

Journal of
Environmental
Management

08

6.No.4

Journal of Environmental Management

A. GILL **Editor-in-Chief**
1620 Acton Street, Berkeley, CA 94702, USA.

Editorial Board

C. ALLAN	Department of Geography and Earth Sciences, University of North Carolina, Charlotte, NC 28227, USA email: cjallan@email.uncc.edu	G. HUANG	CRC Professor and Associate Dean, Faculty of Engineering, University of Regina, Regina, SK S4S 0A2, Canada email: gordon.huang@uregina.ca	J. P. RICHARDS	Institute of Marine Studies, Plymouth University, Drake Circus, Plymouth, Devon, PL4 8AA, UK. email: Jprichards@plymouth.ac.uk
N.-B. CHANG	Department of Civil and Environmental Engineering, University of Central Florida Orlando, FL 32708 email: nchang@mail.ucf.edu	J. MARION	USGS Patuxent Wildlife Research Center, Cooperative Park Studies Unit, Virginia Tech/Department of Forestry, 304 Cheatham Hall (0324) Blacksburg, VA 24061, USA. email: cpsu@vt.edu	M. RUTH	University of Maryland, School of Public Affairs, Van Munching Hall, College Park, MD 20742, USA. email: Matthias_Ruth@umail.umd.edu
R. CULLEN	Commerce Division, P.O. Box 84, Lincoln University, Canterbury, New Zealand. email: cullenr@lincoln.ac.nz	R. K. MORGAN	Department of Geography, University of Otago, P.O. Box 56, Dunedin, New Zealand. email: rkm@hyperperth.otago.ac.nz	H. SAARENMAA	Global Diversity Information Facility-GBIF Secretariat Universitetsparken 15, DK-2100 Copenhagen, Denmark. email: hsaarenmaa@gbif.org
L. G. FIRBANK	Land Use Section, Centre for Ecology and Hydrology, Lancaster Environment Centre, Bailrigg, Lancaster, LA1 4AP. email: lgf@ceh.ac.uk	R. A. PRESTON -WHYTE	School of Life and Environmental Sciences, University of Natal, Durban, South Africa	B. TANSEL	Florida International University, Department of Civil and Environmental Engineering, 10555 West Flagler Street, Engineering Center Miami, FL 33174 USA. email: tanselb@fiu.edu
A. GILBERT	IVM/VU De Boelelaan 1115, 1081 HV Amsterdam. email: alison.gilbert@ivm.vu.nl				

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

658.408 ELS



26869

Journal of environmental management.

CONTROL NUMBER	0000 22817
ACCESSION NUMBER	26869
DATE	29/04/08



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\$ 1223 (valid in North, Central and South America), including air speed delivery. Periodical postage paid at Rahway NJ and additional mailing offices.

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

©2008 Elsevier Ltd. All rights reserved.

This journal and the individual contributions contained in it are protected under copyright by Elsevier Ltd., and the following terms and conditions apply to their use:

Photocopying

Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for non-profit educational classroom use.

For information on how to seek permission visit www.elsevier.com/permissions or call: (+44) 1865 843830 (UK)/(+1) 215 239 3804 (USA).

Derivative Works

Subscribers may reproduce tables of content or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution. Permission of the Publisher is required for all other derivative works, including compilations and translations (please consult www.elsevier.com/permissions).

Electronic Storage or Usage

Permission of the Publisher is required to store or use electronically any material contained in this journal, including any article or part of an article (please consult www.elsevier.com/permissions).

Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the Publisher.

Notice

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

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>

Multivariate statistics to investigate metal contamination in surface soil

Ozan D. Yay^a, Omar Alagha^{b,*}, Gürdal Tuncel^c

^aDepartment of Environmental Engineering, Anadolu University, Eskişehir, Turkey

^bDepartment of Environmental Engineering, Fatih University, Istanbul, 34900, Turkey

^cDepartment of Environmental Engineering, Middle East Technical University, Ankara, 06531, Turkey

Received 17 May 2005; received in revised form 11 December 2006; accepted 11 December 2006

Available online 20 February 2007

Abstract

The modification of soil composition in the urbanized area of Ankara due to wet–dry deposition and pollution-derived particles from the atmosphere is investigated by analyzing 120 surface soil samples, collected from the urbanized area and its un-urbanized surrounding, for major, minor and trace elements. Concentrations of elements from human activity (e.g. Cd, Pb, Cr, Zn, Cu and Ca) in the urbanized area were higher than their corresponding concentrations in global average soil and soil in un-urbanized areas outside the urbanized area. Metal contents in soil were very high in densely populated districts and around some industrial facilities. The only exception was Pb distribution, which was more dispersed, due to the nature of motor vehicle emissions. Alteration of the Cd, Zn, Cu and Cr content of soil was confined to the inhabited and industrial areas, whereas enrichment factors of these elements were close to unity in the remaining study area. Factor analysis identified two polluted soil factor associations. One factor includes elements, such as Zn and Cd, which had high factor scores in inhabited areas and the other factor (high loading of Pb) represents soil polluted by motor vehicle emissions.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Urban soil pollution; Metal deposition; Atmospheric aerosols; Factor analysis

1. Introduction

Heavy metals in soil originate either from weathering of parent material and/or from numerous external contaminating sources (Fergusson and Kim, 1991). In unpolluted regions, parent materials are the primary source of trace elements. Human activities, such as non-point source agriculture activities, release of emissions from nearby point sources such as smelters, and traffic may also affect the chemical composition of soil (Davies and Wixson, 1987; Majdi and Persson, 1989; Haktanir et al., 1995). In urban areas, deposition of pollutants emitted to the atmosphere from point sources such as residential heating, industrial facilities and mobile sources such as traffic are the primary sources of soil pollution.

Trace elements are excellent indicators of the natural variability in soil composition (Huisman et al., 1997; Richard et al., 1999; Blaser et al., 2000; Jeffrey and Hudson, 2001), toxicology, exposure and health risks

associated with soil pollution (Tristan et al., 2000; Welp and Brummer, 1997), the effect of individual sources on soil composition (Landsberger et al., 1999; Hajdu and Licsko, 1999; Jensen et al., 2000; Sutherland et al., 2000) and the potential of organisms to be used as bioindicators of soil pollution (Palmer et al., 1998; Marc et al., 1999; Sandaa et al., 1999; Watmough, 1999).

Sources contributing to soil pollution in an urban/industrial area should be better known in order to control and minimize their effects on soil composition. However, the assessment of the contribution from numerous sources in a large metropolitan area can be a complex task. Multivariate statistical techniques, such as factor analysis (FA), are powerful tools to segregate sources contributing to observed pollution. These techniques have been used both to differentiate between different natural sources that cause variations in soil composition (Carrasco et al., 1991; Chen et al., 1991; Güvenç et al., 2003; Atgin et al., 2000) and to identify pollution sources affecting the metal content of soil (Davies and Wixson, 1987; Xhoffer et al., 1991; Dudka, 1992; Sweet et al., 1993; Davies 1997).

*Corresponding author. Tel.: +90 212 8890810; fax: +90 212 8890906.
E-mail address: oalagha@fatih.edu.tr (O. Alagha).

Pollution maps, which are produced by interpolating measured concentrations of pollutants, may also provide information on the relative contributions of different sources affecting soil composition (Huamain et al., 1999; Bloemen et al., 1995; Frangi and Richard, 1997; Weiss et al., 1994; Atgin et al., 2000; Güvenç et al., 2003).

Although the chemical composition of soil is affected by air pollutants through deposition of air-borne particles, the composition of atmospheric particles is also affected by soil, because re-suspended soil particles are an important component of aerosols, particularly in arid regions or in areas with high density of non-paved roads.

In this study, modification of the elemental composition of soil in the city of Ankara due to deposition of pollution-derived aerosols over many years is discussed. The starting point of the study was the absolute principal component analysis (APCA) performed on atmospheric aerosols in Ankara in 1993 by Yatin (1994) and Yatin et al. (2000). One of the aerosol components determined by FA included both lithophilic and anthropogenic elements, and was classified as polluted surface soil. The high portion of system variance explained by this soil factor suggests that surface soil composition in Ankara may have been significantly modified by deposition of pollution-derived particles from the atmosphere over the years.

Ankara, the capital of Turkey with a population of 4 million, suffered from severe air pollution until 1990. Air quality improved considerably after 1990 owing to the shift from coal to natural gas combustion for residential heating. Prior to 1990, the main source of air pollution in the city was the emissions from residential heating units. Residential heating units apparently are no longer a main source of pollution because of the switch to natural gas. The contribution from industry is not significant, because Ankara is the center of government in Turkey, and industrial activity is fairly limited. Emissions from individual heating units from at least half a million households when coupled with a very low wind speed (annual average 2 m s^{-1} , Yatin et al., 2000) resulted in elevated concentrations of gaseous and particulate pollutants in the atmosphere. Deposition of combustion particles containing anthropogenic elements which have low background concentrations in soil can significantly alter the elemental composition of the soil.

2. Experimental

2.1. Sampling

Soil samples were collected from an area of $\sim 150 \text{ km}^2$, which includes all inhabited and industrial parts of Ankara, but also extends to un-urbanized areas around the city. The study area was divided into $1 \times 1 \text{ km}$ cells and 120 soil samples were collected from every grid intersection. The sample points, distribution of population (DIE, 1999) and important industrial facilities are shown in Fig. 1.

Five soil sub-samples were collected at each grid node from the top 5 cm, and stored in acid-washed polyethylene bags. One of the samples was collected from the intersection point and the remaining four were collected from the four quadrants up to radius of 100 m from the central sampling point.

A composite sample was made in the laboratory by mixing, and thoroughly homogenizing equal amounts of the five sub-samples. If unusual concentrations of elements were observed in any of the homogenized samples, each of the five sub-samples collected from that sampling site was reanalyzed separately to ensure that the observed concentration was a common feature in the 100 m^2 area and not due to local contamination at one of the five sampling points.

To give an idea about the main compositions of the soil types of the study area, a study carried out by Elkarim and Usta (2000) on uncultivated land in the Polatlı area, 70 km from Ankara, was referred to. This area lies within the boundary area of Ankara province and its soil is representative of the study area's soil. Consequently, the Polatlı area soil can be used for comparison and useful to the current study. The soil composition was reported as: clay $36.3 \pm 1.50\%$, silt $39.3 \pm 2.30\%$, sand $24.3 \pm 2.31\%$, pH 7.51 ± 0.51 , CaCO_3 $10.0 \pm 6.20\%$, soil organic matter $1.28 \pm 0.33\%$ and Ammonia-N $5.57 \pm 2.85 \text{ ppm}$ (Elkarim and Usta, 2000).

2.2. Sample preparation

In the laboratory, large items in the soil, such as grass, pieces of wood, etc., were sieved out using a 2 mm sieve, and subsequently dried in a thermostatically controlled oven at 60°C for 48 h. No loss of low volatile element metals (e.g. Hg, As, Se and Sb) was observed when drying was preformed at 30 and 60°C . The percentage recovery of the analyzed elements in SRM materials is given in Section 3.1 in the Results and Discussion. Dried samples were ground using an agate mortar, equal amounts of sub-samples collected at each grid node (100 g) were mixed and the composite sample was homogenized using a RECHT agate homogenizer.

Fifty mg of each soil-homogenized sample was digested with 4 ml of HNO_3 and 1.5 ml HF for $\sim 90 \text{ min}$ at 150°C using PARR, Model 4746 high-pressure decomposition vessels. After digestion, HF was evaporated by placing the sample in the Teflon vessel on a hot plate at 80°C . The vessel was left under a fume hood until about 1 ml of the sample remained, and the residue was diluted to 50 ml. The HNO_3 used in the digestion was produced by sub-boiling-distillation of Merck laboratory grade acid, and the HF used was Merck Ultra-pure grade. Water used in dilution and washing had $17 \text{ M}\Omega$ resistance and was produced by deionization of double-distilled tap water.

A total of 120 digested soil samples were analyzed for 13 major, minor and trace elements, including Al, Ca, Cr, Cu, Fe, Mg, Mn, Zn, Na, K, Pb, Ni and Cd by atomic

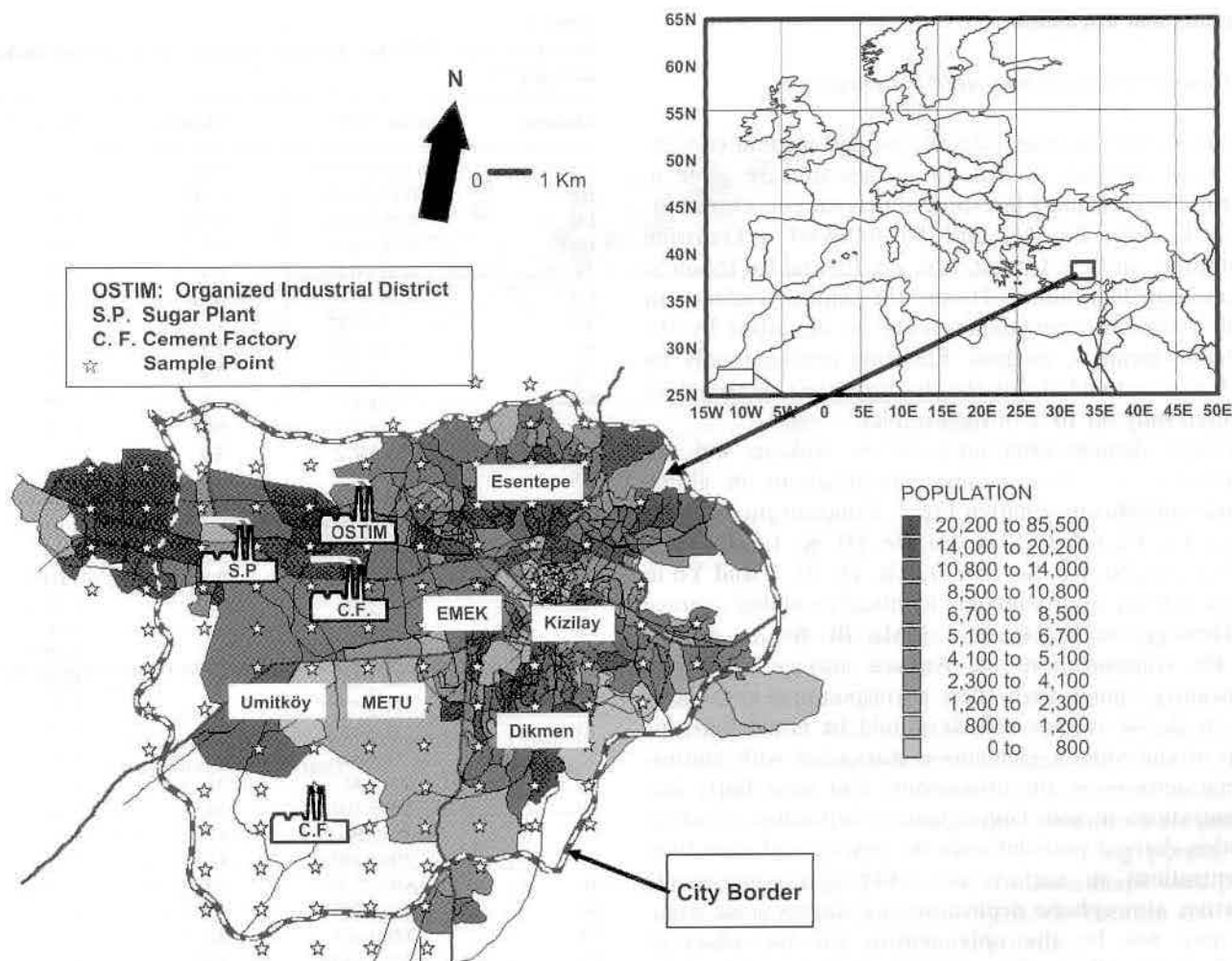


Fig. 1. Sample points, distribution of population and important industrial facilities in the study area.

absorption spectrometry (AAS) using a Perkin-Elmer Model 1100B spectrophotometer, and a Perkin-Elmer Model HGA 700 electrothermal atomization unit.

In addition to the AAS technique and for comparison purposes, 200 mg portions of 10 soil samples were also analyzed in parallel by instrumental neutron activation analysis (INAA) for 40 elements. These elements included: Na, Mg, Al, Cl, K, Sc, Ti, V, Cr, Mn, Fe, Co, Zn, Ga, As, Se, Br, Rb, Sr, Mo, Cd, In, Sb, Cs, Ba, La, Ce, Nd, Sm, Eu, Gd, Tb, Dy, Yb, Lu, Hf, Au, Hg, Th and U. The procedures are described by Olmez, 1989. The sample number for the INAA was not large enough for statistical analysis as in the case of the AAS, but it was useful for method comparison and information about the levels of other trace metals which were not possible to analyze with AAS. Briefly, samples were irradiated twice at the Massachusetts Institute of Technology with a 4.9 MW MITR-II Research Reactor at a thermal neutron flux of $8 \times 10^{12} \text{ n cm}^{-2} \text{ s}^{-1}$. The gamma ray spectrum, after each irradiation, was collected using CANBERRA HPGe detectors, and analyzed by CANBERRA ND900 Genie System software.

Laboratory blanks were routinely prepared and analyzed (one blank for every 5 samples) to monitor the possibility of sample contamination during digestion and subsequent analysis. Accuracy of the AAS measurements was checked by regular analysis of standard reference materials NIST, SRM 1646 (Estuarine sediment) and SRM-2704 (Buffalo River Sediment). Methods of standard additions were used in all AAS analysis.

The percentage recovery of the analyzed elements in SRM materials spanned between 96.9% for Zn and 111.1% for Ca. Typical values found were as follows: Al: 98.7%, Ca: 111.1%, Cd: 97.1%, Cr: 103.1%, Cu: 99.6%, Fe: 101.4%, K: 99.5%, Mg: 100.8%, Na: 101.3%, Ni: 106.5%, and Zn: 96.9%.

The percentage of blank contribution to the average concentration of the measured elements was evaluated and it was found to be: Al: 0.39%, Ca: 1.48%, Cd: 0.29%, Cr: 7.85%, Cu: 11.6%, Fe: 0.29%, K: 0.82%, Mg: 0.63%, Na: 3.78%, Ni: 1.02%, and Zn: 8.97%. It was found that there was no significant contribution from blank values to the measured concentrations in AAS analysis.

3. Results and discussion

3.1. Concentrations of measured parameters

Averages with standard deviations and median concentrations of elements in Ankara surface soil are given in Table 1. The chi-square test showed that the concentrations of Al, K, Mg, Fe, Na and Ni followed a Gaussian distribution, and Ca, Cr, Cu, Mn, Zn, Pb and Cd followed a lognormal distribution. Hence, the central tendency of sample's concentration data should be described by the arithmetic mean or median. Elements analyzed only by INAA were not included in the chi-square test, as they were measured only on 10% of the samples.

Average element concentrations in Ankara soil are compared with corresponding concentrations in global average soil (Mason, 1966) in Fig. 2. Concentrations of Al, Ba, Ca, Ce, Cl, Co, Cr, Dy, Eu, Fe, Hf, K, La, Lu, Mg, Mn, Na, Nd, Ni, Rb, Sc, Sm, Sr, Th, Tb, Ti, V and Yb in Ankara soil are approximately identical to global average soil. However, Se, In, Hg, Sb, Cd, Mo, Br, As, Zn, Au, Cu and Pb concentrations in Ankara surface soil were significantly higher than their corresponding concentrations in global average soil. It should be noted that this group of chalcophilic elements is associated with anthropogenic aerosols in the atmosphere and have fairly low concentrations in soil. Consequently, deposition of aerial pollution-derived particles over the years could alter their concentrations in surface soil. Although conceptually attractive, atmospheric deposition of anthropogenic particles may not be the only reason for the observed concentration differences between local and average soils, because such differences can also be explained by natural sources, such as a unique mineralogy in different parts of the study area.

To test this possibility, concentrations of elements measured in heavily polluted parts of the city were compared with corresponding concentrations in suburbs, where population is scarce and air pollution is low. If the observed differences are due to mineralogy, the high concentrations observed in Ankara soil should be fairly uniform throughout the study area.

Average concentrations of elements at urban and unurbanized sampling stations are given in Table 2. Unurbanized data were estimated by averaging data from 14 sites, which are located at the outskirts of the study area. Urban average data, on the other hand, were calculated from 9 sites, which are located in the most heavily populated districts of the city center around Kizilay area.

Concentrations of Cd, Pb, Cr, Zn, Cu and Ca were higher by a factor of 1.9–3.7 in urban stations, suggesting that the observed differences were not due to a local geologic condition. Concentrations of Mn, K, Ni, Na, Al, Mg and Fe, on the other hand, were comparable in both un-urbanized and urban parts of the study area with a ratio close to 1.0. The *t*-test applied to Cd, Pb, Cr, Zn, Cu and Ca showed that the differences are significant at the 95%

Table 1

Arithmetic mean, RSD and median concentrations of elements in Ankara soil ($\mu\text{g g}^{-1}$)

Elements	Mean $\pm$ SD	Median	Mean/Median
Au	0.047 $\pm$ 0.059	0.02	1.83
Hg	0.17 $\pm$ 0.21	0.06	0.89
Lu	0.28 $\pm$ 0.15	0.32	1.52
Eu	0.71 $\pm$ 0.16	0.7	0.94
Tb	0.56 $\pm$ 0.43	0.4	8.39
Cd	0.58 $\pm$ 0.67	0.41	1.14
Yb	1.5 $\pm$ 0.87	1.7	0.93
Sb	1.7 $\pm$ 1.6	1.2	1.10
Dy	3.5 $\pm$ 1.5	3.8	1.35
Sm	3.4 $\pm$ 1.7	3.6	1.09
In	1.8 $\pm$ 3.7	0.17	1.12
Hf	3.7 $\pm$ 2.2	4.0	0.87
Th	7.0 $\pm$ 3.3	7.6	1.22
Mo	5.2 $\pm$ 5.2	3.0	3.85
Se	7.7 $\pm$ 3.7	7.7	1.92
Co	10 $\pm$ 3	12	1.15
Sc	6.7 $\pm$ 7.8	2.8	1.00
As	13 $\pm$ 18	7.1	1.06
La	25 $\pm$ 13	28	1.06
Nd	35 $\pm$ 33	23	1.07
Ce	46 $\pm$ 25	49	0.99
Br	26 $\pm$ 52	3.1	1.06
Rb	58 $\pm$ 44	51	1.15
V	74 $\pm$ 34	80	1.83
Ni	78 $\pm$ 45	71	0.89
Cr	100 $\pm$ 100	74	1.52
Cl	190 $\pm$ 94	175	0.94
Zn	200 $\pm$ 180	179	8.39
Ba	460 $\pm$ 160	530	1.14
Sr	450 $\pm$ 320	370	0.93
Pb	158 $\pm$ 617	41	1.10
Cu	250 $\pm$ 680	130	1.35
Mn	770 $\pm$ 460	670	1.09
Ti	2,800 $\pm$ 1,600	2,800	1.12
Na	8,700 $\pm$ 4,500	8,200	0.87
Mg	15,000 $\pm$ 58,000	14,100	1.22
K	13,000 $\pm$ 11,000	12,100	3.85
Al	57,000 $\pm$ 13,000	57,300	1.92
Fe	87,900 $\pm$ 35,000	83,000	1.15
Ca	66,900 $\pm$ 60,600	58,000	1.00

confidence level. This test was not applied to the remaining elements, because of the insufficient number of samples which were analyzed by INAA.

Concentrations of elements that have anthropogenic sources in soil are expected to be highly variable from one sampling point to another, because their concentrations in the study area depend on the distances between sampling points and pollution sources. Owing to such a strong variability in their concentrations, these elements should have high relative standard deviations, skewed frequency distributions and high arithmetic mean-to-median ratios. Standard deviations (SD) and average-to-median ratios of elements are displayed in Fig. 3. The elements Au, Mo, Cd, Se, Br, Pb, Hg, Cr, Zn, Cu, Ca, In, Sb and As have the highest SD's and average-to-median ratios as expected, and the chi-square test has shown that they are lognormally distributed within 95% confidence.

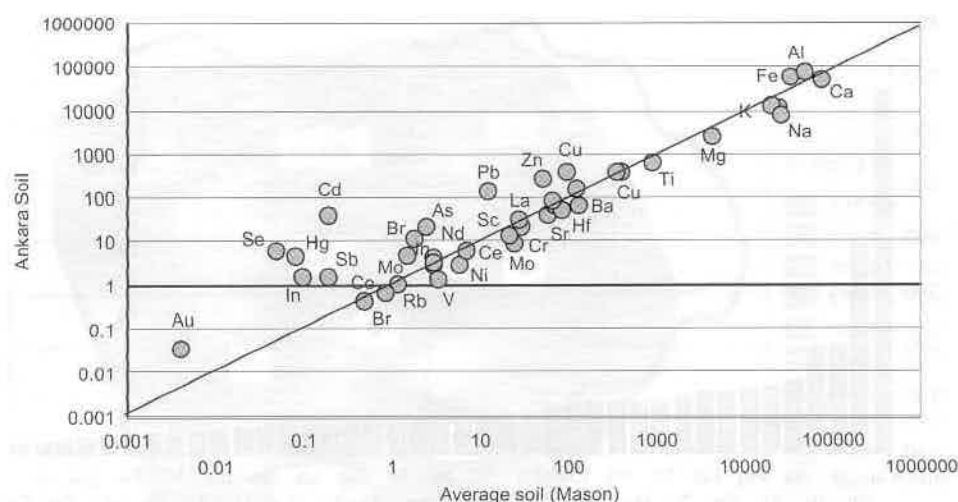


Fig. 2. Comparison of average concentrations of measured elements in Ankara soil and global average soil.

Table 2
Average concentrations of elements at un-urbanized and urban heavily contaminated sites

Element	Un-urbanized average ($\mu\text{g g}^{-1}$)	Urban average ($\mu\text{g g}^{-1}$)	Urban/un-urbanized ratio
Al	63,500 $\pm$ 11,600	63,600 $\pm$ 12,400	1.0
Ca	48,000 $\pm$ 65,000	120,000 $\pm$ 122,000	2.5
Cd	1.16 $\pm$ 0.21	2.15 $\pm$ 1.53	1.9
Cr	87 $\pm$ 59	284 $\pm$ 405	3.3
Cu	36 $\pm$ 17	98 $\pm$ 76	2.7
Fe	90,200 $\pm$ 22,700	85,800 $\pm$ 18,900	1.0
K	11,400 $\pm$ 7,000	13,200 $\pm$ 5,200	1.2
Mg	13,600 $\pm$ 4,500	13,200 $\pm$ 3,300	1.0
Mn	773 $\pm$ 329	1,100 $\pm$ 1,300	1.4
Na	8,700 $\pm$ 4,400	8,900 $\pm$ 3,600	1.0
Ni	71 $\pm$ 49	81 $\pm$ 49	1.1
Pb	51 $\pm$ 82	189 $\pm$ 54	3.7
Zn	65 $\pm$ 28	181 $\pm$ 138	2.8

The remaining elements, Lu, Mn, Eu, Yb, Sc, Th, V, K, Ni, Sm, Na, Al, Nd, Mg, Fe, Ba, Ce, Ti and La have low SDs indicating that their concentrations were fairly uniform throughout the city. They have similar arithmetic mean and median concentrations and all follow the Gaussian distribution within the 95% confidence level.

Higher concentrations of Se, In, Hg, Sb, Cd, Mo, Br, As, Cr, Ca, Cu, Zn, Au and Pb in the urban data set compared to average global soil and their high coefficients of variation imply that these elements are enriched in Ankara surface soil due to aerial deposition of anthropogenic particles over the years.

3.2. Spatial distribution of pollution

The spatial distribution of element concentrations is a useful visual aid to assess the possible sources of

enrichment. If elements are enriched in surface soil due to deposition of pollution aerosols from residential heating, their distribution should follow the trend of population distribution in Ankara. However, if they are enriched due to emissions from industrial sources, their concentrations are expected to be high around those point sources.

Distributions of elemental concentrations were prepared by interpolating data from 120 sampling points, using MapInfo software (1998). Distribution maps were plotted only for elements analyzed by AAS, because the INAA data were insufficient for interpolation.

Distribution patterns of lithophilic elements were significantly different from those of elements related to anthropogenic activities. The distribution of Al (Fig. 4), a typical lithophile, was similar to Mn, V, K, Ni, Na, Mg and Fe. For this group, concentrations were fairly uniformly distributed, and different rock types can explain small variations within the study area. Furthermore, small RSD values and low average-to-median ratios indicate the uniformity of distribution of these elements.

Distributions of elements affected by pollution show more relevant spatial patterns. General features in the distribution of pollution-derived elements are given in Fig. 5 for Cd and Cr. Similar patterns were observed for Cu and Zn.

The distributions of anthropogenic elements have some common features. Their concentrations were high in heavily populated districts of the city, such as Dikmen, Kizilay and Emek (see Fig. 1) and fairly low at areas with low population density. This suggests that emissions from residential heating were the dominant source for the enrichment of these elements in surface soil. These findings agree with the results of previous aerosol and precipitation studies in Ankara where residential heating was also proposed as the primary source of anthropogenic elements in the Ankara atmosphere (Kaya and Tuncel, 1997; Yatın et al., 2000).

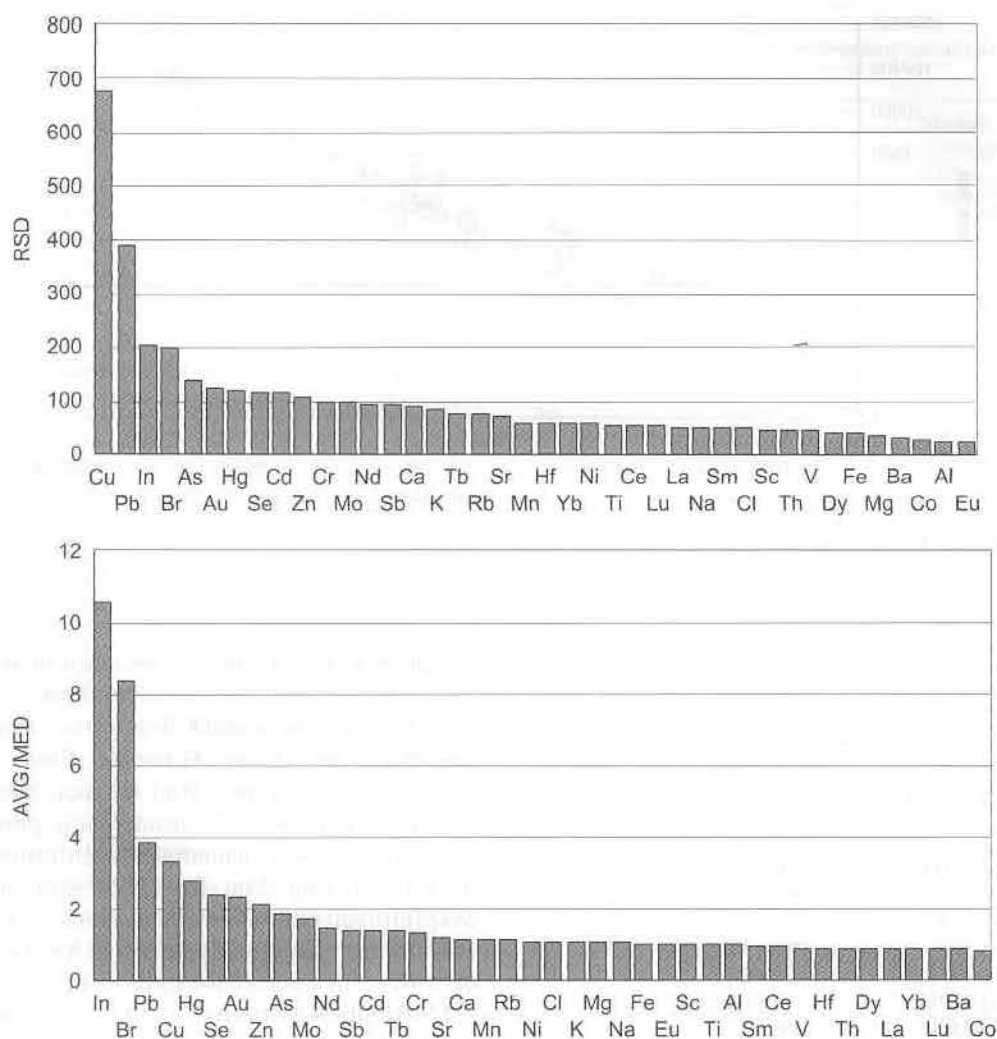


Fig. 3. Relative standard deviations and average-to-median ratios of elements in Ankara soil (given in descending order).

