way upstream. The contribution of the immediate neighborhood of the water sampling points is therefore relatively too small to affect the model. This is somewhat confirmed by the fact that the r^2 is larger for the ECZ than the RZs.

During the dry season, the regression model is quite conclusive with high coefficients of determination. The analysis of variance on the betas also made it possible to reject *H0*: betas = 0. In that period, relatively little sediment is taken in, and, when this occurs, it is not transported very far. This is also the season during which farmers irrigate their fields and when soils are most vulnerable to erosion. The RZ model shows the strongest relationship (0.925) at a buffer distance of 300 m declining to 0.895 at 510 m and 0.831 for the ECZ (Fig. 3a). The actual model (Table 8) shows that the class that contributes most to increased turbidity is barren land followed by agro-pastoral and planted forest. Here the class *forest* is abnormally counted as a contributor but with a much lower weight than barren soil. The strongest inhibitor is predictably riparian forest. Savanna is also an inhibitor but with a rather low weighting factor. Although imperfect, the model still maintains a predictable behavior for most classes (except forest) and appears to be a valid predictor Tables 9,10,11.

5.2. Nitrate and nitrite

In the aquatic environment, nitrogen can be found in various forms: molecular nitrogen, organic nitrogen, ammonium nitrogen, nitrite and nitrate. Domestic and industrial sewers along with animal excrement and fertilizers are the main sources of nitrogen. Nitrite is usually associated with active biological processes influenced by organic pollution (IGAM, 2002). Nitrate is more associated with the use of organic and inorganic fertilizers (Meybeck et al., 1989; Mattikalli and Richards, 1996; Basnyat et al., 1999). Although these two nitrogen sources are often treated together, since they were measured separately, two different sets of models were created. Except for four sampling sites in the middle of the watershed, nitrite and nitrate are higher in the dry season

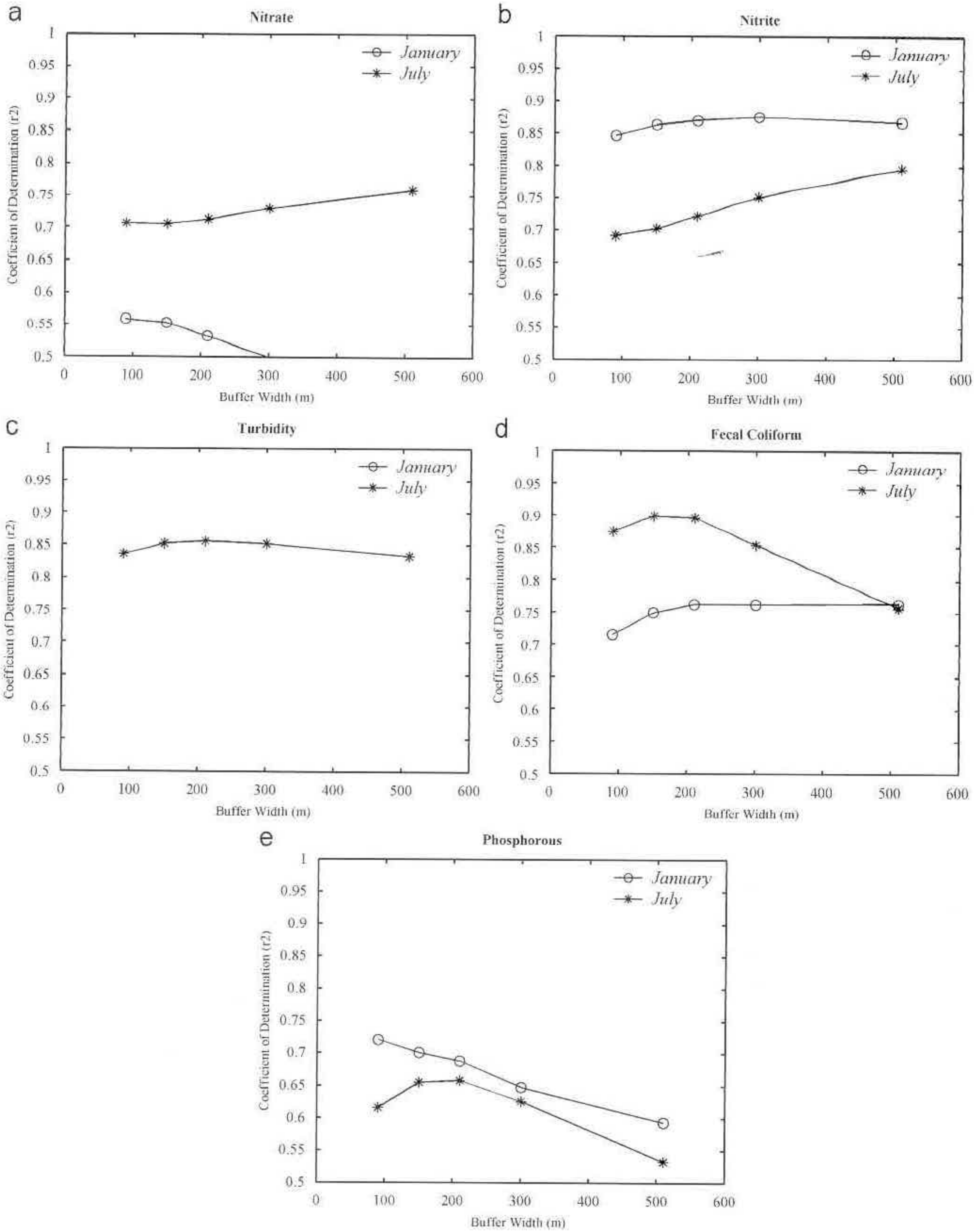


Fig. 3. Plosts of r^2 according to riparian buffer zone width.

Table 8
Multiple regression model for turbidity (dry season/300 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−5.985	−3.98E-03	0.180	−4.75E-04	7.740E-02	8.868E-02	8.433E-02	5.728E-02
Std. err.	2.220	0.029	0.038	0.024	0.029	0.037	0.024	0.022

C = constant; RF = Riparian Forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted Forest; U = Urban.

Table 9
Multiple regression model for nitrate (dry season/510 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−3.294	−2.53E-02	2.486E-02	1.032E-02	5.234E-02	8.074E-02	6.804E-02	2.976E-02
Std. err.	1.918	0.022	0.030	0.018	0.022	0.027	0.017	0.016

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

Table 10
Multiple regression model for nitrate (wet season/90 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	1.186E-02	−1.18E-04	1.756E-04	−6.32E-04	−3.42E-04	4.045E-04	−7.82E-04	9.974E-04
Std. err.	0.090	0.001	0.001	0.001	0.001	0.001	0.001	0.001

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

Table 11
Multiple regression model for nitrite (dry season/510 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−0.313	−4.28E-04	−3.49E-03	4.368E-04	6.766E-03	8.894E-03	7.811E-03	3.259E-03
Std. err.	0.248	0.003	0.005	0.003	0.003	0.004	0.003	0.002

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

(Table 5). The reason for these generally higher concentrations[2] can partly be explained by the fact that high waters of the wet season dilute these chemicals more. Another by no means negligible reason for this seasonal behavior is that farmers do not use much fertilizer during the rainy season as most of it would be washed away.

The models for nitrite show a consistent behavior for both seasons but with a stronger relationship in January (wet) that culminates at 300 m. The dry season models are weaker and less significant in terms of analysis of variance results. The curve shows a steady increase from 90 to 510 m suggesting that it might have continued to increase if a larger buffer had been considered. These results suggest that runoff during the rainy season might be responsible for bringing an increased amount of organic matter from the riparian zone into the stream.

In all three models riparian forest acts as an inhibitor confirming its essential filtering role. For the nitrate, all the other classes have a positive contribution. The fact that forest and savanna have positive weights is not an expected result and probably comes from the various sources of error and the simplified model approach. The wet season nitrite model has, in decreasing order of importance, planted forest, savanna, forest and riparian forest as inhibitors. Here the distance from the stream appears to be less important since the determination coefficients do not decrease much with the increased distance. This is perhaps why the riparian forest is the weakest inhibitor. Urban is the strongest contributor followed by agro-pastoral and barren soil. During the dry season, the model behaves in a less understandable way but the strongest contributors (agro-pastoral, planted forest and urban) are consistent with the expected behavior. Since this model has a weaker coefficient of determination (maximum of 0.772 at 510 m), it is therefore less reliable.

5.3. Phosphorus

Consumption of phosphorus does not appear to affect human health but high phosphorus levels in the water can

[2]The concentrations in Table 5 are considered acceptable for human consumption ($<5\,\text{mg}=L$) but are significantly higher than would be found in a natural environment.

have dramatic effects on aquatic life. Concentrations of over 0.03 mg/L can provoque excessive plant growth, increase demand of oxygen and reduce fish stock (MOE, 1984). Most concentrations presented in Table 5 are over 0.03 mg/L and can be considered excessive. Phosphorus in surface water can come from a variety of sources both point and non-point. It is associated with sewers, animal excrement, fertilizers and some detergents.

Unlike nitrogen, the difference in levels between the wet and dry seasons are small. However, the model for the dry season could not be considered reliable with coefficients of determination well below the 0.6 mark and a level of significance below 90%. The wet season model on the other hand, can be considered acceptable for the first three buffer widths (90, 150 and 210 m). The model for the 90 m buffer is in Table 12 and Fig. 3e and shows that all the land uses except one (agro-pastoral) contribute positively to the level of phosphorus, the strongest being bare soil and planted forest (eucalyptus). Since the model coefficients can only be considered different from zero at 85% (betas≠0 (85%)), we decided to restrict our interpretation to the simple fact that distance from the stream seems to be a determining factor for levels of phosphorus.

5.4. Fecal coliforms

Except for three water sampling stations (and the first three sampling sites with no data in January), the fecal coliform counts are higher during the wet season. Stations 6–8 have the highest counts which are related to the urban area of Belo Horizonte and its surrounding fringe. The lower values during the dry season are probably related to the almost complete absence of rain and runoff.

The models for January are not reliable because the null hypothesis that the betas are equal to zero could not be rejected. The July models, however, were very conclusive with a peak at $r^2 = 0.9$ and rejected $H0$ at 99.5% (Fig. 3f). This peak is reached at 150 m from the stream and then the relationship decreases steadily to $r^2 = 0.76$ at 510 m. The even lower $r^2 = 0.54$ for the ECZ suggests that this trend continues with the increased distance from the stream. Stock grazing near the stream is probably the main explanation for this behavior. It has been observed for instance that in the absence of riparian forest, stock tends to enter the stream more often. This further shows the importance of maintaining or restoring riparian forests.

As can be seen in Table 13, the strongest LULC contribution to fecal coliform counts comes from barren land followed by planted forest, agro-pastoral land and urban areas. These contributors can be merged into "human impact" areas. The inhibitors are riparian forest and savanna showing again the importance of maintaining natural vegetation near streams.

5.5. Discussion

The simplified approach used here was able to show a consistent link between LULC and WQ and that the former can be used to model the latter to some extent. Isolating LULC in riparian zones of different width made it possible to identify some patterns in the spatial behavior of the various water quality parameters with respect to stream distance. It did not, however, bring a significant increase in all the coefficients of determination of the regression models.

The models have also helped identify the LULC classes that have the most impact on WQ. In that respect, riparian forest is the single land use class that shows the most consistent behavior. It is a good inhibitor for turbidity, nitrate, nitrite and fecal coliforms. This behavior is consistent with that observed in many other studies although the exact weight varies widely which can probably be explained by regional differences such as climate, topography, soils, lithology, etc. Another very consistent finding is that barren land is a systematic contributor to the five WQ parameters selected. Its relative role in turbidity

Table 12
Multiple regression model for phosphorus (wet season/90 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−0.661	1.50E-02	1.958E-02	11.133E-03	5.976E-03	−3.57E-03	1.720E-02	3.856E-03
Std. err.	0.554	0.009	0.008	0.006	0.007	0.010	0.006	0.005

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

Table 13
Multiple regression model for fecal coliform (dry season/90 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−179283	2386.6	7513.6	−1899.4	1563.0	2078.3	5116.4	1717.7
Std. err.	152749.7	2409.9	2232.4	1675.8	1840.6	2686.0	1772.1	15001.4

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

exceeds all the other LULC classes. The approach taken here did not however enable us to show this relation for the wet season. During the dry season, the agro-pastoral and planted forest classes consistently contribute to worsening WQ showing that when they are near streams, land conservation practices should be carefully applied. Savanna and secondary forest have a more complex behavior and are sometimes inhibitors, sometimes contributors. It should be considered here that both these classes of mainly open vegetation are often encountered in varying degrees of degradation and we were not able to make the LULC mapping carry this information. The urban class (taken here as a non-point source since only its area is considered) is a consistent contributor but with relatively small weight in the models retained (except phosphorus). This might come as an unexpected behavior but is explained by the fact that only two sampling points are directly affected by large urban centers and by the fact that, being mainly a point pollution source, its area might be of little relevance.

The riparian zone model used in this study was inspired by the study of Basnyat et al. (1999) but was simplified and adapted through cartographic modelling inside a GIS environment. However, unlike the results from Basnyat et al., the differences between the whole ECZs and the RZs are generally small. By subtracting water quality data from the previous point upstream and defining an exclusive contribution zone, we were able to analyze each sampling point as an independent process. The RZs were defined using the main river stream and its major tributaries but other smaller tributaries were left out.

The specific spatial behavior of each parameter with relation to the distance to the stream is perhaps the most interesting finding of the current study. Although by itself, this information does not help in defining legislation for riparian zone preservation, it can be useful for agricultural and environmental agencies giving advice on ideal width depending on the main activities. For instance a wider riparian zone should be preserved where large quantities of nitrogen-based fertilizers are used. On the other hand, the present legislation (30 m on each side of small streams) might be sufficient for grazing land. During the dry season, while the relationship between LULC and turbidity reaches it peak at 300 m, the fecal coliform coefficient of determination culminates at 90 m (or less but this could not be determined) to decrease steadily thereafter. On the contrary, the strength of the relation between LULC and nitrogen (nitrate and nitrite) increases with distance and reaches its peak at the maximum RZ buffer size. This last relation might be due to the fact that the main activities responsible for this parameter are not as significant near the stream. We feel that this aspect of the research, i.e. the spatial behavior of the relation between LULC and the different WQ parameters, merits more attention and should guide specific restrictions in land use at varying distances from the stream instead of the adoption of a single criterion (e.g. preserving 30 m of riparian vegetation) for all activities.

6. Conclusions

The strength of the statistical approach presented here lies in its simplicity and in the fact that it takes advantage of existing WQ analysis and a "coarse" resolution LULC map. It relies on the assumptions that the relation between WQ and LULC can be modelled without taking point pollution sources into account and without considering important hydrological variables like slope and precipitation. Slope could not, in this case, be properly modelled due to the lack of reliable data. Precipitation could not be adequately incorporated into the approach because the water samples were collected throughout the month (January or July) and not during a single day.

The results indicate that the LULC classes can be used to model turbidity, fecal coliforms, nitrogen and, to a lesser extent, phosphorus. The dry season appears to be a better choice for modelling these WQ parameters but more extensive research is needed to determine the effect of precipitation on the simplified approach presented here. They also indicate that each WQ parameter can have a distinct pattern with relation to distance from the stream and that these patterns can help define guidelines for watershed management.

The simplified approach that takes advantage of existing WQ data and LULC of medium-coarse resolution represents an important tool in the Brazilian context. As a developing country, adequate data for monitoring LULC changes and its effect on water quality are often lacking. Future research efforts will focus on incorporating precipitation and slope into the models in a smaller watershed having less disparate characteristics in order to better isolate the effect of LULC.

Acknowledgements

The research presented here is part of the Master's dissertation of Nádia Antônia Pinheiro Santos under the supervision of Dr. Philippe Maillard. The authors are extremely grateful to the *Instituto Mineiro de Gestão das Águas* (IGAM) for making available their water quality data and for providing logistic support and to Luciano V. Dutra from the *Instituto Nacional de Pesquisa Espacial* (INPE) for providing the LANDSAT data. Finally we would like to thank the three anonymous referees who have helped improve the present article with their most appropriate suggestions.

References

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A land use and land cover classification system for use with remote sensing data. Geological Survey Professional Paper 964. US Government Printing Office for the US Department of the Interior, Geological Survey, Washington, D.C.

Basnyat, P., Teeter, L., Lockaby, B., Flynn, K., 1999. The use of remote sensing and GIS in watershed level analyses of non-point source pollution problems. Forest Ecology and Management 128, 65–73.

Biehl, L., Landgrebe, D., 2002. Multispec—a tool for multispectral - hyperspectral image data analysis. Computers and Geosciences 28 (10), 1153–1159.

Bollmann, H.A., Marques, D.M., 2000. Bases para a estrutura»c~ao de indicadores de qualidade de águas. Revista Brasileira de Recursos Hidricos 5 (1), 37–60.

Bruijnzeel, L.A., 1990. Hydrology of Moist Tropical Forests and Effects of Conversion: A State of Knowledge Review. Free University, Amsterdam.

Chavez Jr., P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispetral data. Remote Sensing of Environment 24, 459–479.

Cornish, P.M., 2001. The effect of roading, harvesting and forest regeneration on streamwater turbidity levels in a moist eucalypt forest. Forest Ecology and Management 152, 293–312.

EMATER, 2005. Programas de Desenvolvimento. Empresa de Assistência Técnica e Extensão Rural do Estado de Minas Gerais—Secretaria de Estado da Agricultura—MG. ⟨http://www.emater.mg.gov.br/siteemater/ServProd/Progdesenv.asp⟩ (last accessed August 29, 2005).

Euclydes, H.P., Ferreira, P.A., 2002. Recursos Hidricos e Suporte Ecológico a Projetos Hidroagricolas: Sub-Bacia do Alto e Médio São Francisco. Universidade Federal de Vicosa (Boletim técnico, 6), Vicosa, MG, Brazil.

Evans, R., 2005. Curtailing grazing-induced erosion in a small catchment and its environs, the Peak District, Central England. Applied Geography 25, 81–95.

Fisher, D., Steiner, J., Endale, D., Stuedemann, J., Schomberg, H., Franzluebbers, A., Wilkinson, S., 2000. The relationship of land use practices to surface water quality in the Upper Oconee Watershed of Georgia. Forest Ecology and Management 128, 39–48.

Foster, I.D.L., Ilbery, B.W., Hinton, M.A., 1989. Agriculture and water quality: a preliminary examination of the Jersey nitrate problem. Applied Geography 9, 95–113.

Gardi, C., 2001. Land use, agronomic management and water quality in a small Northern Italian watershed. Agriculture, Ecosystems and Environment 87, 1–12.

IGAM, 2002. Qualidade das Águas Superficiais no Estado de Minas em 2001. Instituto Mineiro de Gestão das Águas. Karr, J.R., Schlosser, J. Water resources and the land-water interface. Science 201 (4352), 229-234.

Karr, J.R., Schlosser, I.J., 1978. Water resources and the land-water interface. Science 201, 229–234.

Mattikalli, N.M., Richards, K.S., 1996. Estimation of surface water quality changes in response to land use change: Application of the export coefficient model using remote sensing and geographical information system. Journal of Environmental Management 48, 263–282.

McCulloch, J.S.G., Robinson, M., 1993. History of forest hydrology. Journal of Hydrology 150, 189–216.

Meybeck, M., Chapman, D., Helmen, P., 1989. Global freshwater quality: a first assessment. Global Environment Monitoring System/UNEP/WHO, Geneva.

MOE, 1984. Water Management: Goals, Policies, Objectives and Implementation Procedures of the Ministry of the Environment. Ministry of the Environment, Government of Canada.

Moore, J.A., Smyth, J., Baker, E., Miner, J., Moffitt, D., 1989. Modeling bacteria movement in livestock manure systems. Transactions of the American Society of Agricultural Engineers 32 (3), 1049–1053.

Nisbet, T.R., 2001. The role of forest management in controlling diffuse pollution in UK forestry. Forest Ecology and Management 143, 215–226.

Olsson, L., Pilesjo, P., 2002. Environmental Modelling with GIS and Remote Sensing, 1st Edition. Approaches to spatially distributed hydrological modelling in a GIS environment. Taylor and Francis, London, GB, Ch. pp. 167–199.

Paula Lima (de), W., Brito Zakia, M.J., 2001. Mata Ciliares: Conservação e Recuperação, 2nd Edition. Universidade de São Paulo: Fapesp, São Paulo, SP, Ch. Hidrologia de matas ciliares, pp. 33–44.

Petersen, R.C., 1992. The RCE: the riparian, channel, and environmental inventory for small streams in the agricultural landscape. Freshwater Biology 27, 295–306.

Polignano, V.M., Alves, A.L., Machado, A.T.M., Lisboa, A.H., 2001. Uma viagem ao Projeto Manuelzão e á Bacia do Rio das Velhas. Universidade Federal de Minas Gerais—Projeto Manuelzão, Belo Horizonte, Brazil.

Ratter, J.A., Ribeiro, J.F., Bridgewater, S., 1997. The brazilian cerrado vegetation and threats to its biodiversity. Annals of Botany 80, 223–230.

Schuft, M.J., Moser, T.J., Wigington Jr., P.J., Stevens Jr., D.L., McAllister, L.S., Chapman, S.S., Ernst, T.L., 1999. Development of landscape metrics for characterizing riparian-stream networks. Photogrammetric Engineering and Remote Sensing 65 (10), 1157–1168.

Sokolonski, H.H., 1999. Manual Técnico de Uso da Terra. IBGE: Departamento de Recursos Naturais e Estudos Ambientais, Rio de Janeiro – RJ.

Sparovek, G., Ranieri, S.B.L., Gassner, A., De Maria, I.C., Schnug, E., Santos (dos), R.F., Joubert, A., 2002. A conceptual framework for the defnition of the optimal width of riparian forests. Agriculture, Ecosystems and Environment 90, 169–175.

Stout, W., Fales, S., Muller, L., Schnabel, R., Weaver, S., 2000. Water quality implications of nitrate leaching from intensively grazed pasture swards in the northeast US. Agriculture, Ecosystems and Environment 77, 203–210.

Tong, S.T.Y., Chen, W., 2002. Modeling the relationship between land use and surface water quality. Journal of Environmental Management 66, 377–393.

Wang, X., 2001. Integrating water-quality management and land-use planning in a watershed context. Journal of Environmental Management 61, 25–36.

ELSEVIER

Journal of Environmental Management 86 (2008) 171–180

Journal of Environmental Management

www.elsevier.com/locate/jenvman

A bioeconomic model of a single-species fishery with a marine reserve

T.K. Kar*, H. Matsuda

Faculty of Environment and Information Sciences, Yokohama National University, 79-7 Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa 240-8501, Japan

Received 10 October 2005; received in revised form 27 November 2006; accepted 1 December 2006
Available online 22 January 2007

Abstract

This study examines the impact of the creation of marine protected areas (MPAs), from both economic and biological perspectives. In particular, we examine the effects of protected patches and harvesting on resource populations. We conclude that protected patches are an effective means of conserving resource populations, even though extinction cannot be prevented in all cases. We discuss the dynamic optimization of a harvest policy by choosing $E(t)$, the harvesting effort, as the dynamic variable. We also discuss the optimal equilibrium harvest policy and explain the biological and bioeconomic interpretations of the results.

Keywords: Marine reserve; Global stability; Limit cycle; Optimal harvesting

1. Introduction

Marine protected areas (MPAs), which we define as spatially well-defined areas where no harvesting takes place, have become a popular approach to managing marine fisheries (Polacheck, 1990; Holland and Brazee, 1996; Sanchirico and Wilen, 1999, 2001, 2005; Bonocoeur et al., 2002; Anderson, 2002; Luck et al., 1998; Rodwell et al., 2002; Neubert, 2003; Hilborn et al., 2006). The impetus for establishing MPAs originally came from countries such as Australia, New Zealand, and the Seychelles, where MPAs have provided many general benefits as a tool for conservation and marine environmental management. It is relatively easy to enforce fishing bans in MPAs, if the necessary will and resources exist: quite simply, any fisherman working in such a reserve is breaking the rules. The more effectively a reserve is functioning, the more carefully these restrictions must be enforced.

Two core objectives have motivated the establishment of most marine reserves: conservation and sustainable provision for human use. The conservation goals include: (i) the conservation of biodiversity; (ii) the conservation of rare and restricted-range species; (iii) the maintenance of genetic diversity; (iv) the maintenance and/or restoration of the natural ecosystem on both local and regional scales; and (v) the conservation of areas vital for vulnerable life stages. The goals relating to human use include: (i) the management of fisheries, (ii) recreation, (iii) education, (iv) research, and (v) the fulfillment of esthetic needs. The current overall enthusiasm for reserves and the recent rapid increase in new policies promoting reserve formation have motivated this paper.

Not everyone supports a major expansion of marine reserves, of course, and fishermen are among the most vocal skeptics. One issue for fishermen is whether they will incur significant costs in the form of lost access to traditional grounds. Whether reserves are likely to be costly to fishermen is a complicated issue. We know, for example, that closing all coastal habitats would essentially cost fishermen all of their current income, and we also know that doing nothing would leave us with the status quo in fisheries management. Several different arguments have been put forth suggesting that reserves at this level would actually benefit fishermen. One argument is that current management methods are not sufficiently strict and are destined to fail over the long run (Ludwig et al., 1993). More interesting arguments have suggested that closing some areas might actually enhance fisheries productivity.

*Corresponding author. Permanent address: Department of Mathematics, Bengal Engineering and Science, University, Shibpur, Howrah 711103, India.

E-mail addresses: tkar1117@gmail.com, t_k_kar@yahoo.com (T.K. Kar).

0301-4797/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.001

This argument is based on the possibility that the existence of larger protected populations of more fecund individuals within reserve boundaries might actually increase the flow of larva outside the boundaries, thus enhancing fishery output in the remaining open areas. MPAs also provide important sites for study and experimentation, including control sites for studies of fishing effects. These have been widely used for education, tourism, and increasing the public awareness and understanding of environmental issues. The net economic benefits from some of these activities can far exceed those from fisheries.

Despite the growing interest in marine reserves (Suman et al., 1999; Gell and Roberts, 2002; Halpern, 2003; Guenette et al., 1998; Brown and Roughgarden, 1997), both the economic benefits and the conservation impact of marine reserves have recently been questioned (Hannesson, 1998; Allison et al., 1998). Conrad (1999) showed that, in the absence of ecological uncertainty and in the context of optimal harvesting, reserves generate no economic benefits to fishermen. Such a result coincides with the perspectives of many fishermen and also some economists. However, Luck et al. (1998) asserted that MPAs can be viewed as a kind of insurance against scientific uncertainty, stock assessments, or regulation errors.

Following the tradition of Clark (1985), we employ a simple modeling approach in this paper. That is, we endeavor to understand complex management issues by developing and analyzing a simple model. Despite the simplifying assumptions used, we still manage to include in the analysis several major biological and management characteristics, including population growth, migration, harvesting effort, and reserve size. To derive analytic results concerning the effects on population level and yield of the division of a population's habitat area into sub-areas, namely a nature reserve (NR) and a harvesting zone (HZ), we construct and analyze a simple logistic growth model. This modeling approach allows us to characterize explicitly the relationships among population protection, yield, reserve size, optimal harvest efforts, and biological parameters, including population growth and migration.

2. Model

2.1. Pre-reserve

We assume that the pre-reserve population is uniformly distributed across its habitat area, with no distinction between patches for mating, spawning, growth, maturation, or protection, for example. We also assume that the population evolves according to the logistic law of growth, which is described by

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{k}\right), \tag{1}$$

where x is the population size at any time t, k the carrying capacity, and r the intrinsic growth rate. Now suppose that the population described by Eq. (1) is subjected to harvesting. We use the catch-per-unit-effort (CPUE) hypothesis (Clark, 1990) to refer to the assumption that the harvest is proportional to the stock level, or that qEx, where E and q are positive constants denoting the harvesting effort and catchability coefficients, respectively.

The equation characterizing the harvested population becomes

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{k}\right) - qEx.$$

If $E > r/q$, the right-hand side is always negative, and a rapid collapse of the resource population will occur. In this case, extinction of the resource population is inevitable, and the ecological environment is destroyed.

2.2. Post-reserve

To protect the population and ecological environment, the region is divided into two patches. The two subpopulations are assumed to be homogeneously distributed across their sub-areas. In our model, we assume that each subpopulation has its own carrying capacity, which is proportionate to its distribution area. We also assume that the total population distribution area equals unity, and that sub-areas 1 (NR) and 2 (HZ) equal s and $1-s$, respectively, where $0<s<1$. If $s=0$, then there is no reserve, and this possibility provides the status quo reference case (for more details, see Anderson (2002)). We assume that fish are free to migrate between the two patches, as well as within each one of them. The net migration between the two areas depends on the difference in the subpopulation densities. We also assume that each subpopulation has its own carrying capacity, which is proportionate to its distribution area.

Thus, our model becomes

$$\frac{dx_1}{dt} = rx_1\left(1 - \frac{x_1}{sk}\right) - \sigma\left(\frac{x_1}{sk} - \frac{x_2}{(1-s)k}\right),$$

$$\frac{dx_2}{dt} = rx_2\left(1 - \frac{x_2}{(1-s)k}\right) + \sigma\left(\frac{x_1}{sk} - \frac{x_2}{(1-s)k}\right). \tag{2}$$

Normalizing the population by dividing the population level by the carrying capacity and considering the harvest in HZ based on a CPUE that is proportional to the population density, we arrive at the growth equations

$$\frac{dx_1}{dt} = rx_1\left(1 - \frac{x_1}{s}\right) - \sigma\left(\frac{x_1}{s} - \frac{x_2}{1-s}\right) = f(x_1, x_2),$$

$$\frac{dx_2}{dt} = rx_2\left(1 - \frac{x_2}{(1-s)}\right) + \sigma\left(\frac{x_1}{s} - \frac{x_2}{1-s}\right) - qEx_2 = g(x_1, x_2), \tag{3}$$

where σ is the migration coefficient.

3. Behavior under harvesting

The behavior of a nonlinear system (3) can be characterized by standard equilibrium, stability, and bifurcation analyses.

In the following theorem, we show that all the solutions of the model under study are bounded.

Theorem 1. *All the solutions of system* (3) *that initiate in* R_2^+ *are uniformly bounded.*

Proof. We define the function

$$W = x_1 + x_2. \tag{4}$$

The time derivative of W along a solution of system (3) is

$$\dot{W} = rx_1\left(1 - \frac{x_1}{s}\right) + rx_2\left(1 - \frac{x_2}{(1-s)}\right) - qEx_2.$$

For each $\tau > 0$, we have that

$$\begin{aligned}\dot{W} + \tau W &= rx_1\left(1 - \frac{x_1}{s}\right) + rx_2\left(1 - \frac{x_2}{(1-s)}\right) \\ &\quad - qEx_2 + \tau x_1 + \tau x_2 \\ &\leqslant \frac{rs}{4}\left(1 + \frac{\tau}{r}\right)^2 + \frac{r(1-s)}{4}\left(1 - \frac{qE}{r} + \frac{\tau}{r}\right)^2. \end{aligned} \tag{5}$$

Thus, there exists a $\mu > 0$ with $\dot{W} + \tau W < \mu$. Applying the theorem on differential inequality (Birkhoff and Rota, 1982), we obtain

$$\begin{aligned}0 &\leqslant W(x_1, x_2) \leqslant (\mu/\tau)(1 - \exp(-\tau t)) \\ &\quad + W(x_1(0), x_2(0)) \exp(-\tau t).\end{aligned}$$

And, for $t \to \infty$, we have $0 \leqslant W \leqslant \mu/\tau$.

Therefore, we have

$$B = [(x_1, x_2) \in R_2^+ : W < \mu/\tau + \varepsilon, \text{ for any } \varepsilon > 0],$$

where B is the region in which all the solutions of system (3) that start in R_2^+ are confined. □

3.1. Equilibrium and stability

We now check whether the model has an equilibrium point with positive subpopulations, as well as whether this equilibrium is stable. The equilibrium for each of the subpopulations requires $(\mathrm{d}x_1/\mathrm{d}t) = (\mathrm{d}x_2/\mathrm{d}t) = 0$, and this obviously implies $\mathrm{d}x/\mathrm{d}t = 0$.

From Eq. (3), we get

$$x_2 = \phi(x_1) = Ax_1(x_1 - B), \tag{6}$$

$$x_1 = \varphi(x_2) = Cx_2(x_2 - D), \tag{7}$$

where the parameters are defined as

$$A = \frac{(1-s)r}{s\sigma}, \quad B = s - \frac{\sigma}{r}, \quad c = \frac{sr}{(1-s)\sigma},$$
$$D = 1 - s - \frac{\sigma}{r} - \frac{qE(1-s)}{r}.$$

Explicit biological interpretations of the parameters A, B, C, and D are not available, but these conditions must be satisfied by the parameters if the equilibrium point is to exist.

Here $A > 0$ and $C > 0$. Curve (6), which expresses the isocline $\mathrm{d}x_1/\mathrm{d}t = 0$, we refer to as S_1, and we refer to curve (7), which expresses the isocline $\mathrm{d}x_2/\mathrm{d}t = 0$, as S_2.

Both of these curves are parabolas. The axis of curve S_1 is parallel to the x_2-axis, and the axis of curve S_2 is parallel to the x_1-axis. If $B > 0$, curve S_1 enters the positive quadrant at $(B, 0)$, and, if $B < 0$, then it enters through the origin. Similarly, if $D > 0$, the curve S_2 enters the positive quadrant at $(0, D)$, and, if $D < 0$, then it enters the positive quadrant through the origin. All of these possibilities are illustrated in Figs. 1–5.

Given these possible cases, it is clear that there is a chance of extinction only when the parameters B and D are both negative (see Fig. 2). Hence, the question of how to avoid extinction by harvesting remains to be answered. With reference to Fig. 1, we consider the slopes of S_1 and S_2 at the origin. If

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1}(0)\Big|_{S_1} < \frac{\mathrm{d}x_2}{\mathrm{d}x_1}(0)\Big|_{S_2}, \tag{8}$$

then there exists a positive equilibrium (x_1^*, x_2^*), while, for the converse of inequality (8), there is no positive equilibrium (Fig. 2).

For the existence of a positive equilibrium, we require, from (6), (7), and (8), that

$$E < E_c = \frac{r}{q} - \frac{s\sigma}{q(s - \sigma/r)(1-s)} > \frac{r}{q}, \tag{9}$$

since $B = s - (\sigma/r) < 0$.

Differentiating E_c with respect to s, we observe that, for a fixed $\sigma \in (0, r)$, E_c increases for $s \in (0, \sqrt{\sigma/r})$, decreases for $s \in (\sqrt{(\sigma/r)}, 1)$, and attains its maximum at $s = \sqrt{\sigma/r}$. When $\sigma \geqslant r$, E_c increases with s. On the other hand, differentiating E_c with respect to σ, we observe that, for fixed $s \in (0, 1)$, E_c decreases with σ.

The dynamic behavior of the equilibria can be studied by computing the variational matrix corresponding to each equilibrium. In particular, the equilibrium (0, 0) is

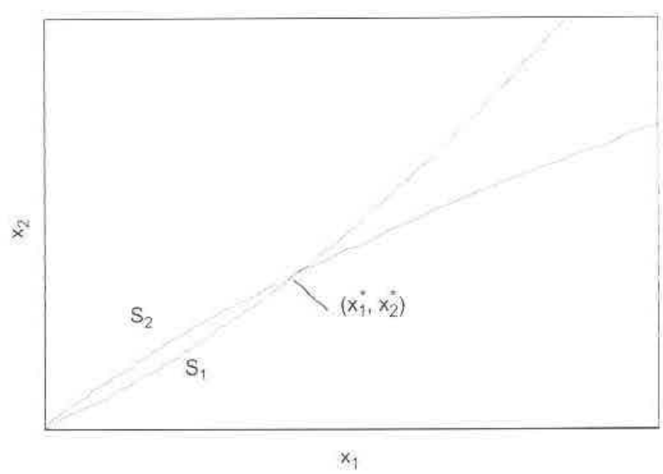


Fig. 1. Case of $B<0$, $D<0$, where a positive equilibrium exists.

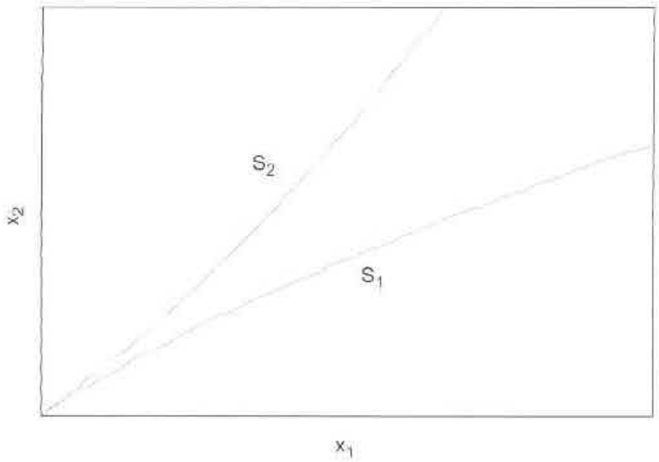


Fig. 2. Case of $B<0$, $D<0$, where no positive equilibrium exists.

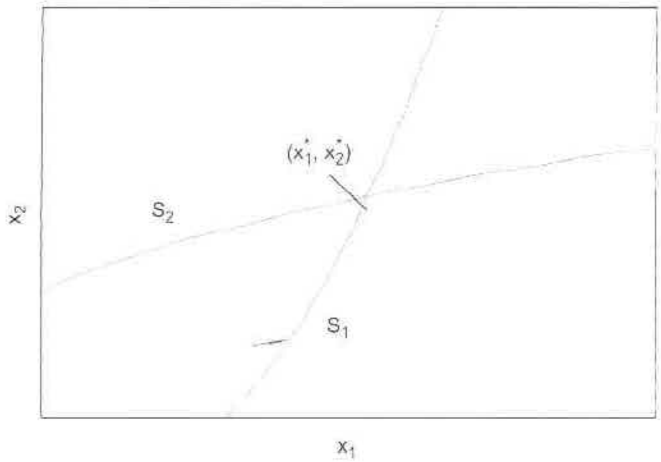


Fig. 5. Case of $B>0$, $D>0$.

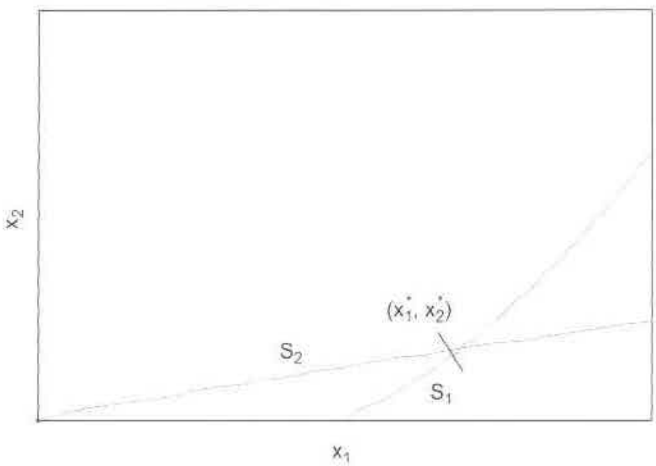


Fig. 3. Case of $B>0$, $D<0$.

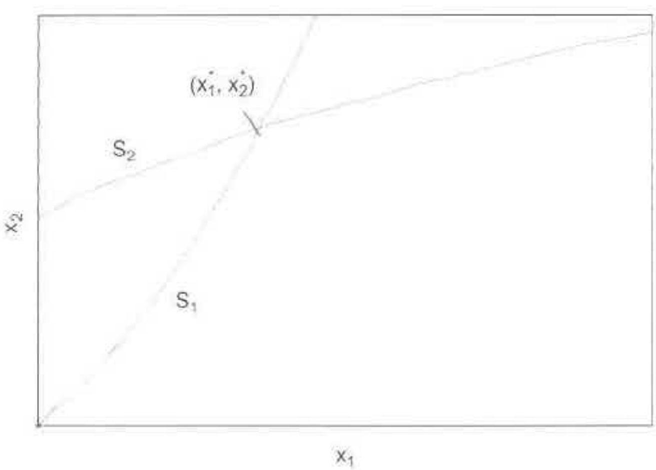


Fig. 4. Case of $B<0$, $D>0$.

unstable. Using the Routh–Hurwitz criteria, it is easy to check that all eigenvalues of the variational matrix corresponding to the interior equilibrium have negative real parts, and, hence, that it is locally asymptotically stable. In addition, the interior equilibrium point is globally asymptotically stable (see appendix).

From system (3), we see that the sustainable yield for the model is given by

$$Y = rx_1\left(1 - \frac{x_1}{s}\right) + rx_2\left(1 - \frac{x_2}{1-s}\right) \tag{10}$$

which indicates that migration does not matter for the sustainable yield level, even though the relative distribution of the population in the two sub-areas does matter.

4. The effects of a protected patch on the population

From Fig. 2, it is clear that, if $E>(r/q) - [s\sigma/q(s-(\sigma/r))(1-s)]$, the introduction of a protected patch is not sufficient to prevent extinction. Therefore, we can conclude that, if E increases in $(0, (r/q) - [s\sigma/q(s-(\sigma/r))(1-s)])$, the interior equilibrium (x_1^*, x_2^*) decreases monotonically and, ultimately, that $(x_1^*, x_2^*) \to (0,0)$. So, if the harvesting effort is large $(E>(r/q) - [s\sigma/q(s-(\sigma/r))(1-s)])$, the population becomes extinct. In the absence of any protected patches, the population inevitably becomes extinct as long as $E \geqslant (r/q)$. In this case, a protected patch is still viable, even though it does not prevent extinction in all situations, because it prevents extinction for

$$\frac{r}{q} < E < \frac{r}{q} - \frac{s\sigma}{q\left(s - \frac{\sigma}{r}\right)(1-s)}.$$

Figs. 6 and 7 clearly indicate that larger reserves always increase biomass; however, a larger reserve does not necessarily increase the harvest. This result is obtained because, although a larger reserve increases spillovers to the fishery, it also reduces the proportion of the population that is subject to harvesting. Putting these pieces together, we can draw the final conclusion that the maintenance of a protected region is practicable in the management of resource populations, although in some cases extinction cannot be prevented.

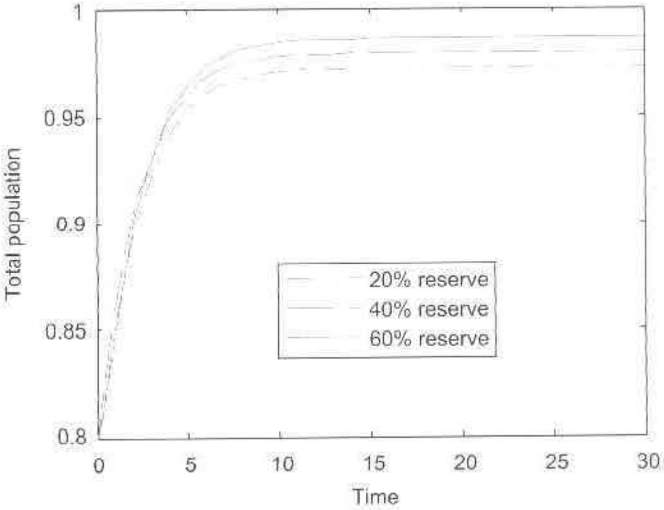


Fig. 6. The relationship between total population and reserve size. The parameter values are $r = 0.5$, $q = 0.1$, $\sigma = 0.4$, and $E = 0.1738$.

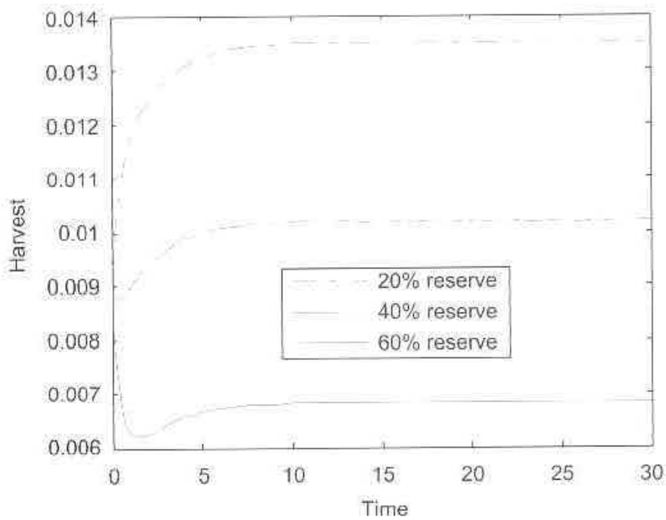


Fig. 7. Harvest against time. The parameter values are the same as in Fig. 6.

In practice, environmental conservation should be valued more highly than the economic benefits of resource management, since it is generally unwise to eliminate any resource population. The ultimate purpose of establishing protected patches for resource populations is to produce greater overall benefits.

5. Optimal harvesting: dynamic optimization

5.1. Problem formulation and necessary conditions

Suppose that a governing body manages the fishery by setting up a reserve zone to control the harvesting of fish. Its objective is to exploit the available resources optimally. The present value J of a continuous time-stream of revenues is given by

$$J = \max_{E} \int_0^\infty e^{-\delta t}(pqx_2(t) - c(x_2))E(t)\,dt, \tag{11}$$

where δ is the instantaneous rate of annual discount, p the price of a unit of fish, and $c(x_2)$ the cost of harvesting. The cost of fishing is assumed to decrease with the stock size, i.e., $\partial c(x_2)/\partial x_2 \leqslant 0$. Thus, our objective is to maximize J subject to the state Eqs. (3) and the control constraint that

$$0 \leqslant E \leqslant E_{\max} < E_c. \tag{12}$$

Applying Pontryagin's maximum principle (Kamien and Schwartz, 1981), we find that the current value Hamiltonian is given by

$$H = (pqx_2(t) - c(x_2))E(t) + \lambda_1 f(x_1, x_2) + \lambda_2 g(x_1, x_2). \tag{13}$$

The equations that form the necessary conditions for a solution are as follows:

- condition for a maximum:

$$\left(\max_E\right) H \Rightarrow \left(\max_E\right)(pqx_2(t) - c(x_2) - \lambda_2 qx_2)E(t),$$
$$0 \leqslant E \leqslant E_{\max}, \tag{14}$$

where $E_{\max}$ is the maximum harvest rate.

- equations of motion for the co-state or shadow prices:

$$\dot{\lambda}_1 = \delta\lambda_1 - \frac{\partial H}{\partial x_1} = \delta\lambda_1 - \lambda_1\frac{\partial f}{\partial x_1} - \lambda_2\frac{\partial g}{\partial x_1} = \delta\lambda_1 - \lambda_1\left(r - \frac{2rx_1}{s} - \frac{\sigma}{s}\right) - \lambda_2\frac{\sigma}{s}, \tag{15}$$

$$\dot{\lambda}_2 = \delta\lambda_2 - \frac{\partial H}{\partial x_2} = \delta\lambda_2 - pqE + \frac{\partial c}{\partial x_2}E - \lambda_1\frac{\partial f}{\partial x_2} - \lambda_2\frac{\partial g}{\partial x_2} = \delta\lambda_2 - pqE + \frac{\partial c}{\partial x_2}E - \lambda_1\frac{\sigma}{1-s} - \lambda_2\left(r - \frac{2rx_2}{1-s} - \frac{\sigma}{1-s} - qE\right), \tag{16}$$

- transversality conditions, since $x_1 \geqslant 0$ and $x_2 \geqslant 0$:

$$\lim_{t\to\infty}\lambda_1 \geqslant 0, \quad \lim_{t\to\infty}\lambda_2 \geqslant 0.$$

Given the linear form of the harvesting cost function, $c(x_2)E$, the Hamiltonian (13) depends linearly on E with coefficient $(pqx_2 - (c(x_2) - \lambda_2 qx_2))$. Consequently, its maximum value is reached for the extremes of E; i.e., the harvest rate must be either 0 or $E_{\max}$. This observation leads to the rule that one must fish as much as possible when the shadow price of fish is sufficiently low $(\lambda_2 < p - (c(x_2)/qx_2))$, and not fish at all when the shadow price is sufficiently high$(\lambda_2 > p - (c(x_2)/qx_2))$. Furthermore, when $\lambda_2 = p - (c(x_2)/qx_2)$, the harvest rate is undetermined. In this case, three solutions for E are possible, namely 0, $E_{\max}$, or $\tilde{E}$, which is the singular control that maintains the condition $\lambda_2 = p - (c(x_2)/qx_2)$. Therefore, the optimal control path will be either "bang–bang" (i.e., harvesting maximally, not harvesting at all, or alternating between the two) or singular (i.e., equating revenues with the shadow price).

We assume that there exists a unique optimal path. After an initial period, the optimal trajectory of the system will approach either an equilibrium or a cycle. Suppose that the optimal path does not approach an equilibrium; in this case, it must cross itself. At this point, the optimal path must continue as it did before, or it will not be unique. This implies a cycle. So, finally, the optimal trajectory will reach either an equilibrium or a cycle. We refer to this state as the end state. The approach path is the beginning of the optimal trajectory prior to the end state being reached.

We devote the remainder of this section to examining the end state. Remember that the harvest rate of an optimal trajectory, and thus of the end state, must be "bang–bang," singular, or a combination of the two. In consequence, the following four end states are conceivable:

(1) *No harvesting*: $E = 0$ and $\lambda_2 \geqslant p - (c(x_2)/qx_2)$. The end state is an equilibrium, and the harvest rate is part of a "bang–bang" control (see Section 5).
(2) *Maximum harvesting*: $E = E_{\max}$ and $\lambda_2 \leqslant p - (c(x_2)/qx_2)$. The end state is an equilibrium, and the harvest rate is part of a "bang–bang" control (see Section 5).
(3) *A singular state*: $E = \tilde{E}$ and $\lambda_2 = p - (c(x_2)/qx_2)$ (see Section 5). A singular harvest rate is applied. This can result in two types of singular equilibria (see Section 5).
(4) A "bang–bang" cycle, which is an oscillation that is controlled by a harvest rate that alternates among the maximum, zero, and possibly a singular harvest rate:

$$E = E_{\max} \quad \text{when } \lambda_2 < p - \frac{c(x_2)}{qx_2},$$
$$E = 0 \quad \text{when } \lambda_2 > p - \frac{c(x_2)}{qx_2},$$
$$E = \tilde{E} \quad \text{when } \lambda_2 < p - (c(x_2)/qx_2).$$

No harvesting (*Case* 1): The first possibility is straightforward. Not harvesting is optimal when, at the equilibrium (x_1^*, x_2^*), the total cost of fishing $c(x_2) + \lambda_2$ exceeds the price of fish. This means that, at any harvest rate, the loss in the social value of fish exceeds the net gain from the fishery.

Maximum harvesting (*Case* 2): The second possibility is to continue harvesting at the maximum level. In this case, the price of fish must exceed the total costs. If the maximum harvest level is relatively small ($E_{\max} < E_c$), then the system asymptotically approaches the equilibrium (x_1^*, x_2^*).

A singular case: The third possibility is an end state in which the total system (state and co-state) remains in a singular state. From condition (14) it follows that, at this end state, we have to satisfy

$$\lambda_2 = p - \frac{c(x_2)}{qx_2}. \tag{17}$$

Substituting (17) and its derivative, $\dot{\lambda}_2 = -(1/q)[(x_2(\partial c/\partial x_2) - c(x_2))/(x_2^2)]g(x_1, x_2)$, into (16) yields the following expression:

$$\lambda_1 = \left[\delta x_2 - \left(pq - \frac{\partial c}{\partial x_2}\right)E - \left(p - \frac{c}{qx_2}\right)\frac{\partial g}{\partial x_2} + \frac{g(x_1, x_2)}{q}\frac{x_2(\partial c/\partial x_2) - c(x_2)}{x_2^2}\right] \Big/ \frac{\partial f}{\partial x_2}. \tag{18}$$

We take the time derivative of expression (18) for λ_1 and substitute it, together with (17), into (15). This action eliminates both shadow prices, and we obtain the following expression, which implicitly defines $\tilde{E}$, the harvest rate in the singular state:

$$\begin{aligned}
&\left[\delta q x_2^4 + x_2^4 \frac{\partial^2 c}{\partial x_2^2} q\tilde{E} - pqx_2^4 \frac{\partial^2 g}{\partial x_2^2} - c\frac{\partial g}{\partial x_2}x^2\right.\\
&+ cx_2^3 \frac{\partial^2 g}{\partial x_2^2} + x_2^3 \frac{\partial^2 c}{\partial x_2^2} g - 2x_2^2 \frac{\partial c}{\partial x_2} g + 2cx_2 g\\
&\left.+\left(x_2^3 \frac{\partial c}{\partial x} - x_2^2 c\right)\frac{\partial g}{\partial x_2}\right] g \frac{\partial f}{\partial x_2}\\
&- \left[\delta x_2 - (pq - \frac{\partial c}{\partial x_2})\tilde{E} - (p - \frac{c}{qx_2})\frac{\partial g}{\partial x_2}\right.\\
&\left.+ \frac{g(x_1, x_2)}{q}\frac{x_2(\partial c/\partial x_2) - c(x_2)}{x_2^2}\right] gqx_2^4 \frac{\partial^2 f}{\partial x_2^2}\\
={}& qx_2^4\left(\delta - \frac{\partial f}{\partial x_1}\right)\left[\delta x_2 - \left(pq - \frac{\partial c}{\partial x_2}\right)\right.\\
&\left.\tilde{E} - \left(p - \frac{c}{qx_2}\right)\frac{\partial g}{\partial x_2} + \frac{g(x_1, x_2)}{q}\frac{x_2(\partial c/\partial x_2) - c(x_2)}{x_2^2}\right]\\
&\times \frac{\partial f}{\partial x_2} - qx_2^4 \frac{\partial g}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)^2\left(p - \frac{c}{qx_2}\right).
\end{aligned} \tag{19}$$

We have ignored the function's arguments so as not to further complicate the expression. Eq. (19) indicates that it is possible to identify a singular harvest rate for every point in the phase diagram. Be aware, however, that $0 < \tilde{E} < E_{\max}$, and, thus, it may not be feasible to find an $\tilde{E}$ for every value of x_1 and x_2.

If a singular harvest rate is employed, an autonomous system results that describes the singular trajectories:

$$\begin{aligned}
\frac{dx_1}{dt} &= rx_1\left(1 - \frac{x_1}{s}\right) - \sigma\left(\frac{x_1}{s} - \frac{x_2}{1-s}\right)\\
\frac{dx_2}{dt} &= rx_2\left(1 - \frac{x_2}{1-s}\right) + \sigma\left(\frac{x_1}{s} - \frac{x_2}{1-s}\right) - q\tilde{E}x_2.
\end{aligned} \tag{20}$$

Here $\tilde{E}$ is the singular harvest rate, which is implicitly given by Eq. (19). From system (20), it follows that, in equilibrium,
$\tilde{E} = (1/q)[r(1 - x_2^*/1 - s) + (\sigma/x_2^*)((x_1^*/s) - (x_2^*/1 - s))]$,
which is independent of time. Thus, the singular harvest rate is constant in equilibrium. Therefore, the equilibria of

the singular system must be equal to the equilibria under fixed harvest rates, as noted in Section 3. Semmler and Sieveking (1994) showed that constant optimal harvesting may push a predator–prey system into cyclical behavior, whereas such a system would reach equilibrium in the absence of harvesting. This result does not hold for the system studied here.

6. Optimal harvesting: equilibrium solutions

We will henceforth analyze a less general case. We take the cost of harvesting to be $c(x_2) = c$ (a constant). The Hamiltonian is given by Eq. (13). From the necessary conditions for optimality (Clark, 1990), the following condition is valid along the optimal solution:

$$\frac{\partial H}{\partial E} = pqx_2 - c - \lambda_2 qx_2 = 0, \tag{21}$$

where the dynamics of the associated shadow prices are described by Eqs. (15) and (16). We observe that, along an optimal equilibrium solution of the problem under consideration, we have

$$\delta\lambda_1 - \lambda_1\left(r - \frac{2rx_1}{s} - \frac{\sigma}{s}\right) - \lambda_2\frac{\sigma}{s} = 0, \tag{22}$$

$$\delta\lambda_2 - pqE - \lambda_1\frac{\sigma}{1-s} - \lambda_2\left(r - \frac{2rx_2}{1-s} - \frac{\sigma}{1-s} - qE\right) = 0. \tag{23}$$

In view of Eq. (21), if the equilibrium solution $(x_1^*(E), x_2^*(E))$ satisfies

$$\delta - r + \frac{2rx_1^*(E)}{s} + \frac{\sigma}{s} \neq 0,$$

then we have

$$\left(\delta - r + \frac{2rx_1^*(E)}{s} + \frac{\sigma}{s}\right)\left[\left(p - \frac{c}{qx_2^*}\right)\right.$$
$$\left.\times\left(\delta - r + \frac{2rx_2^*}{1-s} + \frac{\sigma}{1-s} - \frac{cE}{x_2}\right)\right]$$
$$-\frac{\sigma^2}{s(1-s)}\left(p - \frac{c}{qx_2^*}\right) = 0. \tag{24}$$

Alternatively, along the optimal solution, we have (from (22) and (23))

$$\lambda_1 = \frac{\sigma pqE}{sD}$$

and

$$\lambda_2 = \frac{pqE(\delta - r + (2rx_1^*(E)/s) + (\sigma/s))}{D},$$

where

$$D = \left(\delta + \frac{rx_1^*(E)}{s} + \frac{\sigma}{1-s}\frac{x_2^*}{x_1^*}\right)\left(\delta + \frac{rx_2^*}{1-s} + \frac{\sigma}{s}\frac{x_1^*}{x_2^*}\right)$$
$$-\frac{\sigma^2}{s(1-s)} > 0.$$

At the optimal steady state, the optimal fishing effort can be derived as

$$E^* = \frac{1}{q}\left[r\left(1 - \frac{x_2^*}{1-s}\right) + \frac{\sigma}{x_2^*}\left(\frac{x_1^*}{s} - \frac{x_2^*}{1-s}\right)\right],$$

while the associated profit is

$$J^* = \frac{1}{q}\left[r\left(1 - \frac{x_2^*}{1-s}\right) + \frac{\sigma}{x_2^*}\left(\frac{x_1^*}{s} - \frac{x_2^*}{1-s}\right)\right]\frac{pqx_2^* - c}{\delta}.$$

Our job is now to reach the optimal solution optimally from the initial state $(x_1(0), x_2(0))$. This can be achieved by applying a "bang–bang" control (Pontryagin et al., 1962) to the system, as follows. Define

$$\tilde{E}(t) = \begin{cases} E_{\max} & \text{for } S(t) > 0, \\ E_{\min} & \text{for } S(t) < 0, \end{cases}$$

where $S(t) = pqx_2 - c - \lambda_2 qx_2$. Moreover, let T, be the time at which the path $(x_1(t), x_2(t))$, which is generated via the "bang–bang" control $E(t) = \tilde{E}(t)$, reaches the state (x_1^*, x_2^*). Then, the optimal control policy is

$$E(t) = \begin{cases} \tilde{E} & \text{for } 0 \leqslant t \leqslant T, \\ E^* & \text{for } t > T \end{cases}$$

and the optimal path is given by the trajectory generated by the above optimal control. In view of the global stability property of the interior equilibrium of system (3), we can also reach the singular optimal solution through a suboptimal path by choosing the control policy $E(t)$ to be equal to E^* for all t. The advantage of choosing the optimal path is that it leads to the optimal singular solution more rapidly than does the suboptimal path.

7. Numerical simulations

We now illustrate the case of optimal harvesting at equilibrium numerically by using MATLAB. We specify $r = 0.5$, $s = 0.4$, $c = 0.4$, $q = 0.1$, $\delta = 0.005$, $\sigma = 0.4$, and

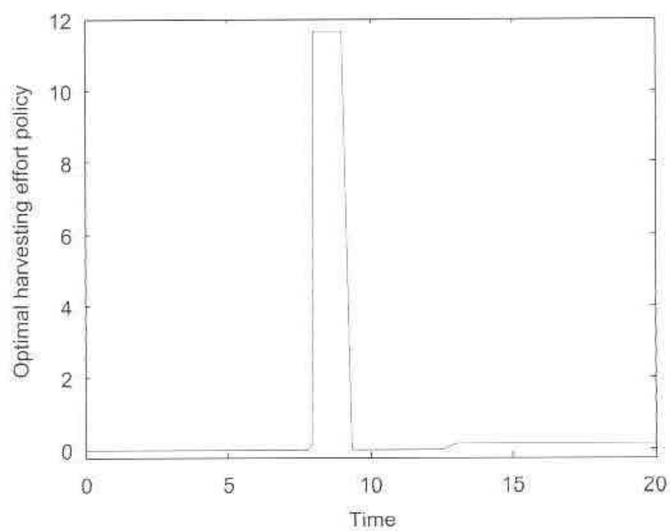


Fig. 8. Optimal harvesting effort policy.

$p = 7$. In addition, we set $E_{\min} = 0$ and $E_{\max} = 11.67$. For the above values of the parameters, we obtain an optimal singular effort of 0.17, an optimal singular stock of 0.39, 0.58, and an optimal singular revenue of 0.34.

The optimal harvest policy is presented in Fig. 8, and the associated optimal and suboptimal paths are presented in Figs. 9 and 10. From Figs. 9 and 10, it is clear that the optimal and suboptimal paths approach their respective singular optimal solutions in both zones.

The details of the optimal and suboptimal paths presented in Figs. 9 and 10 are tabulated in Table 1. From this information, it is clear that the suboptimal paths take longer to reach the optimal singular solution than do the optimal paths.

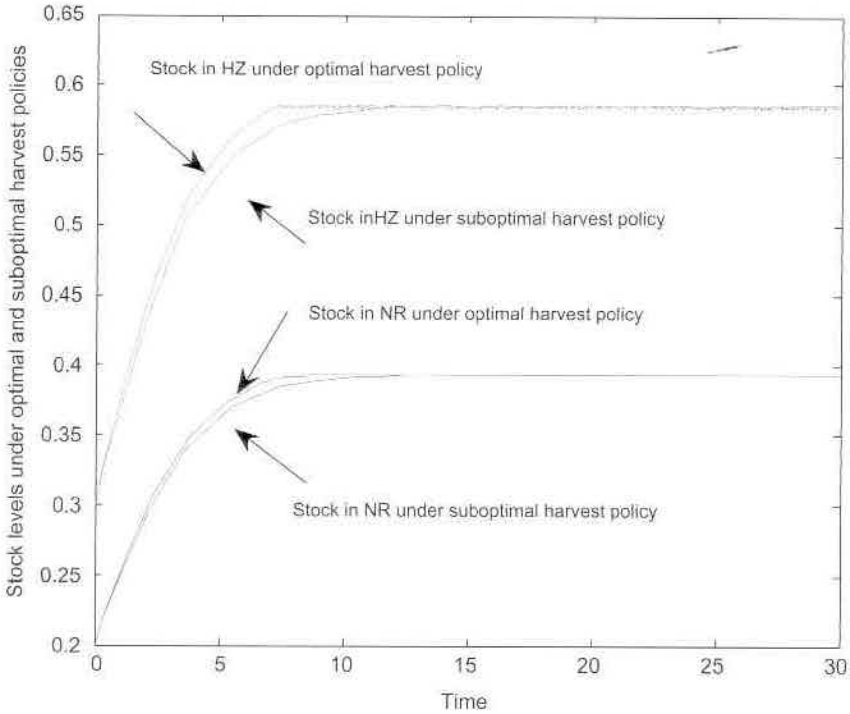


Fig. 9. This figure illustrates the optimal and suboptimal approach paths initiated at (0.2, 0.3).

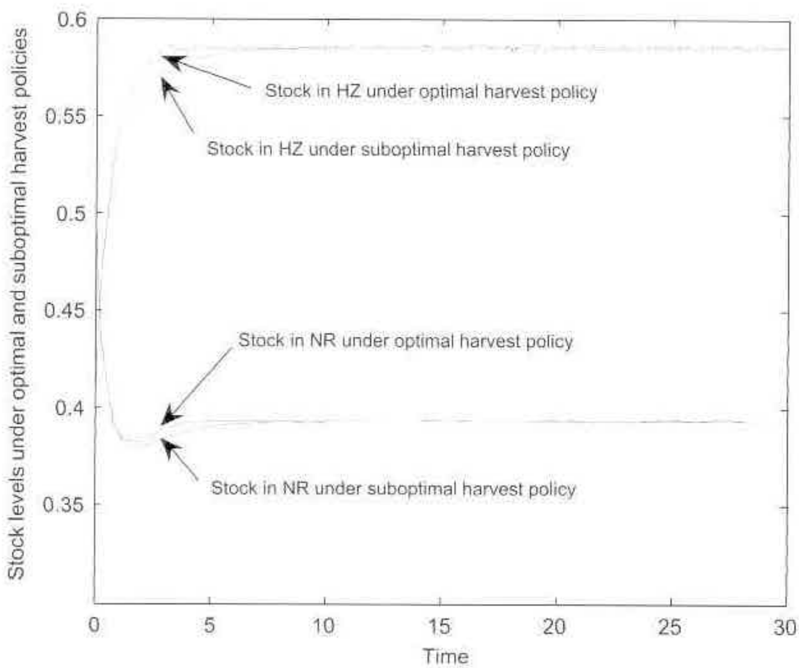


Fig. 10. This figure illustrates the optimal and suboptimal approach path initiated at (0.5, 0.4).

Table 1
Time taken by different approach paths

	Coordinates		Path type	Time taken to reach the end point
	Initial point	End point		
Fig. 9	(0.2, 0.3)	(0.17, 0.58)	Optimal	6.77 units
	(0.2, 0.3)	(0.17, 0.58)	Suboptimal	10.54 units
Fig. 10	(0.5, 0.4)	(0.17, 0.58)	Optimal	2.6 units
	(0.5, 0.4)	(0.17, 0.58)	Suboptimal	5.6 units

8. Concluding remarks

In this paper, we have discussed the impacts of MPAs from both economic and biological perspectives. The results of the analysis indicate that MPAs are a practical means of managing resource populations and are, therefore, beneficial for conserving the ecological environment and resource populations, despite the fact that extinction cannot be prevented in all cases. The establishment of MPAs could be used to maintain a high fish biomass in marine habitats. We have also proved that, when it exists, the interior equilibrium point is globally asymptotically stable. Ecological managers may find it desirable to achieve a unique positive equilibrium that is globally asymptotically stable, and they may use this goal in creating harvesting guidelines and developing the ecosystem in a sustainable manner.

We have also discussed the dynamic optimization of the harvest policy by taking $E(t)$, the harvest effort, as a dynamic variable. The optimal solution in the equilibrium case has also been discussed. Moreover, the biological and bioeconomic interpretations of the results associated with the optimal solutions have been explained.

The simulations (illustrated in Figs. 6 and 7) indicated that a larger reserve always increases resilience; however, a larger reserve does not necessarily increase the harvest. This result is obtained because, although a larger reserve increases spillovers to the fishery, it also reduces the proportion of the population that is subject to harvesting. We have also observed that the optimal and suboptimal paths approach their respective singular optimal solutions in both zones (see Figs. 9 and 10). From Table 1, it is clear that the suboptimal paths take a longer time to reach the optimal singular solution than do the optimal paths.

The methods and results provided in this paper do not constitute a comprehensive analysis. Nevertheless, we believe that our study is an important contribution toward the protection of fisheries, as it approximately characterizes the key level of harvesting.

Acknowledgements

We would like to thank the Japanese Society for the Promotion of Science for its financial support of this research (P05109). We would also like to thank three anonymous referees for their useful comments.

Appendix

To prove that the system is globally stable, we prove that the system does not tend to a limit cycle. Using the Bendixson–Dulac criterion (Strogatr, 1994), we can prove that the system does not have a limit cycle in the phase space. If such a closed trajectory C exists, then $\oint_C \left(\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \cdot \bar{n} \right) \mathrm{d}l = 0$, where $\bar{n}$ is the outward normal on C. The dot product must equal zero, because the trajectory follows C. Green's theorem yields $\iint_S \nabla \cdot \left(g(x_1, x_2) \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \right)$ $= \oint_C g(x_1, x_2) \left(\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \cdot \bar{n} \right) \mathrm{d}l$, where S is the surface enclosed by C.

Therefore, if we can find a function $g(x_1, x_2)$ for which the sign of the integrand is always positive or always negative over at least S, then the surface integral must not be equal to zero. Consequently, this means that C is not a trajectory. Taking $g(x_1, x_2) = 1/x_1 x_2$, we have that

$$\nabla \cdot \left(g(x_1, x_2) \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \right) = -\frac{1}{x_1} \left[\frac{r}{1-s} + \frac{\sigma x_1}{(1-s) x_2^2} \right] - \frac{1}{x_2} \left[\frac{r}{s} + \frac{\sigma x_2}{s x_1^2} \right] < 0.$$

This implies that the system has no limit cycle. Hence, the interior equilibrium is globally asymptotically stable.

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, see:

http://www.textcheck.com/cgi-bin/certificate.cgi?id=4uwDbo

References

Allison, G.D., Lubchenco, J., Carr, M.H., 1998. Marine reserves are necessary but not sufficient for marine conservation. Ecological Applications 8 (Suppl. 1), 79–92.

Anderson, L., 2002. A bioeconomic analysis of marine reserves. Natural Resource Modelling 15 (3), 311–334.

Birkhoff, G., Rota, G.S., 1982. Ordinary Differential Equations. Ginn, Boston.

Bonocoeur, J., Alban, F., Guyader, O., Thebaud, O., 2002. Fish, fishers, seals and tourists: economic consequences of creating a marine reserve in a multi-species, multi-activity context. Natural Resource Modelling 25 (4), 1–25.

Brown, G.M., Roughgarden, J., 1997. A metapopulation model with private property and a common pool. Ecological Economics 22, 65–71.

Clark, C.W., 1985. Bioeconomic Modelling and Fisheries Management. Wiley, New York.

Clark, C.W., 1990. Mathematical Bioeconomics: The Optimal Management of Renewable Resources. Wiley, New York.

Conrad, J.M., 1999. The bioeconomics of marine sanctuaries. Journal of Bioeconomics 1, 205–217.

Gell, F.R., Roberts, C.M., 2002. The fishery effects of marine reserves and fishery closures. World Wildlife Fund. 〈http://www.worldwildlife.org/oceans/fishery_effects.pdf〉.

Guenette, S., Lauck, T., Clark, C.W., 1998. Marine reserves: from Beverton and Holt to the present. Reviews in Fish Biology and Fisheries 8, 251–272.

Halpern, B.S., 2003. The impact of marine reserves: do reserves work and does reserve size matter? Ecological Applications 13 (Suppl. 1), S117–S137.

Hannesson, R., 1998. Marine reserves: what would they accomplish? Marine Resource Economics 13, 159–170.

Hilborn, R., Micheli, F., De Leo, G., 2006. Integrating marine protected areas with catch regulation. Canadian Journal of Fisheries and Aquatic Sciences 63, 642–649.

Holland, D.S., Brazee, R.J., 1996. Marine reserves for fisheries management. Marine Resource Economics 11, 157–171.

Kamien, M.I., Schwartz, N.L., 1981. Dynamic Optimization: The Calculus of Variations and Optimal Control, vol. 4. Elsevier Science Publishers, Amsterdam.

Luck, T., Clark, C.W., Mangel, M., Munro, G.R., 1998. Implementing the precautionary principles in fisheries management through marine reserves. Ecological Applications 8 (1), 72–78.

Ludwig, D., Hilborn, R., Walters, C., 1993. Uncertainty, resource exploitation, and conservation: lessons from history. Science 260, 36–49.

Neubert, M.G., 2003. Marine reserves and optimal harvesting. Ecology Letters 6 (9), 843–849.

Polacheck, T., 1990. Year-round closed areas as a management tool. Natural Resource Modelling 4, 327–354.

Pontryagin, L.S., Boltyonskii, U.G., Gamkrelidre, R.V., Mishchenko, E.F., 1962. The Mathematical Theory of Optimal Processes. Wiley, New York.

Rodwell, L., Barbier, E., Roberts, C., McClanahan, T., 2002. A model of tropical marine reserve-fishery linkages. Natural Resource Modeling 15 (4), 453–486.

Sanchirico, J.N., Wilen, J.E., 1999. Bioeconomics of spatial exploitation in a patchy environment. Journal of the Environmental Economics and Management 37, 129–150.

Sanchirico, J.N., Wilen, J.E., 2001. A bioeconomic model of marine reserve creation. Journal of Environmental Economics and Management 42, 257–276.

Sanchirico, J.N., Wilen, J.E., 2005. Optimal spatial management of renewable resources: matching policy scope to ecosystem scale. Journal of Environmental Economics and Management 50, 23–46.

Semmler, W., Sieveking, M., 1994. On the optimal exploitation of interacting resources. Journal of Economics 59, 23–49.

Strogatr, S.H., 1994. Nonlinear Dynamics and Chaos. Addison-Wesley, Reading, MA.

Suman, D., Shivlani, M., Milon, J.W., 1999. Perceptions and attitudes regarding marine reserves: a comparison of stockholders in the Florida Keys National Marine Sanctuary. Ocean & Coastal Management 42, 1019–1040.

ELSEVIER

Journal of Environmental Management 86 (2008) 181–192

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Classifying environmentally significant urban land uses with satellite imagery

Mi-Hyun Park*, Michael K. Stenstrom

Department of Civil and Environmental Engineering, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA, 90095-1593, USA

Received 7 April 2006; received in revised form 27 November 2006; accepted 5 December 2006
Available online 8 February 2007

Abstract

We investigated Bayesian networks to classify urban land use from satellite imagery. Landsat Enhanced Thematic Mapper Plus (ETM^+) images were used for the classification in two study areas: (1) Marina del Rey and its vicinity in the Santa Monica Bay Watershed, CA and (2) drainage basins adjacent to the Sweetwater Reservoir in San Diego, CA. Bayesian networks provided 80–95% classification accuracy for urban land use using four different classification systems. The classifications were robust with small training data sets with normal and reduced radiometric resolution. The networks needed only 5% of the total data (i.e., 1500 pixels) for sample size and only 5- or 6-bit information for accurate classification. The network explicitly showed the relationship among variables from its structure and was also capable of utilizing information from non-spectral data. The classification can be used to provide timely and inexpensive land use information over large areas for environmental purposes such as estimating stormwater pollutant loads.

Keywords: Urban; Land use classification; Satellite imagery; Remote sensing; Bayesian networks

1. Introduction

Human activity has impacted the quality of receiving waters such as rivers, lakes, and oceans. Urbanization has deteriorated water quality by discharging more pollutants and higher volumes of runoff at accelerated rates. Recent environmental concerns are focusing on diffuse source pollution due to its significant impact on the receiving waters, and progress in managing point source pollution. However, diffuse sources of many pollutants, e.g., suspended solids, nutrients, heavy metals, oil and grease, and pesticides, are difficult to measure and are released over wide areas. As conventional monitoring is impractical, alternative management approaches are needed. Characterizing pollution emission rates using the relationship between land use and runoff water quality is one possible method (Stenstrom et al., 1984; Chiew and McMahon, 1999). These studies used land use information to estimate diffuse source pollution to receiving waters, which is particularly useful in unmonitored watersheds or drainage basins. For small drainage basins, ground surveys can be used to obtain land use information but it is too expensive for large drainage basins or watersheds. Moreover, land uses keep changing due to the rapid population growth and ongoing urbanization in many watersheds. In this case, land use information from public records, that may be updated only once per decade, cannot provide timely information. Therefore, using satellite imagery will be useful to classify land uses in the given watershed or drainage basins and to estimate diffuse source pollution.

For the last three decades, many approaches to land cover classification using satellite imagery have been widely explored. For example, Landsat series imagery has been extensively used since the first satellite was launched in 1972. Satellite imagery has been used to mainly classify the biophysical characteristics of land surface. Land use, on the other hand, which is human activity on the land, is difficult to determine directly from satellite imagery. Although land cover and land use have been used interchangeably in some cases (Anderson et al., 1976),

*Corresponding author. Tel.: +1 310 825 1408; fax: +1 310 206 5476.
E-mail addresses: mhpark@seas.ucla.edu (M.-H. Park), stenstro@seas.ucla.edu (M.K. Stenstrom).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.010

they are not necessarily equivalent in many cases (Brown et al., 2000). One land use can consist of various land covers. Conversely, one type of land cover can be found in different land uses. This is especially true in urban areas because of their heterogeneous characteristics. For instance, residential land use can include different roof materials such as asphalt, clay tiles, wood shingles, or metal. A land use also contains a mixture of different urban features with different proportions. For instance, a pixel of residential land use could contain mixed signatures of buildings, pavement, driveways, and vegetation (i.e., grass yards and trees) (Clapham, 2003). The number of mixed pixels can be reduced using higher-resolution images, such as IKONOS and Quickbird, but these data are generally too expensive for watershed scale management. An example of the latter case is commercial and industrial land uses, which can exhibit similar spectral signatures if the roof or pavements use similar materials (Stefanov et al., 2001).

Land use information is important in many applications, i.e., tax assessment, urban planning, and environmental management. For diffuse source pollution management, successful urban land use classification is essential because the pollutant concentrations and runoff rates are correlated to land use and not land cover. Land cover for two different land uses may be impervious, but pollutant loads and environmental impacts can be quite different. For example, the pollutant loads from a local street are quite different from a school playground even though both could be composed of asphalt. In addition, a simple distinction between urban and non-urban land uses does not provide sufficient information for estimating diffuse source pollutant loading. The challenge of inferring different land use classifications from similar spectral signatures may depend upon using other information or ancillary data. Therefore, a method to infer land use directly from satellite imagery is needed for diffuse source pollution management in urban areas.

Bayesian networks, also called probabilistic networks or belief networks, have not been widely used for satellite image classification. However, they are known as excellent tools in many areas, such as medical informatics, natural language processing, computer vision, and decision making (Charniak, 1991; Russell and Norvig, 1995). Bayesian networks may be successful in satellite imagery application because they have been successfully used to classify images and signals (Malka and Lerner, 2004; Boutell and Luo, 2005; Gurwicz and Lerner, 2005). They were developed for reasoning and inference and are able to handle uncertainty with mixed signatures. Only recently have researchers attempted to adopt Bayesian networks for satellite image classification (Pal et al., 2001; Orun, 2004).

In this research, we used modified classification systems based on US Geological Survey (USGS) Levels I to III and evaluated the performance of Bayesian networks. We investigated the minimum sample size to train the network because there is no universal rule or theoretical background for selecting sample size. We examined the optimal number of variable states (i.e., radiometric resolution) since Bayesian networks are known to be more effective with quantized data. We also investigated the performance of Bayesian networks depending on different classification systems. Our hope is that these results will facilitate the use of Bayesian networks for classifying urban land use from satellite imagery.

2. Background of Bayesian networks

Bayesian networks are capable of automatically estimating the probability of target variable values by means of both probabilistic expression (Bayes' theorem) and graphical representation (graph theory). In the probability theory, $P(T)$, the probability of variable T, which lies in between 0 and 1, is equal to 1 if and only if T is certain (Cowell, 1998). If we have two variables, T and M, conditional probability of M given T is

$$P(M|T) = \frac{P(T, M)}{P(T)}. \tag{1}$$

The product rule of probability for two variables, T and M, is expressed as

$$P(T, M) = P(T|M)P(M) = P(M|T)P(T), \tag{2}$$

which defines the joint probabilities of T and M in terms of conditional and marginal probabilities of individual variables. By rearranging Eq. (2), the Bayesian theorem can be derived as

$$P(T|M) = \frac{P(M|T)P(T)}{P(M)}, \tag{3}$$

where $P(T|M)$ is the posterior probability of T given M, $P(T|M)$ is the likelihood of M given T, and $P(T)$ is the prior probability of T. Eq. (3) can be denoted as

$$\text{posterior probability} \propto \text{likelihood} \times \text{prior probability}. \tag{4}$$

The theorem makes it possible to compute the posterior probability of a target variable from the prior probability and likelihood under the conditional independence assumption. For example, if we have a target variable T and variable M_i, given $i = 3$, such that $P(M_i) \neq 0$ for all i and they are mutually exclusive and exhaustive, we can write their relationship in Bayes' theorem as follows:

$$P(T|M_1, M_2, M_3) = \frac{P(M_1, M_2, M_3|T)P(T)}{P(M_1, M_2, M_3)}. \tag{5}$$

If we can assume that M_1, M_2, and M_3 are conditionally independent given T, we can write:

$$P(T|M_1, M_2, M_3) = \frac{P(M_1|T)P(M_2|T)P(M_3|T)P(T)}{P(M_1)P(M_2)P(M_3)}. \tag{6}$$

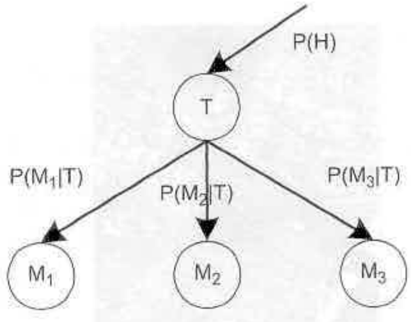


Fig. 1. An example of a Bayesian network.

Being that we are dealing with probability distributions that sum to 1 we can rewrite

$$P(T|M_1, M_2, M_3) = \alpha P(M_1|T)P(M_2|T)P(M_3|T)P(T), \quad (7)$$

where α is set to normalize the probabilities to 1. Because we can estimate the prior probability of $P(T)$ and the likelihoods of $P(M_1|T)$, $P(M_2|T)$, and $P(M_3|T)$ from the given data, we are able to calculate the posterior probability distribution over T for any values of M_1, M_2, and M_3. This equation is graphically represented in Fig. 1.

As shown in Fig. 1, a Bayesian network is a directed acyclic graph (DAG) that consists of nodes and arcs (Pearl, 1988; Neapolitan, 1990). A DAG is a directed graph that does not include any cycles. In a network, a node corresponds to a variable. An arc connects a parent node to a child node and the direction of the arc represents the direction of influence. Therefore, the structure of the network visually shows the relationships between connected nodes. The relationships are quantified by a conditional probability table, in which the probability distribution of a node is conditioned on its parent nodes.

In general, the joint probability distribution between any connected nodes in a network is written as

$$P(X) = \prod_{i=1}^{n} P(x_i|\mathrm{pa}(x_i)), \quad (8)$$

where $X = x_1, \ldots, x_n$ and $\mathrm{pa}(x_i)$ is a parent node of x_i. For example, T is a parent node of M_1, M_2, and M_3 in Fig. 1. In Bayesian networks, there is a conditional independence assumption where a node is conditionally independent of its non-descendant nodes given its parent node. In Fig. 1, M_1 is conditionally independent of M_2, and M_3 given T and vice versa.

$$P(M_1|T, M_2, M_3) = P(M_1|T). \quad (9)$$

3. Study area and data

3.1. Study area

Fig. 2 shows our study areas. The first study area was focused on Marina del Rey and its vicinity. Marina del Rey is located seaward of the Ballona Creek drainage basin, which is the southern part of the Santa Monica Bay watershed. There is growing concern in regards to Santa Monica Bay due to the importance of its water quality and natural resources. It receives the bulk of Los Angeles' treated wastewater and the stormwater runoff from the Los Angeles River via Ballona Creek, which has caused periodic beach closures (NCERQA, 2002). The highly urbanized area in the watershed has impacted the water quality of the Bay. Urban growth has also degraded the Ballona wetlands, which are included in this study area. Marina del Rey is one of the major sources of contamination into the Bay and Ballona wetlands. The size of the study area is approximately 25 km^2 and contains different urban land use features. The area is predominantly composed of residential and open land uses.

Another area was examined to verify the performance of Bayesian network classification. The second area includes six drainage basins adjacent to the Sweetwater reservoir as shown in Fig. 2. The Sweetwater River watershed, located in Southern California near San Diego, is a rapidly urbanizing watershed facing many challenges to water quality and the natural habitat. In order to protect the Sweetwater reservoir, a part of the Sweetwater River watershed, the Sweetwater Authority constructed an urban runoff diversion system immediately adjacent to the reservoir to divert urban runoff. The study area covers 37 km^2 and mainly consists of residential and open land uses.

3.2. Data sets

We used the Landsat Enhanced Thematic Mapper Plus (ETM$^+$) for both study areas. The Marina del Rey image was obtained on August 11, 2002 (path 41 row 36) and the Sweetwater reservoir image was obtained on December 18, 1999 (path 40 row 37). Subimages were extracted for both sites, including all bands except the panchromatic band. The thermal band, band 6, was resampled to match the resolution of the other bands based on the nearest neighbor method using RSI ENVI 4.0.

Three different levels of land use types were identified based on the USGS classification system (Anderson et al., 1976): (1) level I: urban, open, and water; (2) level II: residential, commercial, industrial, transportation, open, and water; and (3) level III: single family residential (SFR), multiple family residential (MFR), commercial, public, light industrial, transportation, open, and water. We also employed a classification system based on the imperviousness that is associated with land uses (Stenstrom and Strecker, 1993). The system is based upon a ground survey of different types of land uses and impervious areas, and was performed by the Los Angeles County Department of Public Works (LADPW) to document flood control benefits. Impervious area is important because it controls the amount of stormwater runoff and therefore the tendency of the land use to pollute receiving waters. Open

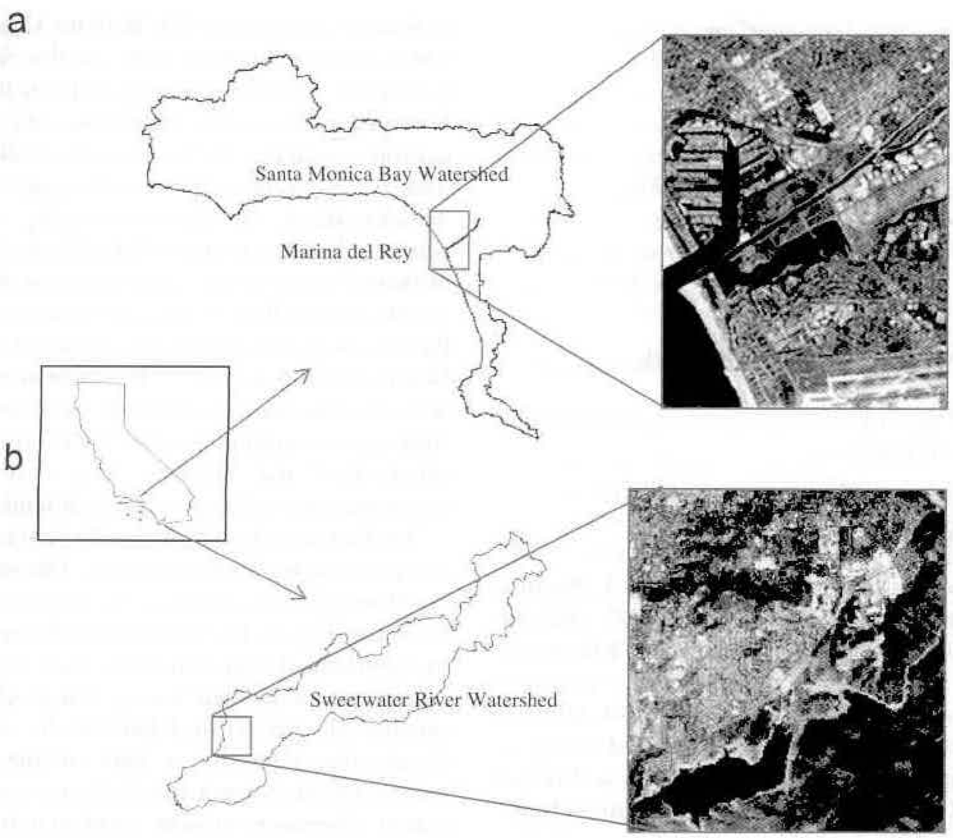


Fig. 2. Location of the study areas: (a) Marina del Rey and Santa Monica Bay watershed; (b) drainage basins into Sweetwater reservoir.

land use is pervious (0%) and produces little runoff. SFR land use is impervious (42%) while MFR land use has even greater roof and paved areas with imperviousness corresponding to 68%. Commercial, industrial, and transportation land uses have the greatest imperviousness (>80%).

Reference data for land use types were obtained from public records and aerial photography. Official land use data were obtained from the Southern California Association of Governments (SCAG, 2003) and San Diego's Regional Planning Agency (SANDAG, 2000). The SCAG land use data were updated from the existing data using digital orthophotography with 1 meter resolution. The SANDAG land use data were updated from the 1990 land use data using satellite images, digital orthophotography, the SanGIS landbase, the County Assessor Master Property Records file, and other ancillary information. However, the minimum land use polygon size from public records was larger than the Landsat pixels. For example, the minimum unit of SCAG data is $6070\,m^2$, which may contain mixed features. All samples in this research were checked against the aerial photography with 1 m resolution to assign sampling pixels into their appropriate categories.

The total number of pixels used for training and testing was 2110 for the Marina del Rey case, which corresponds to 7% of the total data, and 1120 for the Sweetwater Reservoir case, corresponding to 3% of the total data. The test data were excluded from the training data to avoid biased results. Table 2 shows statistical characteristics of the samples.

The coordinate values of each pixel (i.e., longitude and latitude), referred to as spatial ancillary data, were calculated from Landsat ETM$^+$ images using RSI ENVI 4.0 based on WGS 1984 UTM (zone 11N). These data were used to evaluate the benefit of incorporating spatial information of each pixel to classification accuracy.

4. Methods

4.1. Bayesian network structure

We used two structures of Bayesian networks: naïve Bayesian classifiers and Maximum Weight Spanning Trees (MWSTs). Naïve Bayesian classifiers are the simplest Bayesian networks and have only one class node. Every child node contributes to the value of the class node. Naïve Bayesian classifiers are easy to construct, although the relative contribution of each node is hard to identify.

MWSTs are alternative structures that are constructed from data using the optimal dependence tree developed by Chow and Liu (1968). Construction of the network does not require any information other than the mutual information between pairs of variables, which is calculated as follows:

$$\mathrm{MI}(X_i, X_j) = \sum P(x_i, x_j) \log\left(\frac{P(x_i, x_j)}{P(x_i)P(x_j)}\right), \tag{10}$$

where $\mathrm{MI}(X_i, X_j)$ is mutual information of random variables of X_i and X_j. $P(x_i, x_j)$ is a joint probability of x_i and x_j, and $P(x_i)$ and $P(x_j)$ are the probabilities of each random variable. The network was constructed by adding arcs between variable nodes with the greatest mutual information. Cycles or loops of arcs were avoided. The tree structure of MWSTs shows the relationship among variables.

We constructed both network structures by selecting land use as a class node. In both cases, two networks were constructed using spectral data only and using spectral data and spatial data. Prior to classification, the separability of the spectral signatures of the various land uses was evaluated using the Jeffries-Matusita distance (Richards and Jia, 1999), which produces separability indices ranging from 0 to 2 with higher numbers suggesting greater separability.

4.2. Sample size

We investigated the effect of different sample size on classification performance. The minimum training data size can be estimated from the statistical techniques. For example, Fitzpatrick-Lins (1981) suggested the following binomial formula:

$$N = \frac{4p(100 - p)}{\varepsilon^2}, \tag{11}$$

where N is a training data size, p the expected accuracy (%), and ε the allowable error. For instance, with 85% of expected accuracy and 2% of allowable error, the required training data size becomes 1275. Mather (1999) suggested an alternative method to estimate the minimum training data set size, as follows:

$$N = 30nc, \tag{12}$$

where n is the number of spectral bands and c the number of classes. This equation agrees with others (Swain, 1978; Van Genderen et al., 1978; Foody et al., 1995) who found that training data size for each class should be at least $30n$ to form representative training samples. In this case, with 7 spectral bands and 8 classes, at least 1680 training observations are needed. There are other methods of selecting data from each class (Jensen, 1996). An equal or proportional number of data can be selected from each class. The rule of thumb that Congalton (1991) proposed is to collect a minimum of 50 observations for each class.

In order to test the effect of the training data size, we used eight subsets of sample sizes: 500, 750, 1000, 1250, 1500, 1750, 2000, and 2110. The proportion of each class is consistent for all different training data sizes as shown in Table 1. For this analysis, only a naïve Bayesian classifier was used with 256 states (8 bits) for both spectral data and spatial data, using USGS classification system level III (8 land uses) (see also Table 2).

Table 1
Training data for USGS classification systems

Level	USGS land use class	Marina del Rey		Sweetwater reservoir	
		Samples	Ratio (%)	Samples	Ratio (%)
I	Urban	1310	62	525	51
	Open: wetland, barren land	550	26	300	29
	Water	250	12	200	20
	Total	2110	100	1025	100.0
II	Residential	780	37	300	30
	Commercial	230	11	120	11
	Industrial	120	6	55	5
	Transportation	180	9	50	5
	Open: wetland beach	550	26	300	29
	Water	250	12	200	20
	Total	2110	100	1025	100.0
III	SF residential	550	26	200	20
	MF residential	230	11	100	10
	Commercial	100	5	55	5
	Public	130	6	65	6
	Industrial	120	6	55	5
	Transportation	180	9	50	5
	Open: wetland, beach, parks	550	26	300	29
	Water	250	12	200	20
	Total	2110	100	1025	100

4.3. Quantization of variable states

Landsat ETM$^+$ imagery provides 8-bit information per pixel, which is by nature suitable for Bayesian network classification because Bayesian networks often require variables with discrete values. To evaluate the robustness to reduced information, a sensitivity analysis was performed using 3–8 bits. Equal width intervals were used for quantizing radiometric resolution. The equal width interval method breaks the range of the observed data values with k equally sized intervals (Liu et al., 2002; Yang and Webb, 2003). The number of intervals, k, can be selected subjectively. Each interval width is defined as follows:

$$\text{interval width} = \frac{x_{\max} - x_{\min}}{k}, \tag{13}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and the maximum value of a variable x, respectively. This method was applied to all bands and the spatial ancillary data simultaneously. Therefore, spatial ancillary data had the same states as the spectral data when they were quantized. This method has been widely used in many machine learning algorithms because it is simple and effective (Hsu et al., 2000). Again, we used a naïve Bayesian classifier with sample size of 2110 pixels. The class node had eight land uses.

Table 2
Landsat ETM+ image data statistics from samples

Band	Wave length (μm)	Minimum	Maximum	Range	Median	Mean	Standard deviation
(a) Marina del Rey							
1 (Blue)	0.45–0.52	80	228	148	107	113	23
2 (Green)	0.52–0.60	51	216	165	90	94	25
3 (Red)	0.63–0.69	34	253	219	98	100	33
4 (Near IR)	0.76–0.90	13	155	142	64	60	19
5 (IR)	1.55–1.75	7	255	248	96	95	38
6 (Thermal IR)	10.4–12.5	133	210	77	180	175	15
7 (IR)	2.08–2.35	3	255	252	73	74	32
(b) Sweetwater Reservoir Drainage							
1 (Blue)	0.45–0.52	46	127	81	60	64	14
2 (Green)	0.52–0.60	29	111	82	45	48	15
3 (Red)	0.63–0.69	21	127	106	50	51	20
4 (Near IR)	0.76–0.90	12	107	95	48	45	20
5 (IR)	1.55–1.75	9	124	115	63	58	29
6 (Thermal IR)	10.4–12.5	127	191	64	156	153	15
7 (IR)	2.08–2.35	8	123	115	51	46	23

4.4. Accuracy assessment

Overall accuracy and kappa (κ) coefficient (Congalton, 1991; Foody, 2002) were used to assess the Bayesian network performance as follows:

$$\text{Overall accuracy} = \frac{\sum_k n_{kk}}{N}, \tag{14}$$

$$\kappa = \frac{N\sum_k n_{kk} - \sum_k(\sum_j n_{kj}\sum_i n_{ik})}{N^2 - \sum_k(\sum_j n_{kj}\sum_i n_{ik})}, \tag{15}$$

where n_{ij} is the number of pixels with row i and column j and N is the total number of test pixels in the confusion matrix.