In addition to these features, which were observed for all pollution-derived elements, two hot spots were identified where concentrations of most of the anthropogenic elements were high. These two spots cannot be related to population, because they are located at sparsely populated parts of the urbanized area. One of these points is located at the northwest part of Ankara, around a large sugar plant industrial site. The other is at the Ostim Organized Industrial Area, which has many small-scale industrial sites. Obviously, deposition of particles emitted from industries causes elevated concentrations of pollution-derived elements in the soil, but their influence is limited to areas in the immediate vicinity of these facilities.

Distribution of Pb is given in Fig. 4. Unlike the distribution of other pollution-derived elements, high Pb concentrations (compared to suburban sites) were observed in a large part of the study area. Generally, the distribution of high Pb concentrations trends to follow routes of motor vehicle emissions in densely populated areas (Fig. 1). Furthermore, the high Pb concentration was located near the high traffic main roads crossing Ankara city.

3.3. Crustal enrichment factors

Although distribution patterns are useful in determining the variability of human influence on soil composition, they do not provide information on the extent of perturbation. Crustal enrichment factors (EF_c) of elements were calculated using simple spread sheet calculations using the equation given below. The following relation is a good indicator of the extent of soil modification.

$$EF_c = (C_x/C_{Al})_{\text{sample}} / (C_x/C_{Al})_{\text{average soil}}$$

where $(C_x/C_{Al})_{\text{sample}}$ is the ratio of the concentration of a test element to that of Al and $(C_x/C_{Al})_{\text{average soil}}$ is the same ratio with respect to a reference soil. Mason's (1966) global average soil was used as a reference soil, and Al was used as the crustal reference element in EF_c calculations because it is one of the largest components of soil and because of the ease of determination of this element. so, the modification of Al concentration in soil by other anthropogenic sources is difficult. Calculated EF_c 's of elements were then converted to soil modification maps by interpolation, using Map Info

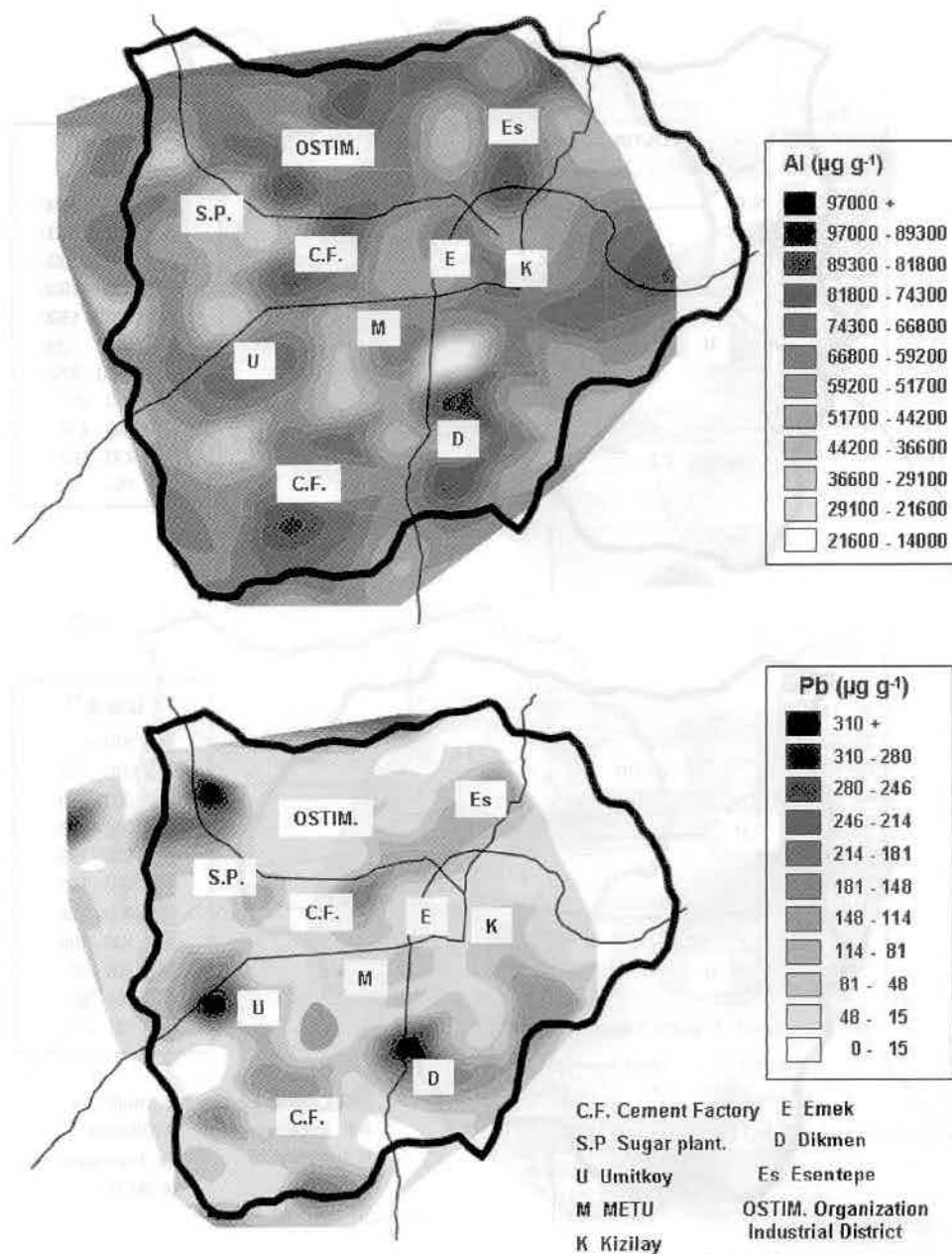


Fig. 4. Distribution of Al and Pb concentration in Ankara soil.

geographic information system software version 4.1. Modification of soil composition is assumed significant when the EF_c value of an element is >10.0 . Smaller enrichments can be due to differences between the compositions of the local soil and the reference soil used in these calculations.

For Cd (as well as Cu and Zn, which are not shown in Fig. 6), the EF_c 's are less than 5.0, which indicates that the soil's composition is not significantly altered by anthropogenic sources in most of the study area. However, fairly high enrichment factors (as high as 50) were observed in densely populated districts of the city (Emek and cement factory), indicating that soil composition has been significantly altered by atmospheric deposition in these regions.

Although the Pb content of the soil has also been modified in populated areas and around industries, high Pb EF_c 's cover a fairly large part of the study area, suggesting extensive soil modification even at unpopulated locations, while no direct link with population density is apparent. Results are displayed in Fig. 6 for Pb and Cd.

The extent of soil modification was also investigated by comparing measured concentrations of elements with predefined pollution classes. There are no well-defined limits to differentiate between clean and polluted soil. Existing criteria are developed for pollutant emissions at a small area where pollutants have gradually accumulated in soil by deposition from the atmosphere. We have

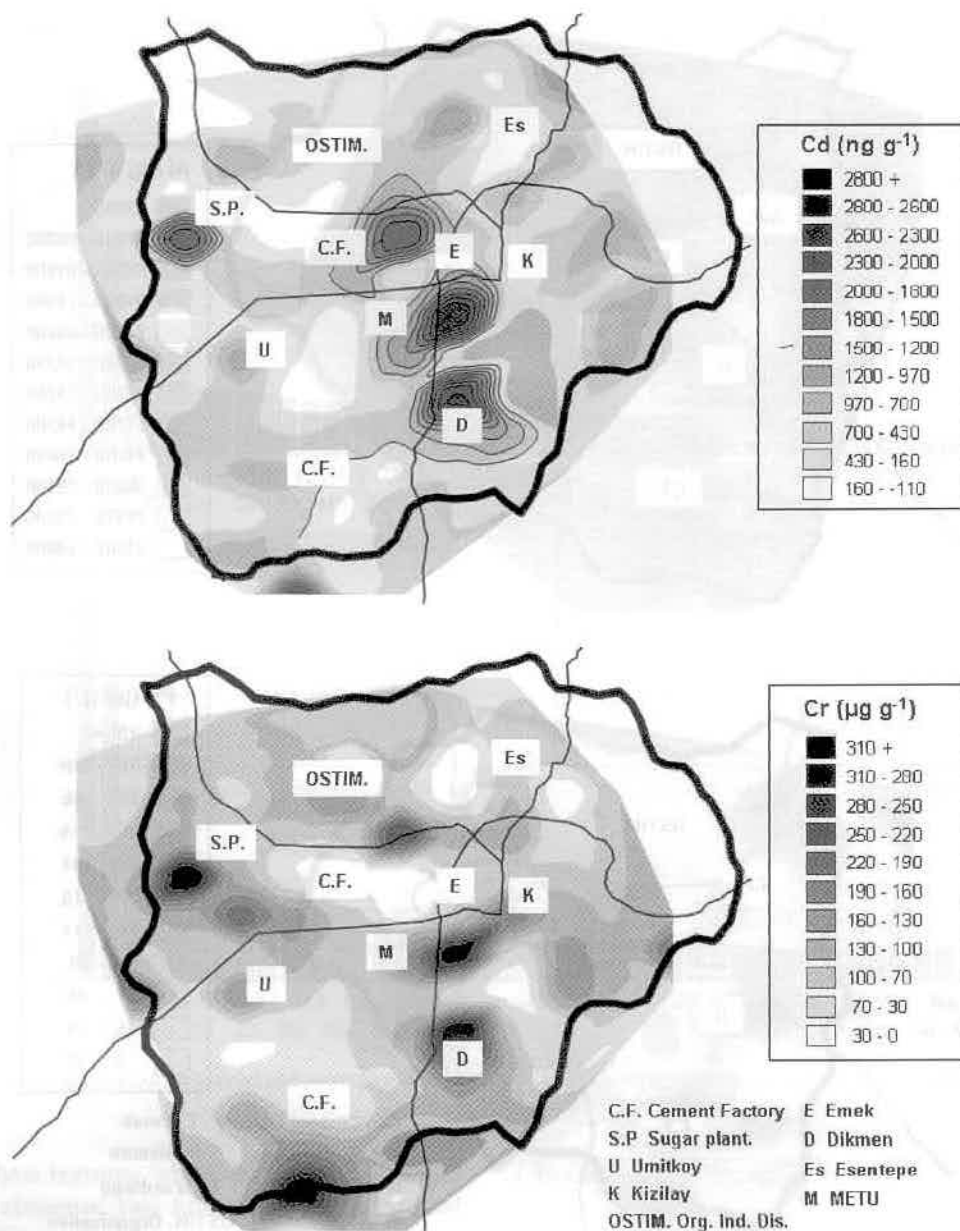


Fig. 5. Distribution of Cd and Cr concentrations in Ankara soil.

developed our own criteria (Table 3), using data from the literature for polluted and clean soils.

Results obtained by comparing concentrations of elements with the limits given in Table 3 are similar to the conclusions reached regarding EF_c distributions. For the elements Cd, Zn, Cu and Cr, soil in most parts of the study area was unpolluted, but soil was “polluted” or “heavily polluted” in districts with high population densities and in areas close to the organized industrial district.

Classification of soil for Pb pollution was different. Soil in a large part of the study area was classified as “polluted” by Pb. Soil was strongly polluted in settlement areas and around major highways.

3.4. Relation between aerosol and soil composition

Detection of a soil contribution in atmospheric aerosols in Ankara, which were enriched with pollution-derived elements, was the starting point of this study (Yatın, 1994; Yatın et al., 2000). Since now it is clear that the composition of surface soil in Ankara has been modified by deposition of anthropogenic particles from the atmosphere, the crustal enrichment factors (EF_c) calculated by Yatın (1994) using Mason's (1966) average soil composition probably overestimate enrichments of chalcophilic elements in the atmosphere.

Enrichment factors of elements were recalculated using both local soil and Mason's (1966) average crustal material

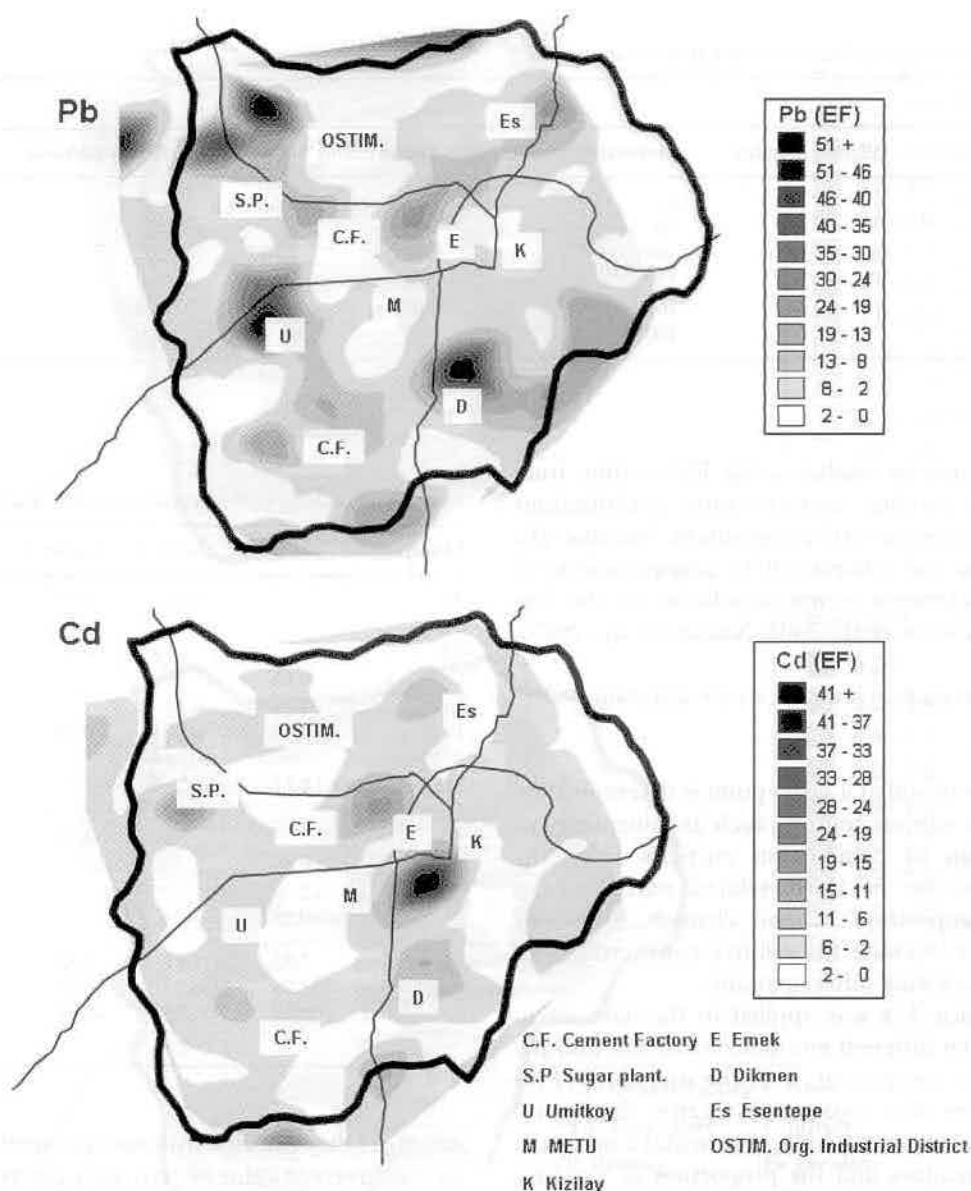


Fig. 6. Distribution of enrichment factors of Pb and Cd.

as a reference. Results for both coarse and fine aerosol fractions collected and analyzed by Yatin (1994) are shown in Fig. 8. Crustal enrichment factors of the elements Cd, Zn, Hg, As, Se, Br and Sb were clearly overestimated in both coarse and fine fractions in the previous studies (Yatin, 1994; Yatin et al., 2000), where Mason's (1966) average soil was used as a reference. This group of elements is highly enriched in fine fraction aerosols even when local soil composition is used in the calculations, thus suggesting that the observed enrichment of elements in the fine aerosol fraction is primarily due to pollution-derived particles in the atmosphere as suggested by Yatin (1994). However, Cd, Zn, As, Se and Sb are not significantly enriched in coarse fraction aerosols when local soil composition is used in the EF_c calculations. Obviously, the observed enrichment of these chalcophilic elements, which was attributed to the attachment of fine anthropogenic particles to coarse

particles in the atmosphere, is due to recycling of polluted surface soil. These observations indicate that EF_c values of elements in urban aerosols have to be evaluated cautiously, because observed enrichment, particularly in coarse fraction aerosols, can be due to recycled polluted surface soil, rather than to contributions of anthropogenic emissions to aerosol mass.

Because source apportionment techniques used in air pollution studies are sensitive to compositions of particles emitted from different sources, the use of the correct soil composition is important in determining the correct source contributions. This is particularly important if whole-filter (particles collected without PM_{10} cut-off) or PM_{10} aerosol (coarse fraction) samples are used in apportionment studies, because whole filter samples include most of the coarse fraction aerosols, where concentrations of anthropogenic elements are determined by recycled surface soil.

Table 3
Pollution classes for anthropogenic elements (concentrations in $\mu\text{g/g}^{-1}$)

Elements	*Pollution class					
	Unpolluted	Slightly polluted	Moderately polluted	Considerably polluted	Heavily polluted	Extremely polluted
Cd	0.3	1	2	3	5	>5
Cu	15	30	50	80	300	>300
Ni	10	30	100	200	400	>400
Pb	30	70	100	500	2500	>2500
Zn	50	100	300	700	3000	>3000
Cr	30	70	100	300	1000	>1000

*Duce et al. (1975).

However, apportionment studies using $\text{PM}_{2.5}$ (fine fraction) particles can produce correct source contributions even with global average soil composition, because the contribution of the soil component to concentrations of pollution-derived elements is not significant in the fine aerosol fraction (Karaca et al., 2005; Samura et al., 2003).

3.5. Factor analysis and origin of elements affecting soil composition

The composition of soil at a given point is determined by contributions from various sources, such as mineralogy of the soil, deposition of combustion particles from the atmosphere, deposition of traffic-related particles, etc. The elemental composition of soil changes from one location to another, because the relative contributions of each of these sources may differ spatially.

A varimax-rotation FA was applied to the data set to differentiate between different components of soil and the distribution of these components over the study area. Five interpretable factors with eigenvalues higher than unity were extracted (Güvenç et al., 2003). Elemental loadings in these factors, eigenvalues and the proportion of variance explained are given in Table 4. The factor loadings or component loadings are the correlation coefficients between the variables (rows) and factors (columns). The squared factor loading is the percent of variance in that variable explained by the factor. Factor loadings >0.25 are regarded as significant in the interpretation of the data.

Factors 2 and 3 have high positive loadings in only lithophilic elements suggesting that they represent unpolluted soil components probably contributed by their mineralogy.

The remaining three factors include pollution-derived elements. Factor 1, which explains 23% of the system variance, has high positive factor loadings on Cr, Ni and Mg, and moderately positive to low loadings on Fe, Cu and Zn. The presence of elements with crustal and anthropogenic origins in the same factor suggests that it represents polluted surface soil.

Factor scores in FA represent the score of each element on the various factors for every sample. Although it is not possible to obtain quantitative information from factor

Table 4
Factor analysis with rotated varimax scores for measured elements

Elements	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Al		0.42	0.58		
Ca				0.26	
Cd				0.73	0.37
Cr	0.82				
Cu	0.28	0.41		0.53	
Fe	0.34	0.67	0.39		
K		0.68			0.34
Mg	0.65	0.21			0.25
Mn		0.60			
Na			0.75		
Ni	0.86				
Pb					0.83
Zn	0.25			0.74	
Eigenvalue	3.03	1.85	1.43	1.10	1.09
% Variance	23.28	14.25	10.99	8.43	8.42
Cumulative	23.28	37.52	48.51	56.95	65.37
% Variance					

scores, as they are dimensionless parameters, normalized to have an average value of zero, they can be useful to identify or to confirm components represented by the factors. The distribution map of factor 1 scores does not resemble the distributions of Cr, Ni, Zn and Cu concentrations discussed previously in Fig. 7. The highest scores are observed at the outskirts of the city where population density is fairly low. Mean crustal enrichment factors of elements in samples that correspond to the highest factor scores over 10 are 1.6 (range 0.9–3.8) for Cr, 1.2 for Cu (ranges between 1.0 and 3.0), 2.2 (range 0.9–4.9) for Zn and 2.0 (range 0.8–4.0) for Ni. Such low EF_c 's and fairly uniform distribution of factor scores indicate that factor 1 is not related to pollution, but represents a certain mineral type of soil which is enriched with Cr and Ni (and to a smaller extent with Cu and Zn) Fig. 8.

Factor 4 is another anthropogenic factor, with high loadings on Zn and Cd and moderate loadings on Cu. Factor 4 scores are very high in the densely populated center of the city and low at un-urbanized areas (Fig. 7). Mean crustal enrichment factors of elements at sampling points that correspond to factor scores over 10 are 8.4

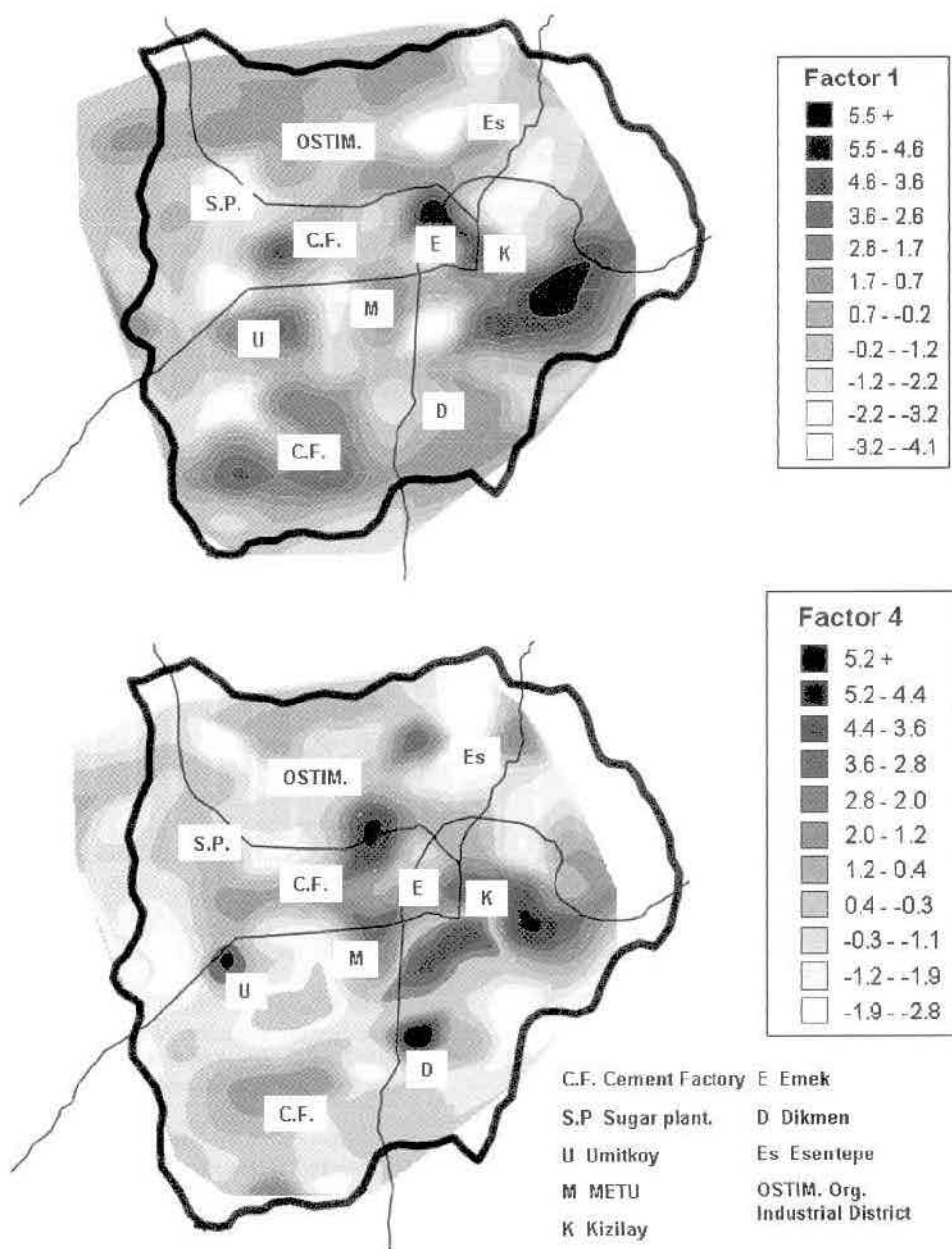


Fig. 7. Factor scores distribution of Factors 1 and 4.

(range 1.1–26), 7.2 (range 1.0–20) and 15.9 (ranges between 1.1 and 56) for Zn, Cu and Cd, respectively. This factor is interpreted as the pollution surface soil factor and represents soil in settlement areas which is polluted by atmospheric deposition of anthropogenic particles. Until recently, Ankara city was dependent on coal for domestic heating, which caused huge atmospheric problems due to the topography and meteorology of this city (Yatin, 1994).

A similar component containing both crustal and anthropogenic elements was also identified in Ankara atmospheric aerosols and attributed to the presence of polluted surface soil components among the collected aerosol population (Yatin, 1994).

Factor 5 has high positive loading on Pb. It most likely represents the component in soil polluted by motor vehicle

emissions. Weak positive loading on Cd possibly explains deposition of particles from worn tires, because Cd is used in the vulcanization process during tire production (Sinha and Banerjee, 1997; Schauer et al., 1996; Anderberg and Stigliani, 1994). Distribution of factor 5 scores in the study area closely resembles that of Pb concentrations, which were discussed previously (Güvenç et al., 2003).

4. Conclusions

The starting point of this study was the detection of a component in atmospheric aerosols in Ankara that appears to be polluted soil. Modification of the chemical composition of soil in Ankara was investigated by analyzing surface soil samples, collected from 120 points in the city and its

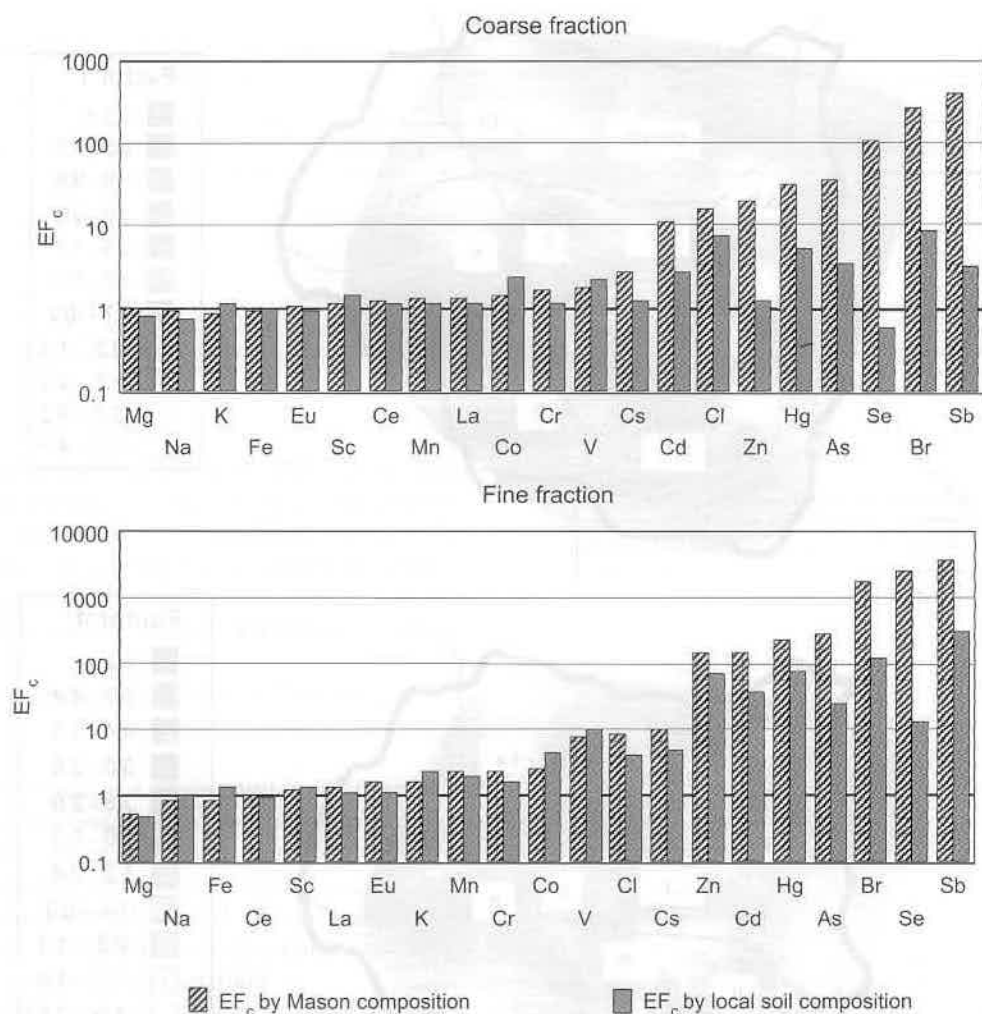


Fig. 8. Coarse and fine fraction of atmospheric aerosol enrichment factors calculated using local (Urban) and average Mason's soils as reference (Al used as soil reference).

surroundings, for major, minor and trace elements using AAS and INAA techniques.

Comparison of measured concentrations of elements in the city with global average soil composition and average measured compositions in un-urbanized stations has demonstrated that concentrations of Cd, Zn, Cu and Pb have been modified by deposition of anthropogenic particles over the years. Distribution maps of their concentrations have shown that modification of soil composition is most pronounced in districts with high population density.

The extent of the influence of atmospheric deposition on soil composition in different parts of the city was examined using crustal enrichment factors of elements. Soil composition has been considerably modified in parts of the study area where population density is high, and close to industrial facilities. However, soil composition is fairly similar to global average soil in scarcely populated areas, which cover most of the study area. Lead was enriched in a larger part of the study area owing to the aerial nature of motor vehicle emissions.

FA revealed five soil components with different compositions. Three of these components, which explained most of the system variance, were ascribed to unpolluted surface soil with different mineralogy. Two soil components were identified as the polluted soil fraction. One of these has high positive loadings on chalcophilic elements with high factor scores in densely populated districts of the city. The other one has high positive loadings on only Pb, representing soil polluted by motor vehicle emissions.

Although enrichment of elements in both coarse and fine fraction aerosols is partly due to the contribution of recycled polluted surface soil, the contribution of the fine soil fraction to aerosol enrichment of elements is not large enough to change the conclusions made by Yatin (1994). However, enrichment of chalcophilic elements in coarse fraction aerosols can be attributed to recycled surface soil. Since source apportionment of air pollution is sensitive to composition of particles emitted from different sources, the use of the correct soil composition is important, particularly if whole-filter or PM₁₀ aerosol samples are used in apportionment studies. Apportionment studies using PM_{2.5}

particles produce better source contribution results even with global average soil composition, which is regularly used in such studies.

Acknowledgment

We thank Ramazan Demir, Duran Karakas for their assistance in the laboratory analysis of samples, and İlhan Olmez of MIT for his assistance with INAA.