The accuracy was assessed using 10-fold cross-validation that is widely used in the machine learning community. In 10-fold cross-validation, the data were randomly divided into 10 subsets of equal size. The networks were trained 10 times with 9 subsets as training data and 1 subset as test data for each time. The average of overall accuracies and κ coefficients from each test were used to assess the accuracies.

Producer's and user's accuracy (Congalton, 1991) were used to assess per class accuracy using a confusion matrix:

$$\text{producer's accuracy} = \frac{n_{ij}}{\sum_i n_{ij}} 100, \tag{16}$$

$$\text{user's accuracy} = \frac{n_{ij}}{\sum_j n_{ij}} 100, \tag{17}$$

where n_{ij} is the number of pixels of each element in the matrix and i and j are the indices for row and column, respectively.

5. Results and discussion

5.1. Training data characteristics

Table 3 shows the Jeffries–Matusita distances between the pairs of land uses. Water and open land uses had excellent separability with all other land uses. SFR was the next most separable. Other land uses had lower separability. In Marina del Rey, the pairs of MFR, commercial, public, and industrial land uses especially had low separability, less than 1. The Sweetwater reservoir case had similar results and the pairs of residential, commercial, public and industrial land uses had low separability. Low separability less than 1 indicates the pair of land uses should be lumped because of the difficulty in separability. The difficulty may be due to similar spectral signatures of roofing materials (Stefanov et al., 2001; Herold et al., 2003).

5.2. Sample size

Fig. 3 shows the classification accuracies as a function of sample size. Accuracies increased monotonically with the sample size—up to 1500 pixels. After 1500 pixels, the accuracies reached a plateau and the increase in accuracy between 1500 and 2110 pixels was trivial (<1%). When incorporating spatial information, the overall accuracies and κ coefficients plateau at approximately 79% and 74%, respectively. Without incorporating spatial information, the classification accuracies decreased by 5%.

The results show that the Bayesian networks performed classification well with only a small sample size. A sample size of 1500 pixels, or 5% of the total data set, was sufficient for classification with acceptable accuracy. The required sample size is comparable to Mather's (1999) suggestion (Eq. (12)), which predicts a minimum training

Table 3
Spectral separability for level III classification using Jeffries–Matusita distance

	SFR	MFR	Commercial	Public	Industrial	Transportation	Open	Water
Marina del Rey								
SFR								
MFR	1.59							
Commercial	1.98	1.13						
Public	1.73	0.72	0.89					
Industrial	1.99	1.51	0.84	0.99				
Transportation	1.99	1.61	1.36	1.56	1.35			
Open	1.87	1.94	1.99	1.94	2.00	1.98		
Water	2.00	2.00	2.00	2.00	2.00	2.00	2.00	
Sweetwater								
SFR								
MFR	1.74							
Commercial	1.43	1.62						
Public	1.29	1.68	1.58					
Industrial	1.64	1.28	1.14	1.50				
Transportation	1.78	1.86	1.61	1.51	1.66			
Open	1.99	2.00	1.98	1.99	1.98	1.99		
Water	2.00	2.00	2.00	2.00	2.00	2.00	2.00	

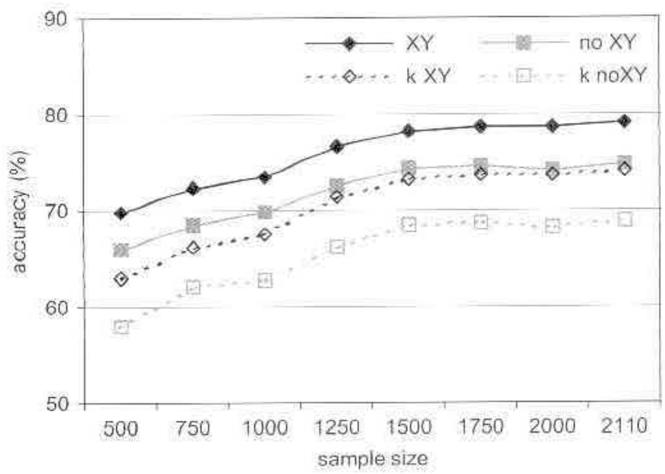


Fig. 3. Performance of Bayesian networks with different sample size. Note that solid lines with closed markers represent overall accuracy and those with open markers represent kappa coefficient.

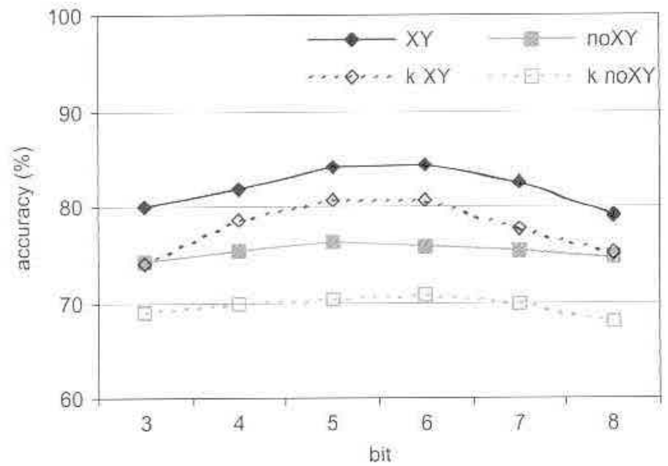


Fig. 4. Performance of Bayesian networks with quantized radiometric resolution. Note that solid lines with closed markers represent overall accuracy and those with open markers represent kappa coefficient.

data size of 1680 pixels (e.g., $30 \times 7 \times 8$). Our results confirm this minimum sample size for Bayesian network classification, although there is no theoretical research on the required size for Bayesian network training data sets. The required training data size may be dependent on the quantization level or the number of variable states. Landsat images contain 8-bit information (256 states) which may impact the required training data size.

5.3. *Quantization of radiometric resolution*

Fig. 4 illustrates the effect of quantization on classification accuracy. Overall accuracies and κ coefficients increased as the number of states increased, achieving a peak when using only 5 or 6 bits of the available 8 bits. At this point, the overall accuracy was 84% and κ coefficient was 81%. After 6 bits, the accuracies decreased as the number of variable states increased. Therefore, classification with 8-bit information was less accurate.

The results show that classification using Bayesian networks of Landsat imagery does not require 8-bit information. Our results show that classification with 3–6-bit information was just as accurate as using 8-bit information, where 5 or 6 bits were optimal. Generally, a large number of quantization levels can represent a continuous variable more accurately, but increases complexity and may introduce inaccuracies. Inaccuracies may be introduced because there may be insufficient training data to represent all pixels for all states. For example, 256 pixel values in the Landsat data with 8 land use states produces 2048 separate cases for the Bayesian network. For the examples in this paper, only 1899 training pixels out of 2110 samples are used, and some states will have no observations. Conversely, a small number of quantization levels may reduce the chance of states with no observations, but may lose information from the original continuous distribution. Malka and Lerner (2004) have shown that the optimum number of quantization levels can depend on the sample size.

The results show that using both extremes of quantization levels was less accurate. Using 5 or 6 bits resulted in more accurate representation of probabilistic tables for Bayesian learning. An advantage of using fewer bits is reduced complexity and computing resources, because the problem dimensionality is proportional to the number of quantization levels to the power of the number of the nodes (Malka and Lerner, 2004). This indicates that reducing the number of quantization levels may provide faster and more accurate learning (Dougherty et al., 1995; Liu et al., 2002; Yang and Webb, 2003). Reduced computing resources may be advantageous for very large problems.

5.4. Different classification systems

Fig. 5 shows the results of classification based on different classification systems. In the Marina del Rey case, naïve Bayesian classifiers provided overall accuracy of 93% and κ coefficient of 86% for level I classification. For level II classification, overall accuracies and κ coefficients were 82% and 76%, respectively. For level III classification, overall accuracies were slightly reduced to 79%. When spatial information was omitted, level II and III classification accuracy decreased by approximately 5%.

Imperviousness classifications were slightly more accurate than level II land use classification even though their class nodes had a similar number of states. In the Marina del Rey case, the overall accuracy and κ coefficient for imperviousness classification were 5% and 7% greater than the corresponding accuracy in level II land use classification. For the Sweetwater reservoir case, overall accuracies and κ coefficients for imperviousness were 10% and 15% greater, respectively, than the corresponding accuracy in

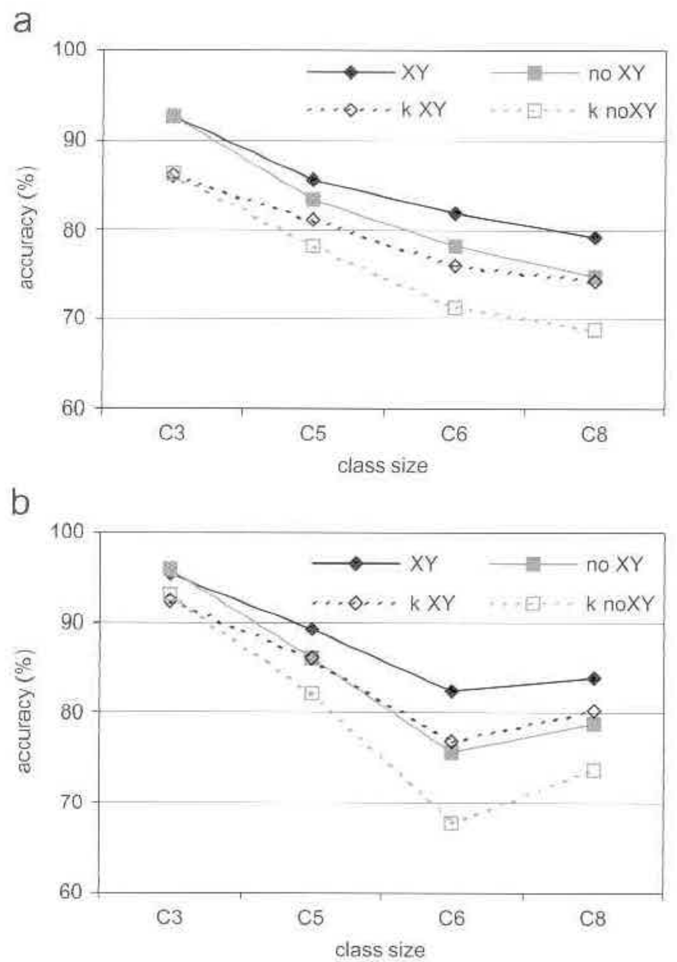


Fig. 5. Performance of Bayesian network classification for different classification systems: (a) Marina del Rey; (b) Sweetwater reservoir. Note that C3 corresponds to USGS land use/land cover level I classification system, C6 to level II system, C8 to level III system, and C5 to imperviousness classification associated with land use.

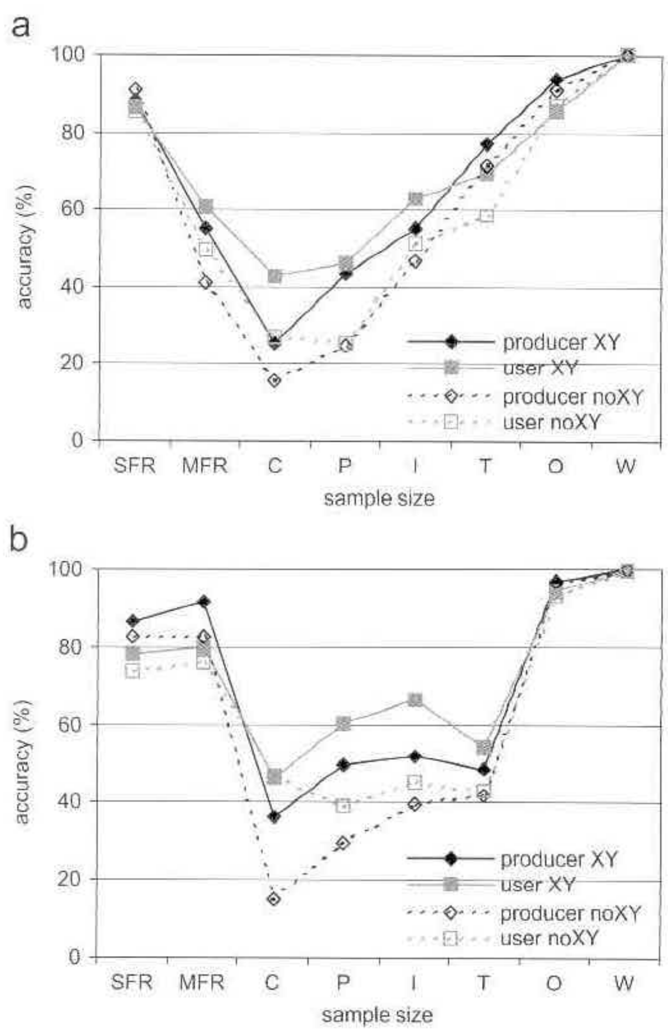


Fig. 6. Per-class accuracy assessment: (a) Marina del Rey; (b) Sweetwater reservoir.

level II land use classification. This is to be expected because the impervious surfaces have less varied spectral signatures.

Fig. 6 shows the producer's and user's accuracies for each land use based on confusion matrices with the level III classification system. In the Marina del Rey case, SFR, open and water land uses were successfully assigned, with accuracies greater than 85%, even without incorporating spatial information. By contrast, the producer's and user's classification accuracies of commercial and public land uses were approximately 50%. In the Sweetwater reservoir case, the accuracies of SFR, MFR, open, and water land uses were all above 85%. The producer's and user's accuracies of commercial land use were also approximately 50%. Omitting spatial information reduced the accuracies of commercial and public land uses by 9–21%. The separability results support this finding in that some urban land uses such as commercial, public, and industrial land uses are difficult to separate due to similar spectral signatures of roofing materials (Stefanov et al., 2001; Herold et al., 2003). However, misclassification among commercial, public and industrial land uses is not a fatal problem for stormwater management because these land uses usually generate similar stormwater pollutant loads and therefore can be combined together into a single class (Park and Stenstrom, 2006). When these three classes are combined into a single class, the overall classification accuracy improved to 86% and omission error of the combined land uses was reduced to 16% (Park, 2004).

In general, the Landsat ETM$^+$ imagery is considered to be suitable for level I or level II classification. In the Marina del Rey study, level III accuracy was only 2–3% lower than level II accuracy, and in the Sweetwater reservoir case, level III classification outperformed level II classification by 1 to 3%. This may be due in part to the differences in land use distribution (the Sweetwater example had less commercial, industrial, and public land use—see Fig. 9). Confusion exists among these land uses because of the inherent similarity of their spectral signatures. An advantage of the Bayesian network approach is its ability to learn zoning effects from spatial information. Therefore, spatial data will be helpful for a higher level of classification using Landsat ETM$^+$ imagery for areas with zoning rules or other rules regulating land use.

5.5. *Bayesian network structures*

Fig. 7 illustrates the two Bayesian network structures. Using naïve Bayesian classifiers, all input variables (bands 1–7) contribute to the value of the class node (land use or imperviousness class). By contrast, the structure of MWSTs shows that bands 2, 5, and 6 mainly contribute to the value of the class node. The importance of bands 2 and 5 is supported by other research. For example, Herold et al. (2003) measured spectral signature and also found

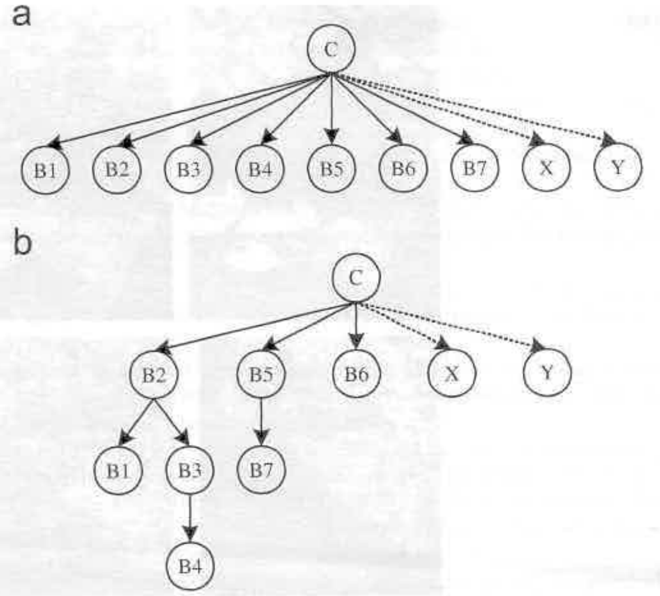


Fig. 7. Bayesian network structures for urban land use classification: (a) naïve Bayesian classifiers; (b) maximum weight spanning trees. The structures with solid lines represent the network with spectral data only, and the structures with dotted lines represent the network including spatial ancillary data.

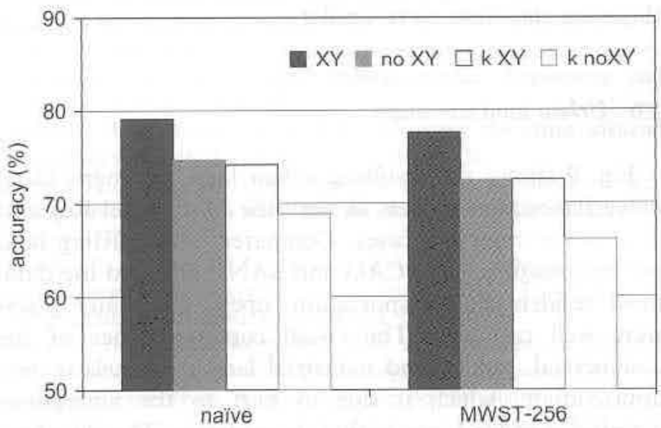


Fig. 8. Classification accuracies with different network structures.

that bands 2 and 5 provided suitable spectral information for separating urban features.

Our results showed high dependency between visible bands (1, 2, and 3) and middle infrared bands (5 and 7), which corresponds to similar findings of strong correlation between these bands (Jensen, 1996). This result implies that information among these bands may be redundant and implies that the network size can be reduced without losing information. Many researchers have not used band 6 for land cover classification, but in our case, including band 6 improved accuracy. Previous studies have reported that thermal information is related to urban surface characteristics (Voogt and Oke, 2003). These results illustrate how the structure of MWSTs can be helpful in understanding the relationships among variables for classification. Fig. 8

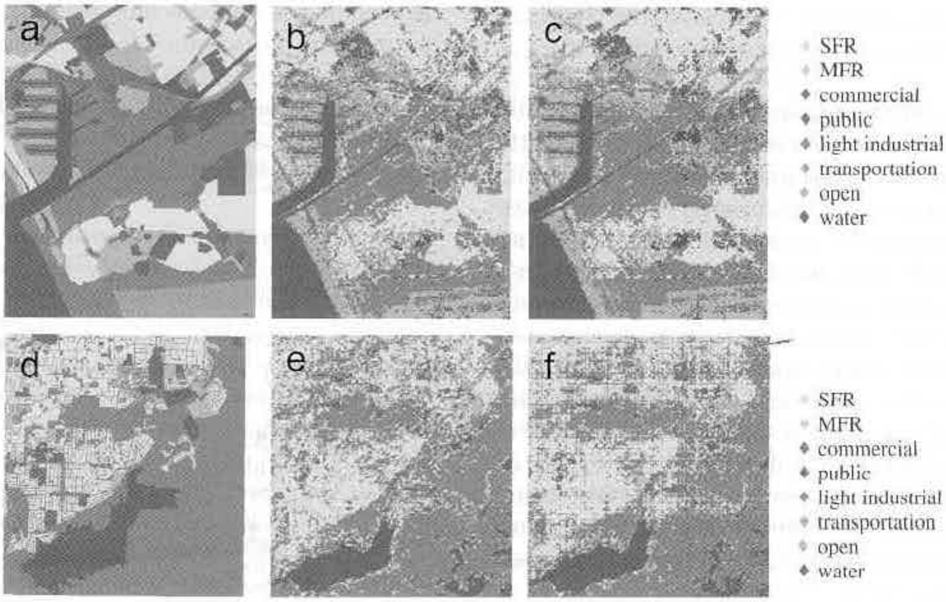


Fig. 9. Urban land use maps: (a) SCAG land use data; (b) classification with spectral data only in Marina del Rey; (c) classification with spectral and spatial data in Marina del Rey; (d) SANDAG land use data; (e) classification with spectral data only in Sweetwater reservoir; (f) classification with spectral and spatial data in Sweetwater reservoir.

shows that the performances of MWSTs and naïve Bayesian classifiers were similar.

5.6. Urban land use maps

Fig. 9 shows the resulting urban land use maps using naïve Bayesian classifiers in both the Marina del Rey and Sweetwater reservoir cases. Compared with existing land use information, i.e., SCAG and SANDAG land use data, most residential, transportation, open, and water pixels were well captured. The visual correspondence of the commercial, public, and industrial land use pixels is only approximate, which is due in part to the ambiguous definitions of land uses within these classes. The classification errors in the pixels belonging to the beach in Marina del Rey were assigned to transportation because beach pixels were undefined in the training data. In the Sweetwater reservoir case, the upper part of the reservoir shows an urban spectral signature because this part of the reservoir was dry at the time the image was taken. The pixels of the shadows in the open area were assigned to water land use because they had not been included in the training data.

Incorporating spatial information improved the visualization of all land uses even when the digital numbers of the spectral signatures were similar. The land use maps in Fig. 9 had less "salt and pepper" effect as the spatial information was added. This is a logical result because the same types of land uses tend to be located together. This occurs because of zoning laws and human nature, i.e., industrial sites are usually not located in SFR areas and vice versa.

The quantized spatial data provided to the Bayesian networks gives them the ability to make classification decisions based upon the locational or neighborhood information associated with each pixel. The number of contiguous pixels assigned to a particular land use having similar properties can be used by a Bayesian network to improve inferences. The roof on a large industrial complex or factory may be composed of the same material as an apartment building (e.g., asphalt), but a Bayesian network can make inferences based upon the number of contiguous pixels associated with the same land use. Although the examples provided in this paper have no way of making direct comparisons of contiguous pixels, the spatial information gives the network the ability to make these inferences. This implies that incorporating other ancillary data, for example, a digital elevation model, might also improve the accuracy of classification. Others have reported the benefits of incorporating ancillary data, such as population and road density information (Greenberg and Bradley, 1997), and contexture information (Stuckens et al., 2000) for land cover classification. We believe that incorporating ancillary data can provide more information to the networks in learning the relationships among variables and in generating more accurate land use maps. Bayesian networks have an important advantage in that they can show which variables contribute more to classification accuracy.

6. Conclusions

This study has shown that Bayesian networks can be used to accurately classify urban land use from Landsat ETM^+ imagery. Both naïve Bayesian classifiers and

MWSTs were evaluated in the Marina del Rey and Sweetwater reservoir drainage basins. The following specific conclusions are drawn:

(1) Bayesian networks were successful in urban land use classification from Landsat ETM$^+$ imagery. Classifying imperviousness, that was related to land use, showed better accuracy due to the low separability of urban features.
(2) Using spatial information, such as X and Y coordinate values of each pixel, improved classification accuracy. The greatest improvement was for commercial, public, and industrial land uses, which had lower per class accuracy. This shows that locational information can enhance highly impervious urban land use classification.
(3) Bayesian networks required only a small sample size for training, equal to approximately 5% of the total number of pixels.
(4) Quantization of variable states (i.e., radiometric resolution) was useful, showing that using all 8 bits were no more accurate than using only 3–6 bits. Using 5 or 6 bits provided the highest accuracy and reduced the required computing resources, which may be important for large problems.
(5) MWSTs have the advantage of showing which variables are most useful for classification, and overcome the "black box" criticism of neural networks, which generally do not provide indication of the cause and effect relationships.

This study can provide a guideline for urban land use classification using Bayesian networks. Bayesian networks handled the uncertainty associated with mixed signatures from urban land use. The success of this study should encourage others to use Bayesian networks for urban land use classification and its application to urban planning and environmental management. Well-identified land use information will improve estimates of environmental quality by identifying significant hot spots.

Acknowledgements

This work was supported in part by the US EPA grant R825831 and by the Sweetwater Authority. The authors thank Prof. Keith Clarke (Department of Geography, University of California, Santa Barbara), Prof. Thomas Gillespie (Department of Geography, University of California, Los Angeles), and Prof. Duncan F. Gillies (Department of Computing, Imperial College London) for their reviewing this manuscript.

References

Anderson, J.R., Hardy, E., Roach, J., Witmer, R., 1976. A land-use and land-cover classification system for use with remote sensor data. US Geological Survey Profession Paper 964.

Boutell, M., Luo, J.B., 2005. Beyond pixels: exploiting camera metadata for photo classification. Pattern Recognition 38 (6), 935–946.

Brown, D.G., Pijanowski, B.C., Duh, J.D., 2000. Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. Journal of Environmental Management 59, 247–263.

Charniak, E., 1991. Bayesian network without tears. AI Magazine 12 (4), 50–63.

Chiew, F.H.S., McMahon, T.A., 1999. Modelling runoff and diffuse pollution loads in urban areas. Water Science and Technology 39, 241–248.

Chow, C.K., Liu, C.N., 1968. Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information theory 14 (3), 462–467.

Clapham, W.B., 2003. Continuum-based classification of remotely sensed imagery to describe urban sprawl on a watershed scale. Remote Sensing of Environment 86, 322–340.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37, 35–46.

Cowell, R., 1998. Introduction to inference in Bayesian networks. In: Jordan, A. (Ed.), Learning in Graphical Models. The MIT Press, Cambridge, MA, pp. 9–26.

Dougherty, J., Kohavi, R., Sahami, M., 1995 Supervised and unsupervised discretization of continuous features. In: Proceedings of the 12th International Conference on Machine Learning, Los Altos, pp. 194–202.

Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map. Photogrammetric Engineering and Remote Sensing 47 (3), 343–351.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80 (1), 185–201.

Foody, G.M., McCulloch, M.B., Yates, W.B., 1995. Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. Photogrammetric Engineering and Remote Sensing 61, 391–401.

Greenberg, J.D., Bradley, G.A., 1997. Analyzing the urban–wildland interface with GIS. Journal of Forestry 95, 18–22.

Gurwicz, Y., Lerner, B., 2005. Bayesian network classification using spline-approximated kernel. Pattern Recognition Letters 26 (11), 1761–1771.

Herold, M., Gardner, M.E., Roberts, D.A., 2003. Spectral resolution requirements for mapping urban areas. IEEE Transaction of Geoscience and Remote Sensing 41 (9), 1907–1919.

Hsu, C.N., Huang, H.J., Wong, T.T., 2000. Why discretization works for naive Bayesian classifiers. In: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 309–406.

Jensen, J.R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. Prentice-Hall, Upper Saddle River, NJ, pp. 248–249.

Liu, H., Hussain, F., Tan, C.L., Dash, M., 2002. Discretization: an enabling technique. Data Mining and Knowledge Discovery 6, 393–423.

Los Angeles County Department of Public Works (LADPW), Alhambra, CA.

Malka, R., Lerner, B., 2004. Classification of fluorescence in situ hybridization images using belief networks. Pattern Recognition Letters 25, 1777–1785.

Mather, P.M., 1999. Computer Processing of Remotely-Sensed Images: An Introduction. Wiley, Chichester, UK.

NCERQA, 2002. Integrated urban watershed analysis: the Los Angeles basin and coastal environment. Institute of the Environment Report #02-02.

Neapolitan, R.E., 1990. Probabilistic Reasoning in Expert Systems: Theory and Algorithms. Wiley, New York, pp. 120–123, 153–190.

Orun, A.B., 2004. Automated identification of man-made textural features on satellite imagery by Bayesian networks. Photogrammetric Engineering and Remote Sensing 70 (2), 211–216.

Pal, C., Swayne, D., Frey, B., 2001. The automated extraction of environmentally relevant features from digital imagery using Bayesian multi-resolution analysis. Advances in Environmental Research 5 (4), 435–444.

Park, M.-H., 2004. Bayesian network application to satellite image classification for stormwater management. Ph.D. Dissertation, University of California, Los Angeles/

Park, M-H., Stenstrom, M.K., 2006. Spatial estimates of stormwater-pollutant loading using Bayesian networks and geographic information systems. Water Environment Research 78 (4), 421–429.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA, pp. 77–133.

Richards, J.A., Jia, X., 1999. Remote Sensing Digital Image Analysis: An Introduction, third ed. Springer, New York, pp. 240–247.

Russell, S., Norvig, P., 1995. Artificial Intelligence: A Modern Approach. Prentice Hall, Upper saddle river, NJ, pp. 436–470, 588–593.

SANDAG land cover and activity center, 2000. 〈http://www.sandag.cog.ca.us/〉.

Southern California Association of Governments, 2003. 〈http://wagsdata.scag.ca.gov〉.

Stefanov, W.L., Ramsey, M.S., Christensen, P.R., 2001. Monitoring urban land cover change; an expert system approach to land cover classification of semiarid to arid urban centers. Remote Sensing of Environment 77 (2), 173–185.

Stenstrom, M.K., Silverman, G.S., Bursztynsky, T.A., 1984. Oil and grease in urban stormwaters. Journal of the Environmental Engineering ASCE 110, 58–72.

Stenstrom, M.K., Strecker, E., 1993. Assessment of Storm Drain Sources of Contaminants to Santa Monica Bay, Vol. I, Annual Pollutants Loadings to Santa Monica Bay from Stormwater Runoff. UCLA-ENG-93-62, Los Angeles, USA.

Stuckens, J., Coppin, P.R., Bauer, M.E., 2000. Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing of Environment 71, 282–296.

Swain, P.H., 1978. Fundamentals of pattern recognition. In: Swain, P.H., Davis, S.M. (Eds.), Remote Sensing: The Quantitative Approach. McGraw-Hill, New York, pp. 136–187.

Sweetwater Authority 〈http://www.sweetwater.org/〉.

Van Genderen, J.L., Lock, B.F., Vass, P.A., 1978. Remote sensing: statistical testing of thematic map accuracy. Remote Sensing of Environment 7 (1), 3–14.

Voogt, J.A., Oke, T.R., 2003. Thermal remote sensing of urban climates. Remote Sensing of Environment 86, 370–384.

Yang, Y., Webb, G.I., 2003. Discretization for naïve-Bayes learning: managing discretization bias and variance. Technical Report 2003/131, School of Computer Science and Software Engineering, Monash University.

ELSEVIER

Journal of Environmental Management 86 (2008) 193–200

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Anaerobic digestion of dairy manure with enhanced ammonia removal

S. Uludag-Demirer[a,*,1], G.N. Demirer[b,1], C. Frear[c], S. Chen[c]

[a]*Department of Industrial Engineering, Çankaya University, 06530 Ankara, Turkey*
[b]*Department of Environmental Engineering, Middle East Technical University, 06531 Ankara, Turkey*
[c]*Department of Biological Systems Engineering, Washington State University, Pullman, WA 99163, USA*

Received 13 February 2006; received in revised form 24 November 2006; accepted 5 December 2006
Available online 25 January 2007

Abstract

Poor ammonia-nitrogen removal in methanogenic anaerobic reactors digesting animal manure has been reported as an important disadvantage of anaerobic digestion (AD) in several studies. Development of anaerobic processes that are capable of producing reduced ammonia-nitrogen levels in their effluent is one of the areas where further research must be pursued if AD technology is to be made more effective and economically advantageous. One approach to removing ammonia from anaerobically digested effluents is the forced precipitation of magnesium ammonium phosphate hexahydrate ($MgNH_4PO_4 \cdot 6H_2O$), commonly called struvite. Struvite is a valuable plant nutrient source for nitrogen and phosphorus since it releases them slowly and has non-burning features because of its low solubility in water. This study investigated coupling AD and controlled struvite precipitation in the same reactor to minimize the nitrogen removal costs and possibly increase the performance of the AD by reducing the ammonia concentration which has an adverse effect on anaerobic bacteria. The results indicated that up to 19% extra COD and almost 11% extra NH_3 removals were achieved relative to a control by adding 1750 mg/L of $MgCl_2 \cdot 6H_2O$ to the anaerobic reactor.

Keywords: Struvite; Anaerobic digestion; Ammonia removal

1. Introduction

Even though anaerobic digestion (AD) is an established farm animal manure bioconversion technology (Hobson and Wheatley, 1993; Lusk, 1998; Demirer et al., 2000; US Environmental Protection Agency, 2002; Güngör-Demirci and Demirer, 2004), there are several key areas of research that must be pursued if AD technology is to be made more effective and economically advantageous for animal manure and therefore allow for wider adoption of the technology. Development of anaerobic processes that are capable of producing reduced ammonia-nitrogen levels in their effluent is one of these areas.

Poor ammonia-nitrogen removal (as well as other components) in methanogenic anaerobic reactors digesting animal manure has been reported as an important disadvantage of AD in several studies (Lusk, 1998; Cheng and Liu, 2002; Martin et al., 2003; Noike et al., 2004). High effluent concentrations of ammonia in anaerobic reactors which can be attributed to the anaerobic bioconversion of proteins contained in animal manure into amino acids and then to ammonia make a posttreatment necessary to remove ammonia before discharge into receiving water bodies. Biological nitrogen treatment processes (nitrification–denitrification) are the most widely used processes for this purpose (Henze and Harremoes, 1978). Nitrification is generally carried out by aerobic, autotrophic bacteria that oxidize NH_4^+ to NO_2^- and NO_2^- to NO_3^- with molecular oxygen as an electron acceptor. These bacterial conversions require a very efficient oxygen supply. NO_2^- and NO_3^- are subsequently reduced to N_2 by denitrifying bacteria that use the NO_x^- as alternative electron acceptors to oxygen, and are most effective in the absence of oxygen. The situation is further complicated, because the autotrophic nitrifiers cannot compete with aerobic heterotrophs for oxygen and other nutrients in the presence of substantial

*Corresponding author. Tel.: +90 312 2844500x4092; fax: +90 312 284 8043.
E-mail address: sibel@cankaya.edu.tr (S. Uludag-Demirer).
[1]At the time of the study, authors were associated with Department of Biological Systems Engineering, Washington State University, Pullman, WA 99163, USA.

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.002

amounts of organic compounds and can therefore easily be overgrown by the heterotrophs. An additional complication is that the denitrifying bacteria must be provided with a suitable electron donor, usually organic compounds (Mulder et al., 1995). Biological nitrogen removal from anaerobically digested animal manures in small-scale operations is not a viable option. This is mainly due to the high cost of oxygen and carbon source supplementation for the nitrification and denitrification steps, respectively, and the requirement of skilled personnel for the operation.

Due to high nutrient content, the liquor produced by the AD process can be used as a liquid fertilizer on the farms on which it was produced. However, the unavailability of land area to which this liquor can be applied, especially in concentrated animal feeding operations, does not enable effective application of the liquid effluent. Therefore, in order to avoid adverse environmental consequences, through overapplication of AD liquor on limited land or leaching of effluent into water supplies, the nutrients must be completely or partially removed. Furthermore, recovery of these nutrients from the anaerobically digested manure is a potential source of revenue, partially offsetting the costs of treatment. This is why control over the point sources of N and P recently shifted from removal to recovery, with a particular emphasis on improving the sustainability of agricultural activities. This was mainly due to the increasing global demand for the nitrogenous fertilizer (from 10 Mt N in 1960 to 90 Mt N in 1998, Mulder, 2003). Therefore, the current attempts are not only to protect the water resources, but also to extract the maximum amounts of N and P from the recoverable sources, such as livestock manure. According to the US Department of Agriculture and Economic Research Service (2001), 1.7 million tons of recoverable N and 0.7 million tons of recoverable P were produced by the livestock and poultry in 1997 in the US.

Under certain conditions ammonia, magnesium, and phosphate react to form a precipitate called struvite ($MgNH_4PO_4 \cdot 6H_2O$). Due to the high concentrations of dissolved orthophosphates and/or ammonia and magnesium ions encountered in the anaerobic treatment of animal manure, there is strong potential for forming the precipitate.

One approach to removing nitrogen and phosphorus from anaerobically digested manure prior to land application is the forced precipitation of magnesium ammonium phosphate hexahydrate ($MgNH_4PO_4 \cdot 6H2O$), commonly called struvite. Struvite is composed of equimolar concentrations of magnesium (Mg^{2+}), ammonium (NH_4^+), and phosphate (PO_4^{3-}). Struvite is an undesirable mineral frequently formed in recycle–flush animal waste management systems. Formation of struvite in flush water recycle pipes has been problematic in liquid manure handling systems because it creates blockages (Booram et al., 1975; Westermann et al., 1985). Therefore, a large portion of struvite research has been directed towards removal and prevention of struvite rather than towards forced precipitation from solution. However, struvite has been found to be a good plant nutrient source for nitrogen and phosphorus since it releases them slowly and has non-burning features because of its low solubility in water (Salutsky et al., 1970; Beal et al., 1999; Miles and Ellis, 2001).

Struvite precipitation, which is common in wastewater systems containing high concentrations of dissolved orthophosphates and free and saline ammonia and magnesium ions, such as in anaerobic fermentation systems, may be the best way to recover nutrients from the slurry after anaerobic treatment. The precipitation of struvite was reported to be an effective method for the recovery of nitrogen and phosphorus in several anaerobically digested wastewaters such as swine and calf wastewater, coke manufacturing, landfill leachate, leather tanning, dairy manure, etc. (Fujimoto et al., 1991; Battistoni et al., 1997, 2000; Li et al., 1999; Schuiling and Andrade, 1999; Munch and Barr, 2001; Nelson et al., 2003; Uludag-Demirer et al., 2005). However, there is only one study reported in which struvite precipitation was carried out during the AD of a food waste in a digester (Lee et al., 2004). The results of this study indicated that struvite precipitation obtained by addition of Mg^{2+} during AD led to 67% and 73% N and P removals, respectively. Therefore, the aim of this study was to couple AD and struvite precipitation in the same reactor to minimize the nitrogen removal costs. To this purpose 24 batch anaerobic digesters containing different doses of $Mg(OH)_2$ and $MgCl_2 \cdot 6H_2O$ as the Mg^{2+} and Na_2HPO_4 as the PO_4^{3-} supply were operated for 56 days. The performance of AD of the dairy manure and the ammonia removal were determined.

2. Materials and methods

2.1. Dairy manure and anaerobic seed cultures

Fresh manure was collected from the Dairy Center at Washington State University in Pullman, WA, and stored at 4 °C prior to use. The composition of the raw dairy manure is presented in Table 1 (Wen et al., 2004).

The mixed anaerobic culture used as seed was obtained from the anaerobic lagoon of the Dairy Center at

Table 1
Composition of raw dairy manure[a] (Wen et al., 2004)

Dry matter (%)	14.60 ± 0.25
Composition (% of dry matter)	
Neutral detergent fiber (NDF)	49.10 ± 1.30
Acid detergent fiber (ADF)	37.83 ± 1.01
Acid detergent lignin (ADL)	11.24 ± 1.02
Hemicellulose (NDF–ADF)	11.27 ± 0.90
Cellulose (ADF–ADL)	26.59 ± 0.28
Lignin (ADL)	11.24 ± 1.02
Total carbon	50.51 ± 1.22
Total nitrogen	3.03 ± 0.58

[a]Data are expressed as mean ± SD of three replicates.

Washington State University in Pullman, WA, and stored at 4 °C prior to use. The mixed anaerobic culture was filtered through a screen of 0.0469 in (1.19 mm) mesh size and concentrated by settling before being used as inoculum. The volatile suspended solids (VSS) concentration of the concentrated seed cultures used was 5460 ± 280 mg/L.

2.2. Experimental setup

Twenty four batch anaerobic reactors (Scientific Instrument Services, Ringoes, NJ) with 160 mL total volume were operated. The effective volume of the reactors was 120 mL. Initially, 50 mL of concentrated anaerobic seed, 30 mL of dairy manure, and varying amounts of $Mg(OH)_2$ or $MgCl_2 \cdot 6H_2O$ as the Mg^{2+} and Na_2HPO_4 as the PO_4^{3-} source were added to the reactors. The initial composition of each reactor as well as initial $Mg(OH)_2$ or $MgCl_2 \cdot 6H_2O$ and Na_2HPO_4 concentrations is given in Table 2. After adding all the necessary components (Table 2), the reactors were flushed with a 25% CO_2 and 75% N_2 gas mixture for 5 min and maintained in a water bath at 35 ± 2 °C. The reactors were mixed manually, once a day after the gas measurement.

The performance of the reactors was monitored by measuring the biogas production and chemical oxygen demand (COD), total solids (TS), total volatile solids (TVS), total Kjeldahl nitrogen (TKN), ammonia nitrogen (NH_3-N), and pH at the end of the experiments.

2.3. Analytical methods

COD, TS, TVS, TKN, NH_3-N, and pH analyses were performed at the WSU Water Quality Lab as described in Standard Methods (APHA, 1995). The gas produced in each reactor was measured using a water displacement device (Demirer et al., 2000).

3. Results and discussion

Fig. 1 depicts the daily gas production (GP) and cumulative gas production (CGP) in each reactor. As is clear from the GP data, addition of struvite precipitating chemicals resulted in a very wide range of effects on the performance of the anaerobic reactors. Some reactors (3, 5, 6, and 9) were almost completely inhibited, while some others (13 and 14) performed similar to the control reactor (Fig. 1). One important observation was about the final pH values of the reactors. The pH values of all the reactors except the ones with almost complete inhibition (Fig. 1, Reactors 3, 5, 6, and 9) were more or less in the acceptable range for AD (6.99–7.69). However, the pH values were

Table 2
The experimental setup

Reactor no.	Volume of stock solution added (mL)			Initial concentration in the reactor (mg/L)		
	$Mg(OH)_2$	$MgCl_2 \cdot 6H_2O$	Na_2HPO_4	$Mg(OH)_2$	$MgCl_2 \cdot 6H_2O$	Na_2HPO_4
B	0	0	0	0	0	0
C1	0	0	0	0	0	0
C2	0	0	0	0	0	0
1	0	0	7.2	0	0	6000
2a	0	0	3.6	0	0	3000
2b	0	0	3.6	0	0	3000
3a	28.8	0	0	4000	0	0
3b	28.8	0	0	4000	0	0
4a	0	28	0	0	7000	0
4b	0	28	0	0	7000	0
5a	28.8	0	7.2	4000	0	6000
5b	28.8	0	7.2	4000	0	6000
6	14.4	0	7.2	2000	0	6000
7a	7.2	0	7.2	1000	0	6000
7b	7.2	0	7.2	1000	0	6000
8	7.2	0	3.6	1000	0	3000
9	7.2	0	0	1000	0	0
10a	0	28	7.2	0	7000	6000
10b	0	28	7.2	0	7000	6000
11	0	14	7.2	0	3500	6000
12a	0	7	7.2	0	1750	6000
12b	0	7	7.2	0	1750	6000
13	0	7	3.6	0	1750	3000
14	0	7	0	0	1750	0

All reactors contained 50 mL seed, all reactors except B (blank) contained 30 mL of manure. The total volume in the reactors was brought to 120 by using water. The concentration of stock solutions were 30, 16.7, and 100 g/L for $MgCl_2 \cdot 6H_2O$, $Mg(OH)_2$ and Na_2HPO_4, respectively. Reactors 2–5, 7, 10, and 12 were run as duplicates.

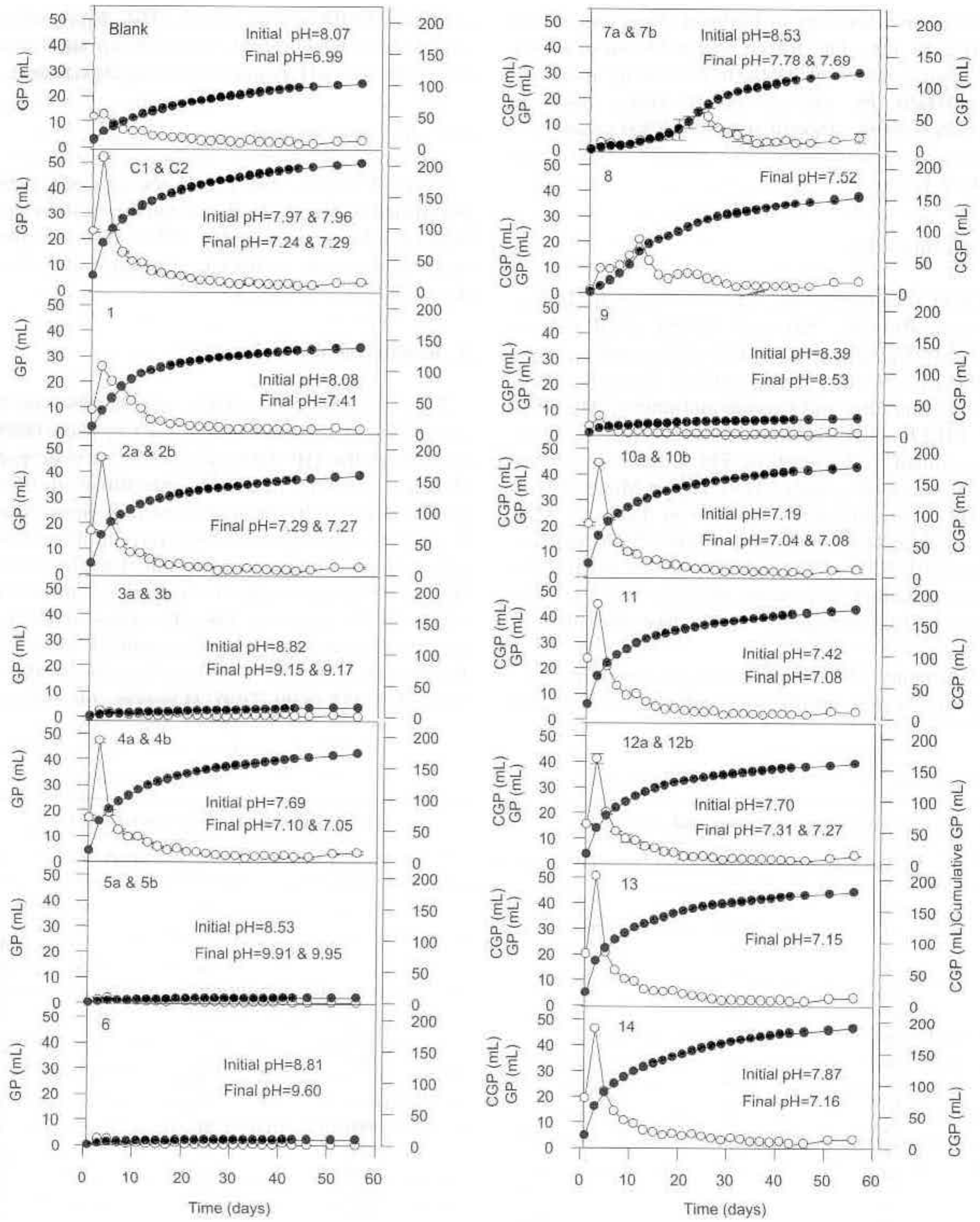


Fig. 1. Gas production in reactors.

extremely high in Reactors 3, 5, 6, and 9 (8.53–9.60). This observation raised the question of the effect of struvite precipitating chemicals on the pH of the anaerobic digesters.

Table 3 depicts the GP, net gas production (NGP, the GP in the reactor minus the GP in the control), and normalized net gas production (NNGP, the NGP in the reactors relative to the control). It is clear from Table 3 that the addition of $Mg(OH)_2$ or $MgCl_2 \cdot 6H_2O$ and Na_2HPO_4 to the anaerobic reactors resulted in significant reductions in the GP. The possible mechanisms resulting in this observation could be the high pH levels inhibiting the methanogens (Speece, 1996) or cation toxicity (Kugelman and Chin, 1971). The IC_{50} (the inhibitory concentration of a chemical resulting in 50% reduction in microbial activity) values for Mg^{2+} and Na^{+} were 1900 and 7400 mg/L, respectively (Kugelman and Chin, 1971). When the maximum Mg^{2+} and Na^{+} concentrations added to the reactors (7000 and 6000 mg/L, respectively) are considered, it is highly probable that the reduction in the GP of the reactors can, at least, be partially explained by cation toxicity. Moreover, Borgerding (1972) reported that the struvite crystals anchor to sludge particles in suspension and this may lead to reduced contact with the substrate and thus reduced GP levels.

Among the reactors tested, Reactors 3 and 9 were treated solely by $Mg(OH)_2$ and it was clear that the basic characteristic of the chemical raised the pH level to the points (>8.50) at which the AD broke down completely. In addition to the pH increase in the reactors, the low solubility of $Mg(OH)_2$ might be another key factor in the break down of the reactors. The $Mg(OH)_2$ slurries may have covered the surfaces of the microorganisms preventing their contact with the organic substrate, hence causing inhibition of GP in the reactors. The addition of $Mg(OH)_2$ accompanied by Na_2HPO_4 to promote the formation of struvite in the reactors did not change the performance of the reactors and the GP in Reactors 5 and 6 was unacceptably low (Table 3 and Fig. 1). This observation indicates that the addition of $Mg(OH)_2$ to form struvite using the already available or added PO_4^{3-} inhibits the AD of the manure as a result of an intolerable increase in the pH.

The reduction in the GP was less than 35% in Reactors 4 and 14, which were spiked only with $MgCl_2 \cdot 6H_2O$. The final pH in these reactors was slightly lower than their initial levels probably due to the acidic property of Cl^- ion in the chemical. The removal of NH_3-N by struvite formation is expected to occur via the use of the available NH_3 and PO_4^{3-} ions in the reactors.

To determine the effect of Na_2HPO_4 addition on the performance of the reactors, the GP in Reactors 1 and 2 can be compared to the GP in the control reactor using the data given in Table 3 and Fig. 1. Although the digestion in these reactors was not complete during the test period, the observed decline in the performance of the reactors is noteworthy. This could be because of dosing the reactors with high concentrations (Table 2) of the chemical resulting in cation (Na^+ ion) toxicity for the AD.

The objective of the study was to combine AD and ammonia-nitrogen removal in the same reactor. To this end, enhanced ammonia-nitrogen removal at the cost of reduced AD performance is not desirable. Therefore, the anaerobic reactors with low GP (lower than 65% relative to the average GP of the control reactors) were not investigated further for COD, TS, TVS, TKN, and NH_3 removals. It has to be underlined at this point that a 35% reduction in the AD performance is neither suggested nor acceptable. However, it can be considered as a loss which can be recovered by pretreatment (Hartmann et al., 2000; Schieder et al., 2000; Angelidaki and Ahring, 2000), phase separation (Demirer and Chen, 2004, 2005a), or enhanced process configuration (Demirer and Chen, 2005b). So it is still noteworthy to investigate the reactors with less than 35% reduction in GP, since their NH_3 removals were superior to that of the control reactor.

Therefore, Reactors 4, 10, 11, 13, and 14 (Fig. 1) were terminated and the contents were analyzed for COD, TS, TVS, TKN, NH_3-N, and pH at the end of the 56th day, when the steady state relative to GP was reached. Table 4 indicates COD, TS, TVS, TKN, and NH_3 concentrations in the reactors at the end of the 56th day. In order to evaluate the performance of these reactors, their COD, TS,

Table 3
The GP, NGP, and NNGP in the reactors

Reactor	GP (mL)	NGP (mL)	NNGP
Blank	101.8	–	–
Control	202.9	101.10	1.000
1	136.5	34.70	0.343
2	157.8	56.00	0.554
3	14.5	−87.40	−0.864
4	173.0	71.20	0.704
5	8.1	−93.70	−0.927
6	9.0	−92.80	−0.918
7	122.2	20.40	0.202
8	153.2	51.40	0.508
9	27.2	−74.60	−0.738
10	175.0	73.20	0.724
11	174.4	72.60	0.718
12	160.4	58.60	0.580
13	180.8	79.00	0.781
14	190.1	88.30	0.873

GP: gas production; NGP: net gas production; NNGP: normalized net gas production.

Table 4
COD, TS, TVS, TKN, and NH_3 concentrations in the reactors at the end of 56th day

Reactor	COD (mg/L)	TS (mg/L)	TVS (mg/L)	TKN (mg/L)	NH_3 (mg/L)
Control	4417.5±746.0	15,140.0±1753.6	9970.0±1286.9	626.0±257.4	359.0±11.3
4	5735.0	13,540.0±1640.5	9180.0±1018.2	511.0±1.4	311.2±5.7
10	5985.0	15,920.0±1725.3	5840.0±1357.6	509.5±201.5	322.0±36.8
11	4560.0	13,220.0	4760.0	444.0	298.0
13	5385.0	15,540.0	7820.0	510.0	277.0
14	3585.0	9980.0	6620.0	548.0	320.4

TVS, TKN, and NH_3 removal data were normalized relative to the control reactor and presented in Table 5.

When Tables 3 and 4 are evaluated together, it can be seen that the GP data represented the COD removal data fairly well except for Reactors 11 and 14. GP in anaerobic reactors is commonly used, especially in routine monitoring, as a substitute of COD (or BOD) measurements because it is an easy and quick method of assessing the biotransformation of organic material in the system. By anaerobic stoichiometry, COD removal and GP data can easily be related (Speece, 1996). However, the sensitivity and dependability of COD measurements is generally considered to be better than gas measurement(s). In this study, COD data were used to evaluate the performance of the reactors (Tables 4 and 5). Furthermore, Demirer and Chen (2005a) reported that the VS and COD data obtained using dairy manure which was taken from the Dairy Center at Washington State University as in the case of this study had a very high correlation.

When the COD removal efficiency is considered, it can be seen that the resultant COD values for all the reactors except 14 were higher than that of the control reactor (Table 4). The COD removals of the Reactors 4, 10, 11, and 13 were 3–35% lower than the control reactor (Table 5). The parameters affecting GP also affect the COD removal efficiency since they are dependent parameters. Therefore, the factors discussed above, namely high pH levels (Speece, 1996), cation toxicity (Kugelman and Chin, 1971), and the struvite crystals anchoring the sludge particles (Borgerding, 1972) might have resulted in the observed reductions in COD removal efficiency. Furthermore, visual observation of the reactors indicated a whitish color of the reactor contents, which was a confirmation of the formation of struvite which is white in color (Doyle and Parsons, 2002). Struvite slurries, which have low solubility and settleability around the pH of 7.0, could possibly reduce the contact of the anaerobic bacteria with the substrate with the additional effect of poor mixing of the reactors.

Unlike the rest of the reactors, the COD removal was improved in Reactor 14 by 19% relative to the control reactor (Table 3). When the COD removals in Reactor 13 and 14 are compared, it can be seen that Reactor 13 removed 22% less COD than the control reactor, while Reactor 14 removed 19% more COD than the control. The only difference between Reactors 13 and 14 was the addition of 3000 mg/L of Na_2HPO_4 to Reactor 13. When this fact is considered along with the superior NH_3 removal in Reactor 13 relative to 14 (23% and 11% more than the control reactor, respectively), it could be speculated that the higher level of struvite precipitates formed in Reactor 13 might have inhibited the anaerobic bacteria by covering their surfaces. Even though it was not possible to differentiate and quantify between the different reactors, this phenomena was visually observed in some of the reactors operated including Reactor 13.

Table 5
Normalized performance of COD, TS, TVS, TKN, and NH_3 removal in reactors with more than 65% GP relative to control

Reactor	COD	TS	TVS	TKN	NH_3
C	1.00	1.00	1.00	1.00	1.00
4	1.30	0.89	0.92	0.82	0.87
10	1.35	1.05	0.59	0.81	0.90
11	1.03	0.87	0.48	0.71	0.83
13	1.22	1.03	0.78	0.82	0.77
14	0.81	0.66	0.66	0.88	0.89

If the concentrations of TKN and NH_3 in the supernatant of the control reactor are used to compare the NH_3 removal in the reactors in which the struvite formation was forced without inhibiting the digestion more than 35%, then the improvement of the process with regard to both TKN and NH_3 removal can be recognized. However, the reduction in the concentration of TKN in the reactor should be analyzed cautiously since TKN measurements include the freely existing NH_3 as well as the NH_3 released from the organic nitrogen as a result of its oxidation by sulfuric acid in the analysis. Therefore, the lower TKN concentrations observed in the reactors reflect in part the decrease in the concentration of NH_3. Moreover, since the difference in TKN and NH_3 concentrations in the reactors and the control reactor was not equal, there should be other mechanisms for the removal of TKN prevailing in the reactors, which could not be investigated further in this study.

The concentrations of TKN and NH_3 were lower in Reactors 4, 10, 11, 13, and 14 than those in the control reactor (Table 4). As stated previously, the removal of TKN and NH_3 was expected in Reactors 4 and 14, which were loaded only with $MgCl_2 \cdot 6H_2O$. The results revealed 18.4% and 12.5% lower TKN concentrations in Reactors 4 and 14, respectively, compared to the control reactor. The concentration of NH_3 decreased by 13.3% and 10.8% in Reactors 4 and 14, respectively, from the control reactor indicating that the influence of Mg^{2+} concentration on the removal of NH_3 was insignificant in the working range of the study. It was also observed that higher concentration of Mg^{2+} in the system inhibited the GP and reduced the COD removal efficiency of the reactors.

The concentration of TKN in the supernatant of Reactors 10, 11, and 13 was found to be 18.6%, 29.1%, and 18.5% lower than that in the control showing an improvement in the process with regard to TKN removal. However, the formation of struvite can be tracked by comparing the concentration of NH_3 in the reactors treated both with $MgCl_2 \cdot 6H_2O$ and Na_2HPO_4 and the control reactor. The concentration of NH_3 in the supernatant of Reactors 10, 11, and 13 decreased more compared to the control as the concentration of $MgCl_2 \cdot 6H_2O$ and Na_2HPO_4 decreased and the maximum observed reduction of 23% occurred in Reactor 13. This clearly indicates the need for optimization of the operational conditions not only based on COD removal but also struvite forming ions concentrations in the reactors.

It is also important to note that one of the major factors controlling the struvite formation and precipitation is the pH of the system, which was not adjusted in this study. The optimum pH for struvite formation to achieve maximum removal of NH_3 is reported as 9.0 in the literature (Nelson et al., 2003; Doyle and Parsons, 2002). Therefore, in the optimization of struvite formation within an anaerobic reactor the pH factor should be investigated within the pH range that the AD can tolerate.

4. Conclusions

Struvite precipitation potential is high in AD mainly due to proper pH values as well as high concentrations of the dissolved component ions (Ohlinger et al., 1998). Therefore, this study investigated coupling AD and controlled struvite precipitation in the same reactor to minimize the nitrogen removal costs and possibly increase the performance of the AD by reducing the ammonia concentrations which have an adverse effect on anaerobic bacteria (Borja et al., 1996; Krylova et al., 1997). The following conclusions can be drawn from the results of this study:

- The addition of $Mg(OH)_2$ as a source of Mg^{2+} inhibits the GP in the reactors and results in complete break down of the AD. Therefore, $MgCl_2 \cdot 6H_2O$ proved to be a better Mg^{2+} supply than $Mg(OH)_2$ for struvite formation.
- The addition of excess $MgCl_2 \cdot 6H_2O$ and Na_2HPO_4 decreased the COD removal performance of the reactors considerably.
- The anaerobic reactors supplemented with struvite precipitating chemicals removed significant amounts of NH_3 (10–23%) relative to the control reactor. The COD removal efficiency of the same reactors ranged between −36% and 19% relative to the control. This is a clear indication of the possibility of using anaerobic digesters for struvite precipitation as well. This could prove to be an advantage only if the anaerobic digestibility of the organic substrate (dairy manure in this study) could be kept at the level of the control reactor.
- Nineteen percent extra COD and almost 11% extra NH_3 removals were achieved relative to the control by only adding 1750 mg/L of $MgCl_2 \cdot 6H_2O$ to the anaerobic reactor.

The study sets forth some of the important aspects to be considered in future studies on the removal of NH_3 in AD reactors by struvite formation followed by its precipitation. For example, the effect of mixing, optimization of $MgCl_2 \cdot 6H_2O$ and Na_2HPO_4 doses, and pH of the system to achieve higher removals of NH_3 should be further investigated. Moreover, the contribution of NH_3 removal via struvite formation in total nitrogen removal should be studied in detail.

References

APHA, 1995. Standard Methods for the Examination of Water and Wastewater, nineteenth ed. Washington, DC.

Angelidaki, I., Ahring, B.K., 2000. Methods for increasing the biogas potential from the recalcitrant organic matter contained in manure. Water Science and Technology 41, 189–194.

Battistoni, P., Fava, G., Pavan, P., Musacco, A., Cecchi, F., 1997. Phosphate removal in anaerobic liquors by struvite crystallization without addition of chemicals: preliminary results. Water Research 31, 2925–2929.

Battistoni, P., Pavan, P., Prisciandaro, M., Cecchi, F., 2000. Struvite crystallization: a feasible and reliable way to fix phosphorus in anaerobic supernatants. Water Research 34, 3033–3041.

Beal, L.J., Burns, R.T., Stalder, K.J., 1999. Effect of anaerobic digestion on struvite production for nutrient removal from swine waste prior to land application. Presented at the 1999 ASAE International Meeting in Toronto, Canada. Paper No. 994042. ASAE, St. Joseph, MI.

Booram, C.V., Smith, R.J., Hagen, T.E., 1975. Crystalline phosphate precipitation from anaerobic animal waste treatment lagoon liquors. Transactions of the ASAE 18, 340–343.

Borgerding, J., 1972. Phosphate deposits in digestion systems. Journal of the Water Pollution Control Federation 44, 813–819.

Borja, R., Sanchez, E., Weiland, P., 1996. Influence of ammonia concentration on thermophilic anaerobic digestion of cattle manure in upflow anaerobic sludge blanket (UASB) reactors. Process Biochemistry 31, 477–483.

Cheng, J., Liu, B., 2002. Swine wastewater treatment in anaerobic digesters with floating medium. Transactions of the ASAE 45, 799–805.

Demirer, G.N., Chen, S., 2004. Effect of retention time and organic loading rate on anaerobic acidification and biogasification of dairy manure. Journal of Chemical Technology and Biotechnology 79, 1381–1387.

Demirer, G.N., Chen, S., 2005a. Two-phase anaerobic digestion of unscreened dairy manure. Process Biochemistry 40, 3542–3549.

Demirer, G.N., Chen, S., 2005b. Anaerobic digestion of dairy manure in a hybrid reactor with biogas recirculation. World Journal of Microbiology and Biotechnology 21, 1509–1514.

Demirer, G.N., Duran, M., Ergüder, T.H., Güven, E., Ugurlu, Ö., Tezel, U., 2000. Anaerobic treatability and biogas production potential studies of different agro-industrial wastewaters in Turkey. Biodegradation 11, 401–405.

Doyle, J.D., Parsons, S.A., 2002. Struvite formation, control and recovery. Water Research 36, 3925–3940.

Fujimoto, N., Mizuochi, T., Togami, Y., 1991. Phosphorus fixation in the sludge treatment system of a biological phosphorus removal process. Water Science and Technology 23, 635–640.

Güngör-Demirci, G., Demirer, G.N., 2004. Effect of initial COD concentration, nutrient addition, temperature and microbial acclimation on anaerobic treatability of broiler and cattle manure. Bioresource Technology 93, 109–117.

Hartmann, H., Angelidaki, I., Ahring, B.K., 2000. Increase of anaerobic degradation of particulate organic matter in full-scale biogas plants by mechanical maceration. Water Science and Technology 41, 145–153.

Henze, C.M., Harremoes, P., 1978. Nitrification and denitrification in waste water treatment. In: Mitchell, R. (Ed.), Water Pollution Microbiology, vol. 2. Wiley, New York, pp. 391–414.

Hobson, P.N., Wheatley, A., 1993. Anaerobic Digestion: Modern Theory and Practice. Elsevier, UK.

Krylova, N.I., Khabiboulline, R.E., Naumova, R.P., Nagel, M.A., 1997. The influence of ammonium and methods for removal during the anaerobic treatment of poultry manure. Journal of Chemical Technology and Biotechnology 70, 99–105.

Kugelman, I.J., Chin, K.K., 1971. Toxicity, synergism and the antagonism in anaerobic waste treatment processes, anaerobic biological treatment processes. In: Pohland, F.G. (Ed.), Advances in Chemistry Series, vol. 105. American Chemical Society, pp. 55–90.

Lee, J.J., Choi, C.U., Lee, M.J., Chung, I.H., Kim, D.S., 2004. A study of NH_3-N and P refixation by struvite formation in hybrid anaerobic reactor. Water Science and Technology 49, 207–214.

Li, X.Z., Zhao, Q.L., Hao, X.D., 1999. Ammonium removal from landfill leachate by chemical precipitation. Waste Management 19, 409–415.

Lusk, P., 1998. Methane recovery from animal manures the current opportunities casebook. National Renewable Energy Laboratory, NREL/SR-580-25145.

Martin Jr., J.H., Wright, P.E., Inglis, S.F., Roos, K.F., 2003. Evaluation of the performance of a 550 cow plug-flow anaerobic digester under steady-state conditions. In: The Ninth International Symposium on Animal, Agricultural and Food Processing Wastes. Research Triangle Park, North Carolina, USA, 12–15 October, pp. 350–359.

Miles, A., Ellis, T.G., 2001. Struvite precipitation potential for nutrient recovery from anaerobically treated wastes. Water Science and Technology 43, 259–266.

Mulder, A., 2003. The quest for sustainable nitrogen removal technologies. Water Science and Technology 48, 67–75.

Mulder, A., van de Graaf, A.A., Robertson, L.A., Kuenen, J.G., 1995. Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. FEMS Microbiology Ecology 16, 177–183.

Munch, E.V., Barr, K., 2001. Controlled struvite crystallization for removing phosphorus from anaerobic digester sidestreams. Water Research 35, 151–159.

Nelson, N.O., Mikkelsen, R.E., Hesterberg, D.L., 2003. Struvite precipitation in anaerobic swine lagoon liquid: effect of pH and Mg:P ratio and determination of rate constant. Bioresource Technology 89, 229–236.

Noike, T., Goo, I.S., Matsumoto, H., Miyahara, T., 2004. Development of a new type of anaerobic digestion process equipped with the function of nitrogen removal. Water Science and Technology 49, 173–179.

Ohlinger, K.N., Young, T.M., Schroeder, E.D., 1998. Predicting struvite formation in digestion. Water Research 32, 3607–3614.

Salutsky, M.L., Duseth, M.G., Ries, K.M., Shapiero, J.J., 1970. Ultimate disposal of phosphate from wastewater by recovery as fertilizer. Chemical Engineering Progress Symposium Series 107, 54–62.

Schieder, D., Schneider, R., Bischof, F., 2000. Thermal hydrolysis (TDH) as a pretreatment method for the digestion of organic waste. Water Science and Technology 41, 181–187.

Schuiling, R.D., Andrade, A., 1999. Recovery of struvite from calf manure. Environmental Technology 20, 765–768.

Speece, R.E., 1996. Anaerobic Biotechnology for Industrial Wastewater. Archae Press, USA.

Uludag-Demirer, S., Demirer, G.N., Chen, S., 2005. Ammonia removal from anaerobically digested dairy manure by struvite precipitation. Process Biochemistry 40, 3667–3674.

US Department of Agriculture, Economic Research Service, 2001. Confined animal production and manure nutrients. ERS Agriculture Information Bulletin No.771.

US Environmental Protection Agency, 2002. Managing manure with biogas recovery systems: improved performance at competitive costs. EPA-430-F-02-004.

Wen, Z., Liao, W., Chen, S., 2004. Hydrolysis of animal manure lignocellulosics for reducing sugar production. Bioresource Technology 91, 31–39.

Westermann, P.W., Safley, L.M., Barker, J.C., 1985. Crystalline build-up in swine and poultry recycle flush systems. In: Agricultural Waste Utilization and Management Proceedings of the Fifth International Symposium on Agricultural Wastes.

ELSEVIER

Journal of Environmental Management 86 (2008) 201–213

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Effectiveness of participatory planning for community management of fisheries in Bangladesh

Parvin Sultana[a,*], Savitri Abeyasekera[b]

[a]Flood Hazard Research Centre, Middlesex University, Queensway, Enfield, EN3 4SA, UK
[b]Statistical Services Centre, The University of Reading, Whiteknights Road, Reading, RG6 6FN, UK

Received 22 November 2005; received in revised form 13 November 2006; accepted 7 December 2006
Available online 8 February 2007

Abstract

This study provides statistical evidence that support for community-based management of resources was more effective when initiated through a process known as participatory action plan development (PAPD). Thirty-six sites were studied where community management of fisheries was facilitated by NGOs. All involved community participation and establishing local fisheries management institutions. However, communities were able to take up more conservation-related interventions and faced fewer conflicts in the 18 sites where a PAPD was the basis for collective action and institution development. This indicates the value and effectiveness of adopting good practice in participatory planning, such as PAPD, which helps diverse stakeholders find common problems and solutions for natural resource management.

Keywords: Community-based management; Participation; Fisheries; Bangladesh

1. Introduction

Participation in planning and development programmes has become increasingly central to international development assistance since the 1980s following Chambers (1983) and adoption of approaches such as Participatory Rural Appraisal that emphasise facilitation and the pre-eminence of local knowledge (Chambers, 1997). However, more recently the mantra of participation has been criticised as a "tyranny" that, for example, reinforces existing social relations or fails to understand local power relations, or has in practice focused on tools rather than empowerment (Cooke and Kothari, 2002; Holmes and Scoones, 2000; Michener, 1998; Mosse, 2002; Nelson and Wright, 1995). The extent to which participatory processes are holistic, build partnerships, build local institutions, create synergies across sectors, foster local ownership, develop local partners and create enabling environments for the transparent and accountable delivery of services at the community level, or replace more legitimate and sustainable existing institutions and processes has been questioned (Cooke and Kothari, 2002).

Participation in the form of community-based management of common pool natural resources has been promoted to improve their management and empower local communities. This has involved using local knowledge, recognising local institutions, establishing common property regimes, and developing partnerships and co-management between communities and government (Berkes et al., 1998; Ostrom, 1990; Pomeroy and Berkes, 1997). All such initiatives depend on community participation, so a major question is how best to initiate such regimes and what participatory planning methods are effective, given the (often considerable) diversity of interests among local communities. In Bangladesh a systematic methodology has been developed for consensus building for floodplain resource management that has been named participatory action plan development (PAPD) (Sultana and Thompson, 2004). The objective of this study was to provide empirical evidence of the effectiveness of PAPD, by comparing community-based fisheries

*Corresponding author. Tel.: +44 880 2 7215983; fax: +44 20 8411 5403.
E-mail address: parvin@dhaka.agni.com (P. Sultana).

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.12.027

management in sites where PAPD had been used with sites where non-PAPD approaches have been used, and to discuss the implications of the findings.

1.1. Participatory action plan development

Sultana and Thompson (2004) summarise the principles and steps in PAPD and how it fits into the overall process of establishing community-based natural resource management. PAPD recognises heterogeneous interests in natural resources (floodplains) and aims to be inclusive of these diverse interests. It involves a series of linked local workshops where different stakeholder groups participate separately and together to develop a plan for management of the common aquatic resources that they depend on (Barr and Dixon, 2001). The process is designed to ensure that poor people's interests are voiced and represented on an equal footing with those of more powerful stakeholders.

Stages four to eight in Box 1 involve two rounds of participatory workshop sessions with separate stakeholder groups followed in each case by a plenary session, and form the actual PAPD set within a wider process. Multi-stakeholder negotiations have been promoted by some as a useful framework in conflicting situations and to resolve deadlock (Janakarajan, 2004), but others have argued that the limitations of such methods include how the agenda is set, and biases against disadvantaged groups who may be manipulated by others (Edmunds and Wollenberg, 2001; Mosse, 1994; Moreyra and Wegerich, 2005). While similar, PAPD is designed not for resolving bi-polar conflicts, or large scale catchment issues, but for helping multiple stakeholder groups find common ground—shared problems and solutions—over which they may cooperate, and where the natural resource base is of modest size—typically up to about 10 villages (Barr and Dixon, 2001). The principle behind PAPD, adapted from Kaner (1996), is that members of any stakeholder category, but especially the disadvantaged (such as fishers in Bangladesh) are better able to express their views separate from other (dominant) categories of people. However, separate workshops will fail to develop a shared understanding of common problems and possible win-win solutions (consensus building). Therefore, PAPD is structured to have two rounds of divergent and convergent sessions. Through this it claims to find solutions that address problems shared by all stakeholders, and has been applied in a number of projects and locations in Bangladesh.

1.2. Community-based fisheries management

The community-based fisheries management project phase two (CBFM-2), supported by the UK Department

Box 1
Participatory Action Plan Development (PAPD) within the community-based fisheries management process

I. Scoping phase (*stages one to three*)
1. Situational analysis (summarising local knowledge)
2. Stakeholder identification and analysis (through key informants)
3. Household census and invitations to PAPD for a random sample of households (stratified by stakeholder categories)

II. Participatory planning phase—PAPD (*stages four to eight*)
4. Problem census (with each individual stakeholder group)
5. Compilation of problem rankings by facilitators (combining stakeholder group rankings)
6. Plenary with stakeholders and local leaders (to review and agree on main problems for solution analysis)
7. Solution and impact analysis (with each individual stakeholder group)
8. Plenary with stakeholders and secondary stakeholders (to present the process, identify feasible solutions, discuss institutional arrangements and next steps)

III. Implementation phase (*stages nine to thirteen*)
9. Develop and adapt community organisations and institutions for resource management
10. Community organisation develops detailed plan to implement solutions agreed in stage eight
11. Problem solving (review and adjust plans with community to mitigate or avoid any adverse impacts)
12. Implementation of action plan
13. Institutionalisation of management arrangements including local policy support.

for International Development and International Fund for Agricultural Development, started in September 2001 and is implemented by the WorldFish Center working with 13 NGO partners and the Department of Fisheries (WorldFish Center, 2003). The project aims to develop and test institutional arrangements for improved fisheries management involving the Government of Bangladesh, community-based organisations (CBOs) and fishing communities, and has worked in 115 water bodies (rivers, lakes and floodplains). Most water bodies are administered by government, and through the project, fishing rights are reserved for communities represented by their CBOs. Thus, it involves a community-based co-management approach, with decision-making devolved to the CBOs, which are formally recognised and advised by government. PAPD was used in 18 areas covering 42 water bodies, mainly by the Center for Natural Resources Studies (CNRS), and also by the WorldFish Center working with three other NGOs: Banchte Shekha, Caritas and Efforts for Rural Advancement. After undertaking PAPD, the NGOs formed CBOs comprising representatives of the different stakeholder groups to implement actions to address their common priority needs, particularly in fishery management, and then supported poorer fishers with training and credit. In the other sites (non-PAPD) the NGOs used their own approaches: reconnaissance studies and often some form of Participatory Rural Appraisal to form savings groups among their target population (essentially fishers) who then received training and credit. The NGOs based membership of the CBOs on these groups or their representatives, and helped the CBOs plan fishery management activities usually without discussion with the wider community.

1.3. Expected implications of PAPD for community-based management

As PAPD is a participatory planning step in the early stage of developing community-based resource management, and is based on consensus building, it was expected to have several positive implications. On this basis the hypotheses outlined and tested statistically in the following sections were formulated. These largely relate to enhancing social capital—the networks, norms and social trust that facilitate coordination and cooperation for mutual benefit (Putnam, 1995) and are grouped under five headings:

(a) Building appropriate community organisations is a major emphasis of CBFM, and as PAPD does not involve a fixed formula for CBO structure it is expected to lead to CBOs that include and involve a wider and more active range of stakeholder representatives. This should enhance local structural social capital—the composition and practices of formal and informal local institutions (Krishna and Schrader, 2000).
(b) Consensus building through PAPD was also expected to bring social capital gains beyond the CBOs, in terms of bonding (social cohesion) and bridging (links or connections across communities and groups, particularly ones that are vertical with local government).
(c) A key consequence of any enhancement in social capital and an indicator for it would be collective action. CBFM by definition involves collective action to sustain, and restore or enhance fisheries.
(d) This collective action is expected to bring benefits in terms of more sustainable fisheries, reversing past degradation and restoring production. Since PAPD identifies other shared natural resource problems and solutions, it could also result in other benefits for the community and individual households.
(e) On the other hand transaction costs—costs of information collection, time and effort spent in decision making, implementing decisions and in monitoring compliance and impacts—can be a significant cost of community-based management (Adhikari and Lovett, 2006). There are arguments that in the long-run co-management and community-based management systems may reduce transaction costs compared with centralised management, for example by establishing norms and high levels of compliance (Abdullah et al., 1998). Yet in the short term establishing community organisations and new management rules may require more time from community members, and this may be higher when multiple stakeholders are involved in consensus building.

Overall PAPD was expected to result in higher levels of trust, harmony, cooperation and collective action compared with CBFM sites without PAPD (Sultana and Thompson, 2004). The study aimed to generate evidence to test this. A further aim was to seek, through case studies, an understanding of the causal links if any between participatory planning and changes in communities and resource management. The study did not consider alternative methods of participatory planning, since no other systematic approaches were used in the CBFM-2 project.

2. Methodology

2.1. Study locations

Within the CBFM-2 project PAPD had been conducted in 18 locations, these formed sampling units for 'with-PAPD'. However, many of these covered more than one water body. In each of them at least one CBO was formed—either a River Management Committee or Beel[1] Management Committee, and in one case seventeen CBOs were formed for a large floodplain. Therefore, data from the respective CBOs (i.e. 59 CBOs surveyed in the 18

[1] Beels are floodplain depressions that usually retain water throughout the year; they may be "closed"—with few connections to other water bodies in the wet season, or "open" with a free flow of water and fish to other water bodies.

Table 1
Distribution of CBFM-2 sites covered by this study by NGO and water body type

	PAPD	Non-PAPD
NGO		
Center for Natural Resource Studies	14	0
Banchte Shekha	2	3
Proshika	0	5
Caritas	1	3
BRAC	0	5
Efforts for Rural Advancement	1	1
Center for Rural and Environment Development	0	1
Water body type		
Open beel	5	12
River	5	3
River and open beel	2	0
Floodplain	6	3
Total	18	18

Beels are floodplain depressions that usually retain water throughout the year.

PAPD units) were aggregated (averaged or summed as appropriate) for each sampling unit. To represent CBFM-2 water bodies without PAPD 'closed beels' (small well defined lakes with few outlets) which are stocked by the fisher community were not considered since PAPD was not conducted in this type of water body. Also excluded were a few sites where there were major problems such as prior legal cases. From the remaining water bodies 18 were purposively sampled to include sites where different NGOs were working in different types of water body (Table 1).

2.2. Data sources

Some project data came from household censuses in 2002, routine quarterly monitoring of project activities up to January 2004, an institutional survey, and transaction cost survey. A substantial part of the data was collected using structured focus group discussions in March–April 2004. Two focus groups were held in each sampled unit. Six to ten members of the CBO committee were invited for one focus group, and for the other, those committee members invited 6–10 of the poorer fishers, preferably from any traditional fisher community of their area, to participate in a separate focus group (Table 2). This method ensured that mixed groups were avoided and scorings made by the participants were for the same issues. Responses from the two focus groups were analysed separately, except where the results were very similar when the data were pooled or averaged across the two groups. The focus groups also generated qualitative information for use in interpreting the results and helped build up an understanding of causation.

2.3. Hypotheses

The study aimed to establish evidence of, and reasons for, any outcome and impact of the PAPD method. It also

Table 2
Number of respondents in the focus group discussions (participatory assessments)

Stakeholder type	PAPD			Non-PAPD		
	MC	Poor fishers	Total	MC	Poor fishers	Total
Full time fisher	36	37	73	30	36	66
Part time fisher	49	47	96	44	46	90
Subsistence fish	51	52	103	38	48	86
Fish trader	6	10	16	1	4	5
Kua/katha owner		5	5		4	4
Fish processor				3		3
Total	142	151	293	116	138	254

MC = member of management committee of community-based organization.
Kua (ditches) and katha (brushpiles) are used as fish aggregating devices by local landowners.

Table 3
Main hypotheses tested in this study

Community-based organization development
(i) PAPD results in more active community-based organizations
(ii) PAPD results in development of community organizations that are more holistic (involve more stakeholder categories)
(iii) PAPD results in community actions involving wider coverage of communities
(iv) PAPD results in faster development of community organizations

Social capital
(v) PAPD results in greater social cohesion
(vi) PAPD results in better links between community organizations and local government

Collective action
(vii) PAPD results in faster uptake of community actions for natural resource management
(viii) PAPD results in more community/collective actions for natural resource management
(ix) PAPD results in community actions that have greater compliance

Benefits
(x) PAPD results in more diverse and appreciated benefits from collective action

Transaction costs
(xi) PAPD requires greater time input from participant communities

aimed to understand causal linkages between this consensus building method and subsequent (development project) outcomes and impacts. This is relevant not only to those involved in community-based fisheries and wetland management in Bangladesh (where a series of projects have supported this type of management in over 400 locations), but also to debate over the merits and dominance of participatory approaches to development.

The main focus was to test the hypotheses in Table 3, comparing CBFM sites with and without PAPD. For some hypotheses several indicators were considered. Each hypothesis was tested, with respect to each corresponding indicator, using a general linear modelling procedure (Grafen and Hails, 2002) that generated an analysis of variance. Other determinants of the response variable were included in the modelling process so that the significance of the PAPD effect could be observed free from possible effects due to other confounding factors. For example, differences between sites may be affected by exogenous factors (such as the type of water body, existing social and user pressures on it, other use of the water body, number of poor fishers involved in decision making, and other development activities), and endogenous factors (the resources and skills available and used by NGOs in establishing CBFM organisations and activities). Case studies were also used (focus group discussions and key informants) to investigate causality of differences encountered and development of organisations/institutions and community actions.

2.4. Possible limitations

Quantitative analysis is inevitably limited by the data available, but in this case also by the timing of the assessment. Many factors will affect the process of developing community-based institutions for natural resource management. The longer the period to assess changes the fuller the picture of the process. On the other hand PAPD is a planning approach used at the initial stage, so it was appropriate to investigate indicators about two years after the PAPDs and before many other factors had influenced the process. Ideally effectiveness would be assessed again some years later, when project support has ended, and when PAPD would be just one of the factors investigated in assessing the effectiveness of CBFM itself. To address partially this limitation we include in the analysis a short update to March 2006.

While planning the analysis, due consideration was given to the majority of sites with PAPD having been supported by one NGO–CNRS. This raises a question whether the PAPD effect, where it was evident, was mainly a reflection of the implementing organisation. The number of sites available to explore this effect was very low. There were only eleven sites where both PAPD and non-PAPD were implemented by the same NGOs (Banchte Shekha, Caritas and ERA). The sample size here is too small to test the statistical significance of the PAPD effect. Nevertheless, the mean values for a few indicators were examined for cases where a PAPD effect was found (see Table 4). This showed positive impacts of the PAPD methodology for a restricted set of sites which exclude CNRS sites, and gives some confidence that the findings are not due to the large number of CNRS sites undertaking PAPD. Moreover, all NGOs posted similarly experienced staff to the field.

Lastly, differences in the project investment by site could be said to explain the apparent effectiveness of PAPD. NGO costs by site were estimated as accurately as possible, based on NGO reports up to 30 September 2003 (the end of

Table 4
Mean values for some indicators, restricted to NGOs, which had sites with and without PAPD[a]

Indicator	PAPD	Non-PAPD
Days to CBO formation	233	409
Number of awareness raising activities	24	6
Percent of poor fishers in CBO	65	22
Mean change in social cohesion (−5 to +5 scale)	4.2	2.2
Score for own and family benefits (0–10 scale)	7.2	4.8
Score for long-term benefits (0–10 scale)	8.5	6.7

[a]Banchte Shekha, Caritas and Efforts for Rural Advancement sites, four PAPD and seven non-PAPD sites.

the second year of the CBFM-2 project). Some difficulties were faced in this process, but accepting these limitations, the total expenditure per PAPD site was found to be about Tk 770,000 (about US$ 13,200). This was higher than in the non-PAPD sites at Tk 463,000. However, the total number of households in the villages covered by PAPD was about 1800 per site compared with about 1100 in non-PAPD sites, so the expenditures per household were very similar and account for the higher per site investment in CBFM with PAPD.

3. Results and discussion

Table 5 summarises the results of comparing sites with and without PAPD implementation with respect to indicators specific for eleven hypotheses. These hypotheses fall into five major groups. Results from each group are considered and discussed in turn below.

3.1. Community-based organisation development

Of the four indicators (see Table 5) used to test the hypothesis that PAPD results in more active CBOs, only one demonstrated a clear benefit under PAPD. This was with respect to the number of awareness raising activities, which was almost four times greater at PAPD sites ($p = 0.002$) compared to non-PAPD sites. There was also a wider range of types of awareness raising activities at the PAPD sites. Although such events also involved the NGOs, this gave strong evidence in support of the hypothesis that PAPD results in more active CBOs in terms of informing others, in the community and beyond, about fisheries management.

Over 70% of CBO members were reported to attend meetings, with some evidence ($p = 0.039$) of a higher percentage of members attending CBO meetings in non-PAPD sites compared to PAPD sites (Table 5). This was not expected. The reason may be because more CBOs in non-PAPD sites included savings and credit in their activities and this requires regular attendance at meetings. Although lacking statistical support, the case studies indicated that following a general consensus formed

Table 5
Results of testing the research hypotheses

Research hypothesis	Indicator used to evaluate the hypothesis	Data source	PAPD	Non-PAPD	Sig. Prob.
Community-based organization (CBO) development					
(i) PAPD results in more active CBOs	Average number of CBO meetings per month	QMR	0.60	1.26	ns
	% Attendance at CBO meetings	QMR	73	80	0.039
	Number of awareness raising activities with organizations outside the CBO	FGD-MC	15.8	4.4	0.002
	% of conflicts resolved by CBO	IM	23	32	ns
(ii) PAPD results in the formation of CBOs that are more holistic, and where poor are better represented	Number of categories of male stakeholders in CBO	QMR	2.8	3.1	ns
	Number of categories of female stakeholders in CBO	QMR	0.72	0.44	ns
	% of CBO comprises poor fishers and landless	QMR	66	35	0.000
(iii) PAPD results in community actions involving wider coverage of communities that perceive benefits	Number of stakeholder categories benefited	FGD-av	7.4	4.6	0.002
	Extent of benefits (1–10 score) for all stakeholders	FGD-av	5.6	4.4	0.008
	Extent of benefits (1–10 score) for fishers	FGD-PF	5.8	4.5	0.021
(iv) PAPD results in faster setting up of CBOs	Number of days taken to form CBO	QMR	263	370	0.004
Social capital					
(v) PAPD results in greater social cohesion and cooperation	Score for change in social cohesion—MC	FGD-MC	4.7	3.1	0.012
	Score for change in social cohesion—fishers	FGD-PF	4.1	2.0	0.007
(vi) PAPD results in better links with local government	Number of times received govt. support	FGD-MC	7.7	4.5	0.019
	Change in attitude score to CBFM in Union Parishad	FGD-MC	2.7	0.9	0.000
	Change in attitude score to CBFM in Upazila	FGD-MC	2.8	1.6	0.000
Collective action					
(vii) PAPD results in faster uptake of community actions for NR management	Number of days from fielding NGO staff to first action	QMR	308	481	0.000
	Number of days from CBO formation to first action	QMR	66	165	0.005
(viii) PAPD results in more community/collective actions for NR management	Number of actions planned and not implemented	QMR	1.0	3.2	0.000
	Number of actions implemented	QMR	4.7	1.7	0.000
(ix) PAPD results in community actions with greater compliance	Number of rules in place	IM	1.9	1.8	ns
	Number of rule breaking incidents	IM	1.0	5.5	0.000
	% of community know rules	IM	84	86	ns
	Total number of conflicts	IM	0.6	8.3	0.008
	Number of internal conflicts	IM	0.2	3.0	0.000
Benefits					
(x) PAPD results in greater benefits and community awareness and concern for collective sustainability and security actions	Number of own benefits	FGD-av	3.0	2.0	0.000
	Own benefit importance (score)	FGD-av	6.8	5.5	0.000
	Number of short term community benefits	FGD-av	2.2	1.9	ns
	Short term benefit importance (score)	FGD-av	6.9	5.9	ns
Time/transaction costs					
(xi) PAPD actions require greater time input from participant communities	Number of hours per person involved in CBFM activities last year	Transaction cost survey	179	391	0.017

Mean values of the indicators and significance level of analysis of variance from general linear model including PAPD as an explanatory factor ($n = 18$ in each group).

ns = test of PAPD versus non-PAPD difference is non-significant at the 5% level.

MC = member of management committee of the community-based organization.

Data sources: QMR = quarterly monitoring report of CBFM-2 project; FGD = focus group discussion undertaken for this study (MC = management committee, PF = poor fishers, av = average); IM = institutional monitoring survey by CBFM-2 in early 2004; transaction cost survey = interviews with a MC member and a PF from each site conducted at the end of the FGD.

through PAPD fewer meetings were needed, whereas in sites without that consensus more meetings were needed to discuss problems and conflicts.

PAPD was also expected to result in the formation of CBOs that include a wider range of stakeholder categories in their membership and a better representation of the poor (based on their participation and recognition of shared concerns in PAPD). There was no significant difference in the mean number of male stakeholder groups or female stakeholder groups included in the CBOs. However, the significantly higher percentage (66%) of the poor (fishers and other landless people) in the CBO membership in PAPD sites is consistent with the percentage of poor households in CBFM-2 sites (70%). The lower representation of the poor in CBOs in non-PAPD sites (35%) may be a consequence of the reported greater influence of Department of Fisheries and NGOs in listing fishers and selecting participants in non-PAPD sites with a bias towards inclusion of local influential people and slightly better off fishers. Thus, there are more poor people in the CBOs in PAPD sites, although whether they are also better represented in decisions depends on the type of people in the executive posts.

It was also hypothesised that more categories of stakeholders would perceive greater benefits in the PAPD sites because the actions taken up address common needs of the community. In all cases fishers were seen as the group benefiting most. But in PAPD sites both management committee members and poor fishers most often ($\geqslant$70% of focus groups) mentioned farmers, poor households, rich people and fish traders as groups who benefit. Farmer's benefit from catching more fish from their own land, while restoration of water for fisheries also provides more water to irrigate their crops. The rich can buy more fish at a cheaper rate. Poor households can fish for food, and when they have less work can sell their catch. Fish traders get fish locally at a cheaper rate. In PAPD sites, a quarter of committee focus groups mentioned women as beneficiaries because in most PAPD sites women are included in the committees.

Although the same types of beneficiary stakeholders were recognised in PAPD and non-PAPD sites, significantly more stakeholder categories ($p = 0.002$) were reported to benefit in the PAPD sites (>7 with PAPD compared with <5 without PAPD, Table 5). The focus groups also scored the extent of benefit for each stakeholder category on a scale of 1–10. The mean score for all types of stakeholder was significantly higher ($p = 0.008$) in PAPD sites (5.6) compared with non-PAPD sites (4.4). So, at this stage of developing CBFM, people from sites with PAPD, seem to benefit more than those in sites without PAPD. The same perceived benefit scores were also analysed just for the fishers since this group was reported to benefit in all sites. The mean score for fishers was higher in PAPD sites at 5.8, compared with 4.5 in non-PAPD sites ($p = 0.021$). Although PAPD has an effect of broadening CBO composition, it seems that the benefit to fishers remains higher in PAPD sites.

While committees can be formed rapidly, the CBFM-2 project aimed to establish capable CBOs representing and accountable to fishers and other stakeholders. In this context PAPD was expected to identify scope for collective action and to facilitate formation of appropriate CBOs and so it was hypothesised that PAPD results in faster setting up of CBOs. The number of days was calculated from the signing of agreements with each concerned NGO for its project involvement to the date that a CBO was first formed. In PAPD sites the time taken to establish a CBO was significantly less ($p = 0.004$): on average it took a year in non-PAPD sites, i.e. about 3 months longer than in PAPD sites (Table 5). In some cases a CBO comprising different types of stakeholders could be established in a flexible way based on greater understanding after PAPD, whereas some NGOs in non-PAPD sites followed a fixed process that required forming and establishing user groups of fishers for some time before the CBO was organised.

3.2. *Social capital*

To measure changes in levels of trust, harmony and cooperation (referred to here as social cohesion) each participant in the focus group discussions was asked to consider whether social cohesion among their stakeholder group had changed since CBFM activities started. A diagram was used to score this change on a scale of -5 to $+5$. The assessments were done separately for committee members and fishers. In general both groups reported increased social cohesion. However, in more than 50% of water bodies with a PAPD the respondents gave scores of four or five, compared with about 40% in non-PAPD sites. Only a few respondents in any site reported a decline in social cohesion. This resulted from differences in opinion over implementation of fishing rules such as a ban on fishing during the breeding season, or ban on using harmful gears. In a few water bodies where professional fishers lease the fishing rights, they exclude others from the community and social harmony had worsened.

In the PAPD sites the members of the management committees and poor fishers in their separate participatory assessments reported significantly greater changes in social cohesion, the difference being greater for the fishers. The case studies indicated that where better-off people were more dominant in the management committees, social cohesion was lower and fishers felt they had little role in management.

Because local elected councillors and officials participated in the PAPD plenary sessions, this might affect bridging social capital in the form of links between community and local government. The Department of Fisheries is a partner in the CBFM-2 project and has a role in co-management in all sites. Hence, all PAPD and non-PAPD sites reported meetings held with government. However, 78% of focus groups in PAPD sites reported support from the local council (Union Parishad), which was always involved in the PAPD plenary sessions, compared with 33% of focus groups in non-PAPD sites.

The focus groups reported that support from local government was mainly in the form of advice and conflict resolution. There was strong evidence ($p = 0.019$) that PAPD has enhanced the frequency of support received by the CBOs from government agencies. PAPD sites had received government help (advice or help in conflict resolution) on average about eight times compared with about five times in non-PAPD sites. This is consistent with the expectation that PAPD strengthens links between the community and secondary stakeholders.

The focus groups were asked to score the change in attitude of local councillors and sub-district officers during 2 years of the project using a -3 to $+3$ scale. There was strong evidence ($p = 0.000$) of a change in attitude. Initial scepticism over community participation in fishery management has given way to positive views. The change in scores was significantly greater (close to 3) in PAPD sites compared with non-PAPD sites (about 1–1.5). Hence, PAPD was associated with greater perceived changes in government attitudes in favour of community-based management. For example, after the PAPD in Shuluar Beel participants were very happy with the response of their Union Parishad Chairman. They kept him as an advisor to the CBO and said they can tap resources through him.