References

- Anderberg, S., Stigliani, W.M., 1994. Integrated approach for identifying sources of pollution: the example of cadmium pollution in the Rhine River Basin. *Water Science and Technology* 29, 61–67.
- Atgün, R.S., El-agha, O., Zararsiz, A., Kocatas, A., Parlak, H., Tuncel, G., 2000. Investigation of the sediments of İzmir Bay: trace elements. *Spectra Chimica Acta B*, 1151–1164.
- Blaser, P., Zimmermann, S., Luster, J., Shotyk, W., 2000. Critical examination of trace element enrichments and depletions in soils: As, Cr, Cu, Ni, Pb, Zn in Swiss forest soils. *The Science of the Total Environment* 249, 257–280.
- Bloemen, M.L., Markert, B., Lieth, H., 1995. The distribution of Cd, Cu, Pb and Zn in topsoils of Osnabrück in relation to land use. *The Science of the Total Environment* 166, 137–148.
- Carrasco, M., Prendez, A., Margarita, G., 1991. Element distribution of some soils of Continental Chile and the Antarctic Peninsula, projection to atmospheric pollution. *Water, Air, and Soil Pollution* 57, 713–722.
- Chen, J.W., Zheng, F., Wu, C., Adriano, Y., Domy, C., 1991. Background concentrations of elements in soils of China. *Water, Air, and Soil Pollution* 57, 699–712.
- Davies, B.E., Wixson, B.G., 1987. Use of factor analysis to differentiate pollutants from other trace metals in surface soils of the mineralized area of Madison County, Missouri, USA. *Water, Air, and Soil Pollution* 33, 339–345.
- Davies, B.E., 1997. Heavy metal contaminated soils in an old industrial area of Wales, Great Britain: source identification through statistical data interpretation. *Water, Air, and Soil Pollution* 94, 85–98.
- DİE, 1999. Monthly Statistics Bulletin. Government Statistics Institute, Ankara.
- Duce, R.A., Hoffman, G.L., Zoller, W.H., 1975. Atmospheric trace elements at remote northern and southern hemisphere sites. *pollution or natural. Science* 187, 59–61.
- Dudka, S., 1992. Factor-analysis of total element concentrations in surface soils of Poland. *The Science of the Total Environment* 121, 39–48.
- Elkarim, A. H., Usta, S., 2000. Assessment of potentially available soil Nitrogen by incubation methods as related to status of some soil-N forms. A, International Symposium on Desertification, Konya-Turkey <http://www.toprak.org.tr/isd/isd_26.htm>.
- Fergusson, J.E., Kim, N.D., 1991. Trace-elements in street and house dusts—sources and speciation. *The Science of the Total Environment* 100, 125–150.
- Frangi, J.P., Richard, D., 1997. Heavy metal soil pollution cartography in northern France. *The Science of the Total Environment* 205, 71–79.
- Güvenç, N., Alagha, O., Tuncel, G., 2003. Investigation of soil multi-element composition in Antalya, Turkey. *Environment International* 1055, 1–10.
- Hajdu, C., Liesko, I., 1999. Environmental pollution caused by a non-ferrous metal smelter. *Water Science and Technology* 39, 165–168.
- Haktanir, K.A., Erpul, S., Tan, G., 1995. Accumulation of traffic related heavy metals around highways. *Tr. Journal of Engineering and Environmental Sciences* 19, 423–431 (in Turkish).
- Huamain, C., Chunrong, Z., Cong, T., Yongguan, Z., 1999. Heavy metal pollution in soils in China: status and countermeasures. *Ambio* 28, 130–134.
- Huisman, J., Vermeulen, F.J.H., Baker, J., Veldkamp, A., Kroonenberg, S.B., Klaver, G.T., 1997. A geological interpretation of heavy metal concentrations in soils and sediments in the southern Netherlands. *Journal of Geochemical Exploration* 59, 163–174.
- Jeffrey, B.R., Hudson, G., 2001. A flexible methodology for the characterization of soils: a case study of the heavy metal status of a site at Dornach. *The Science of the Total Environment* 264, 153–162.
- Jensen, D.L., Holm, P.E., Christensen, T.H., 2000. Soil and groundwater contamination with heavy metals at two scrap iron and metal recycling facilities. *Waste Management and Research* 18, 52–63.
- Karaca, F., Alagha, O., Ertürk, F., 2005. Statistical characterization of atmospheric PM₁₀ and PM_{2.5} concentrations at a non-impacted suburban site of Istanbul, Turkey. *Chemosphere* 59 (8), 1183–1190.
- Kaya, G., Tuncel, G., 1997. Trace element and major ion composition of wet and dry deposition in Ankara, Turkey. *Atmospheric Environment* 31, 3985–3998.
- Landsberger, S., Iskander, F., Basunia, S., Barnes, D., Kaminski, M., 1999. Lead and copper contamination of soil from industrial activities and firing ranges. *Biological Trace Element Research* 71, 387–396.
- Majdi, H.S., Persson, H., 1989. Effects of road-traffic pollutants (lead and cadmium) on tree fine-roots along a motor road. *Plant and Soil* 119, 1–5.
- MapInfo software, version 4.0, 1998.
- Marc, P., Canard, A., Ysnel, F., 1999. Spiders (Araneae) Useful for pest limitation and bioindication. *Agriculture, Ecosystems & Environment* 74, 229–273.
- Mason, B., 1966. *Principles of Geochemistry*, third ed. Wiley, New York.
- Olmez, I., 1989. Instrumental neutron activation analysis of atmospheric particulate matter. In: Lodge, J.P. (Ed.), *Methods of Air Sampling and Analysis*. Springer, Berlin, pp. 121–132.
- Palmer, G., McFadzean, R., Killham, K., Sinclair, A., Paton, G.I., 1998. Use of lux-based biosensors for rapid diagnosis of pollutants in arable soils. *Chemosphere* 36, 2683–2697.
- Richard, B., Maja, L.B., Ingemar, R., Ove, E., Harald, G., 1999. Natural lead concentrations in pristine boreal forest soils and past pollution trends: a reference for critical load models. *Environmental Science and Technology* 33, 3362–3367.
- Samura, A., Omar Al-Agha, O., Tuncel, S.G., 2003. Study of trace and heavy metals in rural and urban aerosols of Uludağ and Bursa (Turkey). *Water, Air, and Soil Pollution* 3 (5), 109.
- Sandaa, R.A., Enger, O., Torsvik, V., 1999. Abundance and diversity of Archaea in heavy-metal-contaminated soils. *Applied and Environmental Microbiology* 65, 3293–3297.
- Schauer, J.J., Rogge, W.F., Hildemann, L.M., Mazurek, M.A., Bend, R.C., Simoneit, R.T., 1996. Source apportionment of airborne particulate matter using organic compounds as tracers. *Atmospheric Environment* 30, 3837–3855.
- Sinha, S., Banerjee, S.P., 1997. Characterization of Haul road dust in an Indian opencast iron ore mine. *Atmospheric Environment* 31, 2809–2814.
- Sutherland, R.A., Tolosa, C.A., Tack, F.M.G., Verloo, M.G., 2000. Characterization of selected element concentrations and enrichment ratios in background and anthropogenically impacted roadside areas. *Archives of Environmental Contamination and Toxicology* 38, 428–438.
- Sweet, C.W., Vermette, S.J., Landsberger, S., 1993. Sources of toxic trace-elements in urban air in Illinois. *Environmental Technology* 27, 2502–2510.
- Tristan, E., Demetriades, A., Ramsey, M.H., Rosenbaum, M.S., Stavrakis, P., Thornton, I., Vassiliades, E., Vergou, K., 2000. Spatially resolved hazard and exposure assessments: an example of lead in soil at Lavrion. *Environmental Research* 82, 33–45.
- Walmough, S.A., 1999. Monitoring historical changes in soil and atmospheric trace-metal levels by dendrochemical analysis. *Environmental Pollution* 106, 391–403.
- Weiss, P., Riss, A., Gschmeidler, E., Schentz, H., 1994. Investigation of heavy metal, PAH, PCB patterns and PCDD/F profiles of soil samples from an industrialized urban area (Linz, Upper Austria) with multivariate statistical methods. *Chemosphere* 29, 2223–2236.

- Welp, L., Brummer, L., 1997. Microbial toxicity of Cd and Hg in different soils related to total and water-soluble contents. *Environmental Safety* 38, 200–204.
- Xhoffer, C., Bernard, P., Van Grieken, R., Auwera, L.V., 1991. Chemical characterization and source apportionment of individual aerosol-particles over the North-Sea and the English-Channel using multivariate techniques. *Environmental Science and Technology* 25, 1470–1482.
- Yatın, M., Tuncel, S.G., Aras, N.K., Olmez, I., Aygun, S., Tuncel, G., 2000. Atmospheric trace elements in Ankara, Turkey: I. Factors Affecting chemical composition of fine particles. *Atmospheric Environment* 34, 1305–1318.
- Yatın, M., 1994. Source apportionment of urban aerosols using receptor modelling approach: an application to Ankara. Unpublished Ph. D. Thesis, Department of Chemistry, Middle East Technical University. 375pp.

Assessment of spatial–temporal patterns of surface and ground water qualities and factors influencing management strategy of groundwater system in an urban river corridor of Nepal

Prakash Raj Kannel^{a,*}, Seockheon Lee^{a,*}, Young-Soo Lee^b

^aWater Environment & Remediation Research Center, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, Republic of Korea

^bDepartment of Environmental Engineering, Kwangju University, Seoul 139-701, Republic of Korea

Received 14 March 2006; received in revised form 15 September 2006; accepted 12 December 2006
Available online 6 February 2007

Abstract

This study examined the spatial–temporal variations and factors influencing the management of groundwater along a section of the Bagmati river corridor in the Kathmandu valley (Nepal). The results showed that rural areas were less polluted than urban areas. In urban areas, the biochemical oxygen demand (BOD), total nitrogen (TN) and total phosphorus (TP) concentrations ranged from 8.41 to 29.74 mg/L, 6.7 to 128.96 mg/L and 0.06 to 1.5 mg/L, respectively. In rural areas, the BOD, TN and TP concentrations ranged from 0.78 to 18.25 mg/L, 4.8 to 11.56 mg/L and 0.07 to 0.65 mg/L, respectively. The level of organics was higher in the pre-monsoon season, while the level of nutrients was higher in post-monsoon season. A comparison of the groundwater and surface water in the upstream rural areas revealed that the TP concentration was higher in the groundwater than in the surface water, which was attributed to the sorption of phosphorus on iron, aluminum or calcium compounds contained in the surface water, which depends upon the temperature, pH and dissolved oxygen. In urban areas, a few wells were found at groundwater levels lower than the corresponding surface water levels and were subjected to a high risk of pollution. Overall, these findings reinforce the notion that the management of surface and ground waters in an integrated approach is essential for attaining sustainable development of groundwater systems.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Spatial variation; Temporal variation; Groundwater quality; Bagmati river

1. Introduction

The Bagmati river Corridor is the principal resource base of municipal water in the Kathmandu valley (Fig. 1), providing almost 92% of the wet season and 60% of the dry season water supply (CBS, 1998). However, the increasing demand for water as a result of an increasing population has placed pressure on the water quality. Moreover, both the surface and groundwater levels have been degraded as a result of unplanned urbanization and inadequate sewerage facilities, which have accelerated the

discharge of untreated domestic liquid waste into the river corridor (MOPE, 2000).

Domestic sewage and industrial effluent appears to be the major contributors to water pollution in the Kathmandu valley. The polluting industrial activities include the processing of agricultural products, textiles and carpets. The intensive use of fertilizers and pesticides in agriculture around the drainage basin of the river had further aggravated the situation. The average use of chemical fertilizers per hectare, such as nitrogen, phosphorus and potassium, has increased considerably in the country from 7.6 kg in 1975 to 26.6 kg in 1998 (MOPE et al., 2001).

The recharge area for groundwater in the valley is shrinking due to urbanization. The urban areas have increased from 26% in 1978 to 46.2% in 1996. Likewise,

*Corresponding authors. Tel.: +82 2 958 5848; fax: +82 2 958 6854.
E-mail addresses: prakash.kannel@gmail.com (P.R. Kannel),
seocklee@kist.re.kr (S. Lee).

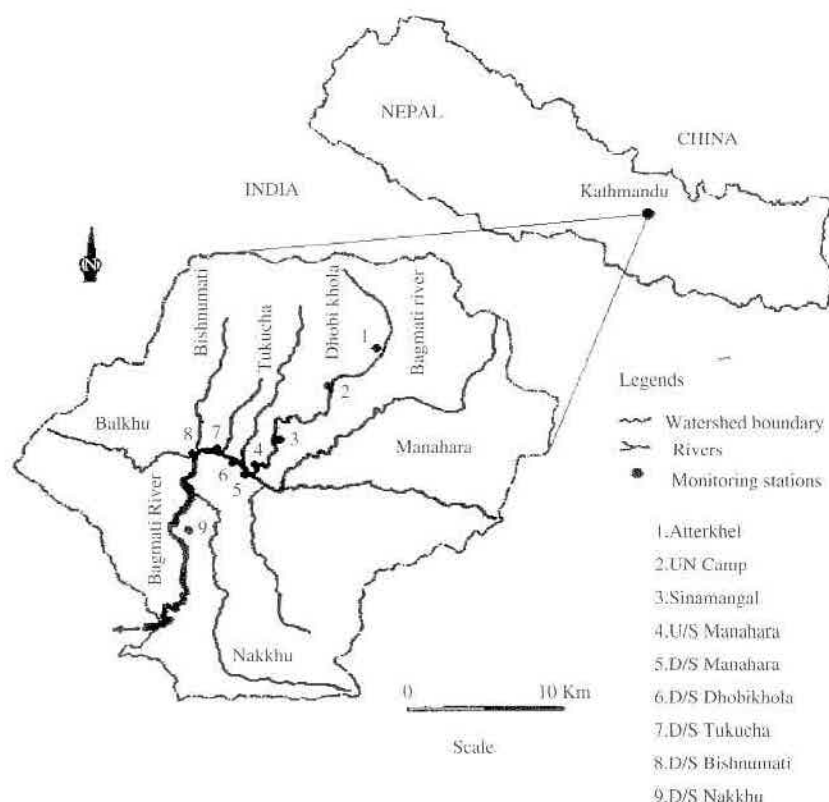


Fig. 1. Map of the Kathmandu valley showing water quality sampling stations along the Bagmati river corridor.

the rural built-up areas have increased from 11.2% to 24% during the same interval (Pradhan, 2000). Groundwater is being heavily used for drinking as well as for other activities, which has resulted in a decline in its water level (MOPE et al., 2001). According to a study by Metcalf and Eddy (2000) in 1999, there was a drop in the pumping water level from 9 to 68.5 m over a few years. The total sustainable level of groundwater withdrawal is approximately 26.3 million liters per day (Stanley et al., 1994). According to the Snowy Mountains Engineering Corporation (SMEC, 1992), the maximum limit of groundwater extraction should be approximately 40.1 million liters per day. However, the total groundwater currently extracted per day is approximately 58.6 million liters (Metcalf and Eddy, 2000). The impact of such water withdrawal from the groundwater reservoir is the infiltration of polluted stream water.

In such circumstances, information on the temporal and spatial trends of the water quality should help in the decision-making process, particularly in developing countries, where there is insufficient data (Massoud et al., 2003). Moreover, there is incomplete understanding of the influence of surface water to reinforce the notion that the management of surface water and groundwater should be integrated for sustainable development. Khatriwada et al. (2002) examined the extent and sources of groundwater contamination in the Kathmandu valley. However, there

have been no studies examining the spatial and temporal variations in groundwater water pollution along the corridor of the Bagmati River. Therefore, this study assessed the state of groundwater pollutions, the spatial-temporal variations and investigated the influential factors of water quality along the river corridor in order to provide the basis for future management strategies.

2. Study area

The study area consisted of an approximately 20 km stretch of the Bagmati River Corridor (Fig. 1), which lies in the Kathmandu valley of the country, Nepal. The valley, which is located in the midland of the Himalayas, lies between 27°32' and 27°49' North and 85°12' and 85°32' East, and is an almost round in shape with a diameter of approximately 30 km E–W and 25 km N–S (Dill et al., 2001). The valley has a central flat part at an elevation of 1300–1400 m above mean sea level (Sharma, 1997). It covers an area of approximately 650 km², with an altitude varying from 1220 to 2800 m above mean sea level (Fujii and Sakai, 2002).

The valley floor has a highly undulating topography with buried ridges of precambrian basement rocks (Jnawali et al., 2002). The valley was originally a lake during Plio/Pleistocene times and had been silted up totally by lacustrine and deltaic river sediment (Shrestha et al.,

1999). The valley basin sediments are mainly loam and are composed of unconsolidated clay, silt, sand and gravel ranging in age from the late Pliocene to the present. The northern area consists mainly of unconsolidated coarse sediment and the southern part consists of thick clay formation and basal gravel (MPPW, 2003). The central part of the basin consists of thick organic mud with sand dominated basal beds underneath (Fujii and Sakai, 2002). The southeast higher part consists of massive gray limestone and the lower formations consists of quartzite and shale with varying degrees of metamorphism (MPPW, 2003).

The valley is in a warm temperate climate zone and receives an annual average rainfall of approximately 1900 mm with temperatures ranging from approximately 0 °C in the winter up to 30 °C in the summer (NWSC, 1992). The valley contains three major cities, Kathmandu, Bhaktapur and Lalitpur, with a total population of the valley being approximately 1.6 million according to the 2001 census (CBS, 2002).

3. Materials and methods

A monitoring network was designed along the Bagmati river corridor (Fig. 1). Nine locations were sampled: three sampling stations, G01, G02, and G09, in rural areas and six stations, G03–G08, in urban areas. Table 1 gives a description of the sampling sites with their locations and types.

The water quality was monitored to determine the chemical and physical parameters: temperature, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), total nitrogen (TN) and total phosphorus (TP) including trace metals (Ca, Fe and Al). The samples were collected in the pre-monsoon (19–20 June) and post-monsoon (2–3 December) seasons and were of 30 h duration. All groundwater

sampling was completed using dip samplers in permanent open dug wells, which were located along the corridor of the river within 200 m from the main river stem. The sampling stations were near the surface water quality sampling sites monitored by Kannel et al. (2005), and the samples were analyzed on the same day and in the same local laboratory.

All the monitored parameters were expressed in mg/L except for the pH, temperature (°C), flow (m³/s), ground water depth (m) and sediment concentration (mg/g). The sampling, preservation, transportation and analysis of water samples were carried out according to the methods described in APHA-AWWA-WPCF (1989). The ground-water depth and the difference in the water levels were measured using a standard tape and fly level.

The BOD samples were taken in sterilized 1000 mL polythene bottles. The samples for the COD, TN and TP were collected in 250 mL polyethylene bottles and preserved with concentrated sulfuric acid to pH < 2. The DO samples were collected in 300 mL bottles by fixing with 1 mL of an alkaline azide solution, 1 mL of a manganese sulfate solution and 1 mL of concentrated sulfuric acid. Prior to transportation, the samples were preserved in an icebox < 4 °C. The water samples for the dissolved metals (each 1 kg) were filtered, and the filtrates were preserved with concentrated nitric acid to pH < 2. The samples were then transported within 6 h to a local laboratory and stored below 4 °C in a refrigerator until needed. The BOD and DO samples were analyzed immediately after reaching the laboratory.

The temperature, pH and DO were measured *in situ*. The temperature was measured using a portable thermometer. The pH was measured with a portable pH-meter. The DO was measured with a portable DO probe as well as in the laboratory by titration with a standard sodium thiosulfate using starch solution as an indicator. The BOD was determined using the azide modification method over a

Table 1
Water quality monitoring stations in the Bagmati river and its tributaries

Types	Stations	Chainage (km)	Locations
Rural	G01	0.000	At right bank of the Bagmati river at about 200 m from the irrigation dam near Atterkhei, Gokarna
Rural	G02	4.875	At right bank at about 50 m from the Bagmati river located approximately 2 km upstream from Wastewater Treatment Plant, Guleswori
Urban	G03	7.900	At left bank of the Bagmati river at about 20 m from Sinamangal steel bridge located at old Baneshwor
Urban	G04	12.120	At about 50 m from the right bank of the Bagmati river located approximately 25 m upstream from the Manahara–Bagmati junction
Urban	G05	12.170	At about 50 m from the right bank of Bagmati river located at 100 m downstream of the Bagmati–Manahara confluence
Urban	G06	14.000	At about 50 m right bank of the Bagmati river located approximately at 100 m downstream of the Thapathali bridge that connects Kathmandu and Lalitpur municipalities
Urban	G07	14.640	At about 75 m from left bank of the Bagmati river located approximately at 200 m downstream from the Thapathali bridge
Urban	G08	15.940	At about 75 m from the right bank of the Bagmati river located just about 35 m downstream of the Bishnumati–Bagmati river junction
Rural	G09	19.800	At left bank of the of the Bagmati river downstream of the Nakkhu khola–Bagmati river junction at a distance of about 200 m from the river Bagmati

5-days incubation period at 20 °C. Within 24 h of sampling, the COD was analyzed using an open reflux method with acidified standard potassium dichromate titrated with ferrous ammonium sulfate. The TP was determined by digestion and a colorimetric method (ammonium molybdenum blue method/ascorbic acid method). Prior to determining the phosphate level, the samples were digested with concentrated nitric and sulfuric acid to convert all the phosphates into the orthophosphate form.

The TN was determined within 7 days of sampling by determining the nitrite-N, nitrate-N and the total kjeldahl nitrogen (TKN) levels. The nitrite-N was determined using spectrophotometry (diazo method). The TKN was determined using the macro-kjeldahl method by following the titration with a standard hydrochloric acid solution. The ammonia nitrogen was determined using the nesslerization method for low concentrations and by titration for higher concentrations.

The trace metals in water were determined by spectroscopic analysis after a preliminary treatment of the preserved samples. The samples were concentrated by slow evaporation over a hot plate in acidic conditions in order to reduce the volume and force the un-dissolved metals into solution. After the preliminary treatment, the level of trace metals was determined by atomic absorption spectrometry (AAS). The aluminum level was determined using eriochrome cyanine R colorimetric method using a UV-visible spectrophotometer. The calcium concentration was determined by titration with a standard EDTA solution using ammonium purpurate as the indicator. The level of iron was determined by atomic absorption.

The trace metals in the sediments were determined by drying the sediment samples and grinding them to pass

through a 150 µm sieve. The sieved samples were dried at 105 °C to remove moisture. After complete drying and cooling, 1 g of the sample was taken in a 250 mL beaker and 30 mL (1+1) HCl was added with 2–3 drops of concentrated HNO₃. The sample was evaporated to dryness over a hot plate. The residue was dissolved in 5 mL HCl and 100 mL water was added. The resulting solution was boiled for a few minutes and the hot solution was filtered through medium textured filter paper. Approximately 3 g of ammonium chloride was added to the filtrate and heated. A dilute ammonia solution was added to turn the solution slightly alkaline. The hydroxide obtained was filtered and ignited at 900–1000 °C for 1 h. The amount of mixed oxide was determined gravimetrically and the hydroxide residue was dissolved in acid and the iron concentration was then estimated by AAS. The aluminum oxide concentration was determined using a subtraction method. The filtrate obtained from the hydroxide precipitation was titrated with a standard EDTA solution after adjusting its volume to 250 mL using Reeder's and Patton's indicator. The amount of calcium was estimated from the volume of EDTA consumed by the fixed sample volume.

4. Results and discussion

Tables 2 and 3 summarize the measured variables in the water and sediment samples. Table 4(a) shows the results of the two-tail *t*-test for the comparison of the water quality between the pre-monsoon and post-monsoon season. Similarly, Table 4(b) gives the result of the *t*-test for the comparison of the water quality between the rural and the

Table 2
Groundwater qualities along Bagmati river corridor

Date	Stations	Locations	Temp (°C)	pH	DO (mg/L)	BOD (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)	GWL (m)	GWL (m)*
19–20 June 2004	G01	Atterkhel	20.0	6.1	1.02	7.12	13.18	4.8	0.14	7.62	1.216
	G02	UN Camp	19.5	6.4	0.89	9.06	16.22	11.56	0.27	3.51	1.315
	G03	Sinamangal	20.0	6.4	0.00	15.65	28.65	6.7	0.15	4.57	−0.765
	G04	U/S Manahara	20.0	7.2	0.57	29.74	45.02	10.68	0.43	—	—
	G05	D/S Manahara	20.0	8.2	0.00	24.88	43.66	10.22	0.12	2.74	4.170
	G06	D/S Dhobi khola	20.5	6.9	0.00	21.44	38.60	10.11	0.24	2.74	−0.185
	G07	D/S Tukucha	20.5	8.6	0.77	19.43	34.90	9.43	0.06	2.06	3.560
	G08	D/S Bishnumati	21.0	9.3	0.62	24.96	49.42	131.0	0.19	2.29	2.069
	G09	D/S Nakkhu	27.0	6.8	0.17	18.25	29.26	7.66	0.44	3.13	7.801
2–3 December 2004	G01	Atterkhel	19.0	6.1	4.30	0.78	7.50	5.89	0.07	7.71	1.126
	G02	UN Camp	18.8	6.2	2.10	9.16	25.00	7.27	0.65	3.34	1.485
	G03	Sinamangal	17.9	6.8	0.26	15.35	21.00	23.73	1.37	3.95	−0.145
	G04	U/S Manahara	19.0	6.7	3.10	10.83	13.50	73.92	1.25	—	—
	G05	D/S Manahara	18.1	6.5	1.04	12.00	19.50	53.63	1.50	2.99	3.920
	G06	D/S Dhobi khola	18.2	6.5	1.50	17.00	26.90	37.81	1.11	2.85	−0.295
	G07	D/S Tukucha	17.0	6.8	2.50	8.41	12.50	17.36	1.41	1.80	3.820
	G08	D/S Bishnumati	19.0	7.0	0.91	9.62	20.00	128.96	0.90	2.79	1.569
	G09	D/S Nakkhu	19.0	6.8	2.04	7.50	12.60	10.61	0.40	3.55	7.381

Note: GWL—groundwater level below local ground surface, GWL*—ground water level with reference to corresponding surface water level, −ve values refers to groundwater level below surface water level.

Table 3
Fe, Al and Ca in water and sediment samples (2–3 December 2004)

S. N.	Locations	Al		Fe		Ca	
		Sediment (mg/g)	Water (mg/L)	Sediment (mg/g)	Water (mg/L)	Sediment (mg/g)	Water (mg/L)
1	Atterkhet, Bagmati river (Location 1)	11.11	0.045	4.34	1.24	3.5	3.23
2	Atterkhet, Dug-well (Location 1)	11.06	0.012	4.2	0.02	2	35.59
3	Sinamangal, Bagmati river (Location 3)	36.4	0.011	18.23	0.04	7	14.56

Al—aluminium, Fe—iron, Ca—calcium.

Table 4(a)
Two sample mean *t*-test (two tail) for groundwater quality data in June and December 2004 along Bagmati river

Parameters	<i>t</i> _{stat}	<i>p</i> -value
Temperature*	3.106	0.013
pH	1.874	0.094
DO*	−3.519	0.006
BOD*	3.024	0.009
COD*	3.308	0.006
TN	−2.257	0.054
TP*	−4.293	0.002

*Significant at 0.05 level (two tail *t*-test).

Table 4(b)
Two sample mean *t*-test (two tail) for groundwater quality data at rural and urban areas along Bagmati River in 2004

Parameters	<i>t</i> -value	<i>p</i> -value
Temperature	0.949	0.379
pH*	−2.790	0.014
DO	1.240	0.255
BOD*	−2.914	0.013
COD*	−2.452	0.027
TN*	−2.346	0.039
TP	−2.121	0.051

*Significant at 0.05 level (two tail *t*-test).

urban areas. Table 5 shows the correlation coefficients for the observed data. Figs. 2–5 gives the plot of the results of the monitored water quality parameters in the pre and post-monsoon seasons.

4.1. State of pollution

The groundwater locations along the Bagmati river corridor showed that the water quality exceeded the permissible limits except for temperature, pH and DO at some locations (Table 2). The temperature ranged from 19.5 to 27 °C in the pre-monsoon season and 17–19 °C in the post-monsoon season. The European Economic Community Standard has a minimum drinking water guideline

Table 5
Correlation coefficients (*r*) for groundwater quality data along the Bagmati river corridor in 2004

Parameters	Temp.	pH	DO	BOD	COD	TN
pH	0.223	—	—	—	—	—
DO	−0.406*	−0.372	—	—	—	—
BOD	0.362	0.669*	−0.701*	—	—	—
COD	0.383	0.730*	−0.669*	0.948*	—	—
TN	−0.258	−0.064	0.112	−0.185	−0.233	—
TP	−0.478*	−0.286	0.211	−0.204	−0.369	0.544*

Note: *Significant at 0.05 level.

value of 12 °C and a maximum of 25 °C. The Canadian drinking water guideline has a recommended value of 15 °C (NHMRC and ARMCANZ, 1996). In the pre-monsoon, the allowable drinking water temperature exceeded these guidelines at station G09. The pH was within permissible range of potable water, 6.5–8.5 (WHO, 1984), except at sites G01–G04, G07–G08 in the pre-monsoon season, and G01–G02 in the post-monsoon season. A DO level >2 mg/L is required in order to prevent anaerobic conditions that can cause bad odors (EMECS, 2001). The DO levels were <2 mg/L at all sites except for G01, G02, G04, G07 and G09 in the post-monsoon season and at all sites in the pre-monsoon season. The BOD concentrations were higher than the permissible limit of 3 mg/L (EMECS, 2001) at all sites except for G01 in the post-monsoon. The maximum acceptable COD level for most industrial uses and for environmental sustainability is 8 mg/L (EMECS, 2001). The COD limit was exceeded at all sites except for one site at G01 in the post-monsoon season.

The TN concentrations were higher than the recommended limit of the Eco-region reference XI (covering the Central and Eastern Forested Uplands of USA) of 0.31 mg/L (US EPA, 2000). However, the high TN value at G08 requires verification. The high permeability of soils and the absence of dense crop cover after applying fertilizer ensure the leaching loss of nitrates to the groundwater (Banaszuk et al., 2005). The sources of nitrogen and phosphorus are industrial, human pollutions and fertilizers. The industrial uses of nitrogen include plastic and metal treatments, textile industries, particleboard and

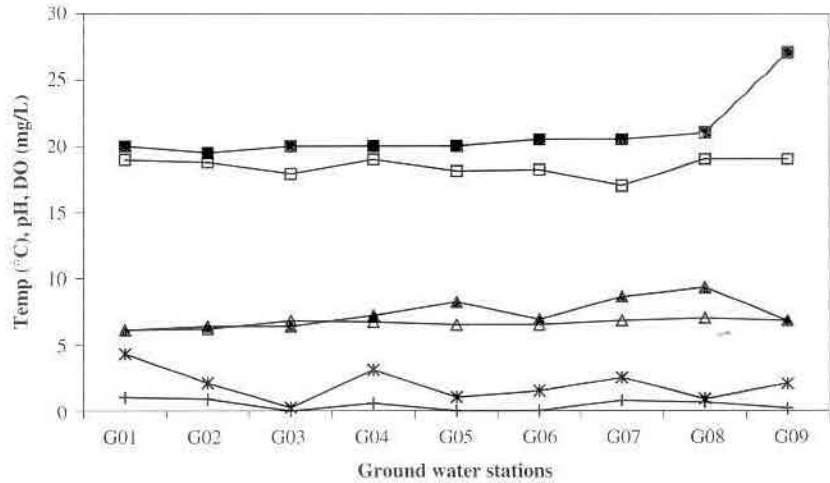


Fig. 2. Temperature, pH and DO levels at the groundwater sampling stations along the Bagmati river corridor on 19–20 June and 2–3 December 2004 [(■) Temp, (▲) pH and + DO in June; (□) Temp, (Δ) pH and (*) DO in December].

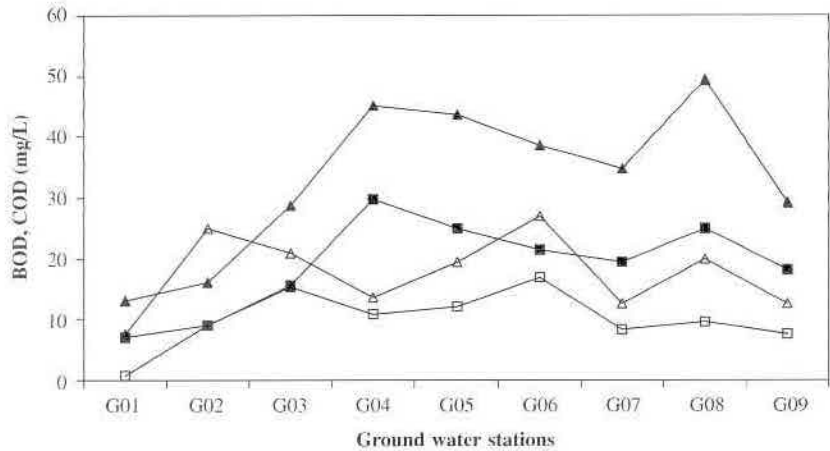


Fig. 3. BOD and COD concentrations at the groundwater sampling stations along the Bagmati river corridor on 19–20 June and 2–3 December 2004 [(■) BOD and (▲) COD in June; (□) BOD and (Δ) COD in December].

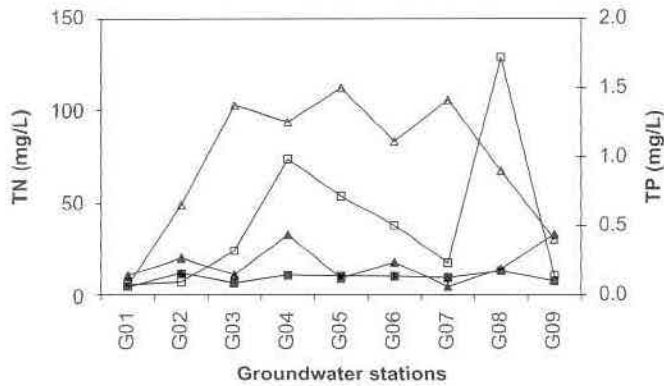


Fig. 4. TN and TP concentrations at the groundwater sampling stations along the Bagmati river corridor on 19–20 June and 2–3 December 2004 [(■) TN and (▲) TP in June; (□) TN and (Δ) TP in December].

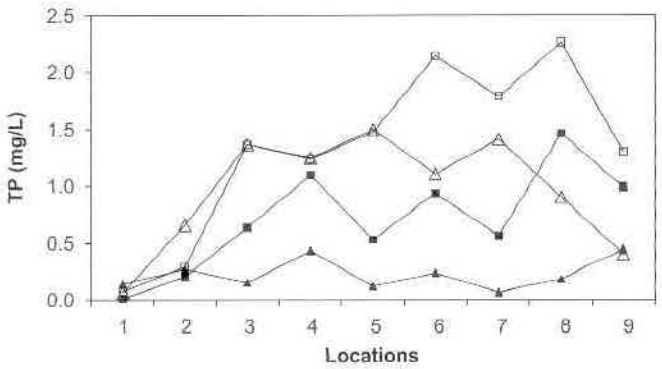


Fig. 5. TP concentrations at surface water stations (Kannel et al., 2005) and corresponding groundwater stations along the Bagmati river on 19–20 June and 2–3 December 2004 [(■) TP in surface water and (▲) TP in groundwater in June; (□) TP in surface water and (Δ) TP in groundwater in December].

plywood, adhesives, feed industries, pharmaceutical industries. The most common nitrogen compounds used are ammonia, nitric acid, urea and ammonium nitrate (Potash

Corporation of Saskatchewan, 2005). The limit of TP concentration is 0.1 mg/L, above which it interferes with coagulation in water treatment plants (US EPA, 1986). The

TP concentrations exceeded the limit at all locations except for G01 in the post-monsoon season and G07 in the pre-monsoon season.

4.2. Temporal variation of water quality

The temperatures (Fig. 2) reflected the seasonal pattern showing higher values in the pre-monsoon season. It varied from 19.5 to 21 °C (excluding station G09) in the pre-monsoon season and 17–19 °C in the post-monsoon seasons. In the pre-monsoon season, the pH units (Fig. 2), BOD and COD concentrations (Fig. 3) were higher, while the DO concentrations (Fig. 2) were lower. The level of nutrients (Fig. 4) was higher in the post-monsoon season (4.8–13.1 mg/L TN and 0.06–0.44 mg/L TP concentrations in pre-monsoon, 5.89–128.96 mg/L TN and 0.07–1.5 mg/L TP concentrations in post-monsoon).

The student *t*-test showed (Table 4(a)) that all the water quality parameters except for the pH and TN were statistically significant ($p < 0.05$). In the pre-monsoon season, the temperature, BOD and COD levels were significantly higher ($p < 0.05$) but the DO and TP levels were significantly lower than in the post-monsoon season ($p < 0.05$).

The reasons for the higher pH in the pre-monsoon season (insignificant, $p = 0.094$) were: (i) under anaerobic conditions, the enhancement of denitrification (microbial reduction of nitrate under anoxic condition) increases the pH (Zilberbran et al., 2001) and/or (ii) the oxidation of organic matter by iron or sulfate reduction results in a higher bicarbonate (a buffering ion) to CO₂ ratio (Park et al., 2006). In the post-monsoon season, nitrification (microbial oxidation of ammonium) under aerobic condition increases acidity (Stumm and Morgan, 1996) leading to a decrease in pH.

The higher BOD, COD and low DO concentrations in the pre-monsoon season are influenced by various factors. Higher rainfall and river flow cause the infiltration of organic matter into the ground water, which decreases the concentration of dissolved oxygen with biodegradation. In addition, the increase in temperature causes a decrease in oxygen solubility, which further reduces the DO concentrations. In the post-monsoon season, the recharge of aerated water oxygenates the shallow groundwater aquifers (Malard and Hervant, 1999) increasing the DO concentrations considerably.

Contamination of the aquifer is accelerated by surface water infiltrating an aquifer when the pumping stations are near the river (Fernando and David, 2005). Aquifer contaminations by rivers carrying highly polluted water have been reported elsewhere (BGS and MOPH, 1994; Kacaroglu and Gunay, 1997).

In the pre-monsoon season, a decrease in DO concentration enhances the de-nitrification process under an anaerobic condition (Pattinson et al., 1998; Carlyle and Hill, 2001), which reduces the nitrogen level considerably (Fig. 4). In the post-monsoon season, the lowering of the

groundwater reservoir level increases the nutrient concentrations (Fig. 4). Significant groundwater abstraction is prevalent in urban areas because the piped water supply is scarce.

4.3. Spatial variation of water quality

The temperature (Fig. 2) showed less variation in rural and urban areas. The pH (Fig. 2) was higher at urban areas. The DO concentration (Fig. 2) was higher in rural areas and decreased as the river entered the urban areas. The BOD, COD (Fig. 3) and TN, TP (Fig. 4) concentrations were higher in urban areas than in the rural areas.

The student *t*-test (Table 4(b)) showed that all the water quality parameter levels except for temperature, DO and TP (marginally significant) were statistically significant ($p < 0.05$). In rural areas, the BOD, COD, TN and pH levels were significantly lower ($p < 0.05$) than in urban areas.

The rural areas, with higher permeability, permit the infiltration of the oxygenated surface runoff and river flow. Less polluted water is introduced in rural areas due to lower anthropogenic activities. Urban areas, allow the infiltration of heavily contaminated surface water via river corridors though paved roads with concrete permitting little infiltrations. In addition, contaminants from leaking sewers, landfill and industrial activities produce a high biological oxygen demand in groundwater (Robins, 2002). The nitrate in groundwater originates mainly from the oxidation of ammonium from the infiltration of effluent in which urea and ammonium prevail over other nitrogen compounds (Zilberbran et al., 2001).

Furthermore, differences in the groundwater quality between the rural and urban areas are influenced mainly by local factors. A groundwater site located in an urban area might have some of the characteristics of a rural area due to the influence of tributaries.

Higher TP concentrations were found in groundwater than in surface water at the upper reaches of the Bagmati River (Fig. 5). According to a study by Kannel et al. (2005), the surface water station at Atterkhel (station B01) near the groundwater station G01 had a TP, DO and pH levels of 0.01, 6.73 mg/L and 7.1 in the pre-monsoon season, and 0.07, 8.5 mg/L and 6.3 in the post-monsoon season, respectively. Similarly, the levels of TP, DO and pH in the surface water station at the UN Camp (station B02) near groundwater station G02 had 0.21, 3.37 mg/L and 6.9 in the pre-monsoon season and 0.29, 6.2 mg/L and 6.5 in the post-monsoon season, respectively.

It was observed that the groundwater stations had higher TP concentrations (0.14 and 0.27 mg/L TP concentrations at G01 and G02, respectively) in the pre-monsoon season (Fig. 5) than the corresponding surface water stations (B01 and B02 have TP concentrations of 0.01 and 0.21 mg/L, respectively). In the post-monsoon season (Fig. 5), only the groundwater station G02 had a higher TP concentration (0.65 mg/L) than the corresponding surface water station

(0.29 mg/L of TP concentration). On the other hand, the groundwater station G01 had an equal TP concentration to that of the surface water station (0.07 mg/L of TP concentrations at both stations). This suggests that the TP concentration is less than or equal to the groundwater at surface water station B01, and was always less than the groundwater at station B02. This suggests that phosphorus level is lower in the surface water than in the groundwater.

The lower TP concentrations in the surface water might be due to the sorption of phosphorus on iron, aluminum or calcium compounds contained in the surface water depending upon temperature, pH and oxygen of water. Aerobic conditions do not favor the release of phosphorus. The high sorption of phosphorus occurs at high dissolved oxygen concentrations (Wang et al., 2003). The sorption of phosphates decreases with increasing pH due to competition with OH^- ion (Lijklema, 1997). The sorption takes place mainly in the pH ranges 2–3 to 6–8 (Zhou et al., 2005). The release of phosphorus increases with increasing temperature (Kim et al., 2004). Hydroxides, oxides and oxy-hydroxides of iron provide sorption sites for phosphorus (Slomp et al., 1996). Gelbrecht et al. (2004) stated that the important processes that reduce the phosphorus concentration in running water include the sorption of soluble reactive phosphorus on Fe (III) hydroxides.

Table 3 shows the Fe, Al and Ca concentrations in sediments and water at three stations: (i) surface water station at Atterkhel (Location 1), (ii) Groundwater station at Atterkhel (Location 1, G01) and (iii) the surface water station at Sinamangal (Location 3). Very high Ca concentration was observed in water sample from groundwater station at Atterkhel (G01), which might be due to agitation of the well by the local people immediately before sampling.

The Al, Fe and Ca concentrations were higher in the surface water sediment samples than in the groundwater samples (Table 3). At the surface water stations, in the pre-monsoon season, the pH range (6–8) and DO concentrations (Fig. 2) favored the sorption of phosphorus on calcium compounds. In the post-monsoon season, higher DO concentration (Fig. 2) and lower temperature favored the sorption of phosphorus on the oxides, hydroxides and oxy-hydroxides of iron or aluminum.

The difference between the level of the surface and ground waters has important implication in the qualities of the groundwater. Lower groundwater levels can receive polluted water from the nearby surface water. Wells G03 and G06 were found below the corresponding surface water levels in both the pre and post-monsoon seasons (Table 2), and received a continuous supply of polluted water from the corresponding river locations. The other wells, at relatively higher levels, contributed base flow to the surface water system.

4.4. Correlation analysis

Table 5 shows the correlation matrix and its significance levels for all the observed data. Because the nine sampling

stations were combined in order to calculate the matrix, the correlation coefficients should be interpreted with caution as they are affected simultaneously by spatial and temporal variations. Nevertheless, some hydro-chemical relationships can be inferred. A high and significant positive correlation was observed between BOD and COD concentrations (correlation coefficient = 0.948) because they are closely related to the contamination of organic matter. Temperature was negatively correlated with DO concentration (correlation coefficient = -0.406, insignificant) because the solubility of the DO concentration decreases with increasing temperature. The BOD and COD concentrations showed negative correlation with the DO concentrations because organic matter is oxidized at the expense of oxygen (Vega et al., 1998). The positive correlation between BOD concentration and pH was attributed to the fact that the production of ammonia and dissolved inorganic carbon during the decomposition of nitrogen-rich organic compounds by bacteria tends to an increase in both the pH and alkalinity (Ferris et al., 1996). The significant correlation between TN and TP concentrations (correlation coefficient = 0.544) is attributed to groundwater pollution from both point and non-point sources.

4.5. Factors influencing management strategies

The water quality of groundwater is affected by many factors including precipitation, surface runoff, groundwater flow, and the characteristics of the catchment area. Seasonal variations in these factors have a strong effect on the concentration of pollutants. The relative water levels between the groundwater and polluted surface waters determine the amount and nature of the deterioration in the groundwater quality along the river corridor.

The introduction of sustainable development and integrated management principles requires the development of new planning tools, which in turn requires a more comprehensive database than is currently available (Schultz, 2000). However, the greatest challenge is to acquire reliable information that will guide the use and protection of water resources. Knowing if certain types of water quality problems are isolated or ubiquitous and whether the conditions are changing spatially or temporally is essential for a proper management plan (Massoud et al., 2006). Furthermore, an effective institutional and regulatory framework is an essential part of a management system. Overall, in addition to the spatial and temporal variation, the relationship between the groundwater and surface water must be considered when assessing the health of the system as well as for setting pollution control strategies.

5. Conclusions

- The groundwater locations along the Bagmati river corridor are exceeding the permissible limits with some exceptions. The water quality is less polluted in rural

areas than in urban areas. In urban areas, the BOD concentration ranged from 8.41 to 29.74. Similarly, the TN and TP concentrations ranged from 6.7 to 128.96 and 0.06–1.5 mg/L, respectively. In rural areas, the BOD concentrations ranged from 0.78 to 18.25 mg/L. Similarly, the TN and TP ranged from 4.8 to 11.56 and 0.07–0.65 mg/L, respectively.

- The TN and TP concentrations are higher in urban areas due to the infiltration of contaminated surface water and the lowering of the groundwater levels as a result of the substantial increase in groundwater abstraction. The level of organics was higher in the pre-monsoon season, while the level of nutrients was higher in the post-monsoon season.
- A comparison of the groundwater and surface water in rural areas revealed the TP concentrations to be higher in the groundwater than in the surface water, which was attributed to the sorption of phosphorus on iron, aluminum or calcium compounds contained in the surface water, which depends on the temperature, pH and dissolved oxygen.
- The groundwater levels of two wells at heart of the municipality were found to be lower than the corresponding surface water levels and are at risk of water pollution.
- Urbanization of the Kathmandu Valley has had a significant impact in deteriorating the surface water and hence the groundwater water quality along the Bagmati River Corridor. These findings should reinforce the notion that the management of surface water and groundwater needs to be integrated in a sustainable development plan of groundwater systems.

Acknowledgement

This work was supported by the Institute of Science and Technology for Sustainability (ISTS) at the Gwangju Institute of Science and Technology (GIST), South Korea.

References

- American Public Health Association (APHA), American Water Works Association (AWWA), Water Pollution Control Federation (WPCF), 1989. Standard Methods for the Examination of Water and Wastewater, 17th ed.
- Banaszuk, P., Wysocka-Czubaszek, A., Kondratiuk, P., 2005. Spatial and temporal patterns of groundwater chemistry in the river riparian zone. *Agriculture, Ecosystems and Environment* 107, 167–179.
- British Geological Survey (BGS), Ministry of Public Health (MOPH), 1994. Impact of urbanization on groundwater: Hat Yai. Final Report. Technical Report WC/94/43, British Geological Survey, Keyworth, Nottingham, UK.
- Carlyle, G.C., Hill, A.R., 2001. Groundwater phosphate dynamics in a river riparian zone: effects of hydrologic flow paths, lithology and redox chemistry. *Journal of hydrology* 247, 151–168.
- CBS, 1998. A Compendium on Environment Statistics. Central Bureau of Statistics, Nepal.
- CBS, 2002. Statistical Pocket Book—Nepal, His Majesty's Government, National Planning Commission Secretariat, Central Bureau of Statistics (CBS), Kathmandu, Nepal.
- Dill, H.G., Kharel, B.D., Singh, V.K., Piya, B., Busch, K., Geyh, M., 2001. Sedimentology and paleogeographic evolution of the intermontane Kathmandu basin, Nepal, during the Pliocene and Quaternary. *Journal of Asian Earth Science* 19, 777–804.
- EMECs, 2001. Water Quality Conservation for Enclosed Water Bodies in Japan 2001, International Center for the Environmental Management of Enclosed Coastal Seas (EMECs); accessed on 20/4/2005, online: <<http://www.emecs.or.jp/>>.
- Ferris, F.G., Stehmeier, L.G., Kantzas, A., Mourits, F.M., 1996. Bacteriogenic mineral plugging. *Journal of Canadian Petroleum Technology* 35, 56–61.
- Fernando, T.W., David, N.L., 2005. Non-agricultural sources of groundwater nitrate: a review and case study. *Water Research* 39, 3–16.
- Fujii, R., Sakai, H., 2002. Paleoclimatic changes during the last 2.5 myr recorded in the Kathmandu Basin, Central Nepal Himalayas, 15th International Himalaya-Karakoram-Tibet Workshop. *Journal of Asian Earth Sciences* 20, 255–266.
- Gelbrecht, J., Lengsfeld, H., Pothig, R., Opitz, D., 2004. Temporal and spatial variation of phosphorus input, retention and loss in a small catchment of NE Germany. *Journal of Hydrology* 304, 151–165.
- Jnawali, B.M., Tuladhar, B., Bahadur, G., Shakya, R.R., 2002. Assessment of barrier potential of sediments for selection of waste disposal sites in the Kathmandu Valley. *Journal of Nepal Geological Society* 27, 69–75 (special issue).
- Kacaroglu, F., Gunay, G., 1997. Groundwater nitrate pollution in an alluvium aquifer Elscisehir urban area and vicinity, Turkey. *Environmental Geology* 31, 178–184.
- Kannel, P.R., Lee, S., Lee, Y., Ahn, K., 2005. Spatial and Temporal Pattern of Water Qualities of Surface Water in Bagmati River, Kathmandu Valley, Nepal. Korea Institute of Science and Technology, South Korea, p. 41.
- Khatiwada, S., Takizawa, S., Tran, S.T.V.N., Inoue, M., 2002. Groundwater contamination assessment for water supply in Kathmandu Valley, Nepal sustainable. *Water Science and Technology* 46 (9), 147–154.
- Kim, L.H., Choi, E., Gil, K., Stenstrom, M.K., 2004. Phosphorus release rates from sediments and pollutant characteristics in Han River, Seoul. *Korea. Science of the Total Environment* 321, 115–125.
- Lijklema, L., 1997. The role of iron in the exchange of phosphate between water and sediments. In: Golterman, H.L. (Ed.), *Interaction Between Sediments and Freshwater*. The Hague, pp. 313–317.
- Malard, F., Hervant, F., 1999. Oxygen supply and the adaptations of animals in groundwater. *Freshwater Biology* 41, 1–30.
- Massoud, M.A., El-Fadel, M., Scrimshaw, M.D., Lester, J.N., 2006. Factors influencing development of management strategies for the Abou Ali River in Lebanon II: seasonal and annual variation. *Science of the Total Environment* 362, 31–41.
- Massoud, M., Scrimshaw, M.D., Lester, J.N., 2003. Qualitative assessment of the effectiveness of the Mediterranean Action Plan: wastewater management in the Mediterranean region. *Ocean & Coastal Management* 46, 875–899.
- Metcalfe, Eddy, Inc., 2000. In: Paper presented at a Seminar on Groundwater and Waste Water, organized by Melamchi Water Supply Development Board, 14 February 2000, Kathmandu, Nepal.
- MOPE, 2000. State of the Environment, His Majesty's Government, Ministry of Population and Environment (MOPE), Kathmandu, Nepal.
- MOPE, ICIMOD, SACEP, NORAD, UNEP, 2001. State of the Environment: Nepal 2001, pp. 119–150.
- MPPW, 2003. Optimizing water use in Kathmandu Valley. Draft Final Report, ADB TA 3700NEP, Ministry of Physical Planning and Works (MPPW), p. 278.
- NHMRC, ARMCANZ, 1996. National water quality management strategy: Australian Drinking Water Guidelines, National Health and Medical Research Council (NHMRC), Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ), p. 376.

- NWSC. 1992. Greater Kathmandu Water Supply Project. Final Report, Nepal Water Supply Corporation, 6B Appendix M, M3–M7.
- Park, J., Robert, A., Sanford, R.A., Craig, M., Bethke, C.M., 2006. Geochemical and microbiological zonation of the Middendorf aquifer, South Carolina. *Chemical Geology* 230, 88–104.
- Pattinson, S.N., García-Ruiz, R., Whitton, B.A., 1998. Spatial and seasonal variation in denitrification in the Swale-Ouse system, a river continuum. *Science of the Total Environment* 210 (1–6), 289–305.
- Potash Corporation of Saskatchewan, 2005. PotashCorp-Nitrogen-PCS Products, assessed on 20 April 2005, available from: <<http://www.potashCorporation.com/nitrogen.htm>>.
- Pradhan, P.K., 2000. Urban growth pattern and food security in Kathmandu Valley. In: Paper Presented at the Workshop on Integrative Strategies for Food Security Improvements in the Kathmandu Valley, 30 April–1 May, Kathmandu, Nepal.
- Robins, N.S., 2002. Groundwater quality in Scotland: major ion chemistry of the key groundwater bodies. *Science of the Total Environment* 294 (1–3), 41–56.
- Sharma, C.K., 1997. Urban air quality of Kathmandu Valley “Kingdom of Nepal”. *Atmospheric Environment* 31 (17), 2877–2883.
- Shrestha, O.M., Koirala, A., Hanisch, J., Busch, K., Kerntke, M., Jäger, S., 1999. Ageo-environmental map for the sustainable development of the Kathmandu valley, Nepal. *GeoJournal* 49, 165–172.
- Slomp, C.P., Van der Gaast, S.J., Van Raaphorst, W., 1996. Phosphorus binding by poorly crystalline iron oxides in North Sea sediments. *Marine Chemistry* 52 (1), 55–73.
- SMEC, 1992. Greater Kathmandu Water Supply Project, Snowy Mountains Engineering Corporation Ltd. (SMEC), Kathmandu, Nepal. Final Report, vol. 3.
- Stanley, I., Macdonald, M., Consultant, E., 1994. Report on the Bagmati Basin Water Management Strategy and Investment Program. East Consult, Kathmandu.
- Schultz, G.A., 2000. Potential of modern data types for future water resources management. *Water International* 25, 96–109.
- Stumm, W., Morgan, J.J., 1996. *Aquatic Chemistry*. Wiley-Interscience, New York.
- US EPA, 1986. Quality Criteria for Water, EPA 440/5-86-001, 1986. US Environmental Protection Agency, Office of Water, Washington, DC.
- US EPA, 2000. Ambient Water Quality Criteria Recommendations: Rivers and Streams in Nutrient Ecoregion XIV. United States Environmental Protection Agency (US EPA), Office of Water, Health and Ecological Criteria Division, Washington, DC. online: <<http://www.epa.gov/>>.
- Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research* 32 (12), 3581–3592.
- Wang, H., Appan, A., Gulliver, A.S., 2003. Modeling of phosphorus dynamics in aquatic sediments: II—examination of model performance. *Water Research* 37, 3939–3953.
- WHO, 1984. Guidelines for Drinking Water Quality, vols. 1 and 2. World Health Organization, Geneva, Switzerland.
- Zhou, A., Tang, H., Wang, D., 2005. Phosphorus adsorption on natural sediments: modeling and effects of pH and sediment composition. *Water Research* 39, 1245–1254.
- Zilberbrun, M., Rosenthal, E., Shachnai, E., 2001. Impact of urbanization on hydro-chemical evolution of groundwater and on unsaturated-zone gas composition in the coastal city of Tel Aviv, Israel. *Journal of Contaminant Hydrology* 50, 175–208.

Evidence of hydrocarbon contamination from the Burgan oil field, Kuwait—Interpretations from thermal remote sensing data

Saif ud din*, Ahmad Al Dousari, Peter Literathy

Environment Sciences Department, Kuwait Institute for Scientific Research, P.O. Box: 24885, Safat—13109, Kuwait

Received 11 March 2006; received in revised form 21 November 2006; accepted 12 December 2006

Available online 8 February 2007

Abstract

The paper presents the application of thermal remote sensing for mapping hydrocarbon polluted sites. This has been achieved by mono-window algorithm for land surface temperature (LST) measurements, using multi-date band 6 data of Landsat Thematic Mapper (TM). The emissivity, transmittance and mean atmospheric temperature were used as critical factors to estimate LST. The changes in the surface emissivity due to oil pollution alter the apparent temperature, which was used as a recognition element to map out oil polluted surfaces. The LST contrast was successfully used to map spatial distribution of hydrocarbon pollution in the Burgan Oil field area of Kuwait. The methodology can be positively used to detect waste dumping, oil spills in oceans and ports, besides environmental management of oil pollution at or near the land surface.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Thermal remote sensing; Hydrocarbon pollution; Land surface temperature; Gulf war; Burgan oil field; Oil lakes; Kuwait

1. Introduction

Kuwait has undergone serious environmental damage due to the formation of “oil lakes” and hydrocarbon surfaces created in the desert following the 1991 Gulf War. There were 810 active wells in Kuwait out of which 730 were damaged or set ablaze during the Gulf War (Kwarteng and Al-Ajmi, 1997). The oil field fires generated huge plumes of black smoke, which eventually settled as soot, tarmat and tarcrete deposits (El-Baz et al., 1994). It was estimated that 2–6 million barrels of oil burned daily, sweeping tons of sulfur dioxide and soot into the atmosphere (Kwarteng and Bader, 1993).

The free flowing crude oil spilled over the desert surface accumulated to form “oil lakes”. These “oil lakes” of Kuwait have been studied in great detail (Al-Besharah et al., 1992; Al-Ajmi et al., 1994; El-Baz et al., 1994; Salam, 1996; Kwarteng, 1998; Al-Dousari, 2001); these are unique and can be considered to be one of the worst environmental disasters of recent times. Following the end of the war, the

damage assessment and reclamation programs were initiated. The Kuwait Oil Company and Kuwait Petroleum Corporation carried out the oil recovery to reduce impact of spilled oil on the local environment. A considerable part of lighter hydrocarbon fraction had evaporated, while the remaining fraction in liquid form was partially recovered. The scars of the damage caused by the spillage of oil remained as soot, tarmats and tarcrete, which are unrecovered denser hydrocarbon fractions.

The tarcrete, tarmats and soot, that affected the underlying soil were subjected to bioremediation which succeeded at places, while the leftover hydrocarbon deposits were covered with mobile sand. This is an unhealthy addition to the arid ecosystem of Kuwait. Over time, the harsh climatic conditions in the Kuwait desert resulted in the onset of hydrocarbon degeneration (Saeed and Al-Mutairi, 2000).

Hydrocarbons on the surface eventually got mixed with soil, and led to changes in the intrinsic properties of the soil, which is manifested as higher land surface temperature (LST) in post-Gulf war period.

Since their formation in 1991, the surface area covered by the “oil lakes”, continues to shrink visibly, leading to the suggestion that there is disappearance of oil polluted

*Corresponding author. Tel.: +965 4989224; fax: +965 5355801.
E-mail address: sdin@kISR.edu.kw (S. ud din).

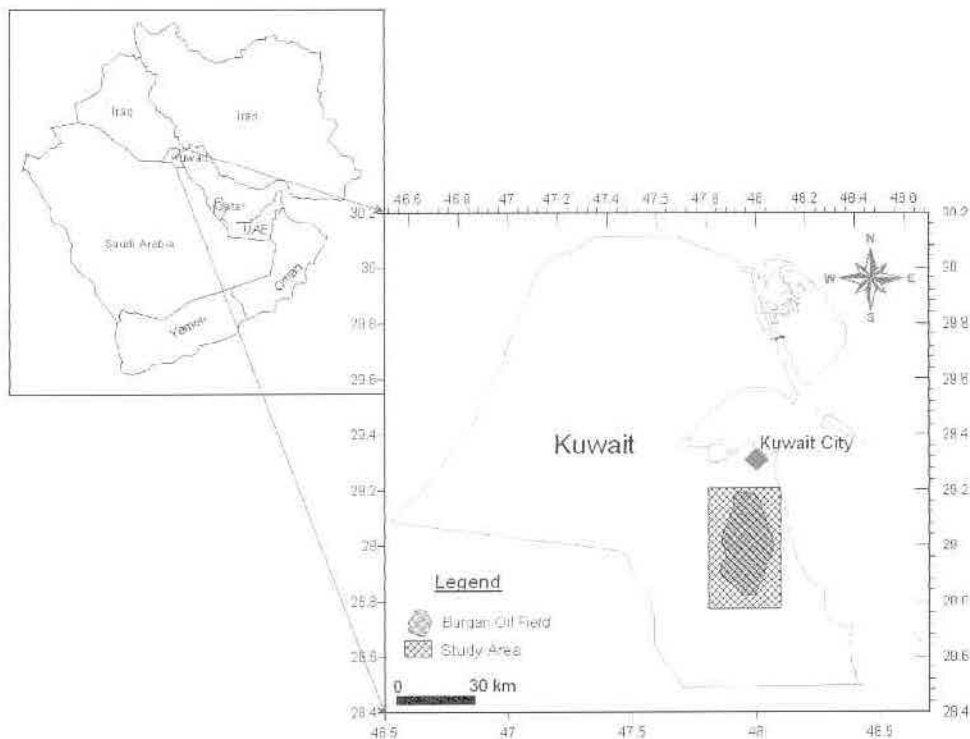


Fig. 1. Location map of the study area covering Burgan and Greater Burgan oil fields, Kuwait.

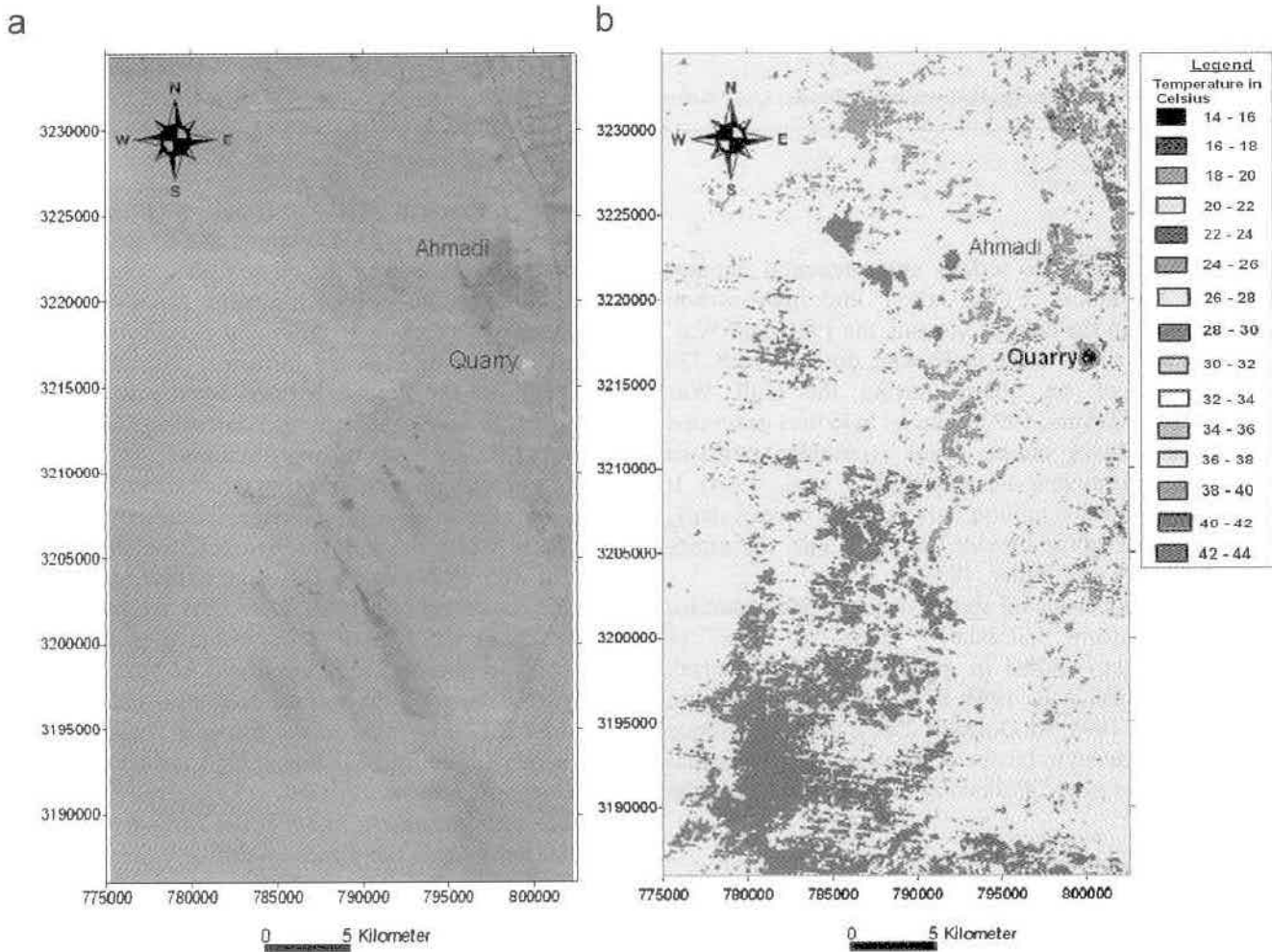


Fig. 2. (a) Landsat TM False color composite bands 4, 2 and 7, (path–row) 165–040 over study area (31 May 1989). (b) land surface temperature image over the study area (31 May 1989).

surfaces. However, this study provides evidence from thermal remote sensing that the spatial spread of hydrocarbon contaminated surface is much more than previously reported (Al-Besharah et al., 1992; Al-Ajmi et al., 1994; Salam, 1996; Kwarteng, 1998). The hydrocarbon polluted surfaces were recognized with the aid of LST contrast and field verifications. The multi-temporal Landsat scenes of 9 June 1989, 30 May 1991, 28 February 1993, 30 August 1995 and 7 April 1998 were selected for this study. All these images are cloud free and are acceptable for image processing. The objective is mapping of spatial disposition of LST pattern, and correlation with oil pollution in the Burgan Oil field, Kuwait.