3.3. Collective action

3.3.1. Community resource management actions

PAPD was expected to result in improved collective action compared with other less holistic, inclusive and structured ways of initiating CBFM. This study supported the hypothesis that 'PAPD results in faster uptake of community actions for natural resources management'. On average PAPD resulted in saving 170 days of NGO facilitation time in achieving the first community actions for natural resource management. Moreover, after formation CBOs in sites with PAPD took about 2 months before their first actions compared with over 5 months in non-PAPD sites. In some PAPD sites, the community influence and awareness generated from the PAPD was so strong that the actions even started straight after the PAPD. In one case the CBO was formed during the last plenary session of the PAPD proper, they planned to stop use of harmful gears, and implemented this within a few days.

The difference in the time from CBO formation to first action was largely because first actions took much longer in open beel and river areas among non-PAPD sites compared to PAPD sites. There appeared to be no effect due to PAPD in floodplain beels. The difference in rivers may be because before the project, most of the fishers had limited access in rivers due to intense fishing effort including many brushpiles[2] made by better off people since the rivers became open access in 1995. PAPD resulted in a general consensus on the problems amongst all local stakeholders including support of local influential people to end harmful fishing, and the formation of CBOs that included these different stakeholders. Therefore they started to remove cross dams and later they banned harmful gear use and restricted brushpiles. In the non-PAPD river sites CBOs were formed exclusively of fishers based on small groups developed by the NGOs. Fishers had difficulty establishing their rights and government recognition of their rights was delayed. Without PAPD a long process of awareness building among the community was needed after the fisher-based CBOs were formed[3].

The hypothesis 'PAPD results in more collective actions for natural resources management' was tested using data on two indicators: number of actions planned but not achieved, and number of actions implemented. Each site has its own unique combination of environment, CBO and its members' interests. Therefore, the plans for fishery management were also different for each site. The typical actions planned by the CBOs include: fish sanctuaries, closed seasons, bans on harmful fishing gears, excavation of silted up areas, stock enhancement,[4] land purchase, tree planting, and making community centres.

Although similar numbers of actions were planned, the CBOs where PAPD was held succeeded in implementing on average three more actions each than the CBOs in non-PAPD sites (Table 5). This is not surprising since the PAPD process generates plans that contain agreed activities. By comparison non-PAPD sites lacked such a clear process for plan development and so identification of planned activities was not systematic or so widely supported. Conflicts in non-PAPD sites also resulted in planned actions not always being undertaken.

3.3.2. Compliance

The hypothesis that PAPD results in community actions with greater compliance was accepted (Table 5). Although after 2 years there were on average the same number of

[2]Brushpiles or *katha* are traditional fish aggregating devices that provide shelter for fish and are fished out 2–3 times in a dry season. They are a way of establishing exclusive ownership of water areas. Generally better off people have the funds and social power to establish them, and then hire traditional fishers to harvest the fish.

[3]"Fisher based CBOs" or "fisher led CBFM" indicate institutional arrangements where people who fish for an income are organised into CBOs. Often with project support they hold fishing rights by paying the annual lease to the government using revolving funds provide by NGOs. Before the same fishery might have been leased to a fisher cooperative but *de facto* control lay in the hands of moneylenders. In "community led CBFM" all local stakeholders, including fishers, are involved in planning and implementing decisions to ensure more sustainable fishing. More often this approach has been adopted where there is no leasing and where a PAPD was used.

[4]Release of fish to augment the natural populations and stocks, this could include carps which do not reproduce and have to be stocked each year, and/or native species that can reproduce and repopulate the water body.

fishing rules in place in CBFM sites with and without PAPD and most (over 80%) of the community were reported in focus groups to be aware of these rules, there were about five times more rule breaking incidents and many more conflicts in the non-PAPD sites than in the PAPD sites.

In all CBFM sites it was expected that participant communities would agree on their management activities, institutions and organisations, and that they could reduce internal conflict among participants. Lower numbers of conflicts within the community at PAPD sites could be attributed to wider consensus linked with PAPD and increased confidence among the committee to negotiate or bargain with others within or outside the local community. Further, the CBFM project has helped establish 'cluster committees' comprising of representatives of adjacent CBOs which coordinate among adjacent water bodies within the cluster area, helping to reduce conflicts with people from neighbouring areas (and more PAPD than non-PAPD sites are in such cluster locations).

In the analysis model both PAPD and water body type (which is associated with property rights and CBO type) had significant effects on the number of rule-breaking incidents. The floodplain beels have always been private land and typically the involved villages recognise common rights among their inhabitants to fish in the wet season, multi-stakeholder committees were established in this environment in both PAPD and non-PAPD sites and had few rule-breaking incidents. Rivers were leased fisheries until 1995, open access since then made implementing rules difficult, consequently under CBFM conflicts were more common, but less so in PAPD river sites. The non-PAPD rivers have fisher-led CBFM, which explains the tendency of others to break rules that were developed by only a certain group.

For example, in Shuluar Beel community members cooperated and complied with decisions made in a PAPD, although in the previous 8 months the NGO made little progress in bringing together the community to manage this floodplain beel. The many ditch owners there previously drained out their ditches to catch all the fish trapped after the wet season. During PAPD, the ditch owners committed not to completely drain the ditches, and some of the ditch owners joined the management committee to continue representing their views. After adopting this rule they reported three to four times more fish in the following monsoon season. However, in the non-PAPD site of Chitra River, the fishers agreed to protect fish in the early monsoon when they move from the river to the nearby beel to breed. But when they found that their neighbours were catching fish at that time in the beel, lacking an underlying consensus the fishers broke their own rules and also started to fish. Thus, the planning process and inclusiveness of PAPD compared to agreements among one group of fishers was critical to the acceptance of local fishing rules.

3.4. Benefits

In the participatory assessments, management committee members and fishers not involved in the committee separately assessed benefits for themselves and for their community in the short term. This was done in terms of the types of benefit they had or expected and their importance on a scale of 1–10. Significantly more types of own benefits were reported in PAPD sites (Table 5). On average respondents reported one more type of own benefit and rated their importance higher at PAPD sites. This was most evident in open beels, the management bodies in PAPD and non-PAPD floodplain sites represent different users and everyone found some benefits. Increases in fish were reported in almost all PAPD sites, and most respondents saw income and knowledge gains. Almost 75% of non-PAPD sites also reported fish catch increases but income and knowledge gains were less common (Table 6).

Table 6
Types of own benefit received or expected in the future identified through participatory assessment (percentage of respondent focus groups reporting)

Types of own benefits	PAPD		Non-PAPD	
	MC	Poor fisher	MC	Poor fisher
Fish increased	94	94	72	72
Income increased	72	61	39	56
Knowledge increased	61	56	33	39
Protein consumption increased	6	6	0	0
Irrigation	22	17	17	11
Income generating activity introduced	33	11	11	0
Fish price decreased	6	6	17	6
Ensured economic uplift	11	17	0	6
Participation	22	6	11	0
Ownership/right over water body	0	11	6	0
Recognition by Government/NGO	6	6	6	0

Multiple responses were possible; the maximum number of responses was 18 for each of the four respondent categories for each type of benefit.
MC = member of management committee of community-based organization.

Short-term community benefits were few and did not differ with PAPD. More long-term community benefits were identified. Focus groups in both PAPD and non-PAPD sites emphasised fishery benefits such as increased biodiversity, conservation measures and limiting fishing effort; but participants where there had been a PAPD regarded community participation and linkages with local government as important long-term benefits, whereas in sites without PAPD establishing ownership and access for fishers were expected.

In case studies, the participants from PAPD sites mentioned increased income as their main benefit—as a result they can send their children to school, and can afford health care and better food, especially for children. They now report consuming more fish. Some of the beneficiaries mentioned increased knowledge through PAPD, training, meetings, and through meeting visitors. They particularly mentioned the PAPD as the first gathering where they freely raised their own problems and proposed solutions that were taken into consideration. They felt that they were given attention and they were not controlled by anyone. The same types of benefits were reported in non-PAPD sites but were not so widespread at the time of assessment. Focus group respondents from non-PAPD sites said that decisions were taken in a big meeting where powerful people dominated decisions and poor people were given less attention.

3.5. Transaction costs

The evidence rejects the hypothesis that PAPD requires participants to spend more time on CBFM. Instead people in PAPD sites appeared to spend less time for community action in the past year. The analysis included both PAPD and respondent type as factors. Combining PAPD and non-PAPD sites, management committee members spent more than double the time of poor fishers, which was expected. This includes time for meetings and the fact that while fishing the committee members, being responsible for the rules set by the committees, are watching to see that rules are not broken. Given the lower incidence of rule breaking and conflict, the reduced time spent on CBFM in PAPD sites suggests that already after about 2 years those management actions that are in place are being observed voluntarily because of the general consensus reached through PAPD. The other factor is that poor fishers have less involvement in NGO groups in these sites than they do in the non-PAPD sites. These savings and credit groups take up time in addition to that related with resource management (of course they also bring direct benefits to those group members in the form of loans and training). Another factor is that a higher incidence of conflicts in the non-PAPD sites means that more time is taken up in addressing conflicts and legal cases in some of these sites.

4. Implications of findings

4.1. Implications for participatory planning

This study shows that broader-based community-led CBFM associated with PAPD was far more effective than narrower fisher-led CBFM associated with many of the non-PAPD sites. The PAPD process differs from less structured approaches such as PRAs as it considers opinions of each stakeholder group separately and then presents each stakeholder group's plans to all stakeholder groups in a plenary session so that all can understand each others' problems and suggestions. Common and uncommon issues are then considered for final planning. In PAPD the scope for powerful people to dominate poor people's views in the planning stage should be minimised by balanced facilitation, since the views of each stakeholder group developed separately are shared with one another in the plenary sessions.

Participation may force the participants to agree with the majority. In some non-PAPD sites it was reported that participants did not say directly what they wanted to avoid antagonising more powerful elders or neighbours, supporting some of the criticisms of participation (Mosse, 1994). However, in PAPD sites the evidence suggests that the opportunity to discuss issues initially, in separate stakeholder groups, helped ensure that actions that benefited the poor were taken. Ensuring participation of poor people (fishers and non-fishers) in the PAPD in homogenous groups let them express their own views without fear and anxiety. Consequently they felt honoured and more confident in group discussions. In addition, through interaction with leaders and officials and exposure to outsiders in the PAPD plenary they became more confident in demanding services.

It is argued that projects influence the way in which people express their needs through participatory methods. However, in PAPD many types of needs were raised by different stakeholder groups. Although the CBFM project focused on addressing wetland and fishery management needs, other community needs were voiced in the PAPDs and in those sites the NGOs addressed some of these needs, for example providing tubewells for drinking water and sanitary latrines.

Concerns over elite domination have been raised regarding CBFM, and regarding actions based on consensus that are perceived as further empowering those vested interests that may have manipulated decision making in the first place (Mohan, 2002). The case studies revealed no evidence that PAPD was associated with elite dominance. In PAPD sites most CBOs (95%) were formed with representatives of different types of stakeholder yet local elites have not taken control for their profit. In non-PAPD sites most CBOs (78%) were formed just of poor fishers, yet in these sites the number of conflicts and court cases has been high, often over attempts by local elites to control the fisheries. This may also be affected by the productivity and lease value of the water bodies, which tended to be lower in PAPD sites[5]. There is evidence during the same period from the Fourth Fisheries Project in Bangladesh that CBOs developed without PAPD but including a wide range of local stakeholders tend to be dominated by elites where there are more valuable resources and funds to be handled by the CBO, for example in water bodies leased from the government and ones with annual stocking of fish (Aeron-Thomas, 2003; Begum, 2004).

4.2. Implications for community-based management

Case studies indicated that the nature of interactions between community and government differ with the use of PAPD. In PAPD sites, government officials were invited in the final plenary of each PAPD, they saw and could make suggestions about or endorse the outcomes, but had not been involved in the stakeholder working groups. At these sites, the PAPD participants proposed who would represent their interests in the CBOs. In non-PAPD sites, participant lists were prepared by the NGOs and endorsed by the Department of Fisheries, which then played a more leading role in the development of plans by that group of participants. Thus, the PAPD plenary session allowed linkages to be established with concerned government officers, after the stakeholder groups had developed their views and proposals. Government officials were happy to see local knowledge, skills, analysis, and proposals presented, and generally made supportive commitments during the plenary, giving the resulting CBOs greater confidence.

The average number of resource management activities planned was about the same for both PAPD and non-PAPD sites, but the number implemented was more than double in PAPD sites. Two examples illustrate the processes involved. In Fatki River the community agreed in PAPD that re-excavation of silted up parts of the river was needed. They discussed the plan with government engineers, and then organised the labour team for implementing the plan. All types of stakeholders were involved in the process and there was no opposition. However, in Dubail Beel (non-PAPD) the participants took decisions on land purchase and excavation to deepen dry season fish habitat, but did not consider the risk that this could not be achieved without funds and consensus; conflicts arose over the land and the plan was not implemented.

[5]The type of water body was significant as a confounding factor for several indicators. Property rights and conflicts are associated with water body type. Open beels were leased to the CBOs, rivers had no lease payment, and floodplains are private land that is a seasonal common pool fishery. Even without a PAPD the CBOs in floodplains tended to be multi-stakeholder, and for some indicators such as collective action the PAPD effect in floodplains was less than for rivers (where open access made it very difficult to introduce management plans and rules without a wide consensus developed through PAPD).

Community-based management in general is expected to minimise conflict and rule breaking and increase cooperation and voluntary compliance among community members. Due to greater social cohesion, higher awareness and better coordination with different agencies, internal conflict among the participants in CBFM sites where there was a PAPD was less and rule-breaking incidences were few.

The initial transaction costs for fishery management were expected to be high because PAPD and implementation of the planned actions were predicted to take more time from participants (for workshops) than in non-PAPD sites. However, this study found some evidence that transaction costs were less in the PAPD sites than in the non-PAPD sites, even including the time taken for the PAPD, which may be associated with achieving greater consensus and voluntary compliance.

The number of days between NGO staff recruitment and CBO formation depended on the capacity and skill of the staff and the approach adopted. In PAPD sites, the CBO formation was faster because the community itself pushed the staff to help form a CBO and implement their planned actions. In non-PAPD sites staff lacked a structured approach to planning, and spent a great deal of time coordinating and organising people for CBO formation. Moreover, the community itself was not so sure about the project objectives, what activity the NGO wanted and the expected outcomes. In the case of PAPD sites, community members could plan what to do under the project at the start and they had an opportunity to put any questions to the implementing NGO, the wider community, the local representatives, and government agencies during the PAPD.

The time scale of this study focuses on PAPD as an initial input to the process of establishing community-based management. The eventual sustainability of CBOs and the fisheries, and the benefits accruing to fishers in the long term are the ultimate indicators of effectiveness. A slow inclusive process might be more or as effective as one based on PAPD. The evidence so far indicates positive changes, including in governance, related to the PAPD sites, and there is some urgency for this because without management changes these fisheries are rapidly degrading. To update the assessment we reviewed project assessments of all the same CBOs conducted in April 2006. These were subjective assessments made by staff not working directly in these sites that considered the role of poor fishers, extent of conflict, linkages with government, absence of elite influence, fish and biodiversity changes, and the management and financial capacity of the CBOs. These were scored from 1 to 10 (highest), and a PAPD effect was still present (mean score for PAPD sites 6.4, for non-PAPD sites 4.0). None of the PAPD sites scored below 4 compared with a third of non-PAPD sites, which appear to be failing to develop effective community-based management. While limited in scope this suggests that PAPD has had a more lasting positive influence on the process.

5. Conclusions

This study provides quantitative evidence that PAPD is effective in terms of more efficient and inclusive participatory planning and initiation of community-based resource management. It also shows that PAPD increases participation of the poor, and brings decisions, actions and benefits faster because of the consensus created.

To design effective community management in a site, all stakeholders' opinions on feasibility, pros and cons of options, and assessment of the benefits and disbenefits to each stakeholder need to be judged. PAPD provides those opportunities. PAPD identified those people interested to mobilise for common interests, and also helped reveal linkages of the villagers with the local power structure and government agencies.

PAPD was very effective in its aim of developing plans in a participatory way keeping in mind all stakeholders' interests. There is strong evidence that plans prepared through the PAPD processes were mostly implemented, and it took less time to implement actions after the planning process and also after CBO formation, saving time and costs in facilitating participatory management of natural resources.

The study assessed changes over only about two years, which was too short to conclude on the sustainability of CBFM including the institutions developed with or without a PAPD. However, there is limited evidence that the CBOs initiated through PAPD have remained more effective, suggesting that PAPD results in longer lasting advantages that should be sustainable. This supports the use of multi-stakeholder planning processes, rather than a focus on specific target groups, and shows the effectiveness at the local level of a structured process that helps to draw out areas of consensus over natural resource problems and solutions.

These benefits can be scaled up within Bangladesh in the natural resources sectors, which are increasingly taking a community-based approach. They are also relevant internationally. Awareness of PAPD has not yet changed the practices of many organisations in Bangladesh. For example, in the CBFM project all partner NGOs were oriented in and participated in a demonstration PAPD in 2002, yet most did not use it widely. One reason is that most of these NGOs have a target group approach that focuses only on households within a specific poverty range and do not see that a more inclusive PAPD approach and consensus can be merged with specific support for poor fishers and other poor resource users. PAPD should not be seen as another technique or formula, but as a starting point, that can be effective when it is followed by flexible and responsive support to address local resource management needs. Further efforts are needed to promote the use of PAPD since this study indicates that the PAPD process provides a structured approach and a focal point for generating local plans and their effective implementation, better community participation, greater unity among stakeholders including the poor, and support from local leaders, compared with community-based management with NGO support but with no PAPD.

Acknowledgements

This paper is based on a study supported by the Natural Resources Systems Programme of the UK DFID (PD131). We thank Christopher Floyd and Paul Thompson for advice on the design of the study, and seven anonymous reviewers for their improvements. The study would not have been possible without the cooperation of the concerned partners of the CBFM-2 project, especially Center for Natural Resource Studies, and the help of many members of the participant communities, for which we are very grateful. This document is in part an output from a project funded by the UK DFID for the benefit of developing countries. The views expressed are not necessarily those of DFID.

References

Abdullah, N.M.R., Kuperan, K., Pomeroy, R.S., 1998. Transaction costs and fisheries co-management. Marine Resources Economics 13, 103–114.

Aeron-Thomas, M., 2003. FFP beneficiary impact monitoring of open water fisheries component: synthesis of key issues from sites covered 2002–3. Fourth Fisheries Project, Department of Fisheries, Dhaka.

Adhikari, B., Lovett, J.C., 2006. Transaction costs and community-based natural resource management in Nepal. Journal of Environmental Management 78, 5–15.

Barr, J.J.F., Dixon, P.-J., 2001. Methods for consensus building for management of common property resources. Final Technical Report of R7562, Centre for Land Use and Water Resources Research, Newcastle University, Newcastle.

Begum, S., 2004. Pro-poor community-based fisheries management: issues and case studies. Fourth Fisheries Project, Department of Fisheries, Dhaka.

Berkes, F., Feeny, D., McCay, B.J., Acheson, J.M., 1998. The benefits of the commons. Nature 340, 91–93.

Chambers, R., 1983. Rural Development: Putting the Last First. Longman, London.

Chambers, R., 1997. Whose Reality Counts? Putting the First Last. IT Publications, London.

Cooke, B., Kothari, U., 2002. The case for participation as tyranny. In: Cooke, B., Kothari, U. (Eds.), Participation: The New Tyranny? Zed Books, London, pp. 1–15.

Edmunds, D., Wollenberg, E.A., 2001. Strategic approach to multi-stakeholder negotiations. Development and Change 32 (2), 231–253.

Grafen, A., Hails, R., 2002. Modern Statistics for the Life Sciences. Oxford University Press, Oxford.

Holmes, T., Scoones, I., 2000. Participatory policy processes: experiences from North and South. IDS Working Paper 113, Institute of Development Studies, Brighton.

Janakarajan, S., 2004. A snake in the grass! Unequal power, unequal contracts and unexplained conflicts: facilitating negotiations over water conflicts in peri-urban catchments. In: Paper Presented at the Conference on Market Development of Water and Waste Technologies through Environmental Economics, 28–29 May 2004, Paris.

Kaner, S., 1996. Facilitator's Guide to Participatory Decision Making. New Society Publishers, British Columbia, and Community at Work, San Francisco.

Krishna, A., Shrader, E., 2000. Cross-cultural measures of social capital: a tool and results from India and Panama. Social Capital Initiative Working Paper Series, World Bank, Washington, D.C.

Michener, V., 1998. The participatory approach: contradiction and co-option in Burkina Faso. World Development 26 (12), 2105–2118.

Mohan, G., 2002. Beyond participation: strategies for deeper empowerment. In: Cooke, B., Kothari, U. (Eds.), Participation: the New Tyranny? Zed Books, London, pp. 153–167.

Moreyra, A., Wegerich, K., 2005. Multi-stakeholder platforms as problems of eating out, the case of Cerro Chapelco in Patagonia, Argentina. In: Paper Presented at ESRC Seminar Series Water Governance—Challenging the Consensus, Seminar 3: Politics, Institutions and Participation, 27–28 June 2005, The Hague.

Mosse, D., 1994. Authority, gender and knowledge: theoretical reflections on the practice of participatory rural appraisal. Development and Change 25 (3), 497–526.

Mosse, D., 2002. People's knowledge, participation and patronage: operations and representations in rural development. In: Cooke, B., Kothari, U. (Eds.), Participation: The New Tyranny? Zed Books, London, pp. 16–35.

Nelson, N., Wright, S. (Eds.), 1995. Power and Participatory Development: The Theory and Practice. IT Publications, London.

Ostrom, E., 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge.

Pomeroy, R.S., Berkes, F., 1997. Two to tango: the role of government in fisheries co-management. Marine Policy 21 (5), 465–480.

Putnam, R.D., 1995. Bowling alone: America's declining social capital. Journal of Democracy January, 65–78.

Sultana, P., Thompson, P., 2004. Methods of consensus building for community-based fisheries management in Bangladesh and the Mekong Delta. Agricultural Systems 82 (3), 327–353.

WorldFish Center, 2003. Community-based Fisheries Management phase 2 Annual Report September 2001–December 2002. WorldFish Center, Dhaka.

Journal of Environmental Management

Journal of Environmental Management 86 (2008) 214–228

www.elsevier.com/locate/jenvman

Are farmers in England equipped to meet the knowledge challenge of sustainable soil management? An analysis of farmer and advisor views

Julie Ingram*

Countryside & Community Research Unit, University of Gloucestershire, Dunholme Villa, The Park, Cheltenham GL50 2RH, UK

Received 29 October 2005; received in revised form 29 October 2006; accepted 7 December 2006
Available online 19 March 2007

Abstract

Concern about the agricultural soil resource in England has led to the introduction of a range of measures, which potentially challenge farmers' knowledge about the soil and its management. Our understanding however of how well-equipped farmers are with regard to effectively carrying out more complex and knowledge intensive sustainable soil management practices is limited. Specifically, by drawing on the concept of scientific and tacit forms of knowledge, this paper examines the knowledge of soils held by farmers through analysis of data collected from semi-structured interviews with farmers and agricultural advisors and supplemented with data from an extensive postal questionnaire survey of advisors. The data indicate that, while farmers are technically well informed, they can often lack the in-depth scientific knowledge required to implement more complex practices such as using the nutrient value of manures. They also reveal that, while most farmers have good knowledge of their own soils, their tacit knowledge of soil management can be weak, notably in relation to cultivation. The paper concludes that although farmers' knowledge about soil and its sustainable management appears in general to be well developed there are some areas, which need to be significantly enhanced and as such require both a policy response and further research effort.

Keywords: Scientific knowledge; Tacit knowledge; Farmer; Agricultural advisor; Soil management; Best management practice; Diffuse pollution; Cultivation

1. Introduction

The centrality of knowledge to agriculture has been highlighted by a number of commentators (Winter, 1997; Morgan and Murdoch, 2000), with knowledge described as the 'fourth factor of production' because of the widely differing knowledge, skills and aptitudes farmers need for producing food (Winter, 1997). This is the case today more than ever before with the emergence of policies encouraging more sustainable farming practices, which are considered to be more complex and more demanding on the skills and knowledge of farmers than conventional farming (Kloppenburg, 1991; Röling and Jiggins, 1994). Mounting evidence of threats to the agricultural soil resource have brought calls for more sustainable management of this vital resource and recent policy developments in Europe and in England mean that demands on farmers' soil management competencies will increase. However, our understanding of the nature and extent of knowledge about this crucial resource held by farmers in England is poorly developed, particularly in comparison with our appreciation of how they manage, and impact, other natural resources such as water, nature and landscape (Lowe et al., 1997; Morris and Potter, 1995; Harrison et al., 1998).

As such, this paper reports the findings of research into the nature and extent of farmers' knowledge about soils through analysis of data collected from semi-structured interviews with farmers and agricultural advisors drawn primarily from those interacting with two projects in England[1] promoting soil management practices: the UK-wide Soil Management Initiative (SMI) and the Landcare

*Tel.: +44 (0) 1242 7141 34/4122.
E-mail address: jingram@glos.ac.uk.

[1]The focus on England rather than the UK as a whole is justified as policy making is devolved to the agricultural departments of the constitutive countries i.e. England, Wales, Scotland and Northern Ireland.

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.036

Partnership in the south-west of the country. The interviews were supplemented with data from an extensive postal questionnaire survey of advisors.

2. Increasing demands on farmers by policy makers

Decline in soil quality in the UK has been largely attributed to intensive arable farming. Subsidies provided as part of the Common Agricultural Policy (CAP) through the 1970–1980s encouraged continuous arable cropping, winter cereals, increased cultivation with heavier machinery, ploughing up of pasture, minimal rotations, the inappropriate use of marginal lands, and overgrazing in upland areas all resulting in negative consequences for the soil (Boardman, 1990; Baldock and Mitchell, 1995; DETR, 1998; Joint Nature Conservation Council, 2002). More recent reports suggest that these practices are continuing and the farming community have been urged to improve their understanding and husbandry of soil (Environment Agency, 2004a, b).

Following calls for more sustainable use of soil (Royal Commission on Environmental Pollution, 1996; Defra, 2002a, b) a number of policy initiatives have emerged in an attempt to counter the increasing threats to the agricultural soil resource both in Europe and in England (DETR, 2001; Commission of the European Communities, 2002). These include 'The First Soil Action Plan for England' (Defra, 2004a); the requirements for farmers to prepare a Soil Management Plan as a condition of receiving the Single Farm Payment[2] (Defra, 2005); and in some cases for payment under the new Environmental Stewardship Scheme. These, together with the introduction of catchment sensitive farming as part of implementation of the EC's Water Framework Directive (Defra, 2004b), all herald a new era of policy concern for soil in England. The farmers' responsibility in achieving the goals associated with these policies is clear, as stated in a popular farming press magazine 'Soil management is something no farmer can afford to ignore' (Farmers Weekly, 2004a).

These policies build on previous initiatives, which have promoted more sustainable 'best management practice' of soil to farmers (MAFF, 1998, 1999a, b; ADAS et al., 2000; Environment Agency, 2001). Best management practices for soil are based on a number of fundamental principles of good soil husbandry including: maintenance of soil structure through enhanced soil organic matter content and protection from compaction, overworking and runoff; as well as the management of soil as a buffer for nutrients by using artificial and organic fertilisers effectively and efficiently. Farmers can lack familiarity and experience with such practices (Park et al., 1997), which are considered to be more knowledge intensive than conventional practices in that they are non-prescriptive and demand attention to detail, observation, as well as an understanding of the scientific principles on which they are based (Röling and Jiggins, 1994; OECD, 2001). Clearly, knowledge is particularly significant given the complexity of the practices and the need for farmers to adapt them to their own soil types. It is to the issue of farmers' knowledge that the discussion now turns.

3. Farmers' knowledge about soil and its sustainable management

Whilst it is accepted that farmers' knowledge of soil management is important little is known about the nature and extent of such knowledge. It has long been recognised that farmers need new knowledge and skills to take on the demands of sustainable agriculture (Winter, 1997) but the suggestion has been that such knowledge is poorly developed due to the continued 'productivist modes of thinking' within the farming community (Curry, 1997; Pyrovetsi and Daoutopoulos, 1999; Wilson, 2001). Today concern remains about whether farmers have the right skills set to deliver the Government's goals for sustainable farming (Defra, 2004c; University of Reading, 2005).

Potentially, at least this knowledge and skills inadequacy extends to soil management. Scientific and policy reports highlighting the degradation of agricultural soils clearly imply that farmers are not managing them sustainably (Boardman et al., 2003; Environment Agency, 2004a, b). A poor knowledge base coupled with lack of experience of complex new technologies and practices has been highlighted as a constraint to more sustainable management of the soil. This has been demonstrated for other knowledge demanding practices, which provide soil and environmental benefits such as integrated farming systems (Morris and Winter, 1999), reduced tillage (Tebrugge and Bohrnsen, 2001; Davies and Finney, 2002; Coughenour, 2003), managing nutrients in manures (Smith et al., 2000, 2001) and organic farming (Burton et al., 1999). Concerns expressed in 1970 about farmers' lack of awareness of soil condition when cultivating (MAFF, 1970) appear to be still valid (Davies and Finney, 2002), while today farmers are reported to have insufficient understanding of the reasons and techniques for soil management (Central Science Laboratory, 2004).

However, this negative view of farmers' competence is questioned by the farming industry who claim that soil is managed sustainably (National Farmers' Union, 1994) and is in 'good heart', arguing that agricultural activities in the UK are synonymous with stewardship and conservation of the resources they rely upon (Ward, 1995). Such nurturing and stewardship are seen to be part of being a 'good' and knowledgeable farmer (McEachern, 1992; Burton, 2004). More generally, farmers regard their own knowledge in managing the environment as very important and often undervalued (Wilson, 1997; Harrison et al., 1998; Seymour et al., 1998). They place high value on their experience and use this as their primary source of knowledge for

[2]As part Common Agricultural Policy (CAP) reforms implemented in 2005, farmers will have to meet new 'cross-compliance' standards designed to protect the environment before they receive their Single Farm Payment.

management decisions (Contant, 1990; Fearne, 1991; Lyon, 1996;Tsouvalis et al., 2000). A growing literature in other countries (Romig et al., 1995; Walter et al., 1997; Bruyn and Abbey, 2003) also demonstrates that farmers have considerable knowledge of their own soil, are able to identify characteristics of soil quality and have developed a rich vocabulary to describe it (van der Ploeg, 1989; Romig et al., 1995; Liebig and Doran, 1999; Tsouvalis et al., 2000; van Rompaey et al., 2001; Curtis et al., 2003).

These different accounts demonstrate the range of views held about farmers' ability to hold and apply knowledge about soil and in part reflect the heterogeneous mix of farmers and their practices. Indeed the temporally and spatially diverse way in which farmers know and understand their farming systems has already been identified (Raedeke and Rikoon, 1997). It is clear however that there is a lack of evidence about farmers' knowledge of soil in England and about their capacity to implement more complex sustainable soil management practices that policy is now demanding.

4. Theoretical considerations

Conceptual approaches to understanding farmers' knowledge in relation to natural resource management have a broad base drawing both on behavioural (Napier et al., 1984; Lichtenberg and Zimmerman, 1999; Ryan et al., 2003) and cultural approaches (Carr and Tait, 1991; Long, 1992; McEachern, 1992; Harrison et al, 1998; Tsouvalis et al., 2000; Burton, 2004); as well as on perspectives that relate knowledge to social and experiential learning (Lyon, 1996; Röling and Wagemaker, 2000; Russell and Ison, 2000).

Developments in sustainable agriculture have brought new understandings of knowledge in the context of farming. A number of commentators have developed the notion that sustainable agriculture is knowledge intensive involving the adoption of technologies that require a high level of management skills, with an emphasis on observation, monitoring and judgement (Röling and Jiggins, 1994; Park et al., 1997; Tebrugge and Bohrnsen, 2001; Coughenour, 2003). Implementing these highly technical practices is thought to require some understanding of the underpinning scientific principles and physical processes (Vanclay and Lawrence, 1994; Pretty, 1995). At the same time, it is considered that sustainable systems and practices are highly dependent upon traditional local and 'ecosystem sensitive' knowledge, with general principles applied in a site-specific way (MacRae et al., 1989; Kloppenburg, 1991; Norgaard, 1984). Arguably, then, the knowledge farmers need for sustainable soil management must consist of both a technical understanding of the principles of soil management or 'scientific knowledge', as well as an intuitive, local or 'tacit knowledge', or at least an ability to interpret technical knowledge in a local context drawing on experience, through observation and monitoring.

The notion of knowledge comprising of scientific and tacit knowledge elements provides a useful framework for this research, where scientific knowledge[3] is understood to be universal, objective and decontextualised and tacit knowledge[4] implicit, indigenous and context dependent resulting from talents, experience and abilities. Scientific knowledge is itself, in part, comprised of what Lundvall and Johnson (1994) call 'know-why',[5] which is the knowledge of principles, rules and ideas of science and technology, it therefore concerns application. 'Know how'[6] (Lundvall and Johnson, 1994) and 'practical knowledge' (Thrift, 1985) share many features of tacit knowledge, being informal and learnt from experience of watching and doing. Characteristics of both scientific and tacit knowledge are highly relevant to soil management, the former because understanding and managing soil requires technical application of scientific principles and decontextualised generalities and the latter because soil is a spatially heterogeneous material 'rooted to place' and is intimately linked through cultivation to farm practices.

Tacit knowledge is fundamentally linked to direct experience and the practical, sensuous and personal skill that develops with attention to a specific place (Hassanein and Kloppenburg, 1995). It is frequently claimed that farmers have an intimate and intuitive tacit knowledge of the soil on their farms and a refined understanding of local spatial and temporal processes, gained through years of walking and cultivating the land (Winklerprins, 1999). Descriptions interpreted through the local environment have been explored and soil-quality assessments are firmly established by farmers in observational field experiences using senses of touch, taste, sight and smell; while soil physical properties become 'known' to farmers through in-field experiences (Romig et al., 1995; Walter et al., 1997; McCallister and Nowak 1999). Winklerprins (1999, p. 151) explores this idea defining 'local soil knowledge' simply as 'knowledge of soil properties and management possessed by people living in a particular environment for some period of time'. Such tacit or local knowledge is considered to be more relevant to sustainable agriculture (Kloppenburg, 1991; Murdoch and Clark, 1994).

[3]Scientific knowledge is also referred to as codified/expert/formal/standardised/codified and institutionally legitimate. It is described as explicit knowledge which can be systematised, written, stored and transferred (Norgaard, 1984).

[4]Tacit knowledge is also referred to as local/lay/indigenous/informal and traditional. It is thought to be 'strongly rooted in place… location specific' (Murdoch and Clark, 1994); and 'has to do with theories, beliefs, practices and technologies that all people have elaborated without direct inputs from the modern, formal and scientific establishment' (McCorkle, 1989).

[5]In Lundvall and Johnson's (1994) typology 'know-what' refers to knowledge about facts, which is largely codified. 'know-why' is the knowledge of principles, rules and ideas of science and technology.

[6]In Lundvall and Johnson's (1994) typology 'know-how' refers to skills, the capability to do something at practical level, as reflected in action and has a significant tacit component.

However, others claim that such local tacit knowledge is exaggerated and distorted and warn against mythologising it suggesting that it can often be nothing more than a set of improvisational capacities summoned by needs (Molnar et al., 1992; Richards, 1993). It has also been argued that indigenous soil knowledge, although still of great value in developing countries (Sillitoe, 1998), has no relevance to modern agriculture where farmers have come to rely heavily on scientific applications in agriculture (Morgan and Murdoch, 2000). For many, science is just as capable, or more capable of finding sustainable solutions (Molnar et al., 1992; Murdoch and Clark, 1994). Farmers in western countries have arguably been assimilating scientific information into their own knowledge for decades. They operate highly technical arable systems incorporating advanced technologies (Ward, 1995; Tsouvalis et al., 2000) and demonstrate adaptations, practical solutions and produce experimental knowledge, sometimes using scientific method (Wilson, 1997; Harrison et al., 1998; Tsouvalis et al., 2000).

Debates about the respective value of scientific and tacit forms of knowledge have lead many researchers to criticise this categorisation and argue that these knowledge forms are fundamentally complementary (Romig et al., 1995; Walter et al., 1997) and that knowledge is comprised of blends of all knowledge forms, that it is heterogeneously constituted (Long, 1992; Murdoch and Clark, 1994; Clark and Murdoch, 1997). Thus, knowledge that farmers hold, or need to hold, about soil and its management is arguably a mixture of both scientific and tacit knowledge as it needs to combine an understanding of new technologies with a new awareness and sensitivity of natural resource management (Röling and Jiggins, 1994; Pretty, 1995). The extent to which farmers are 'equipped' with this blend of knowledge for sustainable soil management is a central concern of this study.

5. Methodology

The qualitative and quantitative data on farmers' knowledge of soils, which are presented in this section, are derived from interviews with a selected group of farmers and agricultural advisors supplemented by an extensive postal questionnaire–survey of a range of advisors. Semi-structured interviews were undertaken with farmers and advisors in the context of two soil management initiatives in England. The UK SMI is an independent organisation which aims to address the problems of soil compaction, structural degradation and erosion by promoting the management of soil structure through appropriate cultivations and practices such as revised plough tillage, well managed reduced or conservation tillage[7] and removal of sub-/surface compaction. The Landcare Partnership, a project piloted by the Environment Agency in the Upper Hampshire Avon catchment in the south-west of England, promotes better farming practices (BFP) to control diffuse farm pollution. Untimely cultivation; maize with late cultivation and often excessive manure application; outdoor pigs; inappropriate manure applications; and lack of nutrient budgeting have all been identified as high risk practices particularly in the areas of more erosive Greensand and of Weald clay, which are exposed in some valleys in this mainly chalk catchment. The BFP proposed all aim to restrict run-off of sediment and loss of nutrients primarily through promoting appropriate cultivation techniques and the use of nutrient management plans to allow for manure nutrient content.[8] The catchment is one of 40 priority catchments recently designated for delivery of catchment sensitive farming by the government.

Selection and sampling of farmers for interviews within both projects was based on attendance at demonstration days and involvement in the project. For SMI all farmers (15 in total) who had attended two recent demonstration days were approached and of these eight agreed to be interviewed. These comprise arable farmers typically from large arable units (>500 acres) in the East and East Midlands regions. In addition, two farmers from the SMI board were interviewed. For Landcare all farmers (eight in total) who had recently attended a recent demonstration event were contacted and five agreed to be interviewed. Their farms are typically mixed with cereal, dairy cattle and sheep and some pig rearing. Two farmers who had provided demonstration sites for the project were also interviewed. Interview details are given in Table 1. The aim of farmer selection was not to extrapolate from a representative sample but to explore through detailed analysis and descriptive narrative the defining characteristics of farmer knowledge, for this reason there is no attempt to typologise the interviewees in the following analysis.

Selection and sampling of advisors for interviews was based on attendance at demonstration days, involvement and potential interaction with the project. For SMI all advisors listed as members (seven in total) and all those who had attended recent events (eight in total) were interviewed. In addition, a sample of 20 advisors

[7] Reduced tillage and conservation tillage refer to non-inversion tillage. The former also called 'minimum or minimal tillage' is 'shallow tillage' (<100 mm without inversion), the latter describes any non-inversion tillage, which leaves at least a third of the soil surface covered by crop residues; it includes direct drilling, shallow and deep tillage (Davies and Finney, 2002). Both are considered by some commentators (UK SMI, 2002) to improve soil structure, biota, soil organic matter and reduce erosion.

[8] To manage manures effectively farmers need to know the nutrient content of applied manures; apply manures evenly and at known rates; rapidly incorporate manures (where appropriate) or use an application technique that will minimise ammonia losses; apply manures in spring (where possible) to reduce nitrate leaching losses; take the nutrient supply from manures into account when calculating inorganic fertiliser additions; and ensure total N per year does not exceed 250 kg/ha.

Table 1
Characteristics of farmers and advisors interviewed and themes discussed

Landcare Partnership	Soil Management Initiative
Farmers interviewed ($N = 7$)	Farmers interviewed ($N = 10$)
1. Mainly arable farm on chalk with winter wheat and arable break crops (field beans and oil seed rape). Some sheep. 500 acres	1. Arable farm, Leicestershire. Combinable crops on loams. 440 acres
2. Arable farm on chalk with winter wheat, oil seed rape and field beans. 350 acres	2. Arable farm, Herts. Combinable crops on chalky soils, Boulder clay. 750 acres.
3. Mixed farm on chalk. Arable rotations with spring barley; suckler cows, indoor pigs. 200 acres	3. Arable farm, Cleveland. Combinable crops on light soils, loams. 520 acres.
4. Mixed farm bordering chalk and Pewsey Vale clay, suckler cows, sheep and arable. 250 acres	4. Arable farm, Rutland. Combinable crops on clayey soils. Reduced tillage. 800 acres
5. Dairy farm, with forage maize and grass silage. 180 acres. Clayey soils, some Greensand.	5. Arable farm, Leicestershire. Loamy soils. Cereals with break crops. 300 acres.
6. Arable farm on mainly chalk with winter wheat, break crops, also sheep. 550 acres	6. Arable farm Leicestershire. Clays with sandy loams. Cereals with break crops (peas or beans). 40 head suckler herd. 450 acres
7. Dairy farmer with some forage maize and grass silage. Clay with some Greensand. 150 acres	7. Arable farm, Kent. Cereals with break crops plus sheep. 1200 acres
	8. Arable farm, Leicestershire. Cereals with break crops (peas or beans). 660 acres
	9. Arable farm, Worcs. Clay loam over clay. Reduced tillage. 600 acres.
	10. Arable farm, Warwickshire. Some sheep. Reduced tillage on Evesham clays. 1000 acres.
Advisors interviewed ($N = 29$)	Advisors interviewed ($N = 35$)
10 partners (agronomists, RDS advisors, conservation advisors, etc.)	7 board members
4 event attendees	8 event attendees
15 local agronomists, merchants, consultants etc	20 agronomists/consultants
Themes discussed in interview	Themes discussed in interview
What is the nature and extent of farmer knowledge about:	What is the nature and extent of farmer knowledge about:
• managing manures and their nutrients	• soil structure and its examination cultivations/appropriate cultivations
• practices leading to erosion	• reduced tillage
• best management practices to prevent run-off	• causes of problems
• sources of information	• sources of information
• the aims of the project	• the aims of the project

were selected for their specialism in combinable crops, cultivation or soil management from the directories of the Association of Independent Crop consultants (AICC) (120 members) and British Institute of Agricultural Consultants (BIAC) (280 members). All Landcare advisor partners (ten in total) were interviewed as were those who had attended a recent demonstration event (four in total). In addition, of the 20 agronomists, seed merchants, farm management companies and consultants operating within the catchment identified using the AICC and BIAC directories, and the local telephone directory, 15 agreed to be interviewed.

The interviews with farmers and advisors were 'semi-structured' in that they were conversations informed by common themes relevant to the issues addressed by the initiatives rather than specific questions (listed in Table 1). Questions were open ended to allow deeper exploration of topics. The interview schedule provided structure and ensured the same issues were discussed in each case, it was supported with reference to relevant publications.[9]

Qualitative data from interviews were complemented by quantitative data from an extensive questionnaire of advisors. Questionnaires were sent to 304 individual advisors in five categories as follows: conservation advisors[10]; Rural Development Service agri-environment

[9]Publications referred to during the interviews for definitions, explanations and lists of best management practice included: Managing Livestock Manure booklets (ADAS et al., 2000); MANNER (MANure Nitrogen Evaluation Routine) a PC decision tool; Best Farming Practices: Profiting from a Good Environment (Environment Agency, 2001); A Guide to Managing Crop (UK SMI, 2002) and Visual Soil Assessment (UK SMI, 2005).

[10]Conservation advisors, mostly in the NGO 'Farming and Wildlife Advisory Group', are farm conservation specialists.

scheme advisors[11]; independent agronomists; ADAS[12] advisors and distributor (commercial) agronomists (FACTS).[13] Due to the different approaches to identifying potential advisors non-probability sampling was used to target certain sectors. In total 163 questionnaires were returned with an average response rate of 40% for the first four categories. It was not possible to estimate a response rate for the FACTS respondents.[14]

In total, 17 farmers and 64 advisors were interviewed within the two projects and 163 advisors surveyed nationally. Advisors were used as the main key informant due to restricted farm access following an outbreak of foot and mouth disease at the time of the study. Advisors are arguably well placed to provide a balanced and well informed opinion about farmers. They have a good understanding of, and regular contact with, farmers; they observe their activities frequently and have one of the closest relationships with farmers and their farms (Angell et al., 1997; Ingram and Morris, 2007). Each advisor can also draw from their experience of advising a number of farmers, typically up to 20 farmers each within a relatively wide geographical area, enabling them to develop a broad impression of the farming community. This contrasts to individual farmers whose experiences are narrower, being tied to a particular farm environment and business. Advisors also arguably provide a more objective view of what is happening on-farm as farmers may be reluctant to own up to poor knowledge and practice. Interviewing advisors as well as farmers also provides different accounts and interpretations which can triangulate and complement each other. The nation-wide survey results add to the multiple sources of evidence and assist in the triangulation of qualitative material.

In the sections presenting the empirical material, the scientific knowledge element of farmers' soil knowledge will be examined, that is, their understanding of the requirements and principles of soil best management practice. The extent of farmers' engagement with using nutrients in manures from interviews in the Landcare initiative gives a more detailed measure of their scientific understanding. Insights into the tacit element of farmer soil knowledge are gained by examining farmer and advisor views of about farmer local knowledge of their fields and their experience, skills and competence in practical soil management. Detailed consideration is given to tacit knowledge of cultivation practices drawing on the interviews from the SMI initiative. Although results are presented separately below it is understood that the two knowledge forms are in reality intimately linked.

[11]Defra's Rural Development Service project officers manage and administer government funded agri-environment schemes.

[12]ADAS, formerly the government advisory service, was privatised in 1997 and now operates as a consultancy, although it still undertakes environmental protection dissemination work for the government.

[13]FACTS provides a national training syllabus and accreditation for arable advisors and farmers. Training includes soil and nutrient management.

[14]Questionnaires were e-mailed on behalf of the author through the FACTS organisation to an unknown number and sector of the membership.

6. Farmers' scientific knowledge of soil

6.1. Farmers' scientific knowledge of soil in relation to soil best management practice

The interviews revealed that there is a range of competencies amongst farmers in relation to scientific knowledge. Many of those interviewed are professional, highly skilled, intelligent graduates, with one farmer FACTS and BASIS[15] trained, and three farmers holding responsible positions within national agricultural organisations. They are competent in preparing their own fertiliser recommendations, using and interpreting research station results and can grasp difficult issues such as the soil nitrogen dynamics. Some advisors' views confirm this, they consider farmers to be technically knowledgeable and as having an understanding of principles that underpin best management practice. As one independent agronomist (A) said '*Most farmers I deal with are aware of good husbandry techniques and go to great lengths to keep soil in good order*'. For them there is clear evidence of good techniques being used. However, this view is not shared by all advisors, as one distributor agronomist's (RP) remark demonstrates: '*They are appalling at soil husbandry, they take their soils for granted and don't necessarily look after them as much as they should do*'. This lack of consensus among advisors about this aspect of farmers' knowledge is reflected in the questionnaire responses with only an average of 40% of all advisors agreeing that farmers are technically well informed about soil management (Table 2). Views among advisors and farmers explored in interviews suggest that, although in a very broad sense farmers practice good husbandry, it is the depth of farmers' technical understanding of soil management that is limited, as one agri-environment scheme advisor (F) explained:

> They are aware of gross errors of management but not aware of more subtle things they can do. There's awareness that the problem exists on one level but not yet sufficient awareness of issues leading to that.

Advisors claim that farmers, although aware of problems, do not necessarily tie them down to their own practise. This was borne out in the Landcare interviews where, although farmers acknowledged that certain management practices lead to increased run-off, none accepted that their individual practices were responsible, attributing the problem instead to other sources such as highways or to their neighbours. Even when the run-off was traced to their own farm, they blamed extreme weather events rather than their own practices. A farmer's (P) remark epitomises

[15]BASIS is an Independent Registration Scheme for the Pesticide Industry and oversees FACTS.

this view, *'I've seen water run of from my fields and it's brown flowing straight into the river. Nothing could be done about it, we're on free draining land if it's washing off here it's washing off everywhere'.*

Soil management is seen by advisors as quite a complicated issue which a number of farmers do not fully understand, particularly with more demanding techniques being introduced. They consider that, although a very important issue, soil management is frequently ignored due to lack of knowledge by farmer. Indeed the majority of advisors (average 74%) responding to the questionnaire thought that lack of knowledge and skills about soil management options was important in explaining farmers' failure to use more sustainable practices (Table 3). Without the knowledge or acquired skill farmers are thought by some advisors to give up and drop back to more familiar intensive production methods:

> I would entirely agree that for more environmentally favourable practices for soil you need more knowledge and skill and if you don't have that or don't develop it quickly then you will give it up and drop back to intensive production methods which have been tried and tested (Farm Manager/Agronomist V).

Some farmers accept they are ignorant in this respect as one (Farmer B) remarked *'I know everything about machinery, a little about crops, but very little about the soil. Even though I work the land, I still don't know enough about soil'*. For many realisation that they lack such understanding was part of their motivation for attending the demonstration days provided by the projects.

6.2. Farmers' scientific knowledge of soil in relation to managing nutrients in manure

A large part of the promotion of soil best management practice in the Landcare initiative has been aimed at using manures as part of the nutrient budget for the farm. This requires some understanding the principles of nutrient dynamics in the soil and being able to estimate amounts, and the nutrient content, of manure so that artificial fertiliser rates can be adjusted accordingly. Interviews with farmers and advisors working in the catchment revealed that this is still very challenging and involves unfamiliar skills for farmers who often lack the experience of high fertility situations. As one independent agronomist (EB) remarked *'What we're doing, advisors and farmers alike, we're all scratching our heads asking how much should we allow for that [manure]?'* A large number of advisors (some 19 out of the 29 interviewed) within the catchment see ignorance of the value of manures as an obstacle claiming that the majority of farmers, dairy farmers in particular, still regard it as a waste product that has to be disposed of rather than as a valuable source of nutrients. They argue that the farmers they work with are often unable or reluctant to measure the amounts of manure used and that manure applications are roughly estimated, spreaders are uneven and manures are often mixed from different animal

Table 2
Advisors' response to the question: to what extent do you agree that most farmers you advise are technically well informed about soil management?

	% of respondents within each advisor category					% average
	Conservation	DEFRA RDS	Independent agronomists	Distributor agronomists	ADAS	
Agree	15	41	44	54	48	40
Neither agree nor disagree	50	50	29	23	39	38
Disagree	35	9	26	23	13	22
No. of valid respondents	32	22	71	13	23	Total 161

Table 3
Advisors' response to the question: How important is lack of knowledge and skills about soil management options as a factor in explaining poor uptake of soil best management practices?

	% of respondents within each advisor category					% average
	Conservation	DEFRA RDS	Independent agronomists	Distributor agronomists	ADAS	
Not important	3	14	13	9	2	8
Neither important nor not important	9	24	24	7	26	18
Important	88	62	63	84	72	74
No. of valid respondents	32	21	70	13	23	Total 159

sources, an agri-environment scheme advisor (F) supports this view:

> They don't know what's in it in many cases, haven't a clue what the analysis is, they are reluctant to do analyses whether slurry or solid, they haven't got a clue as to how much they put on. The whole thing is very hit and miss.

Four farmers interviewed in the catchment confirmed this view accepting that they still view manure as a waste product, rather than an asset, this was particularly the case with dairy and small mixed farms where maize becomes a dumping ground for manure. They also tend to regard nutrient budgeting as a complex process beyond their capabilities. However, progressive and bigger farmers were described by some advisors interviewed in the catchment as more disciplined about accounting for manure, measuring its value as part of their nutritional programme and using more sophisticated spreading machinery. They are understood to be gaining knowledge about the nutrient value of manure, and are increasingly using or, if arable-only, buying in manure, sewage sludge or poultry manure, and have seen benefits to soil structure as well as realised the economic sense of the practice. Three of the farmers (two from mixed farms of about 500 acres) interviewed in the catchment had started to incorporate manures into their nutrient budgeting; one described the benefits:

> Yes definitely we estimate how many tonnes of manure and where it went on. Where I put a lot of manure on last year, I put a lot less N top dressing on, so its helping me again and saving me money. I'm sure manure is all good for the soil structure and composition (Farmer MF).

However, although some farmers are beginning to utilise manure nutrients, an agri-environment scheme advisor for the area suggested that nutrient budgeting is a 'closed book' to 95% of farmers with only the larger arable farmers taking an interest, with perhaps only 5% of farmers and 15% of land being subject to a nutrient balance.

These results suggest that the extent of farmers' scientific knowledge concerning soil management is variable and in some cases not fully developed. Many farmers and advisors interviewed and surveyed point to good husbandry and good technical knowledge as evidence of knowledge about soil management. However, interviews revealed that the depth of this knowledge about soil can be limited. Although advisors suggest that there is 'considerable understanding of the soil', they consider that farmers remain ignorant about the more subtle things, in particular farmers do not always have sufficient knowledge to make the link between certain practices and their consequences suggesting lack of understanding of underpinning principles or 'know-why'. This might explain why only 40% of advisors surveyed (Table 2) agreed that farmers are technically well informed about soil management. In terms of having or acquiring the scientific knowledge needed to incorporate nutrient budgeting into managing manures the larger more progressive farmers appear most likely to be accounting for manures, since it is thought they can afford advice and analysis and are more likely to be experienced, from their arable practices, in nutrient budgeting. Advisors consider that for many farmers this is still a challenge, that they 'haven't got a clue'. Farmers tend to agree, particularly smaller dairy and mixed farmers who are constrained by the size of arable land on which they can spread manure, the condition of the often heavier soil, and their own lack of experience of nutrient budgeting. Given these constraints they are less likely to seek out and develop knowledge on the value of manure.

7. Farmers' tacit knowledge of soil

7.1. Farmers' tacit knowledge of soil in relation to soil best management practice

All the farmers interviewed had some knowledge of soils in terms of their spatial heterogeneity, physical properties and response under different cultivation practices. In support of this an average of advisors 65% agreed that farmers had a good knowledge and understanding of their soils (Table 4). All farmers interviewed appear to have some knowledge of how the soils differ in their fields in terms of depth, texture and drainage, although the level of detail they used varied. Farmers described variation using either a local term, a textural (e.g. sandy, silty) or geological reference, and in some cases a formal classification. While some used vague terms such as 'thick bits' and 'thin bits' others, notably those implementing reduced tillage, knew their fields intimately, even to the point of contesting the soil series map, and some held extensive records of yield for different soil types.

All farmers interviewed also appear to have developed a practical 'working knowledge' of their soils through regular cultivation which enables them to judge its structure and condition. Accordingly, soils were often described by farmers in terms of their ease of cultivation, with terms such as 'light and easy' or 'heavy' used. Some drew relationships between soil texture, structure and soil moisture, distinguishing heavier soils as having better natural structure and better water retention but being more difficult to plough. One farmer who practices reduced tillage emphasised the significance of the weather to these properties:

> The weather is so important, this land, my father taught me, he said 'the one thing that works this land is weather and you will never force it', and he was so right. Because it is heavy land, you have to cultivate dry or drill dry, you can't do both wet and expect to get a good crop (Farmer Y).

In describing farmer knowledge of soil, terms and phrases such as 'intuitive knowledge', 'being in tune with the soil', or 'understanding the soils', which express a less tangible form of knowledge, were commonly used by farmers and advisors. This feel for soil is linked to their central function on the farm, as one advisor remarked '*They [farmers] don't know necessarily about soils but they have an intuition about soils because that's their livelihood*'. Associated with this is the attachment some farmers have for soil, as one advisor commenting on farmers' response to soil erosion said '*It's like loosing their birthright, farmers hate to see the soil running off.*' Those more knowledgeable farmers described experience, long-term observation and record keeping as contributing towards this intuitive or tacit knowledge.

Some advisors and farmers however dispute that farmers have any such attachment or intuitive knowledge of soil. Farmers who consider themselves to be good soil managers are very critical of other farmers, for example, one (Farmer L) said '*They don't have any intuitive feel, no they don't. I don't think they have the slightest interest, not the ones I know anyway!*' A number of advisors also claim that continued poor practice and abuse of soil is evidence of a lack of such knowledge. In accordance with these sentiments, but in contrast to previous comments in favour of farmers' knowledge of their soils, an average of 54% advisors surveyed agreed that most farmers they advised are not concerned with good agricultural practice for soil management (Table 5) supporting one advisor's view that '*Farmers battle on regardless, not really working with soil and the conditions*'. Also an average 79% consider that soil degradation is a problem in English agriculture (Table 6). In support of this more than 55% advisors on average had observed severe compaction, water erosion, capping and poor drainage in the last two years which they attributed to poor soil management (Fig. 1), although most respondents stressed the localised and irregular nature of these incidences with their occurrence often coincident with very

Table 4
Advisors' response to the question: To what extent do you agree that most farmers you advise have a good knowledge and understanding of soils on their farm?

	% of respondents within each advisor category					% average
	Conservation	DEFRA RDS	Independent agronomists	Distributor agronomists	ADAS	
Agree	41	73	63	77	70	65
Neither agree nor disagree	50	27	22	23	30	30
Disagree	9	0	15	0	0	5
No. of valid respondents	32	22	71	13	23	Total 161

Table 5
Advisors' response to the question: to what extent do you agree that most farmers you advise are not concerned with good agricultural practice for soil management

	% of respondents within each advisor category					% average
	Conservation	DEFRA RDS	Independent agronomists	Distributor agronomists	ADAS	
Agree	28	59	70	69	44	54
Neither agree nor disagree	50	41	23	23	52	38
Disagree	22	0	7	8	4	8
No. of valid respondents	32	22	71	13	23	Total 161

Table 6
Advisors' response to the question: to what extent do you think soil degradation is a problem in English agriculture?

	% of respondents within each advisor category					% average
	Conservation	DEFRA RDS	Independent agronomists	Distributor agronomists	ADAS	
No problem exists	0	0	18	0	4	4
Do not know	0	17	17	25	26	17
Yes problem exists	100	83	65	75	70	79
No. of valid respondents	32	23	71	12	23	Total 161

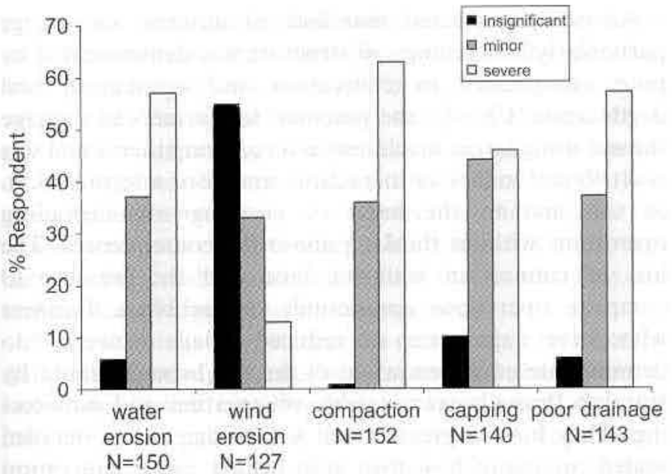


Fig. 1. Advisors response to the question: indicate the severity of any incidences of degradation, which you have observed in the course of your work over the last 2 years which can be attributed to inappropriate land use or poor soil management?

wet conditions. Advisor observations and opinions will clearly be influenced by the type of client, his farming system and soil type, however these figures do suggest that a large number of advisors drawn from a national sample have concerns about soil management, and do concur with interviewee comments.

7.2. Farmers' tacit knowledge of soil in relation to cultivation

The nature and extent of tacit knowledge of soil can be explored through farmers' cultivation since these operations bring the farmer into direct sensual contact, through sight, feel and smell, with the soil. Farmers and advisors interviewed within the SMI project provide an interesting view on the nature of this knowledge. There is universal agreement amongst them that cultivation is a very skilled activity, with important decisions needing to be made about the timing, and the interval between subsequent successive cultivations, as well as the choice of machine. Poor cultivation decisions and practices are regarded by the majority of advisors as the main reason for soil structural degradation including compaction and surface capping which lead to erosion. This is attributed to lack of thought, as one advisor said *'My view is the level of competence in terms of cultivation is not necessarily that great because farmers don't think about what they are doing'*. More specifically farmers' lack of observation and examination of soil is blamed. Although there is thought to be an enormous amount of understanding of soil amongst farmers the main problem, in the opinion of those advisors involved with SMI, is that *'they do not know how to examine their soil to determine how much cultivation is required'*. This lack of inspection is compounded by the pressures to get crops drilled, even when weather conditions are unsuitable, coupled with the availability of very powerful machinery, which compact the soil and cultivate deeper than necessary:

> It's one almighty rush. There is not enough kicking plods, it's more crash bang wallop and getting in before the next job rather than thinking things through. They only look at top, OK they will scuff in with their boot, the surface where the seed is going, to make sure it's in a good condition, but what happens beyond there… they have very rarely gone out and put a spade in the ground and dug a hole to see what's happening (Farmer L).

For those that do have the time or the interest to inspect soils, advisors consider that a further problem arises because farmers are often ill equipped to interpret what they see. The farmers interviewed who had attended the SMI demonstration days agreed that they had limited experience in inspecting the soil and benefited in particular from this element of the demonstration event. Although most of the time farmers are seen to get cultivations right, advisors consider that when things go wrong they can have quite a visible impact both agronomically and environmentally:

> In the days with horse and plough they knew it intimately and couldn't do much damage, now with huge machinery they can do a lot of damage very quickly before they have gained the experience. Even the most in tune farmers can make those mistakes on a big scale (Independent agronomist J).

This is verified by the advisors' comments and survey respondents' observations of localised compaction and water erosion (Fig. 1). Farmers themselves recognise the problems brought by larger machinery, and the pressures soils are under, as one (Farmer P) remarked *'We have huge tractors now and see the damage we are actually doing to the soils'*. There was a suggestion from a number of advisors (nine out of 31) that farmers are more removed from the soil than they used to be. With farms getting bigger, labour a constraint, and the demands of paper work, farmers have less opportunity to visit their fields, many leaving cultivations to their tractor drivers and field walking to their advisors:

> No farmers don't know their soils intimately, they're hopeless. The generation of farmer who knew their soil are those that spent a lot of time in their fields walking up and down but they don't now, they haven't got time for that nor the interest. (Independent agronomist DG).

The inference is that use of large machinery is threatening their knowledge of the soil since this technology removes their physical and sensual contact with the soil, obscuring any visible signs of problems with the subsoil, which may have been detected earlier by someone on foot. However, some farmers defend themselves against such claims, suggesting that it is only on larger farms with hired labour where this is a problem, as one (Farmer M) remarked *'It all depends how much you leave to your* *agronomist. A lot of farmers, myself included, still like to get out there and walk across the field'.*

In contrast to the farmers described above who fail to examine and understand their soils, three farmers interviewed who have been practicing reduced tillage were notable exceptions. Reduced tillage systems require close and regular observation of the surface and subsoil condition and demands a new understanding and awareness of soil. As such these farmers demonstrate an appreciation and an intimate knowledge of their soils. One described his 'awakening' in understanding the soil through inspection and moving away from performing operations through habit:

> Speaking personally ten years ago I never even thought about it. I didn't look at the soil, I didn't dig holes, I didn't look for earth worms. We would subsoil every year before we planted oil seed rape and we wasted thousands of pounds doing that and probably did more harm than good because that was the thing you did, it was habit (Farmer L).

These represent an interesting cohort of farmers who are both knowledgeable about their soils and enthusiastic about the methods they have discovered to manage them sustainably. These farmers tend to be characterised not by type of farm enterprise or farm size but by their commitment and readiness to observe, experiment and learn, which enables them to build up invaluable knowledge over time.

This analysis presents a complex picture of farmers' tacit soil knowledge. Although the advisors surveyed largely agree that farmers are knowledgeable about their own soils, the majority also agree that farmers are not concerned about soil. This might suggest that knowledge of soil does not automatically qualify farmers to be interested in its protection or husbandry or that other factors prevent knowledgeable farmers from demonstrating concern. Advisor interviews reflect this division of views with some considering that farmers do have tacit knowledge, describing them as being 'in tune' or having 'intuitive knowledge' of soil, referring to it as their 'livelihood' or their 'birthright' and others describing farmers as disinterested, not 'in tune with', 'not working with the soil' and not knowing their soils intimately. Farmers themselves tend to acknowledge that, when it comes to soil management, as with any other aspect of farming, there are good farmers and bad farmers. Evidence from farmer interviews reveal that whilst all farmers could offer a description of their soils, these were at different levels of detail, some making very general and sometimes vague comments about variable texture and 'workability', others demonstrating an intimate knowledge. Many acknowledged that, although they knew the spatial patterns of their soils, they lacked a full understanding of soil particularly in relation to its examination and cultivation and accepted that problems occurred, none however accepted that they were 'hopeless' as suggested by some advisors.

Advisors considered that lack of intuitive knowledge particularly concerning soil structure was demonstrated by poor competence in cultivation and consequent soil degradation. Clearly, the potential for farmers to damage the soil using larger machinery is a recurring theme and this is attributed to lack of inspection, and poor interpretation, of soil and to the habit of carrying out degrading operations without thinking about the consequences. The loss of connection with the land and the pressure to complete operations compounds the problem. Farmers who have experience of reduced tillage however do demonstrate an appreciation of the soil brought about by learning through examination, observation and monitoring. They have a greater tacit knowledge and a detailed understanding of how their soils behave under cultivation and/or changing weather. These farmers are highly critical of other farmers' lack of interest and degrading practices.

8. Discussion

This paper describes the views advisors and farmers hold about the nature and extent of farmers' soil knowledge in England. Although the research exposes some conflicting opinions, both within and between the farmer and advisor communities, there is a general consensus that, although farmers are largely knowledgeable, many appear to lack the in-depth scientific and tacit knowledge necessary for carrying out more complex sustainable soil management practices.

The survey results provided a general picture of the extent of advisors' views about farmer soil knowledge while the advisor and farmer interview data elaborate on these views in more depth, exploring the nature of this knowledge in the context of the two projects. The advisors were a valuable and informed source of comment for this research. Their range of views reflect the diversity of situations that they deal with, in terms of their clients, their systems, environment and soil types. They also reflect their own knowledge, standards and interpretations of what constitutes degradation and good soil management. However, correspondence between the views of the different types of advisors who responded to the survey suggest that advisors have a largely unified view about farmers' knowledge. This is with the exception of conservation advisors who are less involved in practical farming matters on the farm and therefore arguably less aware.

Farmers interviewed as part of the SMI initiative represent predominantly larger arable farmers from the eastern parts of the country who are seeking to understand soil management in the context of reducing crop establishment costs through reduced tillage techniques. The Landcare farmers from the south-west have a more diverse farming background and are looking for practices that will lead to both reduced sediment and nutrient loss in response to pressure from the Environment Agency, fisheries and the local public. As such their interests differ, however some commonalties have been revealed. Farmers from both projects are knowledgeable about their own soils however they acknowledge that they are challenged by practices such as managing manures and appropriate cultivations which they recognise as very skilled activities. It is particularly interesting to note that even within a sample of self-selected farmers drawn from those who had attended demonstrations, i.e. motivated individuals, the majority in both projects accept that soil is still a resource they do fully understand and indeed that is why they are seeking knowledge about its management.

The research has shown that the relationship between the different knowledge forms is complex and that scientific knowledge relies considerably on tacit knowledge for interpretation particularly at the level of local farm implementation. Cultivation practices involve highly skilled operations using technical and scientific knowledge but this knowledge needs to be combined with local knowledge of soil and weather conditions to be effective. Similarly, experience is central to tacit knowledge and yet it contributes to farmers being able to confidently apply scientific knowledge as with nutrient budgeting. Results suggest that only by having a full complement of scientific and tacit can farmers be fully 'equipped' to implement soil best management practice. However, for both forms of knowledge there is a sense that farmers have some knowledge but not enough, they are 'in tune but equally …there is considerable ignorance' or 'they are aware of gross errors of management but not aware of more subtle things they can do' or 'there is an enormous amount of understanding about the soil but where farmers fall down is they would not examine it'. This highlights the fact that in many cases farmers do not have sufficient in-depth knowledge that sustainable soil management demands. This lack of knowledge is considered by the majority of advisors surveyed to be important in explaining poor uptake of best soil management practices.

The potential for damage done by larger machinery and poor cultivation decisions is a recurring theme. Back in the 1970s the Strutt report remarked that farmers' lack of knowledge of the composition of the subsoil was leading to mistakes and to unjustified risk taking' (MAFF 1970), it appears that this continues to be the case today as machinery capable of damaging the soil structure is used without sufficient thought given to the subsoil condition. Because of the risks even the most 'in tune' farmers are thought to be 'very capable of making big mistakes'.

Failure to examine the condition of the soil prior to cultivation is seen as a major underlying cause of poor cultivation decisions. Many farmers appear to no longer have the time to walk their fields or inspect the soil and inability to interpret what they see compounds the problem. It has been argued that, before the advent of the productivist era and associated technological evolution, farmers usually had an intimate knowledge of their land holding, its fertility and composition through practices of rotation and ploughing, with local knowledge attuned to the rhythms of nature and tied to the farm and the local ecosystem, of which soil is a central component (Morgan and Murdoch, 2000). The suggestion is that during the intensification of UK agriculture in the period 1950–1980 this farmers' local knowledge base was replaced by specialised and commodified agricultural inputs. This research suggests that the farmers' soil knowledge base may also have been eroded throughout this transition; that the use of hi-tech machinery, compounded by time and labour constraints on the farm, has displaced the traditional relationship farmers had with their soil landscape. Previously, 'in the day with horse and plough they knew the soil intimately' but 'the generation of farmers who knew their soil and spent a lot of time in their fields walking up and down' has been replaced by those who rarely walk the farm and have lost connection with the land. This is with the exception of those SMI farmers who, through the practices of reduced tillage, are rediscovering the soil; they are arguably forgetting the old habits of cultivating without inspection and relearning how to examine and interpret soil conditions, processes that Morgan and Murdoch (2000) claim are necessary to acquire knowledge for sustainable farming. This analysis of course is based on advisors views and, although accepted by some farmers interviewed, was contested by others who claim to walk their fields regularly. However the prospects for UK agriculture, with continued reduction in farm income support, declining crop prices and consequent re-structuring and adjustment, mean that farmers will have less resources, time and labour to address soil management matters. They are also more likely, in striving for efficient systems, to use larger machines or employ contractors to undertake their operations, thereby diminishing their connection with the land even further (GFA-RACE and IEEP, 2003).

In conclusion, there is evidence of some farmers being well equipped to carry out sustainable soil management; however, some areas will have to be significantly enhanced and standardised to meet the new policy challenges, specifically improvement of the 'know-why' element of scientific knowledge, relating poor management practice to consequences, and improvement of the 'know-how' element of tacit knowledge that enables farmers to examine soil and interpret what they see. Achieving these improvements will require different approaches. Enhancing scientific knowledge requires explanation through training and demonstration, as provided by the new BASIS certificate in soil and water management offered to farmers (Farmers Weekly, 2004b), publications (Defra, 2005), and workshops and demonstration events, such as those offered by SMI and Landcare. Enhancing tacit knowledge requires farmers to learn through practice. Practical demonstrations are a popular method for showing farmers how to examine and interpret soil condition; however, this research has revealed the value of experience gained on-farm. Although this has to be an individual endeavour it could be facilitated by competent practitioners. This research has shown that individuals within the farming and advisory communities are well informed, concerned and knowledgeable about soil and as such could support on-farm learning either through farmer discussion groups or one to one advice. Ultimately, farmers need to be encouraged to attend training courses and demonstrations, to seek assistance from advisors and to learn for themselves. Policy demands will provide an impetus for some while the cost savings offered by practices such as reduced tillage and managing manure will provide incentives for others who are searching for ways of surviving in a competitive industry (UK SMI, 2002; Defra, 2004d; Environment Agency, 2005). Enthusiastic farmers, such as those practising reduced tillage, also have a role to play as 'champions' or 'influencers' of good soil management within the farming community.

It has long been recognised that farmers need new skills to take on the new demands of sustainable agriculture but little has been known about soil, this research has gone some way to address this gap in the English context. It will be important to build on this research in the future by consulting a larger sample of farmers to elaborate further the extent and nature of their knowledge about soil and to inform policy by identifying constraints and opportunities for improving this knowledge.

References

ADAS, IGER and SRI, 2000. Managing Livestock Manures. Booklet 1: making better use of livestock manures on arable land. Booklet 2: making better use of livestock manures on grassland. Funded by MAFF.

Angell, B., Francis, J., Chalmers, A., Flint, C., 1997. Agriculture and the rural economy information and advice needs. Report to MAFF by ADAS.

Baldock, D., Mitchell, K., 1995. The implications for soils of the CAP. A report to the Royal Commission of Environmental Protection by the Institute of European Environmental Policy, London.

Boardman, J., 1990. Soil erosion in Britain: costs, attitudes and policies. Social Audit Paper No. 1. Education Network for Environment and Development, University of Sussex.

Boardman, J., Poesen, J., Evans, R., 2003. Socio-economic factors in soil erosion and conservation. Environmental Science and Policy 6, 1–6.

Bruyn, L.A.L.D., Abbey, J.A., 2003. Characterisation of farmers' soil sense and the implications for on-farm monitoring of soil health. Australian Journal of Experimental Agriculture 43 (3), 285–305.

Burton, R.J.F., 2004. Seeing through the 'Good Farmer's' eyes: towards developing an understanding of the social symbolic value of 'Productivist' behaviour. Sociologia Ruralis 44 (2), 195–215.

Burton, M., Rigby, D., Young, T., 1999. Analysis of the determinants of adoption of organic horticultural techniques in the UK. Journal of Agricultural Economics 50 (1), 47–63.

Carr, S., Tait, J., 1991. Differences in the attitudes of farmers and conservationists and their implications. Journal of Environmental Management 32, 281–294.

Central Science Laboratory, 2004. An overview of ELS pilot scheme outcomes and implications for a national scheme. Report to Defra.

Clark, J., Murdoch, J., 1997. Local knowledge and the precarious extension of scientific networks: a reflection on three case studies. Sociologia Ruralis 37 (1), 38–60.

Commission of the European Communities, 2002. Towards a thematic strategy for soil protection. Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions. COM (2002) 179, European Parliament 2003.

Contant, C.K., 1990. Providing information to farmers for groundwater quality protection. Journal of Soil and Water Conservation 45 (2), 314–317.

Coughenour, C.M., 2003. Innovating conservation agriculture: the case of no-till cropping. Rural Sociology 68 (2), 278–304.

Curry, N., 1997. Providing new environmental skills for British Farmers. Journal of Environmental Management 50, 211–222.

Curtis, A., Byron, I., McDonald, S., 2003. Integrating spatially referenced social and biophysical data to explore landholder responses to dryland salinity in Australia. Journal of Environmental Management 68 (4), 397–407.

Davies, B., Finney, J.B., 2002. Reduced cultivations for cereals: research, development and advisory needs under changing economic circumstances. HGCA Research Review No. 48.

Defra, 2002a. The Strategy for Sustainable Farming and Food. Facing the Future. Defra Publications, London.

Defra, 2002b. Farming and Food's Contribution to Sustainable Development—Economic and Statistical Analysis. Defra Publications, London.

Defra, 2004a. First Soil Action Plan for England 2004-06. DEFRA soil activities 〈http://www.defra.gov.uk/environment/landliability/soil/index.htm#Research〉, June.

Defra, 2004b. Developing Measures to Promote Catchment-Sensitive Farming. A joint DEFRA-HM Treasury Consultation. Defra Publications, London.

Defra, 2004c. Learning, Skills and Knowledge Review Final Report 〈http://www.defra.gov.uk/rural/pdfs/lsk/advisor_cpd_project_-plan.pdf〉.

Defra, 2004d. Farm Practice Survey 2004. National Statistics. Defra Publications, London.

Defra, 2005. Cross Compliance Guidance Notes for Soil Management. Defra Publications, London.

DETR, 1998. Draft National Soil Strategy. DETR Publications, London.

DETR, 2001. Draft Soil Strategy for England—a Consultation Paper. DETR Publications, London.

Environment Agency, 2001. Best Farming Practices: Profiting from a Good Environment R&D Publication 23. Environment Agency, Bristol.

Environment Agency, 2004a. Soil, the Hidden Resource. Towards an Environment Agency Strategy for Soil Protection, Management and Restoration—A Consultation Document. Environment Agency, Bristol.

Environment Agency, 2004b. The State of Soils in England and Wales, Bristol.

Environment Agency, 2005. Spotlight on Business. Environment Agency, Bristol.

Farmers Weekly, 2004a. Soil help needed to keep SFP. Farmers Weekly 13 May 2004. Farmers Weekly Interactive. 〈http://www.fwi.co.uk/live/www.fwi.co.uk/live/〉, June.

Farmers Weekly, 2004b. Certificate to suss soil. Farmers Weekly 15 June 2004. Farmers Weekly Interactive. 〈http://www.fwi.co.uk/live/〉, June.

Fearne, A., 1991. Agricultural information: a farmer's point of view. In: Kuiper, D., Röling, N.G. (Eds.), European Seminar on Knowledge Management and Information Technology.

GFA-RACE and IEEP, 2003. The potential environmental impacts of the CAP Reform agreement. Final Report for Defra. Issue: 1.0. Report No: GRP-P-172. GFA-RACE, Cirencester, UK.

Harrison, C.M., Burgess, J., Clark, J., 1998. Discounted knowledges: farmers' and residents' understandings of nature conservation goals and policies. Journal of Environmental Management 54, 305–320.

Hassanein, N., Kloppenburg Jr., J.R., 1995. Where the grass grows again: knowledge exchange in the sustainable agriculture movement. Rural Sociology 60 (4), 721–740.

Ingram, J., Morris, C., 2007. The knowledge challenge within the transition towards sustainable soil management: an analysis of agricultural advisors in England. Land Use Policy 24, 100–117.

Joint Nature Conservation Council, 2002. Environmental Effects of the CAP and Possible Mitigation Measures. Report to DEFRA.

Kloppenburg Jr., J., 1991. Social theory and the de/reconstruction of agricultural science: local knowledge for and alternative agriculture. Rural Sociology 56 (4), 519–548.

Lichtenberg, E., Zimmerman, R., 1999. Information and farmers attitudes about pesticides, water quality and related environmental effects. Agriculture, Ecosystem and Environment 73 (3), 227–236.

Liebig, M.A., Doran, J.W., 1999. Evaluation of farmers' perceptions of soil quality indicators. American Journal of Alternative Agriculture 14 (1), 11–21.

Long, N., 1992. From paradigm lost to paradigm regained. The case of actor-oriented sociology of development. In: Long, N., Long, A. (Eds.), Battlefields of Knowledge: the Interlocking Theory and Practice of Social Research and Development. Routledge, London, pp. 16–43.

Lowe, P., Clark, J., Seymour, S., Ward, N., 1997. Moralising the Environment: Countryside Changes, Farming and Pollution. UCL Press, London.

Lundvall, B.-Å., Johnson, B., 1994. The learning economy. Journal of Industry Studies 1 (2), 23–42.

Lyon, F., 1996. How farmers research and learn: the case of arable farmers of East Anglia, UK. Agriculture and Human Values 13 (4), 39–47.

MacRae, R.J., Hill, S.B., Hening, J., Mehuys, G.R., 1989. Agricultural science and sustainable agriculture: a review of the existing scientific barriers to sustainable food production and potential solutions. Biological Agriculture and Horticulture 6, 173–219.

MAFF, 1970. Modern farming and the soil. Report of the Agricultural Advisory Council on Soil Structure and Soil Fertility. HMSO, London.

MAFF, 1998. Code of Good Agricultural Practice for the Protection of Soil. MAFF Publications, London.

MAFF, 1999a. Controlling Soil Erosion by Water: A Manual for the Assessment of Agricultural Land at Risk of Water Erosion on Lowland England. MAFF Publications, London.

MAFF, 1999b. Controlling Soil Erosion by Water: A Field Guide for An Erosion Risk Assessment for Farmers and Consultants. MAFF Publications, London.

McCallister, R., Nowak, P., 1999. Whole soil knowledge and management: a foundation of soil quality. In: Lal, R. (Ed.), Soil Quality and Soil Erosion. Soil Water Conservation Society, Ankeny, IA, pp. 173–193.

McCorkle, C.M., 1989. Towards a knowledge of local knowledge and its importance for agricultural R&D. Agriculture and Human Values 6 (3), 4–12.

McEachern, C., 1992. Farmers and conservation: conflict and accommodation in farming politics. Journal of Rural Studies 8 (20), 159–171.

Molnar, J.J., Duffy, P.A., Cummins, K.A., Vam Santen, E., 1992. Agricultural science and agricultural counterculture: paradigms in search of a future. Rural Sociology 57 (1), 83–91.

Morgan, K., Murdoch, J., 2000. Organic vs. conventional agriculture: knowledge, power and innovation in the food chain. Geoforum 31, 159–173.

Morris, C., Potter, C., 1995. Recruiting the new conservationists: farmer adoption of agri-environmental schemes in the UK. Journal of Rural Studies 11 (1), 51–63.

Morris, C., Winter, M., 1999. Integrated farming systems: the third way for European agriculture? Land Use Policy 16, 193–205.

Murdoch, J., Clark, J., 1994. Sustainable knowledge. Geoforum 25 (2), 115–132.

Napier, T.L., Thraen, C.S., Gore, A., Goe, W.R., 1984. Factors affecting the adoption of conventional and conservation tillage practices in Ohio. Journal of Soil and Water Conservation, 201–209.

National Farmers' Union, 1994. Submission by National Farmers' Union of England Wales to Royal Commission on Environmental Pollution: Study on Environmental Problems with Soil.

Norgaard, R., 1984. Traditional agricultural knowledge: past performance, future prospects and institutional implications. American Agricultural Economics Association 66, 874–878.

OECD (Organisation for Economic Co-operation and Development), 2001. Adoption of Technologies for Sustainable Farming Systems. Workshop Proceedings Wageningen. The Agricultural University.

Park, J., et al., 1997. Integrated arable farming systems and their potential uptake in the UK. Farm Management 9 (10), 483–494.

Pretty, J., 1995. Regenerating Agriculture: Policies and Practice for Sustainability and Self-reliance. Earthscan, London.

Pyrovetsi, M., Daoutopoulos, G., 1999. Farmers needs for nature conservation education in Greece. Journal of Environmental Management 56 (2), 147–157.

Raedeke, A.H., Rikoon, J.S., 1997. Temporal and spatial dimensions of knowledge: implications for sustainable agriculture. Agriculture and Human Values 14 (2), 145–158.

Richards, P., 1993. Cultivation: knowledge or performance? In: Hobart, M. (Ed.), An Anthropological Critique of Development: The Growth of Ignorance. Routledge, London, pp. 61–78.

Röling, N.G., Jiggins, J.L.S., 1994. Policy paradigm for sustainable farming. European Journal of Agricultural Education and Extension 1 (1–3), 23–43.

Röling, N.G., Wagemaker, M.A.E., 2000. Facilitating Sustainable Agriculture Cambridge. Cambridge University Press, Cambridge.

Romig, D.E., et al., 1995. How farmers assess soil health and soil quality. Journal of Soil and Water Conservation 50, 229–236.

Royal Commission on Environmental Pollution (RCEP), 1996. Sustainable Use of Soil. RCEP Nineteenth Report. HMSO, London.

Russell, D.B., Ison, R.L., 2000. The research-development relationship in rural communities: an opportunity for conceptual science. In: Ison, R.L., Russell, D.B. (Eds.), Agricultural Extensions and Rural Development. Breaking the Traditions. Cambridge University Press, Cambridge, pp. 11–31.

Ryan, R., Erickson, D., de Young, R., 2003. Farmers' motivations for adopting conservation practices along riparian zones in a mid-western agricultural watershed. Journal of Environmental Planning and Management Volume 46 (1), 19–37.

Seymour, S., Turner, R., Gerber, J., Kinsman, P., 1998. Research into Cost Effective Methods of Influencing Attitudes within the Agriculture Community in the Upper Hampshire Avon Catchment. A report commissioned by the Environment Agency.

Sillitoe, P., 1998. Knowing the land: soil and land resource evaluation and indigenous knowledge. Soil Use Management 14, 188–193.

Smith, K.A., Brewer, A.J., Dauven, A., Wilson, D.W., 2000. A survey of the production and use of animal manures in England and Wales. I. Pig manure. Soil Use and Management 16 (2), 124–132.

Smith, K.A., Brewer, A.J., Crabb, J., Dauven, A., 2001. A survey of the production and use of animal manures in England and Wales II. Poultry manure. Soil Use and Management 17 (1), 48–56.

Tebrugge, F., Bohrnsen, A., 2001. Farmers and experts opinion on no-tillage in West Europe and Nebraska. In: Garcia-Torres, L., Benites, J., Martinez-Vilela, A. (Eds.), Conservation Agriculture. A Worldwide Challenge, vol. 1. European Conservation Agriculture Federation Publications, Brussels, pp. 61–71.

Thrift, N., 1985. Flies and germs: a geography of knowledge. In: Gregory, D., Urry, J. (Eds.), Social Relations and Spatial Structures. Macmillan, London.

Tsouvalis, J., Seymour, S., Watkins, C., 2000. Exploring knowledge-cultures: precision farming, yield mapping and the expert–farmer interface. Environment and Planning A 32, 908–924.

UK Soil Management Initiative (SMI), 2002. A Guide to Managing Crop Establishment, Chester.

UK Soil Management Initiative (SMI), 2005. Visual Soil Assessment. UK Soil Management Initiative, Chester.

University of Reading, 2005. Assessment of likely barriers to participation in a framework of continuous professional development for farmers and how these might be overcome. Final Report submitted to Defra.

van der Ploeg, J.D., 1989. Knowledge systems, metaphor and interface: the case of potatoes in the Peruvian highlands. In: Long, N. (Ed.), Encounters at the Interface: A Perspective on Social Discontinuities in

Rural Development, Wageningen Studies in Sociology. The Agricultural University Pudoc, Wageningen, p. 27.

Vanclay, F., Lawrence, G., 1994. Farmer rationality and the adoption of environmentally sound practices: a critique of the assumptions of traditional agricultural extension. Journal of Agricultural Education and Extension 1 (1) 〈www.library.wur.nl/ejae/v1n1t.html〉, June 2004.

van Rompaey, A.J.J., Govers, G., van Heckle, E., Jacobs, K., 2001. The impacts of land use policy on soil erosion risk: a case study in central Belgium. Agriculture, Ecosystems and Environment 83, 83–94.

Walter, G., Wander, M., Bollero, G., 1997. A farmer centered approach to developing information for soil resource management: the Illinois soil quality initiative. America Journal of Alternative Agriculture 12 (2), 64–72.

Ward, N., 1995. Technological change and the regulation of pollution from agricultural pesticides. Geoforum 26 (1), 19–33.

Wilson, G., 1997. Assessing the environmental impact of the environmentally sensitive area scheme: a case for using farmers' environmental knowledge? Landscape Research 22 (3), 303–326.

Wilson, G., 2001. From productivism to post productivism and back again? Exploring the (un)changed natural and mental landscapes of European agriculture. Transactions of the Institute of British Geographers 26 (1), 77–102.

Winklerprins, A.M.G.A., 1999. Local soil knowledge: a tool for sustainable land management. Society and Natural Resources 12, 151–161.

Winter, M., 1997. New policies and new skills: agricultural change and technology transfer. Sociologia Ruralis 37 (3), 363–381.

Journal of Environmental Management 86 (2008) 229–245

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Assessment of a turfgrass sod best management practice on water quality in a suburban watershed

C.E. Richards[a], C.L. Munster[a,*], D.M. Vietor[b], J.G. Arnold[c], R. White[b]

[a]*Department of Biological and Agricultural Engineering, Texas A&M University, 127 Hobgood Building, College Station, TX 77843-2474, USA*
[b]*Department of Soil and Crop Sciences, Texas A&M University, 127 Hobgood Building, College Station, TX 77843-2474, USA*
[c]*USDA—Agricultural Research Service, Grassland, Soil and Water Research Laboratory, Temple, TX, USA*

Received 17 August 2005; received in revised form 9 November 2006; accepted 7 December 2006
Available online 12 February 2007

Abstract

The disposal of manure on agricultural land has caused water quality concerns in many rural watersheds, sometimes requiring state environmental agencies to conduct total maximum daily load (TMDL) assessments of stream nutrients, such as nitrogen (N) and phosphorus (P). A best management practice (BMP) has been developed in response to a TMDL that mandates a 50% reduction of annual P load to the North Bosque River (NBR) in central Texas. This BMP exports composted dairy manure P through turfgrass sod from the NBR watershed to urban watersheds. The manure-grown sod releases P slowly and would not require additional P fertilizer for up to 20 years in the receiving watershed. This would eliminate P application to the sod and improve the water quality of urban streams. The soil and water assessment tool (SWAT) was used to model a typical suburban watershed that would receive the sod grown with composted dairy manure to assess water quality changes due to this BMP. The SWAT model was calibrated to simulate historical flow and estimated sediment and nutrient loading to Mary's Creek near Fort Worth, Texas. The total P stream loading to Mary's Creek was lower when manure-grown sod was transplanted instead of sod grown with inorganic fertilizers. Flow, sediment and total N yield were the same for both cases at the watershed outlet. The SWAT simulations indicated that the turfgrass BMP can be used effectively to import manure P into an urban watershed and reduce in-stream P levels when compared to sod grown with inorganic fertilizers.

Keywords: Manure; BMPs; Urban water quality; Nutrient export; Phosphorus; Turfgrass; Watershed modeling; SWAT; TMDL

1. Introduction

In 2001, the Texas Commission on Environmental Quality (TCEQ) and the United States Environmental Protection Agency (USEPA) approved the recommendations of two separate total maximum daily load (TMDL) assessments that suggested a 50% reduction of soluble reactive phosphorus (SRP) to sections of the North Bosque River in central Texas. One of these sections, at the headwaters of the North Bosque River, is known as the Upper North Bosque River (UNBR) watershed. The UNBR watershed is located in Erath County, the largest milk producing county in the State of Texas (USDA-ARS, 2003). The number of dairies in the watershed constantly changes as a function of feed costs and milk prices (Hauck, 2002), but approximately 80 active dairies and 40 000 cows were distributed throughout the watershed in 2002 (Munster et al., 2004).

McFarland and Hauck (1999) demonstrated that the largest phosphorus (P) loadings to the North Bosque River originated from dairy waste application fields (WAFs). In response to the TMDL recommendations, the State of Texas subsidized manure composting facilities in the UNBR watershed in order to move approximately 50% of the manure off of the dairies (TCEQ, 2003) and reduce the cost of exporting the nutrients out of the watershed. In September 2000, the TCEQ and the Texas State Soil and Water Conservation Board (TSSWCB) began subsidizing the transport of fresh manure from dairies to the composting facilities located in the UNBR and the Leon River watersheds (TCEQ, 2003). This compost has been

*Corresponding author. Tel.: +1 979 847 8793; fax: +1 979 847 8627.
E-mail address: c-munster@tamu.edu (C.L. Munster).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.039

used by the Texas Department of Transportation (TxDOT) to stabilize roadside embankments at construction sites (TCEQ, 2003) and by the Texas Water Resources Institute (TWRI) in cooperation with the US Army to revegetate areas of the Fort Hood Western Training Grounds (TWRI, 2004a). However, new markets that do not require subsidies are needed to utilize the approximately 150,000 m^3 of surplus compost currently available in the watershed (TCEQ, 2003; TWRI, 2004b).

The UNBR TMDL implementation plan states that "land application remains one of the best and most appropriate methods for dealing with large amounts of animal wastes" (TCEQ, 2002). Successful land application is achieved when nutrient transport into surface waters is minimized (TCEQ, 2002) and crop nutrient uptake is maximized so that a large percentage of the applied nutrients can be harvested and exported. The suggested turfgrass sod BMP utilizes P in the composted dairy manure to grow turfgrass at proposed sod farms in the UNBR watershed. The top-dressed (surface applied) manure-grown sod would be harvested an average of 1.5 times per year and each harvest would remove the sod, the composted dairy manure and a thin layer of topsoil. The sod and topsoil would be exported out of the UNBR watershed to suburban developments in nearby watersheds. The value of the turfgrass sod would allow growers to transport the manure nutrients from the dairies to the turfgrass fields and ultimately out of the UNBR watershed. This turfgrass sod BMP has the potential to eliminate the need for state subsidies to move excess manure from impaired watersheds (Hanzlik et al., 2004).

Turfgrass produced with top-dressed composted dairy manure can be sold at a premium because of its unique properties, including accelerated establishment rate and increased cation exchange capacity, aggregation, organic matter, and water holding capacity of the soil (Murray, 1981). Therefore, the increased amount of manure P and organic matter adds value to the manure-grown sod. Import of manure P in sod can eliminate applications of inorganic P fertilizer for establishment and for annual turf maintenance. Previous studies indicated sod transplanted from fields supplied with 190 kg P/ha of manure P raised soil-test P at the receiving site to 130 mg P/kg soil (Vietor et al., 2004). If return of clippings and the dense plant population in the sod layer minimizes annual loss of nutrients after transplanting (Kopp and Guillard, 2002; Kussow, 2004), soil-test P can remain above turf P sufficiency levels for 10–15 years (Carrow et al., 2001). In addition, import of manure P with sod over time could alleviate regulatory constraints similar to partial P fertilizer bans in Minnesota (MAWD, 2003).

Although turfgrass sod is not produced in the UNBR watershed at this time, approximately 5219 ha of suitable sites were identified in Erath County (Munster et al., 2004). In addition, the market for turfgrass sod is expanding in the Dallas/Fort Worth (DFW) metroplex, which is only 160 km from the UNBR watershed (Hall, 1999). Currently, the DFW metroplex purchases and hauls about 60% of transplanted sod from distant locations, including the Texas Gulf Coast and Oklahoma (Munster et al., 2004). The proximity of this growing urban market, which is connected to the UNBR watershed by major roads, favored the expansion of dairy production in the UNBR watershed in the 1980s and 1990s. Munster et al. (2004) estimated approximately 396 440 kg P/year could be exported from Erath County alone if manure was applied at a rate of 200 kg/ha to turfgrass production sites totaling 2643 ha.

Vietor et al. (2004) demonstrated that sod grown with top-dressed manure P can be transplanted without increasing runoff losses of total dissolved P (TDP) when compared to transplanted turfgrass sod fertilized with inorganic P. It was also demonstrated that losses of TDP and total Kjeldahl N (TKN) from turfgrass top-dressed with manure or inorganic fertilizer can approach three times that lost from sod transplanted from fields where composted dairy manure was applied (Vietor et al., 2004). However, the impact of importing this turfgrass sod containing manure nutrients on water quality needs to be evaluated for suburban watersheds.

Bednarz and Srinivasan (2002) simulated the impact of suburban development on flow and sediment yield at the outlet of a suburban stream named Mary's Creek near Fort Worth, Texas. The study predicted increases in flow and sediment yield for Mary's Creek after the construction of a proposed development named Walsh Ranch through simulations of a hydrologic model known as the Soil and Water Assessment Tool (SWAT) (Neitsch et al., 2002).

Limited streamflow, sediment and nutrient data were available for Mary's Creek. A USGS gauging station located at the outlet of Mary's Creek provided historic streamflow data from 1998 to 2002 for model calibration. Moreover, Bednarz and Srinivasan (2002) successfully used this streamflow data for SWAT model simulations of sediment transport. However, previous modeling studies have simulated effects of changes in land management without calibrating the watershed model to measured data (He, 2003; Santhi et al., 2003; Tripathi et al., 2004). In addition, techniques are available for estimating sediment and nutrient loads needed for calibration of watershed models. Tripathi et al. (2003) utilized a pre-calibrated and validated SWAT model to identify critical sub-watersheds to aid in the development of effective management plans in India. Chen et al. (2000) used crop yields and experimental field data to calibrate sediment and nutrient loads in the Environmental Policy Integrated Climate Model or Erosion Productivity Impact Calculator (EPIC) (Williams et al., 1984). Land cover and surface flow were considered the predominant control factors in simulations of sediment and nutrient export from the watershed. Wickham and Wade (2002) similarly demonstrated that land use was a major factor in N and P transport and loss in surface waters. For the Walsh Ranch study, a technique proposed by Bhuyan et al. (2003) was used to calibrate the SWAT model. This technique separates nutrient and sediment losses into stormflow and baseflow losses.

In this study, the SWAT simulations were used to predict nutrient transport responses to two turfgrass import treatments on the Walsh Ranch development. The first treatment was sod transplanted from fields where inorganic fertilizer was applied. The second treatment was turfgrass transplanted from fields where composted dairy manure was applied.

The primary objective of this paper was to assess water quality changes in a suburban watershed due to a turfgrass BMP that imports sod transplanted from turfgrass fields where top-dressed composted dairy manure was applied. This assessment used field data from turfgrass sod field research and the SWAT hydrologic simulation model to analyze changes in flow and sediment and nutrient loading for Mary's Creek in response to the turfgrass BMP.

Table 1
The land-use distribution of the Mary's Creek watershed for major land uses present before and after the construction of the Walsh Ranch development

Land use	Watershed area pre-development (%)	Watershed area post-development (%)
Urban-high density	2.39	2.41
Pasture	19.30	18.85
Range-grasses	40.57	31.30
Forest-mixed	17.88	14.75
Industrial/institutional	0.06	0.72
Transportation/commercial	3.58	8.13
Residential-medium density	11.30	19.03
Residential-low density	4.92	4.81

2. Materials and methods

2.1. Watershed selection

The Mary's Creek watershed with the proposed Walsh Ranch development was chosen to receive the turfgrass grown with composted dairy manure due to its proximity (100 km) to the UNBR. Economically, the distance from the UNBR to Mary's Creek is within an acceptable hauling distance for turfgrass sod (Munster et al., 2004). The Walsh Ranch development is a 2800 ha planned community that is scheduled to begin construction as early as 2020 (W. Frossard, personal communication, 23 June 2003). The development will resemble a small, self-sufficient community with schools, industrial areas, residential sites, public parks, and a community center and will require turfgrass for residential, commercial, and industrial areas. The Walsh Ranch development includes approximately 2800 ha of the Mary's Creek watershed, however, the majority of the Mary's Creek watershed will remain rangeland after construction of Walsh Ranch (Table 1).

Mary's Creek is a perennial stream located west of the DFW metroplex that drains approximately 14 272 ha of predominately range and pasture (Fig. 1). Mary's Creek begins in Parker County and terminates at the Clear Fork of the Trinity River (CFTR) in Tarrant County and is located in the prairie and lakes of Texas on the upper margins of the coastal plain. The terrain is mostly open prairie and rangeland with rolling hills that range from 150 to 250 m in elevation. The climate varies between

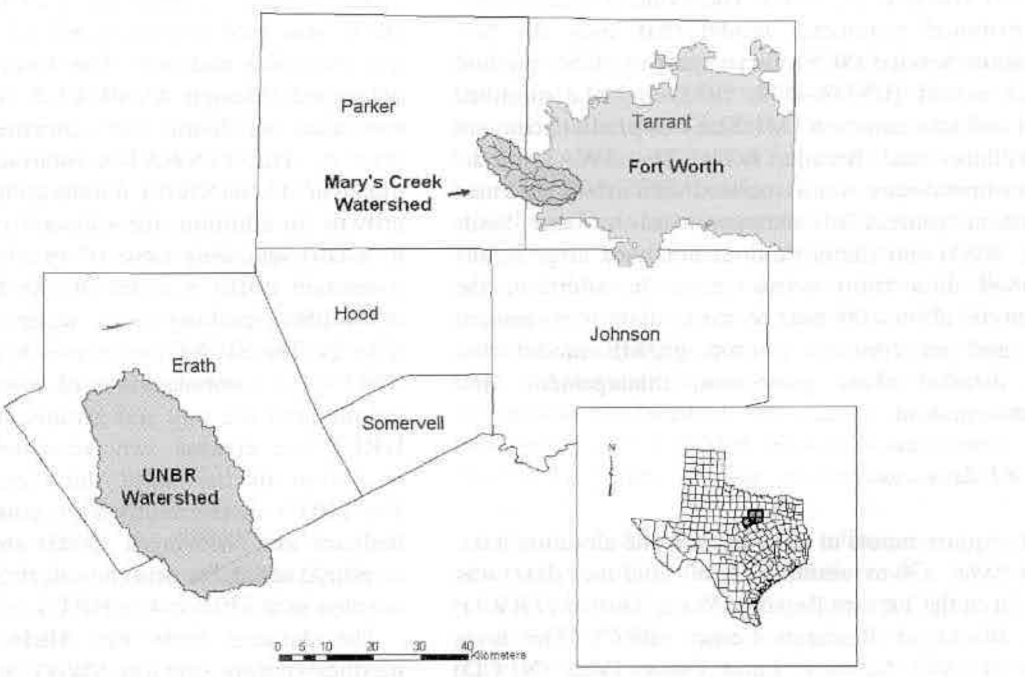


Fig. 1. The location of the Mary's Creek watershed, the UNBR watershed, and Fort Worth, Texas, with county boundaries shown.

subtropical and continental with summers that are hot and humid with 38 °C days common. The winters are characteristically mild with short-lived periods of extreme cold where −7 °C can occur and snowfall is rare. The annual precipitation ranges from less than 500 to more than 1250 mm with an average of approximately 800 mm/year. The majority of the rainfall occurs in the spring (USDA-TAES, 1981) and the historical streamflow records from the United States Geological Survey (USGS) gauging station (08047050) on Mary's Creek reflects this trend with highest flows occurring in March. This gauging station is located near the confluence of Mary's Creek and the CFTR and daily streamflow records were available from June 1, 1998 to September 30, 2002.

Currently, approximately 41% of the land in the Mary's Creek watershed is rangeland and only 22% is allocated to urban land uses (Table 1). Very few nutrients are now applied in the watershed (Jon R. Green, personal communication, 17 October 2004), and there are no wastewater treatment plants that discharge into the stream. Watersheds similar to Mary's Creek in the DFW metroplex area are not typically impaired by nutrients (USGS, 1999).

2.2. The SWAT model

The SWAT 2003 model was used in this study and was interfaced with ArcView 3.2 to integrate geospatial data into the simulations. The SWAT model simulations allowed for the assessment of water quality changes in a developing suburban watershed due to the import of turfgrass sod grown with composted dairy manure. The SWAT model is capable of detecting changes in water yield, sediment, nutrient and pesticide loading due to the effects of land use and agricultural management on a river basin scale (Arnold et al., 1998). The model is a daily time-step, distributed parameter model that uses the Soil Conservation Service (SCS) curve number (CN) method to predict runoff (USDA-SCS, 1972) and the modified universal soil loss equation (MUSLE) to predict sediment yield (Williams and Berndt, 1977). The SWAT model simulates impervious cover associated with urban land uses as consistent sources of sediment and nutrient loads (USEPA, 1983) and therefore does not need large inputs of observed data from urban areas. In addition, the SWAT model allows the user to manipulate management routines and incorporates a crop growth model that includes detailed plant protection, management, and harvest information.

2.3. SWAT data sets

SWAT requires inputs of land use, soil and elevation data. A raster layer (30-m resolution) of land-use data was available from the Tarrant Regional Water District (TRWD) and the Blackland Research Center (BRC). The layer consisted of 1992 National Land Cover Data (NLCD) meshed with a regional Texas Agricultural Experiment Station (TAES) land use map developed from 1997 Landsat 5 imagery. The multi-resolution land characteristics (MRLC) consortium derived the NLCD from Landsat 5 Thematic Mapper satellite imagery. The MRLC classification provided detail about urban land uses and the TAES classification detailed agricultural land uses. The collective map contained both the urban and agricultural data.

Soils data were collected from the Natural Resources Conservation Service (NRCS) which provided detailed Soil Survey Geographic (SSURGO) datasets with scales ranging from 1:12 000 to 1:24 000. These datasets were digitized from published county soil surveys (USDA-NRCS, 1995). A 10-m raster digital elevation model (DEM) of the area and a digitized stream network created by the City of Fort Worth were also available from the BRC.

The SWAT model includes a weather generator function but also allows the user to input weather data. Weather data from the Aledo (Station ID 480129) and Benbrook Dam (Station ID 480691) weather stations were available through the National Climatic Data Center (NCDC). These weather stations were located within an 8 km radius of the Mary's Creek watershed (Fig. 2). Both stations reported daily precipitation totals and the Benbrook Dam station reported daily maximum and minimum air temperature data. The Aledo weather station data spanned the period from 1960 to 2003 and Benbrook Dam weather station data were available for 1990–2003. An extensive SWAT weather database was used to generate relative humidity and solar radiation data based on inputs from regional weather stations near Fort Worth.

2.4. SWAT model configuration

An Arcview 3.2 interface, AVSWAT-X (DiLuzio et al., 2003), was used to process SWAT model inputs for land use, elevation and soil. The 10-m resolution DEM was delineated through AVSWAT-X and a 200 ha threshold was used to divide the watershed into 37 sub-basins (Fig. 2). The AVSWAT-X interface also linked the land use layers to the SWAT databases for land cover and plant growth. In addition, the software integrated the soil layer to a corresponding table of specific soil parameters. The watershed outlet was set at the USGS gauging station (08047050) resulting in a watershed area of 13 976 ha (Fig. 2). The SWAT model uses hydrologic response units (HRUs) to combine areas of a sub-basin into areas of unique land use and soil groups. The threshold at which HRUs are created can be changed based upon the resolution of the input data and the desired output. The HRUs in this study were constructed similar to the Bednarz and Srinivasan (2002) study with the land use threshold set at 5% and the soil threshold set at 10%. This resulted in a total of 470 HRUs.

The datasets from the Aledo and Benbrook Dam weather stations and the SWAT weather generator database were utilized during the SWAT model simulations.

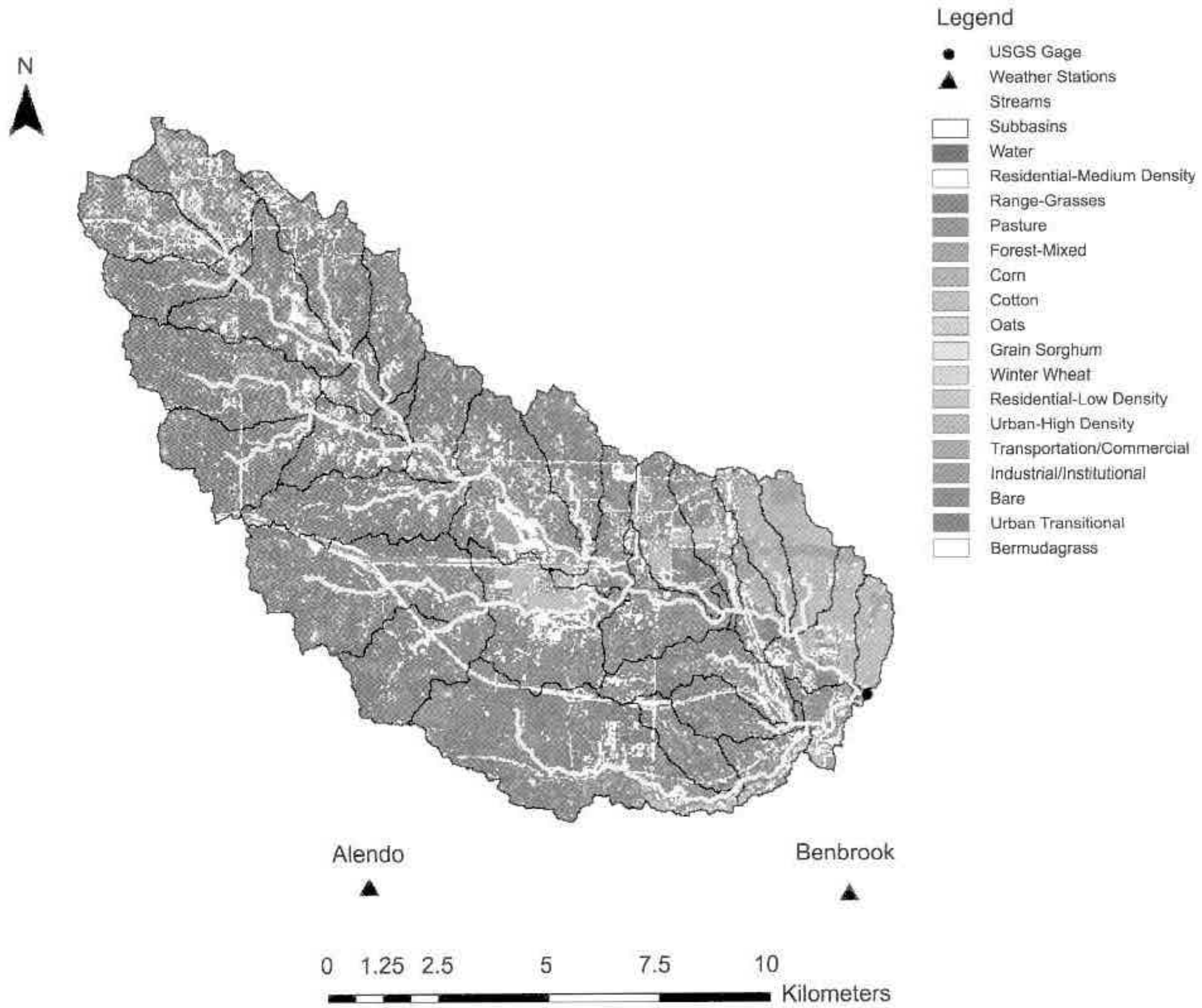


Fig. 2. The stream network and current SWAT land uses in the Mary's Creek watershed with the location of the Aledo and Benbrook weather stations, the USGS stream gage (0804750), and the sub-basins used in the SWAT model simulations also shown. Refer to DiLuzio et al. (2002) for SWAT land use class abbreviations.

The SCS CN method was used to simulate surface runoff and the Priestly–Taylor equation was used to simulate potential evapotranspiration. The Manning's roughness coefficient of the stream channel was set at the SWAT default value (0.014) and potential heat units (PHUs) were used to simulate biomass production. Soils in the watershed do not exhibit preferential flow and therefore the crack flow routine in the model was not activated. SWAT's water quality and in-stream channel degradation routines were not activated due to the lack of field collected data necessary for the use of these routines.

2.5. *SWAT model calibration*

The MRLC/TAES land use map was used to represent land use for the calibration of the SWAT model. The SWAT model was calibrated for flow using historic daily streamflow data from the USGS gage (08047050) over the period from June 1998 to September 2002. A Nash Sutcliffe (NS) statistic (Nash and Sutcliffe, 1970) of greater than 0.50 was used as the criteria for a successful calibration. The NS statistic measures how well the predicted mean agrees with the observed values. An NS value of 1.0 means that the prediction is perfect. The actual simulation period for flow calibration started January 1, 1990 and concluded September 30, 2002 to allow for an adjustment period for model equilibrium among soil, water, and plant processes. The SWAT model was calibrated for average monthly stream flows in a similar manner to previous SWAT model simulation studies in the nearby Bosque River watershed by Saleh et al. (2000), Santhi et al. (2001) and Stewart et al. (2006). After calibration, the predicted monthly average streamflow produced an NS statistic of 0.72 and a root mean square error (RMSE) of 0.54 when compared to the observed monthly average streamflow at the watershed outlet.

Initially, model monthly flow estimates were higher than observed monthly flows. Therefore, the SWAT model parameters were adjusted as shown in Table 2 until the predicted flow was approximately equal to the observed flow. At the start of the model calibration, the base flow fraction was calculated using a base flow filter developed by

Arnold et al. (1995). The base flow alpha factor (ALPHA_BF) was adjusted to 0.158 according to the filter results. Then, to bring the simulated flow rate down further, all CNs (CN2) were adjusted down by a factor of 8, the CN2 limits were adjusted down by 10% and, the soil evaporation compensation factor (ESCO) and the plant water uptake compensation factor (EPCO) were also adjusted down. Temporal adjustments to the peak flows and baseflows were made by increasing the groundwater delay coefficient (GW_DELAY) and increasing the effective hydraulic conductivity of the main channel alluvium (CH_K2). Finally, to accurately simulate the amount of water returning to the stream, the amount of shallow aquifer water that moved into the soil profile (GW_REVAP) was increased and the threshold depth of water in the shallow aquifer for this "revap" to occur (REVAPMN) was decreased.

Table 2
The SWAT model parameters adjusted during the model calibration for stream flow

Parameter	Default value	Calibration value
ALPHA_BF	0.0	0.158
CH_K2	0.0	1.0
CN2	0	−8[a]
EPCO	1.0	0.0
ESCO	0.95	0.01
GW_DELAY	31	93
GW_REVAP	0.02	0.2
REVAPMN	1.0	0.0

[a]All CNs were adjusted downward by −8.

Two annual sediment loading estimations were averaged to estimate an average annual sediment loading to Mary's Creek of 2400 metric tons. The TRWD used sediment removal records below the junction of Mary's Creek and the CFTR to estimate an annual sediment loading of 3200 metric tons in Mary's Creek.

In addition, three separate sources of data were used to estimate sediment load due to stormflow and baseflow as proposed by Bhuyan et al. (2003). Sediment sources for stormflow loadings were assumed to be from urban and rangeland/pasture land uses only. Urban sediment stormflow loads were calculated using a USGS regression equation developed from local data (Baldys et al., 1998). Rangeland/pasture sediment stormflow loads were calculated using event mean concentration (EMC) values based on the average annual stormflow volume (Newell et al., 1992; Baird and Ockerman, 1996). Baseflow sediment data were collected in the summer of 2004 for this study and the average sediment concentration was multiplied by the average annual baseflow volume to calculate the annual baseflow sediment load. The average annual sediment loading from the three data sources was 1600 metric tons (Table 3).

Table 3
The stormflow, baseflow and total average annual sediment load values used to estimate the average annual sediment load at the outlet of Mary's Creek

Source of load	Sediment (tons/year)
Urban storm[a]	570
Rangeland/Pasture storm[b]	820
Baseflow[c]	210
Total	1600

[a]Calculated from USGS regression equation developed by Baldys et al. (1998).
[b]Calculated from event mean concentration (EMC) values (Newell et al., 1992; Baird and Ockerman, 1996).
[c]Observed from baseflow sampling of Mary's Creek conducted May–July 2004.

The SWAT model was calibrated for average annual sediment loading over the period from January 1, 1990 to December 31, 2000. This was the same time period in which sediment loading was estimated (Table 3). The simulated average annual sediment yield after calibration was 2830 metric tons. The SWAT prediction was approximately 18% higher than the calculated average annual sediment yield of 2400 metric tons.

No in-stream nutrient data was available for Mary's Creek and therefore, N and P loads in Mary's Creek had to be estimated. Total N, nitrate and nitrite-N, and total P average annual loads to Mary's Creek were estimated from local urban storm EMC data collected by the USGS (Baird and Ockerman, 1996; Newell et al., 1992), and the baseflow stream samples collected during this study (Table 4). The average annual nutrient load was estimated using the same procedure (Bhuyan et al., 2003).

Table 4
The estimated stormflow, baseflow and total average annual nutrient load at the outlet of Mary's Creek

Source of load	Total N (kg/year)	NO_2 and NO_3 (kg/year)	Organic N (kg/year)	Total P (kg/year)
Urban storm[a]	7700	2690	5010	1930
Rangeland/Pasture storm[b]	17 590	3790	13 800	1400
Baseflow[c]	42 280	140	42 140	12 230
Total	67 570	6620	60 950	15 560

[a]Calculated from USGS regression equation developed by Baldys et al. (1998).
[b]Calculated from EMC values (Newell et al., 1992; Baird and Ockerman, 1996).
[c]Observed from baseflow sampling of Mary's Creek conducted May–July 2004.

The baseflow values for total N, organic N, and total P were disproportionately high compared to the stormflow values due to the higher volume of baseflow per year in Mary's Creek (60%). Also, the baseflow values were heavily weighted to the summer season level of nutrients since baseflow values were calculated from sampling of Mary's Creek conducted May–July 2004.

The SWAT model was calibrated for total average annual N loading over the same period as the sediment calibration (January 1, 1990–December 31, 2000). The average annual organic N yield was estimated by assuming

$$N_{\text{organic}} = N_{\text{total}} - NO_3 - NO_2. \quad (1)$$

The estimated values for total average annual organic N, nitrate-N and nitrite-N loads (Table 4) were used to calibrate SWAT.

The simulated average annual total N yield at the outlet of Mary's Creek after calibration was approximately 11% lower than the calculated average annual total N yield (Table 5). The predicted average annual organic yield after calibration was approximately 13% lower than the calculated average annual organic N yield (Table 5). Lastly, the predicted average annual nitrate- and nitrite-N yields after calibration were approximately 4% higher than the calculated average annual nitrate- and nitrite-N yields (Table 5).

The stream monitoring data for Mary's Creek did not breakdown P into organic and mineral components. Therefore, the SWAT model was calibrated to predict total P (organic and mineral P combined). The simulated average annual total P yield after calibration was approximately 0.1% higher than the estimated average annual total P yield (Table 5).

Without calibration, the SWAT model predicted average annual sediment and nutrient yields greater than the estimated yields. The SWAT model parameters (Table 6) that were adjusted during the nutrient calibration were as follows. The average slope length (SLSUBBSN) was reduced to 5 m (Neitsch et al., 2001), the average slope steepness (SLOPE) was adjusted down to 0.02 m/m and the universal soil loss equation soil erodibility factor (USLE_K1) was decreased by approximately 60% for all soils in the watershed to reduce the HRU contribution of sediment. The biological mixing efficiency (BIOMIX) and

Table 5
The results of the SWAT model calibration for in-stream nutrients at the outlet of Mary's Creek

Constituent	Simulated annual load (kg/year)	Estimated annual load (kg/year)	Difference (%)
Total N	59 940	67 570	−11
NO_2 and NO_3	6880	6620	4
Organic N	53 060	60 950	−13
Total P	15 580	15 560	0.1

Simulated nutrient loads were compared to estimated nutrient loads.

Table 6
The SWAT model parameters adjusted during the model calibration for stream sediment and nutrient loads

Parameter	Default value	Calibration value
SLOPE	0.129	0.020
SLSUBBSN	24.390	5.000
USLE_K1 (all soils)	Various	−60%
BIOMIX	0.92	0.20
ERORGN	0.0	5.0
NPERCO	0.20	0.35
RCN	1.0	0.3
RSDIN	0	10 000
SOL_ORGN	0	10 000
SOL_K (Aledo, Maloterre, Purves)	Various	−100%
SOL_Z1 (Aledo)	101.6	50
SOL_ORGP	0	4000

the organic N enrichment ratio (ERORGN) were increased to improve the ratio of organic N to nitrate- and nitrite-N. The initial soil organic N concentration (SOL_ORGN) and the initial residue cover (RSDIN) were increased to enlarge the organic N yield. The N in rainfall (RCN) and the depth of the top layer of the Aledo soil (SOL_Z1) were decreased to reduce nitrate- and nitrite-N loading to the stream. Also, the saturated hydraulic conductivity (SOL_K) of three soils was reduced in the bottom layers to trap nitrate and nitrite in the soil profile. The nitrogen percolation coefficient (NPERCO) was adjusted to increase the N percolation to the stream from the shallow aquifer and the initial soil organic P concentration (SOL_ORGP) was raised to increase P additions to the stream (Santhi et al., 2001).

2.6. SWAT simulations

2.6.1. SWAT turfgrass transplant routine

Sod is typically transplanted in squares or unrolled in strips to form an instant layer of vegetation. There were no management practices in the SWAT model to simulate this instant addition of soil and biomass. Therefore, a separate turfgrass transplant routine was created that modified the SWAT model management practices to instantly add a layer of soil and mature grass to the soil profile of HRUs that receive transplanted sod. The transplant routine assumed that the layer of soil added had the same characteristics of the soil presently in the HRU. This did not account for the soil characteristics of the soil transplanted with the turfgrass sod, but simplified the analysis of the nutrient import. Soils that are typically transplanted with turfgrass sod would most likely have greater clay content than the soils in the Mary's Creek watershed. Therefore, the SWAT simulations would be conservative with respect to nutrient transport as turfgrass sod with higher clay contents would have increased water holding capacity and decreased nutrient transport capability. The turfgrass import routine required 12 new inputs to the model. A month and day of sod input allowed the user to control the time of the transplant within the model. The additional sod and soil was integrated at the input time into the SWAT routines for runoff, sediment loss, nutrient loss and crop growth.

The addition of the turfgrass transplant routine allowed the SWAT model to simulate the implementation of the turfgrass BMP in the Walsh Ranch development. The SWAT simulations were used to evaluate the effects of importing turfgrass sod fertilized with composted dairy manure on water quality in the Mary's Creek watershed.

2.6.2. Turfgrass treatments

The SWAT model was used to simulate three turfgrass treatments. The treatments included the BMP treatment, a conventional treatment, and the status quo. The BMP and conventional treatments were implemented in the Walsh Ranch development. The status quo simulated only the current land uses in the Mary's Creek watershed.

2.6.2.1. Status quo. The land use classifications for the status quo were not changed from the calibration simulations. The simulation of the current land uses in Mary's Creek provided a control for evaluation of the Walsh Ranch development on water quality. Therefore, both the BMP and conventional treatments could be compared to water quality predictions for current land uses in the Mary's Creek watershed (Table 1).

2.6.2.2. The conventional treatment. The conventional treatment utilized turfgrass sod transplanted from fields grown with inorganic P fertilizer. Additionally, the turfgrass was top-dressed annually with inorganic P fertilizer after transplanting into the Walsh Ranch development. The new SWAT turfgrass transplant routine was used to simulate the import of the inorganic fertilizer-grown sod on residential, commercial and public landscapes planned for the Walsh Ranch development. The physical and chemical properties of the imported turfgrass sod in the conventional treatment were set as found in field experiments. Table 7 shows the summary of the experimental data for conventionally treated sod (Choi et al., 2003; Vietor et al., 2002). The soil organic P content, which was similar between conventional and manure-grown sod, was calculated as the difference between total P and P quantified in soil-test extractions. Conventional fertilizer applications of inorganic N and P were applied to the turfgrass sod as needed for continued growth after transplanting. Inorganic N was applied to the transplanted sod at the rate of 60 kg/ha/year and inorganic P was applied at the rate of 18 kg/ha/year. The conventional treatment transplanted turfgrass sod to approximately 1400 ha of the Walsh Ranch development which affected 25 SWAT model HRUs in the Mary's Creek watershed (Fig. 3).

2.6.2.3. The BMP treatment. In the BMP treatment, turfgrass sod transplanted from fields top-dressed with composted dairy manure was also simulated using the new turfgrass import routine. The manure-grown sod was transplanted into the same residential, commercial and public landscapes in the Walsh Ranch development as was simulated in the conventional treatment. The properties of the transplanted sod were adjusted in the BMP treatment to represent nutrient levels of turfgrass grown with composted dairy manure and inorganic N fertilizer as found in field experiments. Table 7 shows the summary of field data from turfgrass so field research for BMP treated sod (Choi et al., 2003; Vietor et al., 2002, 2004). The manure application during sod production, which supplied P that was largely soluble (75% of total P) in soil-test extractions, increased soluble but not organic P content compared to fertilizer-grown sod at sod harvest. After turfgrass sod was transplanted to the Walsh Ranch development, inorganic N fertilizer was applied as needed (60 kg/ha/year), but no inorganic P fertilizer was added. The turfgrass was placed on the same 1400 ha and in the same 25 HRUs of the SWAT model as simulated in the conventional treatment.

Table 7
The SWAT model inputs used to simulate the conventional and BMP treatments for the installation of turfgrass sod into the Mary's Creek watershed

Turfgrass sod input	Conventional treatment value	BMP treatment value
MON (month)	02 (February)	02 (February)
DAY (day)	01	01
HEATU (heat units to maturity of sod)	3000	3000
SODLAI (leaf area index of sod)	4.0	4.0
SODBION (N content of biomass)	225 kg/ha	244 kg/ha
SODBIOP (P content of biomass)	36 kg/ha	42 kg/ha
SODPPLT (depth of soil added)	25 mm	25 mm
SODORGN (organic N content of soil)	370 kg/ha	540 kg/ha
SODORGP (organic P content of soil)	126 kg/ha	115 kg/ha
SODNO3 (nitrate content of soil)	3 kg/ha	3 kg/ha
SODSOLP (soluble P content of soil)	36 kg/ha	77 kg/ha
SODBIOM (biomass of sod)	18 000 kg/ha	18 000 kg/ha

2.6.3. Simulation procedures

An initial SWAT simulation was performed to demonstrate the effects of the Walsh Ranch development infrastructure (roads, removal of trees, etc.) on streamflow, sediment and nutrient loading without the turfgrass present. The residential, commercial, and public landscapes that would eventually be planted with turfgrass sod were simulated as unfertilized pasture. This simulation predicted

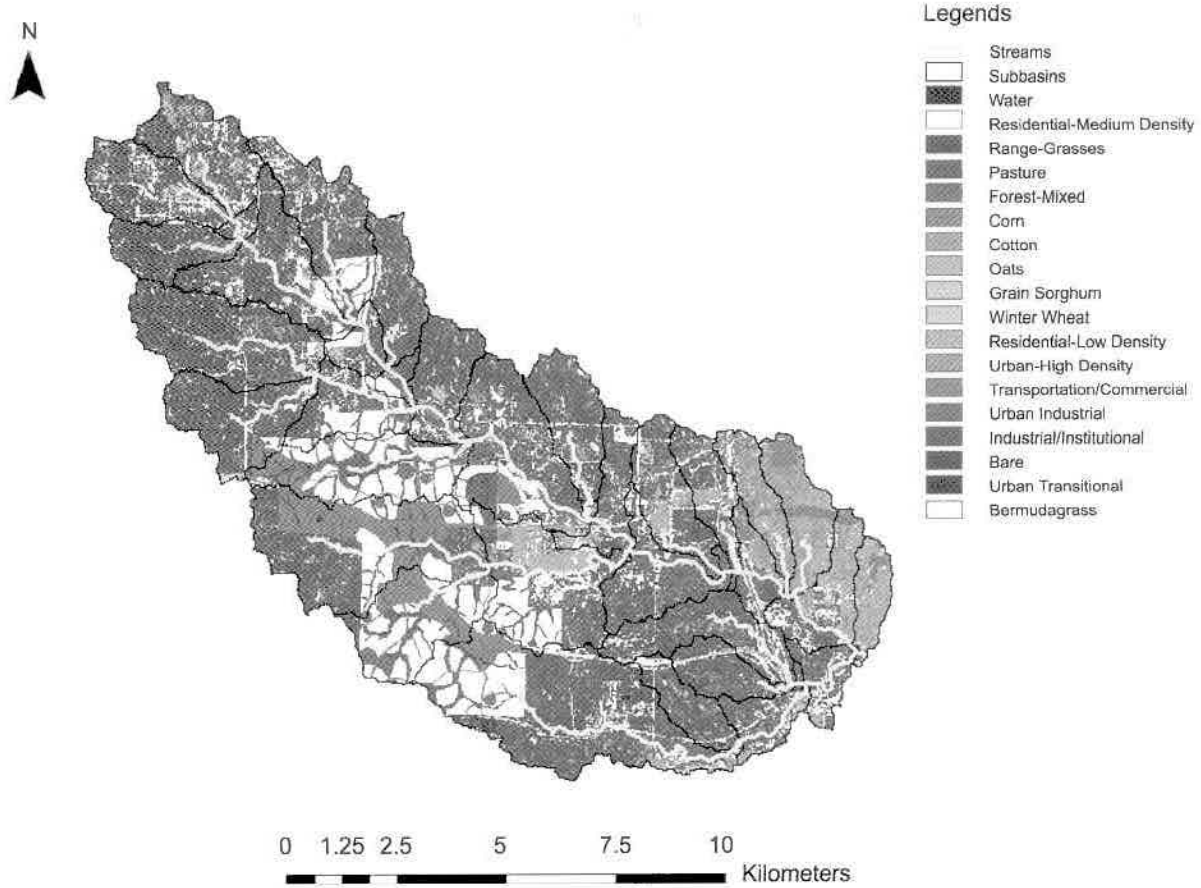


Fig. 3. The Mary's Creek watershed with areas where turfgrass was installed (land-use category Bermudagrass and Residential-Medium Density) in the Walsh Ranch development. This land cover map was used for the SWAT simulations of both the conventional and BMP treatments.

monthly flow and yearly sediment and nutrient loading for a 5-year period (1986–1990) preceding the conventional and BMP turfgrass sod simulations.

Two SWAT simulations were performed to analyze each turfgrass sod treatment. The first model simulation predicted monthly flow and yearly sediment and nutrient loading for a 10-year period (1991–2000). For simulations of the conventional and BMP treatments, the SWAT management files were revised to simulate imports of the contrasting turfgrass sod sources into the Walsh Ranch development on February 1 of year one of the 10-year period (1991). The newly installed turfgrass sod utilized the auto-fertilization and auto-irrigation routines in the SWAT model to ensure that the turfgrass sod was not stressed by lack of nutrients and water. This effectively simulates the management treatment that new sod would receive after transplanting. However, no inorganic P was applied to the BMP treatment after transplanting. For the status quo, no turfgrass sod was installed and land use classifications were not changed. The SWAT model auto-fertilization and auto-irrigation routines were not utilized in the status quo simulation.

A second model simulation was run to predict yearly flow and sediment and nutrient loading from 1950 to 2000 for each sod treatment. These simulations compared long term water quality impacts of the turfgrass BMP to that of the status quo and conventional treatments. The transplant of turfgrass for the conventional and BMP treatments took place on February 1 of year one (1950) and auto-fertilization and auto-irrigation was also used. Again, the land use classifications for the simulation of the status quo were unchanged.

3. Results

3.1. *Influence of development*

Construction of the Walsh Ranch development added 160 ha of impervious cover within the watershed and resulted in an increase of surface runoff. The effect of this additional impervious area on streamflow, sediment and nutrient loads in the Mary's Creek watershed was calculated from a 5-year SWAT simulation from 1986 to 1990. This simulation modeled the Walsh Ranch development with impervious surfaces but without the installation of turfgrass. The green spaces in the development were simulated as unfertilized pasture. The simulation demonstrated the effects of the land use changes in the Mary's Creek watershed that were not related to the turfgrass transplant. This allowed these land-use change effects to be removed from the results of the model simulations after turfgrass was installed in the Mary's Creek watershed. The simulated average increase of streamflow was 0.03 m^3/s. Other increases were 636 tons/year for sediment, 17 838 kg/year for organic N, 1142 kg/year for nitrate-N, and 4965 kg/year for total P at the outlet of Mary's Creek.

3.2. Flow

The 10-year SWAT simulation revealed streamflow was 10% greater for the BMP and conventional turfgrass treatments than for the status quo without any development. The simulated annual streamflow did not differ between the BMP and conventional turfgrass treatments. Simulations of average monthly flow predicted an increase of 0.14 m^3/s/month for the BMP and conventional turfgrass treatments when compared to the status quo (Fig. 4a).

The monthly streamflow increase (0.03 m^3/s/month) caused by the development of the watershed was removed from the BMP and conventional turfgrass treatments as shown in Fig. 4b. As shown in Fig. 4b, the BMP and conventional turfgrass treatments continued to increase streamflow due to the irrigation of the turfgrass. The constant irrigation kept the soil water of the HRUs containing the sod near field capacity resulting in more runoff than the status quo treatment.

The long term, 50-year simulations (1950–2000) of the BMP and conventional turfgrass treatments produced very similar streamflows in Mary's Creek at the watershed outlet (Fig. 5). Compared to the status quo, however, the BMP and conventional treatments increased streamflow 5.3% during the long-term simulation (Fig. 5). The

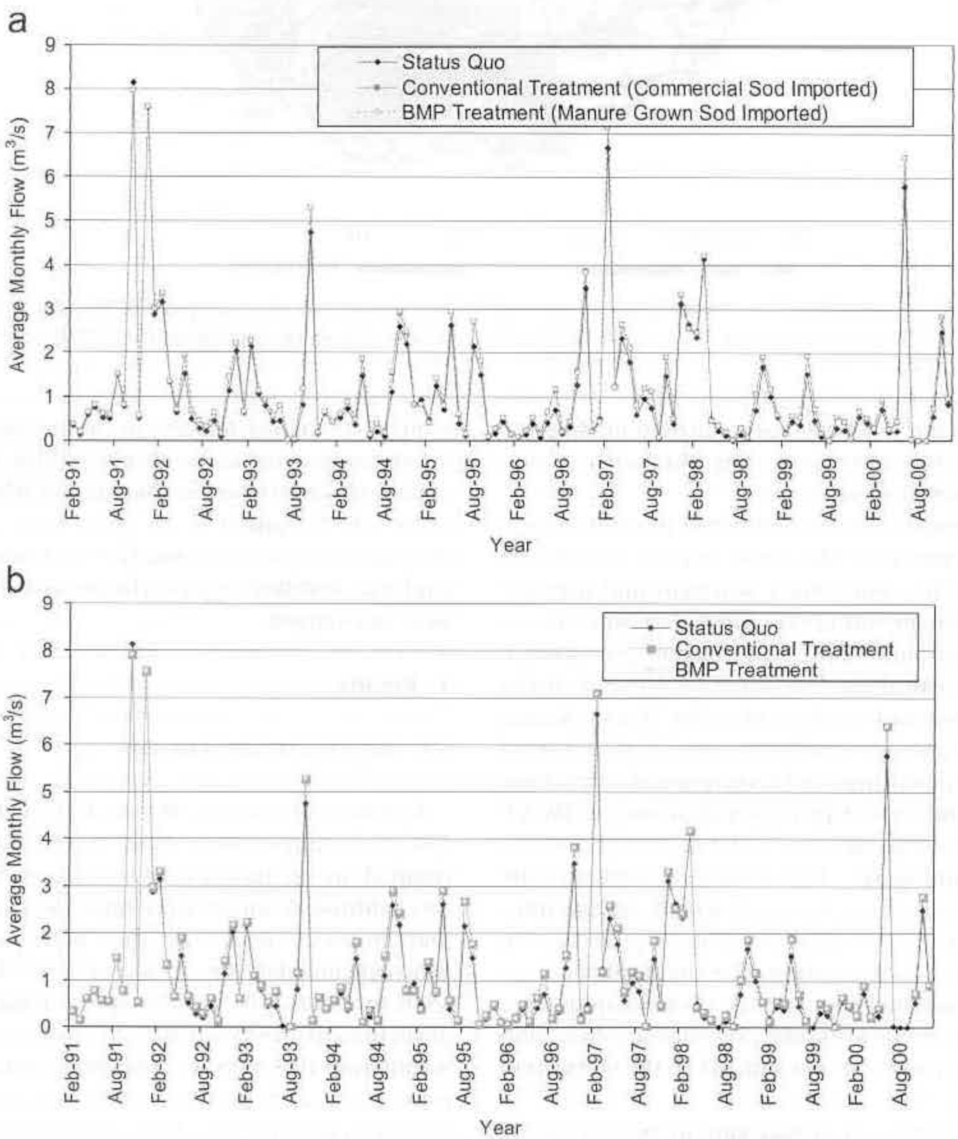


Fig. 4. The simulated average monthly flow for the three treatments at the outlet of the Mary's Creek watershed with the runoff from impervious urban surfaces (a) included in the BMP and conventional treatments, and (b) not included in the BMP and conventional treatments.

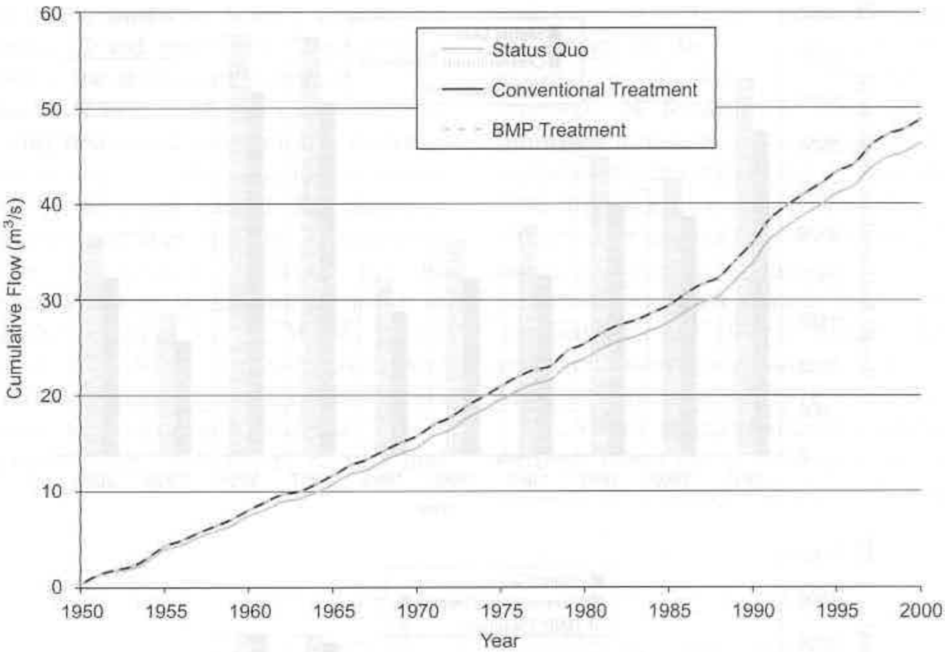


Fig. 5. The cumulative annual SWAT simulated streamflow for the three turfgrass treatments (status quo, conventional and BMP) at the outlet of the Mary's Creek watershed.

influence of the impervious surfaces in the Walsh Ranch development was not factored out of the long-term simulation and caused this long-term increase.

3.3. Sediment

The SWAT simulations indicated both the conventional and BMP turfgrass treatments contributed equally to the sediment loadings of Mary's Creek. The dense growth of turf plants and similar physical properties between manure-grown and conventionally grown turfgrass minimized sediment losses for both treatments (Vietor et al., 2004). Yet, the short term (10 year) simulation demonstrated that the BMP and conventional turfgrass treatments consistently produced greater sediment loads (135 metric tons cumulative) when compared to the status quo, which represented the undisturbed watershed (Fig. 6a). The principal difference between the imported turfgrass sod treatments and the status quo was erosion prior to turfgrass installation due to the increased impervious area within the Walsh Ranch development. The Walsh Ranch development (roads, buildings, sidewalks, driveways, etc.) was in place throughout the 10-year simulation. Similarly, the long-term 50-year simulation indicated that the BMP and conventional turfgrass treatments each contributed a total of 23 710 metric tons more sediment to the stream than the status quo or undisturbed watershed. As postulated for the short-term simulation, the additional sediment loading for both the BMP and conventional turfgrass treatments resulted from erosion before the turfgrass sod was transplanted on disturbed soil and from increased runoff due to the increased impervious areas within the watershed throughout the simulation.

The average sediment load (636 tons/year) caused by the development of the watershed was factored out of the short term simulation. This estimation method revealed the sediment loads contributed by just the turfgrass treatments (Fig. 6b). As shown in Fig. 6b, removing the influence of impervious surfaces in the development demonstrates that the turfgrass sod treatments reduced sediment loading to the stream when compared to the status quo treatment.

Monthly factors such as time of year and plant growth stage may have exerted greater influence on sediment loss in the status quo simulation than in the transplanted sod treatments. However, variation of annual rainfall was not significantly related to variation of sediment load for any of the simulated treatments. The adjusted R^2 values resulting from a regression analysis between variation of annual rainfall and the sediment loads predicted by SWAT for the turfgrass treatments in the Mary's Creek watershed were 0.208 for the transplanted sod treatments (conventional and BMP) and 0.168 for the status quo treatment.

3.4. Nutrients

The simulated increases in streamflow and sediment loading predicted for imports of manure-grown sod (BMP) and fertilizer-grown sod (conventional) were also reflected in the simulated differences in stream nutrient loading between the status quo, BMP and conventional turfgrass treatments in the long term simulations (Table 8).

The simulated in-stream organic N loading differed by 550 300 kg between the status quo and imports of fertilizer-grown (conventional) and manure-grown (BMP) turfgrass sod. Compared to the status quo, the in-stream nitrate-N loading was 42.5% greater for the BMP treatment and

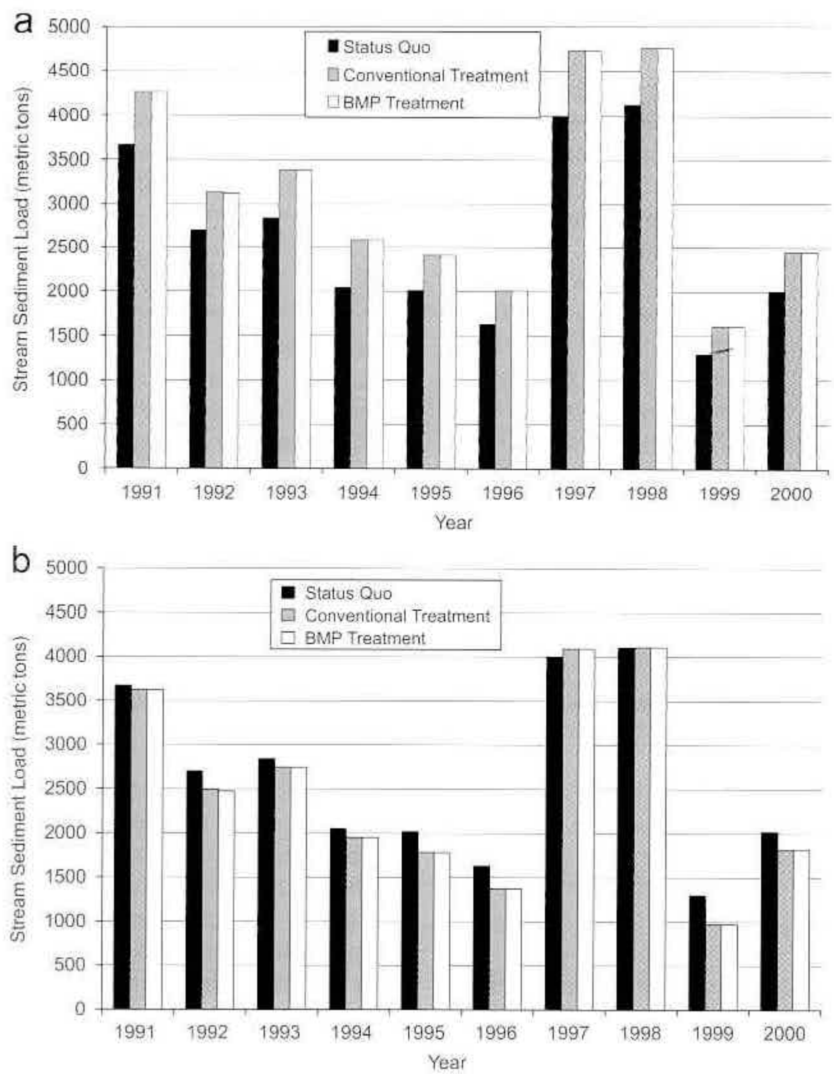


Fig. 6. The SWAT simulated annual stream sediment load for the three turfgrass treatments (status quo, conventional and BMP) at the outlet of the Mary's Creek watershed with (a) the sediment due to increases in runoff from urban impervious surfaces included in the BMP and conventional treatments and (b) the sediment due to increases in runoff from urban impervious surfaces removed from the BMP and conventional treatments.

Table 8
SWAT simulated in-stream nutrient loading at the outlet of the Mary's Creek watershed for the three turfgrass treatments (status quo, conventional and BMP) from 1950 to 2000

	Conventional treatment	BMP treatment	Status quo
Organic N (kg)	2 660 860	2 660 860	2 110 560
Nitrate-N (kg)	484 490	484 930	340 880
Total P (kg)	816 017	804 282	635 200

42.1% greater for the conventional treatment. A portion of the organic N imported with the manure-grown turfgrass sod of the BMP treatment was converted to nitrate-N over time, which led to slightly higher nitrate-N stream loading (0.09%) when compared to the conventional treatment. After imports of fertilizer-grown sod (conventional), total P loading to the stream was 28.5% greater than the status quo treatment. Similarly, predicted P loading for the BMP treatment was 26.6% larger than the status quo. The P fertilizer addition to the fertilizer-grown (conventional) sod increased total P stream loading by 1.5% compared to the BMP treatment.

The short-term simulation allowed a close comparison between the manure-grown (BMP) and fertilizer-grown (conventional) treatments that were imported into the watershed. A linear regression analysis was performed to assess variation of predicted annual sediment load to that of the predicted annual organic N load for the turfgrass treatments. The regression indicated predicted annual sediment load accounted for a significant portion of variation in organic N load among treatments which is to be expected as the organic fraction is attached sediments within the model. The adjusted R^2 values resulting from a regression analysis between in-stream annual sediment load and the annual organic N load predicted by SWAT for the turfgrass treatments at the outlet of Mary's Creek were 0.893 for the transplanted sod treatments (conventional and BMP) and 0.908 for the status quo treatment.

The simulated organic N load in Mary's Creek comparing the status quo and BMP and conventional turfgrass treatments is shown in Fig. 7a. The average in-stream organic N load (17 838 kg/year) caused by increased runoff from impervious surfaces in the development was factored out of the 10-year simulation for the BMP and conventional turfgrass treatments as shown in Fig. 7b. Removing the influence of the Walsh Ranch development revealed that both turfgrass treatments (conventional and BMP) reduced organic N loads to the stream when compared to the status quo treatment. This is due to the reduction of range, unfertilized pasture and forest land uses.

In contrast to organic N, the simulated nitrate-N load in the stream at the outlet increased significantly after the installation of the two turfgrass treatments due to inorganic N fertilization (Fig. 8a). The difference in simulated nitrate-N loads between the status quo and the turfgrass treatments peaked at approximately 30 000 kg in 1992 (Fig. 8a). Low stream flows (Fig. 4) combined with a reduction in application of inorganic N fertilizer lowered the stream nitrate-N load in the conventional and BMP turfgrass treatments during years 1995 and 1996. When summed over the 10-year period, the conventional turfgrass treatment contributed 1620 kg of nitrate N more to Mary's Creek than the BMP turfgrass treatment.

The SWAT model simulated applications of inorganic N fertilizer based on an N stress threshold of 0.9, where 0.0 indicates no plant growth due to N stress and 1.0 indicates

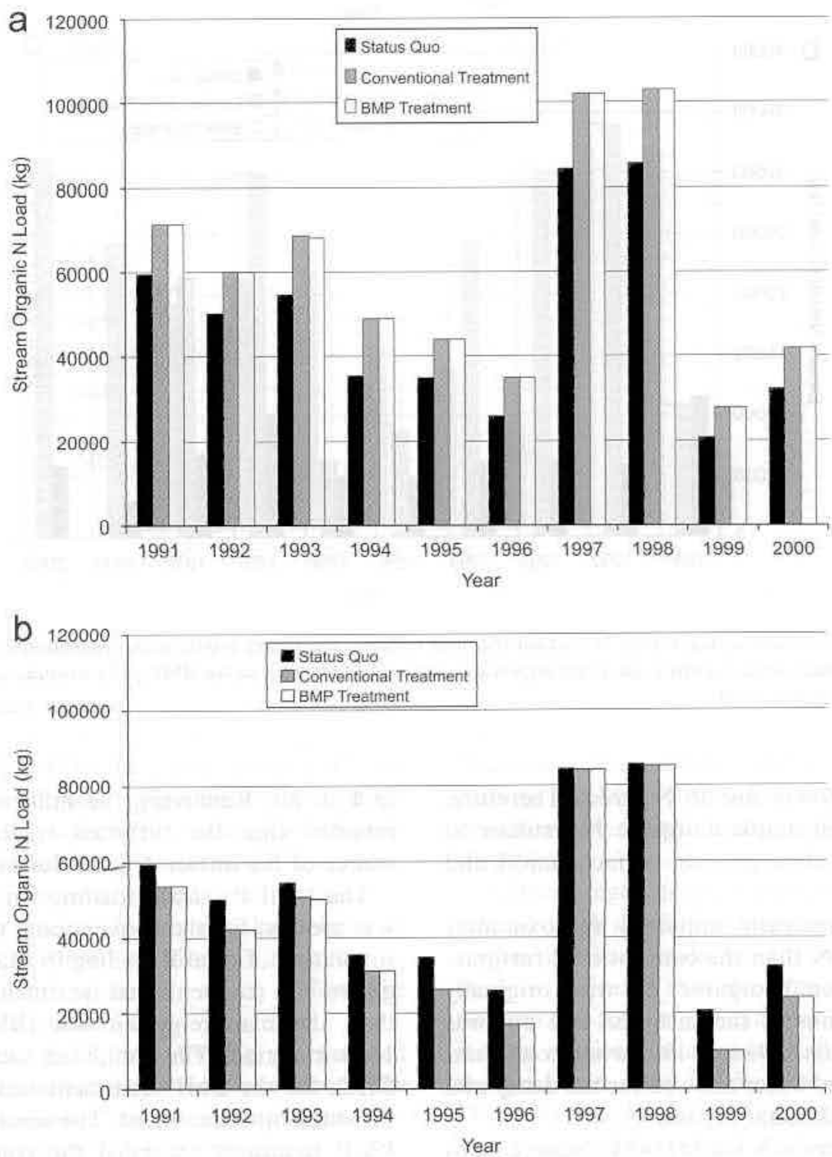


Fig. 7. The SWAT simulated in-stream annual organic N load for the three treatments (status quo, conventional and BMP) at the outlet of the Mary's Creek watershed with increases in runoff from urban impervious surfaces (a) included in the BMP and conventional treatments, and (b) removed from the BMP and conventional treatments.

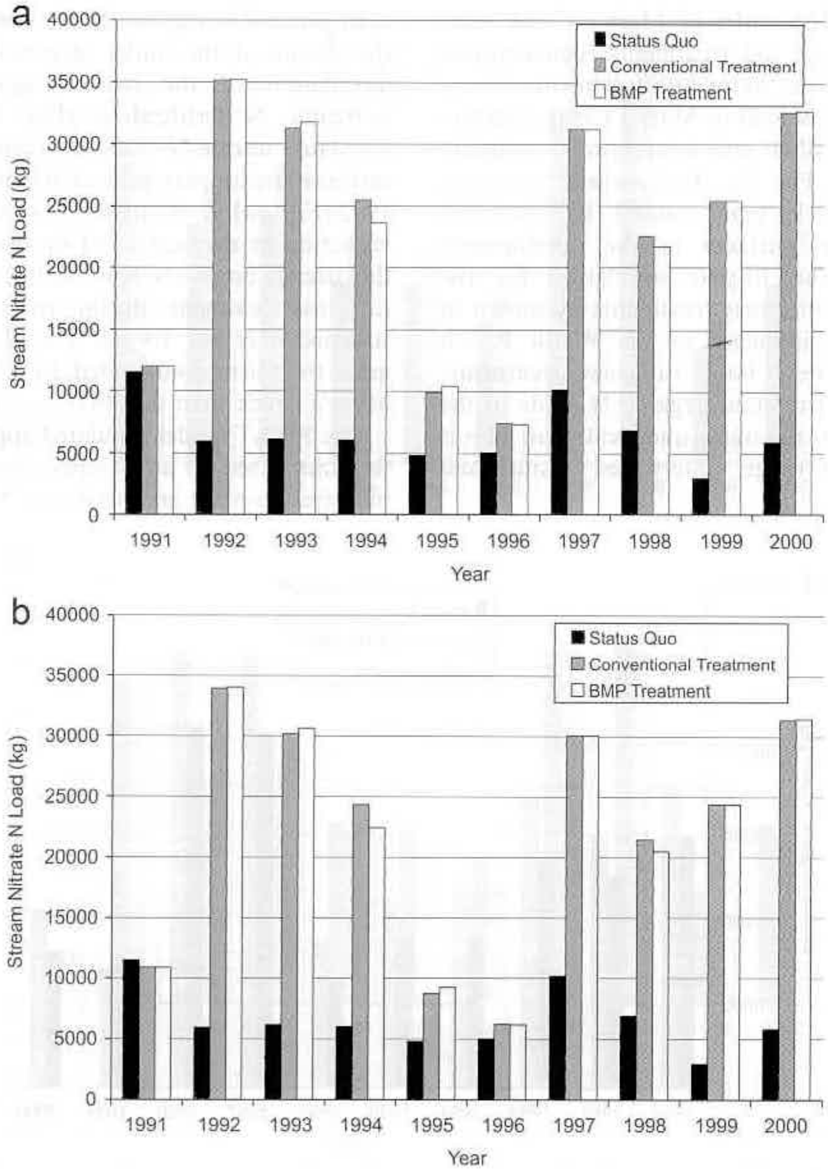


Fig. 8. The SWAT simulated in-stream annual nitrate N load for the three turfgrass treatments (status quo, conventional and BMP) at the outlet of the Mary's Creek watershed with increases in runoff from urban impervious surfaces (a) included in the BMP and conventional treatments and (b) removed from the BMP and conventional treatments.

no reduction in plant growth due to N stress. Therefore, the SWAT model applied ample inorganic N fertilizer to replace N losses due to plant growth, surface runoff and leaching.

The BMP turfgrass treatment imported approximately 170 kg/ha more organic N than the conventional turfgrass treatment. This additional organic N was originally associated with the humus in the turfgrass sod but was eventually released in years 1993 and 1995 when conditions such as the amount of soil water allowed for the decay and mineralization of the additional organic N.

The average stream nitrate-N load (1142 kg/year) caused by increased runoff from urban impervious surfaces in the development was factored out of the 10-year simulation for the BMP and conventional turfgrass treatments as shown in Fig. 8b. Removing the influence of the development revealed that the turfgrass treatments were the major source of the nitrate-N load due to lawn fertilization.

The total P stream loading for the 10-year simulation was greatest for the conventional treatment (Fig. 9a). The simulation of total P loading to Mary's Creek for fertilizer-grown sod (conventional treatment) was 14 843 kg greater than the manure-grown sod (BMP treatment) for the 10-year period. The simulated total P loading to Mary's Creek for the BMP treatment was 69 988 kg greater than the status quo treatment. The simulated total P load of the BMP treatment exceeded the conventional treatment in 1993 only (Fig. 9a) and may be explained as follows. Approximately, 11 kg/ha less organic P was imported with the BMP treatment when compared to the conventional

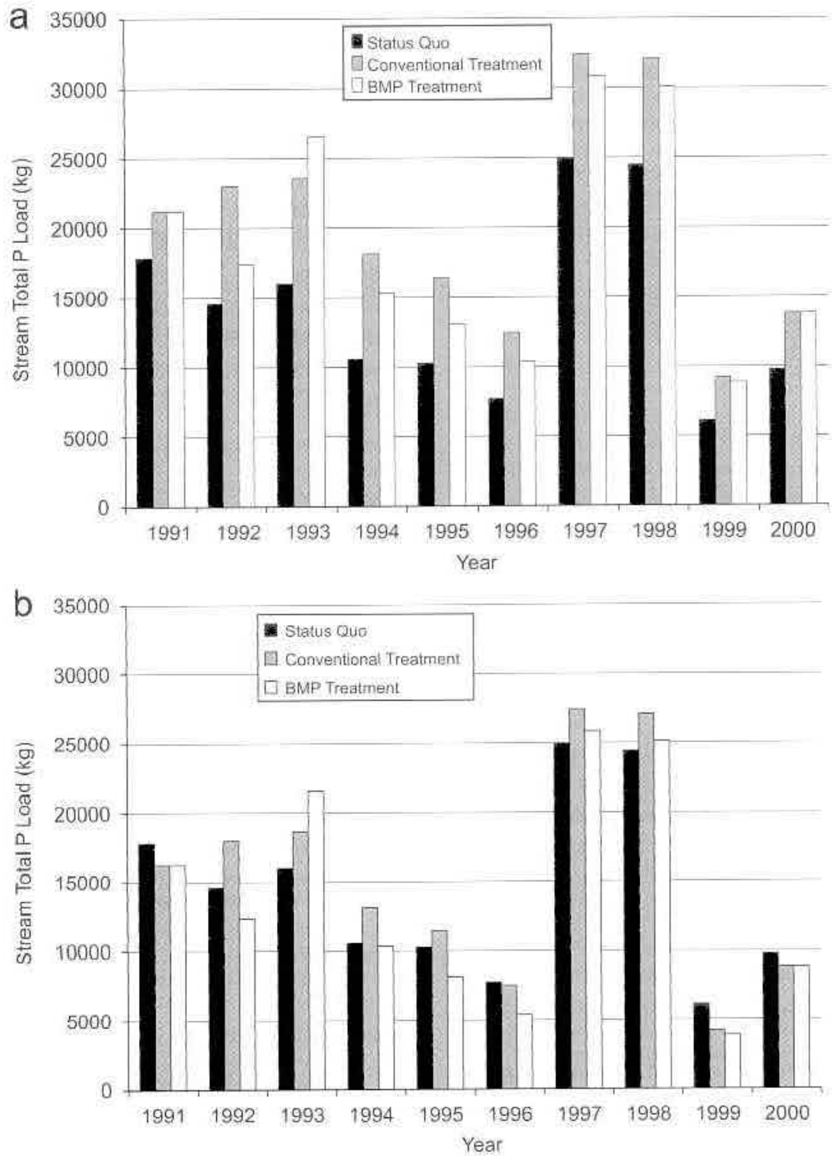


Fig. 9. The SWAT simulated in-stream annual total P load for the three turfgrass treatments (status quo, conventional and BMP) at the outlet of the Mary's Creek watershed with increases in runoff from urban impervious surfaces (a) included in the BMP and conventional treatments, and (b) removed from the BMP and conventional treatments.

treatment. In addition, 41 kg/ha more soluble P and 6 kg/ha more biomass P was imported by the BMP treatment. This additional soluble P was not lost immediately in the BMP treatment, but was released 3 years after the transplant when P was dissolved and transported through surface runoff events. Following this release in 1993, the total simulated P load to Mary's Creek for the BMP turfgrass treatment remained at or below the conventional turfgrass treatment.

The average in-stream total P load (4965 kg/year) caused by increased runoff from urban impervious surfaces in the development was factored out of the 10-year simulation for the BMP and conventional turfgrass treatments as shown in Fig. 9b. Removing the influence of the development revealed that the BMP turfgrass treatment reduced total P loading to Mary's Creek when compared to the status quo treatment. The conventional treatment increased total P loading to Mary's Creek compared to the status quo treatment after the influence of development was removed (Fig. 9b).

A linear regression was performed to relate variation of annual total rainfall amount to annual variation of in-stream nutrient loads for the turfgrass treatments. This analysis indicated that the variation of annual rainfall did not account for a significant portion of variation of the in-stream nutrient loads due to the treatments, except for the predicted nitrate-N load of the status quo treatment (Table 9). The low R^2 for the regression between nitrate-N and rainfall amount for the transplanted turfgrass sod treatments reaffirms that the in-stream nitrate-N loads for these treatments are related more to fertilizer application than streamflow or rainfall amount.

Table 9
The adjusted R^2 values resulting from a regression analysis between annual rainfall and in-stream nutrient loads predicted by SWAT for the turfgrass treatments

Treatments	Adjusted R^2 value		
	Organic N	Nitrate N	Total P
Conventional	0.149	−0.076	0.127
BMP	0.149	−0.076	0.141
Status quo	0.146	0.793	0.164

4. Conclusions

Through model simulation, the proposed turfgrass BMP was found to reduce total P loading to urban streams when compared to conventional commercial turfgrass sod imported and maintained with inorganic P fertilizer. The proposed turfgrass BMP was also found to reduce total P loading to the stream when compared to an undeveloped suburban watershed (the status quo treatment) when the effect of the Walsh Ranch development was factored out of the model results. Losses of total P were 1.1 times higher from the conventionally grown imported sod compared to the proposed turfgrass BMP, yet only 10% of the watershed area was influenced by this treatment. This reaffirms the findings of Vietor et al. (2004) in which commercially top-dressed sod losses of TDP were found to be three times greater than that of transplanted manure grown sod on small plots (100% effective area).

The turfgrass BMP increased in-stream nitrate N loading when compared to the status quo treatment due to increased N fertilizer applications. However, the increase was equivalent to the impact of importing conventional turfgrass sod grown with inorganic fertilizers. This additional in-stream nitrate-N load could be reduced by utilizing urban nutrient BMPs and by homeowner education of proper lawn nutrient application.

The SWAT model simulations indicate that the turfgrass BMP is an effective means of importing manure nutrients from impaired watersheds without raising the in-stream nutrient levels above conventional commercial turfgrass levels. In fact, the turfgrass BMP treatment reduced all in-stream nutrient levels except nitrate-N when compared to the status quo treatment after the effects of increased runoff from impervious surfaces in the development were removed. However, field studies should be conducted to confirm the amount of nutrient loss caused by the transplanted turfgrass sod grown with composted manure. Water quality sampling of a pilot suburban stream, such as Mary's Creek, after receiving turfgrass grown with composted manure would be useful for validating the amounts of nutrient lost from the turfgrass on the watershed scale.

Acknowledgments

The authors would like to thank Woody Frossard and the TRWD for their assistance and data, Nancy Sammons with the United States Department of Agricultural (USDA) Agricultural Research Service (ARS) at the BRC for her help with model development, and Steve Bednarz with the NRCS at the BRC for his assistance and data sharing. This research was supported by the Texas Advanced Technology Program (Project 000517-0073-2001), the USDA-CSREES (Project 2001-51130-11459), the Department of Interior USGS (Project 01HQGR0209), and TWRI.

References

Arnold, J.G., Allen, P.M., Muttiah, R.S., Bernhardt, G., 1995. Automated base flow separation and recession analysis techniques. Ground Water 33 (6), 1010–1018.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. Journal of the American Water Resources Association 34 (1), 73–89.

Baird, C., Ockerman, D., 1996. Characterization of Nonpoint Sources and Loadings to Corpus Christi Bay National Estuary Program Study area. CCBNEP-05. Corpus Christi National Estuary Program, Corpus Christi, TX.

Baldys III, S., Raines, T.H., Mansfield, B.L., Sandlin, J.T., 1998. Urban Stormwater Quality, Event-Mean Concentrations, and Estimates of Stormwater Pollutant Loads, Dallas-Fort Worth area, Texas, 1992–93. USGS Water-Resources Investigations Report 98-4158. United States Geological Survey, Austin, TX.

Bednarz, S.T., Srinivasan, R., 2002. Mary's and Sycamore Creeks Study—Phase I. BRC Report No. 02-11. Blackland Research Center, Temple, TX.

Bhuyan, S.J., Koelliker, J.K., Marzen, L.J., Harrington Jr., J.A., 2003. An integrated approach for water quality assessment of a Kansas watershed. Environmental Modeling and Software 18, 473–484.

Carrow, R.N., Waddington, D.V., Rieke, P.E., 2001. In: Turfgrass Soil Fertility and Chemical Problems: Assessment and Management. Ann Arbor Press, Chelsea, MI, pp. 206–208.

Chen, X., Harman, W.L., Magre, M., Wang, E., Srinivasan, R., Williams, J.R., 2000. Water quality assessment with agro-environmental indexing of non-point sources, Trinity River Basin. Applied Engineering in Agriculture 16 (4), 405–417.

Choi, I., Munster, C.L., Vietor, D.M., White, R.H., Richards, C.E., Stewart, G.A., McDonald, B., 2003. Use of turfgrass sod to transport manure phosphorus out of impaired watersheds. In: Proceedings of 2003 ASAE Total Maximum Daily Load (TMDL) Environmental Regulations II Conference, Albuquerque, NM. American Society of Agricultural Engineers, St. Joseph, MI, pp. 518–526.

DiLuzio, M., Srinivasan, R., Arnold, J.G., Neitch, S.L., 2002. ArcView Interface for SWAT2000. Blackland Research Center, Temple, TX. Available at 〈http://www.brc.tamus.edu/swat/doc.html〉. Accessed on 14 April 2006.

DiLuzio, M., Srinivasan, R., Arnold, J.G., 2003. A GIS-hydrological model system for the watershed assessment of agricultural nonpoint and point sources of pollution. Transactions in GIS, 2004 8 (1), 113–136.

Hall, M.H., 1999. Texas Turfgrass Research: 1999. Consolidated Progress Reports TURF-99-1 thru TURF-99-12. Texas Agricultural Experiment Station, College Station, TX.

Hanzlik, J.E., Munster, C.L., McFarland, A., Victor, D.M., White, R.H., 2004. GIS analysis to identify turfgrass sod production sites for phosphorus removal. Transactions of the ASAE 47 (2), 453–461.

Hauck, L.M., 2002. Investigations of Phosphorus Enrichment and Control in the Lake Waco-Bosque River Watershed: An Overview of Technical and Stakeholder Aspects. Publication No. PR0204. Texas Institute for Applied Environmental Research, Stephenville, TX.

He, C., 2003. Integration of geographic information systems and simulation model for watershed management. Environmental Modeling and Software 18, 809–813.

Kopp, K.L., Guillard, K., 2002. Clipping management and nitrogen fertilization of turfgrass: growth, nitrogen utilization, and quality. Crop Science 42, 1225–1231.

Kussow, W.R., 2004. Phosphorus runoff losses from lawns. Better Crops 88, 12–13.

MAWD, 2003. Legislative Program: 2000 Annual Meeting Resolutions. Minnesota Association of Water Districts, St. Paul, Minnesota. Available at ⟨http://www.mnwatershed.org/rso.htm⟩. Accessed on 13 July.

McFarland, A.M.S., Hauck, L.M., 1999. Existing Nutrient Sources and Contributions to the Bosque River Watershed. Publication No. PR9911. Texas Institute for Applied Environmental Research, Stephenville, TX.

Munster, C.L., Hanzlik, J.E., Vietor, D.M., White, R.H., McFarland, A.M.S., 2004. Assessment of manure phosphorus export through turfgrass sod production in Erath County, Texas. Journal of Environmental Management 73 (2004), 111–116.

Murray, J.J., 1981. Utilization of composted sewage sludge in sod production. In: Sheard, R.W. (Ed.), Proceedings of Fourth International Turfgrass Research Conference. University of Guelph, ON, p. 544.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models—Part I—A discussion of principles. Journal of Hydrology 10, 282–290.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2001. Soil and Water Assessment Tool User's Manual: Version 2000. Blackland Research Center, Temple, TX. Available at ⟨http://www.brc.tamus.edu/swat/doc.html⟩. Accessed on 14 April 2006.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., J.R. Williams, King, K.W., 2002. Soil and Water Assessment Tool Theoretical Documentation: Version 2000. Blackland Research Center, Temple, TX. Available at ⟨http://www.brc.tamus.edu/swat/doc.html⟩. Accessed on 14 April 2006.

Newell, C.J., Rifai, H.S., Bedient, P.B., 1992. Characterization of Nonpoint Sources and Loadings to Galveston Bay. GBNEP Report #15, Webster, TX.

Saleh, A., Arnold, J.G., Gassman, P.W., Hauck, L.M., Rosenthal, W.D., Williams, J.R., McFarland, A.M.S., 2000. Application of SWAT for the Upper North Bosque River Watershed. Transactions of the ASAE 43 (5), 1007–1087.

Santhi, C., Arnold, J.G., Williams, J.R., Hauck, L.M., Dugas, W.A., 2001. Application of a watershed model to evaluate management effects on point and nonpoint source pollution. Transactions of the ASAE 44 (6), 1559–1570.

Santhi, C., Arnold, J.G., Srinivasan, R., Williams, J.R., 2003. A modeling approach to evaluate the impacts of water quality management plans implemented in the Big Cypress Creek watershed. In: Proceedings of 2003 ASAE Total Maximum Daily Load (TMDL) Environmental Regulations II Conference, Albuquerque, NM. American Society of Agricultural Engineers, St. Joseph, MI, pp. 518–526.

Stewart, G.R., Munster, C.L., Victor, D.M., Arnold, J.G., McFarland, A.M.S., White, R.H., Provin, T., 2006. Water quality improvements in the Upper North Bosque River watershed due to phosphorus export through turfgrass sod. Transactions of the ASAE 49 (2), 357–366.

TCEQ (Texas Commission on Environmental Quality), 2002. An Implementation Plan for Soluble Reactive Phosphorus in the North Bosque River Watershed—For Segments 1226 and 1255. State of Texas, Austin.

TCEQ (Texas Commission on Environmental Quality), 2003. Composted Manure Incentive Project. State of Texas, Austin. Available at ⟨http://www.tnrcc.state.tx.us/water/quality/nps/compost/index.html⟩. Accessed on 10 February 2004.

Tripathi, M.P., Panda, R.K., Raghuwanshi, N.S., 2003. Identification and prioritization of critical sub-watersheds for soil conservation management using the SWAT model. Biosystems Engineering 85 (3), 365–379.

Tripathi, M.P., Panda, R.K., Raghuwanshi, N.S., 2004. Development of effective management plan for critical sub-watersheds using the SWAT model. Hydrological Processes, Special Issue: SWAT 2000 Development and Application 19 (3), 809–826.

TWRI (Texas Water Resources Institute), 2004a. Land Restoration: Fort Hood Rangelands & Training Areas. Texas Water Resources Institute, College Station, TX. Available at ⟨http://twri.tamu.edu/projects/LandRestoration.pdf⟩. Accessed on 10 February 2004.

TWRI (Texas Water Resources Institute), 2004b. Dairy Compost Utilization. Texas Water Resources Institute, College Station, TX. Available at ⟨http://twri.tamu.edu/projects/DairyCompost.pdf⟩. Accessed on 10 February 2004.

USDA-ARS (United States Department of Agriculture—Agricultural Research Service), 2003. The Milk Administrator's Report: Southwest Marketing Area. Vol. XXVIV(1).

USDA-NRCS (United States Department of Agriculture—Natural Resource Conservation Service), 1995. Soil Survey Geographic (SSURGO) Data Base: Data Use Information. Misc. Pub. No. 1527. Washington, DC.

USDA-SCS (United States Department of Agriculture—Soil Conservation Service), 1972. National Engineering Handbook. Hydrology Section 4, Chapters 4–10.

USDA-TAES (United States Department of Agriculture—Texas Agricultural Experiment Station), 1981. Soil Survey of Tarrant County, Texas.

USEPA (United State Environmental Protection Agency), 1983. Results of Nationwide Urban Runoff Program Volume 1—Final Report. Springfield, VA.

USGS (United States Geological Survey), 1999. The Quality of our Nation's Waters—Nutrients and Pesticides. USGS Circular 1225. Reston, VA.

Vietor, D.M., Griffith, E.N., White, R.H., Provin, T.L., Muir, J.P., Read, J.C., 2002. Export of manure P and N in turfgrass sod. Journal of Environmental Quality 31, 1731–1738.

Vietor, D.M., Provin, T.L., White, R.H., Munster, C.L., 2004. Runoff losses of phosphorus and nitrogen imported in sod or composted manure for turf establishment. Journal of Environmental Quality 33, 358–366.

Wickham, J.D., Wade, T.G., 2002. Watershed level risk assessment of nitrogen and phosphorus export. Computers and Electronics in Agriculture 37, 15–24.

Williams, J.R., Berndt, H.D., 1977. Sediment yield prediction based on watershed hydrology. Transactions of the ASAE 20 (6), 1100–1104.

Williams, J.R., Jones, C.A., Dyke, P.T., 1984. A modeling approach to determining the relationship between erosion and soil productivity. Transactions of the ASAE 27, 129–144.

ELSEVIER
Journal of Environmental Management 86 (2008) 246–261
Journal of Environmental Management
www.elsevier.com/locate/jenvman

Immersion frying for the thermal drying of sewage sludge: An economic assessment

Carlos Peregrina[a], Victor Rudolph[b,*], Didier Lecomte[a], Patricia Arlabosse[a]

[a]Laboratoire de Génie des Procédés des Solides Divisés (UMR 2392), Ecole des Mines d'Albi Carmaux, Route de Teillet, 81013 Albi CT Cedex 09, France
[b]Chemical Engineering, University of Queensland, St. Lucia QLD 4072, Australia

Received 11 February 2006; received in revised form 13 October 2006; accepted 7 December 2006
Available online 2 March 2007

Abstract

This paper presents an economic study of a novel thermal fry-drying technology which transforms sewage sludge and recycled cooking oil (RCO) into a solid fuel. The process is shown to have significant potential advantage in terms of capital costs (by factors of several times) and comparable operating costs. Three potential variants of the process have been simulated and costed in terms of both capital and operating requirements for a commercial scale of operation. The differences are in the energy recovery systems, which include a simple condensation of the evaporated water and two different heat pump configurations. Simple condensation provides the simplest process, but the energy efficiency gain of an open heat pump offset this, making it economically somewhat more attractive. In terms of operating costs, current sludge dryers are dominated by maintenance and energy requirements, while for fry-drying these are comparatively small. Fry-drying running costs are dominated by provision of makeup waste oil. Cost reduction could focus on cheaper waste oil, e.g. from grease trap waste.

Keywords: Drying; Frying; Oil; Sewage sludge

1. Introduction

1.1. Overview of thermal drying of sewage sludge in the European context

According to the European Environment Agency (EEA) (Christiannsen, 1999), thermal drying or the removal of moisture by evaporation will become an important process step in the disposal of sewage sludge. Indeed, due to a dramatic increase in the volume of wastewater treated in EU countries, about 10.7 millions tons of total dry solids of sewage sludge are generated every year (Bresters et al., 1997). In addition, from 2001 the progressive elimination of landfill as an acceptable method of sewage sludge disposal has resulted in an increase in material going to the two other approved disposal options, namely land spreading and incineration. For both of these methods thermal drying represents a potential intermediate unit operation, having advantage in providing volume reduction, stabilization through inactivation of pathogenic biological organisms and increasing the energy value of the dry material (Grüter et al., 1990; Hasserbrauck and Ermel, 1996).

Three general classes of sludge dryers are reported (Chen et al., 2002): convective or direct dryers, conductive or indirect dryers and mixed dryers.

In *direct drying*, hot gases from the combustion of an external fuel or the dried sludge itself are contacted with the dewatered cake in the dryer to evaporate the remaining water. Some examples of such equipments are drum, rotary, belt, spray and fluidized bed dryers. The heat transfer mechanism is predominantly convective and the evaporated water is mixed into the hot gas stream. For the *indirect drying*, (e.g. thin-film, discs or paddle dryers among others) the heat transfer is mainly conductive and occurs through the walls of the dryer. The heating medium, typically hot gas or thermal oil, is separated from the sludge and the evaporated water is not intermingled with the heating fluid. Although this requires more complicated

*Corresponding author. Tel./fax: +61 7 33 65 41 71.
E-mail addresses: peregrin@enstimac.fr (C. Peregrina), victorr@cheque.uq.edu.au (V. Rudolph).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.035

apparatus, the advantage of this approach is that the latent heat of the evaporated water may be recovered. Mixed dryers try to combine both conduction and convection to evaporate the water. Technical details are available in the literature (e.g. Arlabosse, 2001; Chen et al., 2002; Lowe, 1995; Ressent, 1998).

The thermal dryers currently applied for sewage sludge service have generally been adapted from industrial dryers used in other processes such chemicals, food or pharmaceuticals (Arlabosse, 2001), with mixed results. In most cases, thermal drying is not a cost effective operation because the energy demand is high and the product is a waste material requiring final disposal (Arlabosse, 2001). Drying consumes significant energy (Kudra, 2004), contributes to consumption of non-renewable natural resources and to greenhouse gas production. Efficiencies in energy consumption, e.g. though the use of heat pumps to transform low quality into high quality heat (IEA, 2004), are consequently desirable where possible. However, this is seldom applied in the design of waste thermal dryers which often lack proper optimization and integration in the processes they serve. Nevertheless, a few commercial systems show the potential for efficiency using such technologies, with low energy consumption ranging between 130 and 560 kWh ton^{-1} of evaporated water (Arlabosse, 2001). Moreover, the rapid increase of crude oil prices and concerns about climatic change caused by the greenhouse emissions, support a radical reconsideration of heat pumps in thermal drying (Kudra, 2004).

Two main technical difficulties are regularly referred to for sewage sludge drying (Lowe, 1995; Ressent, 1998):

1. All types of sewage thermal dryers produce vapor containing volatile organic compounds, creating an odor nuisance (Chavez, 2004; Hwang et al., 1995; Lambert et al., 2000; Nurul et al., 1998; Winter et al., 2004), toxicity hazard or even a real risk of explosion (Whipps, 2004).
2. Secondly, at a moderate total solids content typically between 40% and 50%, the sludge undergoes a *plastic phase* (Kudra, 2003; Lowe, 1995). This phase is characterized by exceptionally sticky behavior that complicates movement of the material through the dryers.

In summary, there are three main issues related to current thermal sludge drying processes which require innovative improvement: better energy efficiency; control of the VOCs in the exhaust gas for odor, toxic and explosive chemicals; and avoidance of the problems related to the plastic phase.

1.2. The fry-drying process for thermal drying of sewage sludge

1.2.1. Process concept

Frying is widely used in food processing as a cooking operation mainly to transform the sensory qualities of the foods (Moreira et al., 1999). Nevertheless, frying can also be a very effective drying method for a large variety of products (Vitrac, 2000). Operations using direct contact between the sludge and a liquid as a basis for drying have been reported in the literature, though incompletely (Bress et al., 1987; Kuntschar, 1996; Lue-Hing et al., 1996). The first systematic work about the drying of sewage sludge by immersion frying was presented by Pires da Silva (Silva et al., 2005, 2003). These reported experimental tests which were carried out by immersing a cylinder (about 40 mm length × 20–26 mm diameter) of municipal sewage sludge into bath containing 5 L soybean oil which was maintained at temperatures between 168 and 213 °C, well above the water boiling point. These conditions provided a dried sludge with <5% moisture in about 600 s. Moreover, due to the impregnation of oil the lower heating value (LHV) of fry-dried sludge was 24 MJ/kg which is significantly higher than air-dried sludge with comparable water content (14 MJ/kg).

Peregrina et al. (2006a) have also done similar experiments using waste cooking oils instead of vegetable oils, with a view to improving both the economic and environmental outcomes.

The mass and transfer phenomena occurring during the fry-drying of sewage sludge were characterized (Peregrina et al., 2006a) by four drying stages similar to those involved in frying of foods (Farkas and Hubbard, 2000). There is an initial period of sample heating, followed by boiling, initially at the surface but then proceeding as a front into the interior of the sample. During the third stage, when the water is depleted, oil penetrates into the material and finally there may be a period during which the material undergoes a phase change. The studies reported that the optimal frying temperature is 140–160 °C and that higher rates of vaporization are achieved for smaller individual samples. The waste cooking oil acts both as heating medium during the process and to improve the energy value of the material in the final fried product. The process naturally leads to co-disposal by incineration of these two waste materials.

1.2.2. Technical performance of the fry-drying process

Fry-drying offers many advantages over conventional dryers for processing sewage sludge. It provides for simple direct drying, with the product in direct contact with the heating medium (i.e. frying oil); avoids the *plastic phase* related problems; and the high temperatures completely hygienize the product, reducing handing issues. Moreover, the vapor removed is essentially evaporated water, so the latent heat is easy to recover by condensation.

A significant technical advantage of fry-drying lies in the rapidity of the process. Table 1 compares some typical dryers on the basis of their characteristic heat transfer resistances: the higher the resistance the slower the drying. Convective dryers have thermal resistances an order of magnitude higher than fry-drying and thin layer agitated contact dryers. The latter are mechanically and operation-

Table 1
Limiting thermal resistances for different drying systems

Type of dryer	Limiting mechanisms	Thermal resistance ($m^2\,K\,W^{-1}$)	Reference
Belt dryer	Convection (fluid-to-bed)	$1–3 \times 10^{-2}$	(Léonard, 2002)
Spouted bed dryer	Convection (wall-to-bed)	$0.8–2 \times 10^{-2}$	(Freitas and Freire, 1993)
Paddle dryer	Contact (wall-to-particle)	4×10^{-3}	(Arlabosse, submitted)
Fry-dryer	Contact (oil-to-film)	$3.0–4.4 \times 10^{-3}$	(Peregrina et al., 2006a)
	Contact (oil-to-particle)	$4.5–3.4 \times 10^{-3}$	
Thin-layer dryer	Heating inertia in the wall	$0.5–5 \times 10^{-4}$ ($\xi_W > 1\,kg\,kg^{-1}$)	(Carrère-Gée, 1999)
	Contact (wall-to-particle)	3×10^{-3} ($\xi_W < 1\,kg\,kg^{-1}$)	

ally complex compared with fry-drying. Moreover, there are advantages associated with oil impregnation in the final dried sludge: the higher energy value facilitates incineration; the oil eliminates dusting problems of the dry material; and the physical and chemical changes—taking place during the frying—facilitate storage and transportation.

Oil impregnation makes the final product unsuitable for agricultural land spreading, but this is not especially limiting since most countries are moving to decrease land filling and land spreading of sewage sludge. This will inevitably promote incineration as the preferred disposal method (Christiannsen, 1999; Munck-Kampmann, 2001). As a result a variety of processes (e.g. Werther and Ogada, 1999) have been investigated for recovery of energy from waste incineration (Porteous, 2001), as an element for recycling in a truly integrated waste management hierarchy.

Considering the oil requirement, the food safety problems in 1999 in Europe essentially closed the market for used cooking oils for use in animal feed (Mauvieux et al., 2001) Thus, these food industry by-product oils are becoming increasingly available. From a technical point of view, practically any oil can be used as fry-drying oil, although various economic and environmental issues need to be met. As examples, the high heavy metal content of motor waste oils or the high costs of virgin vegetable oils present drawbacks, which might preclude their utilization in a sewage sludge fry-drying process. However, the grease and oils that are skimmed from the top of the tanks during the pre-treatment and primary treatment from a WWTP might be suitable with some conditioning, and offer the convenience of being collected in the same location as the sludge.

1.2.3. Economic performance of the fry-drying process

The fry-drying process provides environmental and technical advantages, but must also be economic if it is to be accepted in an industrial context. This paper seeks to develop an economic assessment of the process operating at the conditions suggested by the laboratory experiments reported by Peregrina et al. (2006a). The process was simulated using the commercial software package, ASPEN 11.1 (Aspen Technology Inc.) as a basis for the economic evaluation. Three variants of the basic process are examined, based on different energy recovery strategies: a simple condenser to recover low value energy contained in the condensate; a closed heat pump; and an open heat pump. The processes including heat pumps recycle high value energy.

2. Theory and calculation

2.1. Basis of the assessment

The size of plant and context require definition for an economic assessment. The drying capacity depends on the WWTP conditions, product quality, integration into existing plant, the opportunity for using waste heat and economic factors. In this study, the dryer capacity was chosen based on three considerations:

1. Fry-drying is best suited to situations when the final disposal is aimed at incineration of dried material. This leads to the fryer design requirements described by Peregrina et al. (2006b) where mechanically dewatered sewage sludge with 19% total solids is fry-dried until its auto-thermal composition is reached. This is the point at which the dried sludge has just sufficient energy content to sustain its own incineration and minimizes the energy that is necessary for the water drying process. This corresponds to 3.7 kg of water per kg of total bone dry (indigenous) solids and 0.35 kg of waste oil per kg of indigenous total solids, and a LHV of $4.11\,MJ\,kg^{-1}$.
2. Most thermal dryers of sewage sludge work intermittently (Hasserbrauck and Ermel, 1996; Ressent, 1998, 1999), although convective dryers are sometimes in continuous operation (Grüter et al., 1990; Ressent, 1999). This simulation assumes 12 hours per day operation (i.e. one and a half shifts in operation and half a shift for shutdown, startup and handover).
3. The total dry solids processing rate is arbitrarily taken as $1\,ton\,h^{-1}$ on a total dry solids basis. This leads to an annual throughput of 4380 tons, corresponding to a treatment capacity of 219,000 equivalent people. The annual per capita average requirement is estimated as 20 kg total dry solids (Munck-Kampmann, 2001). This represents a common size for the facilities equipped with a thermal dryer (Gross, 1993; Grüter et al., 1990;

Hasserbrauck and Ermel, 1996; Ressent, 1999). The facility would consume 1533 tons of oil per year. For context, Sud Recupération, a waste oil recycler in the South of France and one of the major recyclers in the French market, collects some 6000 tons of waste oil per year, of the estimated total 30,000 tons annually discarded in this region (Sud-Recupération, 2003).

2.2. Economic assessment

The capital cost estimation is done to the "study estimate" or +30% to −20% accuracy level (Turton et al., 2003), being based on a process flow diagram and rough sizing of major process equipment. This preliminary evaluation level may not accurately reflect the final profitability of a plant but is typically applied as a tool for comparison of several processes (Zhang et al., 2003b). This study limits its economic assessment to the evaluation of a total capital investment cost (C_{TC}) and a total manufacturing cost (C_{TM}) (Turton et al., 2003; Ulrich, 1984).

Following the methods described in Turton et al. (2003), C_{TC} is given by the sum of the fixed capital cost (C_{FC}), which represents all the costs associated with building a facility and the working capital cost (C_{WC}), which is the capital required to start up the plant and finance the first few months of operation. The latter is obtained as a fraction (15%) of the C_{FC}. Fixed capital cost consists of three parts:

1. *Total bare module capital cost* (C_{BM}) which is the sum of the costs of the major process equipment;
2. *Contingencies and fees* (C_{CF}) covering unaccounted costs and contractor fees, taken as a fraction of the C_{BM} (18% was used in this study);
3. *Cost associated with auxiliary facilities* (C_{AC}) such as the purchase of land, installation of electricals, utilities etc. This is also applied as a simple multiplier. A value of 30% of the sum of the total C_{BM} and the C_{CF} was used based on the value recommended in Turton et al. (2003)

The first requirement is to determine C_{BM} for which the equipment sizing and materials flow rates were calculated from the ASPEN PLUS^{TM} 11.1 simulation software (ASPEN Tech Inc., USA). Basic equipment prices were obtained from a variety of sources including Peters and Timmerhaus (2001), Turton et al. (2003) and Sud-Recupération (2003). These were updated to 2005 values using the preliminary March 2005 Chemical Engineering Plant Cost Index (CEPCI) (Vatavuk, 2002), where $I_{1990} = 357.6$, $I_{2001} = 397$, $I_{2003} = 402$ and $I_{2005} = 468.1$.

Total manufacturing cost (C_{TM}) is usually divided into three categories: direct manufacturing costs, indirect manufacturing costs and general expenses (Turton et al., 2003; Ulrich, 1984). Charges related to raw materials and labor are considered direct manufacturing costs, while items which are independent of the production rate, belong to the indirect manufacturing costs. Raw material and auxiliary facilities requirements are given by the simulated process streams.

The operator requirements were derived from the flow sheet and major equipment following the method proposed by Alkayat and Gerrard (1984).

General expenses include administrative and distribution costs as well as the research and development charges. Both indirect and general manufacturing expenses are calculated by factoring methods which are widely used in economic assessment (Turton et al., 2003; Ulrich, 1984; Zhang et al., 2003b) and presented in Table 7.

2.3. Process simulation

The procedures for process simulation require the definition of chemical components, selecting a thermodynamic model, determining plant capacity, choosing proper operating units and setting up input conditions such as flow rate, temperature, pressure, (Zhang et al., 2003a). The fry-dryer and the heat pumps which represent the main elements studied in this paper are elaborated further below.

2.3.1. Fry-drying unit

The unusual nature of the frying process, including the absence of thermal equilibrium between the frying oil and the water vapor and the non-conventional materials involved in the process, require the ASPEN PLUS^{TM} 11.1 predetermined blocks to be suitably configured to describe the heat and mass transfers occurring during frying. Consequently, a pseudo-unit operation was used to simulate and calculate the heat and mass balances (see Fig. 1).

According to experimental data (Peregrina et al., 2006a) fry-drying of sewage sludge optimally uses a frying temperature within the range 140–160 °C, and requires a residence time of about 100 s to reach the target autothermal dryness of the product dried sludge (Peregrina et al., 2006b). In this simulation, a frying temperature of 160 °C was assumed. Water evaporation and cooling of the fry-drying oil is simulated using the following strategy:

The sludge (stream 1 in Fig. 1) is defined—according to the ASPEN PLUS^{TM} 11.1 component description—as non-conventional component comprising a mixture of water and total dry matter characterized by its proximate, ultimate and sulfur analysis; the waste oil is also defined as non-conventional component.

The wet sludge [1] is fed into the flash separator (V-101A) where the water evaporation takes place. The Redlich–Kwong–Soave (RKS) equation of state was used to model this part of the process (Twu et al., 1998, 2002). Separately, an auxiliary block (V-101B) calculates the associated evaporative cooling and transfers this information to the flash separator in order to equilibrate the heat and mass balances. An external heat exchanger (E-101)

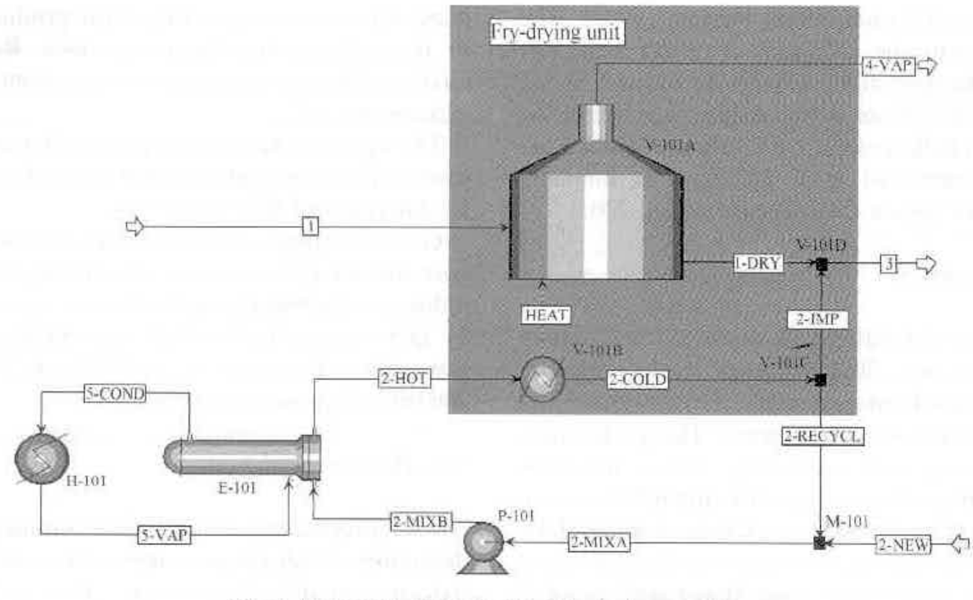


Fig. 1. Flow sheet of the simulated fry-drying operation.

condenses medium pressure steam ($P = 9$ atm (abs) and $T_{sat} = 449$ K) to provide the energy (Q1) for the drying and to raise the temperature of the makeup oil stream (stream 2-NEW in Fig. 1).