The LST monitoring using remotely sensed data is a complex process. Several researchers have developed

algorithms to calculate surface temperature using emitted radiance (Rothery et al., 1988; Collier et al., 1989; Becker and Li, 1995; Richter, 1990, 1991; Gillespie et al., 1998; Nichol, 1998; Urai, 2000; Liang, 2001; Qin et al., 2001; Ou et al., 2002; Dousset and Gourmelon, 2003; Ma et al., 2004).

The basic premise of this study is that each material differs in thermodynamic behavior due to difference in its emissivity. The siliciclastics form the bulk of the surface cover over Kuwait. However, the addition of hydrocarbons to these siliciclastics has altered their composition. The compositional changes are reflected as temperature variations. Landsat Thematic Mapper (TM) data of pre-and post-1991 Gulf war have been processed to obtain the LST. The intra-image LST variation in the sand cover was used



Fig. 3. Ground truth images of the study area. All the pictures are of April 1998 except (a) which is of year 1991: (a) soot covering the study area (b) weathered tarcrete (c) denser tarcrete deposits and (d) oil lakes.

to map the hydrocarbon contamination in the Burgan Oil field area (Fig. 1). Burgan Oil field area was worst affected by the oil spillage in 1991. Approximately 90% of 35.45 km² oil lakes were formed in the study area (Kwarteng, 1999). It was, therefore, considered ideal for this study since it is the most polluted and environmentally challenged surface in the country.

2. Methodology

The spatio-temporal variation in the thermodynamic properties of surface material has been mapped in order to identify hydrocarbon polluted surfaces using Landsat TM data. Emissivity is a strong indicator of compositional variation in silicate minerals which make up the bulk of the earth's surface material. Emissivity affects the apparent temperature due to changes in the thermal properties of materials (conductivity, density, capacity and inertia). There are several algorithms proposed to estimate LST from remotely sensed data. The most common of these are mono-window and split window methods (McClain et al., 1985; Richter, 1990; Pearce et al., 1992; Qin et al., 2001),

the latter is used initially to estimate sea surface temperature.

The LST measurement is a complicated task due to high spatio-temporal variation of surface emissivities and atmospheric water vapor as both affect the thermal radiance reaching the sensor (Liang, 2001).

In the present study, Landsat TM band 6 data with wavelength range of 10.45–12.50 μm was utilized to estimate LST. The Landsat TM has single thermal band, therefore the split window method (Sobrino et al., 1996), and temperature-emissivity method (Gillespie et al., 1998), could not be applied. For measurement of LST from Landsat TM data, the mono-window algorithm, single-channel algorithm, and radiative transfer equations, can be used (Richter, 1990; Qin et al., 2001; Jimenez-Munoz and Sobrino, 2003; Sobrino et al., 2004). In this study the algorithm proposed by Richter (1990) was considered appropriate.

The logic for LST retrieval from satellite data is based on the fact that ground emissivity is known. LST can therefore be calculated by accounting for atmospheric correction. Computation of atmospheric correction is a complex

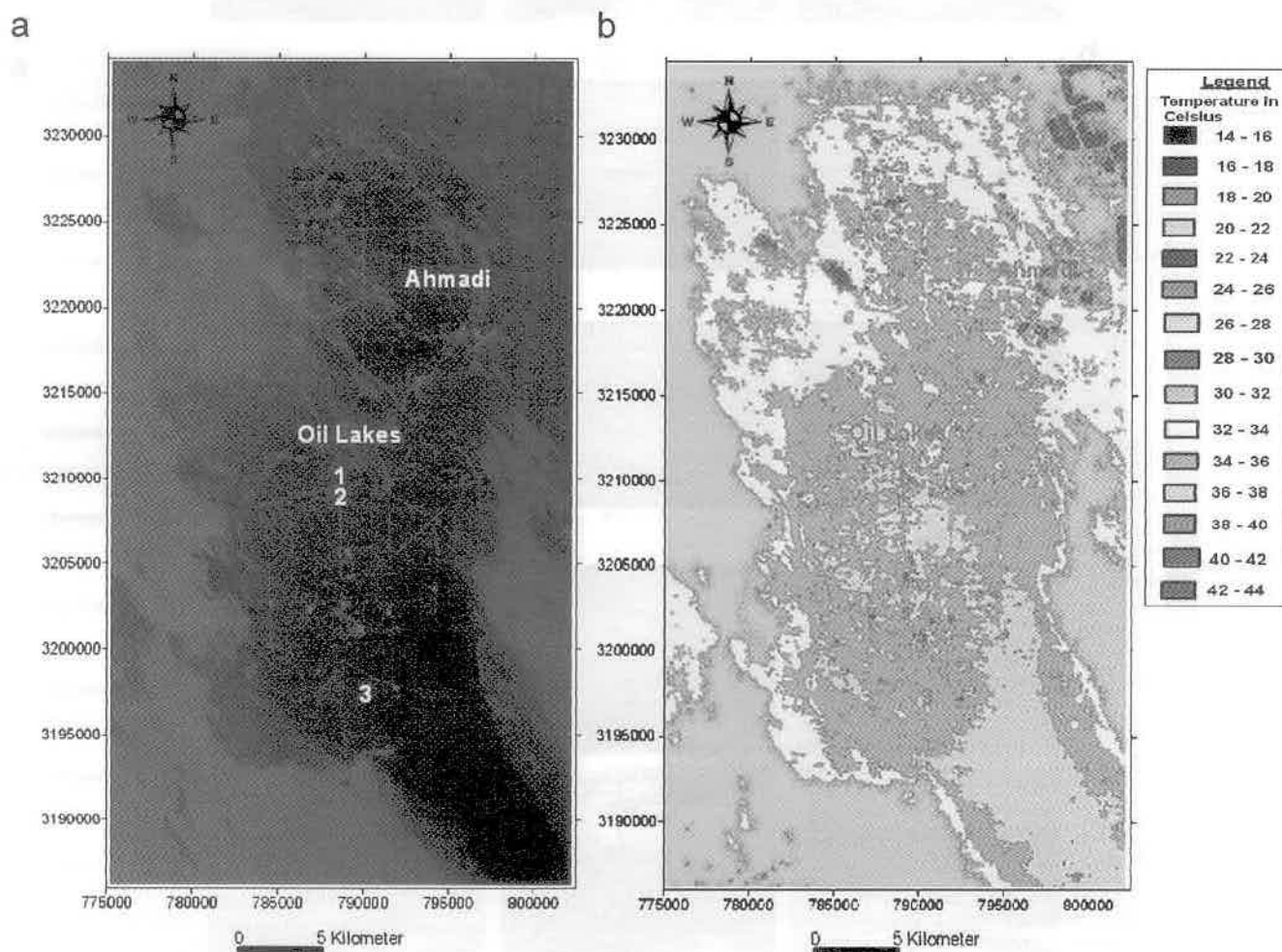


Fig. 4. (a) Landsat TM false color composite bands 4, 2 and 7, (path-row) 165–040 over study area (30 May 1991) and (b) land surface temperature image over the study area (30 May 1991).

process. To simplify the process, however, upward atmospheric thermal radiance and the reflected atmospheric radiance are subtracted from the observed radiance at satellite to compute brightness temperature at ground level. The standard atmospheric profiles provided in ATCOR program of PCI, which is based on LOWTRAN was used to get the atmospheric radiance for surface temperature estimation using band 6 data from Landsat TM.

The mono-window algorithm proposed by Qin et al. (2001) is based on thermal radiance transfer equation to calculate LST. It utilizes transmittance and mean atmospheric temperature to estimate LST.

The LST estimation is done considering the fact that brightness temperature at satellite can be computed by estimation of radiance from DN value and conversion of radiance into brightness temperature. The radiance calculation from DN of TM data utilizes equation developed by Markham and Barker (1986), is

$$L(\lambda) = L_{\min(\lambda)} + (L_{\max(\lambda)} - L_{\min(\lambda)}) Q_{DN} / Q_{\max}, \quad (1)$$

where $L(\lambda)$ is the spectral radiance received at sensor, $Q_{\max}$ the maximum DN value i.e. 255, Q_{DN} the DN value of the

selected pixel i.e. it can be anywhere between 0 to 255, $L_{\min(\lambda)}$ and $L_{\max(\lambda)}$ are minimum and maximum spectral radiance for $Q_{DN} = 0$ and $Q_{\max} = 255$.

The ATCOR Program of PCI uses the constants defined by Schneider and Mauser (1996), where the average wavelength for Landsat TM Band 6 is taken as $11.475 \mu\text{m}$, $L_{\min(\lambda)} = 0.1238 \text{ mW cm}^{-2} \text{ sr}^{-1} \mu\text{m}^{-1}$ for DN value 0 and $L_{\max(\lambda)} = 1.56 \text{ mW cm}^{-2} \text{ sr}^{-1} \mu\text{m}^{-1}$ for DN value 255. Substituting these values in Eq. (1):

$$\begin{aligned} L(\lambda) &= 0.1238(1.56 - 0.1238/255) Q_{DN}, \\ L(\lambda) &= 0.1238 + 0.005632156 Q_{DN}. \end{aligned} \quad (2)$$

The brightness temperature correction for true LST is based on radiative transfer equation (Hurtado et al., 1996). The concept accounts for thermal emittance from an object in accordance with the blackbody theory, which states that

$$B(\lambda)(T) = \frac{C_1}{\lambda^5 (e^{C_2/\lambda T} - 1)}, \quad (3)$$

where $B(\lambda)(T)$ is the spectral radiance of blackbody measured in $\text{W m}^{-2} \text{ sr}^{-1} \mu\text{m}^{-1}$, λ the wavelength in meter, C_1 and C_2

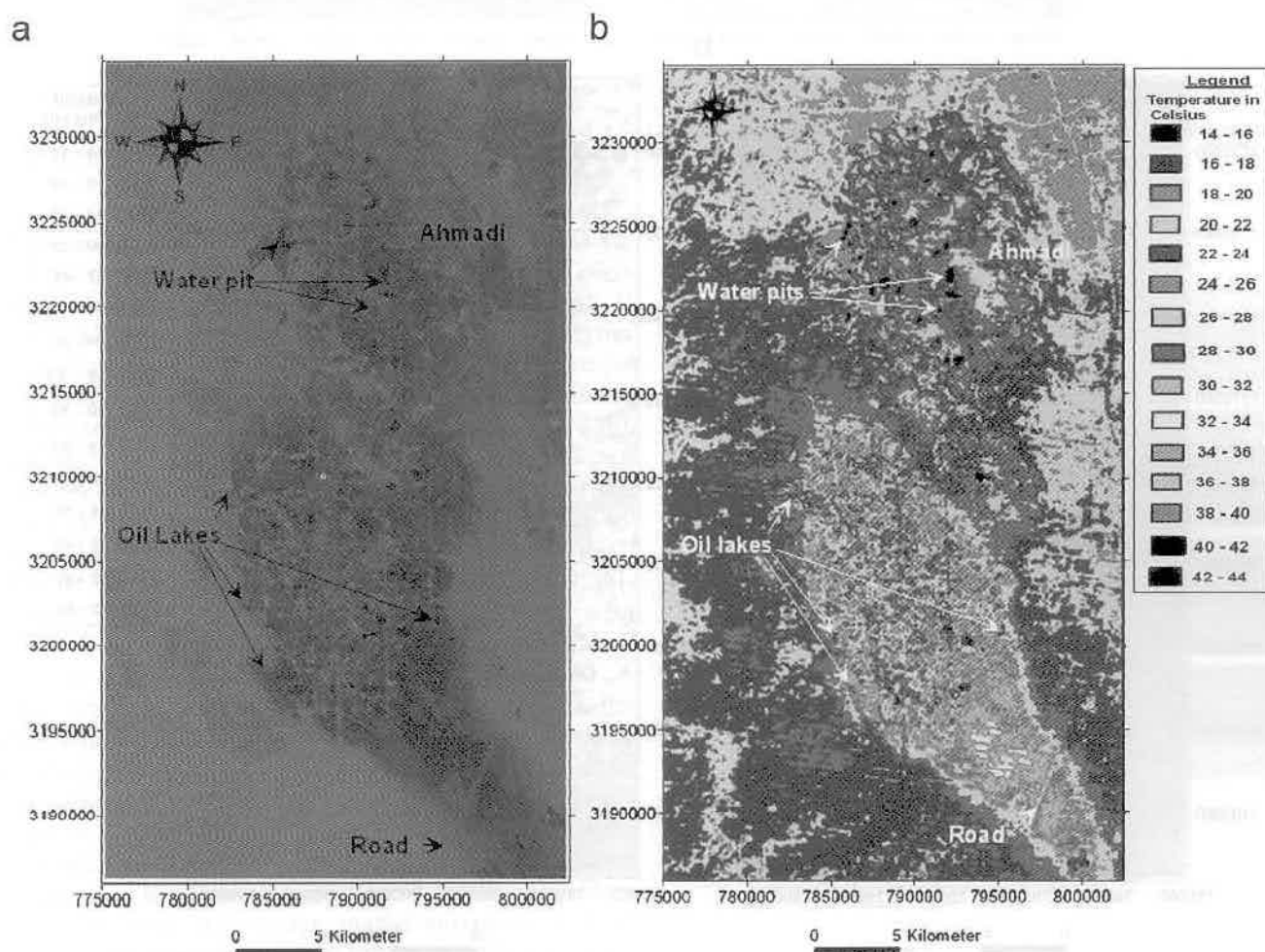


Fig. 3. (a) Landsat TM false color composite bands 4, 2 and 7, (path-row) 165-040 over study area (28 February 1993) and (b) land surface temperature image over the study area (28 February 1993).

are spectral constants, $C_1 = 1.19104356 \times 10^{-16} \text{ Wm}^2$, $C_2 = 1.4387685 \times 10^4 \mu\text{m}^\circ\text{K}$, and T is the temperature ($^\circ\text{K}$).

This concept of blackbody is theoretical. In real life situations, materials everywhere do not behave as blackbodies because of their emissivity:

$$\text{Emissivity} = F_r/F_b,$$

where F_r is the radiant flux exiting a real world body and F_b the radiant flux exiting a black body.

Since the materials in the study area do not behave as blackbodies, emissivities are to be considered which will account for emittance by atmospheric absorption and reflection.

The LST estimates take into account the upward and downward atmospheric radiances. The former is greater than the latter. Their difference in clear sky is within 5°C (Qin et al., 2001). In the present study, the emissivity for the sand in desert environment is assumed to be 0.76, for carbonate lithologies in study area is 0.92, and heavy oil is 0.97. The approximation is taken as

$$B_6(T_a^\infty) = B_6(T_a). \quad (4)$$

To get LST from T_s , the Planck's radiance function is linearized (Qin et al., 2001).

$$L_6 = a + bT_6,$$

where a and b are coefficient valid for temperature range of $0\text{--}70^\circ\text{C}$. a is -67.355351 and b is 0.45860 .

The ATCOR program used for estimation of the LST assumes two surface temperatures T_{s1} and T_{s2} with their corresponding black body temperatures T_{b1} and T_{b2} . The atmospheric correction and elevation functions are included in ATCOR, which provide a very good approximation. The temperature images are derived from ATCORT1.

3. Results and discussion

The digital image processing of Landsat TM band 6 for LST measurement carried out for post-1991 Gulf war images show remarkable pattern coincident with the surface hydrocarbon disposition in the area expressed as variable LST within the image. The area shows an increase in the LST as compared to pre-war image of 1989.

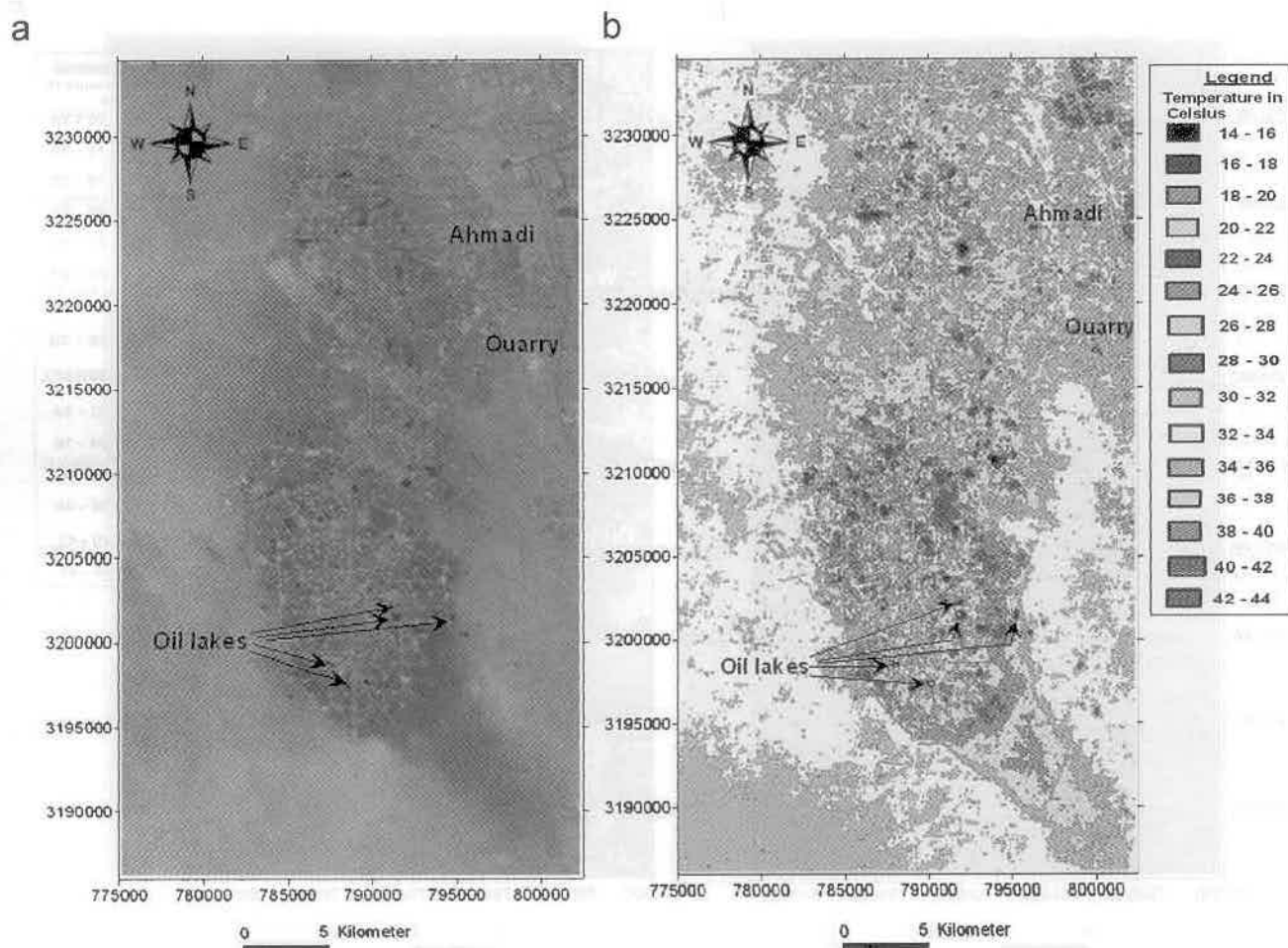


Fig. 6. (a) Landsat TM false color composite over study area bands 4, 2 and 7 (path-row) over study area (30 August 1995) and (b) land surface temperature image over the study area (30 August 1995).

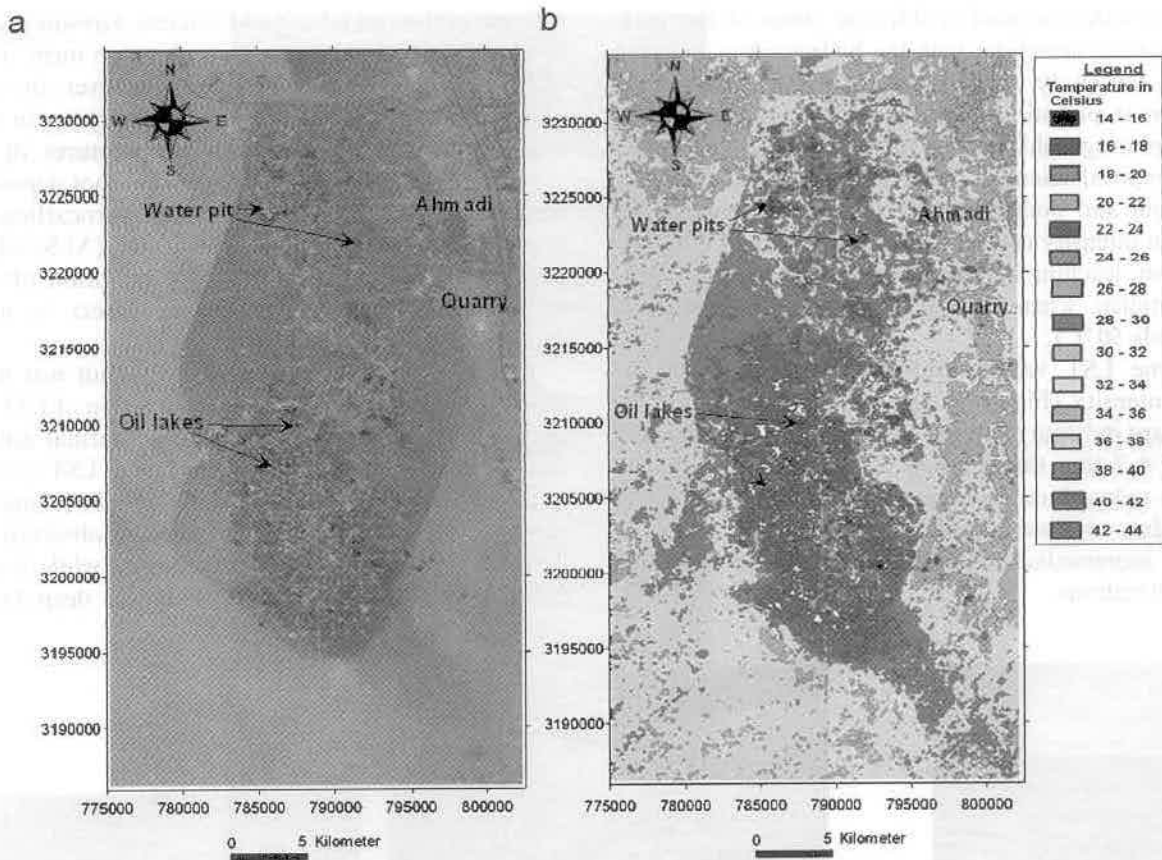


Fig. 7. (a) Landsat TM false color composite over study area bands 4, 2 and 7 (path-row) over study area (7 April 1998) and (b) land surface temperature image over the study area (7 April 1998).

Intra-image LST variations depict oil polluted surfaces, since the emissivity and thermodynamic behavior of oil is different from that of soil or water. The 1989 image (Fig. 2) shows LST differences due to different emissive properties of the lithology and sand. The area exhibit LST variations in the sand cover in the post-1991 Gulf war images, due to hydrocarbon polluted surfaces formed during the 1991 Gulf war. The persistent spatial intra-image LST pattern correlates with the oil lakes, tarmats, and soot fallout (Fig. 4–7). Selective ground truth has been carried out by Kuwait Institute for Scientific Research (Literathy, 1992, 1993; Al Sarawi et al., 1998; Kwarteng, 1998, 1999; Al-Dousari, 2001). The intra-image LST variations were verified from the collated ground truth data (Fig. 3) to verify the temperature anomalies in the images.

The seasonal and micro-climatic implications to the actual temporal response of LST are beyond the scope of this study. The intra-image LST spatial variations, rather than temporal temperature comparison between each image is relevant in this study. The intra-image LST variations correlates very well with the hydrocarbon polluted sites in the study area (Fig. 4–7). The algorithm applied to study intra-image spatial variations in LST shows temperature variations in 1991, 1993, 1995 and 1998 (Fig. 4–7) in the Burgan Oil Field of Kuwait. These figures show similar LST pattern during successive years even

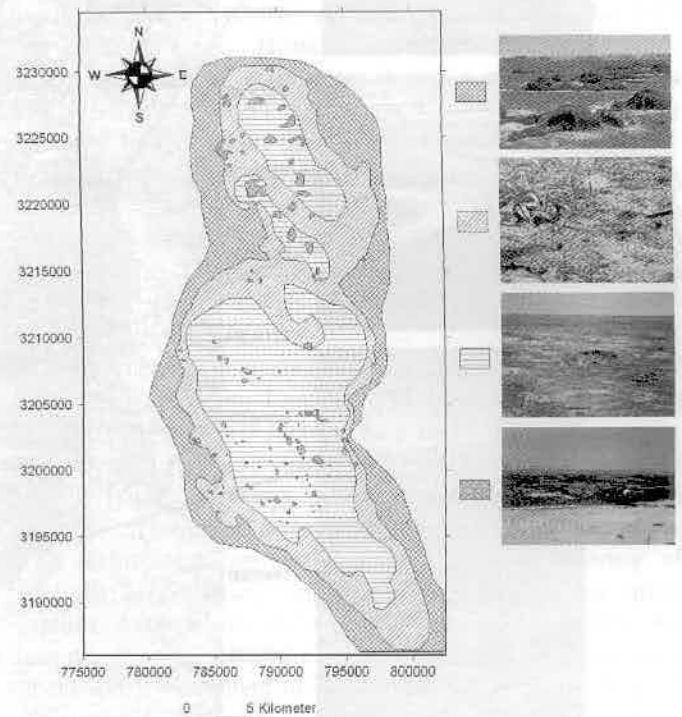


Fig. 8. A sketch map showing different pollution zones within the study area.

though they were obtained at different times of the year. The LST pattern correlates with the hydrocarbon polluted surfaces in the study area. A sketch map is included to show different pollution zones (Fig. 8) and location of some field photographs are also marked on LST map (Fig. 9). The temporal changes have resulted in reduction of spatial extent and pollution intensity of these zones. The reduction in intensity is primarily due to weathering of the hydrocarbon, leaching during rains and downward migration in summers when surface temperature is extremely high (exceeds 50°C).

Intra-scene LST values are directly proportional to pollution intensity (Fig. 10). The higher temperatures in the LST maps indicate relatively higher pollution. It is clear from Figs. 4–7 that the intensity of pollution and spatial extent has reduced in successive years. This decrease is probably due to several reasons: increase in depth of pollutants, bioremediation, sludge recovery, leaching and chemical alterations.

Some of the “oil lakes” and tarcrete deposits are covered by thin veneer of sand which has obscured them. The depth of burial has increased over years; however, the thickness of sand sheet has not increased in recent years in the study area (Saif et al., 2006). High temperatures in summer months imparts mobility to the tar and soot deposits at the surface, and winter rains leach out hydrocarbon fraction which percolate down with the water (Al-Senafy et al., 1997). Consequently, the depth of pollution observed at several places varies from surface veneer to maximum depth of 4 m.

The analyses of pre-and post-1991 Gulf war images of study area shows a very elegant co-relation of LST with the spatial locations of oil lakes, tarcrete, tarmat and soot in the Burgan oil field. Higher intra-image LST corresponds to higher total petroleum hydrocarbon (TPH) and/or near surface burial. Very high values are also observed over the flaring chimneys of the oil separators, while lower LST correlates well with lower TPH and/or deep burial and

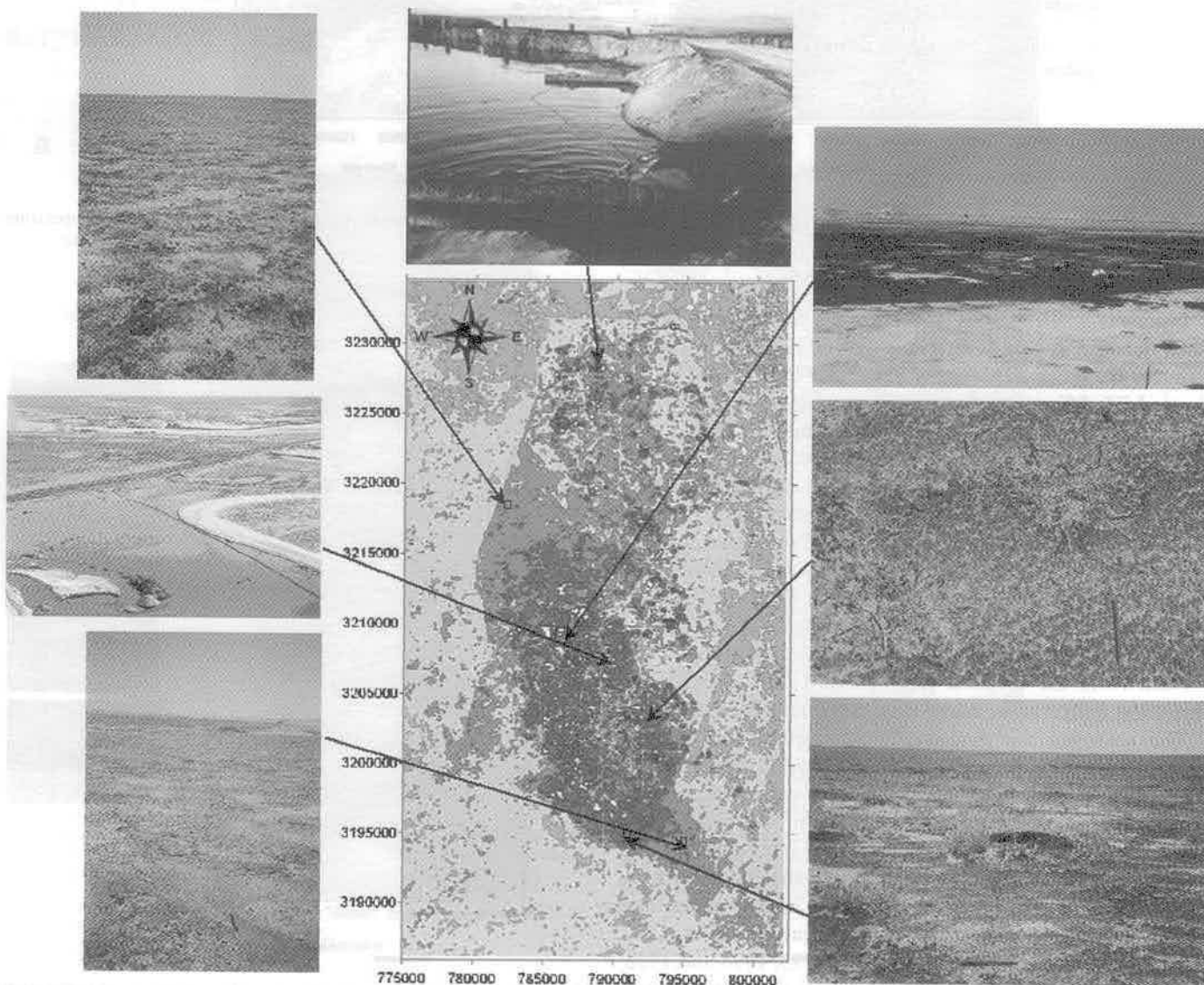


Fig. 9. Location map of some field sites on LST map.