Two additional blocks, a splitter (V-101C) and a mixer (V-101D), simulate the oil impregnation into the partially dried sludge (stream 1-DRY in Fig. 1). This is achieved by separating some oil (2-IMP) from the recycling oil stream (2-COLD) using the splitter (V-101C) and then combining this with the dried sludge using mixer (V-101D) providing stream 3, the final oil impregnated dried sludge.

The model results allow the calculation of the volume of the dryer which holds the frying oil. Equipment associated with the dryer, such as conveyors for the sludge, a fan to remove the volatiles and a circulation pump to circulate oil through the drying vessel and heater, were sized for costing purposes according to empirical data provided by Stier (1996a,b,c), Perry and Green (1984) and Turton et al. (2003).

2.3.2. Simulation of the heat pumps

The heat pump recycles energy from the evaporated water stream to the evaporator. The great majority of heat pumps work on the principle of the vapor compression cycle (IEA, 2004). The main components in such a heat pump system (Fig. 3) are the compressors (C-201 and C-202), the expansion valve (C-203), heat exchangers namely an evaporator (E-201) and condensers (E-202 and E-203), and a recirculating working fluid.

Traditional working fluids for heat pumps were the chlorofluorocarbons, but their chlorine content and chemical stability are harmful to the global environment (Tsai, 2005). Consequently, the European Union has adopted an accelerated elimination schedule for these substances which requires them to be phased out by 2015 (Powell, 2002). As a result, blends of refrigerants or natural working fluids are preferred for new heat pumps applications (IEA, 2004; Riffat et al., 1997).

Among the possible working fluids suggested by the International Energy Agency (IEA), water was chosen for this simulation because it has favorable thermodynamic properties and is intrinsically safe. According to the IEA, water as a working fluid typically operates in the range of temperatures from 80 to 150 °C, although there are some examples where it can work at temperatures up to 300 °C. The major disadvantage with water is its low volumetric heat capacity which means that comparatively large and expensive compressors are required (IEA, 2004).

In order to be able to work within the range of temperatures where water is a useful working fluid, two options present themselves: operating the heat pump near the upper limit, i.e. ~150 °C and using an auxiliary energy source to provide the additional energy to the system; or operating at lower temperature, evaporating below atmospheric pressure and providing additional energy through the vacuum system. This latter process is rather difficult to assess economically, since no industrial references are available for vacuum fryers, even in food applications (Garayo and Moreira, 2002). As a result, only the first alternative has been investigated in this paper.

Within the heat pump, the water–steam equilibrium was calculated using the steam NBS thermodynamic model available in the software. The water mass flow was minimized with the design specification option by targeting the liquid fraction to 1 or 0 depending on the phase change. The exchange surface was determined using the simplified method of the HeatX block, with an appropriate constant overall heat transfer coefficient (Perry and Green, 1984; Sinnott, 1991). Finally, the compressors were calculated by the ASME method provided by the software, using a polytropic compression with an efficiency of 0.65 (Peters and Timmerhaus, 1991). The three energy recovery systems that are examined as case studies in this paper are described below.

2.4. Process design

2.4.1. PROCESS 1: fry-dryer with a condenser as an energy recovery system

PROCESS 1 (see Fig. 2) represents a typical externally heated continuous fryer like those widely used in the food industry (Stier, 1996a). In this configuration, the frying oil is passed through a heat exchanger heated with steam vapor (stream 5-VAP in Fig. 2; $P = 9$ atm (abs), $T = 449.6$ K) and returns directly to the fryer. The frying oil input and output temperatures are 433 and 413 K, respectively. Exhaust vapors (4-VAP) are condensed (4-COND) at 368 K, in the heat exchanger (H-201) using cooling water (6B). Latent energy is not recovered although the hot condensate is potentially available for energy.

To upgrade heat from the exhaust vapor, so that it can be recovered in the process vapor compression may be applied (IEA, 2004) as described in PROCESSes 2 and 3.

2.4.2. PROCESS 2: fry-dryer with a closed heat pump as an energy recovery system

In the closed loop heat pump provided in PROCESS 2 (Fig. 3), water vapor (6A) is formed at $P = 0.7$ atm (abs) and $T_{sat} = 366.7$ K, in the condenser (E-201) using heat supplied by the fry-drying exhaust vapors (4-VAP). This vapor stream is then compressed in the heat pump at $P =$ 4.0 atm (abs) ($T_{sat} = 418.7$ K). Two stages (C-201 and C-202) reduce the maintenance costs of the compression (Turton et al., 2003; Ulrich, 1984). Due to the limited temperature range for the working fluid, the energy recovered in the condenser of the closed loop (E-202) provides only a part (Q4) of the required total heat duty (Q1). Q4 preheats the frying oil (2-PREHOT) up to a temperature of 424 K. The remainder of the heat (Q3) will be provided by a fuel gas powered steam generator (H-101), as in PROCESS 1.

In the closed loop, the high pressure vapor stream (6-C) condenses in the oil pre-heater (E-202) and then is expanded to $P = 0.9$ atm (abs) across a valve (C-203). The expansion, involves the re-vaporization of a fraction of the liquid water. In order to increase the energy recovery capacity of the heat pump, it was proposed to re-condense the working fluid by pre-heating the incoming sewage sludge (1) using an additional heat exchanger (E-203).

2.4.3. PROCESS 3: fry-dryer with an open heat pump for energy recovery (mechanical vapor compression)

PROCESS 3 (see Fig. 4) is formed of the same elements as PROCESS 2, with the difference that, in open systems, exhaust vapor from the fry-drying operation (4-VAP) is directly compressed (to $P = 4$ atm (abs) and $T_{sat} = 479$), before being condensed (4-CONDA) to release heat. This heat is transferred to the frying oil (2-MIXB) through a heat-exchanger (E-202). As in PROCESS 2, the open system temperature limits constrain the system to provide only a part of the required heat duty. Consequently, a fuel gas powered steam generator (H-101) is provided to increase the frying oil temperature to 433 K.

To improve the energy recovery efficiency, it is proposed as in PROCESS 2, that the vapor and liquid water mixture (4-MIX) is passed through a sewage sludge pre-heater (E-203) before being released as hot liquid (4-CONDB) to be used elsewhere in the plant.

3. Results

3.1. Technical process comparisons

The basis and assumptions for the processes examined in this study are summarized in Table 2. The most basic technical comparisons between different processes are on the basis of the number of major processing units in each process (Table 3), the characteristics of the unit operations (Table 4) and the utilities consumed (Table 6).

The simplest fry-drying process is given by the PROCESS 1, which requires only five main items of equipment versus nine and eight for the PROCESSes 2 and 3, respectively (Table 3). The differences in the number of operations comprise the compressors and the supplementary heat exchangers that are required in the heat pumps. The cost of simplicity in PROCESS 1 is lower energy efficiency, requiring the largest heat exchanger and gas powered steam generator (Table 4). The advantage, besides lower capital cost is in personnel, with PROCESS 1 requiring only six operating staff, as opposed to seven for PROCESSes 2 and 3 (Alkayat and Gerrard, 1984).

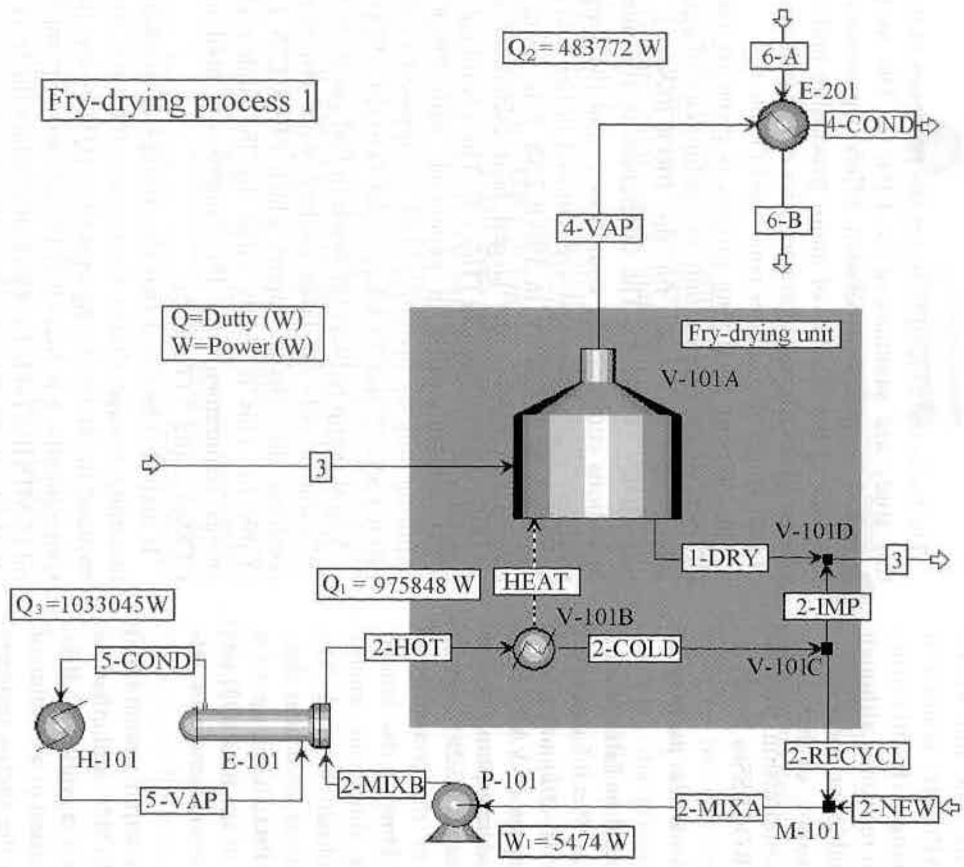
A heat pump reduces the needs in fuel gas by more than 50% and avoids the large cooling water requirement to condense the exhaust vapors which PROCESS 1 creates (Table 6). This is partly offset by the higher electrical energy requirements for the compressors used in PROCESSes 2 and 3 (Table 6).

It may also be noted that the strategy of pre-heating the incoming sewage sludge (1) reduces the total heat duty required in the fry-drying operation (Q1) by less than 5%. Consequently, a reduction in the amount of pumped waste oil (2-MIXB) (Table 6), which determines the fryer volume (V-101A, B, C, D) as well as the pump capacity (P-101) (Table 4), has only a small impact.

3.2. Economic comparisons

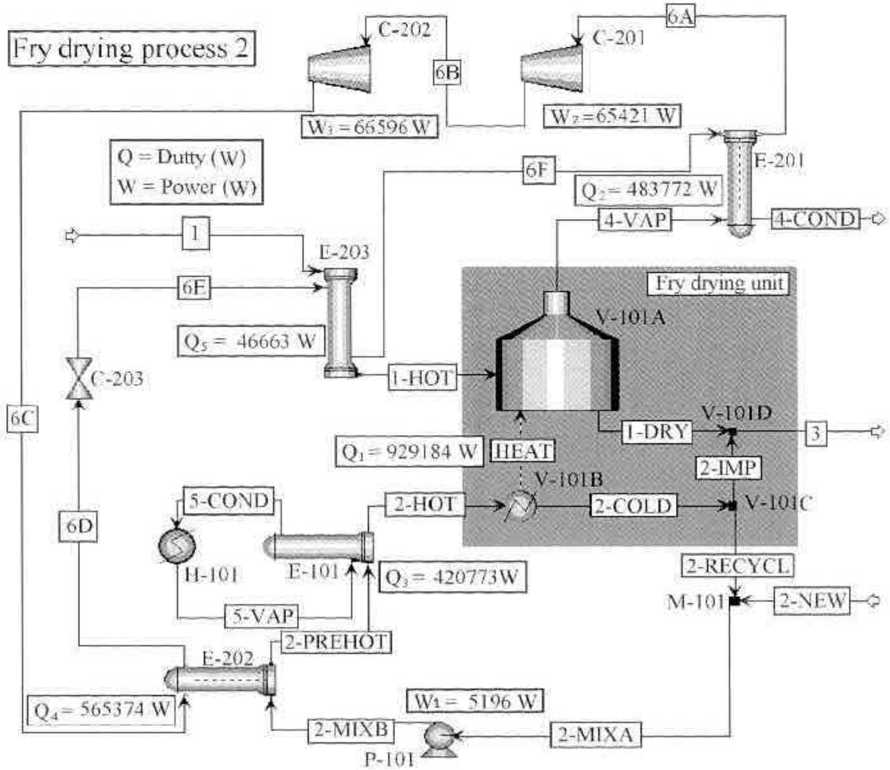
3.2.1. Fixed capital cost

The quantitative use of the capital estimates (Table 5) needs to be qualified by the large errors associated with the simple scaling methods used. Nevertheless, the relative costs may be taken as indicative, since the methodology is consistent. This study finds the fixed capital cost per ton of processed dry solids is estimated as US\$324, US\$275 and US\$232 for PROCESSes 1, 2 and 3, respectively.



Steam ID	1	1-DRY	2-COLD	2-HOT	2-IMP	2-MIXA	2-MIXB	2-NEW	2-RECYCL	3	4-COND	4-VAP	5-COND	5-VAP	6A	6B
From		V-101A	V-101B	E-101	V-101C	M-101	P-101		V-101C	V-101D	E-201	V-101A	E-101	H-101		E-201
To	V-101A	V-101D	V-101C	V-101B	V-101D	P-101	E-101	M-101	M-101			E-201	H-101	E-101	E-201	
Temperature (K)	283.15	375.62	413.15	433.15	413.15	411.96	411.96	283.15	413.15	375.62	368.15	375.62	449.61	449.61	303.15	313.15
Pressure (atm (abs))	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	9.000	9.000	1.000	1.000
Mass flow ($kg \cdot hr^{-1}$)	5263.000	4500.000	68409.56	68409.56	350.000	68409.56	68409.56	350.000	68059.56	5050.000	563.000	563.000	1612.484	1612.484	30798.330	30798.330
Volume flow ($L \cdot sec^{-1}$)	1.376	1.266	25.028	25.028	0.128	25.028	25.028	0.128	24.900	7.585	0.163	263.501	0.539	97.177	8.592	8.625
Component mass flow ($kg \cdot hr^{-1}$)																
Total dry solids	1000.000	1000.000	0	0	0	0	0	0	0	1000.000	0	0	0	0	0	0
Water	4263.000	3700.000	0	0	0	0	0	0	0	3700.006	563.000	563.000	1612.484	1612.484	30798.330	30798.330
RCO	0	0	68409.56	68409.56	350.000	68409.56	68409.56	350.000	68059.56	350.000	0	0	0	0	0	0

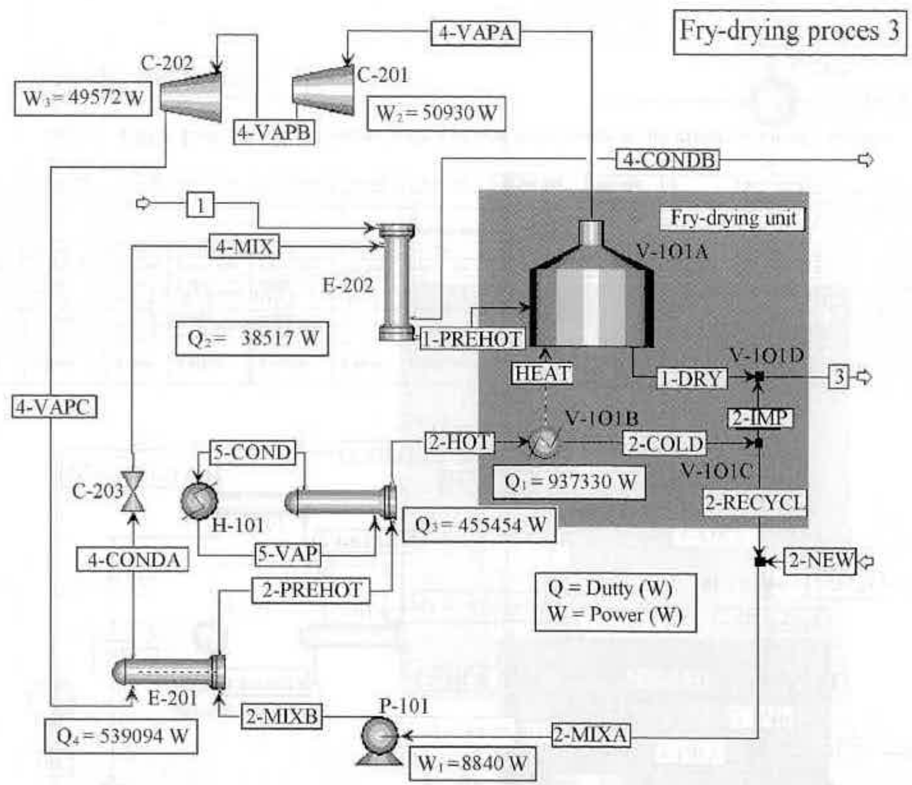
Fig. 2. PROCESS 1 PFD: fry-dryer with a condenser as an energy recovery system.



Steam ID	1	1-DRY	1-HOT	2-COLD	2-HOT	2-IMP	2-MIXA	2-MIXB	2-NEW	2-PREHOT	2-RECYCL
From		V-101A	E-203	V-101B	E-101	V-101C	M-101	P-101		E-202	V-101C
To	E-203	V-101D	V-101A	V-101C	V-101B	V-101D	P-101	E-202	M-101	E-101	M-101
Temperature (K)	283.15	375.62	291.85	413.15	433.15	413.15	411.73	411.73	283.15	422.07	413.15
Pressure (atm (abs))	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Mass flow (kg hr^{-1})	5263.000	4700.000	5263.000	65654.580	65654.580	350.000	65654.580	65654.580	350.000	65654.580	65304.580
Volume flow (L sec^{-1})	1.376	1.266	1.377	24.020	24.020	0.128	24.020	24.020	0.128	24.020	23.892
Component mass flow (kg·hr^{-1})											
Total dry solids	1000.000	1000.000	1000.000	0	0	0	0	0	0	0	0
Water	4263.000	3700.000	4263.000	0	0	0	0	0	0	0	0
RCO	0	0	0	65654.580	65654.580	350.000	65654.580	65654.580	350.000	65654.580	65304.580

Steam ID	3	4-COND	4-VAP	5-COND	5-VAP	6A	6B	6C	6D	6E	6F
From	V-101D	E-201	V-101A	E-101	H-101	E-201	C-201	C-202	E-202	C-203	E-203
To			E-201	H-101	E-101	C-201	C-202	E-202	C-203	E-203	E-201
Temperature (K)	375.62	368.15	375.62	449.61	449.61	366.24	525.82	682.09	418.36	366.24	362.00
Pressure (atm (abs))	1.000	1.000	1.000	9.000	9.000	0.700	1.900	4.000	4.000	0.7	0.7
Mass flow (kg hr^{-1})	5050.000	563.000	563.000	808.744	808.744	557.530	557.530	557.530	557.530	557.530	557.530
Volume flow (L sec^{-1})	7.585	0.163	263.503	0.252	47.687	364.670	193.943	119.662	0.168	35.595	0.160
Component mass flow (kg·hr^{-1})											
Total dry solids	1000.000	0	0	0	0	0	0	0	0	0	0
Water	3700.000	563.000	563.000	808.744	808.744	557.530	557.530	557.530	557.530	557.530	557.530
RCO	350.000	0	0	0	0	0	0	0	0	0	0

Fig. 3. PROCESS 2 PFD: fry-dryer with a closed heat pump as an energy recovery system.



Steam ID	1	1-DRY	1-HOT	2-COLD	2-HOT	2-IMP	2-MIXA	2-MIXB	2-NEW	2-PREHOT
From		V-101A	E-202	V-101B	E-101	V-101C	M-101	P-101		E-201
To	E-202	V-101D	V-101A	V-101C	V-101B	V-101D	P-101	E-201	M-101	E-101
Temperature (K)	283.15	375.62	288.45	413.15	433.15	413.15	411.74	411.74	283.15	421.50
Pressure (atm (abs))	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Mass flow ($kg \cdot hr^{-1}$)	5263.000	4700.000	5263.000	66124.310	66124.310	350.000	66124.310	66124.310	350.000	66124.310
Volume flow ($L \cdot sec^{-1}$)	1.376	1.266	1.377	24.192	24.192	0.128	24.192	24.192	0.128	24.192
Component mass flow ($kg \cdot hr^{-1}$)										
Total dry solids	1000.000	1000.000	1000.000	0	0	0	0	0	0	0
Water	4263.000	3700.000	4263.000	0	0	0	0	0	0	0
RCO	0	0	0	66124.310	66124.310	350.000	66124.310	66124.310	350.000	66124.310

Steam ID	2-RECYCL	3	4-CONDA	4-CONDB	4-MIX	4-VAPA	4-VAPB	4-VAPC	5-COND	5-VAP
From	V-101C	V-101D	E-201	E-202	C-203	V-101A	C-201	C-202	E-101	H-101
To	M-101		C-203		E-202	C-201	C-202	E-201	H-101	E-101
Temperature (K)	413.15	375.62	418.36	375.62	375.62	375.62	499.03	616.15	449.62	449.62
Pressure (atm (abs))	1.000	1.000	4.000	1.000	1.000	1.000	2.200	4.000	9.000	9.000
Mass flow ($kg \cdot hr^{-1}$)	65774.310	5050.000	563.000	563.000	563.000	563.000	563.000	563.000	851.658	851.658
Volume flow ($L \cdot sec^{-1}$)	24.064	7.912	0.170	0.163	21.423	253.501	160.067	108.896	0.265	50.803
Component mass flow ($kg \cdot hr^{-1}$)										
Total dry solids	0	1000.000	0	0	0	0	0	0	0	0
Water	0	3700.000	563.000	563.000	563.000	563.000	563.000	563.000	851.658	851.658
RCO	65774.310	350.000	0	0	0	0	0	0	0	0

Fig. 4. PROCESS 3 PFD: fry-dryer with an open heat pump as an energy recovery system.

Table 2
Basic process simulation conditions

Condition	Assumption	Reference
Plant capacity	4380 tons of total dry solids year^{-1}	
WWTP size	219,000 people	(Munck-Kampmann, 2001)
Thermodynamic model	Redlich Kwong Soave, NBS steam tables and ASME method to calculate polytropic compressors	Aspen Plus 11.1
Residence time	100 s	(Peregrina et al., 2006b)
Range of the frying temperature	413–433 K	(Peregrina et al., 2006a)
Auto-thermal composition of the final fry-dried sludge	$\xi_W = 3.7$ kg water per kg total dry solid	(Peregrina et al., 2006b)
	$\xi_{RCO} = 0.35$ kg oil kg total dry solid	
Final temperature of the condensed fry-drying vapors	368 K	(Gross, 1993)
Overall heat transfer coefficient	$U_0 = 230\ W\ m^{-2}\ K^{-1}$ (sludge–steam)	(Yamahata and Izawa, 1985)
	$U_0 = 250\ W\ m^{-2}\ K^{-1}$ (RCO–steam)	(Perry and Green, 1984)
	$U_0 = 1250\ W\ m^{-2}\ K^{-1}$ (steam–water)	(Perry and Green, 1984)
Medium pressure steam	P = 9 atm (abs)	(Ressent, 1999, Turton et al., 2003)
Operating working fluid in the heat pump	Water between 253 and 423 K	(IEA, 2004)
Efficiencies	Compressor = 0.65	(Peters and Timmerhaus, 1991)
	Pump = 0.80	(Peters and Timmerhaus, 1991)
	Electric motor = 0.90	(Peters and Timmerhaus, 1991)
Oil loop pressure losses	20 psig~1.4 atm	(Turton et al., 2003)
Cooling water operating temperatures	303–313 K	(Turton et al., 2003)
Natural gas lower heating value	47 MJ kg^{-1}	(Poulsen and Hansen, 2003)

Table 3
Main processing units required for each process

Equipment	PROCESS 1	PROCESS 2	PROCESS 3
Fryer vessel	1	1	1
Heat exchangers	2	4	3
Compressors	0	2	2
Pumps	1	1	1
Heaters	1	1	1
Total	5	9	8

Even though PROCESS 1 has the simplest process flow diagram (PFD), the heat requirement is expensive when produced by fired heaters (Ulrich, 1984). Consequently, due to the absence of an energy recovery system, PROCESS 1 requires a much larger and more expensive steam generator, which contributes almost 60% of the total bare module cost. A closed heat pump presents what appears to be an intermediate capital cost, but has the most complicated PFD in terms of the number of unit operations. Also, closed heat pump technology currently available is not well adapted to the fry-drying process, which requires a ΔT of some tens of degrees Kelvin. The configuration demands expensive equipment but delivers a relatively poor energy return. The issues in PROCESS 2 are significantly simplified when water does not need to be re-condensed in a closed loop. Thus, the lowest capital investment, taking care regarding the large error bars, seems to be PROCESS 3. Water evaporated in the fry-drying is increased in temperature to heat the frying oil and then cooled, providing a high energy efficiency within a relatively simple PFD.

3.2.2. *Manufacturing cost*

Direct manufacturing costs were based on the results of the simulation (Table 6) and the costs of raw materials (Sud-Recupération, 2003), salaries (INSEE, 2005) and utilities (Turton et al., 2003). Indirect and general expenses have been calculated using factors typical of chemical engineering plant (Turton et al., 2003). The total manufacturing costs are summarized in Table 7. The assumption is that the fry-dryer will work one and a half shifts per day (i.e. 12 hours per day; 4380 hours a year), giving a process capacity of 4380 t/y dry solids, corresponding to an output of 21244 t/y of sludge dried to the auto-thermal condition.

Within the accuracy of the methods used, the manufacturing costs (C_{TM}) for each process are equivalent and do not provide a basis for distinguishing between the processes. However, the analysis highlights areas of expenditure that offer the biggest opportunities to minimize running costs.

From Table 7, indirect (C_{IM}) and general expenses (C_{GE}) costs represent ~20% and ~18% of the C_{TM}, respectively, in each of the three processes. By comparison, direct manufacturing costs (C_{DM}) contributes 62–64%. Clearly, the initial focus to reduce the manufacturing costs should be C_{DM}.

Unlike most alternative thermal drying processes, the energy cost is not the most important contributor to C_{DM}, (Arlabosse, 2001; Ressent, 1998). The extent of water evaporation that is needed to prepare fry-dried sludge to the auto-thermal condition, and which is supplied by

Table 4
Operating conditions of the main equipment

Equipment	PROCESS 1	PROCESS 2	PROCESS 3
V-101A,B,C,D Fryer vessel	$V = 5.3\,m^3$ $P = 1$ atm (abs) MOC: SS304	$V = 5.0\,m^3$ $P = 1$ atm (abs) MOC: SS304	$V = 5.1\,m^3$ $P = 1$ atm (abs) MOC: SS304
Axial fan	Vol. flow $= 0.26\,m^3\,s^{-1}$ Electric motor = 1.0 kW	Vol. flow $= 0.26\,m^3\,s^{-1}$ Electric motor = 1.0 kW	Vol. flow $= 0.26\,m^3\,s^{-1}$ Electric motor = 1.0 kW
Belt conveyor	Length = 148.9 m Width = 0.35 m Drive power = 9.5 kW	Length = 143.2 m Width = 0.35 m Drive power = 9.2 kW	Length = 144.0 m Width = 0.35 m Drive power = 9.2 kW
E-101 Heat exchanger	*Floating head* Surface $= 163.6\,m^2$ Tube side: 1 atm (abs), SS 304 Shell side: 9 atm (abs), CS	*Floating head* Surface $= 86.0\,m^2$ Tube side: 1 atm (abs), SS 304 Shell side: 9 atm (abs), CS	*Floating head* Surface $= 81.4\,m^2$ Tube side: 1 atm (abs), SS 304 Shell side: 9 atm (abs), CS
E-201 Heat exchanger	*Double pipe* Surface $= 5.0\,m^2$ $P = 1$ atm (abs) MOC: CS	*Multiple pipe* Surface $= 59.5\,m^2$ Tube side: 1 atm (abs), CS Shell side: 0.7 atm (abs), CS	*Multiple pipe* Surface $= 28.6\,m^2$ Tube side: 1 atm (abs), SS 304 Shell side: 4 atm (abs), CS
E-202 Heat exchanger	N/A	*Multiple pipe* Surface $= 23.5\,m^2$ Tube side: 1 atm (abs), SS 304 Shell side: 4 atm (abs), CS	*Double pipe* Surface $= 1.4\,m^2$ $P = 1$ atm (abs) MOC: CS
E-203 Heat exchanger	N/A	*Double pipe* Surface $= 1.9\,m^2$ $P = 1$ atm (abs) MOC:CS	N/A
C-201 Compressor	N/A	*Rotary* Elec. power = 48.8 kW P discharge = 1.9 atm (abs) T discharge = 526 K	*Rotary* Elec. power = 37.6 kW P discharge = 2.2 atm (abs) T discharge = 499 K
C-202 Compressor	N/A	*Rotary* Elec. power = 49.7 kW P discharge = 4.0 atm (abs) T discharge = 682 K Efficiency = 0.65	*Rotary* Elec. power = 36.6 kW P discharge = 4.0 atm (abs) T discharge = 616 K Efficiency = 0.65
H-101 Steam generator	Duty = 850.8 kW $P = 9.0$ atm (abs)	Duty = 455.3 kW $P = 9.0$ atm (abs)	Duty = 481.8 kW $P = 9.0$ atm (abs)
P-101 Pump	*Centrifugal* Elec. power = 4.3 kW Efficiency = 0.8	*Centrifugal* Elec. power = 4.1 kW Efficiency = 0.8	*Centrifugal* Elec. power = 4.2 kW Efficiency = 0.8

electricity or natural gas is only 10–14% of C_{DM}. Oil penetration and adhesion to the fried sludge contributes enormously to its energy value, reducing the drying that is needed. On the other hand, this consumes waste oil in the process—an additional feed (the waste oil) which alternative dryers do not require. This waste oil represents about 52–54% of the direct manufacturing costs, a major contribution to the total manufacturing costs.

The other important expense is the operating labor cost, which is 25–29% of C_{TM} (Table 7), and depends on the number of main processing units in the flow sheet. Comparatively, fry-drying may be expected to be competitive in terms of maintenance and repairs compared with conventional sludge dryers, having less parts in the drying operation, as well as only a small amount of exhaust vapors to manage.

Table 5
Equipment and fixed capital costs of the simulated fry-drying processes

Equipment (in US$, 2006)	PROCESS 1	PROCESS 2	PROCESS 3
V-101A,B,C,D Fryer and accessories	1.8×10^5	1.8×10^5	1.8×10^5
E-101 Heat exchanger	1.9×10^5	1.4×10^5	1.3×10^5
E-201 Heat exchanger	1.3×10^4	1.1×10^5	6.8×10^4
E-202 Heat exchanger	N/A	6.6×10^4	1.4×10^4
E-203 Heat exchanger	N/A	1.9×10^4	N/A
C-201 Compressor	N/A	9.5×10^4	8.0×10^4
C-202 Compressor	N/A	8.4×10^4	7.8×10^4
H-101 Steam generator	5.8×10^5	1.6×10^5	1.6×10^5
P-101 Pump	1.6×10^4	1.6×10^4	1.6×10^4
Total basic module cost: C_{BM_0}	4.1×10^5	3.1×10^5	2.7×10^5
Total bare module cost: C_{BM}	9.8×10^5	8.7×10^5	7.3×10^5
Contingency fees: $C_{CF} = 0.18 \times C_{BM}$	1.8×10^5	1.6×10^5	1.3×10^5
Total module cost: $C_{TM} = C_{BM} + C_{CF}$	1.2×10^6	1.0×10^6	8.6×10^5
Auxiliary facility cost: $C_{AC} = 0.3 \times C_{BM}$	1.2×10^5	9.2×10^4	8.0×10^4
Fixed capital cost: $C_{FC} = C_{TM} + C_{AC}$	1.3×10^6	1.1×10^6	9.4×10^5
Working capital cost: $C_{WC} = 0.15 \times C_{FC}$	1.9×10^5	1.7×10^5	1.4×10^5
Total capital investment $C_{TC} = C_{FC} + C_{WC}$	1.5×10^6	1.3×10^6	1.1×10^6

Table 6
Utility requirements for each process

Equipment	Description	PROCESS 1	PROCESS 2	PROCESS 3
V-101A,B,C,D				
Fan	Elec. power (kW)	1.0	1.0	1.0
Belt conveyor	Elec. power (kW)	10.5	10.2	10.2
Total	Elec. power (kW)	11.5	11.5	11.5
E-101 Heat exchanger	Steam ($kg\,h^{-1}$)	1612.5	808.7	851.7
H-101 Steam generator	Heat duty (kW)	850.8	455.3	481.8
	Fuel gas ($kg\,h^{-1}$)	65.2	34.9	36.9
E-201 Heat exchanger	Cooling water ($kg\,h^{-1}$)	31,000	0	0
C-201 Compressor	Elec. power (kW)	0	48.8	37.6
C-202 Compressor	Elec. power (kW)	0	49.7	36.6
P-101 Pump	Elec. power (kW)	4.3	4.1	4.2

3.2.3. *Economics of thermal drying processes for sewage sludge*

Economic data about current sewage sludge industrial dryers are available from Gross (1993), Grüter et al. (1990), Hasserbrauck and Ermel (1996) and Ressent (1998, 1999). It is difficult to make a strict comparison of the fixed capital and manufacturing costs of installed dryers with those of the processes simulated here, because of the different aims of each drying facility. Nevertheless, the information provides some comparisons that usefully illustrate the economic performance of the competing systems and affords reference points for the evaluation of frying as a process to thermally dry sewage sludge.

Table 8 summarizes a collection of capital costs and capacities reported for some of the sewage sludge thermal drying facilities currently operating in Europe. The fixed capital cost per ton of DS treated per year is US$572-1817, compared with US$232/tDS/y for PROCESS 3 in this study (or US$275/tDS/y and US$324/tDS/y for PROCESSes 2 and 1, respectively). As another comparison, Ressent (1998) reported economic assessments of 13 conventional models for the thermal drying of sewage sludge having annual capacities of 1230–2960 tDS. Converting these results to current values provides an average capital investment for the indirect dryers of ~US$2323/tDS/y and US$1615/tDS/y for the direct dryers.

Ressent (1998) also provides direct manufacturing costs excluding final disposal expenses (i.e. transportation, incineration, etc.) which in current value are ~US$190/tDS/y for both direct and indirect dryers. From Table 7, the comparable costs for the simulated fry-drying processes are about US$202, 199 and 194/tDS/y for PROCESSes 1, 2 and 3, respectively. Although these estimated running costs

Table 7
Total manufacturing costs of the simulated fry-drying processes

Cost (in US$, 2006)	PROCESS 1	PROCESS 2	PROCESS 3
Direct manufacturing cost			
Recycled cooking oil (293 US$ ton^{-1} of RCO)[a]	4.5×10^5	4.5×10^5	4.5×10^5
Operating labor: C_{OL} (23122 US$ year^{-1})[b]	1.4×10^5	1.6×10^5	1.6×10^5
Supervision and clerical labor: $C_{SC} = 0.15 \times C_{OL}$	2.1×10^4	2.4×10^4	2.4×10^4
Natural gas (8.84 US$ GJ^{-1})[c]	1.2×10^5	5.0×10^4	5.4×10^4
Electricity (0.07 US$ kWh)[c]	5.5×10^3	4.6×10^4	3.6×10^4
Cooling water (0.02 US$ m^{-3})[c]	3.9×10^1	0	0
Maintenance and repairs: $C_{MR} = 0.6 \times C_{FC}$	7.4×10^4	6.3×10^4	5.3×10^4
Operating supplies: $C_{OS} = 0.15 \times C_{MR}$	1.1×10^4	9.5×10^3	8.0×10^3
Laboratory charges: $C_{LC} = 0.15 \times C_{OL}$	2.1×10^4	2.4×10^4	2.4×10^4
Patents and royalties: $C_{PR} = 0.03 \times C_{TM}$	4.3×10^4	4.2×10^4	4.0×10^4
Subtotal: C_{DM}	8.9×10^5	8.7×10^5	8.5×10^5
Indirect manufacturing cost			
Overhead, packaging and storage: $C_{OPSL} = 0.16(C_{OL} + C_{MR} + C_{SC})$	1.4×10^5	1.5×10^5	1.4×10^5
Local taxes: $C_{LT} = 0.015 \times C_{FC}$	1.9×10^4	1.6×10^4	1.3×10^4
Insurances: $C_I = 0.005 \times C_{FC}$	6.2×10^3	5.3×10^3	4.4×10^3
Depreciation: $C_D = 0.01 \times C_{FC}$	1.2×10^5	1.1×10^5	8.8×10^4
Subtotal: C_{IM}	2.9×10^5	2.8×10^5	2.5×10^5
General expenses			
Administrative costs: $C_{ADMC} = 0.25 \times C_{ADMC}$	3.5×10^4	3.7×10^4	3.6×10^4
Distribution and selling: $C_{DS} = 0.10 \times C_{TM}$	1.4×10^5	1.4×10^5	1.3×10^5
Research and development: $C_{RD} = 0.25 \times C_{OPSL}$	7.1×10^4	7.0×10^4	6.7×10^4
Subtotal: C_{GE}	2.5×10^5	2.5×10^5	2.4×10^5
Total manufacturing cost: C_{TM}	1.4×10^6	1.4×10^6	1.3×10^6

[a]Converted from the RCO December 2004 price (Sud-Recupération, 2003).
[b]Converted from the 2002 industry worker salary (*ouvrier dans l'industrie*) in France (INSEE, 2005).
[c]Converted from Turton et al. (2003).

Table 8
Approximate fixed capital costs of current sewage sludge thermal drying facilities in Europe

Drying facility	Location	Capacity (ton DS year^{-1})	Published price	Updated price[d] US$, 2006	Ratio (US$ of investment ton^{-1} DS)	Reference
Thin film dryer + disc dryer	Switzerland	8000	4.5×10^{6a}	4.6×10^6	572	(Grüter et al., 1990)
Drum	Germany	7799	15×10^{6b}	11.4×10^6	1459	(Hasserbrauck and Ermel, 1996)
Thin film dryer + disc dryer	Germany	6260	15×10^{6b}	11.4×10^6	1817	(Hasserbrauck and Ermel, 1996)
Drum	France	5093	14×10^{6c}	3.1×10^6	608	(Ressent, 1999)
Belt	Germany	2703	7.8×10^{6c}	1.7×10^6	639	(Ressent, 1999)
Disc	France	1539	9.0×10^{6c}	2.0×10^6	1294	(Ressent, 1999)

[a]Price in Swiss Francs (1.2880CHF = 1US$).
[b]Price in German Mark (1.6195DEM = 1US$).
[c]Price in French Francs (5.4315FRF = 1US$).
[d]Prices updated using the corresponding CEPCI Index and currency, $I_{2005} = 468.1$; $I_{1998} = 389.5$; $I_{1995} = 381.1$ and $I_{1990} = 357.6$.

are quite similar, it may be noted that the contributing factors are quite different (Figs. 5 and 6). Maintenance and energy requirements are very important for indirect and direct dryers, whereas waste oil consumption is much more important for the fry-drying processes. Therefore, one way to improve the fry-drying running costs is to find more economical frying oils, for example trap waste. Another is to find more economical heating sources. Since fry-dried sludge has good calorific characteristics, the final fry-dried sludge could be used to provide energy, e.g. substituting for fossil fuel. This could provide an economic advantage and, since the energy from waste incineration (EfWI) claims to save natural resources (Porteous, 2001), an important environmental outcome might also be achieved. These and other process options require additional design, optimization and evaluation.

Location and integration opportunities also need to be considered. For example, the fry-drying processes simulated in this paper offer only a small volume reduction of the waste, since only small amounts of water are removed and some oil is added, to bring the product to an auto-thermal condition. Since this will impact subsequent transportation costs it would be most advantageous in facilities equipped with an incineration unit or located close to the incinerator.

4. Conclusions

Front end process engineering sizing using ASPEN as the simulation tool has been used as a basis for determining capital and operating costs for fry-drying sewage sludge. Three variants of the process were examined, all based on an annual throughput of 4380 tons of dry solids. The main

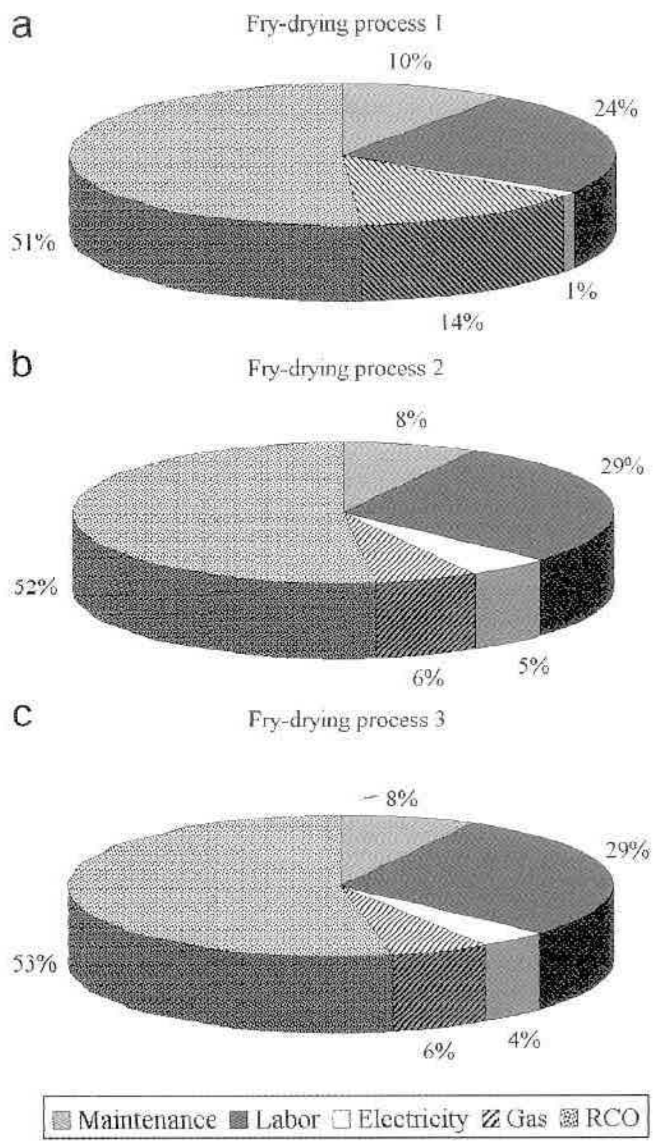


Fig. 5. Direct manufacturing costs for the simulated fry-drying processes.

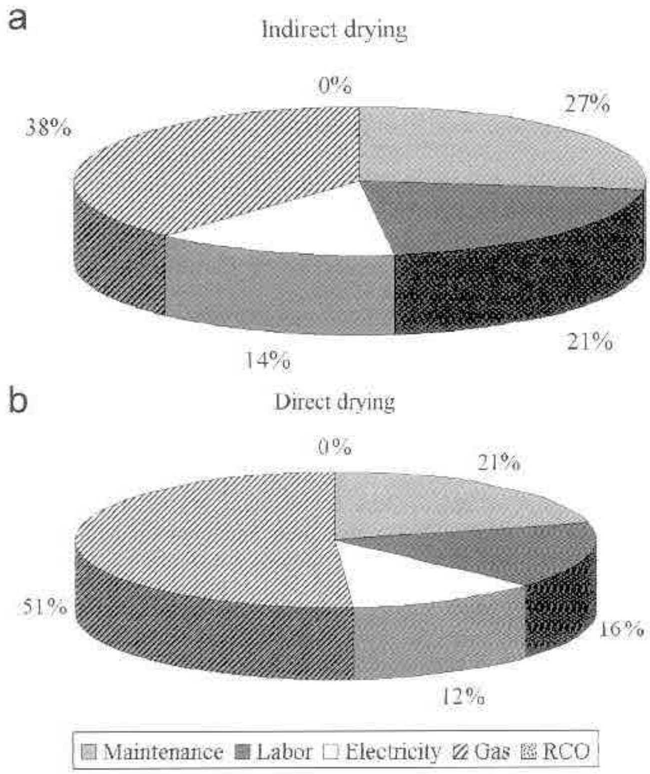


Fig. 6. Direct manufacturing costs for the conventional thermal dryers, according to Ressent (1998).

difference between the three processes was in the exhaust vapors management systems. PROCESS 1 condenses the vapor evaporated from the sludge using cooling water, whereas PROCESSes 2 and 3 used closed and open heat pumps, respectively, to recycle some of the latent energy in the vapor and so improve the energy efficiency of the process. Technically, all three processes appear to be feasible for production of auto-thermal fry-dried sludge and have moderate operating conditions.

The absence of an energy recovery system gives PROCESS 1 the simplest process flowsheet, but requires the largest steam generator and consumes the highest amounts of fuel. Current closed heat pump technology as proposed in PROCESS 2 seems less well adapted to the fry-drying process. Indeed, it has the most complicated PFD and demands the highest electricity consumption. Finally, PROCESS 3 seems to offer a well-adapted technology for fry-drying in which the exhaust vapors are compressed in an open loop to provide process heat without needing re-condensation. This allows an efficient energy recovery in a relatively simple PFD.

Based on comparisons with costs for current sewage sludge thermal drying systems, fry-drying fixed cost per ton of processed dry solids is many times lower than those of conventional dryers. The direct operating costs are about equivalent to conventional processes, although in fry-drying the main variable cost is waste oil, whereas energy and maintenance costs are most important for current processes. This suggests the use of cheaper waste oils, for example grease trap waste, in fry-drying or drying the sludge sufficiently that it may itself be used for heating.

References

Alkayat, W.A., Gerrard, A.M., 1984. Estimating manning levels for process plants. AACE Transactions I.2.1–I.2.4.

Arlabosse, P., 2001. Etude des procédés de séchage des boues urbaines et industrielles. Ecole des Mines d'Albi Carmaux & Association RE.CO.R.D., Albi, France, p. 184.

Arlabosse, P., Chavez, S. Characterization of some sewage sludge properties essential for the design and management of thermal processes Environmental Technology, submitted for publication.

Bress, D.F., Greenfield, B.S., Haug, R.T., 1987. Energy from sludge derived fuels: the hyperion energy recovery system. In: Klass, D., Donald, L. (Eds.), Energy from Biomass Waste, vol. X. Institute of Gas Technology, Chicago, IL, USA, pp. 1173–1182.

Bresters, A.R., Coulomb, I., Deak, B., Matter, B., Saabye, A., Spinosa, L., Utvik, A., 1997. Sludge treatment and disposal: management approaches and experiences. European Environmental Agency & International Solid Waste Association, Copenhagen, Denmark, p. 53.

Carrère-Gée, C., 1999. Indirect drying of thin-film alumina sludge by boiling Application on drum drying. Ph.D. in Energy Transfer, Toulouse, France Université Paul Sabatier, 243pp.

Chavez, S., 2004. Séchage par contact avec agitation de boues résiduaires urbaines Influence de leur origine et des conditions opératoires sur la cinétique de séchage et les caractéristiques des boues sèches et des rejets gazeux. Ph.D. in Applied Sciences, Perpignan, france université de Perpignan, 182pp.

Chen, G., Yue, P.L., Mujumdar, A.S., 2002. Sludge dewatering and drying. Drying Technology 20, 883–916.

Christiannsen, K.M., 1999. Waste annual topic update 1998. European Environment Agency, Copenhagen, Denmark, p. 43.

Farkas, B.E., Hubbard, L.J., 2000. Analysis of convective heat transfer during immersion frying. Drying Technology 18, 1269–1285.

Freitas, L., Freire, J., 1993. Heat transfer in spouted beds. Drying Technology 11, 303–317.

Garayo, J., Moreira, R., 2002. Vacuum frying of potato chips. Journal of Food Engineering 55, 181–191.

Gross, T.S.C., 1993. Thermal drying of sewage sludge. Journal of the Institution of Water and Environmental Management 7, 255–261.

Grüter, H., Matter, M., Oehlmann, K.H., Hicks, M.D., 1990. Drying of sewage sludge. An important step in waste disposal. Water Science and Technology 22, 57–63.

Hasserbrauck, M., Ermel, G., 1996. Two examples of thermal drying of sewage sludge. Water Science and Technology 33, 235–242.

Hwang, Y., Matsuo, T., Hanaki, K., Suzuki, N., 1995. Identification and quantification of sulfur and nitrogen containing odours compounds in wastewater. Water Research 29, 711–718.

IEA, 2004. ⟨http://www.heatpumpcentre.org/⟩, 1 December 2004.

INSEE, 2005. ⟨http://www.insee.fr/fr/ffc/chifcle_fiche.asp?ref_id = NAT-SEF04122&tab_id = 508⟩, 25 September 2005.

Kudra, T., 2003. Sticky region in drying—definitions and identification. Drying Technology 21, 1457–1469.

Kudra, T., 2004. Energy aspects in drying. Drying Technology 22, 917–932.

Kuntschar, W., 1996. Verfahren und Vorrichtung zum Trocknen von Klärschlamm. Process and equipment for the sludge drying, Germany, Patent: DE19639551C1 (30 April 1998).

Lambert, S.D., Beaman, A.L., Winter, P., 2000. Olfactometric characterisation of sludge odours. Water Science and Technology 41, 49–55.

Léonard, A., 2002. Etude du séchage convectif de boues de station d'épuration: Suivi de la texture par microtomographie à rayons X. Ph.D. in Applied Sciences, Université de Liège, Liège, Belgium, 260pp.

Lowe, P., 1995. Developments in the thermal drying of sewage sludge. Journal of the Chartered Institution of Water and Environmental Management 9, 306.

Lue-Hing, C., Matthews, P., Namer, J., Okuno, N., Spinosa, L., 1996. Sludge management in highly urbanized areas. Water Science and Technology 34, 517–524.

Mauvieux, P., Madeline, Y., Gabarda-Oliva, D., Fondin, P., Cholin, B., Picard, J.M., 2001. Les déchets graisseux Quelles solutions et à quels coûts? In: Pollutec 2001. Agence de l'Eau Seine-Normandie, Paris, France, 4 December 2001, 56pp.

Moreira, R., Castell-Perez, M., Barrufet, M., 1999. Deep-fat Frying: Fundamentals and Applications. Aspen Publication, Gaithersburg, MD, USA.

Munck-Kampmann, B., 2001. Waste annual topic update 2000. European Environment Agency, Copenhagen, Denmark, p. 24.

Nurul, A., Hanaki, K., Matsuo, T., 1998. Fate of dissolved odours compounds in sewage treatment plants. Water Science and Technology 38, 337–344.

Peregrina, C., Lecomte, D., Arlabosse, P., Rudolph, V., 2006a. Heat and mass transfer during fry-drying of sewage sludge. Drying Technology 24, 797–818.

Peregrina, C., Lecomte, D., Arlabosse, P., Rudolph, V., 2006b. Life cycle assessment (LCA) applied to the design of an innovative drying process for sewage sludge. Process Safety and Environmental Protection 84, 797–818.

Perry, R.H., Green, D.W., 1984. Perry's Chemical Engineers' Handbook. McGraw-Hill, New York, USA.

Peters, M.S., Timmerhaus, K.D., 1991. Plant Design and Economics for Chemical Engineers, fourth ed. McGraw-Hill, New York, USA.

Porteous, A., 2001. Energy from waste incineration—a state of the art emissions review with an emphasis on public acceptability. Applied Energy 70, 157–167.

Poulsen, T.G., Hansen, J.A., 2003. Strategic environmental assessment of alternative sewage sludge management scenarios. Waste Management and Research 21, 19–28.

Powell, R.L., 2002. CFC phase-out: have we met the challenge? Journal of Fluorine Chemistry 114, 237–250.

Ressent, S., 1998. Etat de l'art sur le séchage thermique de boues urbaines et industrielles, Agence de l'eau Seine-Normandie, Paris, France, p. 358.

Ressent, S., 1999. Campagnes de mesure realisés sur des séchoirs de boues urbaines et industrielles, Agence de l'eau Seine-Normandie, Paris, France, p. 268.

Riffat, S.B., Afonso, C.F., Oliveira, A.C., Reay, D.A., 1997. Natural refrigerants for refrigeration and air-conditioning systems. Applied Thermal Engineering 17, 33–42.

Silva, D.P., 2003. Estudo da secagem de lodo de esgoto através da fritura de imers. Ph.D. in Chemical Engineering, Campinas, SP, Brazil Unidersidade Estadual de Campinas, 180pp.

Silva, D.P., Rudolph, V., Taranto, O.P., 2005. The drying of sewage sludge by immersion frying. Brazilian Journal of Chemical Engineering 22, 271–276.

Sinnott, J.F., 1991. An Introduction to Chemical Engineering Design. Pergamon Press, Oxford, UK.

Stier, R.F., 1996a. Understanding High-Volume Frying. Part 1. Baking & Snack, pp. 70–76.

Stier, R.F., 1996b. Understanding High-Volume Frying. Part 2. Baking & Snack, pp. 51–54.

Stier, R.F., 1996c. Understanding High-Volume Frying. Part 3. Baking & Snack, pp. 41–44.

Sud-Recupération, 2003. ⟨www.sud-recuperation.fr/traitement.html⟩, August 2005.

Tsai, W.-T., 2005. An overview of environmental hazards and exposure risk of hydrofluorocarbons (HFCs). Chemosphere 61, 1539–1547.

Turton, R., Bailie, R.C., Whiting, W.B., Shaeiwitz, J.A., 2003. Analysis, Synthesis, and Design of Chemical Processes, second ed. Prentice-Hall PTR, NY, USA.

Twu, C.H., Coon, J.E., Bluck, D., 1998. Comparison of the Peng–Robinson and Soave–Redlich–Kwong equations of state using a new zero-pressure-based mixing rule for the prediction of high-pressure and high-temperature phase equilibria. Industrial & Engineering Chemistry Research 37, 1580–1585.

Twu, C.H., Sim, W.D., Tassone, V., 2002. Getting a handle on advanced cubic equations of state. Chemical Engineering Progress 98, 58–65.
Ulrich, G.D., 1984. A Guide to Chemical Engineering Process Design and Economics. Wiley, New York, USA.
Vatavuk, W.M., 2002. Updating the CE plant cost index. Chemical Engineering Magazine 62–70.
Vitrac, O., 2000. Carcatérisation expérimentale et modelisation de l'opération de friture. Ph.D. in Chemical Engineering, Paris, France, 326pp.
Werther, J., Ogada, T., 1999. Sewage sludge combustion. Progress in Energy and Combustion Science 25, 55–116.
Whipps, A., 2004. A review of the dangerous and explosive atmosphere regulations in relation to the water industry. Water and Environment 18, 118–120.
Winter, P., Jones, N., Asaadi, M., Bowman, L., 2004. The odour of digested sewage sludge—determination and its abatement by air stripping. Water Science and Technology 49, 185–192.
Yamahata, Y., Izawa, H., 1985. Experimental study on application of paddle dryers for sludge cake drying. In: Fourth International Drying Symposium, Kyoto, Japan, pp. 719–724.
Zhang, Y., Dube, M.A., McLean, D.D., Kates, M., 2003a. Biodiesel production from waste cooking oil: 1. Process design and technological assessment. Bioresource Technology 89, 1–16.
Zhang, Y., Dube, M.A., McLean, D.D., Kates, M., 2003b. Biodiesel production from waste cooking oil: 2. Economic assessment and sensitivity analysis. Bioresource Technology 90, 229–240.

ELSEVIER

Journal of Environmental Management 86 (2008) 262–271

Journal of Environmental Management

www.elsevier.com/locate/jenvman

The influence of institutions and organizations on urban waste collection systems: An analysis of waste collection system in Accra, Ghana (1985–2000)

Julius N. Fobil[a,*], Nathaniel A. Armah[b], Jonathan N. Hogarh[c], Derick Carboo[d]

[a]School of Public Health, College of Health Sciences, P.O. Box LG13, University of Ghana, Legon, Accra, Ghana
[b]UESP Capacity-Building & Training, Ministry of Local Government and Rural Development, PMB, Ministries Post Office, Accra, Ghana
[c]Department of Environmental Science, College of Science, Kwame Nkrumah University of Science & Technology, PMB, University Post Office, Kumasi, Ghana
[d]Department of Chemistry, P.O. Box LG56, University of Ghana, Legon, Accra, Ghana

Received 21 December 2005; received in revised form 14 November 2006; accepted 7 December 2006
Available online 8 February 2007

Abstract

Urban waste collection system is a pivotal component of all waste management schemes around the world. Therefore, the efficient performance and the success of these schemes in urban pollution control rest on the ability of the collection systems to fully adapt to the prevailing cultural and social contexts within which they operate. Conceptually, institutions being the rules guiding the conduct of public service provision and routine social interactions, waste collection systems embedded in institutions can only realize their potentials if they fully evolve continuously to reflect evolving social and technical matrices underlying the cultures, organizations, institutions and social conditions they are designed to address. This paper is a product of an analysis of waste collection performance in Ghana under two different institutional and/or organizational regimes: from an initial entirely public sector dependence to a current mix of public–private sector participation drawing on actual planning data from 1985 to 2000. The analysis found that the overall performance of waste collection services in Ghana increased under the coupled system, with efficiency (in terms of total waste clearance and coverage of service provision) increasing rapidly with increased private-sector controls and levels of involvement, e.g. for solid waste, collection rate and disposal improved from 51% in 1998 to about 91% in the year 2000. However, such an increase in performance could not be sustained beyond 10 years of public–private partnerships. This analysis argues that the sustainability of improved waste collection efficiency is a function of the franchise and lease arrangements between private sector group on the one hand and public sector group (local authorities) on the other hand. The analysis therefore concludes that if such franchise and lease arrangements are not conceived out of an initial transparent process, such a provision could undermine the overall sustainability of private sector initiatives in collection services delivery in the long term, as in the case of the Accra example.

Keywords: Sustainability; Public-private partnerships; Franchise; Waste clearance; Waste collection and pollution control

1. Introduction

The right to health is one of the fundamental human rights, which enjoins nations to guarantee preconditions for healthy lives of their peoples. For example, in Ghana, the key strategy currently adopted to realize this basic requirement for life of all people, both urban and rural populations, is the Primary Health Care (PHC) policy (Armah, 2001; Fobil, 2002). The overall objective of the policy among others, stipulates that, all Government Ministries, Departments and Agencies (MDAs), individually and jointly, provide services and public facilities to protect public health, the environment and settlements, which are vital ingredients for the total development of the country and its people (Armah, 2001). An important

*Corresponding author. Tel.: +233 21 517 506.
E-mail addresses: jfobil@ug.edu.gh (J.N. Fobil), jnhogarh.sci@knust.edu.gh (J.N. Hogarh), dcarboo@ug.edu.gh (D. Carboo).

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.12.038

component of the PHC is the environmental sanitation of which wastes management is a key sub-component. Wastes collection services are indispensable core activities in wastes management. Like any such economic activity, waste collection and its final disposal are governed by neo-classical economic theory of demand and supply. Market conditions that operate within a typical market environment also apply in waste collection, storage, transportation and final disposal (Fobil, 2002). However, it must be noted that the provision of these services cannot be left entirely to the dictates of market forces alone as other external factors such as social and cultural contexts of a given society play part in determining the direction of these market forces.

Generally, a chronic problem confronting all sub-Sahara African cities including Accra in Ghana is a proliferation of squatter and informal settlements, peripheral sprawl, and central city tenements, all lacking basic infrastructure services with poor sanitation (Boadi and Kuitunen, 2005; Nuno-amarteifio, 1995; Bannerman, 2000; Wurapa, 1973; Makoni et al., 2004; Mabogunje, 1995; Satterthwaite, 1993; Von Schirnding, 1996). Indeed, cities in this region are not only rapidly urbanizing, but the urban environments here are fast deteriorating and this kind of urban growth phenomenon may better be described as '*ruralization*" than "*urbanization*" (Bannerman, 2000; Makoni et al., 2004; Mabogunje, 1995; Von Schirnding, 1996; Taiwo, 1996; Majani, 1996). Many of the cities, e.g. Accra in Ghana, Lagos in Nigeria, Kampala in Uganda, Addis Ababa in Ethiopia, Dar-es-Salaam in Tanzania and Nairobi in Kenya are all beset with unhealthy urban environmental conditions and filth (Makoni et al., 2004; Mabogunje, 1995; Von Schirnding, 1996; Taiwo, 1996; Majani, 1996; Kgathi and Bolaane, 2001). Several interventions and urban initiatives meant to improve urban environmental conditions have been attempted in many of these cities without much success and these include the upgrading efforts in Ghana, and urban projects over the years in Kenya, Ethiopia, Senegal, Cote d'Ivoire, and Zambia (Boadi and Kuitunen, 2005; Nuno-amarteifio, 1995; Bannerman, 2000; Makoni et al., 2004; Mabogunje, 1995; Von Schirnding, 1996; Taiwo, 1996; Majani, 1996; Kgathi and Bolaane, 2001).

In many of the cities, the key observable feature is that the collection, transportation and disposal of solid waste has moved from the control of local government authorities to the increased involvement of the private sector (PS) either "spontaneously" in a free market setting or encouraged through local authorities, non-governmental organization (NGOs) or community-based organizations (CBOs) (Nuno-amarteifio, 1995; Post et al., 2001) in a hybrid couple system. This is on account of the wide realization that public sector institutions have failed considerably to perform efficiently in the region. Solid waste management is no longer a (local) government monopoly but now largely a domain, open to various modes of public–private co-operation either through lease agreements or direct franchise arrangements. This hybridized couple system between public and private groups in urban environmental management has both opportunities and potential caveats, which if improperly harmonized, frequently lead to large negative externalities such as poor performance, organizational failures and resource misapplication, and which come with an inevitable consequence of environmental deterioration. Privatization of municipal services often come as a consequence of a multitude of reasons which in most cases is provoked by the onset of poor service delivery by public sector entities. As in the case of what happened in many of the sub-Sahara African cities, widespread internal dissatisfaction by residents and lack of confidence in the service delivery by the local government authorities, provided impetuous for increased pressures from the Bretton Woods institutions which are major sponsors of such urban initiatives, forcing public sector organizations to either completely relinquish or share the role of municipal services delivery to PS groups, often credited with high performance. In the literature, the PS is credited with qualities such as political independence; economic rationality, efficiency, dynamism and innovation, qualities that make it measure up favorably to public sector enterprise (Post et al., 2001). The aim of this paper is to analyze the historical developments in the waste collection system in Accra during the period between 1985 and 2000, highlighting the various economic, institutional, organizational, and market determinants of such historical developments. The paper takes an analytic perspective of waste collection performance in Ghana under two different institutional and/or organizational regimes; from an initial entirely public sector dependence to a current mix of public–private sector participation.

1.1. City of this analysis

The analysis is carried out in Accra, the capital city of the Republic of Ghana. Accra is a coastal city located in the Greater Accra Region, the smallest of the 10 political regions in Ghana (Stephen, 1999). It is, however, the largest of Ghana's 10 leading urban centers, with an estimated population of approximately 1.7 million in 1990 and 2.7 million in 2000 (Carboo and Fobil, 2004). Currently, Ghana's population stands at approximately 20 million; and Accra alone harbors over 30% of the urban population of Ghana and nearly 15% of the country's total population (GSS, 2004). The generation and annual rate of increase of solid waste are high in Ghana. In Accra for example, per capita production of refuse is estimated at 0.40 kg/day (Fobil, 2001). At the current 6% population growth rate, the population of Accra in 2005 was estimated at approximately 3.6 million, with a total refuse generation of 1.44 million kg/day (1440 metric tons per day), which translated into approximately 0.5 million metric tons per year. Nearly 60% by weight of this huge chunk of waste generated was organic (Fobil, 2001), representing 0.3 million metric tons of waste annually. Nearly half of solid waste currently generated in cities in Ghana is not collected

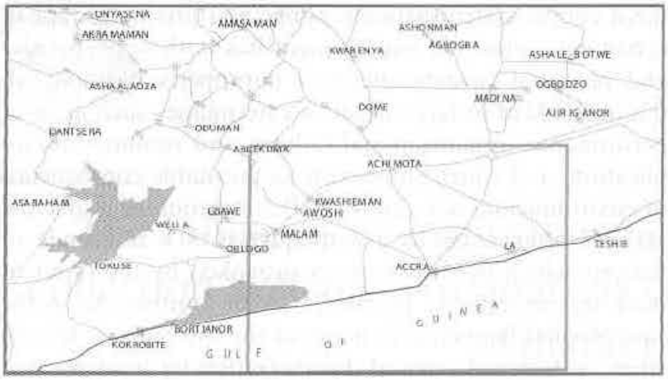


Fig. 1. The Accra Metropolitan Area (AMA), 2005.

(Fobil, 2001; Asomani-Boateng and Haight, 1999) (see Fig. 1).

2. The environmental consequence of institutional and organizational failures in waste collection in Accra

The period between 1957 and 1985 in Ghana witnessed a general deterioration and breakdown of public sanitation facilities in the major cities notably the regional capitals of the country (Nuno-amarteifio, 1995; Laryea, 2000; Martinson et al., 1999). This created complete environmental chaos in all urban centers. The deterioration in sanitation services and the systemic degradation of urban environmental quality came as a consequence of institutional failures (Boadi and Kuitunen, 2005; Nuno-amarteifio, 1995). During this period, Accra like many other cities in Ghana, experienced a serious decline in environmental sanitation, particularly in waste management (Nuno-amarteifio, 1995; Bannerman, 2000; Laryea, 2000). This was a result of the failure of institutions to respond to the planning needs of a modern city (Nuno-amarteifio, 1995). Many reasons have been cited for the systemic failures of institutions in sub-Saharan Africa. These may broadly be grouped into two, namely—human and technical factors (Fobil, 2002, 2001; Makoni et al., 2004; Mabogunje, 1995; Majani, 1996; Agunwamba, 2001, 2003; Anikwe and Nwobodo, 2002; Bruggemann, 1971; Darko and Fletcher, 1998; EPA-Ghana, 1997; Gulis et al., 2004; Gwebu, 2003; Holmes, 1983; Leschber, 2002; Makule, 2000; Mbuligwe, 2004; Mtani, 2004). Institutions are generally conceived as the rules of the game in a society or more technically, as the humanly devised constraints that shape human interaction (North, 1999). Institutions then tend to structure incentives in human exchange, whether political, social, or economic. Institutions may be contrasted from organizations in that; they do not provide a physical structure to human interaction (North, 1999). However, they set non-physical limits to societal actions. Institutions may appear feasible, but if they are not supported by clearly defined organizational arrangements, they may well be counter-productive.

2.1. Human factors

Indiscriminate erection of structures by individuals and private estate developers at unapproved sites is a common phenomenon in many parts of Accra, culminating in lack of appropriate access routes to many of these areas. This is a result of non-enforcement of physical planning and land-use laws in the city and thus, giving the built environment a confused and chaotic outlook. Accra, like many other cities in sub-Sahara Africa, is then a settlement of poorly laid structures, which lacks access routes in many of its residential enclaves. Lack of access routes makes garbage removal and lifting rather difficult if not impossible through motorized waste lifting system, a situation, which provides for rapid accumulation of solid wastes as seen in many residential areas in the city today. It is then common to find that many households in the city are not served with waste collection services.

By 1985, the problem of urban environmental sanitation had reached such a level that action could not be delayed any longer. There were huge accumulations of solid waste at several locations within most of the communities in Accra, more especially in the high density, low-income residential areas of the metropolis. The situation was the same with respect to liquid waste.

The rapid rate of urban growth made it very difficult for city authorities to meet the demand for critical public services such as potable water, sewers, public latrines, public water closet toilets, and other sanitation facilities which made it very difficult to deal with the large volume of liquid wastes from homes. Overflowing public toilets became daily norms rather than the exception. Indeed, many of these public facilities fell into long periods of misuse, abuse and disrepair, resulting in high build up of sludge in the septic tanks. This reduced the available volume of the septic tanks, and caused the tanks to overspill rapidly. In such cases, it was not unusual for such public toilets to be closed to the public. A closure often led to defecation around the facility and in open places in the neighborhoods.

The private facilities in homes, industry, shops or markets, and offices or institutions were not left out. For lack of maintenance of the sanitation equipment, only few cesspool emptiers were available and serviceable, resulting in long waiting period for desludging services to be provided.

Private pan latrine carriers[1], employed by the Accra Metropolitan Assembly (AMA) took advantage of the inadequacy of vehicles to extort from users extra payment for themselves, for emptying overflowing pan latrines, irrespective of whether or not the pan was properly registered with the Assembly. The solid waste that remained uncollected for several months often found

[1]Private pan latrine carriers are private individuals who made living by going from house-to-house to manually lift human excreta using head-carrier containers.

routes into open drains thereby blocking these drains and runoff, and became a veritable recipe, not only for breeding of mosquitoes and flies, but also factories from which emanated obnoxious odors into the atmosphere with their offensive smells.

2.2. Technical and administrative factors

A study was conducted to diagnose the cause of these unacceptable and health-threatening environmental conditions with the assistance of GTZ (of the government of the Federal Republic of Germany) during the period 1983–1985 (Armah, 2001). The study found that the roles of sub-units and sub-departments within the AMA (the sole public provider of sanitation services) were not properly clarified leading to overlap of responsibilities in most cases and in some other cases complete absence of administrative oversight. This situation often developed into serious institutional and practical conflicts as has been conceptualized in Figs. 2 and 3. For instance, it was found that as the key roles were not well defined among units making up the AMA, a potential danger of running into mal-functioning of the financial system developed, and which may have led to cash-flow problems (Armah, 2001). This was because the different units scrambled for budget allocations to the detriment of the goals of the larger organization. Often, cash-flow difficulty led to irregular and low wages as well as low salaries among the workforce, which may then have resulted in job dissatisfaction, low waste collection efficiency, poor performance and corruption among the personnel and staff. There was inadequate coordination among the bodies, which had equal powers under different management. The lack of coordination and overlap of responsibilities created misunderstanding and confrontation among the groups (Armah, 2001).

The main findings of the study showed that there were the following institutional weaknesses:

- fragmentation of the waste management functions among several departments of the Assembly e.g.,
  - Metro Health Department delivered the services of solid waste collection and haulage to disposal sites, pan latrine service, desludging of septic tanks, and the Operation and Maintenance (O&M) of public toilets,
  - Metro Engineer's Department was responsible for the construction and maintenance of the sanitary facilities,
  - Mechanical Engineer's Department had responsibility for the repair and maintenance of the sanitation vehicles and equipment, the management of treatment and final disposal sites (landfills, compost plant and fecal treatment plants), the engagement and assignment of drivers and operators to the Health Department for services delivery,
- the waste management activities were not effectively coordinated and harmonized, resulting in overlapping in some areas, and other gray areas not assigned to any of the departments,
- inadequate numbers and/or often, inappropriate fleet and equipment resulting in maintenance problems which were exacerbated by the lack of a standardized policy on

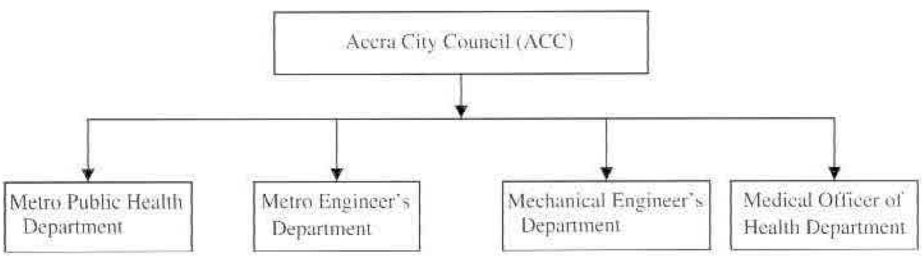


Fig. 2. Previous institutional arrangement (1957–1985).

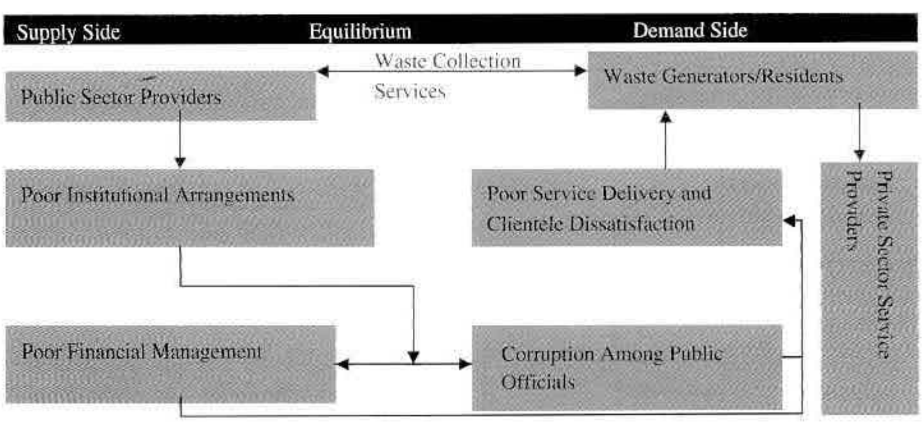


Fig. 3. Institutional and economic analytic model in waste collection services.

their operations. In most times, there were more than one model of vehicles in the sanitation fleet deployed in the municipal authority,

- serious cash-flow problems, which adversely affected the effective and efficient operation and maintenance of sanitary fleet and equipment, and this in turn, resulted in their unreliability and poor availability for services delivery. Also, service fees were charged only for house-to-house solid waste collection and this constituted less than 10% of the services delivered,
- complete absence or low involvement of PS group in waste collection services provision,
- a dearth of qualified, well-trained and motivated professionals, technicians, operation and maintenance staff to undertake with efficiency the planning, management, operation and maintenance of the services.

3. Remedial actions to correct institutional failures in AMA

In the short and medium term, a Board of Directors was composed to oversee the operations and services to be delivered by a commercially oriented entity, wholly owned by the AMA. However, to meet the exigencies in the long term, a Waste Management Department (WMD) was set up in 1985 with the sole responsibility to collect, store, transport and dispose off wastes generated in the city. In the same year, a Metro Public Health Department (MPHD) was established by re-structuring the erstwhile Medical Officer of Health Department (MOHD) and the Mechanical Engineers Department (MED). The restructuring together with other internal re-organization led to clear definition and differentiation of roles among various actors within the WMD (see Fig. 4).

The WMD was mandated to execute the waste management portfolio, consisting of the following:

- management of solid and liquid wastes,
- management of public toilets, public cleansing, namely street and pavement sweeping, drain cleansing and lawn-mowing and the repair and maintenance of sanitation vehicles and equipment.
- the WMD was also permitted to use the service fees it collected to fund its operations. This laid the foundation for the establishment of "financial autonomy" for the new WMD,
- the MPHD, with the mandate for the execution of the other components of environmental sanitation, supported the WMD with the enforcement of the relevant bye-laws, education of the public on environmental health issues, and monitoring of waste management services delivered by the WMD to the various communities of the metropolis, and
- the Legal Department supported the WMD to draft appropriate bye-laws to facilitate delivery of services, the enforcement of compliance and the prosecution of offenders.

Additionally, a Waste Management Advisory Committee (WMAC) was established with oversight responsibility for waste management and reporting to the Executive Committee of the AMA on the activities of the WMD. Membership of the WMAC was made up of:

- two Assembly Members, who were serving on the then Sub-Committee of Environmental Sanitation and Health, one of whom was appointed Chairman,

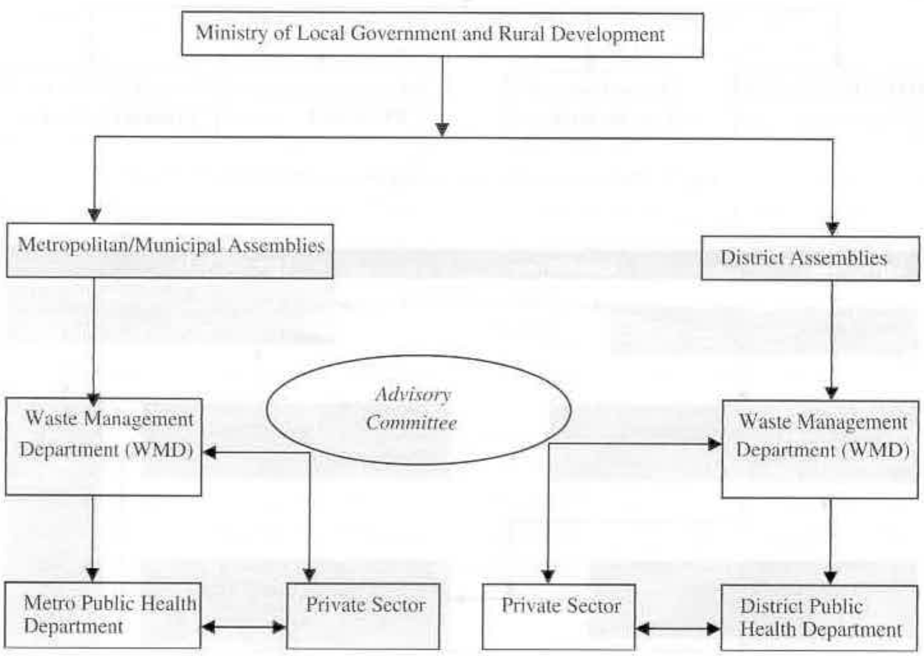


Fig. 4. New institutional arrangement (1985-Present).

- head of the Metro Finance Department,
- head of Metro Works Department,
- head of MPHD,
- head of Department of Town and Country Planning,
- head of the Environmental Health Division of the Ministry of Health,
- representative of the Council for Scientific and Industrial Research (CSIR),
- representative, School of Engineering (Environmental Department) of the University for Science and Technology,
- head of the WMD as Member/Secretary.

3.1. Introduction of public–private partnerships in waste collection in Accra, Ghana

Public–private partnerships (PPPs) are described in the literature as the transfer and control of the responsibility of service provision (e.g. public services such as healthcare, utilities, water supply and sanitation, etc.) from public sector entities, either wholly or partly, to the privately own establishments (Armah, 2001; Majani, 1996; Kgathi and Bolaane, 2001; Bruggemann, 1971; Darko and Fletcher, 1998; Gwebu, 2003; Mtani, 2004; Forsyth, 2004; Massoud and El-Fadel, 2002; Eliah, 2000; Fobil and Atuguba, 2004). The concept involves a wide range of PS participation in public services and serves as a potential strategic management tool for improved performance in public services delivery (Mbuligwe, 2004; Massoud and El-Fadel, 2002). The concept was introduced in Ghana alongside many other public sector reforms such as the Structural Adjustment Program (SAP), Economic Recovery Program (ERP), etc., so as to salvage the systemic down-turn in performance of public institutions. There is strong consensus among development experts that PPP is a "magic" solution to most inefficiency problems in public institutions. Therefore conclusions in development literature are almost always unanimous that the deployment of PPPs in public services provision in developing countries can cause a twitch in development trends by reason of the following:

- improved performance of the public sector by employing innovative operation and maintenance methods,
- reduced and stabilized costs of providing services by ensuring that work activities are performed by the most productive and cost effective means,
- improved environmental protection by dedicating highly skilled personnel to ensure efficient operation and compliance with environmental requirements, and
- access to private capital for infrastructure investment by broadening and deepening the supply of domestic and international capital (Mbuligwe, 2004; Mtani, 2004; North, 1999; Massoud and El-Fadel, 2002; Nhapi et al., 2003).

While in development literature, the concept of PPPs implementation largely subsumes effective cost recovery, efficient financial management, improved economies of scale, cost, improved public accountability, better institutional management, etc., the Accra analysis may likely point out that, for PPPs to achieve long-term sustainability, it may require active and continuous examination of rendered services to determine whether they are more appropriately and effectively performed by the PS. And that partial devolution of controls and responsibilities by local authorities may eventually lead to a complex interplay of external forces, which provides for the return of state controls and un-necessary political interference. Nonetheless experiences from countries where these have been tried show that the improved performance of public institutions through the implementation of PPPs can be counter-productive if the cultural and traditional contexts of the societies in which they are implemented are ignored at the design stages (North, 1999; Forsyth, 2004; Massoud and El-Fadel, 2002; Eliah, 2000; Fobil and Atuguba, 2004; Nhapi et al., 2003).

4. Impacts of the corrective measures

The outcomes of reorganization of these Institutions, Departments and private sector participation in waste collection are presented in Tables 1–3 and Fig. 4 (Fobil, 2002). The major observation was that the institutional rearrangement led to improved institutional efficiency and

Table 1
Planning data, targets and actual performance for liquid waste management, Accra (5-year development plan, 1996–2000)

Year	1996	1997	1998	1999	2000
Population					
Total	1,521,566	1,579,386	1,639,403	1,701,700	1,766,364
Growth rate (%)	3.8	3.8	3.8	3.8	3.8
Waste generation					
l/cap.day	0.44	0.44	0.44	0.44	0.44
m^3/yr	244,531	253,823	263,468	273,480	283,872
Targets (m^3/yr)					
AMA	90,000	91,500	91,000	91,000	91,000
Private sector	22,000	24,000	28,000	32,000	36,000
Total	112,000	115,500	119,000	123,000	127,000
% Coverage (%)	46	46	45	45	45
% Private sector (%)	20	21	24	26	28
Performance (m^3/yr)					
AMA	62,230	62,600	39,700	38,980	#N/A[a]
Private sector	21,030	21,100	26,440	53,540	#N/A
Total	83,260	83,700	66,140	92,520	#N/A
% Coverage (%)	34	33	25	34	#N/A
% Private sector (%)	25	25	40	58	#N/A

[a]N/A means non applicable, as data was not available.

Table 2
Planning data, targets and actual performance for solid waste management (1992–1995)

Year	1992	1993	1994	1995
Population				
Total	1,314,490	1,363,126	1,413,562	1,465,864
Growth rate (%)	3.7	3.7	3.7	3.7
Waste generation				
kg/cap.day	0.51	0.51	0.51	0.51
Tonnes/yr	244,860	253,920	263,315	273,057
Targets (T/yr)				
AMA	194,000	207,000	214,000	222,000
Private sector	2500	5000	7000	9000
Total	196,500	212,000	221,000	231,000
% Coverage (%)	80	83	84	85
% Private sector (%)	1	2	3	4
Performance (T/yr)				
AMA	195,840	153,490	199,715	110,800
Private sector	4660	18,730	23,640	29,500
Total	200,500	172,220	223,355	140,300
% Coverage (%)	82	68	85	51
% Private sector (%)	2	11	11	21

Table 3
Basic planning data for solid waste management in Accra (5-year development plan, 1996–2000)

Year	1996	1997	1998	1999	2000
Population					
Total	1,521,566	1,579,386	1,639,403	1,701,700	1,766,364
Growth rate (%)	3.8	3.8	3.8	3.8	3.8
Waste generation					
kg/cap.day	0.55	0.55	0.55	0.55	0.55
Tonnes/yr	305,664	317,279	329,335	341,850	354,841
Targets (T/yr)					
AMA	162,000	167,000	177,000	182,000	187,000
Private sector	75,000	90,000	105,000	120,000	135,000
Total	237,000	257,000	282,000	302,000	322,000
% Coverage (%)	78	81	86	88	91
% Private sector (%)	32	35	37	40	42

Source: WMD, AMA.

financial management. This in turn rekindled enthusiasm in the personnel and staff and led to improved service delivery. Furthermore, the introduction of private service providers resulted not only in improved collection performance, but also in increased collection coverage. This led to substantial waste clearing with corresponding improvements in urban environmental performance.

By the agreement between the PS group and the AMA, the PS paid some percentage of the income they made from the waste collection services to the Assembly. The total amount made from waste collection fees minus the percentage given to the AMA gave the net income of the PS groups. Percentage increase in coverage (% served) and frequency of garbage removal together gave the performance index, which was measurable from daily collection quantities in kilograms. Whether or not the PS scheme were successful depended on the margin of economic rent (which represented its net income). If profit margins were huge, then it meant there was enough economic incentive to encourage private individuals to continue with the scheme. The schemes then would be sustainable if profits offset operation costs. The Assembly made a saving of ¢46.32 million in the first year for deploying 319 conservancy laborers to the PS. In addition, the operators paid to the Assembly ¢164.52 million for storage, haulage and final disposal facilities it provided. In the immediate past, the Assembly would have earned within the same period ¢21.60 million from the 4500 registered pans paying ¢400.00 per month each. A plausible explanation of the low returns if AMA were to provide these services was that the AMA had not been efficient in its revenue generation. For example, flat rates that were charged for waste collection by the AMA were not realistic in economic terms as compared to charge per kilogram weight of waste collected by PS entities.

More pans were registered by the service providers, thereby reducing the numbers of unregistered pan users, which hitherto had been serviced by the laborers of the Assembly, without any payment of service fees to the AMA.

The experience was that fees were paid directly to the service providers for their services with low default rates of less than 15%, even though there was no effective law enforcement as there were no appropriate bye-laws to support the policy of PS participation in the delivery of the service.

Table 1 describes a 5-year development plan on liquid waste collection services in Accra. It shows targets and actual performance of liquid waste collection service providers. Targets were set based on the quantity of liquid wastes to be collected and institutional as well as logistics capacity of the metropolitan assembly (i.e. in terms of resources: human, financial, trucks, etc.). There were wide differences between AMAs's targets and actual performance because the assembly had over the several years of operation become so weary and fatigued that performance declined considerably. It maintained its ambitious targets, yet work capacity dwindled resulting in the observed widening gaps between targets and actual performance. On the other hand, the PS started with low targets because of restricted access to workforce but gained momentum over the years as the AMA gradually relinquished responsibility of waste collection service delivery to the PS which had performed creditably well within a short time bringing about substantial improvement in collection coverage. Although, much credit had been bestowed to the PS of having been successful in waste collection service delivery, it was not without shortfalls. There had been complains from residents about high rates of default of waste removal which generated substantial public nuisance. Additionally, we note that the PS involvement in waste collection only begun to come under experimentation in Ghana in 1985, an important observation was that it had shown to outperform the AMA in the liquid waste service delivery. Actual performance of PSP in desludging services increased steadily since 1996 with output beyond expectation from mid-term of the 5-year plan period. This was as a result of a large influx of cesspit emptiers brought in by resident businessmen. Also the returns on investment were so lucrative that Ghana nationals living abroad procured second-hand vehicles and shipped them to their relations to invest in private waste collection services so as to generate income and to lessen the burden of remittances to relations at home. The AMA reduced its duties/fees to a mere cost of supervision and monitoring.

Table 2 shows a cautious combination implementation of both AMA's traditional system of solid waste collection and the PS collection experimentation over a 4-year stretch from 1992 to 1995. During these pilot trials, the AMA held to its targets. However, with time, the PS gained some confidence from residents and submetro authorities, which reflected in the increased collection coverage and performance during the period under review. From the figures in Table 2, the AMA seemed justified in the cautious step it took in the PSP pilot program because of the low equipment holding of the contractors. Performance, however, increased progressively over the period as a result of acquisition of additional vehicles and equipment. Heap collection with hired tipper trucks undertaken early in 1992 accounted for the high AMA performance. However, the pilot provided a useful guide in planning for Phase II for PS participation in solid waste collection services for all income levels and for the two service types, namely house-to-house and communal container services.

In the second and expanded phase of PS participation in solid waste collection, the PS groups received tremendous support, which saw the transfer of greater solid waste collection service responsibility onto them. This was shown by a substantial increase in coverage of solid waste collection in the PS (see Table 3). However, the rapid growth in PS involvement and the subsequent increase in performance soon declined due to the reluctance of the assembly to monitor the private groups and political interference in the organization of waste collection service sector.

5. Lessons learnt: Ghana's experience of PPP in waste collection

The observations noted and lessons learnt from the implementation of PPP in waste management in Ghana are that:

- in identifying stakeholders, an assembly must undertake a careful study of the areas to be covered by the service, which should include any NGOs and CBOs, corporate and non-corporate entities operating within the communities,
- consultations with the identified stakeholders to discuss and establish roles and responsibilities,
- the Assembly must establish within its structures the lead organization and supporting units to plan, manage and monitor the delivery of the services as the Client and Regulator of the waste management responsibility,
- resources must adequately be provided for the monitoring of privatized services if value for money and the satisfaction of all stakeholders were to be achieved through PS participation in waste management,
- the Assembly's Strategic Waste Management Plan should be developed, clearly indicating the Plan period, the projections, targets, priorities and packaging of the services to be delivered and the areas for PS participation,
- preparation of appropriate and relevant documentation (tenders, agreements, etc.) and procedures for contract/franchise procurement; franchise services should be prepared well ahead of call for tender wherever possible,
- development of regulatory mechanisms to ensure participation of beneficiaries, efficient delivery of services and application of sanctions to penalize infractions and defaults backed by relevant and enforceable bye-laws of the Assembly, and
- undertake risk analysis to pre-empt conditions that may derail the PS participation process, typically political interventions at various levels. This may usefully be accomplished in conjunction with the Assembly's Strategic Waste Management Plan, which can be given legal force through its adoption by the Assembly.

6. Conclusions

The institutional arrangement for waste collection and management in Accra and within the country at large is elaborate. Hierarchically, the Ministry of Local Government, Rural Development and Environment is responsible for nation-wide drawing of guidelines for waste management. Under a decentralized policy, Metropolitan, Municipal and District Assemblies (DAs) are responsible for the implementation of these guidelines with some modifications if needed to meet local conditions. The status of waste collection and management in the Accra metropolis is unimpressive. It lacks long-term planning as well as faces an undue political interference. It does not seem to have canvassed the support of community members and residence. However, although the collection schemes have had some implementation drawbacks and imperfections, the institutional restructuring and the involvement of private sector groups had achieved some gains in the medium term (especially within the first 15 years after reforms) in the following respect:

- that there had been marked improvement in quality of services, especially in the Operation and Management (O&M) of public toilets.

- that service coverage for solid waste collection and disposal had improved from 51% of accessible areas in 1998 to about 91% in the year 2000. However, improved coverage did not necessarily result in waste clearing as many of the private entities defaulted,
- that the private sector in the waste sub-sector had shown growth in numbers as well as in capacity,
- that the Assembly had relinquished the role of direct involvement in delivery of services and had developed skills and strategies for effective supervision and monitoring, and finally
- that the residential areas that remained to be linked to the sewer system had been leased out on contract to private sector waste service providers who are now performing well above what AMA did prior to the privatization program.

In spite of the short-term gains, the achievements made seemed to have waned and the enthusiasm of private sector groups appeared to have eroded and its performance had substantially dwindled in the long term. After 20 years of reforms, the conditions that prevailed during the era preceding the institutional reforms and the private sector participation in waste management sector begun to re-emerge and appeared to have already engulfed the low-income communities. Whilst acknowledging that a formal evaluation of the performance of the waste management sector institutional reforms and the participation of private sector groups had not been conducted, it may, however, be concluded from the analysis of post-reform trends, institutional data, and synthesis of public documents that the private sector participation probably did not show long-term sustainability because of the following lapses in the design and implementation:

- that lack of transparency in the award of waste collection contracts, franchises, and leases may have weakened the selection process and that the award of contracts was not necessarily made to qualified contractors,
- that ill-defined implementation plans for public–private sector mix of waste collection system may have provided for its practical execution tremendously difficult and/or cumbersome,
- that non-monitoring of private waste collection contractors, franchisees and leases, and lack of enforcement of contract terms by the metro and sub-metro authorities may have given way for poor services delivery in the long term,
- that lack of qualified manpower within the metropolitan authority may have led to poor planning and execution of effective and sustainable programs and compromised their sustainability,
- that political interference in the administrative matters of the metro authority may have led to loss of initial focus of the public–private sector initiative in Accra and thus compromising their long-term sustainability, and lastly,
- that the complete absence and lack of involvement of the community members and residents: for whom the waste collection services are provided, during privatization of waste collection services, in waste management decisions and program design as was obvious in contract award documents may have led to lack of support of the program.

References

Agunwamba, J.C., 2001. Analysis of socioeconomic and environmental impacts of waste stabilization pond and unrestricted wastewater irrigation: interface with maintenance. Environmental Management 27 (3), 463–476.

Agunwamba, J.C., 2003. Analysis of scavengers' activities and recycling in some cities of Nigeria. Environmental Management 32 (1), 116–127.

Anikwe, M.A., Nwobodo, K.C., 2002. Long term effect of municipal waste disposal on soil properties and productivity of sites used for urban agriculture in Abakaliki, Nigeria. Bioresource Technology 83 (3), 241–250.

Armah, N.A., 2001. Private Sector Participation in Waste Management in Accra, a Case Study. Carl Bro Intl, Accra, pp. 1–37.

Asomani-Boateng, R., Haight, M., 1999. Reusing organic solid waste in urban farming in African cities—a challenge for urban planners. Third World Planning Review 21 (4), 411–428.

Bannerman, R.R., 2000. Conflict of technologies for water and sanitation in developing countries. Schriftenreiche des Vereins fur Wasser-, Boden-, und Lufthygiene 105, 167–170.

Boadi, K.O., Kuitunen, M., 2005. Environment, wealth, inequality and the burden of disease in the Accra metropolitan area, Ghana. International Journal of Environmental Health Research 15 (3), 193–206.

Bruggemann, H.B., 1971. Some practical aspects of water and waste water management and pollution control. Royal Society of Health Journal 91 (3), 144–146.

Carboo, D., Fobil, J.N., 2004. Physico-chemical analysis of municipal solid waste (MSW) in the Accra metropolis. West African Journal of Applied Ecology 5 (2), 116–117.

Darko, E.O., Fletcher, J.J., 1998. National waste management infrastructure in Ghana. Journal of Radiological Protection 18 (4), 293–299.

Eliah, E., 2000. Work from waste. World Work (34), 18–20.

EPA-Ghana, 1997. Guidelines for the construction and implementation of sanitary landfills. In: EPA of Ghana Quarterly, Accra, pp. 12–25.

Fobil, J.N., Atuguba, R.A., 2004. Ghana: changing urban environmental ills in slum communities. International Journal of Environmental Policy and Law 34 (4–5), 206–215.

Fobil, J.N., 2001. Factors to be considered in design of an integrated municipal solid waste management in the Accra metropolis. In: Environmental Science Programme, Faculty of Science. University of Ghana, Legon, Accra, p. 169.

Fobil, J.N., 2002. Municipal wastes collection and urban environmental management in Accra, Ghana. In: International Symposium on Environmental Pollution Control and Waste Management (EPCOWM'2002). Tunis, INRST and JICA, Tunisia.

Forsyth, T., 2004. Building deliberative public–private partnerships for waste management in Asia. In: Conference on Democratic Network Governance. Roskilde University, Roskilde University, Denmark.

GSS, 2004. Ghana Statistical Service Information.

Gulis, G., et al., 2004. Health status of people of slums in Nairobi, Kenya. Environmental Research 96 (2), 219–227.

Gwebu, T.D., 2003. Population, development, and waste management in Botswana: conceptual and policy implications for climate change. Environmental Management 31 (3), 348–354.

Holmes, J.R., 1983. Metropolitan waste management decisions in developing countries. Journal of the Royal Society of Health 103 (1), 25–32.

Kgathi, D.L., Bolaane, B., 2001. Instruments for sustainable solid waste management in Botswana. Waste Management Research 19 (4), 342–353.

Laryea, N.O., 2000. Challenges for health and water resources in the Birim districts of eastern Ghana. Schriftenreiche des Vereins fur Wasser-, Boden-, und Lufthygiene 105, 53–57.

Leschber, R., 2002. International report: sludge management and related legislation. Water Science and Technology 46 (4–5), 367–371.

Mabogunje, A.L., 1995. The environmental challenges in Sub-Saharan Africa. Environment 37 (4), 4–9, 31–35.

Majani, B.B., 1996. Coping with urban growth and development through environmental planning and management (EPM): the sustainable Dar es Salaam project. Urban Health Newsletter 1 (28), 26–31.

Makoni, F.S., et al., 2004. Impact of waste disposal on health of a poor urban community in Zimbambwe. East African Medical Journal 81 (8), 422–426.

Makule, D.E., 2000. Pollution of water sources due to poor waste management—the case of Dar-es-Salaam. Schriftenreiche des Vereins fur Wasser-, Boden-, und Lufthygiene 105, 117–121.

Martinson, F.E., Marfo, V.Y., Degraaf, J., 1999. Hepatitis E virus seroprevalence in children living in rural Ghana. West African Journal of Medicine 18 (2), 76–79.

Massoud, M., El-Fadel, M., 2002. Public–private partnerships for solid waste management services. Environmental Management 30 (5), 621–630.

Mbuligwe, S.E., 2004. Assessment of performance of solid waste management contractors: a simple techno-social model and its application. Waste Management 24 (7), 739–749.

Mtani, A., 2004. Governance challenges and coalition building among urban environmental stakeholders in Dar es Salaam, Tanzania. Annals of the New York Academy of Sciences 1023, 300–307.

Nhapi, I., Gijzen, H.J., Siebel, M.A., 2003. A conceptual framework for the sustainable management of wastewater in Harare, Zimbabwe. Water Science and Technology 47 (7–8), 11–18.

North, C.D., 1999. Institutions, institutional change and economic performance. In: Alt, J., North, D. (Eds.), The Political Economy of Institutions and Decisions. Cambridge University Press, Cambridge.

Nuno-amarteifio, N., 1995. The Accra experience. Countdown to Istanbul 1 (5), 14.

Post, J., Broekema, J., Obirih-Opareh, N., 2001. Trial and error in privatisation, experiences with urban solid waste collection in Accra (Ghana) and Hyderabad (India). In: Space and Place in Development Geography Conference, Utrecht.

Satterthwaite, D., 1993. Securing water for the cities. People Planet 2 (2), 13.

Stephen, C., 1999. Urban environment, health and poverty in developing countries: an analysis of differentials using existing data. In: Faculty of Medicine, University of London, London, UK.

Taiwo, D., 1996. Creating a strong working relationship between the governments and the community for sustainable development of Ibadan city, Nigeria. Urban Health Newsletter 1 (28), 70–76.

Von Schirnding, Y.E., 1996. Environmental planning and management in Greater Johannesburg. Urban Health Newsletter (28), 77–86.

Wurapa, F.K., 1973. Typhoid in Accra: a follow-up study of typhoid fever patients at the Korle-Bu Teaching Hospital. Ghana Medical Journal 12 (2), 184–188.

ELSEVIER

Journal of Environmental Management 86 (2008) 272–281

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Spatial and statistical assessment of factors influencing nitrate contamination in groundwater

Marco Masetti[a,*], Simone Poli[b], Simone Sterlacchini[c], Giovanni P. Beretta[a], Arianna Facchi[d]

[a] *Department of Earth Sciences, "Ardito Desio", University of Milan, Via Mangiagalli 34, 20133 Milan, Italy*
[b] *Department of Environmental and Territorial Sciences, University of Milan-Bicocca, Piazza della Scienza 1, 20126 Milan, Italy*
[c] *Institute for the Dynamic of Environmental Processes, National Research Council (CNR-IDPA), Piazza della Scienza 1, 20126 Milan, Italy*
[d] *Institute of Agricultural Hydraulics, University of Milan, Via Celoria 2, 20133 Milan, Italy*

Received 13 December 2005; received in revised form 7 November 2006; accepted 7 December 2006
Available online 12 February 2007

Abstract

The weights of evidence (WofE) modeling technique has been used to analyze both natural and anthropogenic factors influencing the occurrence of high nitrate concentrations in groundwater resources located in the central part of the Po Plain (Northern Italy). The proposed methodology applied in the Lodi District combines measurements of nitrate concentrations, carried out by means of a monitoring net of 69 wells, with spatial data representing both categorical and numerical variables. These variables describe either potential sources of nitrate and the relative ease with which it may migrate towards groundwater. They include population density, nitrogen fertilizer loading, groundwater recharge, soil protective capacity, vadose zone permeability, groundwater depth, and saturated zone permeability. Once conditional dependence problems among factors have been solved and validation tests performed, the statistical approach has highlighted negative and positive correlations between geoenvironmental factors and nitrate concentration in groundwater. These results have been achieved analysing the calculated statistical parameters (weights, contrasts, normalized contrasts) of each class by which each factor has been previously subdivided. This has permitted to outline: the overall influence each factor has on the presence/absence of nitrate; the range of their values mostly influencing this presence/absence; the most and least critical combination of factor classes existing in each specific zone; areas where the influence of impacting factor classes is reduced by the presence of not impacting factor classes. This last aspect could represent an important support for a correct land use management to preserve groundwater quality.

Keywords: Groundwater; Nitrate; Weights of evidence

1. Introduction

Nitrate contamination and the associated health concerns are among the most common problems adversely affecting groundwater quality worldwide (Canter, 1996). Nitrate (NO_3^-), is a groundwater contaminant of particular importance due to its widespread presence in aquifers and potential health and environmental impacts. Nitrate occurs naturally from mineral sources and animal wastes, and anthropogenically as a by-product of agriculture and human wastes. Madison and Burnett (1985) list the following, among others, as major anthropogenic nitrate sources: fertilizers, feedlots, dairy and poultry farmings, sewage systems and septic tank drainages. According to Baker (1990), nitrate concentrations in areas with a shallow water table may respond rather quickly to recharge and dilution events, depending on the pathways of water movements and nitrate sources. In fact, nitrate moves almost as a conservative tracer in the vadose zone (Bodier et al., 1993), mostly because of its high solubility and its tendency to be repelled by negatively charged soil particles (Keeney, 1986). Elevated nitrate concentrations in drinking water have been associated with low oxygen levels in the blood of infants, known as methemoglobinemia (Coinly, 1945), as a result from reduction of ingested nitrate to nitrite. A population-based case-control study in Nebraska found long-term consumption (30 years) of community

*Corresponding author. Tel.: +39 02 503 15507; fax: +39 02 503 15494.
E-mail address: marco.masetti@unimi.it (M. Masetti).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.023

water with average levels of nitrate 4 ppm was positively associated with a two-fold risk for non-Hodgkin's lymphoma (Weisenberg, 1990). Increased nitrogen levels have also shown negative effects on surficial water by expediting eutrophication and decreasing dissolved oxygen levels.

Determining areas where groundwater is at risk of nitrate contamination and which factors mainly influence nitrate presence in groundwater represents an important step in managing and protecting this resource and human health. In this study, both natural and anthropogenic factors influencing the occurrence of high nitrate concentration in groundwater have been considered and analyzed. Population density, nitrogen fertilizer loading, groundwater recharge, soil protective capacity, vadose zone hydraulic conductivity, groundwater depth, and saturated zone hydraulic conductivity are the factors used in this study. The model has been applied in the Lodi District (Northern Italy), an important agricultural territory, where the urban development is becoming increasingly relevant. The District is located within the Po Plain, a wide area identified as a Nitrate vulnerable zone by the European Community, as a result of the implementation of EU Nitrate Directive (91/976/EC), which aimed to prevent and reduce nitrate water pollution from agricultural sources.

Many authors have already correlated nitrate concentration with specific explanatory variables using GIS and/or statistical methods. Among others, Eckardt and Stackelberg (1995), Nolan et al. (2002), and Tesoriero and Voss (1997) used different logistic regression models to assess aquifer susceptibility to nitrate contamination considering natural and anthropogenic factors. Gardner and Vogel (2005) employed both logistic and Tobit regression methods to determine the relationship between groundwater nitrate and land use, while Lake et al. (2003) combined in a GIS a model of nitrate leaching with some hydrogeological features to identify areas of groundwater vulnerable to nitrate pollution. By determining the susceptibility to nitrate contamination and the overall importance of single factors influencing nitrate distribution in groundwater, studies outlined the existence of relationships which often vary locally and not always as expected. In this study the Weights of Evidence (WofE) modeling technique has been used to find and analyze correlations among selected factors and nitrate concentrations in groundwater. WofE brings along some important considerations, that helps to go deeper in the comprehension of these correlations: (1) the importance of each single factor class can be evaluated, allowing to determine the range of values influencing nitrate concentration, both positively and negatively. This permits to define, within each factor, guide values useful for managing future land use, safeguarding groundwater resources (e.g. maximum fertilizer loadings, maximum irrigation amount acceptable in each specific area); (2) the combination of two or more classes belonging to different factors, showing strong correlation with nitrate concentration, can be calculated; (3) it is also possible to determine where a specific impacting class of a factor commonly correlated with high nitrate concentration can be eventually neutralized by the presence, in the same area, of no-impacting classes.

WofE has been extensively applied in the field of mineral resources (Agterberg et al., 1993; Bonham-Carter et al., 1989; Carranza and Hale, 2000; Porwal et al., 2001) and then for landslide susceptibility (Lee et al., 2002; Sterlacchini et al., 2004; Van Westen et al., 1997). This work follows previous applications of the model performed by the authors (Alberti et al., 2001; Masetti et al., 2005) and represents, together with the Florida Aquifer Vulnerability Assessment project (Baker et al., 2002; Arthur et al., 2005), one of the first experiences in the application of the WofE in the hydrogeological field.

2. Methods

2.1. Study area

The Lodi District is located within the central part of the Po Plain (Northern Italy) at an average altitude of 64 m a.s.l. (Fig. 1). It covers an area of 715 km^2, mainly occupied by an extensive agricultural activity. Urban zones, about 10% of the whole district, are progressively increasing, leading to a consistent increment of population density in many areas. The population census of year 2001 (ISTAT, 2001) reports that about 80% of municipalities of the district have registered, in the last 10 years, an increase of population which is higher than 25% for about a third of these municipalities.

Sediments constituting this sector of the Po Plain are mainly composed by gravel and sand with frequent vertical and lateral changes to finer fractions, delimitating different aquifers. The study has been focused on the unconfined aquifer having a thickness of about 30 m. This aquifer has a shallow water table, mainly between 2 and 10 m depth, with seasonal changes of $\pm$1–2 m, correlated to infiltration processes due to precipitation and irrigation (Facchi et al., 2004) . The aquifer represents a groundwater reservoir providing a high sustainable yield and it is used for public, agricultural, and industrial water supply.

The territory is bounded on three sides by important rivers which exert a general draining effect on groundwater of the unconfined aquifer (Fig. 2a): Adda River on the east, Po River on the south and Lambro River on the west. Since the year 2000 the local Environmental Agency (ARPA) has monitored nitrate in groundwater in 69 wells. Well depth ranges from 13 to 27 m, while the depth to the bottom of the screen level from 12 to 25 m with an average screening length of 7 m; all wells are within the unconfined aquifer. Groundwater sampling follows a standard protocol which includes: (a) water pumping until temperature, conductivity and pH stabilized, (b) collection of water samples in glass containers at a temperature of 4 °C; (c) delivery to laboratory within 4 h. Samples have been analyzed by ARPA laboratory according to Italian Environmental Agency guidelines (APAT, IRSA-CNR,

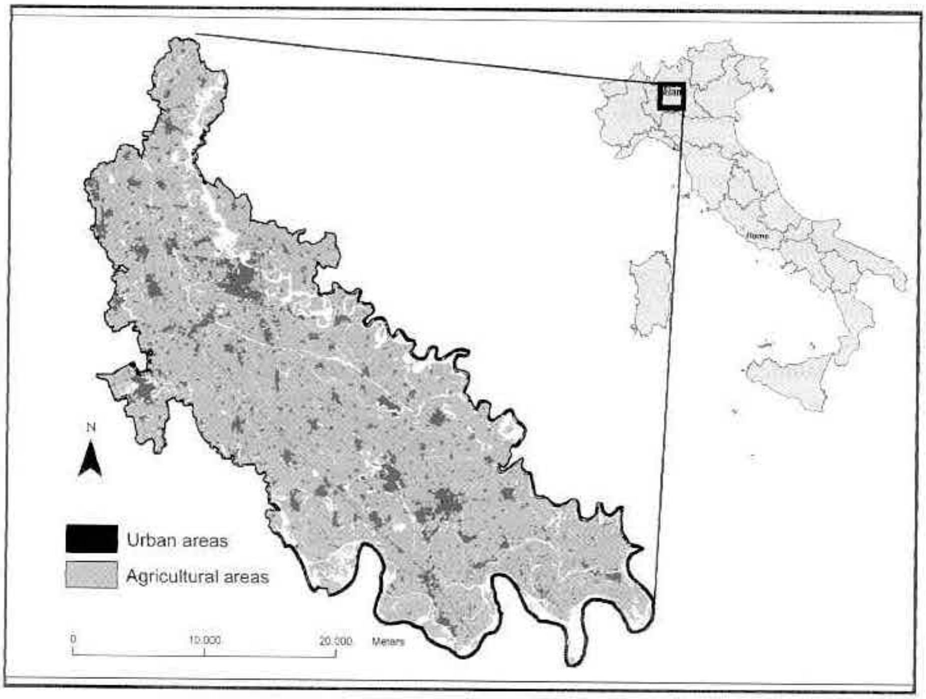


Fig. 1. Location of the study area.

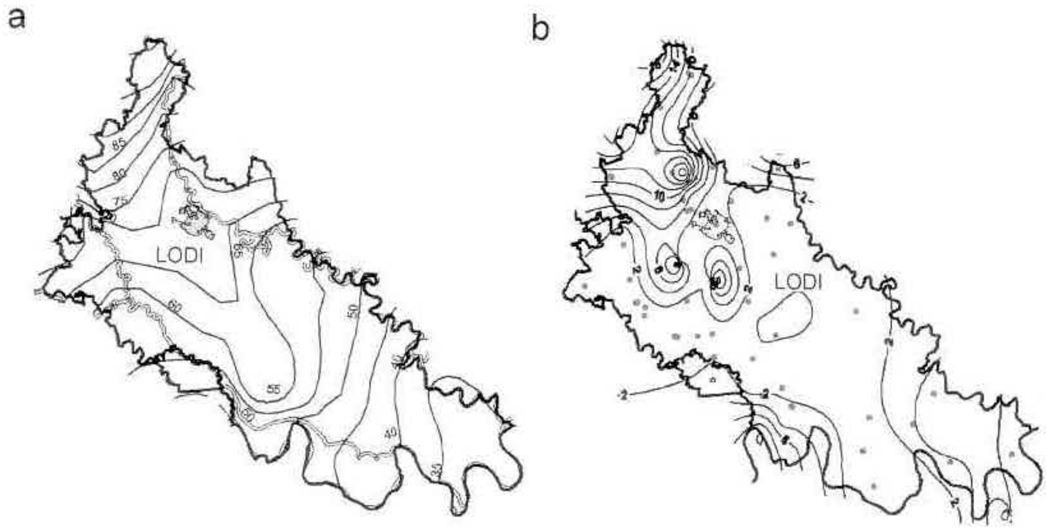


Fig. 2. (a) Piezometric map (levels in m a.s.l.) and (b) nitrate concentration map (values in mg/l) with monitoring points in gray.

2003), based on APHA standard method (1998). Measured NO_3^- concentrations in groundwater range from a minimum of 0.25 mg/l to a maximum of 25.5 mg/l, with a median value of about 5.0 mg/l. Nitrate distribution (Fig. 2b), measured at the beginning of 2003, shows the northern area as the most impacted sector, where concentrations are normally higher than 10 mg/l. From north to south a general decrease of nitrate occurs and concentrations are below 10 mg/l.

Analysis of historical data indicates a similar spatial trend in groundwater nitrate concentrations. Important changes are probably due to very local and transitory episodes of contamination which do not affect the general nitrate distribution in groundwater. Basic statistics of recent years are shown in Table 1.

2.2. *The WofE modeling technique*

The WofE modeling technique (Bonham-Carter et al., 1988) provides an effective tool for the integration of different hydrogeological information. In particular, this application exploits the exploratory power of WofE (Harris et al., 2003) in identifying single factors and their

Table 1
Main statistics of nitrate concentration in groundwater in recent years

Year	No. of samples	Mean (mg/l)	Median (mg/l)	Standard deviation (mg/l)	Minimum (mg/l)	Maximum (mg/l)
2000	67	4.6	4.8	3.4	0.25	22.5
2001	69	4.7	5.1	4.2	0.50	23.5
2002	68	4.3	4.7	3.8	0.50	25.5
2003	69	4.5	4.8	3.6	0.25	20.5

combinations that influence nitrate concentration in groundwater.

In effect, WofE is principally based on map-correlation and map-integration processes, in order to define relationships between spatial layers and combine predictor factors in supporting a hypothesis. It combines known occurrences of a phenomenon (training points or response variable) with available spatial data (predictor factors or explanatory variables) so as to perform either an analysis of the influence each factor has on the response variable and a predictive response. In particular, each training point is the location of what has to be predicted. In this work, this term represents "impacted" wells (wells with nitrate concentration above a threshold value). Moreover, patterns (or predictor factors) correspond to the variables inserted into the analysis, according to a specific conceptual model. Direct or indirect relations with sites have to be identified and analyzed.

The model uses a log-linear form of the Bayesian probability model and it is based on the concepts of prior and posterior probabilities. The prior probability $P\{D\}$ that an event occur per unit area is calculated as the total number of occurrences (training points, representing "impacted" wells) over the total study area, obtaining a probability of occurrence without considering any patterns.

This initial estimate can be later increased or diminished in different areas by the use of other evidences. If the spatial distribution of an evidence (B) is present (or absent), the probability of finding a new event given the presence (or absence) of the new evidence can be expressed as a conditional (or posterior) probability:

$$P\{D|B\} = P\{D\}\frac{P\{B|D\}}{P\{B\}},$$

$$P\{D|\overline{B}\} = P\{D\}\frac{P\{\overline{B}|D\}}{P\{\overline{B}\}}.$$

In this work, posterior probability gives a measure of aquifer susceptibility to nitrate contamination.

The positive and negative weights can be calculated for a pattern (Bonham-Carter, 1994):

$$W^{+} = \log_e \frac{P\{B|D\}}{P\{B|\overline{D}\}}, \quad W^{-} = \log_e \frac{P\{\overline{B}|D\}}{P\{\overline{B}|\overline{D}\}}.$$

When both prior and posterior probabilities are equal, $W^{+} = W^{-} = 0$ indicating that there is no association between an event and a prediction pattern.

For each predictor map, the difference between W^{+} and W^{-} defines the contrast ($C = W^{+} - W^{-}$), an overall measure of the degree of the spatial association between each class of the predictor factors and the response variable (the occurrence of an event). This parameter is very important for accepting or rejecting a class of a predictor variable. A positive value suggests that the pattern is a useful predictor and it has a direct relationship with the presence of "impacted" wells.

The Studentized value of C (normalized contrast) is calculated as the ratio of C to its standard deviation, $C/s(C)$. It provides a useful measure of significance of the contrast because of the uncertainties related to weights and missing data (Bonham-Carter, 1994; Raines, 1999). In this work, two criteria have been followed to consider this value as significant: each pattern must have a normalized contrast superior than 1.96 and for each pattern a justification from a hydrogeological point of view is also considered.

The WofE modelling technique has been applied using the Arc SDM® extension (Spatial Data Modeller) within the ArcView® 3.2 platform (Kemp et al., 2001). This extension improves integration of spatial datasets, the assessment of conditional independence problems, the evaluation of response theme uncertainty, and the execution of validation tests.

In particular, overall test of conditional independence and χ^2-test may be performed, defining whether conditional dependence problems can occur or not: the first is used as an overall assessment of conditional independence among datasets, while the latter is useful for analyzing conditional independence among couples of input factors, which could be conditionally dependent from a hydrogeological point of view.

In order to analyze correlations between explanatory variables and response variable, nitrate concentrations must be firstly converted into a binary form, establishing a threshold value which clearly separates events from no-events, that is wells with concentration higher and lower than a pre-fixed threshold value, respectively. The use of descriptive statistical measures to determine threshold value is widely accepted and used in literature (Panno et al., 2006 for a complete review). In this study the threshold has been chosen equal to 5 mg/l on the base of the median values in different years, as reported in Table 1.

2.3. Selected factors

In order to perform the analysis, both natural and anthropogenic factors have been selected. Maps representing spatial distribution of groundwater depth, vadose zone hydraulic conductivity and saturated hydraulic conductivity have been computed, while maps of groundwater recharge, soil protective capacity, population density and nitrogen fertilizer loading have been derived from existing sources of information.

Groundwater depth, vadose zone hydraulic conductivity and saturated hydraulic conductivity were determined through the collection and the critical review of well boring logs, where hydraulic tests had been performed. One hundred thirty wells have been selected according to their spatial distribution, the thoroughness of boring log description, and hydraulic test validity. Soil protective capacity has been obtained from the local Agency of Services for Agriculture and Forest (ERSAF, 2004). This variable describes soil capacity to reduce water-soluble polluting substances leaching from the surface and it is related to filtering and buffering capacity, due to both mechanical and biological/microbiological activity allowing degradation. Basic soil properties, as hydraulic conductivity, formation of saturated levels, grain size, pH, and cation exchange capacity (CEC) are considered. The model divides soils in three protective capacity classes: high, moderate, and low according to the combination of parameters.

Groundwater recharge map is the result of a comprehensive modeling study of surface and groundwater resources conducted for the Muzza-Bassa Lodigiana Irrigation District, which includes the Lodi District (Facchi et al., 2004, 2005). A distributed, integrated surface water–groundwater simulation system has been implemented and applied considering both rainfall and irrigation amounts. The system couples the conceptual vadose zone model ALHyMUS (Facchi et al., 2004, 2005) with the groundwater model MODFLOW (McDonald and Harbaugh, 1988). The simulation outcomes have been verified for the years 1999 and 2000 using the available data on hydrometric level in rivers as well as groundwater level in alluvial aquifer system. Recharge has been calculated on a year basis in mm/yr.

Nitrogen fertilizer loading have been carried out considering both organic and chemical load (Acutis and Provolo, 2003) which can be correlated to breeding and agricultural activities, respectively. Therefore, this factor only considers anthropogenic sources of nitrate not related to urban areas; nitrate contribution deriving from these areas, mainly due to septic tanks and leakage from sewer systems, have been expressed through the population density (Nolan et al., 2002) of 2001 (ISTAT, 2001).

Population density has been calculated on a municipal basis. So, a single value of this parameter has been associated with each polygon, representing the spatial distribution of each municipality.

3. Results and discussion

3.1. Susceptibility map and statistical validation

Each explanatory variable was trasformed into categorical data using different classification techniques, which created classes representing appropriate intervals of measured values.

For each predictor factor, weights and contrasts were firstly evaluated and then critically interpreted with consideration to their standard deviations and studentized contrasts. Through analysis of statistical parameters, the influence of some classes in determining low or high susceptibility values emerged. Moreover, according to hydrogeological considerations, some classes of predictor variables, not presenting an effective significance of the contrast value, have been reclassified, obtaining outcomes with an improved StudC values of calculated parameters (Table 2).

For example, in the first simulations vadose zone hydraulic conductivity and groundwater depth seemed not to be powerful in discriminating among susceptible and not susceptible areas. However, by combining the two factors and deriving a new one, which qualitatively represents the inverse of the time that water takes to reach groundwater table from the topographic surface, best results were obtained (see Fig. 3). This factor has been defined "1/ travel time".

Fig. 3 represents the spatial distribution of groundwater susceptibility to nitrate contamination in the Lodi District. Susceptibility classes have been defined on the base of posterior probability values according to this scheme: LOW indicates areas where Post prob < Prior prob, MODERATE where Prior prob < Post prob < 1.5*Prior prob, HIGH where 1.5*Prior prob < Post prob < 2*Prior prob and VERY HIGH where post prob > 2*Prior prob. These classes were chosen to better describe proximity to the prior probability value, that represents the main threshold between susceptible and not susceptible areas. Areas where Post prob is higher than Prior prob will be indicated in the text as "impacted areas". The most susceptible sectors (about 15% of the entire area, corresponding to "High" and "Very High" classes) are

Table 2
Example of how confidence can be improved: the case of nitrogen fertilizer loading

Classes of nitrogen fertilizer loadings	StudC values	Generalization	StudC values of the new classification
1	0	1	
2	−9.12	1	−9.79
3	6.05	3	6.05
4	−0.24	4	
5	7.44	4	5.28
6	4.06	5	4.06

located in the north-western part of the study area, while "Low" susceptibility classes, representing more than 60% of the total area, cover the central and southern sectors. The success rate curve has been calculated in order to evaluate the robustness of the final model (Chung and Fabbri, 1999), plotting on *X*-axis the cumulative percentage of susceptible area (starting from the high to low susceptibility values) and on the *Y*-axis the cumulative percentage of "impacted" wells. It shows how well the model is able to properly identify the training "impacted" wells in relation to the distribution of susceptibility values.

This curve (Fig. 4) shows that 85% of "impacted" wells fall within the 20% of the most susceptible area, while the Correctly Classified Points Index (percentage of "impacted" wells falling within impacted areas) is equal to 92%. These two parameters testify for a good quality of results (Van Westen et al., 2003). Because of the similarity in chemical data obtained in different years, as mentioned in Section 2.1, a complete validation test was not performed.

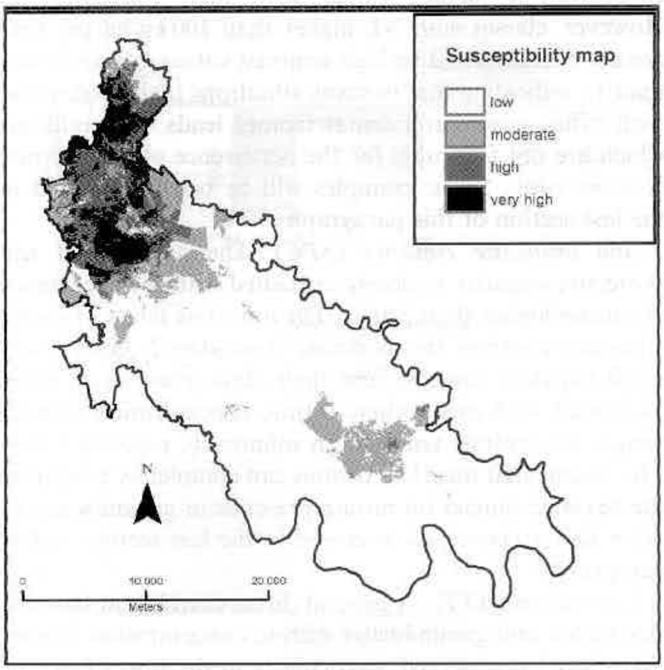


Fig. 3. Map representing classes of groundwater susceptibility to nitrate contamination.

3.2. Analysis of influencing factors

The influence of factors on nitrate presence/absence in groundwater were analyzed considering both their spatial distribution within the most "impacted" area and the positive and negative weights and contrasts (*C*). The distribution of factor classes in the total area and in the "impacted" area (area with posterior probability higher than prior probability), were compared by means of histograms. For example, groundwater recharge and soil protective capacity are represented in Fig. 5.

These factors identify two typical but different trends: an increase and a decrease of the portion of the "impacted" area with respect to the total area, by increasing class values. These graphs give a preliminary idea of the influence of these factors in relation to the nitrate occurrence in groundwater. It is important to note the lack of the higher soil protective class in the "impacted" area; this highlights the importance of this factor in controlling nitrate movement from surface to groundwater table.

Contrast values for each factor class were then analyzed to find: (1) correlation with nitrate concentration higher than 5 mg/l in groundwater and (2) the range of factor values which mostly affects the occurrence of this concentration. Figs. 6 shows histograms of *C* values with their standard deviations.

Groundwater recharge (GWR): A clear positive correlation between the increase of GWR and the occurrence of high nitrate concentration in groundwater is shown. Classes characterized by values lower than 500 mm/yr are associated with low concentration while classes with values

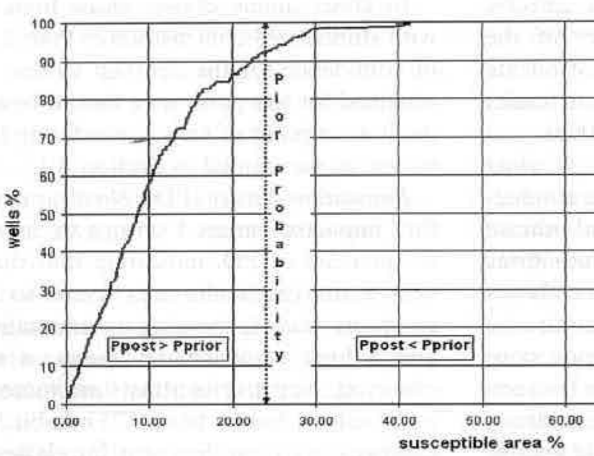


Fig. 4. Success rate curve. It shows that about 85% of "impacted" wells are inside the 20% of the entire area. This value coincides with areas characterized by a Post prob > Prior prob.

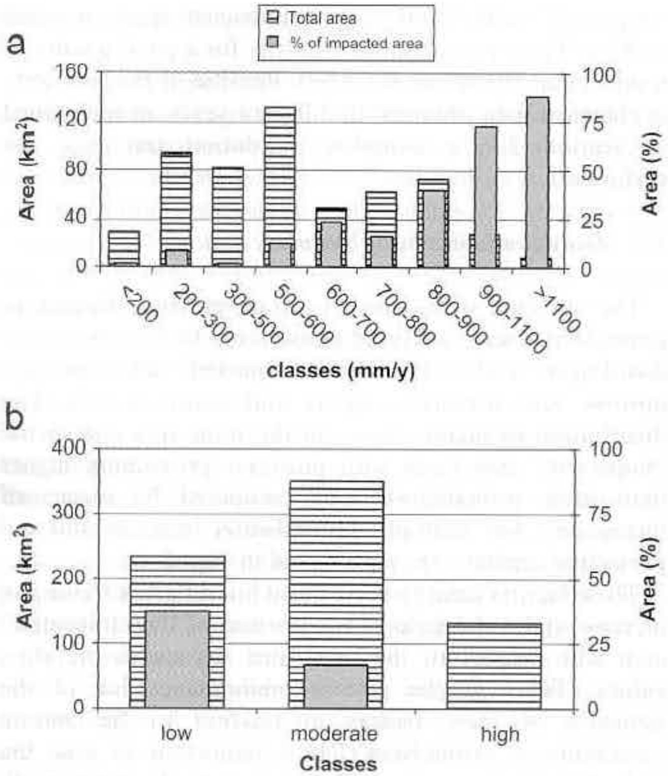


Fig. 5. Classes of groundwater recharge (a) and soil protective capacity (b): distribution in the total area (histograms with horizontal lines) and percentage of impacted area (gray columns).

higher than 900 mm/yr with high concentration, respectively. Classes in the range between 500 and 900 mm/yr have contrast values close to zero, indicating a general small influence on the presence of nitrate. The increase of GWR can influence nitrate concentration in groundwater in two opposite ways: with an increment of nitrate mass transport to groundwater through the vadose zone and with a better effectiveness in contaminant dilution. The conditions indicating the prevalence of one of the two processes are strictly site specific and should be checked out for every single study area. Results show that the progressive increase of nitrate mass transport directly related to GWR is not balanced by the effect of the increase in dilution, causing a general increment of nitrate concentration above the threshold value. Different results may be obtained in other areas (Masetti et al., 2005).

Saturated zone hydraulic conductivity (*KS*): A clear negative correlation is shown between hydraulic conductivity of the saturated zone of the aquifer and nitrate concentration higher than 5 mg/l. Classes with values lower than 5E−4 m/s show positive contrast values while classes with values higher than 1E−3 m/s have negative contrasts, resulting associated mostly with areas with nitrate concentration lower than 5 mg/l. Classes in the range between 5E−4 and 1E−3 m/s have contrasts close to zero, indicating a very small influence on the presence of nitrate greater than 5 mg/l. This trend can be correlated by two phenomena, both linked to the increase of groundwater velocity, and directly related to saturated hydraulic conductivity:

(1) the increasing effect of contaminant dispersion;
(2) a decrease in the residence time of contaminant in groundwater thus preventing the accumulation of contaminant below the surficial sources.

These factors can interact one another causing a general decrease in groundwater nitrate concentration.

Nitrogen fertilizer loading (NL): A clear direct correlation exists between an increase of NL and the occurrence of groundwater nitrate concentration greater than 5 mg/l. The value of 100 kg/ha per year can be considered a critical threshold; lower values are associated with low nitrate concentration, higher values with high concentration. However, classes with NL higher than 100 kg/ha per year are not characterized by high contrast values (always lower than 1), indicating that in many situations their association with other geoenvironmental factors leads to conditions which are not favorable for the occurrence of high nitrate concentration. Some examples will be briefly analyzed in the last section of this paragraph.

Soil protective capacity (*SPC*): The decrease of soil protective capacity is clearly explained with the occurrence of nitrate higher than 5 mg/l. The low class (class 1) shows a positive contrast, the moderate class (class 2) gives a very small negative contrast and high class (class 3) is never associated with areas where nitrate concentration exceeds 5 mg/l; its contrast tends to an infinitively negative value. This means that this class of soils can completely neutralize the negative impact on nitrate presence in groundwater of other factor classes, as discussed in the last section of this paragraph.

1/travel time (*TT*): A general direct correlation between this factor and groundwater nitrate concentration greater than 5 mg/l is observed, particularly in the lowest and the highest range values. The average classes confirm this general positive correlation but with some strong exceptions.

In effect, many classes show high standard deviation, with studentized contrast lower than 2, indicating low level of confidence for the contrast values. In any case, results obtained for this parameter should be analyzed only from a qualitative point of view, considering the derivation of this factor, as mentioned in Section 3.1.

Population density (PD): No clear correlation is seen with the "impacted" areas. Contrast values vary irregularly with the increase of PD, indicating that the influence of urban nitrate sources in the area seems to be small. However, excepting class 1, showing an anomalous positive contrast and a high standard deviations, a slight trend can be observed. Negative contrasts are more frequent for classes with values lower than 275 inhabit./km^2 while positive contrasts are more frequent for classes with values higher than 275 inhabit./km^2. This contrast distribution should be carefully taken into account considering the increase of

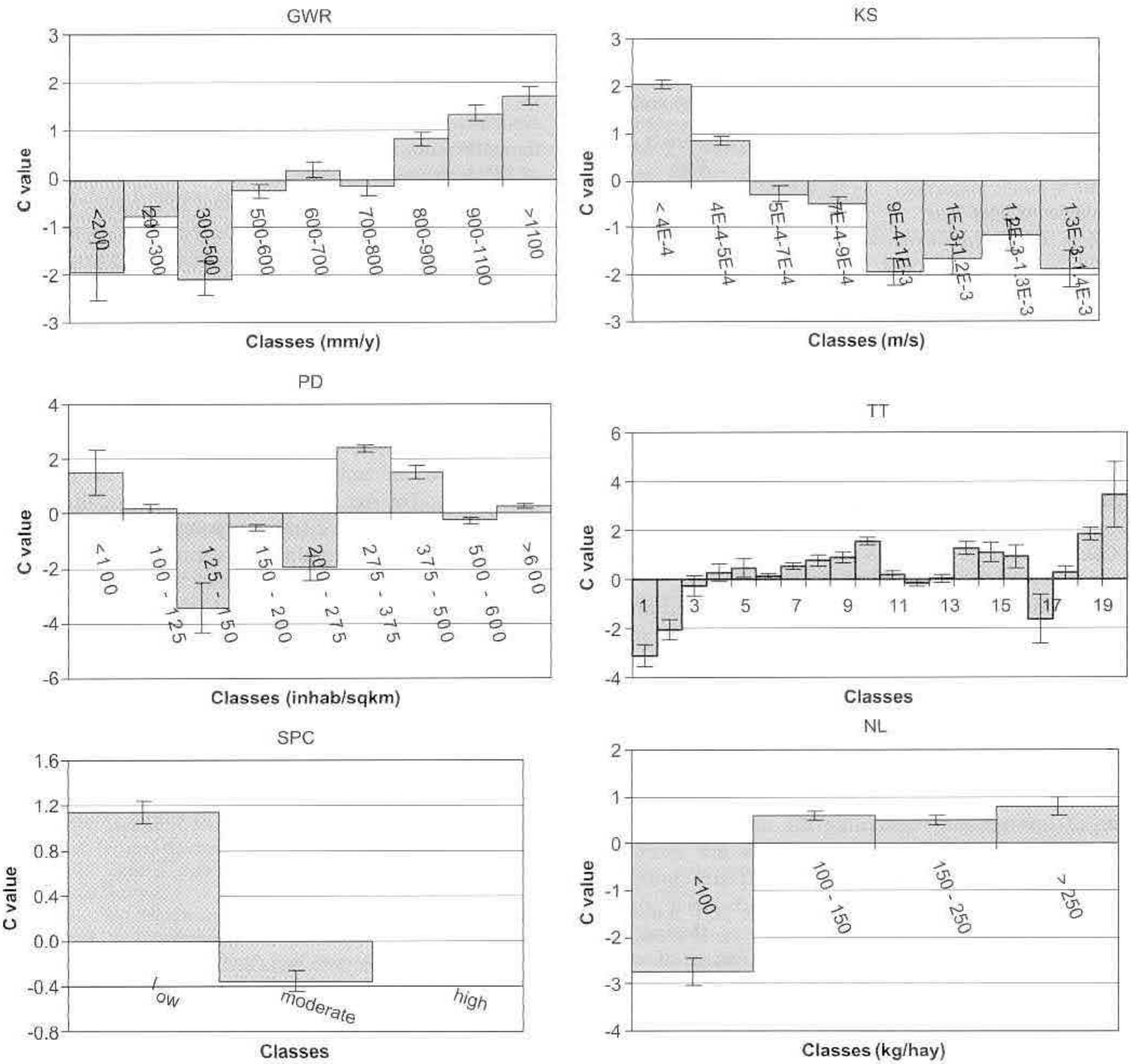


Fig. 6. Contrast and standard deviation values for each factor (GWR = groundwater recharge, KS = saturated hydraulic conductivity, PD = population density, TT = inverse of travel time, SP = soil protective capacity, NL = nitrogen fertilizer loading).

population occurring in the area and the consequent increase in PD.

A compared analysis of contrasts of the single factor classes (Fig. 6) shows that the most influencing factors are GWR, saturated zone hydraulic conductivity and soil protective capacity.

Although the exact shape of the trends may differ, they all have high negative contrast values for the not impacting classes and high positive contrast values for the impacting classes, which are correlated with nitrate concentration greater than 5 mg/l.

The WofE model allows also to perform an analysis of the importance of a combination of factor classes which can be found in different zones of the study area. For example, Table 3 shows the combination of the most important factor classes associated with the lowest and highest probability to find nitrate concentration greater than 5 mg/l. PD and inverse of travel time are not reported because of their wide range of values associated to these combinations.

Other important and more detailed analysis can be carried out on specific parameters. Soils with high protective capacity were not related to areas of nitrate concentration greater than 5 mg/l; however a correlation could be evidenced when these soils were associated with NL higher than 100 kg/ha per year in 36 km^2. This means that in these zones, despite the use of a medium–high load of fertilizer, groundwater is not significantly impacted by

Table 3
Combination of main factors most and least influencing nitrate concentration in groundwater >5 mg/l

Parameter	High probability	Low probability
Nitrogen fertilizer loading	>150 (kg/ha yr)	<100 (kg/ha yr)
Soil protective capacity	Low	High
Saturated hydraulic conductivity	<5E−4 (m/s)	>1E−3 (m/s)
Groundwater recharge	>900 (mm/yr)	<500 (mm/yr)

nitrate sources. This phenomenon also happens in zone characterized by moderate soil protective capacity and NL higher than 150 kg/ha per year when associated with GWR lower than 500 mm/yr and hydraulic conductivity of saturated soils higher than 1E−3 m/s. These results indicate the critical need for planning the location of agricultural practices, requiring high nitrogen load, with minimal impact on groundwater quality.

4. Conclusion

This study presents an analysis of natural and anthropogenic factors influencing nitrate occurrence in groundwater using the WofE modeling technique as a data exploratory tool. This technique allows the use of both categorical and numerical variables and combines the spatial distribution of each variable with the actual nitrate occurrence in groundwater above a selected threshold value. The outcomes of the WofE application outline a series of information concerning the impact that factors have on this occurrence: factor weights are specifically calculated for the study area and not arbitrarily chosen; the range of values for each single factor which has a positive, negative or no impact on groundwater is obtained; the combination of factors, again with their range of values, which results most or least critical in exceeding the threshold value, is determined for each specific area. The obtained results provide a useful tool to evaluate the potential effect of development on groundwater quality for future planning of land use, especially when regarding the decision of agricultural practice in a specific area.

GWR and saturated hydraulic conductivity are the variables which show the best correlation with nitrate presence/absence greater than 5 mg/l. The increase of GWR causes a general increase in nitrate concentration, indicating that in the unsaturated zone the mass transport process prevails over the dilution process; this is particularly efficient in areas where GWR is higher than 900 mm/yr. On the contrary, in the saturated zone the dilution process becomes very important, as shown by contrast values of saturated hydraulic conductivity classes; values of KS higher than 1E−3 m/s are associated with a lower probability to find nitrate concentration above 5 mg/l.

Soil protective capacity is another important parameter; zones where this parameter is high show a strong control on nitrate presence in groundwater given that they are never associated with nitrate concentration higher than 5 mg/l, even in areas of high NL. Many zones with moderate soil protective capacity, associated with high NL, are correlated with groundwater nitrate <5 mg/l if associated to other specific classes of factors, as GW lower than 500 mm/yr and KS higher than 1E−3 m/s. The impact of PD is not so well defined.

In this study, WofE has been used to combine the spatial distribution of wells with nitrate concentration above a threshold value with available predictor factors or explanatory variables. The value of 5 mg/l has been chosen on a simple statistical basis using the median value of nitrate concentration in groundwater. It is our opinion that the model may be effectively applied with different threshold values, selected on the base of regulatory limits, but always considering the actual nitrate concentrations. If the value is too high, the number of training points available could decrease to an unrepresentative sample of the entire population. On the contrary, if the nitrate threshold is too low, the number of training points could increase in a considerable way, becoming an over representative sample, which could not adequately identify low susceptible areas. In these cases the outcomes of the model should be probably inaccurate in determining the spatial distribution of groundwater susceptibility to nitrate contamination.

Acknowledgment

We would like to thank Gary Lee Raines of United States Geological Survey for providing valuable and constructive comments.

References

Acutis, M., Provolo, M., 2003. Stime dei carichi diffusi di Azoto, Fosforo e Fitofarmaci da agricoltura nelle acque di superficie della Lombardia. IReR, Piano di Tutela delle Acque della regione Lombardia.

Agterberg, F.P., Bonham-Carter, G.F., Cheng, Q., Wright, D.F., 1993. Weights of evidence modelling and weighted logistic regression for mineral potential mapping. In: Davis, Herzfeld, U.C. (Eds.), Computer in Geology, 25 Years of progress. Oxford University Press, Oxford, pp. 13–32.

Alberti, L., De Amicis, M., Masetti, M., Sterlacchini, S., 2001. Bayes' rule and GIS for evaluating sensitivity of groundwater to contamination. In: Proceedings of the International IAMG Conference 2001, Cancun, Messico.

Arthur, J.D., Baker, A.E., Cichon, J.R., Wood, A.R., Rudin, A., 2005. Florida aquifer vulnerability assessment (FAVA): contamination potential of Florida's principal aquifer system. Report submitted to the Division of Water Resource Management Florida Department of Environmental Protection, p. 148.

Baker, A.E., Cichon, J.R., Arthur, J.D., Raines, G.L., 2002. Florida aquifer vunerability assessment. Geological Society of American Abstracts with programs, vol. 34(6), p. 346.

Baker, D.B., 1990. Groundwater quality assessment through cooperative private well testing: an Ohio example. Journal of Soil and Water Conservation 45, 230–235.

Bodier, M.W., Frank, K.D., Spalding, R.F., 1993. Nitrate-N movement in a fine textured vadose zone. Journal of Soil and Water Conservation 48, 350–354.

Bonham-Carter, G.F., 1994. Tools for map analysis: map pairs. In: Merriam, D.F. (Ed.), Geographic Information Systems for Geoscientist: Modelling with GIS. Pergamon Press, New York, pp. 221–265.

Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F., 1988. Integration of geological datasets for gold exploration in Nova Scotia. Photogrammetric Engineering and Remote Sensing 54 (11), 1585–1592.

Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F., 1989. Weights of evidence modelling: a new approach to mapping mineral potential. In: Agterberg, F.P., Bonham-Carter, G.F. (Eds.), Statistical Applications in the Earth Sciences. Geological Survey of Canada paper 89–9.

Canter, L.W., 1996. Nitrates in Groundwater. Lewis Publisher, New York.

Carranza, E.J.M., Hale, M., 2000. Geologically constrained probabilistic mapping of gold potential, Baguio district, Philippines. Natural Resources Research 9, 237–253.

Chung, C.F., Fabbri, A.G., 1999. Probabilistic prediction models for landslide hazard zonation. Photogrammetric Engineering & Remote Sensing 65 (12), 1389–1399.

Coinly, H.H., 1945. Cyanosis in infants caused by nitrates in well water. Journal of the American Medical Association 129, 112.

Eckardt, D.A., Stackelberg, P.E., 1995. Relation of groundwater quality to land use on Long Island, New York. Groundwater 33 (6), 1019–1033.

ERSAF, 2004. Strumenti ed indirizzi per la gestione multifunzionale dei suoli agricoli (SIGMA). Piano di Tutela delle Acque della regione Lombardia.

Facchi, A., Ortuani, B., Maggi, D., Gandolfi, C., 2004. Coupled SVAT–groundwater model for water resources simulation in irrigated alluvial plains. Environmental Modelling and Software 19 (11), 1053–1063.

Facchi, A., Gandolfi, C., Ortuani, B., Maggi, D., 2005. Simulation supported scenario analysis for water resources planning: a case study in Northern Italy. Water Science and Technology 51 (3–4), 11–18.

Gardner, K.K., Vogel, R.M., 2005. Predicting ground water nitrate concentration from land use. Ground Water 43 (3), 1–10.

Harris, D.V., Zurcher, L., Stanley, M., Marlow, J., Pan, G., 2003. A comparative analysis of favorability mappings by Weights of Evidence, Probabilistic Neural Networks, Discriminant Analysis and Logistic Regression. Natural Resources Research 12 (4), 241–255.

ISTAT, 2001. General Population and Housing Census. 〈http://dawinci.istat.it/pop/〉.

Keeney, D.R., 1986. Nitrate in Groundwater: Agricultural Contribution and Control. In: Proceeding of the Agricultural Impacts on Groundwater conference. National Water Well Association, Columbus, OH.

Kemp, L.D., Bonham-Carter, G.F., Raines, G.L., Looney, C.G., 2001. Arc-SDM: arcview extension for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural network analysis. User manual.

Lake, I.R., Lovett, A.A., Hiscock, K.M., Betson, M., Foley, A., Sunnenberg, G., Evers, S., Fletcher, S., 2003. Evaluating factors influencing groundwater vulnerability to nitrate pollution: developing the potential of GIS. Journal of Environmental Management 68, 315–328.

Lee, S., Choi, J., Min, K., 2002. Landslide susceptibility analysis and verification using the Bayesian probability model. Environmental Geology 43, 120–131.

Madison, R.J., Burnett, J.O., 1985. Overview of the occurrence of nitrate in groundwater of the United States. US Geological Survey Water Supply Paper 2275, 93–105.

Masetti, M., Poli, S., Sterlacchini, S., 2005. Aquifer vulnerability assessment using weights of evidence modelling technique: application to the province of Milan, Northern Italy. Proceedings of IAMG 2005: GIS and Spatial Analysis, vol. 1, pp. 499–504.

McDonald, M.G., Harbaugh, A.L., 1988. A modular three-dimensional finite-difference ground-water flow model. US Geological Survey Techniques of Water-Resources Investigations, p. 586.

Nolan, B.T., Hitt, K.J., Ruddy, B.C., 2002. Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States. Environmental Science and Technology 36 (10), 2138–2145.

Panno, S.V., Kelly, W.R., Martinsek, A.T., Hackley, K.C., 2006. Estimating background and threshold nitrate concentrations using probability graph. Ground Water 44 (5), 697–709.

Porwal, A., Carranza, E.J.M., Hale, M., 2001. Extended weights-of-evidence modelling for predictive mapping of base metal deposit potential in aravalli province, Western India. Exploration and Mining Geology 10 (4), 273–287.

Raines, G.L., 1999. Evaluation of weights of evidence to predict epithermal-gold deposits in the Great Basin of the Western United States. Natural Resources Research 8 (4), 257–276.

Sterlacchini, S., Poli, S., Masetti, M., 2004. Spatial integration of thematic data for predictive landslide mapping: a case study from Oltrepo Pavese area, Italy. In: Lacerda, Ehrlich, Fontoura, Sayao (Eds.), Proceedings of the IX International Symposium on Landslides. Landslides: Evaluation and Stabilization. Taylor & Francis Group, London, pp. 109–115.

Tesoriero, A., Voss, F., 1997. Predicting the probability of elevated nitrate concentrations in the Puget sound basin: implications for aquifer susceptibility and vulnerability. Ground Water 35 (6), 1029–1039.

Van Westen, C.J., Rengers, N., Terlien, M.T.J., Soeters, R., 1997. Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. Geologische Rundschau 86, 404–414.

Van Westen, C.J., Rengers, N., Soeters, R., 2003. Use of geomorphological information in indirect susceptibility assessment. Natural Hazards 30, 399–419.

Weisenberg, D.D., 1990. Environmental epidemiology of non-Hodgyn lymphoma in Eastern Nebraska. American Journal of Industrial Medicine 18 (3), 303–305.

Journal of Environmental Management 86 (2008) 282–290

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Size distributions of fine and ultrafine particles in the city of Strasbourg: Correlation between number of particles and concentrations of NO_x and SO_2 gases and some soluble ions concentration determination

Estelle Roth[a], Damaris Kehrli[b], Karine Bonnot[b], Gwénaëlle Trouvé[b,*]

[a]Groupe de Spectrométrie Moléculaire Atmosphérique (GSMA), Moulin de la Housse, BP 1039, 51687 REIMS Cedex 2, France
[b]Laboratoire Gestion des Risques et Environnement, Université de Haute Alsace, 25 rue de Chemnitz, 68200 MULHOUSE, France

Received 26 September 2005; received in revised form 13 November 2006; accepted 7 December 2006
Available online 2 February 2007

Abstract

An Electrical Low Pressure Impactor (ELPI) was used during spring and autumn 2003 in the centre of Strasbourg for the measurement of atmospheric aerosols size distribution. The concentration of NO_x and SO_2 in air was simultaneously measured with specific analysers. Samples were collected in the range 0.007–10 μm in equivalent aerodynamic diameter size. Number distributions are representative of a pollution originating from urban traffic with a particle size distribution exhibiting a nucleation mode below 29 nm and an accumulation mode around 80 nm in size. A mean particle density equal to 39000 ± 35000 total particles per cm^3 with a size ranging from 7 to 10 μm was obtained after a sampling period of 2 weeks in spring. About 86.9% of the number of particles have an aerodynamic diameter below 0.1 μm and 13.1% between 0.1 and 1 μm. Correlation coefficients between the number of particles impacted on each ELPI plate and gas concentrations (SO_2 and NO_x) showed that the numbers of particles with diameter between 0.10 and 0.62 μm are highly related to the NO_x concentration. This result indicates that particles are traffic induced since NO_x is mainly emitted by cars as shown by measurements on various sites. Particles are less clearly correlated to the SO_2 concentration.

Particle analysis on different ELPI plates for a sampling period of 2 weeks in autumn showed high level of soluble NO_3^-, SO_4^{2-} and NH_4^+ ions. Indeed, up to 90% b.w. of these three species were found in the particle range 0.1–1 μm. The formation of particulate NH_4NO_3 is favoured by high NO_x concentration, which induces the formation of gaseous HNO_3.

Keywords: Aerosols; PM1; PM0.1; NO_x; SO_2; ELPI

1. Introduction

Urban aerosol is a mix of a background aerosol resulting from long distance transport and "fresh" locally emitted particles. The latter ones are mainly composed of organic compounds, elemental carbon, water, metal oxides and salts containing ammonium, sulphate or nitrate species depending on the source of emission (Cass et al., 2000). The scientific literature is well documented on chemical and physical characterization of TSP, PM10 and PM2.5 by collecting the aerosol at its source or in urban air (Ohta et al., 1998; Qin et al., 1997, 2003; Kleemann et al., 2000; Temesi et al., 2001, Alastuey et al., 2004; Putaud et al., 2004). However, few studies on precise chemical characterization of ultra fine particles (with diameter below 0.1 μm) in terms of heavy metals and ions concentration are available (Hildemann et al., 1991; Hughes et al., 1998; Cass et al., 2000; Pakkanen et al., 2001). Although these very small particles are not representative of the total suspended mass of particles, they largely dominate in number and may be responsible for some adverse health effects due to air-pollutant exposure. Ultra fine particles get deposited in the alveolar part of the lung where the adsorption efficiency for trace elements is as high as 60–80% (Linlay et al., 1997; Renwick

*Corresponding author. Tel.: +33 3 89 32 76 63; fax: +33 3 89 32 76 61.
E-mail addresses: Estelle.Roth@univ-reims.fr (E. Roth), Damaris.Kehrli@uha.fr (D. Kehrli), Karine.Bonnot@wanadoo.fr (K. Bonnot), gwenaelle.trouve@uha.fr (G. Trouvé).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.022

et al., 2001; Becker et al., 2005). Thus, particle size distribution studies in ambient air are of prime importance. Indeed, particle size distribution has already been studied in traffic-dominated places and PM10 number in the range 10^4–10^5 particles cm^{-3} were reported. (Shi et al., 1999; Weijeis et al., 2004; Gouriou et al., 2004; Pirjola et al., 2006). The numbers of particles may vary over a large range according to several parameters like traffic density. During high traffic episodes, particles may reach values equal to 10^6 particles cm^{-3} in immediate road vicinity (Gouriou et al., 2004) and decreases rapidly with increasing distance from the road (Gouriou et al., 2004; Gramotnev and Ristovski, 2004; Pirjola et al., 2006). Wind speed was also found to exert an effect on particles number which decreases at measurement points near to the roadside (less than 10 m) due to polluting gas dilution (Pirjola et al., 2006).

In this study, an urban aerosol was collected with an Electrical Low Pressure Impactor (ELPI) in the centre of the city of Strasbourg (France) during 2 weeks in spring and in autumn of 2003. This technique enables precise particle size distribution determination by measuring particles number concentrations for 12 ELPI collection plates between 7 nm and 10 µm. The period of time from in 2003 was devoted to particle size measurement and NO_x and SO_2 analysis in order to establish possible relationships between the particle size and the content of polluting gases in air.

Chemical speciation of species present in or on aerosols was determined during the autumn period of investigation by placing filters on each stage. Although chemical speciation of particles is primarily depending on their formation process, it also will be influenced by particle interaction with gaseous pollutants and/or gas to particle conversion. NO_3^-, SO_4^{2-}, NH_4^+ are the major water-soluble species concerned by the gas to particle conversion process: the oxidation of SO_2 leads to secondary sulphate accumulation on particles while NO_x leads to gaseous HNO_3 which may react with ammonium to produce NH_4NO_3 particles. Several studies have clearly demonstrated that particulate nitrate formation is linked to NO, NO_2, NH_3 and reactive organic gas (ROG) in local air (Kleeman and Cass, 1999). American source control strategies target the emissions of polluting gases, which react in the atmosphere to produce secondary aerosol species including sulphates, nitrates, ammonium ions and secondary organic aerosol. These include direct abatement in emission of SO_2, NH_3, ROG and NO_x. Kleeman and Cass (1999) demonstrated that these strategies which were originally designed to reduce ambient ozone concentration through the control of emissions of ROG and NO_x would also reduce fine particle mass concentrations primarily by reducing the nitrate aerosol (Kleeman and Cass, 1999).

In order to check the effect of polluting gas concentration on the chemical composition of aerosols, the major soluble ions fraction (NO_3^-, SO_4^{2-}, NH_4^+, Cl^-, K^+, Na^+, Ca^{2+}) was analysed on the particles accumulated on the 12 impactor stages during the autumn period of analysis.

2. Materials and methods

2.1. Sampling site

Field experiments were conducted in the centre of the French city of Strasbourg where high-density motor-vehicle traffic is prevailing. Two sampling periods, the first one from April 7 to April 20 (called spring period) and the second one from October 21 to November 5 (called autumn period) were carried out. In both the cases, the sampling site was located at less than 10 m from a large boulevard and at about 3.5 m above ground.

The two sampling periods (autumn and spring) were like homogenous, since we checked that averages of concentration of gases and number of particles were in the same order of magnitude during both periods.

2.2. Aerosol and gas sampling

Atmospheric particles were sampled at about 3.5 m from ground, using an ELPI (Dekati LTD). The cascade impactor system separates the particle matter following aerodynamic equivalent cut-off diameter at 50% efficiency in twelve particle size fractions ranging from 7 nm to 10 µm (Table 1). Before entering the impactor stages, particles are charged in a positive unipolar particle charger (corona charger) according to their Stockes diameter. After being charged by the corona charger, the atmospheric particles are introduced in the cascade impactor in order to be classified owing to their inertia and their aerodynamic diameter. A multistage electrometer enables to count the charged aerosol particles. Current is simultaneously measured for the 12 impactor stages and directly converted by the electrometer in particles number and concentrations using mathematical algorithms (Marjamäki et al., 2000). Particle number and size measurements with this particular sampling technology (ELPI) implies that the particle charging is function of the Stockes diameter and their

Table 1
Properties of the impaction stages of the ELPI

Stage	Holes	Aerodynamic diameter 50% (µm)
Filtration	—	0.007
1	69	0.0288
2	58	0.0571
3	21	0.0949
4	19	0.157
5	27	0.264
6	50	0.384
7	48	0.617
8	20	0.954
9	17	1.61
10	14	2.41
12	1	4.02
13	1	9.97

separation is function of the aerodynamic diameter of particles. Particle number and concentrations calculated accounts for this particularity (Marjamäki et al., 2000). Sampling was set at a constant flow rate of 9.81 L min^{-1}. Number distributions were sampled with the ELPI every 15 min.

NO_x analyses were carried out with an AC31 M analyser (Environnement SA) using a chemiluminescence's detector. The detection limit is about 0.35 ppb with a resolution time of 20 s. SO_2 measurements were done with an AF21 M analyser (Environnement SA) using UV fluorescence with a detection limit of about 1 ppb. NO_x and SO_2 measurements were done every 15 min on the sample site in order to operate at the same interval of time as for particle measurements.

2.3. Analytical procedure

For analytical determination of the soluble chemical fraction on particles, the aerosol was accumulated on quartz filters (Millipore) placed on each impactor stage during the autumn period. After collection, each filter was untidily placed in a Petri cell before being swathed in aluminium paper. These samples were stored in tightly closed plastic bags and at a temperature equal to 4 °C waiting for further chemical analysis. Water-soluble species (sulphate, nitrate, chloride, ammonium, potassium, calcium, sodium ions) were extracted with 4 mL deionised water in an ultrasonic bath for 60 min.

Capillary electrophoresis (WATERS QUANTA 4000) was used for ion quantification and applying the procedures described by Romano and Krol (1993) and Chen and Cassidy (1993) for anions and cations, respectively. Analytical procedures of anions were determined in a 4.60 mM chromate electrolyte solution buffered at pH 8 with a solution containing a mixture of 0.65 mM of boric acid and 0.50 mM of myristiltrimethylammonium hydroxide (OFM-OH$^-$). OFM OH$^-$ was purchased from Waters SA. A separation potential of 20 kV was applied after either 30 s hydrostatic injection at 2 kPa or 15 s with an electromigration injection potential of 2 kV when the electro migration injection mode was selected. The electro migration injection mode allows the determination of traces. The detection of anions was carried out by UV inverse spectroscopy at 254 nm.

Cation content was quantified in an electrolyte at pH 4.5 containing a mixture of 2-OH-2,4,6-cycloheptatrienone (tropolone), 18 crown-6-ether and UV CAT 2 (a crown-ether additive patented and purchased from Waters SA). A separation potential of 20 kV was applied after 30–60 s hydrostatic injection at 2 kPa. UV-inverse spectroscopy at 185 nm was used for detection of cations. Limits of determination (LOD) of ions were determined by using the same experimental protocol on five unexposed Teflon representing "blanks" (BL). Limits of detection and determination were assumed to be equal to three times and 10 times the value of standard deviation of the blanks, respectively. When a measured concentration was found below the LOD, the value was assumed to be zero.

3. Results and discussion

3.1. Particle size distribution

A mean value equal to $3.9 \times 10^4 \pm 3.5 \times 10^4$ particles cm^{-3} is obtained for the spring sampling period as shown in Table 2. It is close to the value reported by Shi et al. (1999) who measured aerosols in Birmingham (UK)

Table 2
Number densities (particles cm^{-3}) in every granulometric slide for the weekdays morning rush (6h30–8h00 a.m.), the weekends morning (6h30–8h00 a.m.), for a complete 2-weeks period during spring and for the background and the upper levels

Aerodynamic diameter of particles (nm)	Weekdays 6h30–8h00 a.m. (mean values)	Weekends 6h30–8h00 a.m. (mean values)	Two weeks (mean values)	Background level[a]	Upper level[b]
7 < Dae < 28.8	38,000	18,000	26,300	5450	86,000
28.8 < Dae < 57.1	6000	2300	4350	1100	15,000
57.1 < Dae < 94.9	6000	2300	4300	1000	15,300
94.9 < Dae < 157	4100	1600	3000	630	10,750
157 < Dae < 264	1600	800	1300	290	4200
264 < Dae < 384	1050	460	720	180	2100
384 < Dae < 617	540	220	360	100	1000
617 < Dae < 954	170	60	110	25	270
954 < Dae < 1610	34	11	22	3.3	62
1610 < Dae < 2410	6	1.8	3.5	0.07	11
2410 < Dae < 4020	1.6	0.4	0.8	0.02	2.9
4020 < Dae < 9970	0.6	0.5	0.4	0.04	1.3
Total particle number 7 < Dae < 9970	5.7×10^4	2.6×10^4	3.9×10^4	0.9×10^4	13.5×10^4

[a]The background level is estimated from the mean of the 10% lower values in 2-weeks sampling.
[b]The upper level is estimated from the mean of the 10% higher values in 2-weeks sampling.

also using an ELPI equipment. The number densities of particles for aerodynamic diameters between 7 nm and 10 μm vary from 0.9×10^4 to 13.5×10^4 particles cm^{-3} for the background and the upper level, respectively. Levels detected on the Amsterdam ring motorway were in the same range: $0.8 \times 10^4 cm^{-3}$ for the background level and 13.5×10^4 particles cm^{-3} during rush hours (Weijers et al., 2004). Since the particle concentration exponentially decreases when the distance from a busy road increases up to 60 m (Gramotnev and Ristovski, 2004), the immediate nearness of the sampling site to the road in the centre of Strasbourg (less than 10 m) implies a high influence of the traffic on the aerosol nature with a particle size distribution decreasing from 7 nm to 10 μm. Indeed, high numbers of particles with aerodynamic diameter ranging from 7 to 29 nm are found as shown in Table 2. This result confirms the importance of granulometric characterization of particles having a high potential of penetration in lungs. The mean number of particles is about 26300 cm^{-3} and may reach 86000 cm^{-3} in highly concentrated episodes (Table 2). Number size distributions represented in Fig. 1 in particles/cm^3/dlogDae from data in Table 2 show two modes. Size distribution data (number of particles, mass of particles, concentration of species in $ng\,m^{-3}$ is commonly normalized over a wide range of diameters. The units are divided by dlogDae (aerodynamic diameter) in the following figures. Particle numbers and concentrations correspond to the surface of the bar in the histograms. This unit is commonly used to draw logarithmic chemical distributions of ions, of particles (Pirjola et al., 2006; Harris and Maricq, 2001; Pakkanen et al., 2001).

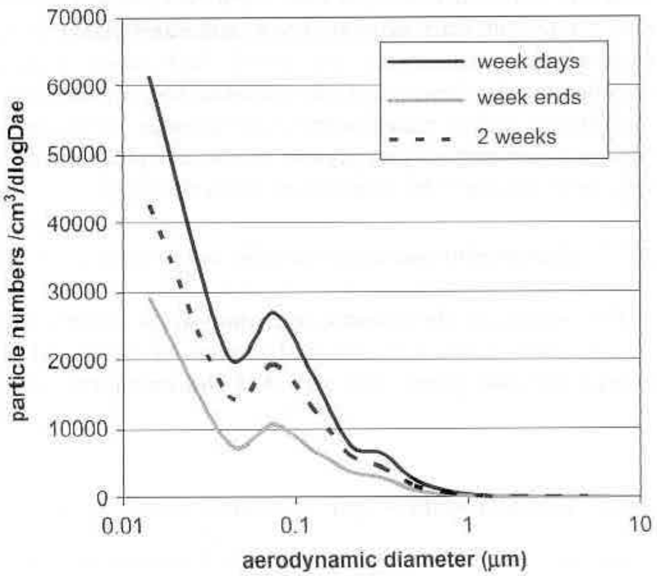


Fig. 1. Distribution of number of particles corresponding to a weekend mean, weekdays mean and a mean over 2 weeks during the spring period.

A strong nucleation mode for the smallest particles (7–29 nm) is observed. Also, an accumulation mode peaks at 60 nm whatever the event (mean of particles numbers over 2 weeks, mean on weekdays or weekends). Pirjola et al. (2006) using a better resolving system (50 channels between 3 and 50 nm with an SMPS against one channel for the 7–29 nm size range for the ELPI) reported distributions exhibiting a nucleation mode in the range 14.2–16.6 nm depending on the distance from road. Moreover, Gouriou et al. (2004) have registered distributions peaking at 60 nm using an ELPI equipment. High concentration of this size fraction in urban air mainly corresponds to diesel distributions. In this case, the nucleation mode does not appear in the distributions of particles collected directly at the tailpipe of cars (Harris and Maricq, 2001). In fact, the nucleation mode resulting from gas to particle conversion depends on both the dilution ratio and the amount of surface available for adsorption of volatile organics (Gouriou et al., 2004). In our case, the aerosol is not sampled at emission but after dilution explaining the presence of a nucleation mode in addition to the accumulation one.

A total of 86.9% of the numbers of particles have an aerodynamic diameter below 0.1 μm and 13.1% between 0.1 and 1 μm as shown in Table 3. A similar ultra-fine particle mode (<0.1 μm) ratio (82–87%) was found in Basel (Junker et al., 2000). The proportion of granulometric slides (PM0.1; PM0.1–1 and PM1–10) is nearly constant whatever the events considered (Table 3). This indicates that the relative aerosol composition regarding these large size fractions is constant: only the number of particles is changed due to anthropogenic activities. The weekdays mean morning rush is 1.5–2 times higher and the weekends mean around 6–8 a.m. is 1.5–2 times smaller than

Table 3
Percentage number of PM0.1, PM0.1–1 and PM1–10 for the weekdays morning rush (6h30–8h00 a.m.), the weekends morning (6h30–8h00 a.m.), the mean on 2 weeks and the background and the upper level during spring

Particle diameter (nm)	Weekdays 6h30–8h00 a.m. (mean values) (%)	Weekends 6h30–8h00 a.m. (mean values) (%)	Two weeks (mean values) (%)	Background level[a] (%)	Upper level[b] (%)
PM0.1	86.9	87.8	86.4	86.0	86.3
PM0.1–1	13.1	12.2	13.5	14.0	13.6
PM1–10	0.07	0.05	0.07	0.04	0.06

[a]The background level is estimated from the mean of the 10% lower values in 2-weeks sampling.
[b]The upper level is estimated from the mean of the 10% higher values in 2-weeks sampling.

the 2 weeks number densities (whatever the granulometric slides) as shown in Table 2.

3.2. Influence of gas type and concentrations to gas–particle conversion

3.2.1. Relation between number of particles and gas concentrations

In order to study to what extent the concentration of particle is related to gas production processes (exhaust gases of engines, heating processes), the correlation coefficients between the number of particles impacted on each ELPI plate and gas concentrations (SO_2 and NO_x) were determined and are given in Fig. 2. A set of 1320 instantaneous measures of particle numbers and gas concentrations was taken to have a good accuracy in the correlation coefficient determination. This set of measures was done every 15 min during the spring period between April 7 and 20, 2003.

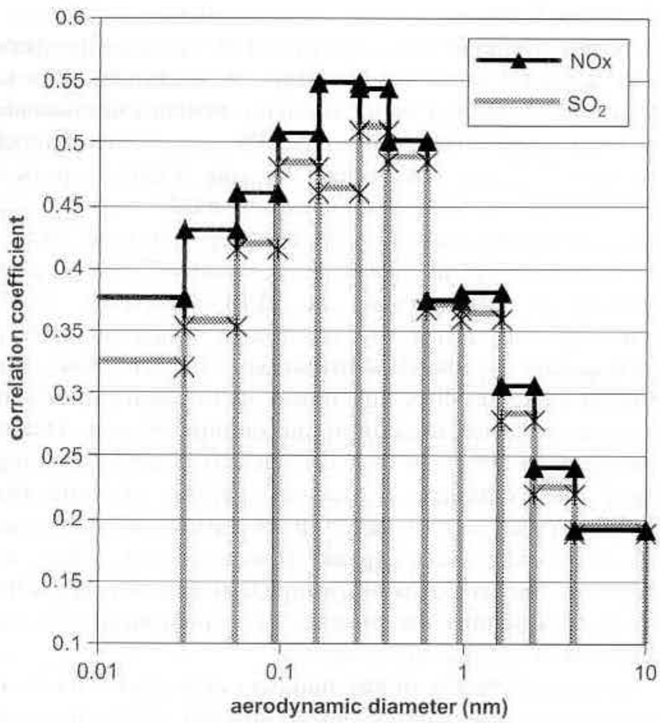


Fig. 2. Correlation coefficients between number of particles and SO_2 and NO_x concentrations over the spring sampling period.

Fig. 2 shows that the number of particles from 0.10 to 0.62 μm diameter is highly related to NO_x concentrations: the correlation coefficients are higher than 0.5.

NO_x are mainly emitted by cars according to the location of the measurement site. According to Harris and Maricq (2001) the particle size distribution corresponding to automotive emission is in the range 0.02–0.30 μm with a maximum between 0.06 and 0.10 μm. (Harris and Maricq, 2001). The particles emitted by motors are emitted together with a high level of produced NO_x, which induces gas–particle conversion to a significant extent. Accordingly, the particles and especially the finest ones may coagulate and therefore grow in size.

The number of particles is somewhat less correlated with the SO_2 concentration than with NO_x concentration (Fig. 2). In fact, the emission of SO_2 originates mainly from the sulphur present in the fossil fuel (coal, lignite, petroleum coke, heavy fuel oil, domestic heating oil, diesel oil). Hence, urban transportation contributes very little to SO_2 emission and may be present on the sampling site after transport through the atmosphere (Citepa, 2005).

3.2.2. Quantitative analysis of soluble ions on particles

The results of the quantitative analysis of soluble ions are shown in Table 4, in which the values of LOD and the blanks are also given. The reported concentration values

Table 4
Sized concentrations of soluble anions (ng m^{-3}), limits of determination (LOD in ng m^{-3}), blank values (BL in ng m^{-3}) and cations/anions ratio (neq m^{-3}) during autumn

Diameter (nm)	NH_4^+	K^+	Ca^{++}	Na^+	Cl^{**}	$SO4^{**}$	NO_3^{**}	Ratio Σcat/Σan[a]
7 < Dae < 28.8	0	0	29.8	31.0	7.2	26.6	76.7	1.42
28.8 < Dae < 57.1	0	0	0	0	10.3	8.0	18.7	0.00
57.1 < Dae < 94.9	0	0	0	0	0	15.3	7.6	0.00
94.9 < Dae < 157	12.4	0	0	0	0	41.6	14.7	0.62
157 < Dae < 264	24.7	0	0	0	0	54.9	32.2	0.83
264 < Dae < 384	50.8	0	0	0	0	124.1	76.4	0.74
384 < Dae < 617	278.1	11.5	6.6	0	0	241.2	281.0	1.68
617 < Dae < 954	299.5	6.3	6.2	0	0	354.4	449.8	1.16
954 < Dae < 1610	52.2	6.2	13.8	0	0	127.6	44.1	1.09
1610 < Dae < 2410	0	0	17.4	0	0	13.0	17.1	1.59
2410 < Dae < 4020	0	0	32.6	0	0	8.0	23.2	3.90
4020 < Dae < 9970	0	3.1	7.5	0	0	8.0	34.7	0.75
Dae > 9970	0	0	16.8	0	0	8.0	41.0	1.19
LOD	6.0	3.0	6.0	31.0	7.0	8.0	8.0	
BL	3.0	1.0	3.0	3.0	2.0	2.0	2.0	

[a] Ratio is equal to: Σcations/Σanions, expressed in neq m^{-3}.

are the average of three determinations by capillary electrophoresis. There are high levels of NO_3^- (1076 ng m^{-3}), SO_4^{2-} (1023 ng m^{-3}) and NH_4^+ (718 ng m^{-3}) associated with the particles size range between 7 nm and 10 µm. Also, a low concentration of Ca^{2+} (114 ng m^{-3}) and a negligible one for Na^+ (31 ng m^{-3}), K^+ (24 ng m^{-3}) and Cl^- (18 ng m^{-3}) are observed. Moreover, Mg^{2+} ions were not detected whatever the granulometric slide. The PM10 particulate nitrate level is typical of an urban atmosphere: Qin and Oduyemi (2003) detected 1107 ng m^{-3} of NO_3^- and Kaneyasu et al. (1995) reported 1200 and 900 ng m^{-3} of NO_3^- in two different towns in Japan during March 1988, respectively. Chloride and sodium contents in PM10 are mostly affected by sea aerosol. As an example, Kaneyasu et al. (1995) found 20 times more Na^+ and 30 times more chloride in a town near the ocean. A low concentration close to the LOD was measured for sodium and chloride ions in the centre of Strasbourg. Considering the geographic location of the city of Strasbourg, which is far from the sea a low concentration of Na^+ is expected. Water-soluble size distributions of K^+, Ca^{2+}, Na^+, Cl^- are shown in the Fig. 3. Indeed, Na^+ is only detected in the lowest stage with aerodynamic diameter ranging from 7 to 29 nm. Chloride is similarly distributed among the stages with aerodynamic diameters ranging between 7 and 57 nm (Fig. 3) and K^+ is mostly related to particles with aerodynamic diameters ranging from 384 to 954 nm (Fig. 3). Echelar et al. (1998) linked the presence of potassium and sodium in fine particles to biomass burning. Accordingly, the fact that wood is one of the major fuel source for heating of individual houses in the region Alsace where the city of Strasbourg is located may indeed explain the origin of the alkaline ions on the particles. The Vosgean solid mass constitutes an important firewood natural resource.

Water-soluble size distributions of ammonium, sulphate and nitrate are given in the Fig. 4. NH_4^+, SO_4^{2-} and NO_3^- ions exhibit a similar distribution with a maximum for particles with aerodynamic diameters between 617 and 954 nm (Fig. 4). These ions are predominantly associated to fine particles ranging from 0.1 and 1 µm in size. For instance, 92.5% b.w. of the total NH_4^+, 81% of the total SO_4^{2-} and 76.8% of the total NO_3^- contents are detected in this size range. Most of the sulphate and nitrate was formed through SO_2 and NO_x conversion in the atmosphere. Nitrates and sulphates are commonly found in particles larger than 0.2–0.3 µm in diameter and the highest concentration of water soluble inorganic species is found in particles larger than 0.5 µm in diameter. Hughes et al. (1998) previously described this chemical distribution in atmospheric particle during wintertime conditions in Pasaden . The contribution of sulphate and nitrate ions in ultra fine particles is in the same order of magnitude for the Strasbourg and Helsinki areas (Pakkanen et al., 2001).

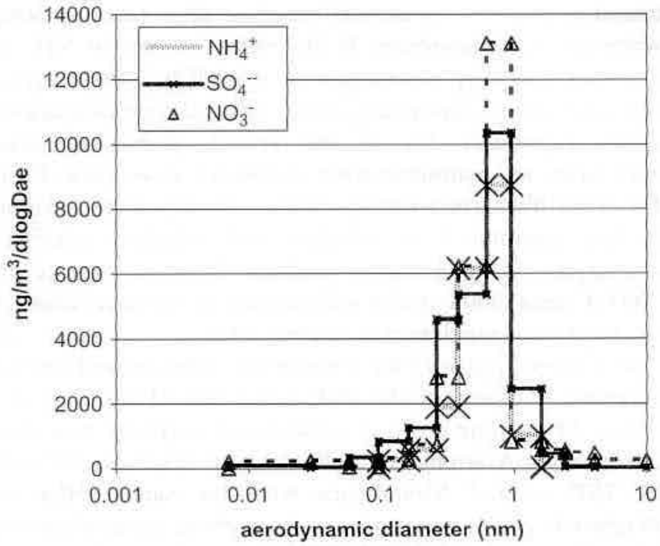


Fig. 4. Soluble ammonium, sulphate, nitrate concentration for different granulometric slides collected in the autumn sampling period.

Concentration value for NH_4^+ in ultra fine particles (diameter lower than 0.08 µm) was measured below the limit of determination. Taking into account the measurement uncertainty NH_4^+ could not be associated to the speciation of NH_4Cl since chloride was only detected in the smallest particles (7–29 nm in diameter) (Figs. 3 and 4). The low concentrations of nitrate and the absence of ammonium in particles with diameters larger than 2.5 µm may also be explained by the possible evaporation of NH_4NO_3 during the ELPI sampling. Pakkanen et al. (1999) previously demonstrated that a loss of a substantial fraction of nitrate may occur by its evaporation during the aerosol sampling with a Berner Low-pressure Impactor.

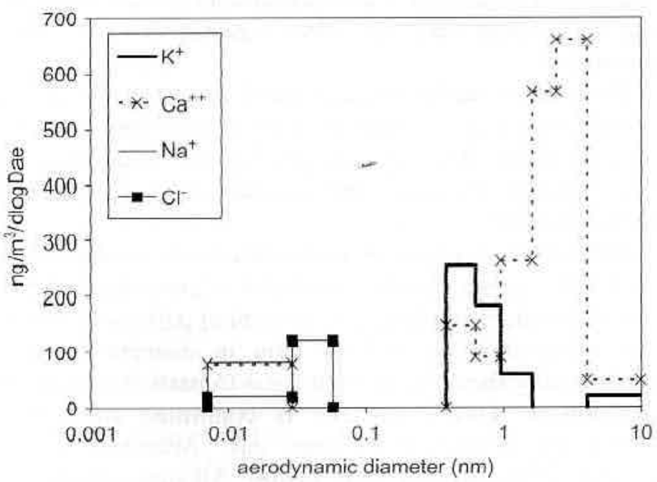


Fig. 3. Soluble potassium, calcium, sodium and chloride concentration for different granulometric slides collected during autumn sampling period.

The similar size distributions of ammonium, sulphate and nitrate ions in Fig. 4 suggest the existence of two mineral salts under the speciations of NH_4NO_3 and $(NH_4)_2SO_4$. This is particularly observed in the size range 0.1–1 µm. Most of the ammonium comes from NH_3

emission released by animal breeding after reaction with water in the atmosphere. If the major sources of NH_3 is livestock and crop production (around 97%), the remaining part comes essentially from ground transportation, which represents 2% of the French global emission and from the manufacturing industries (less than 1%). Emission from road transportation steadily increases due to the equipment of vehicles with selective catalytic converters. NH_3 emission reached 753 kt in France in 2003 (Citepa, 2005) and it will increase in the near future if no additional regulation is implemented.

In Europe, much of the ammonium sulphate and nitrate sulphate is found in the PM2.5 fraction (Putaud et al., 2004). More than 95% of ammonium sulphate was also observed by Alastuey et al. (2004) in the fraction $<0.7\,\mu m$ of TSP around Monanegra with the same order of magnitude for ammonium and sulphate concentration, respectively (Alastuey et al., 2004). According to the recent observation of Ynoue and Andrade (2004), nitrate and ammonium sulphates are also present in a simple unimodal size distribution over the Sao Paulo Metropolitan area in Brazil,. Consequently, PM1 measurements should register the main secondary inorganic species arising from high SO_2 and NO_x sources.

Ca^{2+} ions are rather associated to larger particles (384–9970 nm in diameter) with a maximum frequency in the diameter range 954–2410 nm. For instance, 69% Ca^{2+} b.w. is found on coarse particles (diameter between 1 and 10 μm) and 23.3% b.w. is located on PM0.1 particles. Moreover, it is well established that Ca species are essentially contained in coarse natural particles (5% b.w. in natural particles) originating from calcium silicates from soils and representing the major source of coarse particles in the atmosphere (Tanaka et al., 1981). However, recent leaching experiments from natural and anthropogenic aerosols show that more than 80% of Ca, K, Ni, Cu, Zn, S and Br are water-soluble species and are associated with anthropogenic sources as fly ashes (Desboeufs et al., 2001, 2005). These elements are mostly associated with soluble anions like carbonate, sulphate, nitrate and. Indeed, Clarke and Karani (1992) have indicated that calcium carbonate is a major constituent of particles in the atmosphere and therefore, one may assume that in the present study Ca is associated to other species than silicates although the particle size is quite large. Table 4 shows an excess of cations in the charge balance for particles with diameters above 0.4 μm in diameter. The occurrence of a high Ca^{2+} concentration clearly evidences the unbalanced electrical charge which therefore may be attributed to the presence of $CaCO_3$ in the coarser particles as already indicated by Clarke and Karani (1992). Also, $CaCO_3$ can be converted into $Ca(NO_3)_2$ after reaction with gaseous nitric acid as mentioned by Harrison and Kitto (1990). The presence of sulphate in the coarsest particle fraction also may be due to the presence of calcium sulphate (Querol et al., 1998). Sulphate segregation is due to a different mechanism of formation in which ammonium sulphate is mainly formed by a nucleation–condensation process, whereas calcium sulphate is formed by reaction of H_2SO_4 on the surface of coarse $CaCO_3$ particles (Querol et al., 1998). The mass balance expressed in neq m^{-3} done on the seven analysed ions points to a lack of cations for the electro neutrality completion on particles smaller than 0.4 μm in diameter. This lack of cations for small diameters range may be due to traces of cationic water soluble species of metals like Cu^{2+}, Fe^{2+}, Fe^{3+}, etc. These metals may come from anthropogenic carbonaceous matrices like fly ashes which exhibit a higher solubility and a larger dissolution rate in water than those originating from natural sources mainly composed of alumina silicates.

4. Conclusions

Two aspects of the aerosol near a busy road from the centre of Strasbourg were investigated following its separation in an ELPI in 12 size slides from 7 nm to 10 μm: the number concentrations size distributions and chemical characteristics regarding the soluble ions fraction composition. Particulate pollution in the city of Strasbourg is presented first in terms of particles numbers for a spring sampling period and secondly, in terms of chemical speciation for an autumn sampling. Total particles number range (10^4–10^6 particles cm^{-3}) is in agreement with previous studies for rush hours, background and mean levels (Weijers et al., 2004; Gramotnev and Ristovski, 2004). Number distributions show clearly two modes with a strong nucleation mode for the smallest particles (7–29 nm) and an accumulation mode peaking at 60 nm. It looks like the typical signature size distribution for diesel and gasoline engine exhaust.

The average ultra fine particle number in the atmosphere of Strasbourg during 2-weeks period with aerodynamic diameter between 0.007 and 0.1 μm represents 86.4% of the total fraction size. By making the ratio between two consecutive size ranges, it comes out that the relative aerosol composition is about constant whatever the events (traffic rush, night, etc.). For particles up to 1 μm diameter, this ratio varies only for $\pm 18\%$ against 60% for coarse particles.

Correlation studies between particles and polluting gas composition link the particles with aerodynamic diameter ranging from 0.10 to 0.62 μm to the occurrence of NO_x emitted by cars. The correlation coefficients are higher than 0.5.

The soluble fraction of particulate matter collected on the ELPI stages is mainly constituted of ammonium nitrate and ammonium sulphate. The content of particulate nitrate and ammonium centred on 1 μm in diameter may be explained by the occurrence of a gas to particle conversion mechanism. This hypothesis is confirmed by a high correlation coefficient between gas concentration and number of particles of 1 μm in size. Alkaline cations and chloride are present in the solid phase at a very low concentration close to the limit of determination. Calcium is the only alkaline-earth compound present in the soluble fraction of the particles.

Acknowledgements

The authors whish to thank Mr. J. Kleinpeter and Mr. E. Rivière from the "Association pour la Surveillance et l'étude de la Pollution atmosphérique en Alsace" (Strasbourg, France) for their kind technical contribution and helpful discussions.

References

Alastuey, A., Querol, X., Rodriguez, S., Plana, F., Lopez-Solar, A., Ruiz, C., Mantilla, E., 2004. Monitoring of atmospheric particulate matter around sources of secondary inorganic aerosol. Atmospheric Environment 38, 4979–4992.

Becker, S., Mundandhara, S., Devli, R.B., Madden, M., 2005. Regulation of cytokine production in human alveolar macrophages and airway epithelial cells in response to ambiant air pollution particles: further mechanistic studies. Toxicology and Applied Pharmacology 207, 269–275.

Cass, G.R., Hughes, L.A., Bhave, P., Kleeman, M.J., Allen, J.O., Salmon, L.G., 2000. The chemical composition of atmospheric ultrafine particles. Philosophical Transactions of the Royal Society of London A 358, 2581–2592.

Centre Interprofessionnel Technique d'Etude de la Pollution Atmosphérique (Citepa), 2005. Inventaire des émissions de polluants atmosphériques en France. Séries sectorielles et analyses étendues, Paris, France.

Clarke, A.G., Karani, G.N., 1992. Characterisation of the carbonate content of atmospheric aerosols. Journal of Atmospheric Chemistry 14, 119–128.

Chen, M., Cassidy, M.R., 1993. Separation of metal ions by capillary electrophoresis. Journal of Chromatography 640, 425–431.

Desboeufs, K.V., Losno, R., Colin, J.L., 2001. Factor influencing aerosol solubility during cloud processes. Atmospheric Environment 35, 3529–3537.

Desboeufs, K.V., Sofikitis, A., Losno, R., Colin, J.L., Ausset, P., 2005. Dissolution and solubility of trace metals from natural and anthropogenic aerosol particulate matter. Chemosphere 58, 195–203.

Echelar, F., Artaxo, P., Martins, J.V., Yamasoe, M., Gerab, F., Maenhaut, W., Holben, B., 1998. Long-term monitoring of atmospheric aerosols in the Amazon Basin: source identification and apportionment. Journal of Geophysical Research 103D, 31849–31864.

Gourion, F., Morin, J.P., Weill, M.E., 2004. On-road measurements of particle number concentrations and size distributions in urban and tunnel environments. Atmospheric Environment 38, 2831–2840.

Gramotnev, G., Ristovski, Z., 2004. Experimental investigation of ultrafine particle size distribution near a busy road. Atmospheric Environment 38, 1767–1776.

Harris, S.J., Maricq, M.M., 2001. Signature size distributions for diesel and gasoline engine exhaust particulate matter. Aerosol Science 32, 749–764.

Harrison, R.M., Kitto, A.M., 1990. Field intercomparison of filter pack and denuder sampling methods for reactive gaseous and particulate pollutants. Atmospheric Environment 28A, 2633–2640.

Hughes, L.S., Cass, G.R., Gone, J., Ames, M., Olmez, I., 1998. Physical and chemical characterization of atmospheric ultrafine particles in Los Angeles area. Environmental Science and Technology 32 (9), 1153–1161.

Hildemann, L.M., Markowski, G.R., Cass, G.R., 1991. Chemical composition of urban sources of fine organic aerosol. Environmental Science and Technology 25, 744–759.

Junker, M., Kasper, M., Röösli, M., Camenzind, M., Künzli, N., Monn, C., Theis, G., Braun-Fahrländer, C., 2000. Airborne particle number profiles, particle mass distributions and particle-bound PAH concentrations within the city environment of Basel: an assessment as part of the BRISKA Project. Atmospheric Environment 34, 3171–3181.

Kaneyasu, N., Ohta, S., Murao, N., 1995. Seasonal variation in the chemical composition of atmospheric aerosols and gaseous species in Sapporo, Japan. Atmospheric Environment 29, 1559–1568.

Kleeman, M.J., Cass, G.R., 1999. Effect of emissions control strategies on the size and composition distribution of urban particulate air pollution. Environmental Science and Technology 33, 177–189.

Kleemann, M.J., Schauer, J.J., Cass, G.R., 2000. Size and composition distribution of fine particulate matter emitted from motor vehicles. Environmental Science and Technology 34, 1132–1142.

Linlay, W.H., Stapleton, K.W., Zuberbulher, P., 1997. Fine particle fraction as measure of mass depositing in the lung during inhalation of nearly isotonic nebulized aerosols. Journal of Aerosol Science 28, 1301–1309.

Marjamäki, M., Keskinen, J., Chen, D.R., Pui, D.Y.H., 2000. Performance evaluation of the electrical low-pressure impactor (ELPI). Journal of Aerosol Science 31, 249–261.

Ohta, S., Hori, M., Yamagata, S., Murao, N., 1998. Chemical characterization of atmospheric fine particles in Sapporo with determination of water content. Atmospheric Environment 32 (6), 1021–1025.

Pakkanen, T.A., Hillamo, R.E., Aurela, M., Andersen, H.V., Grundahl, L., Ferm, M., Persson, K., Karlson, V., Reissell, A., Royset, O., Floisland, I., Oyola, P., Ganko, T., 1999. Nordic intercomparison for measurement of major atmospheric nitrogen species. Journal of Aerosol Science 30, 247–263.

Pakkanen, T.A., Kerminen, V.M., Korhonen, C.H., Hillamo, R.E., Aarnio, P., Koskentalo, T., Maenhaut, W., 2001. Urban and rural ultrafine (PM0.1) particles in the Helsinki area. Atmospheric Environment 35, 4593–4607.

Pirjola, L., Paasonen, P., Pfeiffer, D., Hussein, T., Hämeri, K., Koskentalo, T., Virtanen, A., Rönkkö, T., Keskinen, J., Pakkanen, T.A., Hillamo, R.E., 2006. Dispersion of particles and trace gases nearby a city highway: mobile laboratory measurements in Finland. Atmospheric Environment 40, 867–879.

Putaud, J.P., Raes, F., Van Dingenen, R., Baltensperger, U., Brüggemann, E., Facchini, M.C., Decesari, S., Fuzzi, S., Gehrig, R., Hüglin, C., Laj, P., Lorbeer, G., Maenhaut, W., Mihalopoulos, N., Müller, K., Querol, X., Rodriguez, S., Schneider, J., Spindler, G., Ten brink, H., Torseth, K., Wiedensolher, A., 2004. A European aerosol phenomenology-2: chemical characteristics of particulate matter at kerbside, urban, rural and background sites in Europe. Atmospheric Environment 38, 2579–2595.

Qin, Y., Oduyemi, K., 2003. Chemical composition of atmospheric aerosol in Dundee, UK. Atmospheric Environment 37, 93–104.

Qin, Y., Chan, C.K., Chan, L.Y., 1997. Characteristics of chemical compositions of atmospheric aerosols in Hong Kong: spatial and seasonal distributions. Science of the Total Environment 206, 25–27.

Querol, X., Alastuey, A., Lopez-Soler, A., Plana, F., Puicercus, J.A., Ruiz, C.R., mantilla, E., Juan, R., 1998. Seasonal evolution of atmospheric suspended particles around a coal-fired power station: chemical characterization. Atmospheric Environment 32, 719–731.

Renwick, L.C., Donaldson, K., Clouter, A., 2001. Impairment of alveolar macrophage phagocytosis by ultrafine particles. Toxicology and Applied Pharmacology 172, 119–127.

Romano, J., Krol, J., 1993. Capillary electrophoresis, an environmental method for the determination of anions in water. Journal of Chromatography 640, 403–412.

Shi, J.P., Khan, A.A., Harrison, R.M., 1999. Measurements of ultrafine particle concentration and size distribution in the urban atmosphere. Science of the Total Environment 38, 2831–2840.

Tanaka, S., Darzi, M., Winchester, J.W., 1981. Elemental analysis f soluble and insoluble fractions of rain and surface waters by particle-induced X-ray emission. Environmental Science and Technology 15, 354–357.

Temesi, D., Molnar, A., Meszaros, E., Feczko, T., Gelencser, A., Kiss, G., Krivacsy, Z., 2001. Size resolved chemical mass balance of aerosol particles over rural Hungary. Atmospheric Environment 34, 4347–4355.

Weijers, E.P., Khlystov, A.Y., Kos, G.P.A., Erisman, J.W., 2004. Variability of particulate matter concentrations along roads and motorways determined by a moving measurement unit. Atmospheric Environment 38, 2993–3002.

Ynoue, Y.R., Andrade, M., 2004. Size-resolved mass balance of aerosol particles over the Sao Paulo Metropolitan Area of Brazil. Aerosol Science and Technology 38, 52–62.

Journal of Environmental Management 86 (2008) 291–296

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Removal of *o*-nitrophenol from water by electrochemical degradation using a lead oxide/titanium modified electrode

Farid R. Zaggout*, Nasser Abu Ghalwa

Chemistry Department, Al-Azhar University-Gaza, P.O. Box 1277, PNA-Via, Israel

Received 30 September 2005; received in revised form 4 December 2006; accepted 7 December 2006
Available online 6 February 2007

Abstract

This study examined *o*-nitrophenol removal from aqueous solutions by electrochemical oxidation employing a modified electrode. The modified electrode was produced by electrodepositing lead oxide onto a titanium substrate. Following electrochemical oxidation of *o*-nitrophenol-containing solutions, the remaining *o*-nitrophenol concentration and chemical oxygen demand (COD) values were determined. The optimum parameters were current density of $40\,mA\,cm^{-2}$, pH of 2.47, 60 min of electrolysis time, $4\,g\,L^{-1}$ NaCl electrolyte solution and temperature of 30 °C. Under these optimum conditions of electrochemical degradation using a lead oxide/titanium modified electrode complete removal of *o*-nitrophenol and COD was achieved.