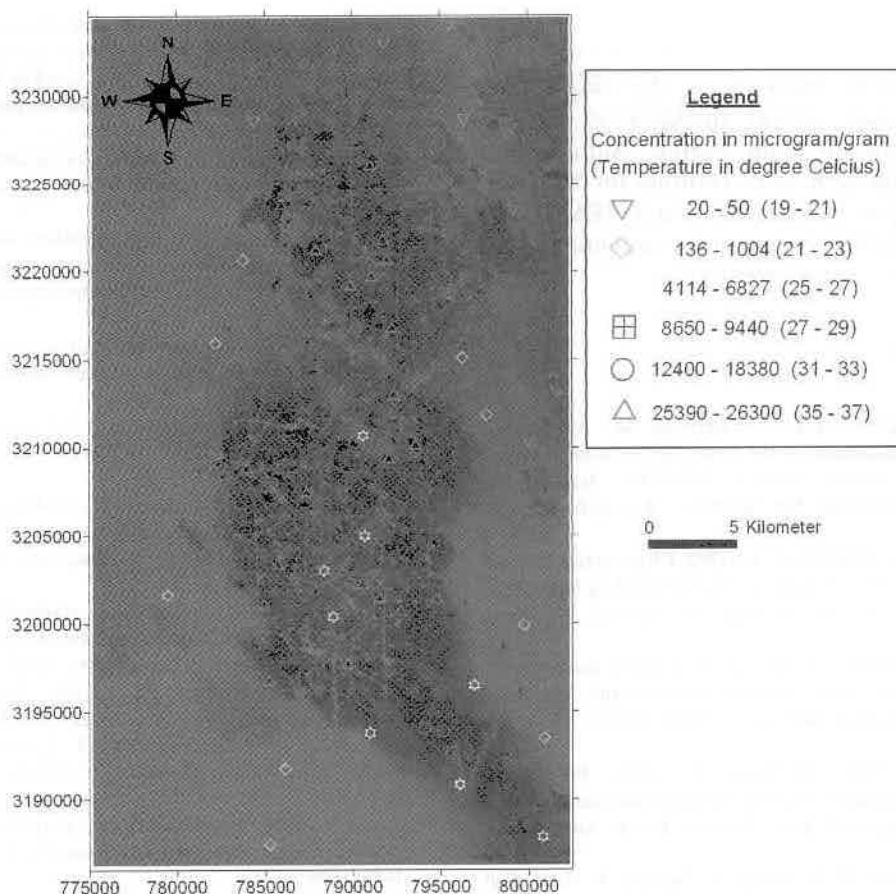


Fig. 10. TPH concentrations and LST in study area.

water pits. The TPH concentration in the uncontaminated soil varies from 7 to $20 \mu\text{g g}^{-1}$ whereas the “oil lakes” shows a TPH concentration of around $25000 \mu\text{g g}^{-1}$. These “oil lakes” correspond to the high-temperature zones in LST maps. The tarmats are solidified thick sludge deposits with TPH concentration of $12,000\text{--}19,000 \mu\text{g g}^{-1}$. They correspond to a slightly lower temperature than “oil lakes”. There exists a good correlation between the hydrocarbon concentrations and LST in all images.

4. Conclusion

The thermal band of Landsat TM has not been optimally used for LST measurements, primarily due to complexities in various algorithms. The simplified methodology adopted here will be useful for mapping hydrocarbon pollution at, and near surface, in the arid ecosystems. The adopted algorithm is based on three critical factors: emissivity, transmittance, and mean atmospheric temperature to estimate LST. The emissivity of sand is estimated as 0.76, while for heavy oil is 0.97 in the present study. ATCOR standard atmospheric profiles were used for atmospheric correction. However, Qin et al. (2001) demonstrated that transmittance dependence of LST is minimum in low water vapor conditions as is the case in the

study area. The possible error in estimation of the LST is $\pm 0.9^\circ\text{C}$ to 1.1°C . However, in this study the accuracy of the actual LST estimation is not a primary concern since the relative LST patterns are sufficient to identify different hydrocarbon polluted surfaces. The actual LST measurement and calibration is part of an ongoing research.

The soil sample analyses show a very good correlation between the TPH concentrations and the temperature anomalies, with relatively high TPH concentrations representing higher LSTs.

It has been demonstrated in this study that LST measurements using Landsat TM data can be useful in mapping oil polluted surfaces as a result of changes in the emissivity of the sand surface. The methodology presented in this paper can be adopted for detecting oil polluted surfaces on land and as spills in oceans and near the ports. The technology can be used for quick monitoring of spatially large areas, using similar algorithm for other satellite data, where the temporal coverage is quick. An important application can be in near real time monitoring of oil spills. The effect of darkness will not affect the LST observations, so waste dumping during night-time can be easily identified for pollution control and environmental management.

Acknowledgements

Authors are extremely thankful to the anonymous reviewers who suggested critical modification to the manuscript for making it more useful. Authors are thankful to management of Kuwait Institute for Scientific Research for supporting the present study (EM037K). Thanks are also due to Mr. A. Jacob for providing some excellent photographs.

References

- Al-Ajmi, D., Misak, R., Khalaf, F.I., Al-Sudairawi, M., Al-Dousari, A.M., 1994. Damage assessment of the desert and coastal environment of Kuwait by remote sensing (VT001C). Report no. KISR 4405, Kuwait Institute for Scientific Research, Kuwait (unpublished).
- Al-Besharah, J., Salman, M., Al-Matrouk, F., 1992. Characterization and quantification of reclaimable oil from oil lakes formed in Kuwait oil fields. Report No. 4081, Kuwait Institute for Scientific Research, Kuwait.
- Al-Dousari, A., 2001. Analysis of change in the tarcrete layer on the desert surface of Kuwait using satellite imagery and cell-based modeling. Ph.D. Thesis Remote Sensing and GIS, Boston University (unpublished).
- Al-Sarawi, M., Massoud, M.S., Al-Abdali, F., 1998. Preliminary Assessment of oil contamination levels in soils contaminated with oil lakes in the Greater Burgan oil fields, Kuwait. *Water, Air and Soil Pollution* 106, 493–504.
- Al-Senafy, M.N., Viswanathan, M.N., Senay, Y., Sumait, A., 1997. Soil contamination from oil lakes in northern Kuwait. *Journal of Soil Contamination* 6 (5), 481–494.
- Becker, F., Li, Z., 1995. Surface temperature and emissivity at various scales: definition, measurement and related problems. *Remote Sensing Reviews* 12, 225–253.
- Collier, P., Runacres, A.M.E., McClatchey, J., 1989. Mapping of very low surface temperatures in the Scottish Highlands using NOAA AVHRR data. *International Journal of Remote Sensing* 10 (9), 1519–1529.
- Dousset, B., Gourmelon, F., 2003. Satellite multi-sensor data analysis of urban surface temperatures and landcover. *ISPRS Journal of Photogrammetry and Remote Sensing* 58, 43–54.
- El-Baz, F., Abuelgasim, A., Koch, M., Pax-Lenney, M., Lambin, E., Al-Doasari, A., Marr, P., Ryherd, S., Morency, R., 1994. Detection by satellite images of environmental change due to the Gulf War. In: El-Baz, F., Makharia, R.M. (Eds.), *The Gulf War and the Environment*. Gordon and Breach Science Publishers, Lausanne, pp. 1–24.
- Gillespie, A.R., Rokugawa, S., Matsunaga, T., Cothren, J., Hook, S., Khale, A., 1998. A temperature and emissivity separation algorithm for advanced spaceborne thermal emission and reflection radiometer (ASTER) images. *IEEE Transaction of Geoscience and Remote Sensing* 36, 1113–1126.
- Hurtado, E., Vidal, A., Caselles, V., 1996. Comparison of two atmospheric correction methods for Landsat TM thermal band. *International Journal of Remote Sensing* 17, 237–247.
- Jimenez-Munoz, J.C., Sobrino, J.A., 2003. A generalized single channel method for retrieving land surface temperature from remote sensing data. *Journal of Geophysical Research* 108 (Doi:10.1029/2003JD003480).
- Kwarteng, A.Y., 1998. Multitemporal remote sensing data analysis of Kuwait's oil lakes. *Environment International* 24 (1/2), 121–137.
- Kwarteng, A.Y., 1999. Remote sensing assessment of oil lakes and oil-polluted surfaces at the Greater Burgan oil field, Kuwait. *International Journal of Applied Earth Observation and Geoinformation* 1 (1), 36–47.
- Kwarteng, A.Y., Al-Ajmi, D., 1997. Satellite Remote Sensing Applications in the State of Kuwait. Kuwait Institute for Scientific Research, Kuwait, 101pp.
- Kwarteng, A.Y., Bader, T.A., 1993. Using satellite data to monitor the 1991 Kuwait oil fires. *Arabian Journal of Science and Engineering* 18, 95–115.
- Liang, S., 2001. An optimization algorithm for separating land surface temperature and emissivity from multispectral thermal infrared imagery. *IEEE Transactions on Geoscience and Remote Sensing* 39 (2), 264–274.
- Litherathy, P., 1992. Environmental consequences of the Gulf War in Kuwait: Impact on water resources. *Water Science Technology* 26, 21–30.
- Litherathy, P., 1993. Considerations for assessment of environmental consequences of the 1991 Gulf War. *Marine Pollution Bulletin* 27, 349–356.
- Ma, Y., Menenti, M., Tsukamoto, O., Ishikawa, H., Wang, J., Gao, Q., 2004. Remote Sensing parameterization of regional land surface heat fluxes over arid area in northwestern China. *Journal of Arid Environments* 57, 117–133.
- McClain, E.P., Pichel, W.G., Walton, C.C., 1985. Comparative performance of AVHRR-based multichannel sea surface temperatures. *Journal of Geophysical Research* 90, 11587–11601.
- Markham, B.L., Barker, J.L., 1986. Landsat MSS and TM post calibration dynamic ranges, atmospheric reflectance and at satellite temperature. EOSAT Landsat Technical Notes 1, Earth Observation Satellite Company, Lanham, Maryland, pp. 3–8.
- Nichol, J.E., 1998. Visualisation of urban temperatures derived from satellite images. *International Journal of Remote Sensing* 19 (9), 1639–1649.
- Ou, S.C., Chen, Y., Liou, K.N., Cosh, M., Brutsaert, W., 2002. Satellite remote sensing of land surface temperatures: application of the atmospheric correction method and split window technique to data of ARM-SGP site. *International Journal of Remote Sensing* 23 (24), 5177–5192.
- Pearce, A.F., Prata, A.J., Manning, C.R., 1992. Comparison of NOAA/AVHRR-2 sea surface temperatures with surface measurements in coastal waters. *International Journal of Remote Sensing* 10, 37–52.
- Qin, Z., Karnieli, A., Berliner, P., 2001. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel–Egypt border region. *International Journal of Remote Sensing* 22 (18), 3719–3746.
- Richter, R., 1990. A fast atmospheric correction algorithm applied to Landsat TM. *International Journal of Remote Sensing* 11, 159–166.
- Richter, R., 1991. Atmospheric Correction of Landsat TM, MSS, and SPOT Images: ATCOR Users Manual. DLR-IB 552-14/90, Oberpfaffenhofen: German Aerospace Research Establishment <http://www.rese.ch/atcor/atcor_paper.html>.
- Rothery, D.A., Francis, P.W., Wood, C.A., 1988. Volcano monitoring using short wavelength infrared data from satellites. *Journal of Geophysical Research* 93 (B7), 7993–8008.
- Saeed, T., Al-Mutairi, M., 2000. Comparative composition of polycyclic aromatic hydrocarbons (PAHs) in the sea water soluble fractions of different Kuwaiti crude oils. *Advances in Environmental Research* 4, 141–145.
- Salam, A.J.A., 1996. The oil lakes environment disaster. In: A. Al-Awadi, M.T. Balba, C. Kamizawa (Eds.), *Proceedings. Symposium on Restoration and Rehabilitation of the Desert Environment*. Elsevier Scientific Publishing Co., Amsterdam, pp. 117–137.
- Saif, uddin, Al-Dousari, A., Al-Ghadban, A.N., Aritoshi, M., 2006. Use of interferometric techniques for detecting subsidence in the oil fields of Kuwait using synthetic aperture radar data. *Journal of Petroleum Science and Engineering* 50 (1), 1–10.

- Schneider, K., Mauser, W., 1996. Processing and accuracy of Landsat TM data for lake surface temperature measurement. *International Journal of Remote Sensing* 16, 2111–2124.
- Sobrino, J.A., Li, Z.L., Stoll, M.P., Becker, F., 1996. Multi-channel and multi angle algorithm for estimating sea and land surface temperature with ASTER data. *International Journal of Remote Sensing* 17 (11), 2089–2114.
- Sobrino, J.A., Jimenez-Munoz, J.C., Paolini, L., 2004. Land surface temperature retrieval from Landsat TM 5. *Remote Sensing of Environment* 90, 434–440.
- Urai, M., 2000. Volcano monitoring with Landsat TM short-wave infrared bands: The 1990-1994 eruption of Unzen Volcano, Japan. *International Journal of Remote Sensing* 21 (5), 861–872.

Environmental restoration of invaded ecosystems: How much versus how often?

Ram Ranjan*

International Agricultural Trade and Policy Center, Department of Food and Resource Economics, University of Florida, USA

Received 6 March 2006; received in revised form 3 December 2006; accepted 12 December 2006

Available online 6 February 2007

Abstract

This paper derives the optimal level of restorative efforts required to restore environments degraded by invasive species invasion. Specific attention is focused on a case when restoration faces the risk of failure through relapse of the restored environment caused by repeat invasions. The level of restored environment may also play a role in its future improvement or susceptibility to failure. The tradeoff between the optimal level of environmental quality and number of restorative efforts required to attain that given environmental quality is highlighted.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Environmental restoration; Resilience; Restoration failure; Invasive species

1. Introduction

Invasive species are a noticeable source of biodiversity degradation (Glowka et al., 1994). Lately, invasive species have become a subject of widespread concern due to the enormous economic and environmental damages that they inflict upon society (Pimentel et al., 1991, 2000). While a number of options exist to prevent the advent of invasive species, none of them are foolproof. Once the species have invaded a given eco-system, steps could be taken to either control them in part or eradicate them. However, it is rarely economically or physically viable to eradicate them. Yet, in most cases the invaded environment could be restored to a certain extent in order that society may continue to derive economic and environmental services from it.

Recent studies on the economics of invasive species management include those by Shogren (2000), Knowler and Barbier (2000), Olson and Roy (2002), Eiswerth and Van Kooten (2002), Perrings (2003), Eiswerth and Johnson (2002), Perrings (2002), Horan et al. (2002), Rouget et al. (2004), Settle and Shogren (2002), Van Wilgen et al. (2004), etc. While these studies focus mostly on the optimal

combination of prevention and control options, one possible option is to take restoration measures to bring the invaded eco-system close to its pre-invaded state through control of the pest population and recovery of the eroded environment. Eradication can also be considered as an option, though as an extreme form of restoration. Restoration may be more relevant for recovery of habitats such as grasslands, lakes, wetlands, etc., which provide environmental services to the society in their un-invaded states.

This paper looks at the important issue of the extent of optimal restoration of an invaded environment that provides economic amenities to society, but is faced with the threat of repeated invasions.¹ Possibility of re-invasion, which might lead to failure of the restoration project, is considered. While it is possible for a restoration project to fail independently of any invasion, this possibility is ruled out here to concentrate on the issue of failures caused by repeated invasions only. This also allows us to focus on the important issue of extent of restoration when the risk of failure may be related to the level of restoration. It is possible for restoration projects to fail due to many reasons that are not invasive species related. However, such risks

* Tel.: +352 392 1881 326; fax: +352 392 9898.

E-mail address: rranjan@ifas.ufl.edu.

¹ The level of amenities depends upon the level of environmental quality, which in turn would be affected by the extent of restoration.

only enter as an exogenous parameter if incorporated within our model and do not add any additional insight over the key question of building positive stock of resilient biodiversity.

Current work on restoring invaded ecosystems has been mostly confined to the field of restoration ecology. Yet, there are significant issues of economic importance that come into play while deciding the extent of restoration. Total restoration may neither be feasible nor desirable for most invaded ecosystems due to the high costs associated with achieving and maintaining them. Further, restored eco-systems face the risk of falling back into degraded states from repeated invasions. Therefore, restoration efforts that do not incorporate this possibility of failure are bound to lead to inefficient outcomes. Most restoration efforts, after the initial investment, require substantial subsequent efforts to constantly monitor and fight the invasives for sustained periods of time. This is an essential feature of restorative efforts that are specifically targeted against repeated invasions. The restored environment may face continuous threats from repeated invasion even as restoration efforts are underway.²

Experimental work on restoration ecology has revealed that degraded eco-systems may be resilient to restoration efforts owing to changes in landscape connectivity and changes in native species pools from invasion by exotics (Suding et al., 2004). Suding et al. point out that understanding the nature of feedback between the abiotic and biotic factors in the degraded state of the environment is crucial for the success of restoration projects. For instance, biotic factors and abiotic factors may interact in the post-invasion scenario reinforcing each other in such a fashion that reverting back to the pristine state becomes difficult. See Suding et al. for some examples on this interaction. These feedbacks often lead to degraded states that are resilient to restorative efforts. By simply restoring the abiotic factors and omitting the biotic factors, the pristine level of environmental quality may not be achieved. In such cases, either due to loss of dominant species, or significant changes in the environment, there is a regime change which is non-linear in nature and which might exhibit hysteresis (Mayer and Rietkerk, 2004 and references therein). Suding et al. cite several reasons for this resilience effect seen in degraded environments. These include the 'species effects', 'trophic interactions' and 'landscape connectivity'. The species effect occurs when the new species has changed the characteristics of the invaded environment in its favor (and against the native species) eventually carving out a niche for itself. Thus, by doing so it makes the degraded environment much more resilient to restoration efforts. One case is that of grasses in Hawaii woodlands that make the environment fire prone by altering the nitrogen cycle, thus eliminating the native shrubs (Suding et al., 2004). Similarly, habitat destruction can cause a reduction in the

native seed-source, thus making the restoration of the native species much more difficult. Finally, trophic interactions between the herbivores in the Savannas have led to the replacement of the woody grasses with the inedible invasive shrubs. In fact, most systems that exhibit strong or moderate abiotic regimes, such as arid grasslands, rangelands, wetlands, etc. are prone to regime shifts that lead to resilience in restoration (Didham et al., 2005).

A crucial economic issue is then over the extent of restorative efforts to be undertaken per period when the environment exhibits resilience and the risk of failure of restoration projects is real. For instance, invasive species that make grasslands more prone to frequent fires may be countered by planting other species that compete with them and are also less prone to frequent fires. However, in case of a fire break-out species of both kinds would get eliminated, thereby negating all the previous efforts of restoration. Another related issue is over the level of restorative efforts when risks are stock dependent. In the above case, the more the species of fire-resistant kind are planted; the lower would be the risk of failure of restorative efforts. Further, higher stock of fire-resistant species may exhibit stock-dependent resilience, i.e., once a threshold level of fire-resistant species has been reached, there may be a sharp decline in the level of other restorative efforts required to preserve the level of restored environment.

Restoration and resilience improving measures under risk have been found to be at the center of issues that concern invaded ecosystems in the ecology literature. However, these issues also make the economic analysis fairly complicated, as the non-linear attributes of the ecological processes must be included in a traditional cost-benefit approach. Currently there are very few known applications of restoration risks on invasives species in the economics literature. For instance see Hatzistahis et al. (1998).

In this paper, a model of environmental restoration is designed that incorporates the risk and resilience effects associated with environmental restoration. The issue of how much restoration effort to undertake is then looked at in an inter-temporal cost-benefit analysis setting. When risk of failure may be stock dependent, the question of how much restoration versus how often becomes relevant, as the costs of continual but lower restoration must be weighed against the costs of less frequent but larger restorative efforts leading to a higher environmental quality. This also determines under what circumstances a more resilient state is desirable given the higher costs associated with its attainment. The literature on the management of ecosystems faced with non-linear dynamics and regime changes provides guidance over choice of consumption paths and control efforts that might prevent costly reversals from one state to another (Maler et al., 2003). This paper extends that analysis and applies to the case of invaded ecosystems and their restoration when multiple equilibria are possible. Significantly, valuable insights are derived over the optimal level of restoration of invaded environments faced with

²For example, invasive plant species may survive through the next season through their seeds, which may be hard to eliminate.

threats of repeated invasion, the risk of which might be stock-dependent. The model also provides insights into the tradeoff between the levels of restoration and the time required to attain it. Numerical simulations reinforce these analyses.

The paper first starts with a deterministic model, where restoration efforts are not faced with the threat of failure, in order to understand the role of resilience associated with environmental restoration. The analysis delves into the existence of multiple equilibria with respect to environmental restoration. Next, risk of failure is introduced into the model. Finally, the tradeoff between the level of restoration and the frequency of failure of restoration is taken up in the above setting.

2. Basic model

Consider a degraded environment that could provide recreational and environmental benefits upon restoration. The stock of environmental quality (q) evolves as

$$\dot{q} = \alpha I + \eta \frac{q^a}{q^a + b} - \delta q. \quad (1)$$

Environmental quality stock q improves from restoration efforts net of any natural rate of decay δ . The amount of environmental quality lost to decay increases as the level of environmental quality improves. This assumption is made in order to make unlimited improvements in environmental quality difficult.³ Furthermore, it also captures the fragile balance that certain complex ecosystems like wetlands and lakes exhibit and which are prone to degradation from slightest perturbations caused by human and natural forces.⁴ For instance, the restoration of Florida everglades requires amongst other things introduction of native species and restoration of water flow to prevent decay of vegetation, etc. It is likely that as the quality of restoration is improved through an increase in stock of native species, a larger quantity of the same would get lost from the stagnating vegetation if water flow is not improved.

There may be multiple options available for restoration; but in order to simplify things, we assume that it is possible to combine these options together into a single restoration variable (I). The choice of a single restorative variable is justifiable in the sense that in most cases restorative efforts do not necessarily have to add in environmental stock from outside. Most of the time restorative efforts are simply about removing the cause of trouble. Even the degraded

environments may have a capacity to grow back to their full potential, but are often overshadowed by the negative forces such as pests that cause its degradation through a complex interaction with natural forces such as fire, droughts, floods, diseases, etc. One particular example is the case of Buffel grass invasion in Queensland, Australia on the native species such as the Brigalow and Gridgee. Buffel grass pastures increase the risk of fire amongst these native species, and the more fire-infested the surrounding gets, the higher is the density of Buffel grass over time. Thus, in a positive feedback relationship with fire and the native species, Buffel grass has been able to replace most of these native species over time (Butler and Fairfax, 2003). Other examples of models involving resilience in grasslands can be found in Perrings and Walker (1997, 2004).

α is a parameter that transforms restoration efforts into stock of environmental quality. The second term on the right-hand side in Eq. (1) captures a sharp upward jump exhibited in environmental quality once a threshold level is crossed. This term captures the resilience aspect of degraded ecosystems. Conventionally, resilience has been defined in two ways in the ecology literature. First one, termed as the '*engineering resilience*' defines it as the speed of bouncing back of any perturbed system (Pimm, 1984). The other one, termed the '*ecological resilience*', is about the amount of stress that the system can tolerate before flipping from its original state to another stable but degraded state (Holling and Meffe, 1996; Carpenter and Cottingham, 1997). In certain cases a system might both be *resistant* and *resilient* to external shocks. *Resistance* in the literature is defined as the ability of the system to sustain a shock while remaining unchanged, as compared to resilience, which refers to its ability to revert back to the original state.

In this paper we follow the '*ecological resilience*' definition to model the impact of restoration. Parameters η , a and b define the rate and magnitude of this effect. Restoration induced jump in environmental quality is defined in a positive sense here, as beyond a certain threshold of environmental restoration the environment shifts into a better state and is more responsive to restoration efforts. Alternatively, this formulation mandates that a willful restorative perturbation in environmental quality would not lead a system out of its degraded state unless some threshold is crossed. Consequently, resilience here is defined in terms of an improvement in the ability of the system, through enhanced biodiversity or other possible changes, to fight back invasive species and remain uninvaded. Maler et al. (2003) use the functional form as in Eqs. (1) and (2) in their paper to model the negative impact of a pollutant such as an input of phosphorous in a lake which could lead to hysteresis effect once a threshold level of the stock of phosphorous is crossed. In this paper, however, we have used the same functional form to model resilience effect. With an increase in the stock of environmental quality stock (q), it is the

³It also allows us to explore steady-state properties of the system. When the decay parameter is absent from Eq. (1), it is possible to make unlimited improvements in environmental quality. However, situations where the costs of restoration are significant and the time horizon not too long, the net benefits from the environment would converge, thus making an expected cost-benefit analysis feasible.

⁴For instance, Keohane et al. (2000) incorporate both natural decay and recovery of the environment while modeling the optimal restoration problem.

restored environment that exhibits failure or resistance to revert back and therefore becomes resilient.

Instantaneous benefits $m(q)$ are derived from environmental quality.⁵ The cost of restoration $c(l)$ is convex in restorative efforts, thus making unlimited restoration prohibitive. Let μ be the shadow price of environmental quality and r the social discount rate. Here we assume that the time horizon of the inter-temporal optimization problem is infinite. The results of this analysis could also be applied to restoration projects with a finite time horizon as long as the discount rate is positive. Society maximizes benefits from environmental quality net of restoration costs as

$$\text{Max} \int_0^{\infty} [m(q) - c(l)]e^{-rt} dt \quad (2)$$

subject to the constraints posed on environmental quality evolution as given by Eq. (1). The current value Hamiltonian (cvh) is written as

$$\text{cvh} = m(q) - c(l) + \mu \left(\alpha l + \eta \frac{q^a}{q^a + b} - \delta q \right). \quad (3)$$

First-order condition with respect to restorative efforts implies that per unit cost of restoration must be equated to the shadow value of that marginal unit of restoration:

$$\frac{c'(l)}{\alpha} = \mu. \quad (4)$$

Co-state variable μ evolves as

$$\dot{\mu} = -m'(q) + (r + \delta)\mu - \mu \frac{\eta a b q^{a-1}}{(q^a + b)^2}. \quad (5)$$

From (4) and (5), the time path of restorative efforts could be derived as

$$\dot{l} = \frac{-m'(q)\alpha}{c''(l)} + \frac{c'(l)}{c''(l)} \left(r + \delta - \frac{\eta a b q^{a-1}}{(q^a + b)^2} \right). \quad (6)$$

In a steady state, restorative efforts and environmental quality are held constant. From (1) to (6) we get

$$\alpha l + \frac{\eta q^a}{q^a + b} = \delta q, \quad (7)$$

$$\frac{\alpha m'(q)}{c''(l)} = \frac{c'(l)}{c''(l)} \left(r + \delta - \frac{\eta a b q^{a-1}}{(q^a + b)^2} \right). \quad (8)$$

Eqs. (7) and (8) define a relationship between environmental quality and restorative efforts, which could be solved to derive their steady-state values. The isoclines for which the levels of restorative effort and environmental quality are constant are represented in Fig. 1 below.⁶

⁵These benefits are ecological benefits that do not deplete from public consumption. Ecosystems such as grasslands, forests and fisheries are also subjected to direct harvests that lead to a reduction in the environmental stock. This has not been modeled here as the primary goal of restoration may not be immediate consumption in most cases.

⁶The shapes of the cost and benefit curves are assumed to be non-linear, the relevant parameters of which are shown in Appendix A.

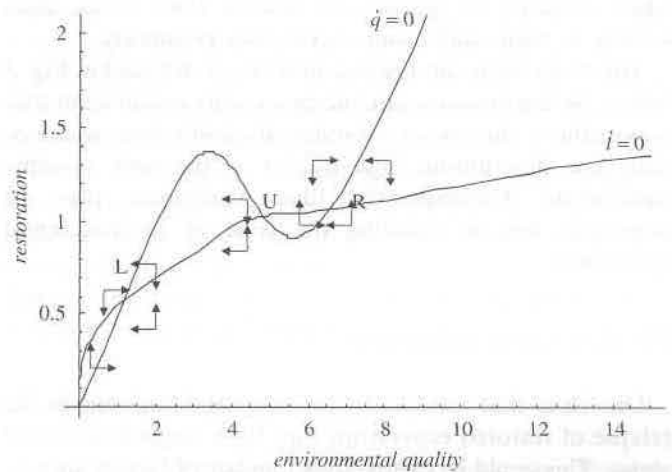


Fig. 1. Steady-state levels of environmental quality and restoration.

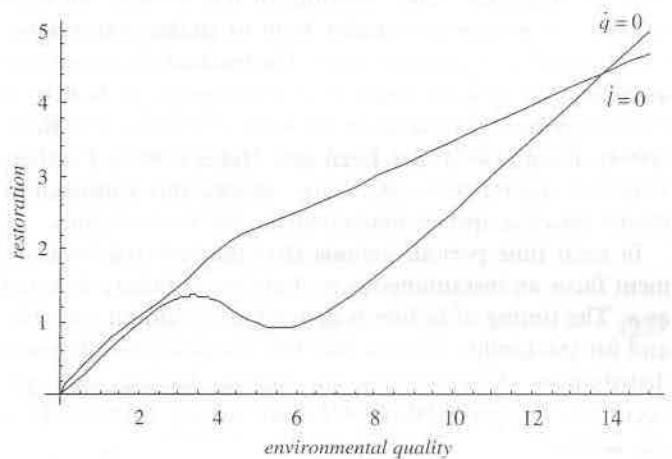


Fig. 2. A case of resilient steady state (high stock benefits).

Note that there exist three possible equilibria L , U and R , the low, middle and the high environmental qualities, respectively. Of the three, the low and the high equilibria are the stable ones with the middle one being unstable. The resilience effect is depicted by a jump in environmental quality once environmental quality crosses the threshold given by the crest in the $\dot{q} = 0$ curve. The state below this threshold is the degraded state. Also note that R is the resilient equilibrium as environmental quality can be reduced significantly without letting the system flip back to the low-quality steady state. The threshold, below which environmental quality falls into the 'degraded' state, is given by region on the right-hand side of the unstable equilibrium. The state left to this threshold is the high-quality state or the resilient state. The 'high equilibrium', which is the resilient state, may not be possible to reach from a degraded state in all cases. If benefits from environmental restoration are lower than the costs incurred, or if the discount rate is high, or if the resilience effect is not very significant, the high equilibrium may not be desirable. Similarly, the low equilibrium may not be desirable under certain circumstances. Fig. 2 depicts a case

where benefits of restoration exceed their costs, thus leading to high equilibrium as the only possibility.

The effect from varying discount rate is depicted in Fig. 3 below. As the discount rate increases, only equilibrium that is possible is the low environmental quality one, whereas, with low discounting high quality is the only possible equilibrium. Consequently, time preferences play an important role in deciding the level of environmental restoration.

2.1. Restoration with relapse

One issue that restoration projects are faced with is the relapse of restored ecosystems into their original degraded states. This could be caused by a number of factors such as renewed infestations which could be seasonal, climate-induced or man-made. Further, once the system flips back into the degraded state, restoration has to start all over again as environmental quality built up in the past is gone. This possibility is explored next. The methodology used for modeling the risk of invasion in this system is that of a Poisson process, based upon the work of Clarke and Reed (1994), Reed (1988) and Reed and Heras (1992). Further, Knowler and Barbier (2005) have applied this approach to model invasive species management under uncertainty.