Keywords: Modified electrode; *o*-nitrophenol; Electrochemical oxidation; Wastewater treatment

1. Introduction

The efficiency of eliminating organic pollutants from water by electrochemical processes is dependent upon the type of electrode that is used and the potential/current conditions that are applied (Abu Ghalwa and Zaggout, 2006; Vase et al., 1998). When sodium chloride (NaCl) solution is electrolyzed in an individual cell during wastewater treatment by indirect electrochemical oxidation, a disproportionate amount of Cl_2 forms (Eq. (1)) at the anode, relative to the amount of OH^- ions generated at the cathode (Schmittinger, 1986; Newman and Tiedemann, 1978; Callwell, 1981; Novak et al., 1982).

$$Cl_2 + 2OH^- \rightarrow ClO^- + Cl^- + H_2O. \tag{1}$$

If the decomposition potential of a solvent is greater than that of a targeted impurity or if a suitable electrocatalyst can be found, then direct electrochemical oxidation can be used to remove that impurity from wastewater (Ibl and Vogt, 1981; Tilak et al., 1981; Gutman and Murphy, 1981; Kyriacous and Jannakoudis, 1986; Fry and Britton, 1986; Baizer and Lund, 1985).

Traditional methods (Hamza and Hamoda, 1980) of dealing with textile industry wastewater consist of various combinations of biological, physical and chemical methods. These have become inadequate due to the large variability of the composition of textile industry wastewaters. Treatment of textile wastewater from a large dyeing and finishing mill by a continuous process of combined chemical coagulation, electrochemical oxidation and activated sludge treatment, was investigated by Sheng and Peng (1996). The treatment efficiency of the combined process was assessed by measurements of chemical oxygen demand (COD) and color (turbidity) reduction. The effects of several operating variables, namely wastewater flow rate, conductivity, pH, applied current and amount of polyaluminum chloride (PAC), upon electrochemical oxidation efficiency of the textile wastewater were also examined (Vase et al., 1998). Electrochemical treatment of textile wastewater with a high Cl^- ion concentration employing Ti/RuO_2, Ti/Pt and Ti/Pt/Ir electrodes was investigated by Polacro et al. (1999). The efficiency of the three electrodes obeyed the following order in removal of organic materials present in water $Ti/RuO_2 > Ti/Pt > Ti/Pt/$

*Corresponding author. Tel.: +972 0599 344981.
E-mail addresses: zaggout2004@yahoo.com (F.R. Zaggout), n.ghalwa@alazhar-gaza.edu (N. Abu Ghalwa).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.033

Ir. It has been reported that *para*-nitrophenol was degraded by electroxidation using Bi/PbO_2 to determine the mechanism of the reaction (Borras et al., 2003). The United States Environmental Protection Agency has listed *o*-nitrophenol as a priority pollutant (Keith and Telliard, 1979). It may be released into the environment via wastewater and from fugitive emissions that are produced during dye production and synthesis. The oxidation of phenol has been widely studied due to its extensive presence in industrial wastewaters (Borras et al., 2003). *o*-Nitrophenol serves as a chemical intermediate in the production of pigments, rubber chemicals, lumber preservatives, photographic chemicals, and fungicide agents. The objective of this work was to develop an electrochemical system for highly efficient removal of *o*-nitrophenol from wastewater.

2. Materials and methods

2.1. Reagents

All chemicals and reagents used in this study were used as received. NaCl and NaOH were analytical grade. HCl (37%), $Pb(NO_3)_2$, NaF, HNO_3, H_2SO_4 (96%) and *o*-nitrophenol were laboratory grade. The chemicals and reagents used in these experiments were purchased from Aldrich. Distilled water was used throughout the work.

2.2. Preparation of the modified electrode

Pretreatment of the titanium sheet (IMI 115), 1 cm × 1.4 cm, was carried out according to the following procedures. The titanium sheet was polished on 320-grit paper strips using water as a lubricant, and subsequently treated with 1 mm siliceous paste blasting. The sheet was then degreased in 40% NaOH, cleaned in a hot 1:1 (v/v) mixture of HNO_3 and H_2SO_4 and finally washed in water. The treated surface was immersed in a boiling aqueous solution of oxalic acid (15%) until the TiO_2 had dissolved. Active metal oxide coating was then carried out immediately to minimize TiO_2 formation (Polacro et al., 1999).

The electrodeposition of PbO_2 was performed under a constant anodic current (100 mA, 30 min) from a 0.1 M HNO_3 solution containing 0.5 M $Pb(NO_3)_2$ and 0.04 M NaF. During electrolysis, the potential ranged from 1.5 to 1.8 V. Electrodeposition was carried out for 30 min, which produced a PbO_2 loading of about 14 mg cm^{-2} (Polacro et al., 1999). In order to verify the reproducibility of the electrode preparation, the experiments were repeated in triplicate. Our results demonstrated good reproducibility (within 5%) of the current efficiencies measured on the three electrodes. Furthermore, the electrodes were stable enough to be used in consecutive runs. The analyses of solutions revealed that no metal dissolution occurred during electrolysis.

2.3. Electrolysis

The absorption maximum (λ_{max}) of *o*-nitrophenol occurred at 450 nm. The cell used was made of transport Perspex (150 cm^3). The anode (modified electrode 1 cm × 1.4 cm) was supported in a vertical position midway between and parallel to the stainless steel cathodes. The distance between the cathode and the anode was 3 cm and 50 ml of sample solution was used in each experiment. Electrical current was provided by connection to a direct current power supply (model GP4303D, LG Precision Co. Ltd., Korea). Optimum operation conditions of the electrocatalytic oxidation process were determined by assessing degradation of *o*-nitrophenol and COD removal. Current was measured with a digital multimeter (Kyoritsu, model 1008, Japan). The following operation conditions were the variables examined: type of conductive electrolyte, current density (mA cm^{-2}), duration of electrolysis (min), pH, temperature (°C), conductive electrolyte concentration (g L^{-1}) and initial nitrophenol load concentration (mg L^{-1}).

2.4. Analyses

Treatment efficiency was evaluated by determining remaining *o*-nitrophenol concentration (mg L^{-1}) and COD (mg O_2 L^{-1}), via a closed reflux titrimetric method (Greenberg et al., 1992; Awad and Abu Ghalwa, 2005). A UV–vis spectrophotometer (1601 Shimadzu) was used to measure *o*-nitrophenol absorbance. Colorimetric analysis of the remaining *o*-nitrophenol concentration was based upon measured decreases in the absorbance of the nitrophenol solution as a result of the electrochemical oxidation process. The absorbance of remaining *o*-nitrophenol was measured at λ_{450nm}. All experiments were performed in triplicate and the standard deviation was computed and indicated in the plots as error bars with a confidence limit of 95%. The leached lead that could be present in the water after degradation was determined using a Flame Atomic Absorption Spectrometer. The instrument was a Shimadazu Atomic Absorption Spectrometer model (AA-6601F), and a homogeneous acetylene–air flame was used. The optimum conditions of the instrument were: wave length 283.3 nm, slit width 0.5 mm and lamp current 8–300 mA. A calibration curve was prepared by using a standard solution of lead. The absorbance of each standard was measured after the deionized water as a blank. A calibration curve of absorbance versus concentration was prepared at the time the samples were to be analyzed.

2.5. Effect of the type of conductive electrolyte

Electrochemical oxidation removal efficiency of *o*-nitrophenol and COD was determined in HCl (4 g L^{-1}), NaOH (4 g L^{-1}), and NaCl (4 g L^{-1}) at pH (2.47) solutions. The following operating conditions were used during these experiments: 40 mA cm^{-2} current density, 30 °C temperature and initial o-nitrophenol load concentration of $100\,mg\,L^{-1}$. The reactions were allowed to proceed for 60 min.

2.6. Effect of current density

The treatment processes were carried out for 60 min under the following conditions: pH 2.47, 30 °C, initial o-nitrophenol load concentration of $100\,mg\,L^{-1}$ and NaCl concentration of $4\,g\,L^{-1}$.

2.7. Effect of pH

In order to investigate whether o-nitrophenol removal is pH dependant, the pH of the solution was varied by adding drops of HCl and NaOH solution using a pH-meter while the other conditions were kept constant. The reactions were carried out for 60 min with an initial load concentration of $100\,mg\,L^{-1}$, a current density of $40\,mA\,cm^{-2}$, a temperature of 30 °C and a NaCl concentration of $4\,g\,L^{-1}$.

2.8. Effect of electrolysis time

To assess the effect of electrolysis time, experiments were conducted with operating treatment conditions that were consistent with those described above (current density of $40\,mA\,cm^{-2}$, pH 2.47, 30 °C, NaCl $4\,g\,L^{-1}$ and initial load concentration $100\,mg\,L^{-1}$), but with different electrolysis times ranging from 0 to 90 min.

2.9. Effect of NaCl concentration

Different concentrations of the electrolyte NaCl ranging between 1 and $6\,g\,L^{-1}$ were used to remove o-nitrophenol from solutions with a load concentration of $100\,mg\,L^{-1}$, pH 2.47, current density of $40\,mA\,cm^{-2}$ and temperature of 30 °C.

2.10. Effect of temperature

Similar to the previous experiments, a current density of $40\,mA\,cm^{-2}$, pH of 2.47, initial load concentration of $100\,mg\,L^{-1}$, NaCl concentration of $4\,g\,L^{-1}$ and 60 min electrolysis time were used to test the effect of temperature on removal efficiency. These experiments included variable o-nitrophenol load concentration in the range of $10–300\,mg\,L^{-1}$.

3. Results and discussion

3.1. Effect of the type of conductive electrolyte

The removal efficiency of o-nitrophenol and COD was determined. The results are expressed in terms of remaining o-nitrophenol concentration ($mg\,L^{-1}$) and COD removal ($mg\,O_2\,L^{-1}$). As shown in Table 1, NaCl was the most effective conductive electrolyte for nitrophenol removal. Therefore, NaCl was used as the conductive electrolyte in subsequent experiments. These experiments also demonstrated the feasibility of using a PbO_2/Ti anode for o-nitrophenol removal by electrochemical oxidation. The simulated effluents contained mainly NaCl as the conductive electrolyte, and had a pH of 2.47.

Table 1
Effect of conductive electrolyte type on o-nitrophenol and COD removal using a PbO_2/Ti modified electrode

Conductive electrolyte	Remaining conc. of o-nitrophenol ($mg\,L^{-1}$)	COD ($mg\,O_2\,L^{-1}$)
NaCl	0.00	0.00
HCl	54	80
NaOH	43	64

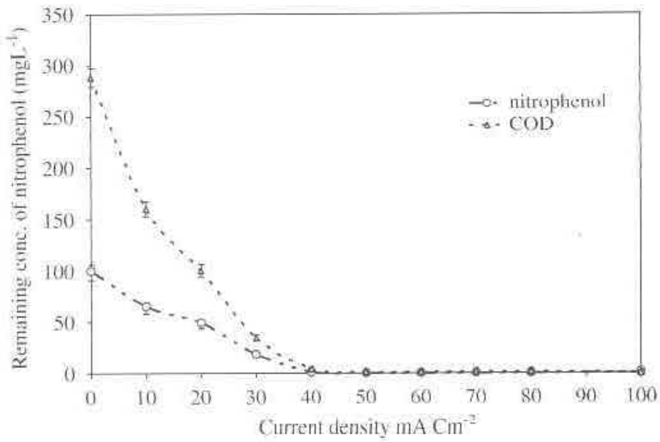


Fig. 1. Effect of current density on the removal of o-nitrophenol and COD at initial o-nitrophenol load concentration $100\,mg\,L^{-1}$, NaCl $4\,g\,L^{-1}$, pH 2.47, 60 min time of electrolysis and temperature of 30 °C.

3.2. Effect of current density

The removal processes of o-nitrophenol were carried out for 60 min. As shown in Fig. 1 o-nitrophenol removal and COD reduction were greatly increased with increases in current density. The efficiency data became asymptotic at current densities above $40\,mA\,cm^{-2}$. For this reason, the current density of $40\,mA\,cm^{-2}$ was considered optimum.

3.3. Effect of pH

The pH of the solution was varied while the other conditions were kept constant. As shown in Fig. 2, maximum removal of o-nitrophenol and COD was achieved in the pH range of ~2–4. Therefore solutions at pH 2.47 were used in subsequent experiments.

3.4. Effect of electrolysis time

To investigate the effect of electrolysis time, experiments were conducted with operating treatment conditions that were consistent with those described previously but with

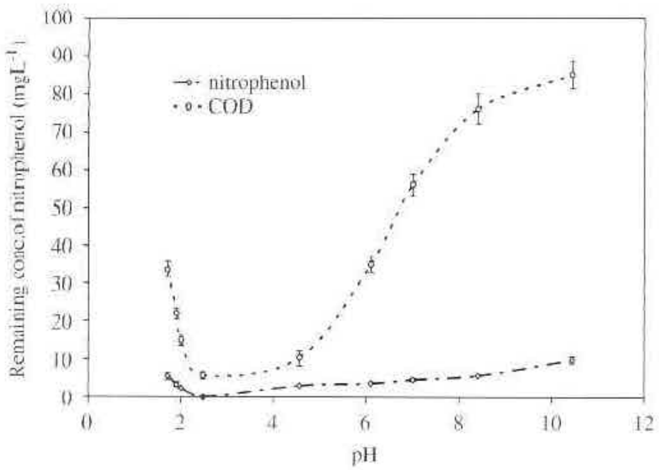


Fig. 2. Effect of pH value on the removal of *o*-nitrophenol and COD at initial *o*-nitrophenol load concentration $100\,mg\,L^{-1}$, 10 min time of electrolysis, NaCl $4\,g\,L^{-1}$, current density $40\,mA\,cm^{-2}$ and temperature of 30 °C.

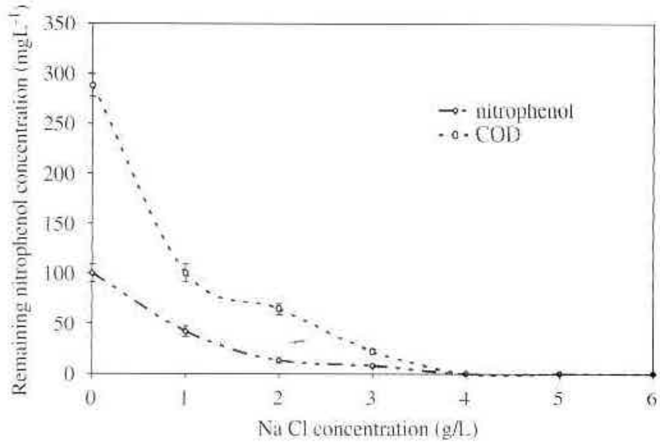


Fig. 4. Effect of NaCl concentration on the removal of *o*-nitrophenol and COD at initial load concentration $100\,mg\,L^{-1}$, current density $40\,mA\,cm^{-2}$, pH 2.47, 60 min time of electrolysis and temperature of 30 °C.

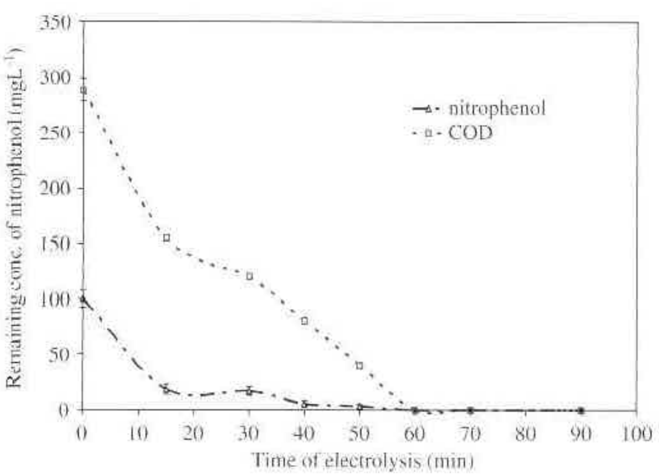


Fig. 3. Effect of time of electrolysis on the removal of *o*-nitrophenol and COD at initial *o*-nitrophenol load concentration $100\,mg\,L^{-1}$, NaCl $4\,g\,L^{-1}$, pH 2.47, current density $60\,mA\,cm^{-2}$ and temperature of 30 °C.

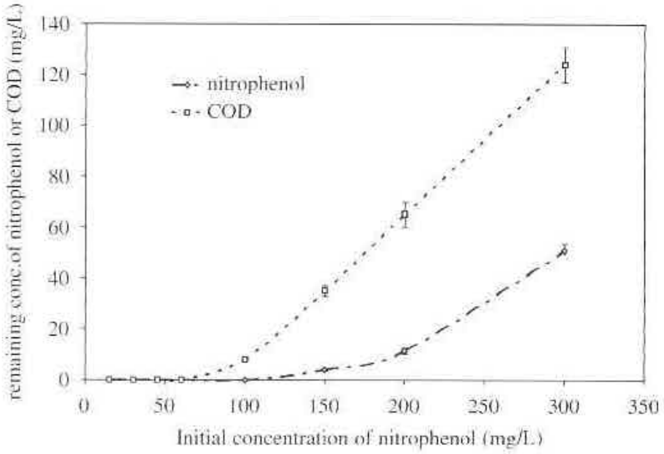


Fig. 5. Effect of initial *o*-nitrophenol concentration on the removal of *o*-nitrophenol and COD at current density $40\,mA\,cm^{-2}$, 60 min time of electrolysis, NaCl $4\,g\,L^{-1}$, pH 2.47 and temperature of 30 °C.

different electrolysis times ranging from 0 to 90 min. As shown in Fig. 3, the maximum removal from water for both *o*-nitrophenol and COD were achieved by reactions that proceeded for at least 60 min. Therefore, the electrolysis time of 60 min was taken as optimum for the removal of *o*-nitrophenol.

3.5. *Effect of NaCl concentration*

The effect of NaCl concentration on the removal of *o*-nitrophenol was investigated. As shown in Fig. 4, experiments employing 60 min reaction periods under the above conditions revealed that *o*-nitrophenol removal efficiency increased with increasing NaCl concentration up to $4\,g\,L^{-1}$. At NaCl concentrations above $4\,g\,L^{-1}$, no additional removal was achieved.

3.6. *Effect of initial o-nitrophenol concentration*

A current density of $40\,mA\,cm^{-2}$, pH of 2.47, temperature of 30 °C, NaCl $4\,g\,L^{-1}$, time of electrolysis 60 min and a variable *o*-nitrophenol concentration in the range of $10–300\,mg\,L^{-1}$ were used. The results from Fig. 5 indicate that the change of initial concentration had no effect on the *o*-nitrophenol removal below $100\,mg\,L^{-1}$. However, with an initial *o*-nitrophenol concentration above $60\,mg\,L^{-1}$, the COD removal efficiency was reduced. When the *o*-nitrophenol concentration exceeded $100\,mg\,L^{-1}$, removal of *o*-nitrophenol was incomplete.

3.7. *Effect of temperature*

The effect of temperature in the removal of *o*-nitrophenol from water was investigated. It was found that there is no effect of temperature in the range of 30–70 °C on the removal of *o*-nitrophenol or the removal of COD. For this reason, 30 °C was established as the optimal operating temperature.

3.8. Optimal conditions

From the results described above, it can be concluded that the optimum operating conditions for the treatment of *o*-nitrophenol with a PbO_2/Ti modified electrode are as follows: current density of $40\,mA\,cm^{-2}$, pH of 2.47, temperature of 30 °C, reaction time of at least 60 min and NaCl concentration of $4\,gL^{-1}$. Considering the effect of initial concentration on the removal of *o*-nitrophenol and COD, it can be concluded that the PbO_2/Ti modified electrode provided a high level of efficiency in the treatment process. It is noteworthy that using this electrode as the anode in the electrochemical oxidation process enabled complete removal of concentrated *o*-nitrophenol and COD to be attained.

It can be assumed that in the present study *o*-nitrophenol removal from the aqueous solutions proceeded by indirect rather than direct electrochemical oxidation. The electrocatalytic oxidation process was carried out at relatively low current densities of $40\,mA\,cm^{-2}$. At this low current density, enough Cl_2 and ClO^- are generated in the solutions to drive the oxidation process. These species are powerful oxidizing agents, capable of oxidizing and removing *o*-nitrophenol from an aqueous solution (Vase et al., 1998; Abu Ghalwa and Zaggout, 2006). It is worth mentioning that the odors of chlorine and hypochlorite were perceptible during the experiments that used NaCl as the conductive electrolyte. However, when either NaOH or HCl were used as the conductive electrolyte, the *o*-nitrophenol and COD removal efficiencies were relatively low, indicating that the amount of oxygen, and/or intermediate species, formed at these low current densities were not enough to achieve complete removal of *o*-nitrophenol and COD. The presence of chlorine and hypochlorite was smelt during electrolysis reactions. This indicates that electrogenerated Cl_2 and ClO^- may play a main role in the electrocatalytic oxidation process of *o*-nitrophenol and its removal from solutions. The mechanism of electrogeneration from a solution containing chloride ions involves two steps, the first of which is the primary oxidation of chloride ions to chlorine at the anode surface according to (Cristina Bocca and Cerisola, 2000):

$$2CL^- \Rightarrow CL_2 + 2e^-. \tag{2}$$

This is followed by the secondary solution phase reaction resulting in the formation of hypochlorous acid:

$$CL_{2(aq)} + H_2O \Rightarrow HCLO + CL^- + H^+, \tag{3}$$

which undergoes dissociation into hypochlorite and hydrogen ions:

$$HCLO \Rightarrow CLO^- + H^+. \tag{4}$$

The results of the present investigation have shown that the degradation of *o*-nitrophenol on the lead dioxide anode occurred very efficiently in the presence of NaCl as a conductive electrolyte. The degradation rate of this compound in NaOH was not as good as that in the NaCl containing solution. However, the lead dioxide electrode performed deficiently in the electrochemical degradation of the investigated organic pollutants, in the presence of HCl.

In the NaCl solution, the oxidation of the *o*-nitrophenol could occur directly via reaction of this organic compound with the electrogenerated hydroxyl radicals adsorbed on the lead dioxide surface. In addition indirect electrochemical oxidation may possibly have taken place involving the participation of the electrogenrated hypochlorite ions produced from the naturally occurring chloride ions in the processed water according to Eqs. (2)–(4). This, in turn, contributed to the higher degradation rate of the organic pollutants.

The reduced degradation efficiency in the presence NaOH could be attributed primarily to the absence of chloride which ensures that the electrocatalytic degradation of *o*-nitrophenol in the conductive electrolyte occurs primarily by direct oxidation on the lead. It was reported in the literature that in the absence of chloride the oxygen evolution was the major anodic reaction, but the oxygen formation in solution did not produce significant oxidation (Do and Yeh, 1996). On the other hand, the degradation of the *o*-nitrophenol observed in HCl solution could be correlated partially to the absence of chloride that, once again, guarantees a direct oxidation process. The poor performance of the lead dioxide electrode in the presence of HCl might also be attributed to either growth of an adherent film on the anode surface that poisoned the electrode or production of stable intermediates that could not be further oxidized by direct electrolysis in this conductive electrolyte.

As indicated in the above discussion, the presence in the conductive electrolyte of ions (such as chloride) capable of generating species of powerful oxidizing power and the conditions at the electrode surface play a paramount rule in the electrocatalytic degradation of organic pollutants on modified electrodes.

4. Conclusion

The experimental PbO_2/Ti electrode removed *o*-nitrophenol from aqueous solutions efficiently and with a short electrolysis time under conditions of a low current density where the electric cost was $0.32\,kW\,h\,L^{-1}$, a mild temperature and with NaCl as the conductive electrolyte. The optimum operating conditions of the treatment process were a current density of $40\,mA\,cm^{-2}$, a temperature of 30 °C, an initial *o*-nitrophenol load concentration of $100\,mg\,L^{-1}$, and a 60 min duration of electrolysis. The percentage of COD removal using the PbO_2/Ti electrode was 95% which is more than that in reported studies using Bi-PbO_2/Pt/Ti or Pt/Ti electrodes which was 70–80% COD removal (Iniesta et al., 2002). The electrode was stable throughout the experiments and no decrease in its efficiency was evident.

Acknowledgments

The authors would like to express their gratitude and appreciation to Acadmey Educational Development (AED) and The Palestinian Ministry of Higher Education for their financial support of this project.

References

Abu Ghalwa, N.A., Zaggout, F.R., 2006. Electrodegradation of methylene blue dye in water and wastewater using lead oxide/titanium modified electrode. Journal Environmental Science and Health (Part A) 41, 2283–2297.

Awad, H.S., Abu Ghalwa, N.A., 2005. Electrochemical degradation of acid blue and basic brown dyes on Pb/PbO_2 electrode in the presence of different conductive electrolyte and effect of various operating factors. Chemosphere 61, 1327–1335.

Baizer, M., Lund, H., 1985. Organic Electrochemistry. Dekker, New York.

Borras, C., Laredo, T., Scharifker, B.R., 2003. Competitive electrochemical oxidation of *p*-chlorophenol and *p*-nitrophenol on Bi doped PbO_2. Electrochemica Acta 48, 2775–2780.

Callwell, D.L., 1981. In: Bockris, J.O'M., Conway, B.E., Yeager, E., White, R.E. (Eds.), Comprehensive Treatise of Electrochemistry, vol. 2. Plenum Press, New York, p. 105.

Cristina Bocca, M.P., Cerisola, G., 2000. Electrochemical treatment of wastewater containing polyaromatic organic pollutants. Water Research 34, 2601–2605.

Do, J.S., Yeh, W.C., 1996. Partial electrooxidative degradation of phenol with in situ electro generated hydrogen peroxide and hypochlorite. Journal of Applied Electrochemistry 26, 673–678.

Fry, A.J., Britton, W.E., 1986. Topics in Organic Electrochemistry. Plenum Press, New York.

Greenberg, A.E., Clesceri, L.S., Eaton, A.D., 1992. Standard Methods for the Examination of Water and Wastewater, 18th ed. American Public Health Association, Washington, pp. 5.7–5.10.

Gutman, F., Murphy, O.J., 1981. In: White, R.E., Bockris, J.O'M., Conway, B.E. (Eds.), Modern Aspects of Electrochemistry, vol. 15. Plenum Press, New York, p. 1.

Hamza, A., Hamoda, M.F., 1980. Multiprocess treatment of textile wastewater. In: Proceedings of the 35th Purdue Industrial Waste Conference, Lafayett, Indiana.

Ibl, N., Vogt, H., 1981. Comprehensive Treatise of Electrochemistry, vol. 2. Plenum Press, New York, p. 167.

Iniesta, J., Exposito, E., Gonzalez-Garcia, J., Montiel, V., Aldaz, A., 2002. Electrochemical treatment of industrial wastewater containing phenols. Journal of the Electrochemical Society 149 (5), D57–D62.

Keith, L.H., Telliard, W.A., 1979. Priority pollutants I—a perspective view. ES&T special report. Environmental Science and Technology 13 (4), 415.

Kyriacous, D.E., Jannakoudis, D.A., 1986. Electrocatalysis for Organic Synthesis. Wiley, New York.

Newman, J., Tiedemann, W., 1978. In: Gersicher, H., Tobias, C.W. (Eds.), Advances in Electrochemistry and Electrochemical Engineering, vol. 11. Wiley, New York.

Novak, D.M., Tilak, B.V., Conway, B.E., 1982. In: Conway, B.E., Bockris, J.O'M. (Eds.), Modern Aspects of Electrochemistry, vol. 18. Plenum Press, New York, p. 195.

Polacro, M., Palmas, S., Renoldi, F., Mascia, M., 1999. On the performance of Ti/SnO_2 and Ti/PbO_2 anodes in electrochemical degradation of 2-chlorophenol for wastewater treatment. Journal of Applied Electrochemistry 29, 147–151.

Schmittinger, P., 1986. Chlorine. In: Ullmann's Encyclopaedia, vol. 6A. VCH, Weinheim p. 399.

Sheng, H.L., Peng, C., 1996. Continuous treatment of textile wastewater by combined coagulation, electrochemical oxidation and activated sludge. Water Research 30, 587.

Tilak, B.V., Lu, P.W., Colman, J.E., Srinivasan, S., 1981. In: Bockris, J.O'M., Conway, B.E., Yeager, E., White, R.E. (Eds.), Comprehensive Treatise of Electrochemistry, vol. 2. Plenum Press, New York, p. 1.

Vase, R., Sawant, S.B., Pangarkar, V.G., 1998. Electrochemical oxidation of *p-t*-butyl toluene to *p-t*-butyl benzaldehyde. Journal of Applied Electrochemistry 28, 623–626.

Journal of Environmental Management 86 (2008) 297–307

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia

Mahbub Hussain[a,*], Syed Munaf Ahmed[b], Walid Abderrahman[c]

[a]*Earth Sciences department, King Fahd University of Petroleum & Minerals, Dhahran, P.O. Box 1133, KFUPM, Dhahran-31261, Saudi Arabia*
[b]*Water Resources Division, Abunayyan Group, Riyadh, Saudi Arabia*
[c]*Center for Environment and Water, Research Institute, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia*

Received 23 September 2005; received in revised form 2 December 2006; accepted 7 December 2006
Available online 22 June 2007

Abstract

A multivariate statistical technique, cluster analysis, was used to assess the logged surface water quality at an irrigation project at Al-Fadhley, Eastern Province, Saudi Arabia. The principal idea behind using the technique was to utilize all available hydrochemical variables in the quality assessment including trace elements and other ions which are not considered in conventional techniques for water quality assessments like Stiff and Piper diagrams. Furthermore, the area belongs to an irrigation project where water contamination associated with the use of fertilizers, insecticides and pesticides is expected. This quality assessment study was carried out on a total of 34 surface/logged water samples. To gain a greater insight in terms of the seasonal variation of water quality, 17 samples were collected from both summer and winter seasons. The collected samples were analyzed for a total of 23 water quality parameters including pH, TDS, conductivity, alkalinity, sulfate, chloride, bicarbonate, nitrate, phosphate, bromide, fluoride, calcium, magnesium, sodium, potassium, arsenic, boron, copper, cobalt, iron, lithium, manganese, molybdenum, nickel, selenium, mercury and zinc. Cluster analysis in both Q and R modes was used. Q-mode analysis resulted in three distinct water types for both the summer and winter seasons. Q-mode analysis also showed the spatial as well as temporal variation in water quality. R-mode cluster analysis led to the conclusion that there are two major sources of contamination for the surface/shallow groundwater in the area: fertilizers, micronutrients, pesticides, and insecticides used in agricultural activities, and non-point natural sources.

Keywords: Cluster analysis; Logged irrigation water; Agriculture; Contamination; Water types

1. Introduction

Since surface water is scarce or almost non-existent in arid regions like Saudi Arabia, efforts are being continuously made to use ground water for irrigation and greening of the landscape in such regions. The tradition of agricultural activities in northeast Saudi Arabia is ancient and involves a number of oases including Al Hasa and Hofuf. Increased agricultural activities in the area have resulted in higher demand for groundwater in the region, and as a consequence of these agricultural activities, the groundwater table has declined significantly in recent years (Ministry of Agriculture, 1984). Increased use of irrigation water has resulted in higher salinity of the soil. In addition, the presence of an aquitard in the shallow subsurface in the area has introduced a new problem—water logging. The study area is a large farm (~100 km^2) located ~150 km northwest of the city of Dammam, Saudi Arabia.

Irrigated agriculture in the study area started in 1985 when several deep wells were drilled in the Paleocene age Umm-Er-Radhuma aquifer for irrigation purposes. At present, there are over 75 wells located in the area. Several crops including wheat, barley, alfalfa and rhodus are routinely cultivated on this farm. Soil type at the irrigation project is representative of the soil in Saudi Arabia with a sand size fraction often exceeding 90% of the total soil constituents. Based on the soil taxonomy scheme proposed by the Natural Resource Conservation Service of the

*Corresponding author. Tel.: +966 3 8604916; fax: +966 3 8602595.
E-mail addresses: mahbub@kfupm.edu.sa, mahbubferoze52@yahoo.com (M. Hussain).

0301-4797/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.jenvman.2006.12.007

United States Department of Agriculture (formerly the Soil Conservation Service), in 1999, the soils of eastern Saudi Arabia including the study area can be classified as aridisols soil. Since such soils have very little water-containing capacity and are very low in moisture, irrigation is required to support cultivated crops. Because of excessive irrigation and the presence of a well-documented aquitard in the Neogene undifferentiated Hofuf/Dam Formation, the low-lying areas in the farm have been flooded forming several ponds. Also, because of high evaporation and evapotranspiration (ET) rates the concentrations of hydrochemical variables were found to be higher at the time of sampling. Preliminary analysis carried out by the hydrologists of the Research Institute (RI) of King Fahd University of Petroleum & Minerals (e.g., Al-Assar, 1992) indicated elevated levels of several contaminants related to the agricultural activities on the farm. These elevated levels, to some extent, can also be attributed to high evaporation and ET rates in the area.

In recent decades, there has been a rapid increase in non-point sources of groundwater pollution i.e. pollution, entering the system over a wide area. Martinez and Albiac (2004) noted that non-point source pollution from agriculture has increased significantly in several European Union countries including Spain over recent decades due to large-scale investments in product-enhancing technologies that involve the intensive use of machinery, industrial fertilizers and pesticides. Agriculture is now considered to be the dominant source of non-point pollution of groundwater (Jones, 1997). The commonly found contaminants in groundwater due to agriculture are nitrate, chloride, sodium, calcium, magnesium, ammonia, phosphate and trace elements (George et al., 1987; Burkhart and Kolpin, 1992; Fetter, 1992; Spalding and Exner, 1993; Beke et al., 1993).

Many different sources and processes can be responsible for the contaminants polluting the surface and groundwater. Detailed hydrochemical research is needed to evaluate the different processes and mechanisms involved in polluting water (Helena et al., 1999). Conventional classification techniques like Stiff and Piper diagrams only consider selected major and minor ions in determining the chemical quality of the groundwater in an aquifer. In view of the limitations of these methods and the availability of an increasing number of chemical parameters in recent years, a wide range of statistical techniques are now required for proper analysis of data (Ashley and Lloyd, 1978; Dalton and Upschurch, 1978; Usunoff and Guzman, 1989; Olmez et al., 1994; Voudouris et al. 2000; Guler et al., 2002; Reghunath et al, 2002; Hussein, 2004; Mahlknecht et al., 2004). It has been shown that multivariate statistical techniques can be very useful tools for interpretation of such water quality data.

This paper documents the geochemical analysis of collected water samples from several irrigation logged water ponds in eastern Saudi Arabia and the utilization of cluster analysis in classifying waters. In addition, attempts were made to identify the main contamination sources (natural and anthropogenic).

2. The study area

The study area is located at Al-Fadhley in the Eastern Province of Saudi Arabia. The study area is bounded on the north by latitude 26°50′31″N, on the south by latitude 26°34′34″N, on the east by longitude 49°16′15″E, and on the west by longitude 48°59′58″E (Fig. 1). The study area is around 10 km from west to east and 9 km from north to south. The area is covered by Quaternary-age gravel, sand and silt and overlies the Neogene Hofuf, Dam and Hadrukh formations. The soil type in the study area is typical of Saudi soil and consists of over 90% sand. According to Al-Assar (1992), the shallow aquifer in the irrigation project area is very sandy and consists of over 90% sand. Several crops including wheat, barley, alfalfa and rhodus are cultivated in these fields. The study area, as part of the northeastern region of Saudi Arabia, is in the desert belt but experiences relatively more rainfall (av. 70–100 mm/yr) than the areas further to the west. This rainfall occurs primarily in the winter season as a result of irregular incursions of humid and cold polar air modified along the coast of the Arabian Gulf.

3. Sampling and analysis

The study is based on a total of 34 water samples collected from the two largest lakes at the farm. The sampling was conducted at the end of the dry season (September–October 1998), when maximum concentrations were expected, and at the end of the rainy season (February–March 1999), when maximum dilution was anticipated. Seventeen samples were collected for both summer and winter. Fig. 2 shows the locations of the sampling points. The samples were analyzed following methods outlined in the American Public Health Association manual (APHA, 1992). Water samples were collected in stopper-fitted polyethylene bottles and refrigerated at 4 °C in order to be analyzed as soon as possible. Conductivity, temperature and pH were measured in situ using a portable water tester.

The water samples were analyzed for major and some of the minor ions and trace elements. Anions analyzed include sulfate, chloride, bicarbonate, nitrate, phosphate, bromide and fluoride; cations include calcium, magnesium, sodium and potassium; trace elements include arsenic, boron, copper, cobalt, iron, lithium, manganese, molybdenum, nickel, selenium, mercury, and zinc. Anions were quantified using Ion Chromatography. Bicarbonates were determined by titration. Mercury was quantified using cold vapor AAS, cations were determined by ICP at the Material Characterization Laboratory (MCL) at the Research Institute (RI), King Fahd University of Petroleum & Minerals. The details of the analytical data are given in Table 1. TDS, major ions, minor ions and trace elements

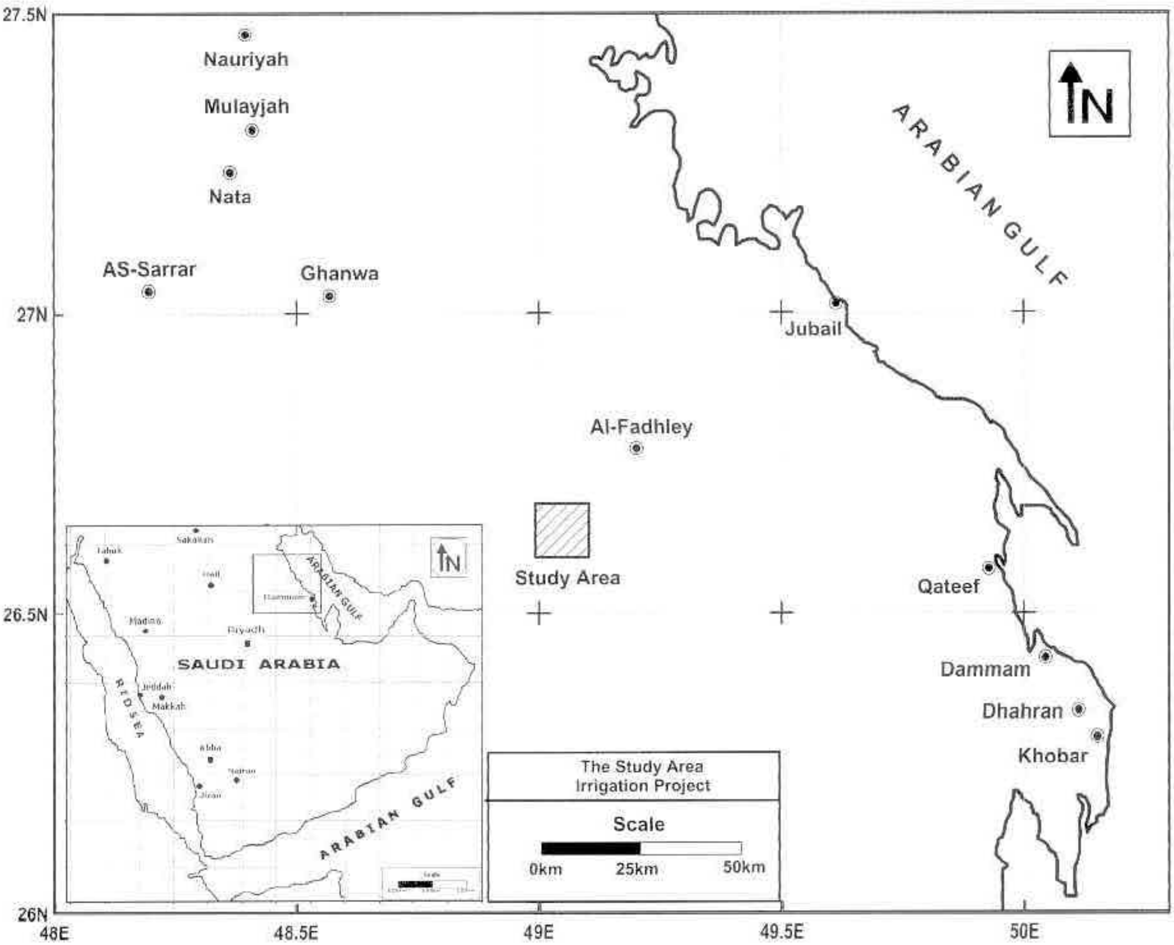


Fig. 1. Location map of the Study Area.

are in mg l^{-1} whereas conductivity is given in deci-Siemens per meter (dS/m), and pH in standard units.

4. Conventional classification of water: piper diagram

Piper diagrams, also called trilinear diagrams (Piper, 1953), are drawn by plotting the proportions (in equivalents) of the major cations ((Ca^{2+}, Mg^{2+}, ($Na^{+}+K^{+}$)) on one triangular diagram, the proportions of the major anions (Alkalinity ($CO_3+HCO_3^-$), Cl^-, SO_4^{2-}) on another, and combining the information from the two triangles on a quadrilateral. The position of this plotting indicates the relative composition of a groundwater in terms of the cation–anion pairs that correspond to four vertices of the field. The main drawback that Piper and trilinear diagrams have is their plottings show the chemical character of a groundwater according to the relative concentration of its constituents, but not according to the absolute concentrations.

All the water samples collected in summer and winter were plotted on a Piper diagram (Fig. 3). The employing of the water classification scheme of Back and Hanshaw (1965) showed that the waters in the study area are sulfate–chloride dominated. Piper diagrams would be able to classify the water at the study area into only one type. The objective of the study was to assess the environmental impacts of agricultural practices in the area and also to determine water quality in detail and not just in terms of major and minor ions. Trace elements are the critical aspect of this study and are not included in conventional methods. Guler et al. (2004) in their study have compared all the multivariate statistical methods with almost all of the conventional/graphical water classification schemes available. Huge numbers of both surface and groundwater samples were collected in their study carried out in the south Lahontan region of southeastern California. They concluded that multivariate statistical methods are very effective tools for classifying a huge number of samples with many hydrochemical variables. The present study overcomes the limitations of using the conventional Piper diagrams. A multivariate statistical technique called cluster analysis was applied to the hydrochemical data. Most importantly, most of the trace elements and nitrates were included in the cluster analysis and they did play a critical role in classifying the water at the study area into three types as discussed in the following sections.

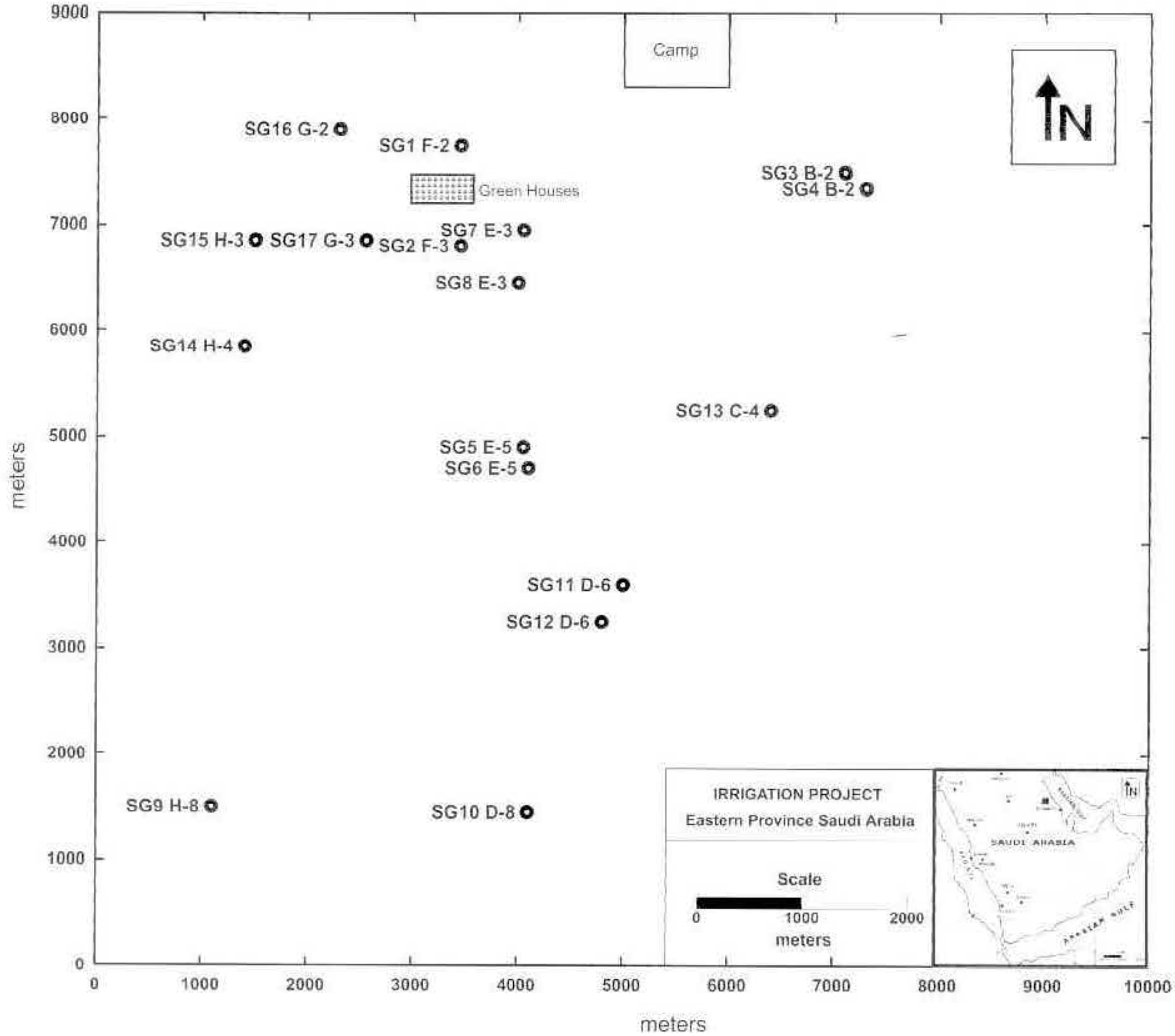


Fig. 2. Map showing the location of the sampling points at the study area.

5. Q-mode cluster analysis

Cluster analysis is a technique for grouping individuals or objects into unknown groups. There are two types of cluster analysis: R- and Q-modes (Brown, 1998). This technique can be used to group the commonly collected water quality data, where each cluster indicates the water of a particular quality. In the present study, Q-mode cluster analysis was performed on the water chemistry data to group the samples in terms of water quality (Chandrasekhram et al., 2001; Grande et al., 2003). Hydrochemical results of all samples were statistically analyzed by using the software STATISTICA®. The weighted pair-group method was applied and Euclidean distance was chosen as a measure of similarity. The data were classified in a simple and direct manner with results presented as dendograms (Figs. 4 and 5). Based on an imaginary horizontal line (phenon line) on the cluster scale three distinct groups or clusters in summer as well as in winter samples were recognized. Table 2 also shows the cluster number and its members for both of the seasons.

Clusters of samples listed in Table 2 indicate that each cluster has a water quality of its own which is different from the other clusters. If the example of total dissolved solids (TDS) is taken for the summer, cluster 1 includes: SG9-H8, SG10-D8, SG13-C4, SG16-G2, where the TDS concentration ranged from 3000 to 6000 mg/l which is the characteristic of brackish water. Cluster 2 includes the samples: SG1-F2, SG2-F3, SG3-B2, SG4-B2, SG5-E5, SG6-E5, SG7-E3, SG8-E3, SG11-D6, SG12-D6, SG15-H3, where the TDS mostly ranged from 20 000 to 30 000 mg l^{-1} which is the characteristics of moderate saline water and cluster 3 includes samples: SG14-H4, SG17-G3, where the TDS ranged from 35 000 to 50 000 mg l^{-1} which is the characteristic of highly saline water. Similarly, by considering the other water quality variables (major ions, minor ions and trace elements) the water quality or chemistry associated with each cluster can be assessed in detail.

After dividing the samples into three different clusters and by considering the samples and their locations for each cluster, zones were developed (I, II, and III). Each of these

Table 1
Analytical data of the samples collected in summer (a) and winter (b)

Sample ID[a]	pH	Cond	TDS	Alk	HCO_3	Br	Cl	F	NO_3	PO_4	SO_4	Ca	K	Mg	Na	As	B	Cu	Fe	Li	Mn	Hg	Mo	Ni	Se
SG1 F-2	8.06	36.2	31500	143	174	11.7	12100	3.9	42.7	<0.2	5460	1020	260	807	7150	0.19	9.95	<0.003	<0.02	0.47	0.018	<0.2	0.18	<0.010	<0.08
SG2 F-3	7.77	34.8	30000	169	207	10.1	11200	4.3	4.6	3.7	5390	1110	212	805	6580	<0.03	9.23	0.066	<0.02	0.45	0.012	<0.2	0.09	<0.010	<0.08
SG3 B-2	7.98	6.02	4850	187	228	2.7	1430	2.6	8.5	4.6	996	318	42	148	916	<0.03	1.3	<0.003	<0.02	0.09	0.04	<0.2	<0.01	<0.010	<0.08
SG4 B-2	8	6.01	4610	187	228	3.4	1590	2.3	8.4	<0.2	1080	313	35	146	894	<0.03	1.36	0.02	0.31	0.09	0.008	<0.2	0.03	<0.010	<0.08
SG5 E-5	7.8	34.9	30000	216	264	11.6	11000	4.2	5	<0.2	5400	1450	315	1000	7460	0.19	9.54	<0.003	0.16	0.45	0.012	<0.2	0.09	<0.010	<0.08
SG6 E-5	7.79	34.6	30000	212	259	7.6	11300	4.3	5.6	5.8	5460	1200	229	838	6400	<0.03	9.58	0.062	<0.02	0.45	0.012	<0.2	0.09	<0.010	<0.08
SG7 E-3	7.81	34.5	30000	209	255	9.6	11000	4.2	5.6	<0.2	5280	1190	221	835	6460	<0.03	9.38	<0.003	<0.02	0.44	0.012	<0.2	0.09	0.07	<0.08
SG8 E-3	7.8	34.6	29800	205	251	9.5	11100	4.2	5.5	<0.2	5280	1140	204	806	6290	<0.03	9.47	0.066	0.13	0.45	0.012	<0.2	0.09	0.08	<0.08
SG9 H-8	7.68	6.84	6050	106	129	3.7	1520	0.45	7.8	<0.2	1930	517	69.2	191	815	<0.03	1.64	<0.003	0.2	0.11	0.2	<0.2	0.03	<0.010	<0.08
SG10 D-8	8.26	5.18	4570	138	168	3.15	1090	0.45	102	107	1300	406	46.4	146	546	<0.03	1.1	<0.003	<0.02	0.08	0.02	<0.2	<0.01	<0.010	<0.08
SG11 D-6	7.73	29.1	26800	304	371	18.7	9260	2	4.8	<0.2	5420	1100	263	1230	4340	0.24	8.27	<0.003	0.1	0.5	0.058	<0.2	<0.01	<0.010	<0.08
SG12 D-6	7.77	29	25600	307	375	20.5	9240	2.1	<2	3.7	5480	1110	276	1250	4390	0.25	8.35	0.054	<0.02	0.51	0.056	<0.2	<0.01	<0.010	<0.08
SG13 E-4	8.82	7.4	5950	91	110	3.75	1610	0.65	5.1	1.35	1940	394	70.2	288	850	<0.03	1.74	<0.003	<0.02	0.13	<0.001	<0.2	<0.01	<0.010	<0.08
SG14 H-4	8.27	43.7	35700	112	137	11.7	15300	3.5	383	1.65	5760	1070	237.2	1030	8820	0.23	11.12	0.058	0.09	0.52	<0.001	<0.2	<0.01	<0.010	<0.08
SG15 H-3	7.3	26.7	21600	255	311	7.9	8500	2.7	388	4.6	4000	1010	151	621	4710	0.18	6.47	0.04	<0.02	0.28	<0.001	<0.2	0.15	0.04	<0.08
SG16 G-2	8.57	4.53	3440	90	110	2.15	883	0.2	69.7	<0.2	1170	250	62.9	118	573	<0.03	1.23	0.014	<0.02	0.09	0.006	<0.2	0.03	0.02	<0.08
SG17 G-3	8.3	56.9	48200	79	96	16.5	21400	5.4	568	<0.2	7690	1240	348	1340	12800	0.024	18.9	<0.003	<0.02	0.73	0028	<0.2	0.4	0.08	<0.08

Sample ID[b]	pH	Cond.	TDS	Alk	HCO_3	Br	Cl	F	NO_3	PO_4	SO_4	Ca	K	Mg	Na	As	B	Co	Cu	Fe	Li	Mn	Mo	Ni	Se	Zn
SG1 F-2	8.55	31.1	26400	83.6	102	24	10800	9	65	<0.2	5320	1054	242.17	860.7	6941.13	<0.059	10	0.24	0.075	<0.01	0.49	0.02	0.19	<0.014	<0.1	0.017
SG2 F-3	8.2	28.8	24200	103	125	7.4	9170	14.8	464	<0.2	4360	1002	186.27	680.2	6633.22	<0.059	10	0.3	0.1	<0.01	0.41	0.02	0.239	<0.014	<0.1	0.017
SG3 B-2	7.7	4.17	3590	298	364	2.61	890	0.98	5.48	<0.2	1070	398	28.1	163.5	494.6	<0.059	1.48	0.23	0.077	<0.01	0.14	0.01	0.047	<0.014	<0.01	0.011
SG4 B-2	7.65	4.43	3880	241	294	1.93	929	2.15	<2	0.95	1190	397	33.87	167.1	510.5	<0.059	1.35	0.23	0.087	<0.01	0.15	0.03	0.051	<0.014	<0.01	0.024
SG5 E-5	8.55	33.1	28900	130	159	19.9	13100	9.7	76.8	<0.2	6170	1150	208.4	1102	6933.32	<0.059	11.52	0.42	0.13	<0.01	0.55	0.02	0.227	<0.014	<0.01	0.029
SG6 E-5	8.63	33.2	30800	117	143	21.1	12100	13.3	77.9	<0.2	5740	1143	208	1104	6954.2	<0.059	11.55	0.46	0.15	<0.01	0.58	0.02	0.242	<0.014	<0.01	0.033
SG7 E-3	7.88	28.3	24600	255	312	14.6	9820	3.1	43.8	<0.2	5120	1.98	192.63	819.3	6251	<0.059	0.99	<0.1	0.04	<0.01	0.43	0.01	0.126	<0.014	<0.01	<0.006
SG8 E-3	7.89	28.5	24200	245	299	20	10000	6.3	47.3	<0.2	5180	1101	189.4	816.3	6083	<0.059	9.03	0.15	0.05	<0.01	0.46	0.01	0.13	<0.014	<0.01	<0.006
SG9 H-8	7.54	4.24	3460	279	341	3.23	824	2.25	1.29	0.93	1150	383	43.63	154.1	608	<0.059	1.06	<0.1	0.022	<0.01	0.076	0.04	0.11	<0.014	<0.01	<0.006
SG10 D-8	8.16	5.33	4840	141	172	2.7	1150	1.02	126	<0.2	1490	466	60	200	656.23	<0.059	1.15	<0.1	0.026	<0.01	0.087	0	0.019	<0.014	<0.01	<0.006
SG11 D-6	7.62	23.2	21600	304	370	20.6	8230	14.5	7.2	<0.2	4840	1069	232.24	1177	4.56	<0.059	7.63	0.18	0.048	<0.01	0.49	0.01	0.04	<0.014	<0.01	<0.006
SG12 D-6	7.62	23	21400	306	373	23.5	7990	5.7	6.4	<0.2	4710	4058	232.42	1163	4.41	<0.059	7.6	0.21	0.045	<0.01	0.495	0.01	0.045	<0.014	<0.01	<0.006
SG13 E-4	7.92	15	13700	281	342	9.55	4380	3.1	5.55	<0.2	3960	981.5	164.72	704.6	2527	<0.059	5.34	<0.1	<0.006	<0.01	0.301	0.01	1.1	<0.014	<0.01	<0.006
SG14 H-4	8.38	38.7	39500	116	142	11.6	14800	9.3	359	<0.2	5660	1122	232.51	1138	9310.32	<0.059	13.06	0.32	0.097	<0.01	0.694	0.02	0.249	<0.014	<0.01	<0.006
SG15 H-3	8.02	37.3	34400	200	244	14.9	14000	6.8	530	<0.2	5080	1225	234.57	1029	8681.1	<0.059	10.86	0.3	0.079	<0.01	0.64	0.02	0.24	<0.014	<0.01	<0.006
SG16 G-2	8.25	24.8	20600	132	161	10	8290	15.4	148	5	4270	926	193.25	739.4	5141.27	<0.059	8.01	<0.1	<0.006	<0.01	0.33	<0.001	0.071	<0.014	<0.01	<0.006
SG17 G-3	8.42	43.5	37600	97.1	118	11.6	14900	9.6	620	<0.2	5990	1235	273.56	1168	10826	<0.059	16.33	0.3	0.11	<0.01	0.74	0.32	0.32	<0.014	<0.01	<0.006

Note: Values less than (<) (example: <0.02) indicates detection limit (DL) used for that particular analyte.

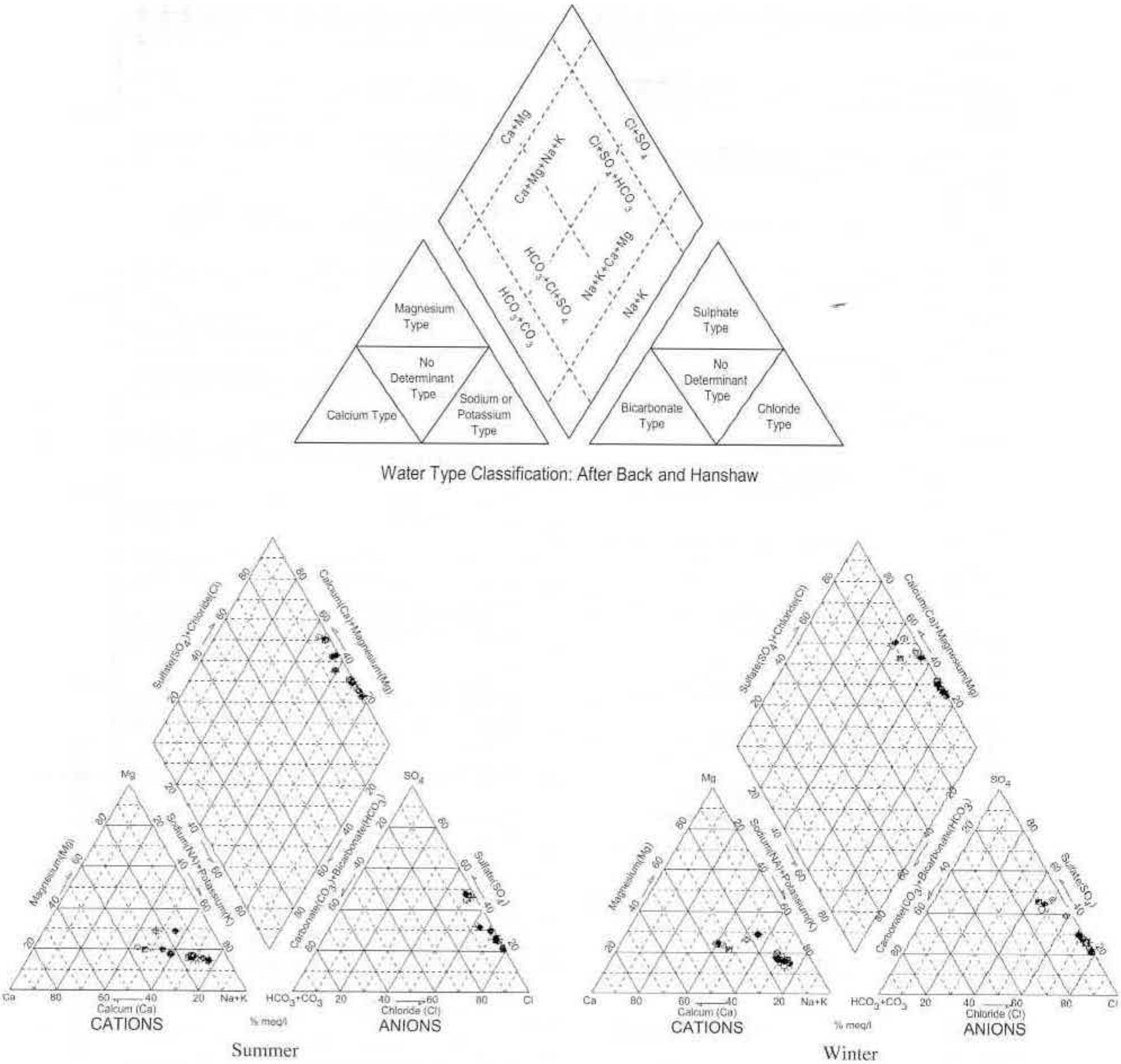


Fig. 3. Piper diagrams for summer, winter season and water type classification by Back and Hanshaw (1965).

zones corresponds to its respective cluster, and has a water quality which is different from the other zones. Figs. 6 and 7 show the distribution of clusters or zones or types of water quality at the study area.

5.1. Water types in summer

5.1.1. Water Type-I

This water is basically chloride and sulfate dominated, however; calcium and sodium are also present, but in a relatively low concentration. Bicarbonate is the lowest. As far as trace elements are concerned, this water has the lowest concentrations of most of the trace elements compared to other waters of the study area, boron being the highest while copper was the lowest. It also has low concentrations of lithium and manganese. This water type has the highest level of manganese. The Type-I water type is found in the north-western, central-east and southern parts of the study area (Fig. 6).

5.1.2. Water Type-II

Water Type-II is sodium and chloride dominated and it also has low concentrations of sulfate and calcium. Again, bicarbonate is at its lowest concentration. Among trace elements, boron concentration is highest and manganese is lowest. Type-II water is found in the central and north-eastern parts of the study area (Fig. 6).

5.1.3. Water Type-III

This water is sodium and chloride dominated but also contains low concentrations of sulfate, calcium and magnesium. Bicarbonate is the lowest. Based on the major

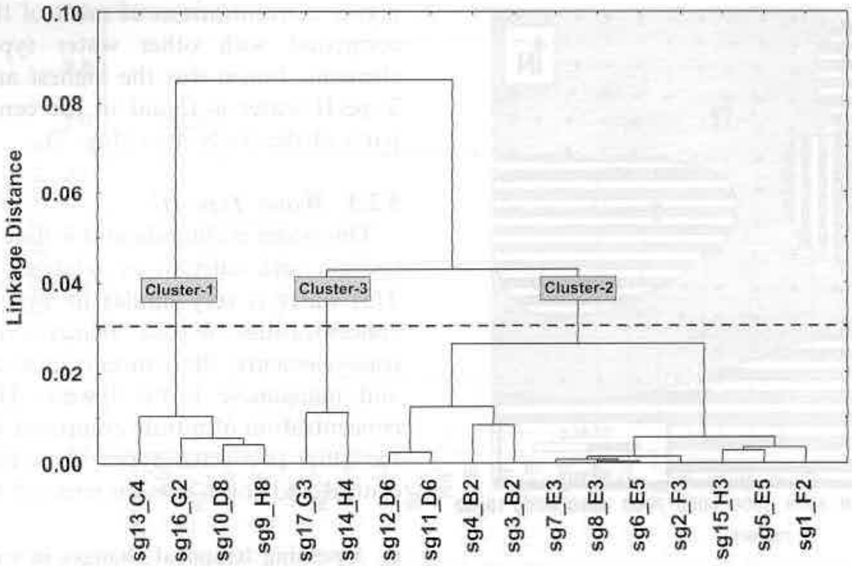


Fig. 4. Tree diagram from cluster analysis in Q-mode for water samples collected in summer shows three distinct groupings, comprising of samples given in Table 2.

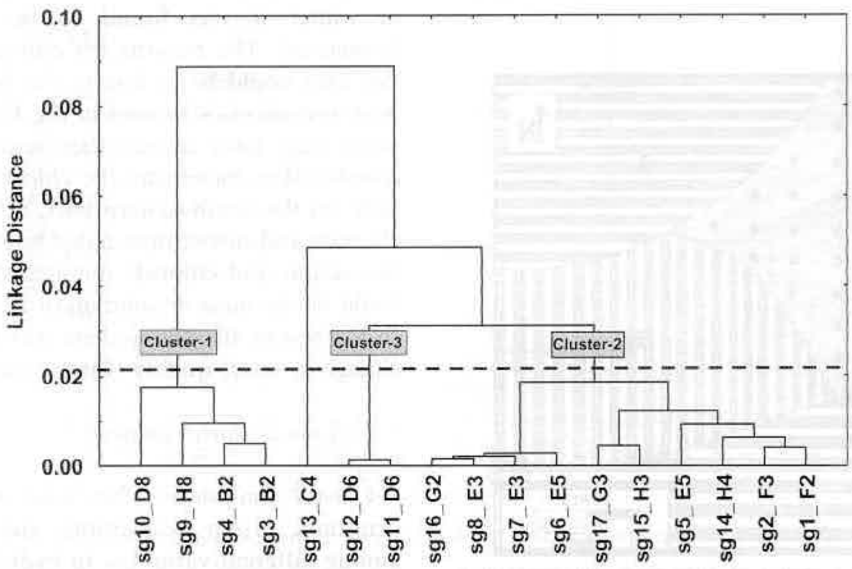


Fig. 5. Tree diagram from cluster analysis in Q-mode for water samples collected in winter shows three distinct groupings, comprising of samples given in Table 2.

Table 2
Clusters and their members in Q-mode for the samples collected in both seasons

Cluster number	Members
Summer	
1	SG9-H8, SG10-D8, SG13-C4, SG16-G2
2	SG1-F2, SG2-F3, SG3-B2, SG4-B2, SG5-E5, SG6-E5 SG7-E3, SG8-E3, SG11-D6, SG12-D6, SG15-H3
3	SG14-H4, SG17-G3
Winter	
1	SG3-B2, SG4-B2, SG9-H8, SG10-D8
2	SG1-F2, SG2-F3, SG5-E5, SG6-E5, SG7-E3, SG8-E3 SG14-H4, SG15-H3, SG16-G2, SG17-G3
3	SG11-D6, SG12-D6, SG13-C4

ion concentrations this water is similar to water Type-II. However, because of high concentrations of nitrate and some trace elements, this water should be considered a different type. In addition, this water has higher nitrate and arsenic concentrations when compared with other water types of the study area. Type-III water is found in the central and north-western parts of the study area (Fig. 6).

5.2. Water types in winter

5.2.1. Water Type-I

This water is chloride and sulfate dominated. It also has sodium and calcium but in relatively lower concentrations and the concentration of bicarbonate is lowest among the water types tested. Among the trace elements identified, the

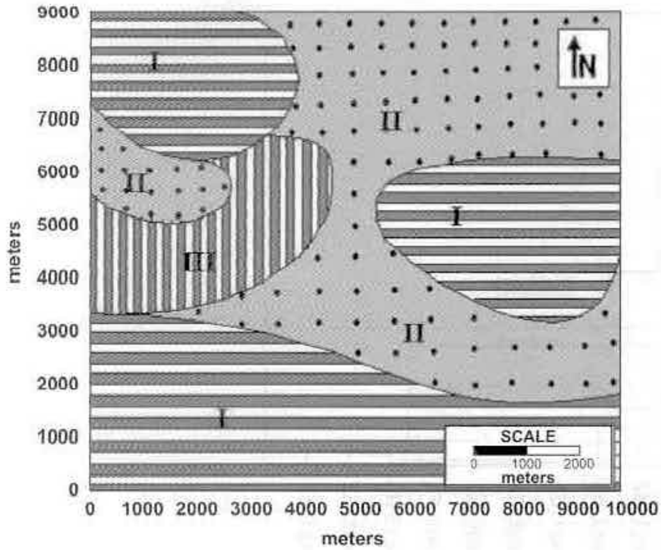


Fig. 6. Distribution of three clusters (I, II and III) from the Tree diagram of Fig. 4 showing spatial zonation of waters at the study area in summer. Each zone indicates water quality of its own.

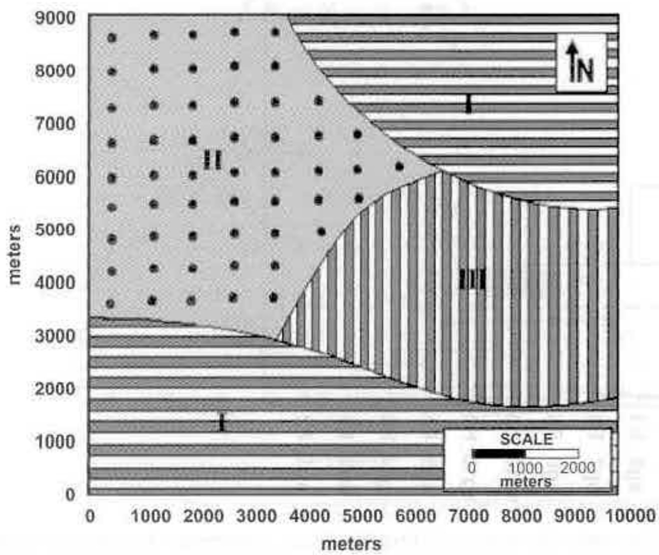


Fig. 7. Distribution of three clusters (I, II and III) from the Tree diagram of Fig. 5 showing spatial zonation of waters at the study area in winter. Each zone indicates water quality of its own.

concentration of boron was highest and manganese was the lowest. The distribution of the trace elements in water Type-I is lower than those of other water types of the area. Type-I water is found in the southern and north-eastern parts of the study area (Fig. 7).

5.2.2. *Water Type-II*

While water Type-II is sodium and chloride rich, it also contains appreciable amounts of sulfate and calcium. The concentration of bicarbonate was lower than those of other water types identified in the area. This water type has higher concentrations of most of the trace elements when compared with other water types. Among the trace elements, boron was the highest and zinc was the lowest. Type-II water is found in the central and north-western parts of the study area (Fig. 7).

5.2.3. *Water Type-III*

This water is chloride and sulfate dominated but also has sodium and calcium in relatively lower concentrations. This water is very similar to Type-I water except for the concentrations of trace elements and nitrate. Among the trace elements, the concentration of boron is the highest and manganese is the lowest. This water has a lower concentration of nitrate compared to the concentrations in the other two water types. Type-III water is found in the central and south-eastern parts of the study area (Fig. 7).

6. Assessing temporal changes in water quality

Water in the north-western part of the study area in summer was basically chloride and sulfate dominated but in winter it was found to be sodium and chloride dominated. The reasons for chloride replacing sulfate in this area could be to due to the huge discharge of water from greenhouses located in the Field F2. The discharged water may have caused the existing salts in the soil to dissolve thus increasing the chloride concentration. Similarly, in the north-eastern part, in summer the water was chloride and sulfate dominated but the in winter it changed to sodium and chloride dominated. The reasons for this could be the same dissolution process as mentioned earlier. In the rest of the places there was no significant temporal change in water quality (Figs. 6 and 7).

6.1. *R-mode cluster analysis*

Cluster analysis is often used in scientific research to determine group associations and to assess the affinity among different variables. In hydrological studies R-mode cluster analysis specifically has been used in determining the association of different water quality parameters and ultimately, the sources and processes with which they are associated (Hopke et al., 1976; Voudouris et al., 2000; Hussein, 2004; Mahlknecht et al., 2004).

In the present study, cluster analysis in R-mode was performed on pH, TDS, alkalinity, conductivity, calcium, magnesium, sodium, potassium, chloride, sulfate, bicarbonate, nitrate, bromide, fluoride, phosphate, boron, copper, iron, lithium, manganese, mercury, molybdenum, nickel, selenium, zinc, cobalt, and arsenic. The weighted pair-group method was applied and Euclidean distance was used as the measure of similarity. Before analyzing for clusters, the data were normalized by generating a correlation coefficient matrix. Cluster analysis revealed two distinct groups or clusters for summer as well as for winter data (Figs. 8 and 9). From these figures, different clusters and their members were extracted as follows.

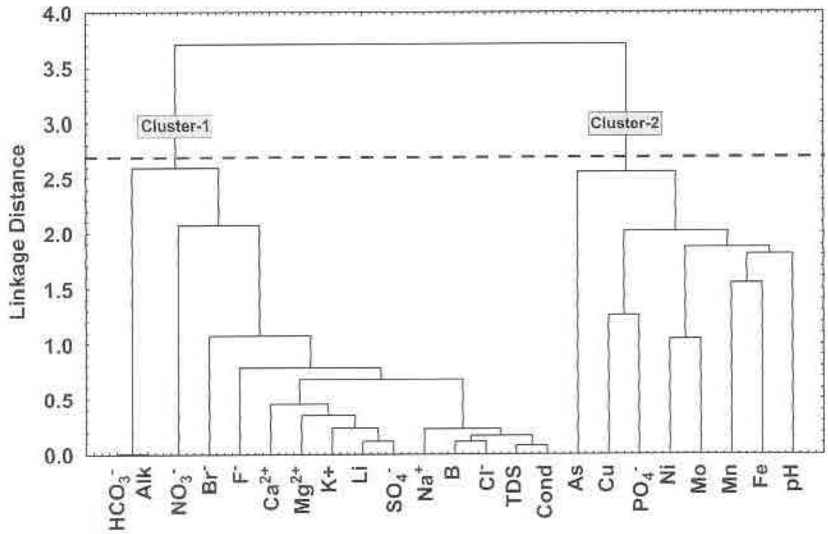


Fig. 8. Tree diagram for 23 variables from cluster analysis in R-mode shows two major groups consisting of variables given in Table 3 (summer).

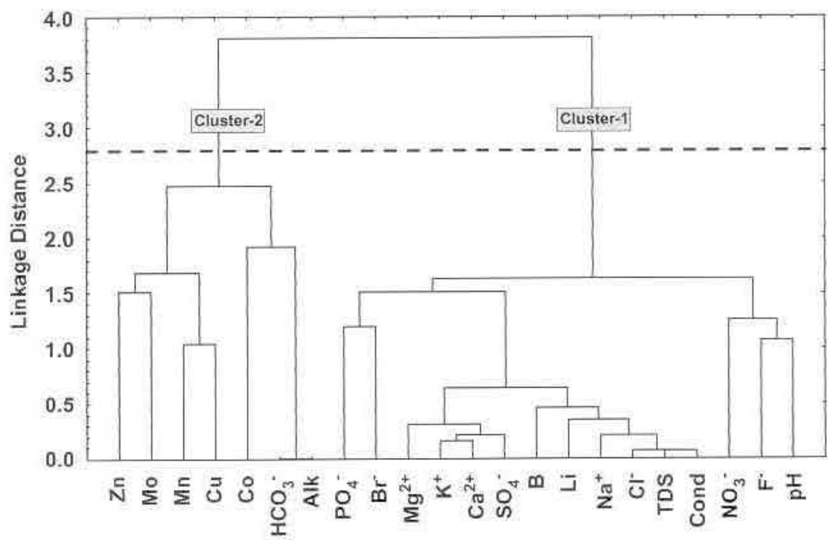


Fig. 9. Tree diagram for 23 variables from cluster analysis in R-mode shows two major groups consisting of variables given in Table 3 (winter).

For summer cluster 1 includes: bicarbonate, alkalinity, nitrate, bromide, fluoride, calcium, potassium, lithium, sulfate, sodium, boron, chloride, TDS and conductivity. Cluster 2 includes: arsenic, copper, phosphate, nickel, molybdenum, manganese and iron. For the winter data cluster 1 comprised the following variables: phosphate, bromide, magnesium, potassium, calcium, sulfate, boron, lithium, sodium, chloride, TDS, conductivity, nitrate, fluoride, and cluster 2 was found to contain: cobalt, copper, alkalinity, bicarbonate, manganese, molybdenum, and zinc.

As shown in tree diagrams prepared for the summer and winter data (Figs. 8 and 9) and also (Table 3) there are two distinct clusters: one comprising of all major and minor ions and the other consisting of trace elements. Therefore the tree diagrams suggest that there are basically two pollutant poles: major-minor ions pole and a trace elements pole. The term "pole" was used by Rapin (1980) in order to demonstrate the anthropogenic effects on sediment from the Bay of Nice and Villefranche-sur-Mer on the French Mediterranean coast. Chemical parameters included in his study were: Al^{3+}, Fe^{3+}, K^+, Si^{4+} and some other trace elements included were Ni^{2+}, V^{4+}, Cr^{3+}, Ti^{4+}, Hg^{2+}, Pb^{3+}, Cd^{2+} and Zn^{2+}. As a first step the correlation coefficient matrix was generated as it is necessary to perform the clustering operation. He found Al^{3+}, Fe^{3+}, K^+, Si^{4+} (clay minerals) clustering together in one group and Ni^{2+}, V^{4+}, Cr^{3+}, Ti^{4+}, Hg^{2+}, Pb^{3+}, Cd^{2+} and Zn^{2+} in another group. From this clustering he concluded that these two groups which he named poles (clay pole and pollutant pole) are the main anthropogenic contaminants in the sediments. The authors of the present study decided to retain the term "pole" and attempt to explain these two poles. The authors have concluded that the source of trace elements is most probably due to the application of trace elements as micronutrients in irrigation farming, as no geologic or lithologic source could be established in the area. The other pole (major-minor ions pole), which consists of all ions, however, simply cannot be attributed to a single source; there could be many natural sources and anthropogenic processes controlling this pole. To get a clear idea of the sources, a sequential extraction method called factor analysis was performed, and the results documented in several publications including Ahmed (1999) and Ahmed et al. (2005). However, to understand and distinguish the natural and anthropogenic sources and geochemical evolution of water in the area, water samples were also collected from the deep Umm-Er-Radhuma (UER) aquifer. Lithologically, UER consists of calceranitic limestones with some dolomitic limestone, dolomite, minor marl, shale and anhydrite (Powers et al., 1966; ITALCONSULT, 1969; GDC 1980). Isotopic age determinations of UER aquifer groundwater indicate that it is fossil water, estimated to be about 10 000–28 000 years old (Edgell, 1997). Table 4 shows the analytical data of 10 samples collected during the study. TDS, major ions, minor ions and trace elements are in mg l^{-1} whereas conductivity is given in deci-Siemens per meter (dS/m), and pH in standard units. Dissolution of Sylvite (KCl), Halite (NaCl), and Anhydrite ($CaSO_4$) minerals and aquifer material may be responsible for the ions' existence in deep groundwater.

Table 3
Clusters and their members in R-mode for the samples collected in both seasons

Cluster number	Members
Summer	
1	HCO_3^-, Alkalinity, NO_3^-, Br^-, F^-, Ca^{2+}, K^+, Li, SO_4^{2-}, Na^+, B, Cl^- TDS and Conductivity
2	As, Cu, PO_4, Ni, Mo, Mn, Fe
Winter	
1	Br^-, PO_4^-, Mg^{2+}, K^+, Ca^{2+}, SO_4^{2-}, B, Li, Na^+, Cl^-, TDS, Conductivity, NO_3^-, and F^-
2	Co, Cu, Mn, Mo, Zn, HCO_3^- and Alkalinity

Table 4
Analytical data for the samples collected from deep Aquifer (UER)

Sample ID	pH	Cond.	TDS	Alk	HCO_3	Br	Cl	F	NO_3	PO_4	SO_4	Ca	K	Mg	Na	B
DG1-B5	7.65	2.22	1870	131	160	1.34	413	0.77	19.1	<0.2	479	182	13.5	65	238	0.43
DG2-B7	7.55	2.21	1830	131	160	1.1	418	0.76	19.8	0.39	473	179	13.8	64	231	0.42
DG3-E8	7.33	2.2	1840	133	163	1.24	415	0.74	19.7	0.48	473	172.33	12.7	59.9	203	0.43
DG4-F7	7.3	2.2	1850	135	165	1.33	417	0.79	19.4	<0.2	481	185	13.7	65.6	228	0.41
DG5-F8	7.31	2.2	1850	133	163	1.07	417	0.82	19.6	<0.2	475	173	12.4	60	202	0.43
DG6-H8	7.49	2.21	1900	133	162	1.35	410	0.8	18.8	0.38	482	186	13.5	66.1	230	0.43
DG7-H6	7.4	2.18	1770	133	163	1.15	404	0.94	19.5	<0.2	487	187	13.2	65	226	0.43
DG8-H5	7.4	2.19	1840	133	163	1.21	398	0.92	19.7	0.6	482	189	13	65	226	0.43
DG9-C3	7.39	2.2	1920	133	163	0.91	412	0.73	18.5	0.59	488	191	13.5	66.6	230	0.44
DG10-E3	7.4	2.36	1860	133	163	1.41	446	0.76	19.7	0.65	494	191	14.3	65.8	258	0.45

Note: Values less than (<) (example: <0.02) indicates detection limit (DL) used for that particular analyte. All trace elements were found to be below DL.

The main source of the ions Ca^{2+}, Mg^{2+}, Na^+, Cl^-, and SO_4^{2-} in surface water (ponds), which are controlled by the ions pole, is their extensive use in the study area in the form of fertilizers. To some extent their presence is also due to excess irrigation to maintain the soil moisture, as the elements Ca^{2+}, Mg^{2+}, K^+, and SO_4^{2-} are naturally occurring elements in Saudi groundwaters (Sadiq and Alam, 1997). Also the leaching of soil material, mixing of existing salts in soil, and high evaporation and ET rates also lead to very high concentrations of ions that contribute to further deterioration of the water quality of the ponds formed. According to the Water Atlas of Saudi Arabia, the area Al-Fadhley is considered to be one of the hottest parts in Saudi Arabia. The temperature reaches up to 48 °C in summer with very high evaporation and ET rates. The mean monthly evaporation has been estimated to be 150 mm in December–January and 350 mm in June–July (Ministry of Agriculture, 1984; unpublished reports of Ministry). ET during the cold winter season is estimated to be 3–4 mm/yr, while for the hot summer season the values can be as high as 12–13 mm/yr (Water Watch and others, 2005; unpublished report).

7. Conclusions

Agricultural activities have resulted in the release of several toxic metals including arsenic, copper, cobalt, molybdenum, and zinc in waters of the study area as no lithologic source could be found in the area. The TDS level was also found to be very high.

As winter is the irrigation period in the study area when fertilizers and pesticides are used in huge quantities, agricultural impacts were found to be more severe in winter than in summer as expected.

Cluster analysis in Q-mode resulted in three major water types (Cl^--SO_4^{2+}-dominated, Na^+-Cl^--dominated but also containing low levels of SO_4^{2+} and Ca^{2+}, and Na^+-Cl^--dominated with lower levels of SO_4^{2+} Ca^{2+}, and Mg^{2+}) for both the seasons (summer and winter) sampled.

Q-mode cluster analysis indicated recognizable aerial changes in water quality in different seasons.

Cluster analysis in R-mode resulted in two pollutant poles: an ions pole and a trace elements pole.

Acknowledgments

The authors gratefully acknowledge the logistic and field support received from King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia in completing the project. Special thanks to Research Institute, KFUPM for helping us in water analysis. Thanks are also due to Water Resources Division, Abunayyan Group, Riyadh, Saudi Arabia for allowing time and resources in preparing the manuscript. The authors would also like to acknowledge three anonymous reviewers for their critical review of the manuscript and suggestions for further improvements.

References

Ahmed, S.M., 1999. Quality assessment of irrigated water at SHADCO irrigation Project at Al-Fadhley, Eastern Province Saudi Arabia. Master's Thesis, Department of Earth Sciences, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, pp. 131

Ahmed, S.M., Hussain, M., Abderrahman, W., 2005. Using multivariate factor analysis to assess surface/logged water quality and source of contamination at a large irrigation project at Al Fadhli, Eastern Province, Saudi Arabia. Bulletin of Engineering Geology and the Environment 64, 319–327.

Al-Assar, R.S., 1992. Numerical Simulation of Groundwater in the Umm-Er-Radhuma Aquifer at SHADCO Project, Eastern Province Saudi Arabia. Master's Thesis, Department of Civil and Environmental Engineering, King Fahd University of Petroleum & Minerals: Dhahran, Saudi Arabia, pp. 119

APHA, AWWA, WPCF, 1992. Standard Methods for Examination Water and Waste Water Analysis, 18th ed. American Public Health Association, Washington, DC, USA.

Ashley, R.P., Lloyd, J.W., 1978. An example of the use of factor analysis and cluster analysis in groundwater chemistry interpretation. Journal of Hydrology 39, 355–364.

Back W and Hanshaw, B.B., 1965. Chemical geohydrology. Advances in Hydroscience 2, 49–109.

Beke, G.J., Entz, T., Graham, D.P., 1993. Long term quality of shallow groundwater at irrigated sites. Journal of Irrigation and Drainage Engineering 119, 8–23.

Brown, C., 1998. Applied Multivariate Statistics in Geohydrology and Related Sciences, first ed. Springer, Berlin, pp. 248.

Burkhart, R.M., Kolpin, D.W., 1992. Hydrologic and land-use factors associated with Herbicides and Nitrates in near surface aquifers. Journal of Environmental Quality 22, 646–656.

Chandrasekhram, J.K.D., Burner, D.S.Z., Stueben, D., 2001. Arsenic contamination in groundwater, Murshidabad district, West Bengal. In: Manjunath, S. (Ed.), Proceedings of Water-Rock Interaction, Vol. 10. A.A. Balkema Publishers, Leiden, The Netherlands.

Guler, C., Thyne, G.D., McCray, J.E., Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry. Hydrogeology Journal 10, 455–474.

Dalton, M.G., Upschurch, S.B., 1978. Interpretation of hydrochemical facies by factor analysis. Groundwater 16, 228–233.

Edgell, H.S., 1997. Aquifers of Saudi Arabia and their geological framework. The Arabian Journal for Science and Engineering 22, 4–31.

Fetter, C.W., 1992. Contaminant Hydrogeology. Macmillan Publishing Co., New York, pp. 458.

George, V.S., Bouwer, H., Peter, J.W., 1987. Irrigation effects in Arizona and New Mexico. Journal of Irrigation and Drainage Engineering 113, 30–48.

Grande, J.A., Borrego, J., Torre, M.L.D., Sainz, A., 2003. Application of cluster analysis to the geochemistry zonation of the estuary waters in the tinto and odiel rivers (Huelva, Spain). Environmental Geochemistry and Health 25, 233–246.

Groundwater Development Consultants (GDC), 1980. Umm-Er-Radhuma study: Bahrain assignment: Demeter house, station road, Cambridge, CBI 2RS, unpublished report to ministry of agricultural and water, Riyadh.

Helena, B.A., Vega, M., Barrado, E., Pardo, R., Fernandez, L., 1999. A case of hydrochemical characterization of an alluvial aquifer influenced by human activities. Water Air and Soil Pollut. 112, 365–387.

Hopke, P.K., Gladney, E.S., Gordon, G.E., Zoller, W.H., Jones, A.G., 1976. The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. Atmospheric Environment 10, 1015–1025.

Hussein, M.T., 2004. Hydrochemical evaluation of groundwater in the Blue Nile Basin, eastern Sudan, using conventional and multivariate techniques. Hydrogeology Journal 12, 12–144.

ITALCONSULT, 1969. Water and agricultural development studies for area-IV, eastern province, Saudi Arabia: unpublished report to ministry of agricultural and water, Riyadh.

Jones, J.A.A., 1997. Global Hydrology. Addison-Wesley Longman Ltd., Essex, pp. 399.

Mahlknecht, J., Steinich, B., Navarro de Leon, I., 2004. Groundwater chemistry and mass transfers in the Independence aquifer, central Mexico, by using multivariate statistics and mass-balance models. Environmental Geology 45, 781–795.

Martinez, Y., Albiac, J., 2004. Agricultural pollution control under Spanish and European environmental policies. Water Resources Research 40, w10501.

Ministry of Agriculture and Water, 1984. Water Atlas of Saudi Arabia, Riyadh, Kingdom of Saudi Arabia, pp. 112

Olmez, L., Jack, W.B., Villaume, J.F., 1994. A new approach to understanding multiple-source groundwater contamination: factor analysis and chemical mass balance. Water Research 28, 1095–1101.

Piper, A.M., 1953. A graphical procedure in the geochemical interpretation of water analysis. USG.S. Ground Water Note 12.

Powers, R.W., Ramirez, L.F., Redmond, C.D., Elberg Jr., E.L., 1966. Sedimentary geology of Saudi Arabia. US Geological Survey Professional paper. 560-D, pp. 147.

Rapin, F., 1980. Anthropogenic effects on sediment from Bay of Nice and Villefranche-sur-Mer on the French coast. Ph.D. Dissertation, University of Geneva, pp. 221.

Reghunath, R., Sreedhara, M.T.R., Raghavan, B.R., 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Research 36 (10), 2437–2442.

Sadiq, M., Alam, I., 1997. Metal concentration in a shallow groundwater aquifer underneath petrochemical complex. Water Research 26, 3089–3096.

Spalding, R.F., Exner, M.E., 1993. Occurrence of nitrate in groundwater—a review. Journal of Environmental Quality 22, 392–402.

STATISTICA® 5.0 for Windows, 1998. StatSoft, Inc.,Tulsa OK.

USDA, Natural Resources Conservation Services, 1999. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. Agriculture Handbook No. 436, pp. 871.

Usunoff, E.J., Guzman, A., 1989. A multivariate analysis in hydrochemistry: an example of the use of factor and correspondence analyses. Groundwater 27, 27–34.

Voudouris, K., Panagopolous, A., Koumanatakis, J., 2000. Multivariate statistical analysis in the assessment of hydrochemistry of the Northern Korinthia Prefecture alluvial system (Peloponnese, Greece). Natural Resources Research 9 (2), 135–146.

Water Watch, UNDP, MOWE and World Bank, 2005. Historic groundwater abstractions at National scale in the Kingdom of Saudi Arabia. An independent remote sensing investigation, pp. 127.

ELSEVIER

Journal of Environmental Management 86 (2008) 308–318

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Using a choice experiment to measure the environmental costs of air pollution impacts in Seoul

Seung-Hoon Yoo[a,*], Seung-Jun Kwak[b], Joo-Suk Lee[b]

[a]*School of Business and Economics, Hoseo University, 268 Anseo-Dong, Cheonan, Chungnam 330-713, Republic of Korea*
[b]*Department of Economics, Korea University, 5-1 Anam-Dong, Sungbuk-Ku, Seoul 136-701, Republic of Korea*

Received 5 February 2004; received in revised form 4 December 2006; accepted 7 December 2006
Available online 8 March 2007

Abstract

Air pollution, a by-product of economic growth, has been incurring extensive environmental costs in Seoul, Korea. Unfortunately, air pollution impacts are not treated as a commercial item, and thus it is difficult to measure the environmental costs arising from air pollution. There is an imminent need to find a way to measure air pollution impacts so that appropriate actions can be taken to control air pollution. Therefore, this study attempts to apply a choice experiment to quantifying the environmental costs of four air pollution impacts (mortality, morbidity, soiling damage, and poor visibility), using a specific case study of Seoul. We consider the trade-offs between price and attributes of air pollution impacts for selecting a preferred alternative and derive the marginal willingness to pay (WTP) estimate for each attribute. According to the results, the households' monthly WTP for a 10% reduction in the concentrations of major pollutants in Seoul was found to be approximately 5494 Korean won (USD 4.6) and the total annual WTP for the entire population of Seoul was about 203.4 billion Korean won (USD 169.5 million). This study is expected to provide policy-makers with useful information for evaluating and planning environmental policies relating specifically to air pollution.

Keywords: Air pollution; Environmental cost; Choice experiment; Multinomial logit model

1. Introduction

Air pollution is a by-product of economic growth and has become an increasingly critical issue in Korea. In fact, the air pollution problem in Seoul, which is the capital of Korea, has reached a critical level and thus is an imminent issue that needs to be resolved. The pollution levels of Seoul have been reported to often exceed the international air quality standards of air pollutants. The nitrogen dioxide (NO_2) level in Seoul is about 0.036 ppm, which is nearly two times higher than 0.21 ppm, the standard level of the World Health Organization (WHO). Furthermore, the particulate level in Seoul is about 76 $\mu g/m^3$, which is nearly four times higher than those of major industrialized cities such as New York (22 $\mu g/m^3$), London (27 $\mu g/m^3$), and Sydney (18.5 $\mu g/m^3$).

According to the Bank of Korea (2002), air pollution abatement and control expenditures in Korea have been radically increasing.[1] The air pollution abatement and control expenditures in 2001 amounted to 1.3 trillion Korean won (USD 1.1 billion),[2] which is an increase of 1.5 times that in 1994. Moreover, the air pollution abatement and control expenditures of the Korean government in 2001 were 73 billion Korean won (USD 60.8 million), which is 2.5 times higher than those in 1994.

To alleviate this air pollution problem, many policies, regulations, and laws have been enforced. In 1997, the Korean Ministry of Environment designated Seoul as an air pollution restriction zone, and, furthermore, the Seoul Metropolitan Government has implemented an ozone (O_3)

*Corresponding author. Tel.: +82 41 560 8344; fax: +82 41 563 0135.
E-mail addresses: shyoo@hoseo.edu (S.-H. Yoo), sjkwak@korea.ac.kr (S.-J. Kwak), leejoosuk@korea.ac.kr (J.-S. Lee).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.008

[1]Each year, the Bank of Korea reports the pollution abatement and control expenditure to be used as basic data for Green GDP. It includes only actual expenditure incurred for pollution prevention, abatement, and control.

[2]At the time of the survey, USD 1 was approximately equal to 1200 Korean won.

forecasting system where the ozone density is estimated and published in a majority of the daily newspapers the next day. In addition, for those days when the level of ozone density is significantly high, the general public is advised to refrain from outdoor activities and to reduce unnecessary driving.

In 2002, the Korean Ministry of Environment adopted an air quality improvement policy to be implemented in the metropolitan areas. This policy was intended to reduce the level of total suspended particulates (TSP) to 40 $\mu g/m^3$. The policy instruments included levying emission charges, installing filtering systems for large diesel vehicles, and regulating the pollutant quantities of power plants and factories. In 2003, the Clean Air Act was enacted. Its objective is to raise the level of air quality in Seoul within 10 years to the level of the Organization for Economic Cooperation and Development's (OECD) 30 member nations. This act is composed of various initiatives and measures to promote the control of air pollution and encourage the use of less-polluting vehicles.

In evaluating any policy or regulation that would reduce air pollution, it is useful to compare the cost of implementing the policy with the environmental cost of air pollution impacts. If the cost for improving air quality is greater than the environmental cost of air pollution impacts, the general public is mostly likely to be against such policies or regulations to control air pollution. Thus, it would be helpful to know the cost of air pollution, but there is currently no means to calculate the cost of air pollution in monetary value. In other words, without any justified method for computing the cost, policies and regulations for controlling air pollution will be subject to public scrutiny. Meanwhile, proponents of improving air quality such as environmental groups and numerous residents tend to overestimate the environmental cost of air pollution impacts. On the other hand, opponents including industrial plants emitting air pollutants who may be damaged by the regulation or restriction for improving air quality tend to underestimate them. Thus, in order to make an informed public decision and to evaluate the economic efficiency of any policy or regulation, it is necessary to estimate the cost of air pollution impacts by relying on non-market valuation methods.

There have been a number of studies that attempted to measure the cost of air pollution impacts by employing various non-market valuation methods. In Korea, Kim et al. (1998) estimated the average household's annual mean willingness to pay (WTP) for reducing asthma caused by air pollution in Seoul for 1 day by using the contingent valuation method (CVM). Yoo and Chae (2001) estimated the average household's annual mean WTP for reduction of O_3 in Seoul. Kim et al. (2003) calculated the households' marginal WTP for a 4% improvement in mean sulfur dioxide (SO_2) concentrations by using a hedonic price model. In other countries, Alberini et al. (1997) calculated the median WTP of Taiwanese to prevent illness caused by air pollution. In the United States, Delucchi (2000) applied the CVM to estimate the health cost, the visibility cost, the material-damage cost, and the forests and crops damage cost of motor vehicle pollution. Delucchi et al. (2002) applied the hedonic price model to estimate the total cost of anthropogenic TSP pollution. Brajer and Mead (2004) applied a health effect function to estimate the cost for improving air quality to meet the WHO standard in China.

Most of these studies used non-market valuation methods such as the hedonic price model and CVM. Even though these studies have been successful in estimating the cost of air pollution, there is still room for improvements (Hanley et al., 1998). Above all, if there are various kinds of impacts of air pollution, each impact of air pollution should be separately evaluated. However, all these methods have limitations in separating the values placed on individual impacts of air pollution.

In this study, we attempt to apply a choice experiment which offers a promising opportunity to explore the value placed on air pollution improvement in Seoul by considering various attributes related to air pollution such as mortality, morbidity, soiling damage, and poor visibility. This study is significant in that it is the first study ever to apply a choice experiment to measure the environmental costs of air pollution in a newly industrialized Asian nation, where the adverse effects of air pollution may in fact be more severe than in the developed countries. Although the number of studies about air pollution is increasing, there is still limited related data for newly industrialized nations. Even though there are numerical data for newly industrialized nations, they are usually based on approximation and extrapolation and are therefore less reliable than those referring to the developed countries.

The remainder of this study is organized as follows. Section 2 explains the current status of air pollution in Seoul. Section 3 reviews the methodology of the choice experiment. Section 4 presents statistical models to derive costs of air pollution impacts. Section 5 discusses the results. Some concluding remarks are made in the final section.

2. Current status of air pollution in Seoul

According to the environmental white book of the Korean Ministry of Environment, SO_2, carbon monoxide (CO), NO_2, TSP, and O_3 are the main air pollutants in Seoul. In addition to these, some quantities of sulfate, benzene, acetaldehyde, formaldehyde, and dioxins, which are hazardous to human health, have been detected in air in Seoul. In the 1970s and 1980s, SO_2, CO, and TSP were reported as the main air pollutants. In response to the increasing seriousness of air pollution, residents of Seoul started to use natural gas and low-sulfur diesel, which release fewer pollutants, and industrial facilities were relocated outside Seoul. Such efforts to reduce the levels of these pollutants within the city have shown to be quite productive. For example, as shown in Table 1, the average

Table 1
Air pollution trends in Seoul

Year	Air pollutants				
	SO_2 (ppm)	CO (ppm)	NO_2 (ppm)	TSP (μg/m^3)	O_3 (ppm)
1994	0.019	1.5	0.032	88	0.014
1995	0.017	1.3	0.032	78	0.013
1996	0.013	1.2	0.033	72	0.015
1997	0.011	1.2	0.032	68	0.016
1998	0.008	1.1	0.030	59	0.017
1999	0.007	1.1	0.032	66	0.016
2000	0.006	1.0	0.035	65	0.017
2001	0.005	0.9	0.037	71	0.015
2002	0.005	0.7	0.036	76	0.014
Current standard	0.020	9.0	0.050	70	0.006

Source: Korean Ministry of Environment (2004).

levels of SO_2, CO, and TSP decreased by 73.6%, 53.4%, and 13.6%, respectively, over the past 9 years (1994–2002).

However, in the 1990s, O_3 emerged as the main air pollutant. The level of O_3 increased from 0.009 ppm in 1990 to 0.014 ppm in 2002. In addition, Table 1 shows that the level of NO_2 did not change significantly and TSP has increased again since 2001. In spite of the imposition of more stringent air quality regulations, the radical increases in vehicles and energy consumption in Seoul have made it difficult to reduce air pollutants. It should be noted that the total number of vehicles in Seoul has increased from 1,193,633 in 1990 to 2,691,431 in 2002.

Air pollution in Seoul is known to be a cause of some negative effects including the health of the general public of Seoul. The negative health effects are demonstrated by the increased hospital admissions due to the exacerbation of cardiac and respiratory diseases and the increased mortality rate. In Korea, many studies have been done and numerous reports have been published highlighting the deadly health effects of air pollution. Hanyang University Hospital reported that the number of patients, especially those whose age is over 65 years, could be decreased by one-half if the air pollutants in Seoul were decreased by one-third of the current level. The Gyeonggi Research Institute insisted that the air pollution in Seoul is the main cause of the increase in hospital admissions due to respiratory diseases by 13,121 annually. Furthermore, the Korean Ministry of Environment predicted that mortality in Seoul caused by TSP will be increased from 1940 in 2000 to 4000 in 2020.[3]

In addition to the adverse health effects, air pollution in Seoul is regarded as the cause of so-called nuisance or aesthetic effects such as soiling damage and poor visibility (Elliott et al., 1997). In 2002, the annual average pollution level of fine dust in Seoul was 76 μg/m^3, which is the worst among the OECD's 30 member nations. At the time, it was reported that many residents in Seoul experienced discomfort as a result of the soiling damage such as the increase in the frequency and time consumed for cleaning and doing the laundry (Watson and Jaksch, 1992). Recently, the Korean Ministry of Environment reported that the average number of hazy days due to air pollution was about 100 days per year in Seoul. Furthermore, the distance of visibility in Seoul was only 10.9 km in 2000 which is shorter than that in other cities of Korea such as Ulsan (16 km) and Taegu (13.9 km).

3. Choice experiment: methodology and methodological issues

3.1. Methodology

The choice experiment encompasses a variety of multi-attribute preference elicitation techniques widely used by market researchers to evaluate potential new products and new markets for existing products (Garrod and Willis, 1997). The choice experiment is a suitable method for valuing environmental goods with multi-attributes (Baarsma, 2003). Recently, this approach was employed as an alternative to CVM and to complement other preferred methods such as the hedonic price model and travel cost method. In addition, the National Oceanic and Atmospheric Administration (NOAA) in the United States has included this approach in its recent rule-making governing natural-resource damage assessments for oil spills (Johnson and Desvousges, 1997).

The choice experiment has a number of advantages as follows. Above all, it is easier than other valuation methods in estimating the value of each attribute that makes up an environmental good. This is useful because many policies are more concerned with changing attribute levels, rather than losing or gaining the environmental good as a whole (Hanley et al., 1998). It allows respondents to systematically evaluate trade-offs among multiple environmental attributes or among environmental and non-environmental attributes. This trade-off process may encourage respondent introspection and facilitate consistency checks on response patterns (Johnson and Desvousges, 1997). In addition, as it does not ask for the WTP of respondents, it reduces the number of protest responses, especially those involving tax increases or willingness to accept environmental degradation in return for payment. It also increases the amount of information obtained from each respondent, thus reducing the required sample, and hence reducing the costs of the survey.

3.2. Objects to be valued and attributes

Damage from air pollution in Seoul and people's perceptions of this problem have increased the need for adopting new policies relating to air quality improvement. For practical, acceptable, and flexible policies for people

[3]These reports are not available. Only a summary was published in newspapers.

Table 2
Attributes and levels of air pollution impacts

Attributes	Descriptions	Levels
Mortality	Decrease in annual new persons per 10 million of life lost due to lung cancer cases	0#
		2500
		5000
Morbidity	Decrease in annual new persons per 10 million suffering from respiratory disease cases	0#
		250,000
		500,000
Soiling damage	The percentage of decrease in laundry and cleaning costs as well as time cost incurred for do-it-yourselfers	0#
		10
		20
Poor visibility	The kilometer of the increased visibility	0#
		5
		10
Price	Willingness to pay for alleviating air pollutants through increasing traffic charge, electric charge, or ecological tax	0#
		5000
		10,000
		15,000

Note: # indicates the current level of each attribute.

and policy-makers, it is important to know the environmental costs of each air pollution impact. These costs can be assessed in monetary terms through the evaluation of various hypothetical policies that consider various attributes of air pollution impacts and their levels. Thus, this study assumes various hypothetical policies composed of various attributes and estimates the environmental costs of air pollution impacts by assessing these hypothetical policies.

To identify the important attributes of air pollution impacts, we selected a preliminary set of attributes that was derived by extensive literature reviews (Ottinger et al., 1990; Korea Meteorological Administration, 1993; Desvousges et al., 1995; European Commission, 1995; Rowe et al., 1995; Korea Electric Power Corporation, 1997). Then, we reviewed and revised them through extensive interviews with policy analysts, researchers and professors. In addition, we used a focus group to discuss the understanding of and reaction to the attributes and levels in order to refine them. This focus group's input helped us to identify the most important and meaningful attributes. The final attributes were selected based on the five criteria as listed below.

First, the attributes should be independent or nearly independent of one another (Kwak et al., 2001). Second, there should only be a small number of attributes, preferably not more than six because trade-offs become difficult to understand and to show to respondents in comprehensible form if there are too many attributes (Phelps and Shanteau, 1978). Third, attributes should be describable by combining simple explanations and visual instruments such as photographs, charts, and pictures. Fourth, attributes should be scientifically meaningful and important facts about air pollution should not be omitted. Fifth, attributes should have some meaning to people and relate to their reasons for having WTP to avoid environmental impact arising from air pollution. By using these five screening criteria, we identified the following four attributes of air pollution impacts: mortality, morbidity, soiling damage, and poor visibility. Table 2 shows these attributes, as well as the price attribute, and how each level of attributes was defined.

3.2.1. Mortality

The level of attribute for mortality is described as the decrease in annual new person per 10 million of life lost due to lung cancer. To determine the level of attribute, we referred to many previous studies conducted on similar issues. According to Lave and Seskin's (1979) report, if sulfate, which is one of the types of air pollutants, increases by 1 $\mu g/m^3$, the total mortality rate would increase by 6.7 per 100,000 persons. Moreover, Kim (1991) found that if the air pollutants increased by 10%, the total mortality rate would increase by 29 per 100,000 persons. Kim (1982) further reported that the main cause of death was lung cancer. The lowest bound of this attribute is zero if there is no new additional air pollution and the highest bound of this attribute is 5000 per 10 million persons if air pollution is reduced.

3.2.2. Morbidity

The level of attribute for morbidity is the decrease in annual new person per 10 million suffering from respiratory cases. According to Kim (1991), if the air pollutants increase by 10%, the total mortality rate would increase by 32 per 1000 persons. Air pollutants have been reported to cause respiratory diseases such as pneumoconiosis, bronchitis, and asthma. The lowest bound of this attribute is zero if there is no new additional air pollution and the highest bound of this attribute is 500,000 per 10 million persons if air pollution is reduced.

3.2.3. Soiling damage

Soiling damage by air pollutants results in increased household laundry and cleaning costs as well as an increase in the time spent on such activities (Watson and Jaksch, 1992). Thus, the level of attribute is the percentage of decrease in laundry and cleaning costs as well as time spent on such activities. The lowest bound of this attribute is zero if there is no new additional air pollution and the highest bound of this attribute is 20% if air pollution is reduced.

3.2.4. Poor visibility

Visibility means the maximum distance from which a material's shape and color can be seen. Although there may be various reasons for poor visibility such as fog, rain, moisture, and wind, in most cases, the poor visibility in Seoul has been found to be mainly due to air pollutants. The level of attribute is the distance of visibility in kilometers. As the maximum visibility in Seoul is nearly 35 km, we decided that the lowest bound of this attribute is zero if the level of air pollution remains unchanged and the highest bound of this attribute is 10 km if air pollution is reduced.

3.2.5. Price

The price attribute includes monthly electric utility payment and transportation expenditures, which are likely to be familiar items to respondents in general. They have a plausible connection with air quality, in that electric utilities and cars are the main sources of air pollutant emissions.[4] Despite its high level of familiarity and obvious connection with the environmental good, this price attribute may encourage respondents to restrict their WTP amounts to the range associated with a fair or customary expenditure (Mitchell and Carson, 1989). Therefore, we included an ecological tax. The following explanation on price was inserted in the questionnaire.

"To guarantee the success of hypothetical policies presented below, your household should pay a given amount in higher monthly electric utility payment, transportation expenditure, and ecological tax for the air pollution control policy of government".

We decided the levels of the price attribute through a pretest. The lowest bound of this attribute is zero, and the highest bound of this attribute is 15,000 Korean won if the respondent wants the additional air pollution reduced.

3.3. Choice sets

A key problem encountered in a choice experiment is information overload—too many alternatives with too many complex attributes. Thus, a data generating process for instances where there are too many alternatives is essential. This process would rely on carefully designed choice tasks that help to reveal the factors influencing the choice.

In designing a choice experiment, it is important to carefully define the attribute space (including attribute and range) such that the attribute space includes the portion relevant for the policy questions being asked. Furthermore, a choice experiment involves the use of statistical design theory to construct choice sets which can yield coefficient estimates that are not confounded by other factors. In this study, we employed the "orthogonal main effects design" which is effective in terms of isolating the effects of individual attributes on the choice .[5] The ability to "design in" this orthogonality is an important advantage over the revealed preference random utility models, where attributes in reality are often found to be highly correlated with one another (Hanley et al., 1998). The orthogonal main effects design was implemented by using the SAS 8.0 package.

In this study, each attribute was classified into 3 levels, except for price which was classified into 4 levels.[6] In the choice experiment's questions, there were 3 alternatives of which 2 represented the improved air quality and the other represented the fixed status quo. Then, there were $3^4 \times 4 \times 3^4 \times 4$ possible combinations of attributes and levels to form the choice sets. However, since it was impractical to ask respondents to choose from all combinations, we drew a subset of all choice sets for estimating coefficients and drew 48 choice sets. They were then divided into 6 sets of 8 choices each. Fig. 1 shows the example of the choice set that was actually used. Each respondent was presented with 8 choice sets and was asked to choose among the status quo and 2 alternatives.

3.4. Questionnaire design

We prepared a survey questionnaire with the assistance of experts at a polling firm and tested it with a focus group to see how much potential respondents understood the questions. The final version reflected the inputs of the focus group as well as the advice provided by the experts at the polling firm who were assigned to organize the fieldwork.

The final survey questionnaire comprised of 3 sections. The first part was intended to measure respondents' general

[4]For deciding payment vehicle, we referred to the study of Johnson and Desvousges (1997) which used monthly electric utility payment and transportation expenditures for payment vehicles.

[5]An anonymous reviewer commented that there is a problem with the main effects design employed in this study as it assumes that no interactions occur between attributes. In other words, in our analysis, it is impossible to distinguish what the main effect really is from what an interaction aliased with the main effect is. By simply repeating the same main effects design over and over again, it makes any interactions that may be present be always aliased with the same main effects. Even though the main effects design is commonly used in the literature, it is desirable to implement statistical design of choice sets which can fully consider interaction effects between attributes (e.g. see Bullock et al., 1998).

[6]An anonymous reviewer pointed out that our analysis relating to the design comprising of 3 levels per attribute fails to consider the central level and thus may be wasteful, and so suggested that it would be more desirable to set the effects of each attribute to be non-linear rather than linear.

	Alternative A	Alternative B	
Decreased number of life lost due to lung cancer compared with status quo	0	2,500	
Decreased number of patient due to respiratory disease compared with status quo	250,000	0	
The percentage of decrease in laundry and cleaning costs as well as time cost incurred for do-it-yourselfers compared with status quo	0	20	
The kilometers of the increased visibility compared with status quo	10,000	10,000	
Additional monthly tax for alleviating air pollutants	10,000	10,000	
Check with √ the only available alternative which you prefer among Alternative A, B or status quo	□ A	□ B	□ Status quo

Fig. 1. A sample choice set used in this study.

concerns about air pollution, to familiarize them with the attributes of the air pollution impacts being evaluated, and to elicit information about their past experiences with these attributes. To enhance respondents' understanding, a color photograph of a hazy day was inserted into this section. The second part contained choice experiment analysis questions designed to elicit respondents' WTP for alleviating air pollution by estimating trade-offs between price and other attributes. The final part elicited the socioeconomic information of the respondents such as income, age, education, and so on.

3.5. Survey method

Since this study is the first study that used a choice experiment for evaluating the environmental costs of air pollution impacts in Seoul, it was not clear whether the respondents had fully understood the trade-offs between price and other attributes of air pollution described in the scenario. Therefore, we conducted person-to-person interviews where we gave detailed questions to respondents in order to obtain higher response rates.

In all, we conducted 654 person-to-person interviews, from which we obtained a total of 600 data sets. The survey was carried out only in Seoul and among the heads of households or housewives between the age of 20 and 65 years, taking into account the characteristics of the households. Sampling and fieldwork were done by the interview experts of a professional polling firm, Hana Marketing Service, Inc., which is located in Seoul. The interviews were done on randomly selected respondents at their homes to maximize the scope for detailed questions and answers.

4. Model

4.1. Random utility model

Choice experiments share a common theoretical framework with other valuation approaches. Thus, in this study, the random utility model is used to explain individual choices by specifying functions for the utility derived from the available alternatives. This function can be estimated with a multinomial logit (MNL) developed by McFadden (1973). MNL assumes that choices are consistent with the independence from irrelevant alternatives (IIA) property which states that for any individual, the ratio of choice probabilities of any 2 alternatives is entirely unaffected by the systematic utilities of any other alternatives.[7] According to this framework, the indirect utility function U_{ij} for each respondent i who chooses alternative j in the choice set C_i can be expressed as

$$U_{ij} = V_{ij}(Z_{ij}, S_i) + e_{ij}. \tag{1}$$

The indirect utility function U_{ij} can be decomposed into the deterministic part V_{ij}, which is typically specified as a function of the attributes Z_{ij} in alternative j chosen by the respondent i and the respondent i's characteristic S_i, and the stochastic part e_{ij}, which represents the unobservable influence on individual choice. Furthermore, if $U_{ij} > U_{ik}$ for all $j \neq k$ in the choice set C_i, the probability that respondent

[7] An anonymous reviewer pointed out that it is necessary to check if the assumption of IIA holds. However, as this study used the orthogonal main effects design, it is almost impossible to test IIA. In addition, to the best of the authors' knowledge, most choice experiment studies did not check whether the assumption of IIA holds.

i will choose alternative j is given by

$$Pr(j|C_i) = Pr(V_{ij} + e_{ij} > V_{ik} + e_{ik}) = Pr(V_{ij} - V_{ik} > e_{ik} - e_{ij}). \quad (2)$$

In order to deal with this probability, it is necessary to know the distribution of the error term e_{ij}. A typical assumption is that they are independently and identically distributed with an extreme-value (Weibull) distribution, which implies that the probability of any particular alternative j being chosen as the most preferred can be expressed in terms of the logistic distribution (McFadden, 1973). This probability can be expressed as

$$Pr(j|C_i) = \frac{\exp(V_{ij})}{\sum_{k \in C_i} \exp(V_{ik})}. \quad (3)$$

Each respondent's multinomial responses obtained from the questions of the choice experiment scenarios were interpreted as the choice results for the respondents' utility maximization. In this study, each respondent was given 8 choice sets and asked to choose among 3 alternatives including the status quo alternative. The choice results for alternative j of the respondent i were either "yes" or "no". The log-likelihood function can be written as

$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{3} (y_{ij} \ln[Pr(j|C_i)]), \quad (4)$$

where y_{ij} is a binary variable, 1 when the respondent i chooses alternative j among 3 alternatives and 0 otherwise, and N is the total number of respondents. The parameters of this log-likelihood function are estimated by maximum likelihood estimation.

4.2. Utility function and MWTP

4.2.1. Model without covariates

The utility function of the model without covariates, with the exception of the error term e_{ij}, can be expressed as a linear function of an attribute vector (Z_1, Z_2, Z_3, Z_4, Z_5) = (mortality, morbidity, soiling damage, poor visibility, price). It includes one alternative-specific constant (ASC), which represents a dummy for the respondent's choosing the status quo alternative in the choice set. ASC captures the utility of an alternative that the attributes fail to capture (Adamowicz et al., 1994).

$$V_{ij} = \mathrm{ASC}_j + \beta_1 Z_{1,ij} + \beta_2 Z_{2,ij} + \beta_3 Z_{3,ij} + \beta_4 Z_{4,ij} + \beta_5 Z_{5,ij}, \quad (5)$$

where β's are the parameters to be estimated for each attribute that influences the respondent's utility. If we are calculating MWTP from the status quo level of each attribute and assume that all the other variables remain constant, we can obtain the following MWTP estimates by totally differentiating Eq. (5) and omitting i for brevity:

$$\begin{aligned} \mathrm{MWTP}_{Z_1} &= -(\partial V/\partial Z_1)/(\partial V/\partial Z_5) = -\beta_1/\beta_5, \\ \mathrm{MWTP}_{Z_2} &= -(\partial V/\partial Z_2)/(\partial V/\partial Z_5) = -\beta_2/\beta_5, \\ \mathrm{MWTP}_{Z_3} &= -(\partial V/\partial Z_3)/(\partial V/\partial Z_5) = -\beta_3/\beta_5, \\ \mathrm{MWTP}_{Z_4} &= -(\partial V/\partial Z_4)/(\partial V/\partial Z_5) = -\beta_4/\beta_5. \end{aligned} \quad (6)$$

The MWTPs of each attribute represent the marginal rate of substitution between the price and each attribute.

4.2.2. Model with covariates[8]

In order to explain preference heterogeneity and WTP variations among individuals, it is useful to use alternative model specifications where some individual-specific variables (socioeconomic, attitudinal, and past-experience variables) are taken into account. The individual-specific variables include the respondent's income, age, the education level, and a dummy variable for the respondent's smoking in the utility function.

Two versions of alternative models with covariates have been suggested. First, Greene (2002) proposed multiplying demographic variables by dummy variables for each choice within a set. However, if respondents need to answer multiple questions, it would be impractical to use this model because separate dummies would have been required for each distinct alternative to the status quo alternative faced by each individual. Second, Gordon et al. (2001) presented an idea of making the individual-specific variables interact with ASC terms in the utility function. Since this is ideally suited for our data, we chose to interact the four individual-specific variables with ASC. This can be formulated through the following utility function:

$$\begin{aligned} V_{ij} = {} & \mathrm{ASC}_i + \beta_1 Z_{1,ij} + \beta_2 Z_{2,ij} + \beta_3 Z_{3,ij} + \beta_4 Z_{4,ij} \\ & + \beta_5 Z_{5,ij} + \beta_6 \,\mathrm{ASC}_i\, \mathrm{Income}_i \\ & + \beta_7 \,\mathrm{ASC}_i\, \mathrm{Age}_i + \beta_8 \,\mathrm{ASC}_i\, \mathrm{Education}_i \\ & + \beta_9 \,\mathrm{ASC}_i\, \mathrm{Smoke}_i, \end{aligned} \quad (7)$$

where β_1–β_5 are the parameters to be estimated for each attribute which influences the respondents' utility and the range from β_6 to β_9 are the parameters to be estimated for individual-specific variables multiplied by ASC.

The definitions and sample statistics of the covariates used in this study are presented in Table 3. The mean monthly household income of the sample in this study was 2.44 million Korean won (USD 2032). The reported mean income in Seoul for 1997 was 2.47 million Korean won (USD 2058). Therefore, the mean income of the sample in this study approximates the reported mean income of the population very closely at least.

With respect to age, the mean age of the sample in this study was 40.16, and with respect to education, the mean education level in years was 12.64. This means that the median person in the sample had some high-school

[8]According to an anonymous referee's suggestion, we used the mean-centered covariates.

Table 3
Definitions and sample statistics of covariates

Variables	Definition	Mean	Standard deviation
Income	The respondent's monthly income (10,000 Korean won)	243.86	76.34
Age	The respondent's age	40.16	9.90
Education	The respondent's education level in years	12.64	2.54
Smoke	Dummy for the respondent's smoking (1 = yes, 0 = no)	0.35	0.48

education background. Furthermore, the mean of the dummy for smoking was 0.35 which means 65% of the sample were non-smokers.

4.2.3. *The full quadratic model*

The third model that this study estimated is a variation of the full quadratic model. This can be formulated through the following utility function:

$$\begin{aligned} V_{ij} = {} & \mathrm{ASC}_i + \beta_1 Z_{1,ij} + \beta_2 Z_{2,ij} + \beta_3 Z_{3,ij} + \beta_4 Z_{4,ij} + \beta_5 Z_{5,ij} \\ & + \beta_6 Z_{1,ij}^2 + \beta_7 Z_{1,ij} Z_{2,ij} + \beta_8 Z_{1,ij} Z_{3,ij} \\ & + \beta_9 Z_{1,ij} Z_{4,ij} + \beta_{10} Z_{1,ij} Z_{5,ij} \\ & + \beta_{11} Z_{2,ij}^2 + \beta_{12} Z_{2,ij} Z_{3,ij} + \beta_{13} Z_{2,ij} Z_{4,ij} + \beta_{14} Z_{2,ij} Z_{5,ij} \\ & + \beta_{15} Z_{3,ij}^2 + \beta_{16} Z_{3,ij} Z_{4,ij} + \beta_{17} Z_{3,ij} Z_{5,ij} \\ & + \beta_{18} Z_{4,ij}^2 + \beta_{19} Z_{4,ij} Z_{5,ij} \\ & + \beta_{20} Z_{5,ij}^2 n. \end{aligned} \quad (8)$$

5. Results

5.1. *Estimation results of the model without covariates*

The estimation results of the model without covariates are presented in Table 4. All coefficients on the attributes in the indirect utility function are statistically significant at the 5% level. Moreover, their signs are consistent with our expectations. For instance, the coefficients of the four attributes, with the exception of price, are positive. It means that as the level of these attributes increases, the probability of choosing alternatives rather than the status quo alternative increases. The coefficient of the price is negative, which confirms that increasing levels of price have a negative effect on utility.

5.2. *Estimation results of the model with covariates*

The estimation results of the model with covariates are presented in Table 5. Some of the coefficients of covariates are not significant at the 5% level. The coefficient of Smoke and that of Age are insignificant at the 5% level. However, the coefficient of Income and that of Education, that is, β_6 and β_8 are positive and statistically significant at that level. This means that respondents' utility for the status quo alternative decreases with the increase in income and education level of respondents. Especially, the air quality is a normal good and therefore the significant and positive coefficient for income had been expected and conforms to the economic theory.

Table 4
Estimation results of the model without covariates

Variables[a]	Coefficients	t-values[b]
ASC	−2.0676	−25.18**
Mortality (per 1000 persons)	0.0180	2.81**
Morbidity (per 10,000 persons)	0.0223	17.33**
Soiling damage	0.0061	2.12*
Poor visibility	0.0277	4.89**
Price (per 1000 Korean won)	−0.1396	−21.10**
Number of observations	4800	
Log-likelihood	−4349.40	

Notes: * and ** indicate statistical significance at the 5% and 1% levels, respectively.
[a]The variables are defined in Table 2.
[b]The degree of freedom for the value is 4794 (= 600 × 8−6).

Table 5
Estimation results of the model with covariates

Variables[a]	Coefficients	t-values[b]
ASC	−2.0604	−22.92**
Mortality (per 1000 persons)	0.0180	2.81**
Morbidity (per 10,000 persons)	0.0223	17.32**
Soiling damage	0.0060	2.11**
Poor visibility	0.0277	4.88**
Price (per 1000 Korean won)	−0.1396	−21.05**
Income	0.0010	3.14*
Age	0.0004	0.17
Education	0.0223	2.04*
Smoke	0.0353	0.70
Number of observations	4800	
Log-likelihood	−4339.31	

Notes: * and ** indicate statistical significance at the 5% and 1% levels, respectively. We used the mean-centered covariates.
[a]The variables are defined in Tables 2 and 3.
[b]The degree of freedom for the value is 4789 (= 600 × 8−11).

We also present the estimation results of the full quadratic model in Table 6. Only 12 of 20 coefficients are significant at the 5% level. Especially, the signs of β_6 and β_{11} are negative. This means that marginal utility decreases even though the total utility increases with the increase in the level of mortality and morbidity.

5.3. *WTP estimates of each attribute*

The household's MWTP of respondents for obtaining one unit increase from the less preferred level of each attribute can be calculated by using Eq. (6). The results of

Table 6
Estimation results of the full quadratic model

Variables[a]	Coefficients	t-values[b]
ASC	−1.3147	−6.01**
Mortality	0.2560	6.12**
Morbidity	0.0663	10.21**
Soiling damage	0.0412	2.65**
Poor visibility	0.0781	2.60**
Price	−0.1541	−3.53**
Mortality^2	−0.0030	−2.05*
Mortality·morbidity	0.0014	0.18
Mortality·soiling damage	0.0024	2.75**
Mortality·poor visibility	−0.0147	−3.72**
Mortality·price	−0.0495	−1.34
Morbidity^2	−0.0086	−9.65**
Morbidity·soiling damage	0.0052	3.33**
Morbidity·poor visibility	0.0011	0.35
Morbidity·price	−0.0062	−1.45
Soiling damage^2	−0.0010	−1.77
Soiling damage·poor visibility	−0.0005	−0.55
Soiling damage·price	−0.0003	−3.05**
Poor visibility^2	−0.0025	−1.19
Poor visibility·price	0.0008	0.52
Price^2	0.0025	1.16
Number of observations	4800	
Log-likelihood	−4175.76	

Notes: * and ** indicate statistical significance at the 5% and 1% levels, respectively. We used the mean-centered covariates.
[a]The variables are defined in Tables 2 and 3.
[b]The degree of freedom for the value is 4779 (= 600 × 8-21).

Table 7
Marginal willingness to pay (MWTP) estimates and their confidence intervals for the model without covariates

Attributes	MWTP (t-values)	95% confidence intervals
Mortality	0.1289 (2.79)**	[0.0393–0.2184]
Morbidity	0.0160 (18.55)**	[0.0143–0.0176]
Soiling damage	43.40 (2.14)*	[3.53–84.23]
Poor visibility	198.73 (5.09)**	[121.62–276.79]

Notes: The unit is Korean won per month. The t-values are computed by the use of the delta method. * and ** indicate statistical significance at the 5% and 1% levels, respectively.

MWTP estimates of the model with no covariates are shown in Table 7. [9] For example, the average MWTP for mortality per household is 0.1289 Korean won. Its t-value is 2.81, which means that the hypothesis that the MWTP is zero at the 1% level is rejected. Moreover, in order to allow for uncertainty, the confidence intervals for the point estimate of MWTP and the 95% confidence intervals for each attribute are presented.[10] It is interesting to compare the MWTP for a decrease in laundry costs with current average laundry costs. In Korea, as most households wash their clothes themselves, their average laundry cost is below 5000 Korean won (USD 4.17) per month. Therefore, we can guess that MWTP for a decrease in laundry costs is approximately the current average laundry costs.

The WTP estimates of this study provide preliminary information on the benefits of realistic policy scenarios. We assume that a 10% reduction in the levels of concentration of major pollutants in Seoul will result in a decrease in the mortality rate to 29 per 100,000 persons and the morbidity rate to 32 per 1000 persons according to the results of Kim (1991). Next, the household's monthly WTP for a 10% reduction in the levels of concentration of major pollutants in Seoul is 5494 Korea won (USD 4.6). Its 95% confidence interval ranges from 4690 Korean won (USD 3.9) to 6265 Korean won (USD 5.2).[11]

In this case, the monthly WTP for all households in Seoul can be calculated by multiplying the households' monthly WTP by the number of households of Seoul. Then the yearly WTP for all households of Seoul can be calculated by multiplying the monthly WTP by the number of months per year. According to the Korea National Statistical Office, there were 3,085,976 households in Seoul in 2000. Multiplying this by the households' yearly WTP yields the total WTP which is about 203.4 billion Korean won (USD 169.5 million). This result is much greater than the result of other Korean estimates of benefits from improving the air quality. For example, in Yoo and Chae (2001), the annual benefit of O_3 reduction in Seoul was estimated to be between 55.6 billion Korean won (USD 46.3 million) and 70.6 billion Korean won (USD 58.8 million).

6. Concluding remarks

This study was motivated by the need for more quantitative information to help policy-makers to take appropriate actions to ameliorate the air pollution problems. The choice experiment was used to measure environmental costs of individual attributes of air pollution impacts. Overall, the survey was relatively successful in eliciting MWTP values for reducing multiple environmental impacts from air pollution in Seoul. The respondents' choice works for selecting a preferred alternative were found to be within their ability, and the MWTP estimates were statistically different from zero.

In the case of the model without covariates, the mean estimate of households' monthly MWTP was about 0.129

[9]In the model with covariates, the inclusion of selected socioeconomic variables may distort MWTP. In the full quadratic model, some coefficients, which are insignificant statistically, may also distort MWTP. Therefore, we do not report the MWTP estimates for the full quadratic model as well as the model with covariates.

[10]To calculate the 95% confidence intervals for each attribute, the Monte Carlo simulation technique of the Krinsky and Robb (1986) was used.

[11]It can be interpreted as both the benefit from air pollution improvement policy and the environmental cost of the air pollution impacts because this WTP is the cost in relation to the air quality deterioration as well as the benefit from air quality improvement.

Korean won for a reduction in the mortality rate of 1 in 10 million persons. With respect to morbidity rate, the mean estimate of households' monthly MWTP was about 0.0161 Korean won for a reduction of 1 in 10 million persons. Furthermore, the mean estimate of households' monthly MWTP was about 43.40 Korean won per 1% reduction in the soiling damage whereas the mean estimate of households' monthly MWTP was about 198.73 Korean won per 1 km reduction in visibility. According to the model with covariates, as the income and the education level of respondent increase, the respondent's utility for status quo alternative decreases. Therefore, we can expect that the more the people earn and the more educated they are, the more interested they would be in improving air quality.

This study provides insight for both research and policy-making. For research purposes, this study demonstrated the feasibility of extending the use of a choice experiment, at least to the air pollution of a newly industrialized Asian nation. In addition, the study demonstrated that choice experiment is a suitable method for valuing each attribute that makes up the air pollution impacts. From a policy-making perspective, this study provides useful information to help policy-makers in developing and implementing more appropriate policies to deal with air pollution problems. It also illustrates that there is a substantial non-market MWTP to mitigate multiple environmental impacts from the air pollution and that respondents place different values on the changes in attributes from the status quo to specific levels. Lastly, the results from this study provide a useful framework for incorporating such quantitative information in the evaluation of various policies with regard to air pollution.

Acknowledgments

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2002-041-B00122). The authors like to thank the three anonymous reviewers for their valuable comments and constructive suggestions. It is acknowledged that any errors in this paper are the authors' sole responsibility.

References

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining revealed and stated preference methods for valuing environmental amenities. Journal of Environmental Economics and Management 26, 271–292.

Alberini, A., Cropper, M., Fu, T., Krupnick, A., Liu, J., Shaw, D., Harrington, W., 1997. Valuing health effects of air pollution in developing countries: the case of Taiwan. Journal of Environmental Economics and Management 34 (2), 107–126.

Baarsma, B.E., 2003. The valuation of the Ijmeer reserve using conjoint analysis. Environmental and Resource Economics 25, 343–356.

Bank of Korea, 2002. Annual Report of the pollution abatement and control expenditure. Seoul, Republic of Korea.

Brajer, V., Mead, R.W., 2004. Valuing air pollution mortality in China's cities. Urban Studies 41 (8), 1567–1585.

Bullock, C.H., Elston, D.A., Chalmers, N.A., 1998. An application of economic choice experiments to a traditional land use—deer hunting and landscape change in the Scottish highlands. Journal of Environmental Management 52 (4), 335–351.

Delucchi, M.A., 2000. Environmental externalities of motor vehicle use in the US. Journal of Transport Economics and Policy 34, 135–168.

Delucchi, M.A., Murphy, J.J., McCubbin, D.R., 2002. The health and visibility cost of air pollution: a comparison of estimation methods. Journal of Environmental Management 64 (2), 139–152.

Desvousges, W.H., Johnson, F.R., Banzhaf, H.S., Russell, R.R., Fries, E.E., Dietz, K.J., Helms, S.C., 1995. Assessing Environmental Externality Costs for Electricity Generation. Prepared for Northern States Power Company (Minnesota). Triangle Economic Research, Durham, NC.

Elliott, S.J., Kreuger, P., Cole, D., Hall, R., Voorberg, N., Thorne, S., Wakefield, S., 1997. Perceptions of air pollution: the North Hamilton survey. Final Report. Prepared for the Hamilton-Wentworth air quality initiative.

European Commission, 1995. Externalities of energy: external project for the directorate general XII. Metroeconomia, CEPN, IEK, Eyre Energy-Environment, ETSU, Ecole des mines.

Garrod, G.D., Willis, K.G., 1997. The non-use benefits of enhancing forest biodiversity: a contingent ranking study. Ecological Economics 21, 45–61.

Gordon, J., Chapman, R., Blamey, R., 2001. Assessing the Options for Canberra Water Supply: an Application of Choice Modeling in the Choice Modeling Approach to Environmental Valuation. Edward Elgar.

Greene, W.H., 2002. Econometric Analysis, fourth ed. Prentice-Hall, London.

Hanley, N., Wright, R.E., Adamowicz, W., 1998. Using choice experiments to value the environment. Environmental and Resource Economics 11, 413–428.

Johnson, F.R., Desvousges, W.H., 1997. Estimating stated preferences with rated-pair data: environmental, health, and employment effects of energy programs. Journal of Environmental Economics and Management 34, 79–99.

Kim, C.W., Phipps, T.T., Anselin, L., 2003. Measuring the benefits of air quality improvement: a spatial hedonic approach. Journal of Environmental Economics and Management 45, 24–39.

Kim, T.Y., Kwak, S.J., Um, M.J., 1998. Valuation of health impacts owing to air pollution. Korea Journal of Resource Economics 8 (1), 1–26 (in Korean).

Kim, Y.E., 1991. The influence of air pollution to citizens' health. Seoul National University Master Degree Dissertation (in Korean).

Kim, Y.S., 1982. The study of relation between air pollution and mortality. Journal of Korea Health Association 8 (2), 25–38 (in Korean).

Korea Electric Power Corporation, 1997. A Study on the Social Costs of Electric Utilities. Electric Resources Planning Section, Seoul, Republic of Korea (in Korean).

Korea Meteorological Administration, 1993. The Potential Effects of Climate Change in the Korean Peninsula: a Climate Scenario. Meteorological Research Institute, Seoul, Republic of Korea.

Korean Ministry of Environment, 2004. Annual Report of ambient air quality in Korea, 2003, Seoul, Republic of Korea.

Krinsky, I., Robb, A., 1986. On approximating the statistical properties of elasticities. Review of Economics and Statistics 68, 715–719.

Kwak, S.J., Yoo, S.H., Kim, T.Y., 2001. A constructive approach to air quality valuation in Korea. Ecological Economics 38, 327–344.

Lave, L.B., Seskin, E.P., 1979. Air pollution, climate, and home heating: their effects on US mortality rates. American Journal of Public Health 62 (7).

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P., (Ed.), Frontiers in Econometrics. New York.

Mitchell, R.C., Carson, R.T., 1989. Using Surveys to Public Goods: the Contingent Valuation Method. Resources for Future, Washington, DC.

Ottinger, R.L., Wooley, D.R., Robinson, N.A., Hodas, D.R., Babb, S.E., 1990. Environmental Costs of Electricity. Oceana, New York.

Phelps, R.H., Shanteau, J., 1978. Livestock judges: how much information can an expert use? Organizational Behavior and Human Performance 21, 209–219.

Rowe, R.D., Lang, C.M., Latimer, D.A., 1995. New York Environmental Externalities Cost Study. New York.

Watson, W., Jaksch, J., 1992. Air pollution: household soiling and consumer welfare losses. Journal of Environmental Economics and Management 9, 248–262.

Yoo, S.H., Chae, K.S., 2001. Measuring the economic benefits of the ozone pollution control policy in Seoul: results of a contingent valuation survey. Urban Studies 38 (1), 49–60.

Journal of Environmental Management 86 (2008) 319–328

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Kinetic research on the sorption of aqueous lead by synthetic carbonate hydroxyapatite

HuanYan Xu[a], Lei Yang[a,b], Peng Wang[a,*], Yu Liu[c], MingSheng Peng[d]

[a] School of Municipal and Environmental Engineering, Harbin Institute of Technology, Harbin 150090, China
[b] Department of Applied Chemistry, Harbin Institute of Technology, Harbin 150090, China
[c] School of Material Science and Technology, Wuhan Institute of Technology, Wuhan 430073, China
[d] Department of Earth Sciences, Zhongshan University, Guangzhou 510275, China

Received 24 February 2006; received in revised form 22 June 2006; accepted 8 December 2006
Available online 23 March 2007

Abstract

The sorption of aqueous lead on carbonate-hydroxyapatite (CHAp) is a complicated non-homogeneous solid/water reaction, which from the kinetic point of view has two stages. In the first stage, the reaction rate is so fast and the kinetic pathway so intricate that further research is required. In the second stage, the reaction rate slows down and the reaction process follows that of a first-order kinetic equation. Experimental results show that the relationship between the reaction rate constant k_1 and temperature T agrees with the Arrhenius equation, and that the activation energy of sorption (E_a) is 11.93 kJ/mol and the frequency factor (A) is 2.51/s. The reaction rate constant k_1 increases with the Pb^{2+} initial concentration and decreasing pH, but with increasing CHAp dosage. X-ray diffraction (XRD), scanning electron microscopy with energy dispersion spectrum (SEM-EDS) and toxicity characteristic leaching procedure (TCLP) tests indicate that the main sorption mechanism is dissolution-precipitation, in conjunction with surface sorption.

Keywords: Carbonate-hydroxyapatite; Lead ion; Sorption kinetics; Mechanism; TCLP tests

1. Introduction

Lead can be toxic for humans and animals at some concentrations and it is a serious public health issue worldwide. It is a widespread constituent of the earth's crust, and its concentration in soil ranges from 2 to 200 mg/kg and averages 16 mg/kg (Ma et al., 1994). Even low levels of lead in waters are a concern, primarily because of bioaccumulation in the food chain. The number of children presenting with high lead concentrations in blood remains a problem (Arnich et al., 2003). Significant lead contamination of ambient air, soil, dust and wastewater provides exposure which, along with other pathways, will result in lead accumulation in the body. The consequence of debilitating health caused by lead accumulation in humans (and particularly young children) is well-known (NEPC, 1998a). Blood lead is an effective biomarker for lead exposure, for although lead also accumulates in bone, teeth and other tissues, the effects of lead on the human nervous system correlate with blood lead levels (Younes, 1995). Not only are young children more likely to absorb lead (they absorb approximately 50% of all lead inhaled or ingested, compared to approximately 10% for adults) but the effects at lower levels are more likely to be greater and ongoing (Anthony, 2003). Extensive reports and reviews are available on the systemic, immunological, neurological, developmental, reproductive, genotoxic and carcinogenic effects of acute to chronic lead exposure by the inhalation, oral and dermal routes of administration (Freeman et al., 1996). Therefore, considerable effort has been made to treat Pb-containing wastes. Although various methods of treatment, such as coagulative precipitation, adsorption onto chelating resins, ion exchange, reverse osmosis and so on, have been used to remove lead from soils and wastewater, cost-effectiveness and environmental friendliness make *in situ* immobilization and geochemically reactive barriers two promising approaches that have the

*Corresponding author.
E-mail address: xhy7587@hotmail.com (P. Wang).

0301-4797/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2006.12.011

potential to remove lead from solution and/or stabilize lead in soils, wastes and waters.

Resources and environment are the two most popular topics in earth sciences. Minerals are one of the most important resources on earth, and they have been used and investigated in earlier times for their properties as resources. Nowadays, minerals are also extensively used and investigated because of their environmental properties, giving rise to the field of environmental mineralogy. Environmental mineralogy is a new, interdisciplinary subject that investigates not only how minerals cause environmental problems, but also how they resolve them. As far as the latter is concerned, such minerals as zeolite (Chlopecka and Adriano, 1997; Chang et al., 2004a, b; Pitcher et al., 2004; Ören and Kaya, 2006), goethite (Trivedi and Axe, 2001; Mustafa et al., 2004; Pohlmeier and Lustfeld, 2004; Lackovic et al., 2004), magnetite (Stipp et al., 2002; Chang et al., 2004a, b; Martínez et al., 2004), perlite (Alkan and Dogan, 2001; Doğan et al., 2004; Acemioğlu, 2005) and so forth, are used to resolve various environmental pollutions. Apatites have been the object of considerable attention (Miyake et al., 1990; Jeanjean et al., 1996; Chen et al., 1997a, b; Fuerstenau et al., 1997; Admassu and Breese, 1999; Manecki et al., 2000; Mavropoulos et al., 2002; Arnich et al., 2003; Mavropoulos et al., 2004; Gómez del Rio et al., 2004; Peld et al., 2004; Lee et al., 2005; Raicevic et al., 2005; Knox et al., 2006) because of their peculiar crystal-chemistry characteristics, and it has been proposed that they can be used to remove heavy metal ions, including lead, from contaminated wastewaters (Mavropoulos et al., 2004). They can also be used as a barrier to minimize heavy metal migration from heavy-metal-containing solid wastes or soils (Chen et al., 1997a, b).

The efficiency with which apatite removes aqueous lead has encouraged extensive fundamental research into the thermodynamics and mechanisms of lead removal. Two main mechanisms, ion exchange between Pb^{2+} and Ca^{2+} at the apatite lattice (Suzuki et al., 1984; Miyake et al., 1990; Shashkova et al., 1999; Sugyiama et al., 1999) and apatite dissolution followed by lead phosphate precipitation (Ma et al., 1994; Chen et al., 1997a, b; Laperche et al., 1996, 1997; Zhang et al., 1998), have been put forward. At present, neither of the mechanisms has been clearly identified because of the lack of analytic testing techniques. These results have only been confirmed by X-ray diffraction (XRD) (Mavropoulos et al., 2004). At the same time, thermodynamic research, which could help us to analyze the mechanisms for removing lead by apatite, has not been carried out. Prasad and co-workers have made outstanding contributions to thermodynamic research into the sorption of heavy metal ions by low-grade rock phosphate (francolite) (Prasad et al., 2000, 2002; Prasad and Saxena, 2004). They observed that the sorption kinetics appeared to be mainly controlled by liquid-film diffusion. External-diffusion and film-diffusion models were tested to evaluate mass-transfer coefficients and film-transfer constants at different initial concentrations of metal cations. The applicability of the Langmuir and Freundlich adsorption isotherms in each case of metal ion adsorption was studied separately at different temperatures. But factors affecting the reaction rate constant have not been discussed in detail. Likewise, only XRD has been used for mechanism research. Natural rock phosphate consists not only of fluorapatite but also of many other minerals, such as calcite, quartz, dolomite and clays, which, according to the literature, can also remove aqueous lead ion (Huang and Fuerstenau, 2001; Rouff et al., 2005; Hizal and Apak, 2006). Therefore, fluorapatite and many other minerals must be considered when analyzing the mechanism.

In this study, synthetic hydroxyapatite (HAp, $Ca_{10}(PO_4)_6(OH)_2$) is used for the sorption of aqueous lead ion. In particular, we focus on the kinetic process of the sorption, consisting of the kinetic equation and the factors that affect the reaction rate constant. The sorption mechanism is analyzed using XRD, scanning electron microscopy with energy dispersion spectrum (SEM-EDS) and toxicity characteristic leaching procedure (TCLP) tests. From the point of view of kinetics, and using several testing techniques, we try to gain further insight into the sorption of lead ion by HAp.

2. Experimental

2.1. Hydroxyapatite preparation

The hydroxyapatite used in the experiments was synthesized by a sol–gel method (Sun et al., 1992), which takes place in a constant-temperature water retort, and the reaction can be described by the following chemical equation:

$$10Ca(OH)_2 + 6H_3PO_4 = Ca_{10}(PO_4)_6(OH)_2 + 18H_2O.$$

The chemical reagents used were all analytical grade. The synthesis procedure can be described as follows: the suspension of 1.0 mol/L $Ca(OH)_2$ was titrated by slowly adding 0.6 mol/L H_3PO_4. The reaction vessel was heated to 40 °C and stirred intensively. In the course of the reaction, pH was measured and the titration rate was controlled in order to keep the suspension pH above 7. Then the suspension was kept under the mother liquid conditions for 24 h. The precipitates were filtrated with doubly distilled water three times through a Whatman 41 filter paper, dried at 100 °C for 12 h and calcined at 600 °C for 2 h. Finally, the calcined sample was manually ground and the particles with a size of <74 μm were separated by sieving and used in these experiments. The synthetic method is similar to that reported by Smičiklas and co-workers. The main difference is that they used a nitrogen atmosphere to prevent CO_2 from being absorbed and incorporated into the hydroxyapatite crystal (Smičiklas et al., 2000).

2.2. Testing techniques

XRD was used to identify mineral phases and estimate crystallinity. Measurements were made on a Shimazu XD-5A instrument with Cu-Kα radiation ($\lambda = 1.54056$ Å) at 30 mA and 35 kV in the Bragg–Brentano θ–θ geometry. Samples used for XRD analysis were air-dried, crushed, and mounted on a glass sample holder. The XRD patterns were obtained using a graphite monochromator, divergence and scattering slits of 1°, a detector slit of 0.15 mm, and a step scanning rate of 2° 2θ/s.

Fourier transform infrared spectroscopy (FT-IR) was used to confirm the chemical groups in the Hap structure. Measurements were made on a SP3-300 IR spectrometer after Hap samples were mixed with 300 mg of spectroscopic grade KBr and ground in an agate mortar. Spectra were obtained over a range of 400–4000 cm^{-1} with a resolution of 2 cm^{-1}. The estimated wave-number accuracy for sharp bands was about 1 cm^{-1}.

A scanning electron microscope (SEM) equipped with X-ray energy dispersive spectroscopy (EDS) was used to determine the crystal morphology and chemical element composition. Measurements were made on a JSM-5500 SEM instrument with digital imaging processing. The samples were mounted on stainless stubs using double-stick tape and coated with gold. The SEM was operated at 25 kV for secondary electron (SE) imaging and 20 kV for EDS analysis. Backscattered electron (BSE) images were obtained to observe the Pb-rich phases.

2.3. Sorption experiments

The amount of lead retained in the solid sample was determined by batch experiments. A lead-bearing solution of 1000 mg/L was obtained from analytical-grade acetic lead, subsequently diluted to the required concentrations. The pH was adjusted by adding sodium hydroxide and nitric acid to the lead solutions. Then, HAp was added to several 100 mL fractions of the lead solution and mechanically shaken for some time at room temperature. After filtration, the lead concentration of the filtrate was determined by spectrophotometric methods with dithizone, using a 752-type ultraviolet-visible spectrophotometer. For all sorption tests, blank experiments were carried out with the same experimental procedure to check the extent of lead sorption by the glass flasks, which must be deducted. In order to check the reproducibility of the results, random tests were made under different experimental conditions.

2.4. TCLP tests

The United States Environmental Protection Agency (US EPA) Land Disposal Restriction (LDR) Program sets treatment standards for the waste from land disposal. The treatment standards are often based on the constituent concentrations in the TCLP waste extract. Thus, the TCLP, which replaced the extraction procedure (EP), is commonly used either to determine whether a waste is hazardous or to determine whether a treated waste meets the treatment standards for land disposal. The procedure is designed to study the mobility of both organic and inorganic matter present in liquid, solid and multiphasic wastes and, thus, must be considered in the development of treatment technologies. It can also be used in desorption studies (Chen et al., 1997a, b). In our study, the TCLP test method was applied to determine the leaching behavior of lead from carbonate-substituted hydroxyapatite (CHAp) solid residues after lead sorption. For this purpose, 100 g of sample was placed in a plastic bottle together with 2000 mL of leach solution, a sodium acetate/acetic acid buffer solution, the pH of which was 4.93. The mixture was then agitated at 18 rpm for 24 h. The mixture is filtered through a 0.45 μm glass fiber filter and the filtrate obtained was analyzed (USEPA, 1990).

3. Results and discussion

3.1. Hydroxyapatite characterization

The FT-IR spectrum is shown as Fig. 1, from which it can be observed that there are two strong absorption bands at 3575.7 and 645 cm^{-1}, assigned to the stretching mode and bending mode of the OH group in the HAp structure. Absorption double bands due to the CO_3 group in the HAp structure are observed at 1455.0 and 1411.5 cm^{-1}. These bands have been mentioned in other literature (Tang et al., 2003), and are different from the single band of carbonate, which indicates that there are CO_3 groups that are the result of the dissolution of $CO_{2(g)}$ in the reaction solution, the formation of CO_3^{2-}, and subsequent substitution with PO_4 groups in the HAp structure. The XRD pattern (Fig. 2) shows that the sample has the typical hexagonal structure of hydroxyapatite (matching JCPDS 09-0432), with unit-cell parameters $a = 0.9400$ nm, $c = 0.6879$ nm, $c/a = 0.7318$ and $V = 0.52645$ nm^3, calculated by the

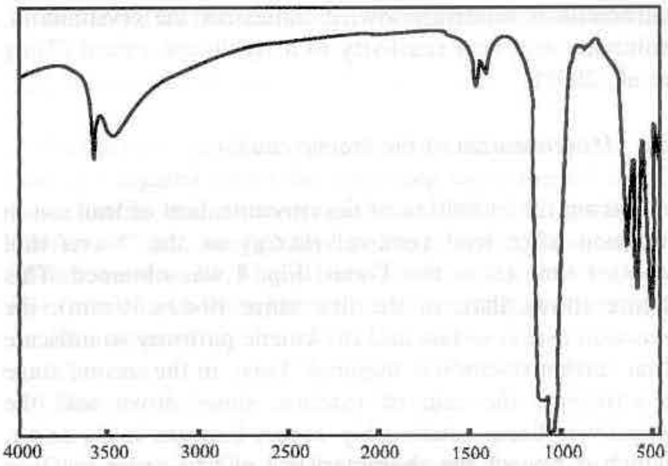


Fig. 1. IR spectrum of CHAp synthesized by the sol–gel method.

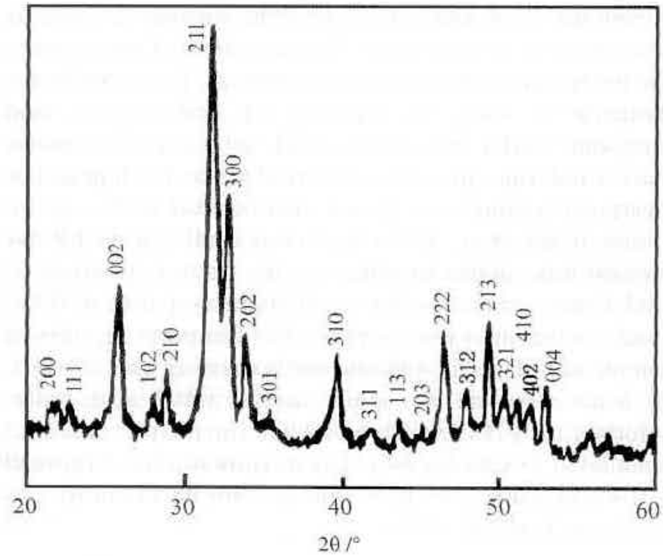


Fig. 2. XRD pattern of CHAp synthesized by the sol–gel method.

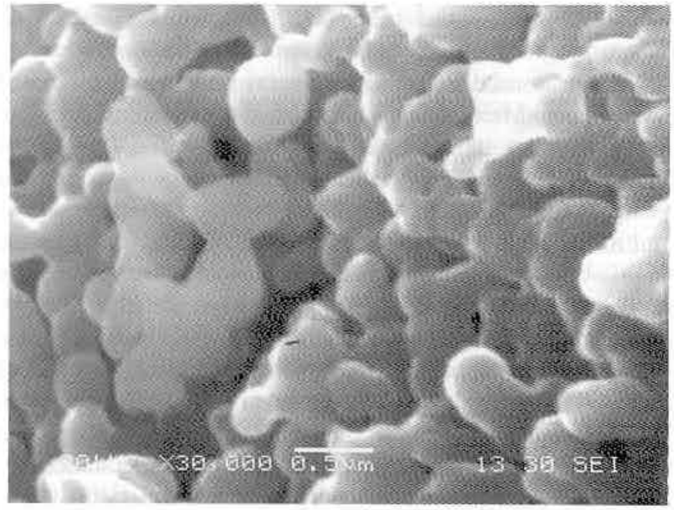


Fig. 3. SEM micrograph of CHAp synthesized by the sol–gel method.

parameter refinement program. The poor crystallinity, which is caused by the substitution of PO_4 groups by CO_3 groups (Liu and Comodi, 1993), is confirmed by the XRD result. The FT-IR and XRD results in this study are in accordance with that of the ultra-fine bioresorbable carbonated hydroxyapatite produced by Murugan and Ramakrishna (2006). Observation by SEM (Fig. 3) revealed that the crystals are on the nanoscale, ranging from 50 to 200 nm. At the same time, the sample has a large specific surface area ($51.86\,m^2/g$), measured by the single-point BET method (using liquid N_2), which can help in the sorption of lead ion. The wet chemical method was used to analyze the chemical composition of the specimen and the results are as follows: P_2O_5, 35.81%; CaO, 45.19%; and loss on ignition, 18.77%. From the results, it can be calculated that the molar ratio Ca/P of the specimen is 1.60, less than that of stoichiometric hydroxyapatite (1.67), which also shows that there is substitution between CO_3 and PO_4 in the HAp lattice. So, the sample synthesized is CHAp. Although the amount of carbonate is relatively low, it influences the crystallinity, solubility and acid reactivity to a significant extent (Tang et al., 2003).

3.2. Determination of the kinetic equation

Taking the logarithm of the concentration of lead ion in solution after lead removal ($\ln C_R$) as the Y-axis and contact time (t) as the X-axis, Fig. 4 was obtained. This figure shows that, in the first stage ($0<t<10$ min), the reaction rate is so fast and the kinetic pathway so intricate that further research is required. Then, in the second stage ($t>10$ min), the rate of reaction slows down and the negatively linear relationship occurs between $\ln C_R$ and t, which is one of the characteristics of first-order reaction kinetics. Consequently, the reaction process follows a first-order reaction kinetic equation, in the second stage, and

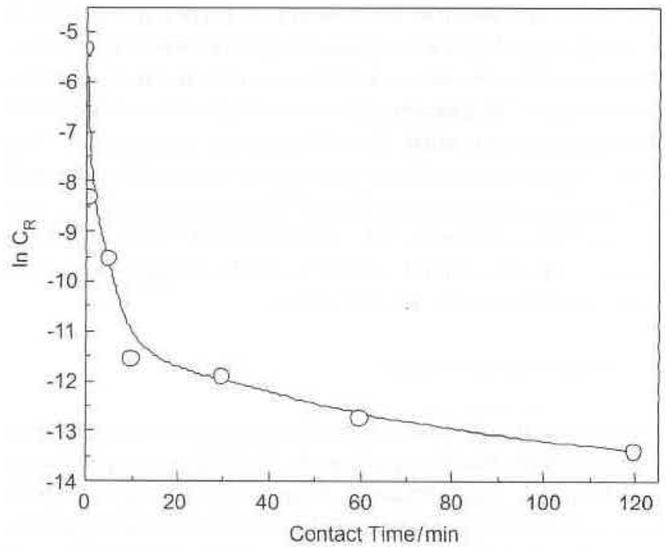


Fig. 4. Curve between $\ln C_R$ and contact time.

can be described as the following equation:

$$\ln C_R = -k_1 t + B,$$

where B is an integral constant and k_1 is a reaction rate constant. All in all, the sorption of aqueous lead on CHAp is a complicated non-homogeneous solid/water reaction, which can be described in two kinetic stages. In the first stage, the kinetic pathway is intricate, and requires further research. In the second stage, the reaction rate slows down and the reaction process follows a first-order kinetic equation.

3.3. Impact of temperature on k_1

The impact of temperature T on the reaction rate constant k_1 is shown in Fig. 5, and the experimental conditions were the following: the pH was 2, the dosage of CHAp was 5 g/L, the initial lead concentration was 1000 mg/L and the contact time was 30 min. It was

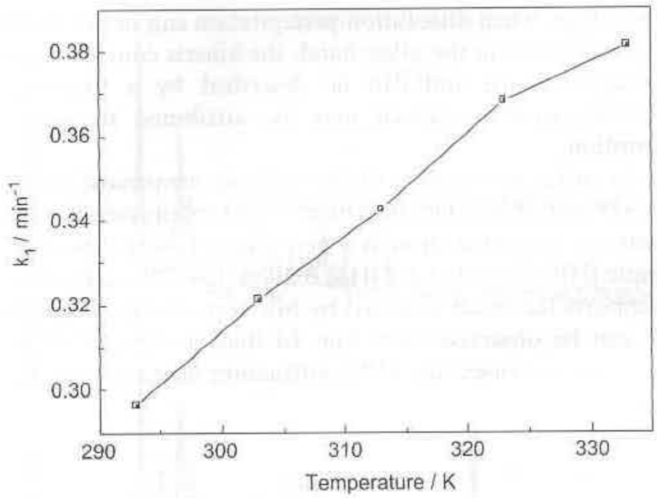


Fig. 5. Effect of the temperature on constant k_1.

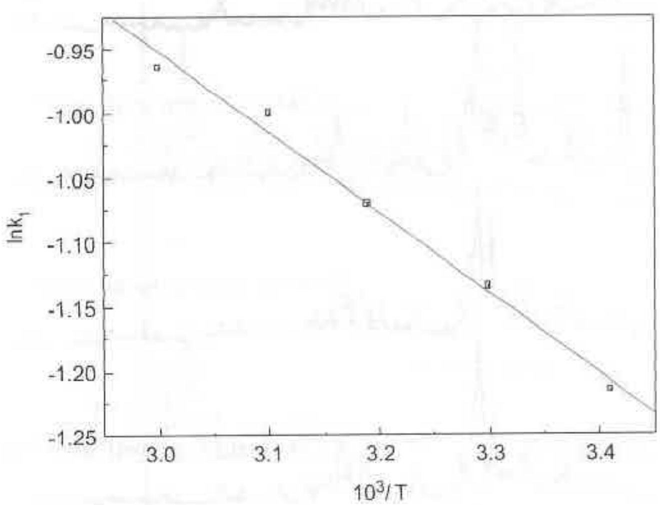


Fig. 6. Fitting curve between $\ln k_1$ and $1/T$.

observed that the reaction rate constant k_1 increased as the temperature increased. In all the chemical reactions, the relationship between temperature and k_1 is complex, and can be described by five different equations. In our study, the relationship between temperature and k_1 followed the Arrhenius equation, because of the linear relationship between $\ln k_1$ and $1/T$ (see Fig. 6). After fitting, the linear equation was obtained as

$$\ln k_1 = -623.22/T + 0.91712$$

with a reliable coefficient $R = 0.9945$. Comparing this equation with the Arrhenius equation, it was calculated that the activation energy of sorption (E_a) was 11.93 kJ/mol and the frequency factor (A) was 2.51 s^{-1}.

3.4. Impact of initial lead concentration on k_1

Under the experimental conditions of pH 2, dosage of CHAp of 5 g/L, temperature of 298 K and contact time of 30 min, we investigated the impact of initial lead concentration on the reaction rate constant k_1. This is shown in Fig. 7, where it can be observed that the reaction constant k_1 decreased as the initial lead concentration increased.

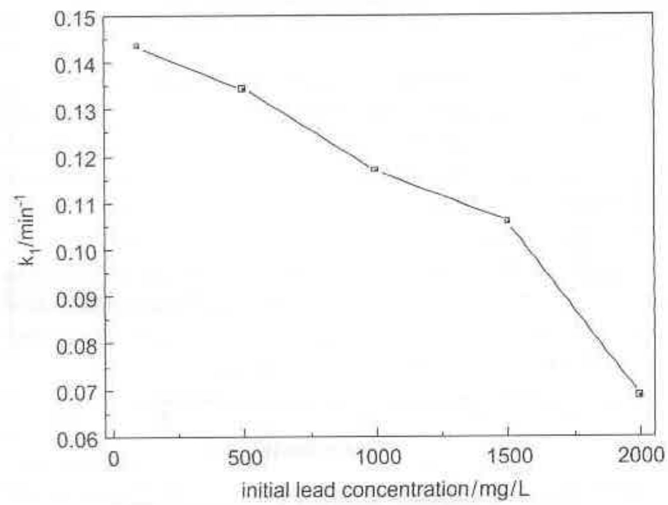


Fig. 7. Relationship between initial Pb^{2+} concentration and k_1.

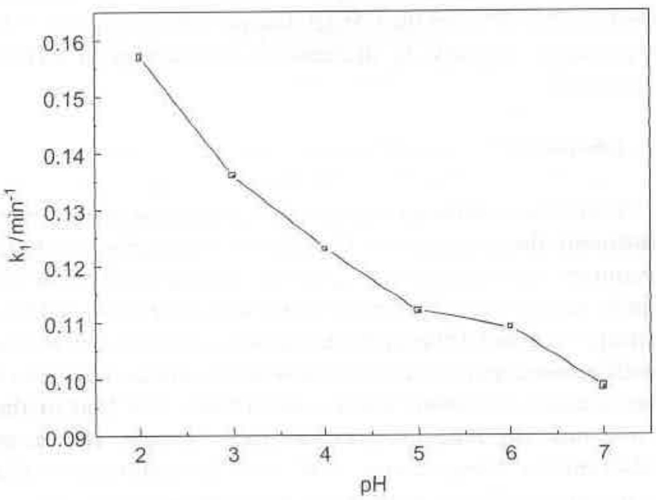


Fig. 8. Relationship between pH and k_1.

3.5. Impact of pH on k_1

The impact of pH on the reaction rate constant k_1 was also investigated under the following experimental conditions: initial lead concentration of 1000 mg/L, dosage of CHAp of 5 g/L, temperature of 298 K and contact time of 30 min (see Fig. 8). The pH and the reaction rate constant k_1 exhibited positive linear relativity: namely, k_1 increased, as pH increased.

3.6. Impact of dosage of CHAp on k_1

The impact of the CHAp dosage on the reaction rate constant k_1 is shown in Fig. 9 (experimental conditions: lead initial concentration: 1000 mg/L, pH = 2, temperature:

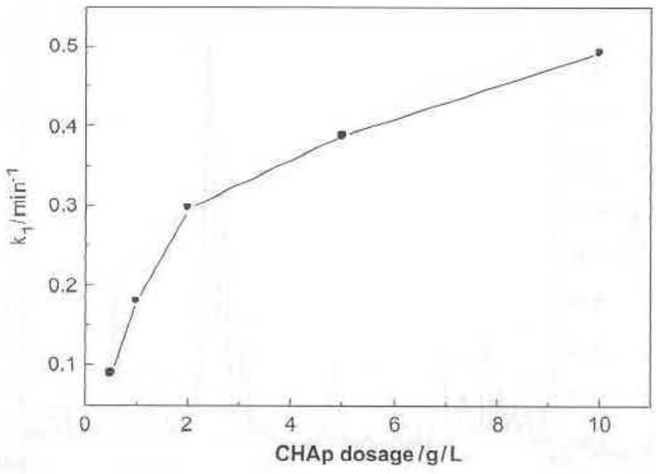


Fig. 9. Relationship between CHAp dosage and k_1.

298 K, and contact time: 30 min). There was a positive linear relationship between the CHAp dosage and the reaction rate constant k_1: namely, k_1 increased as the dosage of CHAp increased.

4. Discussion

As mentioned above, the removal of aqueous lead ion by carbonate-hydroxyapatite (CHAp) is a complicated non-homogeneous solid/water reaction, which is affected by such factors as pH, initial lead concentration, CHAp dosage, contact time and temperature. It is a feature of multiphase reactions that reaction occurs at the interface of two phases, or within a particular phase, and that at the same time the reactant pervades the interface. Hence, in addition to temperature, pH and concentration, the reaction rate of the multiphase reaction is determined by the size and character of the interface and the sorption of the reactant. Generally speaking, multiphase reactions consist of the following processes:

(i) the reactant pervades onto the interface;
(ii) the reactant molecule is adsorbed onto the interface;
(iii) the adsorbed molecule reacts on the interface;
(iv) the end-product is desorbed;
(v) the product separates from the interface through permeation.

The reaction rate of the multiphase reaction is controlled by the rate of the slowest process mentioned above. Different procedures make the reaction behave differently, because pervasion, sorption and reaction follow different rules.

In our study, the poor crystallinity and the nanoscale mean that CHAp can diffuse very fast in acid aqueous solution. At the same time, lead ions in solution can be quickly adsorbed at the water/CHAp interface, followed by the dissolution-precipitation or ion-exchange reaction. The rapid reaction rate complicates the kinetic course in the first stage, when dissolution-precipitation can occur. In the second stage, on the other hand, the kinetic course is more straightforward and can be described by a first-order kinetic equation, which may be attributed to surface sorption.

In order to confirm the hypothesis mentioned above, XRD and SEM-EDS were used. XRD results (see Figs. 10 and 11) show that there is a new phase: hydroxypyromorphite [HPY, $Pb_{10}(PO_4)_6(OH)_2$, JCPDS 02-0700 card], which supports the result reported by Mavropoulos et al. (2004). It can be observed from Fig. 10 that, as the pH of the solution increases, the HPY diffraction peaks get weaker,

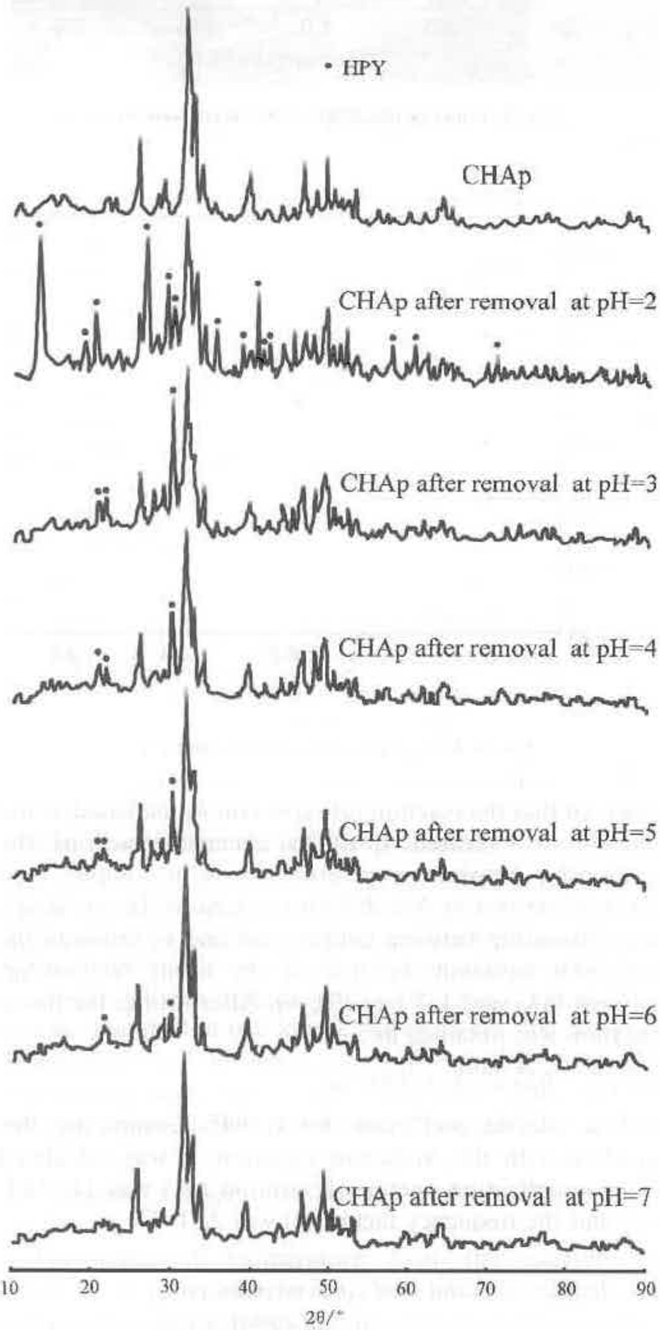


Fig. 10. XRD patterns of CHAp after removal of lead ions in different pH conditions.

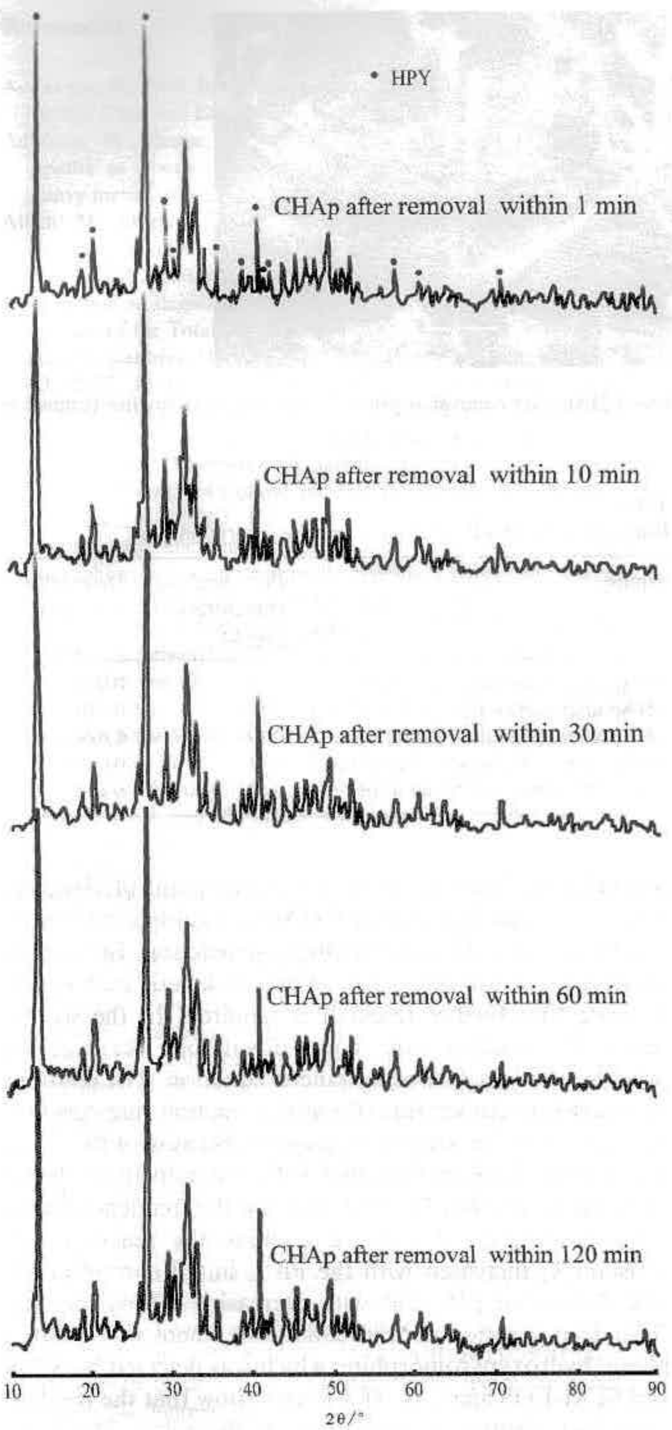


Fig. 11. XRD patterns of CHAp after removal of lead ions with different contact times.

which indicates that the amount of newly formed HPY decreases. Meanwhile, the solubility of CHAp decreases as the pH solution increases (Xu, 2004), so it can be concluded that the removal of aqueous lead ion is closely related to the dissolution of CHAp and the following precipitation of HPY, which indicates that the removal mechanism is dissolution-precipitation. At the same time, judging from the XRD result (see Fig. 11), the dissolution-precipitation reaction is so rapid that HPY can form in 1 min.

SE imaging of SEM (Fig. 12) shows that, at $pH = 7$, there was no change in CHAp crystal morphology before or after aqueous lead ions were removed, but that there was a needle and sheet crystal morphology after removal at $pH = 2$, which is in agreement with the results of XRD. In addition, the characteristics of element distribution can be obtained from the SEM back-scattering electron composition image (see Fig. 13), in which there is a bright needle and sheet area, where the Pb-rich phase is found. EDS analysis (Fig. 14) shows that there was a large amount of lead in the solid phase after the lead ions were removed at $pH = 2$.

In addition, both the solubility and the lattice energy of Pb-bearing hydroxyapatite are lower than those of hydroxyapatite (Crannell et al., 2000; Flora et al., 2004):

Solubility:	$Pb_{10}(PO_4)_6(OH)_2$ $\log K_{sp} = -125.6$
	$Ca_{10}(PO_4)_6(OH)_2$ $\log K_{sp} = -76.30$
Lattice energy:	$Pb_{10}(PO_4)_6(OH)_2$ 33518 kJ/mol
	$Ca_{10}(PO_4)_6(OH)_2$ 34092 kJ/mol

According to the rule of solubility and the least energy principle, the precipitation of $Pb_{10}(PO_4)_6(OH)_2$ can occur spontaneously. So dissolution-precipitation is preferred to ion-exchange for the lead sorption mechanism in acid conditions ($2 < pH < 6$).

The TCLP test results are shown in Table 1. When the pH of the sorption solution increased, the Pb^{2+} leaching concentration of CHAp after removal also increased: that is to say, the lower the pH of the sorption solution, the better the environmental stability and desorption ratio of CHAp after the lead ions have been removed. Otherwise, all the Pb^{2+} leaching concentrations in the TCLP tests were below the legal standard of 5.0 mg/L, which shows that after lead ion removal the environmental stability of CHAp is better. In addition, the results of the TCLP experiments show that surface sorption is one of the removal mechanisms during the removal of lead ions by CHAp, and that the lower the pH of the sorption solution, the lower the desorption ratio.

So, it can be concluded that, besides dissolution-precipitation, surface sorption is one of the lead sorption mechanisms. As $H_2PO_4^-$ is the dominant phosphate species in the range $pH < 7$, for the dissolution-precipitation, these reactions could be represented as

$$Ca_{10}(PO_4)_6(OH)_2 + 14H^+ \rightarrow 10Ca^{2+} + 6H_2PO_4^- + 2H_2O,$$

$$10Pb^{2+} + 6H_2PO_4^- + 2H_2O \rightarrow Pb_{10}(PO_4)_6(OH)_2 + 14H^+$$

as indicated by many scholars (Ma et al., 1994; Chen et al., 1997a, b; Laperche et al., 1996, 1997; Zhang et al., 1998). Although they also considered dissolution-precipitation as the main mechanism, they have no supporting evidence from XRD or SEM-EDS, or TCLP tests.

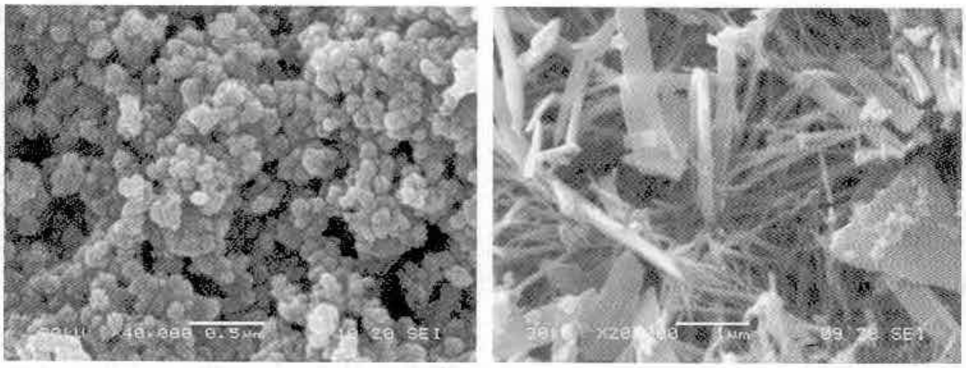


Fig. 12. Secondary electron images of SEM for CHAp after removal of lead ions (CHAp after removal at pH = 7(× 40,000)) (CHAp after removal at pH = 2(× 20,000)).

Fig. 13. Back-scattering electron composition images of SEM for CHAp after removal of lead ions at pH = 2.

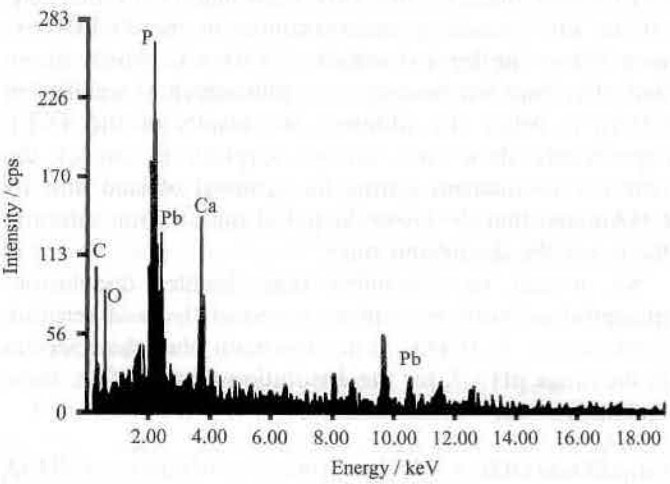


Fig. 14. EDS composition analysis for CHAp after removal of lead ions at pH = 2.

5. Conclusion

As an environmental mineralogical material, CHAp can effectively remove lead ions in aqueous solution under different experimental conditions and there are several potential applications. From the kinetic point of view, the removal of aqueous lead by CHAp is a complicated, non-homogeneous solid/water multiphase reaction. In the first stage, the reaction rate is so fast and its kinetic pathway so intricate that further research is required. In the second stage, the reaction rate slows down and the reaction process follows a first-order kinetic equation. The relationship between temperature T and the reaction rate constant k_1 agrees with the Arrhenius equation, because of the linear relationship between $\ln k_1$ and $1/T$. The activation energy of sorption (E_a) was 11.93 kJ/mol and the frequency factor (A) was 2.51 s^{-1}. At the same time, the reaction rate constant k_1 increased with the Pb^{2+} initial concentration and decreasing pH, and with increasing CHAp dosage. After lead sorption in acid conditions, there was another phase: hydroxypyromorphite, which was detected by XRD and SEM-EDS analysis. TCLP tests show that the residues after lead sorption are environmentally stable. The lower the sorption solution pH, the lower the desorption ratio. So, it is proposed that the lead removal mechanism is dissolution-precipitation, in conjunction with surface sorption.

Table 1
Data about the TCLP experiment and the desorption ratio

Sample	Pb^{2+} leaching concentration (mg/L)	Desorption ratio (%)
CHAp after removal of lead at pH = 2	0.11	0.011
CHAp after removal of lead at pH = 3	0.16	0.016
CHAp after removal of lead at pH = 4	0.19	0.029
CHAp after removal of lead at pH = 5	0.41	0.064
CHAp after removal of lead at pH = 6	0.69	0.132

Acknowledgement

The authors thank Baohua Zhang and Lei Li (Center for Analysis and Testing, Wuhan Institute of Technology) for the Scanning Electron Microscopy and Infrared Spectroscopy measurements, respectively.

References

Acemioğlu, B., 2005. Batch kinetic study of sorption of methylene blue by perlite. Chemical Engineering Journal 106, 73–81.

Admassu, W., Breese, T., 1999. Feasibility of using natural fishbone apatite as a substitute for hydroxyapatite in remediating aqueous heavy metals. Journal of Hazardous Materials B 69, 187–196.

Alkan, M., Dogan, M., 2001. Adsorption of copper(II) onto perlite. Journal of Colloid and Interface Science 243, 280–291.

Anthony, L.M., 2003. An assessment of the effectiveness of lead pollution reduction strategies in North Lake Macquarie, NSW, Australia. The Science of the Total Environment 303, 125–138.

Arnich, N., Lanhers, M.C., Laurensot, F., Podor, R., Montiel, A., Burnel, D., 2003. In vitro and in vivo studies of lead immobilization by synthetic hydroxyapatite. Environmental Pollution 124, 139–149.

Chang, C.F., Chang, C.Y., Chen, K.H., Tsai, W.T., Shie, J.L., Chen, Y.H., 2004a. Adsorption of naphthalene on zeolite from aqueous solution. Journal of Colloid and Interface Science 277, 29–34.

Chang, C.M., Wang, Y.J., Lin, C., Wang, M.K., 2004b. Novel predicting methods for the removal of divalent metal ions by magnetite/amorphous iron oxide composite systems. Colloids and Surfaces A: Physiochemical and Engineering Aspects 234, 1–7.

Chen, X.B., Wright, J.V., Conca, J.L., Peurrung, L.M., 1997a. Effects of pH on heavy metal sorption on mineral apatite. Environmental Science and Technology 31, 624–631.

Chen, X.B., Wright, J.V., Conca, J.L., Peurrung, L.M., 1997b. Evaluation of heavy metal remediation using mineral apatite. Water, Air and Soil Pollution 98, 57–78.

Chlopecka, A., Adriano, D.C., 1997. Influence of zeolite, apatie and Fe-oxide on Cd and Pb uptake by crops. The Science of the Total Environment 207, 195–206.

Crannell, B.S., Eighmy, T.T., Krzanowski, J.E., Dykstra Eusden Jr., J., Shaw, E.L., Francis, C.A., 2000. Heavy metal stabilization in municipal solid waste combustion bottom ash using soluble phosphate. Waste Management 20, 135–148.

Doğan, M., Alkan, M., Türkyilmaz, A., Özdemir, Y., 2004. Kinetics and mechanism of removal of methylene blue by adsorption onto perlite. Journal of Hazardous Materials B 109, 141–148.

Freeman, G.B., Dill, J.A., Johnson, J.D., Kurtz, P.J., Parham, F., Matthews, H.B., 1996. Comparative absorption of lead from contaminated soil and contaminated salts by weanling fischer 344 rats. Fundamental and Applied Toxicology 33, 109–119.

Fuerstenau, M., Zhong, K., Hu, W., Liu, Y., 1997. Remediation of heavy metal ions utilizing collophane. Minerals Engineering 10 (11), 1245–1251.

Gómez del Río, J.A., Morando, P.J., Cicerone, D.S., 2004. Natural materials for treatment of industrial effluents: comparative study of the retention of Cd, Zn and Co by calcite and hydroxyapatite. Part I: batch experiments. Journal of Environmental Management 71, 169–177.

Hizal, J., Apak, R., 2006. Modeling of copper(II) and lead(II) adsorption on kaolinite-based clay minerals individually and in the presence of humic acid. Journal of Colloid and Interface Science 295, 1–13.

Huang, P., Fuerstenau, D.W., 2001. The effect of the adsorption of lead and cadmium ions on the interfacial behavior of quartz and talc. Colloids and Surfaces A: Physicochemical and Engineering Aspects 177, 147–156.

Jeanjean, J., Fedoroff, M., Faverjon, F., Vincent, U., Corset, J., 1996. Influence of pH on the sorption of cadmium ions on calcium hydroxyapatite. Journal of Materials Science 31, 6156–6160.

Knox, A.S., Kaplanm, D.I., Paller, M.H., 2006. Phosphate sources and their suitability for remediation of contaminated soils. Science of the Total Environment 357, 271–279.

Lackovic, K., Angove, M.J., Wells, J.D., Johnson, B.B., 2004. Modeling the adsorption of Cd(II) onto goethite in the presence of citric acid. Journal of Colloid and Interface Science 269, 37–45.

Laperche, V., Traina, S.J., Gaddam, P., Logan, T.J., 1996. Chemical and mineralogical characterizations of Pb in a contaminated soil: reactions with synthetic apatite. Environmental Science and Technology 30, 3321–3326.

Laperche, V., Logan, T.J., Gaddam, P., Traina, S.J., 1997. Effect of apatite amendments on plant uptake of lead from contaminated soil. Environmental Science and Technology 31, 2745–2753.

Lee, Y.J., Elzinga, E., Reeder, R.J., 2005. Sorption mechanisms of zinc on hydroxyapatite: systematic uptake studies and EXAFS spectroscopy analysis. Environment Science and Technology 39, 4042–4048.

Liu, Y., Comodi, P., 1993. Some aspects of the crystal-chemistry of apatites. Mineralogical magazine 57, 709–719.

Ma, Q.Y., Logan, T.J., Traina, S.J., 1994. Effects of NO_3^-, Cl^-, F^-, SO_4^{2-} and CO_3^{2-} on Pb^{2+} immobilization by hydroxyapatite. Environment Science and Technology 28, 408–418.

Manecki, M., Maurice, P.A., Traina, S.J., 2000. Uptake of aqueous Pb by Cl^-, F^-, and OH^- apatites: mineralogical evidence for nucleation mechanisms. American Mineralogist 85, 932–942.

Martinez, M., Giménez, J., Pablo de, J., Rovira, M., Duro, L., 2004. Sorption of selenium(IV) and selenium(VI) onto magnetite. Applied Surface Science 252, 3767–3773.

Mavropoulos, E., Rossi, A.M., Costa, A.M., Perez, C.A.C., Moreira, J.C., Saldanha, M., 2002. Studies on the mechanisms of lead immobilization by hydroxyapatite. Environmental Science and Technology 36, 1625–1629.

Mavropoulos, E., Rocha, N.C.C., Moreira, J.C., Rossi, A.M., Soares, G.A., 2004. Characterization of phase evolution during lead immobilization by synthetic hydroxyapatite. Materials Characterization 53, 71–78.

Miyake, M., Watanabe, K., Nagayama, Y., Nagasawa, H., Suzuki, T., 1990. Synthetic carbonate apatites as inorganic cation exchangers: exchange characteristics for toxic ions. Journal of Chemical Society, Faraday Transaction 86 (12), 2303–2306.

Murugan, R., Ramakrishna, S., 2006. Production of ultra-fine bioresorbable carbonated hydroxyapatite. Acta Biomaterialia 2, 201–206.

Mustafa, G., Singh, B., Kookana, R.S., 2004. Cadmium adsorption and desorption behaviour on goethite at low equilibrium concentrations: effects of pH and index cations. Chemosphere 57, 1325–1333.

Flora, N.J., Yoder, C.H., Jenkins, H.D.B., 2004. Lattice energy of apatites and the estimation of $\Delta H_f(PO_4^{3-},g)$. Inorganic Chemistry 43, 2340–2345.

NEPC, 1998a. Final Impact Statement: National Environment Protection Measure for Ambient Air Quality. National Environment Protection Council, Canberra (Chapter 12).

Ören, A.H., Kaya, A., 2006. Factors affecting adsorption characteristics of Zn^{2+} on two natural zeolites. Journal of Hazardous Materials B 131, 59–65.

Prasad, M., Saxena, S., 2004. Sorption mechanism of some divalent metal ions onto low-cost mineral adsorbent. Industrial and Engineering Chemistry Research 43, 1512–1522.

Prasad, M., Saxena, S., Amritphale, S.S., Chandra, N., 2000. Kinetics and isotherms for aqueous lead adsorption by natural minerals. Industrial and Engineering Chemistry Research 39, 3034–3037.

Prasad, M., Saxena, S., Amritphale, S.S., 2002. Adsorption models for sorption of lead and zinc on francolite mineral. Industrial and Engineering Chemistry Research 41, 105–111.

Peld, M., Tõnsuaadu, K., Bender, V., 2004. Sorption and desorption of Cd^{2+} and Zn^{2+} ions in apatite-aqueous systems. Environment Science and Technology 38, 5626–5631.

Pitcher, S.K., Slade, R.C.T., Ward, N.I., 2004. Heavy metal removal from motorway stormwater using zeolites. Science of the Total Environment 334–335, 161–166.

Pohlmeier, A., Lustfeld, H., 2004. Reaction rates of heavy metal ions at goethite: relaxation experiments and modeling. Journal of Colloid and Interface Science 269, 131–142.

Raicevic, S., Kaludjerovic-Radoicic, T., Zouboulis, A.I., 2005. In situ stabilization of toxic metals in polluted soils using phosphates: theoretical prediction and experimental verification. Journal of Hazardous Materials B 117, 41–53.

Rouff, A.A., Reeder, R.J., Fisher, N.S., 2005. Electrolyte and pH effects on Pb(II)–calcite sorption processes: the role of the $PbCO^0_{3(aq)}$ complex. Journal of Colloid and Interface Science 286, 61–67.

Shashkova, I.L., RatKo, A.I., Kitikova, N.V., 1999. Removal of heavy metal ions from aqueous solutions by alkaline-earth metal phosphates. Colloids and Surfaces A: Physiochemical and Engineering Aspects 160, 207–215.

Smičiklas, I.D., Milonjić, S.K., Pfendt, P., Raičević, S., 2000. The point of zero charge and sorption of cadmium () and strontium () ions on synthetic hydroxyapatite. Separation and Purification Technology 18, 185–194.

Stipp, S.L.S., Hansen, M., Kristensen, R., Hochella Jr., M.F., Bennedsen, L., Dideriksen, K., Balic-Zunic, T., Léonard, D., Mathieu, H.-J., 2002. Behavior of Fe-oxides relevant to contaminant uptake in the environment. Chemical Geology 190, 321–337.

Sugyiama, S., Fukuda, N., Matsumoto, A., Hayashi, H., Shigemoto, N., Hiraga, Y., Moffat, J.B., 1999. Interdependence of anion and cation exchanges in calcium hydroxyapatite: Pb^{2+} and Cl^-. Journal of Colloid and Interface Science 220, 324–328.

Sun, S.Z., Lei, J.H., Wu, J.X., 1992. The synthesis of hydroxyapatite by sol–gel method. Inorganic Salt Industry (3), 24–25 (in Chinese).

Suzuki, T., Ishigaki, K., Miyake, M., 1984. Synthetic hydroxyapatites as inorganic cation exchangers, part 3: exchange characteristics of lead ions (Pb^{2+}). Journal of Chemical Society: Faraday Transaction 1, 3157–3165.

Tang, R.K., Henneman, Z.J., Nancollas, G.H., 2003. Constant composition kinetics study of carbonated apatite dissolution. Journal of Crystal Growth 249, 614–624.

Trivedi, P., Axe, L., 2001. Ni and Zn sorption to amorphous versus crystalline iron oxides: macroscopic studies. Journal of Colloid and Interface Science 244, 221–229.

USEPA, 1990. Toxicity characterisation leaching procedure (TCLP), EPA Method 1311, Washington, USA.

Xu, H., 2004. Study on the removal behavior and mechanism of aqueous Cd ion by apatites. Ph.D. Thesis, Department of Earth Sciences, Zhongshan University, Guangzhou (in Chinese).

Younes, M., 1995. The role of biomarkers in derivation of WHO guidance values for air pollutants. Toxicological Letters 77, 189–190.

Zhang, P., Ryan, J.A., Yang, J., 1998. In vitro Pb solubility in the presence of hydroxyapatite. Environmental Science and Technology 32, 2763–2768.

Continued from outside back cover

Journal of
Environmental *Management*

January 2008 Volume 86 Number 1

Contents

Continued on inside back cover

Abstracted/indexed in the abstract and citation database SCOPUS®. Full text available on ScienceDirect®.

Note to contributors. Online submission is available for this journal; to submit your article online, go to http://ees.elsevier.com/jema. Instructions on how to prepare a manuscript to the journal can be found in the Guide for Authors on the journal homepage at http://www.elsevier.com/locate/jenvman.

This journal is printed on acid free paper

0301-4797(200801)86:1;1-#