In each time period, assume that the restored environment faces an instantaneous probability of failure denoted as p . The timing of failure is defined as a random variable, and for tractability assume that this variable has a Poisson distribution. Conditional on failure having not yet occurred, the probability of failure in any interval ds is pds where

$$\int_0^t pds = \lambda(t). \quad (9)$$

If T is a stochastic variable that represents the time of failure, the cumulative probability density function for

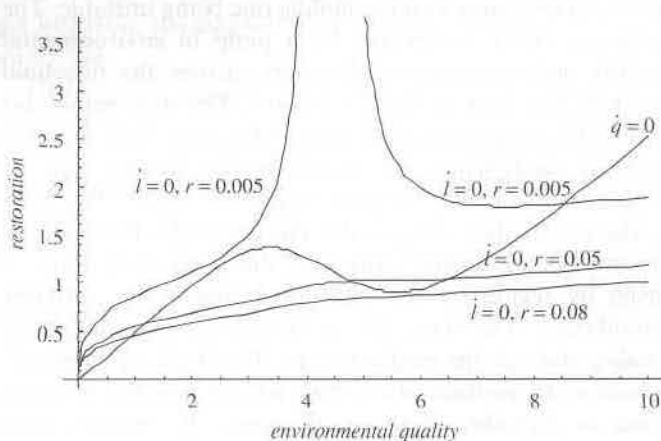


Fig. 3. Restorative effort isoclines for various discount rates. Note: In the above figure, there are three isoclines pertaining to three different discount rates. The isoclines for the discount rate .005 are split into two parts, whereas the isoclines for higher levels of discount rates (.05 and .08) are continuous.

failure is $F(t) = \Pr(T < t)$ and $F(t) = 1 - e^{-\lambda(t)}$, given the Poisson specification. The survivor function is $S(t) = \Pr(T > t) = 1 - F(t)$ which is defined as the probability of project survival up to time period t .

The manager is faced with the challenge of incorporating such possibilities into her optimization framework. The manager's task is to maximize her long-term value $V(q_0)$, starting with an initial level of quality q_0 as

$$V(q_0) = \text{Max} \int_0^\infty (m(q) - c(l) + pV(q_0))e^{-\lambda(t)-rt} dt \quad (10)$$

subject to (1), where p is the constant hazard rate of invasion characterized by a Poisson process. In Eq. (10), the term $pV(q_0)e^{-\lambda(t)-rt}$, when integrated up to infinity, captures the expected value from the system flipping back into the original state at time t and the manager having to start all over again. $V(q_0)$ represents the value function from starting all over again from the initial level of environmental quality q_0 . Eq. (10) in its extended form can be re-written as

$$V(q_0) = \text{Max} \left[\int_0^\infty (m(q) - c(l))e^{-\lambda(t)-rt} dt + \int_0^\infty (pV(q_0))e^{-\lambda(t)-rt} dt \right], \quad (11)$$

given that $\lambda(t) = pt$, the above relation can be further simplified as

$$V(q_0) = \text{Max} \int_0^\infty (m(q) - c(l))e^{-\lambda(t)-rt} dt + pV(q_0) \frac{1}{r+p}. \quad (12)$$

Eq. (12) reduces to

$$V(q_0) = \frac{r+p}{r} \text{Max} \int_0^\infty (m(q) - c(l))e^{-pt-rt} dt. \quad (13)$$

Setting up the current value Hamiltonian for the above problem, we get

$$(m(q) - c(l))e^{-\lambda(t)} \frac{r+p}{r} + \xi \left(\alpha l + \frac{\eta q^a}{q^a + b} - \delta q \right), \quad (14)$$

where ξ is the shadow price of environmental quality

The first-order condition with respect to restorative effort yields

$$\xi = \frac{1}{\alpha} c'(l) e^{-\lambda} \frac{r+p}{r}. \quad (15)$$

Let $\xi e^{\lambda} = \beta$, be the adjusted shadow price of environmental quality. The rate of evolution of the shadow price is determined by the no-arbitrage condition as

$$\dot{\xi} = -m'(q) e^{-\lambda} \frac{r+p}{r} + \xi \left(-\eta \frac{abq^{a-1}}{(q^a + b)^2} + \delta + r \right). \quad (16)$$

Therefore, the rate of change of the adjusted shadow price β is given by

$$\dot{\beta} = -m'(q) \frac{r+p}{r} + \beta \left(-\eta \frac{abq^{a-1}}{(q^a + b)^2} + \delta + r + p \right). \quad (17)$$

In steady state, $\dot{\beta} = 0$, implying

$$\beta = \frac{m'(q)}{(-\eta(abq^{a-1}/(q^a + b)^2) + \delta + r + p)} \frac{r + p}{r}. \quad (18)$$

Substituting for β from (15) above, we get the steady-state relationship between restoration efforts and environmental quality as

$$c'(I) = \frac{m'(q)\alpha}{(-\eta(abq^{a-1}/(q^a + b)^2) + \delta + r + p)}. \quad (19)$$

Note that in the no-risk case derived before, the steady-state evaluation of Eq. (6) would yield

$$c'(I) = \frac{m'(q)\alpha}{(-\eta(abq^{a-1}/(q^a + b)^2) + \delta + r)}. \quad (20)$$

Eq. (20) is similar to Eq. (19) except for the extra term p in the denominator of Eq. (19). When restoration efforts are faced with an ever-present constant exogenous risk of invasion, this risk acts as an additional discounting term. Consequently, the above analysis reveals that the steady-state restorative efforts would be lower in the case when there is a risk of relapse as compared to no-risk case.

Note that in the above Eqs. (19) and (20); the increment in resilience from a change in stock serves as an adjustment to the discount rate which is further augmented by the natural rate of decay of environmental quality. From the way this resilience effect has been specified in the model, some interesting implications can be deduced for the optimal restoration path. The level of environmental quality shows a sharp jump upwards once a certain threshold level has been reached. This can be discerned from Fig. 4 below where once point A has been reached the next possible level of environmental quality is above point B, as points A and B both correspond to the same level of environmental quality. Due to this reason, as long as environmental quality is lower than this threshold, the

resilience effect will not be significant. Therefore, the discounting effect brought by a change in resilience due to enhanced environmental stock, kicks in only beyond that threshold level of stock. As a consequence, the change in the optimal steady-state level of restoration effort and environmental quality from some external disturbance in parameters would be significantly higher if the steady state is closer to this threshold. In lay terms, the incentives for restoration efforts are higher; the closer the system is to the flipping threshold. Next we consider the case when risk of relapse may be stock dependent.

2.2. Stock-dependent risk

In the stock dependent risk case, the relationship between the hazard rate and environmental quality is specified as

$$\dot{\lambda} = p(q) \quad (21)$$

The value function can be specified as before:

$$V(q) = \text{Max} \left[\int_0^\infty (m(q) - c(I) + p(q)V(q))e^{-\lambda - rt} dt \right] \quad (22)$$

which can be further expanded for a starting level of environmental quality as

$$V(q_0) = \text{Max} \int_0^\infty (m(q) - c(I))e^{-\lambda - rt} dt + \text{Max} \int_0^\infty (p(q)V(q_0))e^{-\lambda - rt} dt. \quad (23)$$

Rewriting above we get

$$V(q_0) = \text{Max} \frac{\int_0^\infty (m(q) - c(I))e^{-\lambda - rt} dt}{(1 - \int_0^\infty (p(q))e^{-\lambda - rt} dt)} \quad (24)$$

subject to the equations of motion for the hazard function as given by (21) and the environmental stock as given by (1). It is not very straightforward to analytically perform dynamic optimization on the above problem as it is not in a standard format; therefore, we take recourse to numerical simulations.⁷

⁷In order to reduce the problem into a standard framework one may define two more state variables as z_1 and z_2 , whose rate of change is defined as

$$\dot{z}_1 = (m(q) - c(I))e^{-\lambda - rt}, \quad (25)$$

$$\dot{z}_2 = p(q)e^{-\lambda - rt}. \quad (26)$$

Now the above problem in Eq. (23) reduces to:

$$V(q_0) = \text{Maximize} \frac{z_1}{(1 - z_2)} \quad (27)$$

subject to $\dot{q} = \alpha I + \eta \frac{q^a}{q^a + b} - \delta q$, $\dot{\lambda} = p(q)$, $\dot{z}_1 = (m(q) - c(I))e^{-\lambda - rt}$ and $\dot{z}_2 = p(q)e^{-\lambda - rt}$. The current value Hamiltonian of the above problem that would maximize $V(q_0)$ is defined as

$$\frac{z_1}{(1 - z_2)} + \gamma_1 \left(\alpha I + \eta \frac{q^a}{q^a + b} - \delta q \right) + \gamma_2 p(q) + \gamma_3 (m(q) - c(I))e^{-\lambda - rt} + \gamma_4 (m(q) - c(I))e^{-\lambda - rt}. \quad (28)$$

The first-order conditions along with the equations of motion for the co-state variable would yield a time path for the restorative efforts and the environmental quality. Though, this approach is presented here for academic

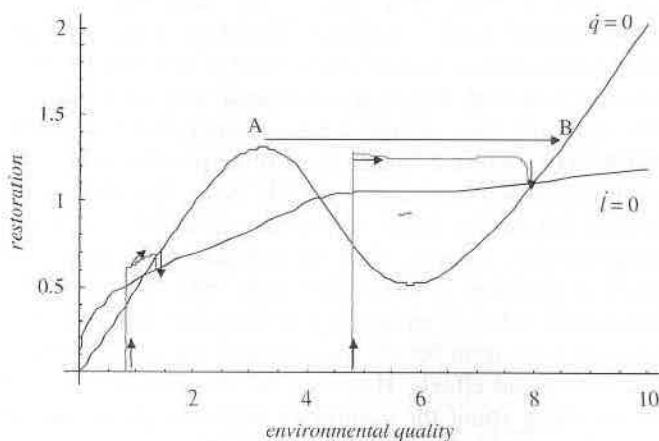


Fig. 4. Time paths of restorative efforts from two starting levels of environmental quality when hysteresis effect is substantial ($\eta = .06$, $r = .03$ and $q_0 = .8$ and 4.8). Note: The direction of evolution of environmental quality from their starting levels is illustrated through the arrows. The transition path has been solved through discrete dynamic optimization using GAMS.

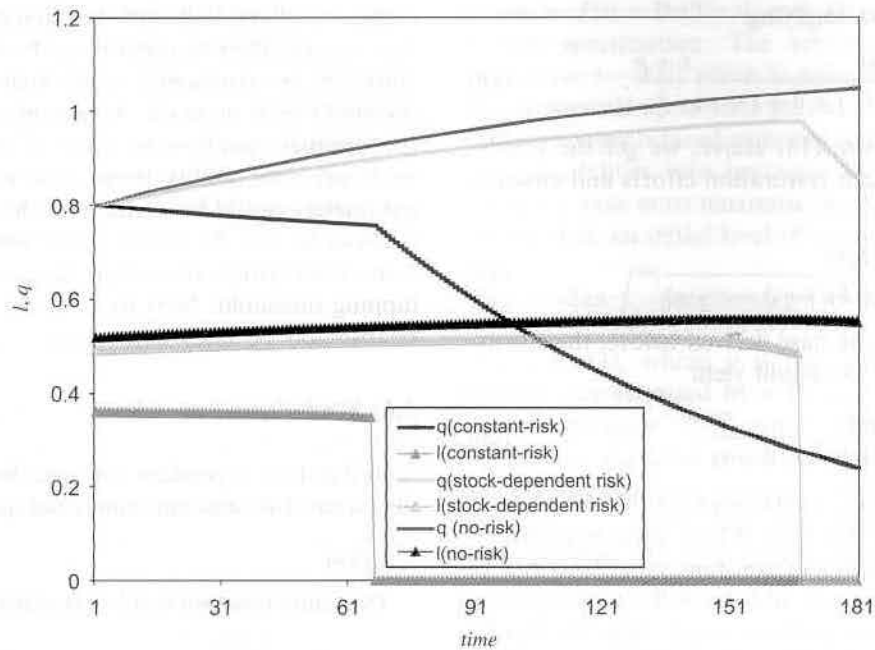


Fig. 5. Restoration and quality levels under no-risk, constant-risk and quality stock-dependent-risk. Note: In the above figure, restoration is marked with thicker lines whereas environmental quality is shown in thinner lines.

Fig. 4 shows the time paths of environmental quality from a slight increase in the hysteresis effect of $\eta = .06$ from the base case (see Appendix A for base case values) and with a lower discount rate of $r = .03$. We can see that the higher steady-state equilibrium becomes attainable as the starting level of environmental quality increases ($q_0 = 4.8$).

Next we compare the time paths of restorative efforts under constant and environmental quality-dependent risk of failure with the no risk case. In the constant risk case, the hazard rate is assumed to be .1, whereas in the environmental quality-dependent risk case the hazard function is defined as

$$p = p_0(1 - \beta \times q(t)), \text{ where } p_0 = .01 \text{ and } \beta = .1 \text{ \& } q(t) < 10. \quad (29)$$

In the above equation p_0 is the constant component of the hazard rate. Clarke and Reed (1994) refer to p_0 as the policy-independent component of the hazard rate, as it is an exogenous parameter. Whereas, $(1 - \beta \times q(t))$ is the policy-dependent component that is endogenous to environmental quality in which β is the parameter that determines the effect of stock of environmental quality on hazard rate. Fig. 5 shows the time paths of restorative efforts and environmental quality under the three cases for a starting level of environmental quality of .8. Note that the highest level of environmental quality is attained when there is no risk of failure. Under a constant risk of failure, environmental quality attained is the lowest, whereas the

endogenous-risk case has a higher environmental quality than the constant-risk case. The effect of risk is primarily to discount the future benefits from environmental quality. However, when the risk is stock dependent, environmental quality is increased to capitalize on its risk reducing impacts.

2.3. How much versus how often

When restoration projects are faced with the risk of collapsing back into a degraded state, the question of how much effort to put in becomes important. If the ecosystem keeps collapsing into the degraded state time and again, it may take a long time before the desirable level of environmental stock is attained. Therefore, it may happen that systems which require a low level of restorative effort but are faced with high risks of reversal may take a longer time to reach their steady states as compared to systems which may require a higher level of restorative effort but are faced with a lower level of risk. Note that the risk of project failure has been optimally accounted for in the above models, meaning that the optimal restoration plan involves choosing a restoration path that maximizes the discounted value of expected long-term net benefits. This expected long-term benefit incorporates the risk of failures and subsequent efforts. However, the above models do not say anything about the number of times the project would fail before a steady state is reached. The time taken to reach the steady state when there are no setbacks to the restoration project would differ from the case when setbacks happen. The actual time taken to reach a desirable level of environmental restoration would depend upon the number of times relapses happen before the goal is

(footnote continued)

interest, the numerical simulation approach is much simpler to solve through GAMS.

achieved. This concept is explored further in the setting of the model described above.

Let t^* be the time it takes for the ecosystem to reach the steady-state level of environmental quality without collapsing, when there is a constant risk of reversal back to the initial degraded level.⁸ In the presence of a constant risk of reversal, the expected time $E(t^*)$ taken to reach the steady state would be given by

$$E(t^*) = \int_0^{t^*} \{s + E(t^*)\} p e^{-ps} ds. \quad (30)$$

Note that in the above formulation, once the system reverts back into the degraded state it has to start all over again, and therefore, would take the same amount of expected time thereafter. s is the time at which the restoration project fails, thus sending the system back to its initial level. Also, s ranges from 0 to t^* , and the probability of failure is exponentially distributed with hazard rate p . Moreover, the system faces risk of reversal even after the steady state has been reached, but, by the optimal nature of that steady state it would mean that restorative efforts and environmental stock maximize expected benefits at that level of risk. Integrating the above term we get

$$E(t^*) = \left[-s p e^{-ps} - \frac{e^{-ps}}{p} + E(t^*)(-e^{-ps}) \right]_0^{t^*}. \quad (31)$$

Solving the above equation one derives the expected time as

$$E(t^*) = \frac{e^{pt^*}}{p} - \frac{1}{p} - t^*. \quad (32)$$

Fig. 6 below plots the contours for $E(t^*)$ for a range of values of the hazard rate p and t^* . Note that the expected time taken is much higher when either t^* or p are higher.

The case of stock-dependent risks is slightly complicated. The time taken to reach the steady state without any interruptions is a function of the discount rate, the marginal benefits and costs of restoration, rate of decay of environmental stock and the resilience parameter. For example, a higher rate of discount would require a lower environmental stock and thus would take lesser time to reach as compared to a case when the benefits from environmental stock are higher or the costs of restoration are lower. Whereas, a lower rate of discount would make the resilient state more desirable thus requiring more time to traverse. Similarly, a lower level of p_0 (the constant component of the hazard rate) would make a higher environmental quality attractive. This is shown in Fig. 7 below, where maximum possible level of environmental quality falls with an increase in p_0 . However, it can be numerically verified that the expected time to steady state $E(t^*)$ is actually lower for the case when p_0 is .1

⁸Analytically, in most steady-state problems it may take an infinite amount of time for the system to reach the steady state. But, for practical purposes, t^* can be calculated using dynamic optimization software such as GAMS.

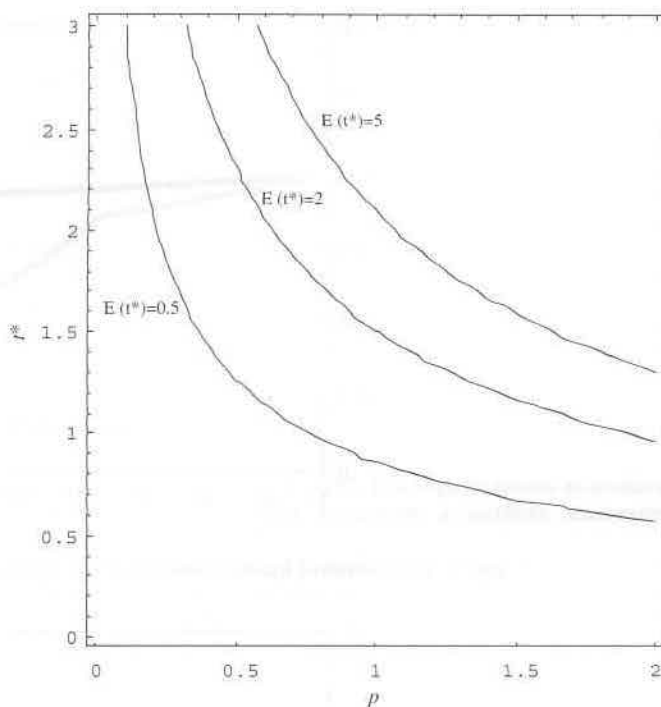


Fig. 6. Contours of expected time to steady state in presence of risk

(420 time units) as compared to the case when p_0 is .5 (37279 time units). That is, even though the desirable level of environmental quality increases with a reduction in risks, the expected time taken to reach it falls. Using a time horizon of 250, the time to reach the steady state without relapses, (t^*), is 211 units for $p_0 = .1$ and 162 units for $p_0 = .5$. This is because the steady-state level of environmental quality falls as p_0 rises. Note that even as (t^*) decreases as risk increases, the expected time to steady state, $E(t^*)$ increases. This is so as an increase in p_0 also increases the number of relapses, thus increasing the total expected time.

If the hazard rate falls sharply with a marginal increase in the environmental stock, it would reduce the expected number of relapses over the same period, as the expected duration for a single relapse increases. This would have an effect of reducing the expected time to reach the steady state. However, the negative effect of environmental quality on the hazard rate would also make it beneficial to strive for a higher environmental quality as the hazard rate comprises one of the elements of adjusted discount rate as derived in Eq. (14) above. This is shown in Fig. 8 below where an increase in θ , parameter which determines the influence of environmental quality stock on hazard rate, leads to an increase in the maximum possible environmental stock. It can also be numerically verified that the expected time to steady state, $E(t^*)$, decreases as θ increases even though the time taken to reach the steady state without a relapse, (t^*), is higher for higher levels of θ .

The net impact that a stock of environmental quality could have on project success through its influence on hazard rate is slightly complicated. On one hand, it offers the incentive to attain higher stock of environmental

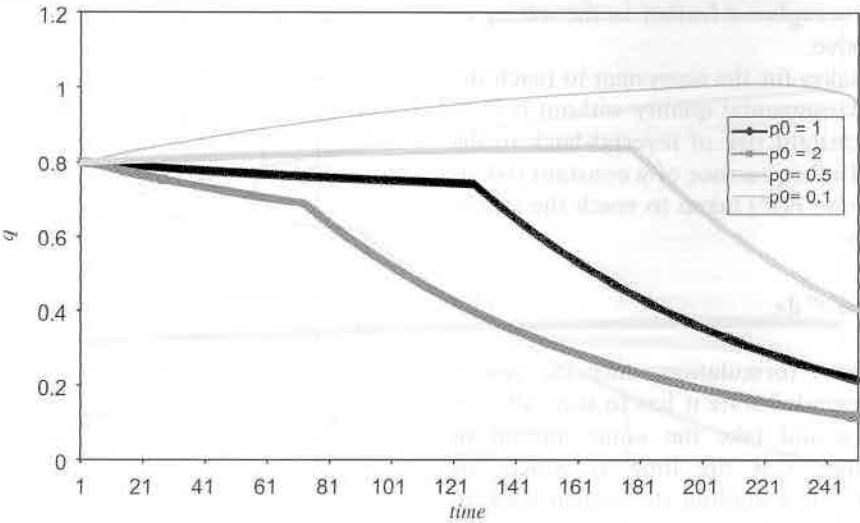


Fig. 7. Environmental quality levels under varying levels of policy-independent component of the hazard rate (p_0).

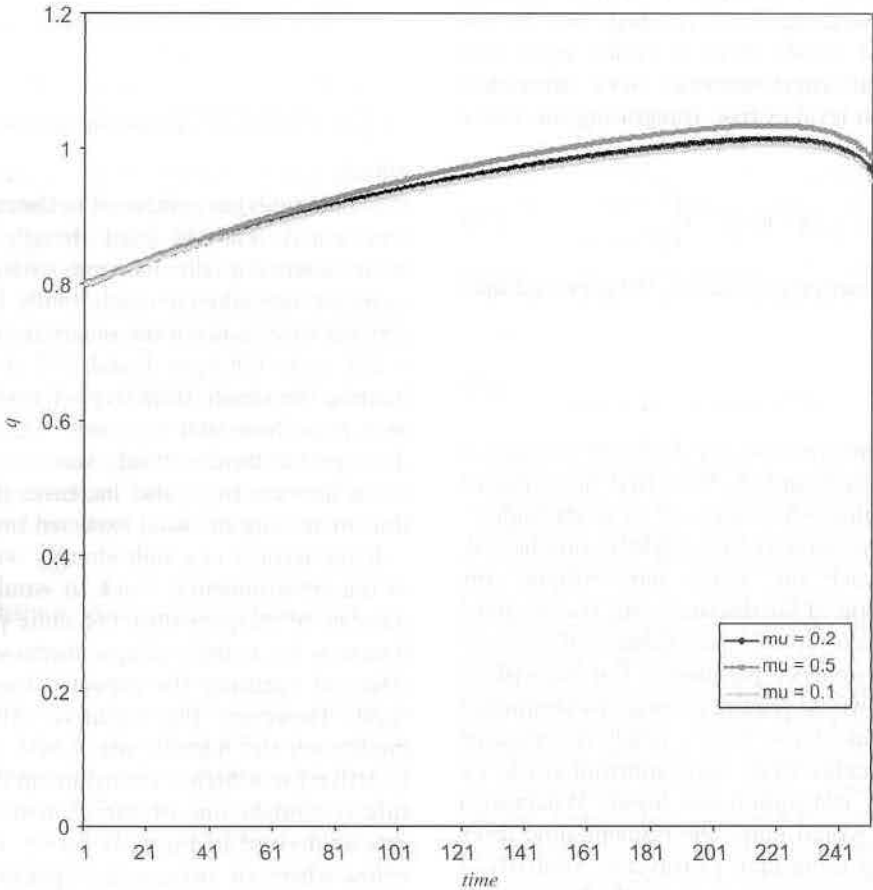


Fig. 8. Environmental quality levels under θ (μ).

quality, as a higher quality yields direct utility and also reduces the risk of relapse. On the other hand, a higher goal for environmental stock also means that a higher restoration effort is required to reach there. If costs are convex in restoration efforts, restoration efforts may need to be stretched over a longer period of time. Consequently, the more time that is required for reaching the steady state, the

higher would be the expected number of relapses. Therefore, a tradeoff between how much quality to strive for and how many failures in order to reach it is highlighted in the case of stock-dependent risks of restoration.

The issue of expected time to steady state is important to policy makers as one important goal of restoration projects is to bring the system back to a level at which it could be

exploited for direct economic uses. In the case when consumption of environmental quality leads to a reduction in its stock, additional restorative efforts will need to be taken in order to maintain the optimal steady-state level.

3. Conclusion

In this paper, the role of restoration measures in improving environmental quality was looked at through the application of the concept of resilience. Optimal restoration efforts were derived when environmental quality impacted the risk of failure of restoration projects. It was shown that the environmental and economic parameters determined the desirability of the level of resilience, and a highly resilient environment might not always be either feasible or desirable.

The tradeoff between the extent of restoration and the number of failures was derived. It was shown that the expected time to reach the desirable state in the event of multiple relapses is a function of both the hazard rate (p) and the time taken to reach the steady state without a relapse (t^*). This relationship between p and t^* is convex. This implies that the expected time to steady state under the possibility of relapses could be same for high risks of collapse but lower t^* and low risks of collapse but a higher t^* . Note that t^* could be low due to the influence of several factors such as the discount rate, benefits and costs of restoration, etc. It also turned out that no straightforward derivation of expected time to steady state is possible when risks are stock dependent.

There exist several other challenges to restoration projects. Some even oppose the idea of human interferences in degraded environments. Holling and Meffe (1996) in an influential paper argue in favor of natural disturbances that help build the resilience of a system rather than human interventions that shield it against them. Conflicting opinions exist towards the choice of restoration tools, with some even claiming that exotic species themselves may play a beneficial role in restoration of the environment as human interferences lead to further disturbances (Antonio and Meyerson, 2002). However, when restorative options are available and their advantages are clear, it may be worthwhile to apply them, especially when the benefits from their restoration span economic and environmental goods. In case of environments invaded by alien species, the need for restoration is an urgent one, as invasive species pose serious threats of extinction of valuable native ecosystems. It must also be kept in mind that restoration projects need to incorporate longer time horizons and utilize the resilience effects offered by higher levels of environmental quality in order to be able to ward off current and future threats of invasion.

Appendix A

See Tables A1 and A2.

Table A1
Parameters used for simulation

Parameters	a	b	γ	δ	n	η	α	ρ
Values	6	10,000	2	.01	4	.05	.02	.05

Table A2
Functional forms of cost and benefit functions

$$m(q) = q^{\eta}$$

$$c(I) = I^{\alpha}$$

References

- Antonio, D.C., Meyerson, L.A., 2002. Exotic plant species as problems and solutions in ecological restoration: a synthesis. *Restoration Ecology* 10 (4), 703–713.
- Butler, D.W., Fairfax, R.J., 2003. Buffel grass and fire in a Grigg and Brigalow woodland: a case study from Central Queensland. *Ecological Management and Restoration* 4 (2), 120–125.
- Carpenter, S.R., Cottingham, K.L., 1997. Resilience and restoration of lakes. *Conservation Ecology* [online] 1 (1), 2 Available from the Internet. URL: <<http://www.consecol.org/vol1/iss1/art2>>.
- Clarke, R.H., Reed, W.J., 1994. Consumption/pollution tradeoffs in an environment vulnerable to pollution related catastrophic collapse. *Journal of Economic Dynamics and Control* 18, 991–1010.
- Didham, R.K., Watts, C.H., Norton, D.A., 2005. Are systems with strong underlying abiotic regimes more likely to exhibit alternative stable states? *Oikos* 100 (2), 409–416.
- Eiswerth, M.E., Johnson, W.S., 2002. Managing non-indigenous invasive species: insights from dynamic analysis. *Environmental and Resource Economics* 23, 319–342.
- Eiswerth, M.E., Van Kooten, G.C., 2002. Uncertainty, economics, and the spread of an invasive plant species. *American Journal of Agricultural Economics* 84, 1317–1322.
- Glowka, L., Burhenne-Guilmin, F., Synge, H., 1994. A Guide to Convention on Biological Diversity. Environmental Policy and Law Paper 30; IUCN Publication Services Unit, Cambridge, UK.
- Hatzistathis, A., Goudelis, G., Zagats, T., 1998. Conclusions from the reforestation in Greece and perspectives. *East-West Series in Economics, Business and the Environment* 1 (2), 61–70.
- Holling, C.S., Meffe, G.K., 1996. Command and control and the pathology of natural resource management. *Conservation Biology* 10, 326–327.
- Horan, R.D., Perrings, C., Lupi, F., Bulte, E.H., 2002. Biological pollution prevention strategies under ignorance: The case of invasive species. *Am. J. Agric. Econ.* 84 (5), 1303–1310.
- Keohane, N.O., Van Roy, B., Zeckhauser, R.J., 2000. Controlling stocks and flows to promote quality: the environment, with applications to physical and human capital. NBER Working Paper, <<http://www.nber.org/papers/w7727.pdf>>.
- Knowler, D., Barbier, E., 2000. The economics of an invasive species: a theoretical model and case study application. In: Perrings, C., Williamson, M., Dalmazzone, S. (Eds.), *The Economics of Biological Invasions*. Edward Elgar Publication, Cheltenham, UK.
- Knowler, D., Barbier, E., 2005. Importing exotic plants and the risk of invasion: are market-based instruments adequate? *Ecological Economics* 25, 341–354.
- Maler, K.G., Xepapadeas, A., de Zeeuw, A., 2003. The economics of shallow lakes. *Environmental and Resource Economics* 26 (4), 603–624.
- Mayer, A.L., Rietkerk, M., 2004. The dynamic regime concept for ecosystem management and restoration. *BioScience* 54 (11), 1013–1020.
- Olson, L., Roy, S., 2002. The economics of controlling a stochastic biological invasion. *American Journal of Agricultural Economics* 84, 1311–1316.

- Perrings, C., 2002. Biological invasions in aquatic systems: the economic problem. *Bulletin of Marine Science* 70 (2), 541–552.
- Perrings, C., 2003. Mitigation and adaptation strategies in the control of biological invasions. Paper Presented at the Fourth BIOECON Workshop, Venice, August.
- Perrings, C., Walker, B., 1997. Biodiversity, resilience and the control of ecological-economic systems: the case of fire-driven rangelands. *Ecological Economics* 22 (1), 73–83.
- Perrings, C., Walker, B., 2004. Conservation in the optimal use of rangelands. *Ecological Economics* 49, 119–128.
- Pimental, D., Lach, L., Zuniga, R., Morrison, D., 1991. Environmental and Economic Costs Associated with Non-Indigenous Species in the United States. College of Agriculture and Life Sciences, Cornell University, Ithaca, NY <http://www.news.cornell.edu/releases/jan99/species_costs.htm>.
- Pimental, D., Lach, L., Zuniga, R., Morrison, D., 2000. Environmental and economic costs of non-indigenous species in the United States. *Bioscience* 50 (1), 53–65.
- Pimm, S.L., 1984. The Complexity and Stability of Ecosystems. *Nature, Non-Linear Economic Dynamics*. Springer, Berlin.
- Reed, W.J., 1988. Optimal harvesting of a fishery subject to random catastrophic collapse. *IMA Journal of Mathematics Applied in Medicine and Biology* 5, 215–235.
- Reed, W.J., Heras, H.E., 1992. The conservation and exploitation of vulnerable resources. *Bulletin of Mathematical Biology* 54 (2/3), 185–207.
- Rouget, M., Richardson, D.M., Nel, J.L., Le Maitre, D.C., Ego, B., Mgid, T., 2004. Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions* 10, 475–484.
- Settle, C., Shogren, J.F., 2002. Modeling native-exotic species within Yellowstone lake. *American Journal of Agricultural Economics* 84 (5), 1323–1328.
- Shogren, J., 2000. Risk reduction strategies against the explosive invader. In: Perrings, C., Williamson, M., Dalmazzone, S. (Eds.), *The Economics of Biological Invasions*. Edward Elgar Publication, Cheltenham, UK, pp. 56–69.
- Suding, K.N., Gross, K.L., Houseman, G.R., 2004. Alternative states and positive feedbacks in restoration ecology. *Trends in Ecology and Evolution* 19.
- Van Wilgen, B.W., de Wit, M.P., Anderson, H.J., Le Maitre, D.C., Kotze, I.M., Ndala, S., Brown, B., Rapholo, M.B., 2004. Costs and benefits of biological control of invasive alien plants: case studies from South Africa. *South African Journal of Science* 100, 113–115.

Evaluation of the temporal variations of air quality in Taipei City, Taiwan, from 1994 to 2003

Shuenn-Chin Chang, Chung-Te Lee*

Graduate Institute of Environmental Engineering, National Central University, 300 Zhongda Rd, Zhongli 32001, Taiwan (ROC)

Received 3 March 2006; received in revised form 28 November 2006; accepted 12 December 2006

Available online 12 February 2007

Abstract

Data collected from the five air-quality monitoring stations established by the Taiwan Environmental Protection Administration in Taipei City from 1994 to 2003 are analyzed to assess the temporal variations of air quality. Principal component analysis (PCA) is adopted to convert the original measuring pollutants into fewer independent components through linear combinations while still retaining the majority of the variance of the original data set. Two principal components (PCs) are retained together explaining 82.73% of the total variance. PC1, which represents primary pollutants such as CO, NO_x, and SO₂, shows an obvious decrease over the last 10 years. PC2, which represents secondary pollutants such as ozone, displays a yearly increase over the time period when a reduction of primary pollutants is obvious. In order to track down the control measures put forth by the authorities, 47 days of high PM₁₀ concentrations caused by transboundary transport have been eliminated in analyzing the long-term trend of PM₁₀ in Taipei City. The temporal variations over the past 10 years show that the moderate peak in O₃ demonstrates a significant upward trend even when the local primary pollutants have been well under control. Monthly variations of PC scores demonstrate that primary pollution is significant from January to April, while ozone increases from April to August. The results of the yearly variations of PC scores show that PM₁₀ has gradually shifted from a strong correlation with PC1 during the early years to become more related to PC2 in recent years. This implies that after a reduction of primary pollutants, the proportion of secondary aerosols in PM₁₀ may increase. Thus, reducing the precursor concentrations of secondary aerosols will be an effective way to lower PM₁₀ concentrations.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Pollutant trend; Primary pollutants; Secondary pollutants; Principal component analysis; PM₁₀

1. Introduction

Due to successful economic development leading to dense population and an increasing number of motorized vehicles, several cities in Taiwan are facing the problem of degrading air quality. The Taiwanese government implemented new ambient air-quality standards in 1992, and in recent years, various governmental agencies have dedicated themselves to improving air quality with obvious accomplishments. For example, the percentage of PSI (pollutant standards index) greater than 100, which is adapted from the United States Environmental Protection Agency (US EPA) as a criterion for ambient air quality, decreased from

4.01% in 1994 to 1.96% in 2003 in Taipei City. Among the criteria pollutants, O₃ and PM₁₀ are the two major sub-index pollutants that cause poor air quality as determined by the PSI being greater than 100 (Taiwan EPA, 2005).

In order to improve air quality and to assess air pollution control measures, the variations of pollutant concentrations must be understood. Generally, meteorological changes will result in greater variations of air pollutant concentrations than variations of emissions. For example, most of the daily variations in PM₁₀ concentrations in the Netherlands are shown to be mainly caused by large-scale meteorological processes and less by variations in emissions (van der Wal and Janssen, 2000). In addition to the accumulation of pollutants due to poor atmospheric dispersion, the transboundary transport of pollutants and secondary aerosols may contribute significantly to

*Corresponding author. Fax: +886 3 4221602.

E-mail address: ctlee@cc.ncu.edu.tw (C.-T. Lee).

poor regional air quality. In Hong Kong, the compositions of winter particulates show that they are mainly transported from southeastern China (Ho et al., 2003). In the rural regions of eastern Spain, both PM_{10} and TSP concentrations demonstrate seasonal variations. In Eastern Spain, the highest particulate concentration in summer is caused by sandstorms from Africa; the second highest concentration is related to secondary aerosols caused by regional ozone episodes (Rodríguez et al., 2002). In Taiwan, PM_{10} concentrations have relatively greater variations with high PM_{10} concentrations in March and May, mainly due to the dusts brought over by the northeastern wind from mainland China after the outbreak of sandstorms (Yang, 2002; Lin, 2001; Chou et al., 2005; Lin et al., 2005).

Ozone is one of the major pollutants causing poor air quality in Taipei City, like many of the metropolises in the world. Ozone concentrations depend on photochemical reactions of precursors, regional-scale horizontal transportation, stratospheric transportation, and surface subsidence, as well as the titration effect (Tao et al., 2005). Vukovich (1994) pointed out that more than 65% of the year-to-year variation in ozone concentrations is contributed by meteorological changes. In Taipei City, the occurrence of high O_3 concentration is always associated with strong insolation in the morning, a rapid temperature rise, less cloud cover, and/or no precipitation (Liu et al., 1994).

Principal component analysis (PCA) has been extensively used in meteorological science and air pollution research, such as in analyzing variations of particulate sulfate, meteorological and air-quality data, source apportionment, atmospheric chemical reactions, atmospheric dispersion, and the effect of seasonal variations on air pollution (Henry and Hidy, 1979; Henry and Hidy, 1982). PCA is also used in studying spatial and temporal variations (Yu and Chang, 2001; Eder, 1989; Eder et al., 1993; Ashbaugh et al., 1984; Hooper and Peters, 1989; Lehman et al., 2004). Buhr et al. (1992) applied PCA to the analysis of ratios of atmospheric CO , NO_x , and SO_2 to determine the sources of various pollutants in order to re-evaluate the rationality of the source inventory. Moreover, Buhr et al. (1995) used PCA to resolve sources of NO_x and examine the influence of various pollution sources on ozone variations. Vukovich and Sherwell (2003) applied PCA to the analysis of the correlation between daily ozone maximum concentrations ($O_{3, \max}$) and meteorological parameters.

In this study, PCA is performed on the daily mean concentrations of SO_2 , CO , NO_x , PM_{10} , and $O_{3, \max}$ to identify primary and secondary pollutants, as well as to examine their long-term tendencies and seasonal variations in Taipei City. In order to exclude the abnormal impact of high PM_{10} events on long-term air-quality tendencies (Hosiokangas et al., 1999), PCA is also used to sort out transboundary transportation, such as sandstorm events in this study.

2. Method

2.1. Area of study

Taipei City, with a population of 2.62 million living in 272 km², is located in a basin in northern Taiwan (Fig. 1) and is the target area of this study. With 700,000 cars and more than 1 million motorcycles, the city has 6536 motorized vehicles per square kilometer. The emission inventory from the Taiwan Environmental Protection Administration (TEPA) shows that except for the emissions from three municipal incineration plants, there is no significant stationary source in Taipei City; most of the CO and NO_x come from vehicle emissions which constitute 98% and 77% of the total CO and NO_x emissions, respectively.

2.2. Sources of data

TEPA established the TAQMN (Taiwan Air-Quality Monitoring Network) in September 1993, which includes 74 monitoring stations around the island. There are five ambient air-quality monitoring stations within Taipei City, which are located on top of school buildings, 15 meters above the ground. The monitoring data used in this study include hourly values and a daily average of SO_2 , CO , NO_x , PM_{10} , O_3 , and wind speed.

The data are collected using commercial monitoring instruments designated by the US EPA as an equivalent or reference method and manufactured by the US Thermo Environmental Instruments, Inc. (Franklin, MA, USA). The instruments included the following: model 43 pulsed fluorescent SO_2 analyzer (EQSA-0276-009) for SO_2 , model 48 gas filter correlation ambient CO analyzer (RFCA-0981-054) for CO , model 42 chemiluminescence NO – NO_2 – NO_x analyzer (RFNA-1289-074) for NO_x , model 49 UV Photometric Ambient O_3 Analyzer (EQOA-0880-047) for

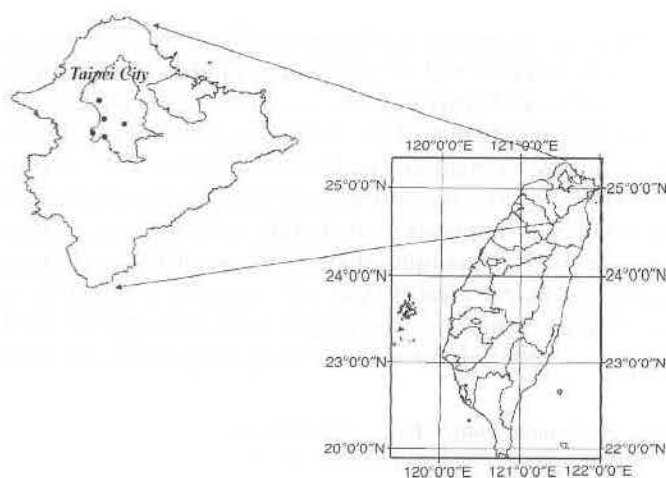


Fig. 1. Geographic locations of the five air-quality monitoring stations (shown in dots) in Taipei City, Taiwan.

O_3 , and model 650 (formerly: Wedding & Associates) PM_{10} Beta gauge automated particle monitor (EQPM-0391-081) for PM_{10} . The TAQMN instruments were well operated and maintained to ensure data quality. Scheduled quality control procedures included daily zero and span checks, biweekly precision checks, quarterly multiple-point calibrations, and data validation. Additionally, TEPA contracted an independent quality assurance program and instrument performance audit to assure data quality.

2.3. Data processing

In order to avoid the influence of diurnal pollutant variations on the study and obtain spatial representation, we calculated the average of the daily pollutant averages obtained at all five stations for a long-term trend study. In addition, only daily averages with more than 16 effective hours per day were considered valid in this study.

For the study of ozone characteristics, we adopted the spatial average of $O_{3, \max}$. Eder et al. (1993) used $O_{3, \max}$ to analyze the spatial and temporal variability of ozone over non-urban regions in the eastern US. This concept is based on the consideration that an extreme concentration of ozone has a much higher adverse impact on humans or the environment. In a day, $O_{3, \max}$ frequently occurs between 1:00 and 3:00 p.m. when ozone distribution is rather uniform in the boundary layer. Thus, the data collected near the surface represent the average concentration for the entire layer. At night, when photochemical reactions stop, the NO emitted from motor vehicles will titrate daytime ozone. Hence, using $O_{3, \max}$ instead of the daily average will better index the seriousness of air pollution.

2.4. Statistical analysis

Factor analysis is useful in identifying variations of recurring and independent modes if they exist in the data set (Vukovich and Sherwell, 2003). In this study, we performed PCA to determine the variation patterns of five pollutants, i.e., SO_2 , CO, NO_x , PM_{10} , and O_3 , from 1994 to 2003, based on 18,250 records (5×3650 daily averages). The time series plots of all the variables analyzed in this study are shown as Fig. 2. PCA is carried out using commercial software (Statistica Rel. 6.0, Statsoft Incorporated, 2001) installed in a personal computer.

The main objective of PCA is to convert the original variables into fewer independent components through linear combinations while still retaining the majority of the variance of the original data set. However, the exact number of retained components remains a subject of controversy (Lehman et al., 2004). When conducting the study of $O_{3, \max}$ for non-urban regions in the eastern United States, Eder et al. (1993) applied the Scree test, which looks for a discontinuity in the curve defined by plotting the percentage variance explained for each component in the order extracted, to retain six principal components (PCs) with 64% cumulative variance of the

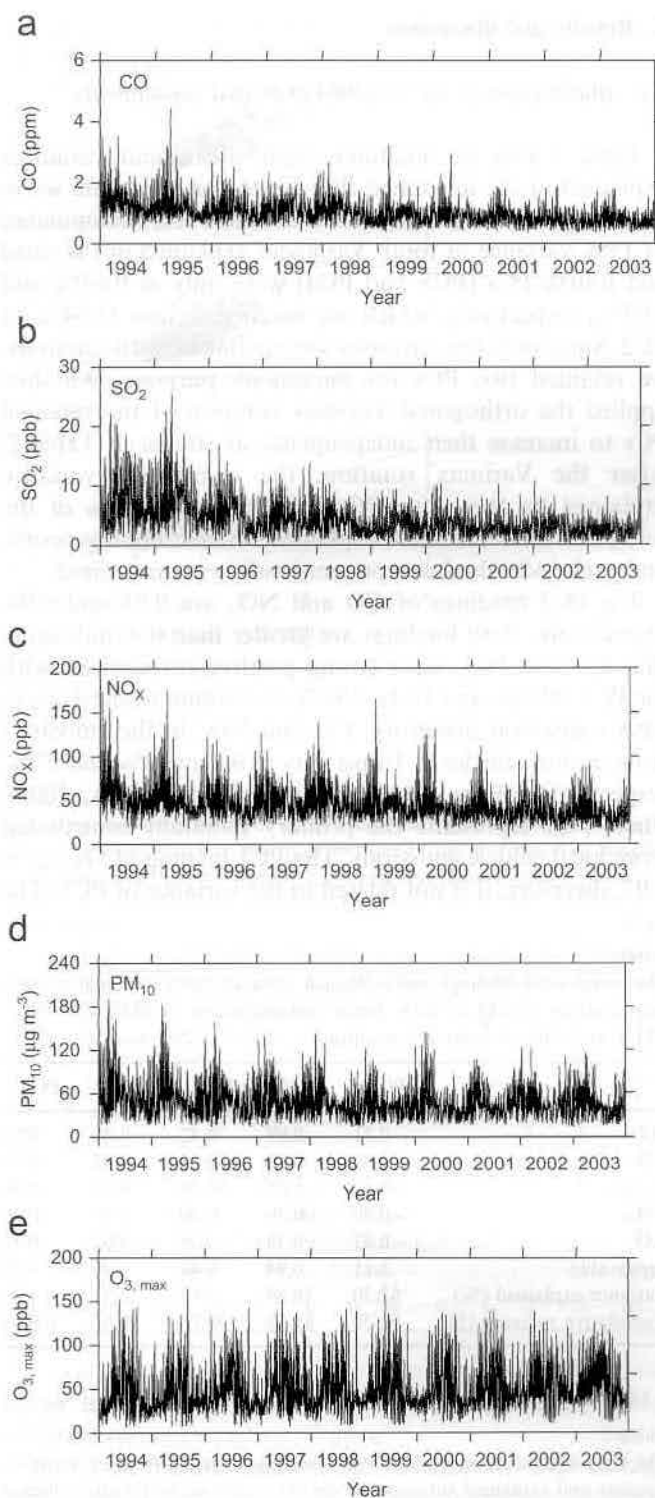


Fig. 2. Time series of the pollutants adopted in the air-quality analysis of Taipei City from 1994 to 2003 (a) CO, (b) SO_2 , (c) NO_x , (d) PM_{10} and (e) $O_{3, \max}$.

total. Vukovich and Sherwell (2003) used PCA to analyze $O_{3, \max}$ with meteorological factors for the Baltimore–Washington corridor and adopted the 75% variance rule (the explained cumulative variance of retained components exceeding 75% of the total).

3. Results and discussion

3.1. Implication of the retained principal components

Table 1 lists the loadings, eigenvalues, and variances explained of the unrotated PCs for the original data set in this study. The first two PCs (PC1 and PC2) accumulated 81.13% variance of total. Variances explained in the third and fourth PCs (PC3 and PC4) were only at 9.63% and 7.91%, respectively, which are much less than 18.84% of PC2. Since only five variables were included in the analysis, we retained two PCs for parsimony purpose. We then applied the orthogonal Varimax rotation of the retained PCs to increase their independence as shown in Table 2. After the Varimax rotation, the cumulative variance explained by these two PCs remained at 81.13% of the total, but their variances explained become homogeneous, and each individual component loading is maximized.

The PC1 loadings of CO and NO_x are 0.95 and 0.94, respectively. Both loadings are greater than 0.9, indicating that CO and NO_x have strong positive correlations with the PC1 (Henry and Hidy, 1982). According to the Taiwan EPA's emission inventory, CO and NO_x in the emissions from motor vehicles in Taipei City make up 98% and 77%, respectively, of the total emissions (Taiwan EPA, 2005). Thus, PC1 represents the primary pollutant contributed from local vehicle emissions. The PC1 loading of O_3 is 0.07; therefore, it is not related to the variance of PC1. The

PC2 loading of O_3 is 0.95, which represents the contribution of photochemical reactions. The PC1 loading of SO_2 is 0.72 implying that most SO_2 in Taipei City has a variance similar to CO and NO_x ; however, some SO_2 might be contributed by sources other than local motor vehicles, e.g., those from neighboring Keelung City and even from transboundary transportation (Chang et al., 2000; Lin, 2001; Lin et al., 2005; Chang and Lee, 2006a). It should be noted that the PC1 loading of PM_{10} is 0.64, but its PC2 loading is slightly lower at 0.52. The PM_{10} loading in both PC1 and PC2 indicate that in addition to local sources of emissions, other factors such as the photochemical reactions in producing secondary aerosols (Rodríguez et al., 2002), and yellow sand (Lin, 2001; Lin et al., 2005) also cause PM_{10} to vary significantly.

3.2. Separation of transboundary transport from the long-term trend in PM_{10}

Recently, Asian sandstorms were determined to have caused high PM_{10} concentrations in Taiwan (Lin, 2001; Lin et al., 2005). In order to track down the control measures put forth by the authorities, the transboundary transport of PM_{10} from outside sources must be eliminated from the analysis of the long-term trend or seasonal variation of PM_{10} in Taipei City.

Previous results show that a significant increase of PC2 loading in PM_{10} may indicate contributions from secondary aerosols produced in strong photochemical activities. Based on our previous study (Chang and Lee, 2006b), the O_3 at 60 ppb is an appropriate index for separating low from high photochemical activity in Taipei City. By using this criterion, we think that the enhancement of secondary aerosol in PM_{10} might be found for the level of O_3 higher than 60 ppb, while less PM_{10} is contributed from secondary aerosol for O_3 less than 60 ppb. Furthermore, it is noted that local pollutants tend to accumulate under low wind speed; in contrast, the yellow sand is transported by Mongolian high-pressure systems with strong winds to Taipei City. Thus, scatter plots of PM_{10} and PC1 scores were segregated by ozone concentration and wind speed (Fig. 3).

When the wind speed is less than 1 m s^{-1} , or when O_3 is higher than 60 ppb and the wind speed is less than 2 m s^{-1} , PM_{10} PC1 scores are larger than for at 0.64 as shown in Fig. 3(a), (b) and (d), and Table 2. When O_3 is less than 60 ppb and the wind speed is higher than 1 m s^{-1} , or when O_3 is higher than 60 ppb and the wind speed is higher than 2 m s^{-1} , there is a weaker correlation between PM_{10} and PC1 scores than for the full data set at 0.64 as shown in Fig. 3(c), (e) and (f), and Table 2. The correlation is the least when O_3 is less than 60 ppb, and the wind speed is greater than 2 m s^{-1} as shown in Fig. 3(e). As PC1 represents local emission contribution, a smaller correlation coefficient between PM_{10} and PC1 scores indicates that the high PM_{10} concentrations are affected

Table 1

The component loadings and explained variance from principal component analysis (PCA) of daily mean concentrations of SO_2 , CO, NO_x , PM_{10} , and daily maximum concentration of O_3 (18,250 records)

	PC1	PC2	PC3	PC4	PC5
SO_2	-0.81	-0.03	0.42	0.40	0.02
CO	-0.91	0.28	-0.23	0.02	-0.19
O_3 , max	-0.46	-0.83	-0.28	0.10	0.01
PM_{10}	-0.80	-0.20	0.30	-0.47	0.00
NO_x	-0.87	0.36	-0.28	-0.02	0.17
Eigenvalue	3.11	0.94	0.48	0.40	0.07
Variance explained (%)	62.30	18.84	9.63	7.91	1.33
Cumulative variance (%)	62.30	81.13	90.76	98.67	100.00

Table 2

The first two component loadings resulting from PCA after Varimax rotation and explained variance for the five major air pollutants collected in Taipei City from 1994 to 2003

	PC1	PC2
SO_2	0.72	0.37
CO	0.95	0.13
O_3 , max	0.07	0.95
PM_{10}	0.64	0.52
NO_x	0.94	0.04
Eigenvalue	2.73	1.33
Variance explained (%)	54.60	26.53
Cumulative variance (%)	54.60	81.13

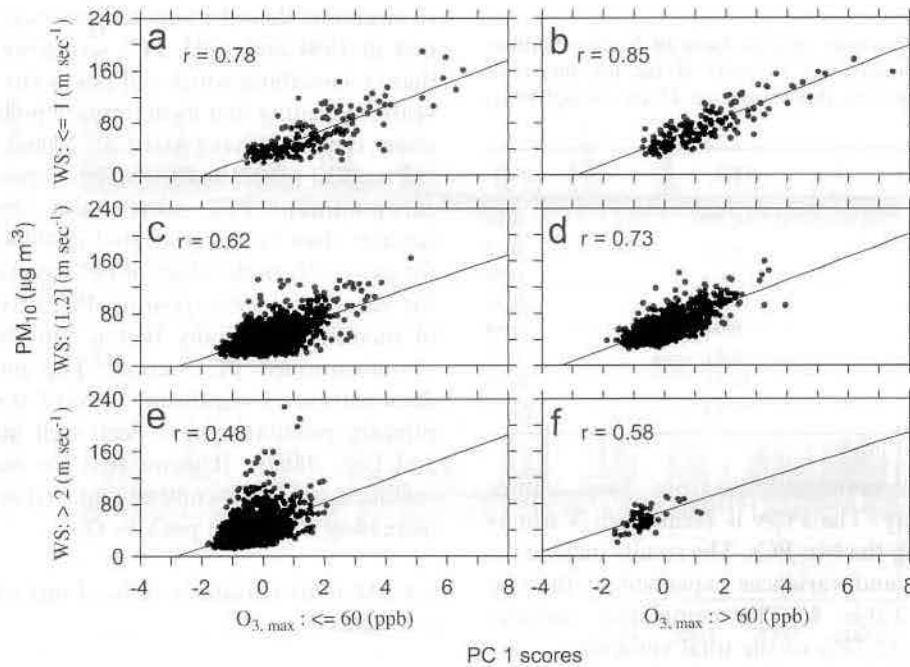


Fig. 3. Scatter plots of PM_{10} associated with PC1 scores at different O_3 and WS levels.

by transboundary transport and not by local emissions sources.

The PC scores are standardized with zero average and a standard deviation of 1. Higher PC1 scores show a greater influence of local pollution emissions. According to Fig. 3(a) and (b), the PM_{10} concentrations increase with high PC1 scores at wind speed less than 1 m s^{-1} no matter what $O_{3, \text{max}}$ levels are, and the correlation coefficients are 0.78 and 0.85, respectively. This indicates that the PM_{10} concentrations are accumulated mainly from local source contributions at low wind speeds for different $O_{3, \text{max}}$ levels. In contrast, the correlation coefficient between the PM_{10} concentrations and PC1 scores decreases with higher wind speeds than 1 m s^{-1} at different $O_{3, \text{max}}$ levels as shown in Fig. 3(c)–(f). It is acknowledged that a portion of PM_{10} is contributed by secondary aerosols when $O_{3, \text{max}}$ is higher than 60 ppb. For $O_{3, \text{max}}$ being less than 60 ppb, the transboundary transport of Asian yellow sand instead of photochemical secondary aerosols may contribute to PM_{10} concentrations. Fig. 3(c) and (e) show that when PC1 scores exceed 1.5, PM_{10} increases with higher PC1 scores in a linear fashion. However, when PC1 scores are less than 1.5, the PM_{10} concentrations have more scattered relationships with PC1 scores. Since lower PC1 scores show less influence of local emissions, and higher $O_{3, \text{max}}$ concentrations may result in secondary aerosol formation, the cases of high PM_{10} concentrations but with lower PC1 scores and lower $O_{3, \text{max}}$ concentrations may be classified as originating from transboundary transport such as from Asian yellow sand. In order to reduce the influence of high PM_{10} concentrations caused by transboundary transport on the long-term trend, a criterion of the PC1 score being less than 1.5 and $O_{3, \text{max}}$

Table 3

The temporal distribution of the 47 days of high PM_{10} concentrations caused by transboundary transport identified from a criterion of the PC1 score being less than 1.5 and $O_{3, \text{max}}$ being less than 60 ppb

Year	Month												Total
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
1994		1		1					1				3
1995			1		1					1			3
1996					3						1		4
1997	1		1										2
1998				1							1	1	3
1999	1	1		2									4
2000	2		3	4	1							1	11
2001	1		4	2									7
2002	2	3	2										7
2003	1										2		3
Total	8	5	11	10	5	0	0	0	0	1	2	5	47

being less than 60 ppb is applied to screen out PM_{10} concentrations exceeding $97 \mu\text{g m}^{-3}$.

Based on the abovementioned criterion, 47 days of high PM_{10} concentrations caused by transboundary transport have been identified, as shown in Table 3. The year with the greatest number of these days was 2000, with 11 days of high PM_{10} concentrations, followed by 2001 and 2002, both with seven. Most of the selected days occurred from January to May and October to December, with the highest monthly numbers being 11 days in March and 10 days in April. This is in agreement with the time period when Taiwan is under the influence of air masses moving from the Asian Continent. After removing these individual cases, the results of the long-term tendencies or seasonal patterns will better

Table 4

The first two component loadings resulted from PCA after Varimax rotation and explained variance for 10 years of the five major air pollutants collected in Taipei City after sorting out 47 days of high-PM₁₀ events

	PC1	PC2
SO ₂	0.73	0.36
CO	0.95	0.12
O _{3, max}	0.08	0.96
PM ₁₀	0.70	0.51
NO _x	0.94	0.04
Eigenvalue	2.81	1.33
Variance explained (%)	56.15	26.58
Cumulative variance (%)	56.15	82.73

represent the source contributions from local source activities in Taipei City. The PCA is rerun with Varimax rotation while retaining the two PCs. The results include the loadings, eigenvalues, and variances explained of the two PCs, as shown in Table 4. The cumulative variance explained amounts to 82.73% of the total variance.

The results of the PCA after deducting the influence of transboundary transport show that the component loadings of 0.95 and 0.94 for NO_x and CO, respectively, are still the two highest PC1 loadings. The component loading of O_{3, max} is 0.08 in the PC1 and 0.96 in PC2. For SO₂, the component loading in the PC1 and the PC2 are 0.73 and 0.36, respectively. These results are similar to those obtained without deducting the influence of transboundary transport. However, the PM₁₀ loading in PC1 increases to 0.70 and slightly decreases to 0.51 in the PC2. This indicates that after deducting the influence of transboundary transport, PC1 will better represent PM₁₀ pollution from local emissions, while the PC2 still corresponds to PM₁₀ produced from photochemical reactions.

3.3. Trend of local and photochemical pollutions

As the magnitude of PC scores is the linear combination of PC loadings from various pollutants (Lehman et al., 2004), positive PC scores indicate that the pollutant concentration for a day is higher than the mean concentration, and vice versa for the negative PC scores. A distance-weighted least-squares smoothing procedure (the influence of individual points decreases with the horizontal distance from the respective points on the curve) (Statistica Rel. 6.0, Statsoft Incorporated, 2001) is used to smooth out the PC scores; the results will provide valuable references for evaluating long-term trends.

Fig. 4 shows the time series of the two rotated PC scores. Fig. 4(a) illustrates the variations of PC1, which represents local pollution. The positive scores occurred mostly in the early 5 years (1994–1998) and decreased yearly, while the negative scores increased annually. Since the average and standard deviation of PC scores are 0 and 1, respectively, for those PC scores three times greater than the standard deviation, therefore, the air pollutant concentration is shown

to be greater than the long-term average. Fig. 4(a) also shows that in 1994 and 1995, PC1 scores were frequently greater than 3 something which did not occur in the most recent 3 years, indicating that local primary pollutants have been well under control (Chang and Lee, 2006a).

Fig. 4(b) plots the PC2 scores representing photochemical pollution. PC2 scores also demonstrate extremes (greater than 3) that appeared in 1994 and 1995. Although the extremely high values of PC2 scores did not increase for the years after 1995, positive PC2 scores show a tendency of increasing annually that is grossly different from the aforementioned PC1 scores. The moderate peak in O₃ demonstrates a significant upward trend even if the local primary pollutants have been well under control (Chang and Lee, 2006b). It seems that the current strategy of air pollution control should be adjusted in order to control the increasing moderate peak in O₃.

3.4. Monthly variations of local and photochemical pollutions

The time series of the two PC scores demonstrate obvious temporal variations, as shown in Fig. 4. If the variation in pollution caused by anthropogenic emission sources in a basin is small such as in Taipei City, seasonal variations of various pollutants will be dominated by changes in meteorological conditions. This is an analogous to the accumulation of pollutants found in Milan due to persistent thermal inversions (Marcazzan et al., 2001). Yang (2002) also reported that under the influence of low temperatures and fewer rainy days in winter, the PM₁₀ concentration was observed to be high in southern Taiwan.

In order to show monthly variations, we made the box plots of daily PC scores in a like month (around 300 values) in Figs. 5(a) and 6(a) and calculated the median of the PC scores for a like day (a 10-value average) over the 10-year period in Figs. 5(b) and 6(b). Using the median instead of the mean will prevent the dominance of any single value in determining the typical day over the 10-year period (Eder et al., 1993). Fig. 5(a) shows a larger fluctuation of PC1 scores from January to April, while the median of PC1 scores shows positive scores at the same period, as shown in Fig. 5(b). After May, most days have negative PC scores. This indicates that incidents of air pollution caused by the local emissions mostly occur from January to April. When Taipei City sits in the warmer and relatively calm region before a cold front, pollutants frequently accumulate as was the case in Tel Aviv, Israel (Graham et al., 2004). An accumulation of pollutants also occurs with the decreasing northeastern wind during the change of seasons. However, the period between late January and early February is exceptional. Chinese New Year holidays always fall within this period, when residents living in Taipei City return to their hometowns, thus causing an empty period of about 6 days and leading to an apparent reduction of primary pollutants. Hence, PC1 scores show uncommon negative values from the end of late January to early February.

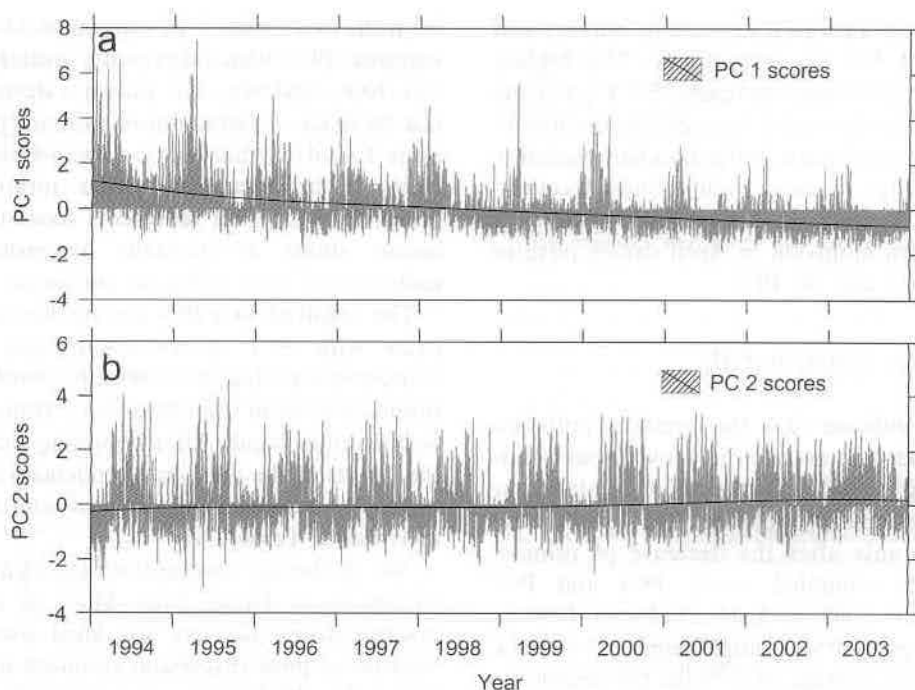


Fig. 4. Time series of the daily standardized principal component scores for the two retained principal components. The bold line represents the distance-weighted least-squares smoothing curve of PCs.

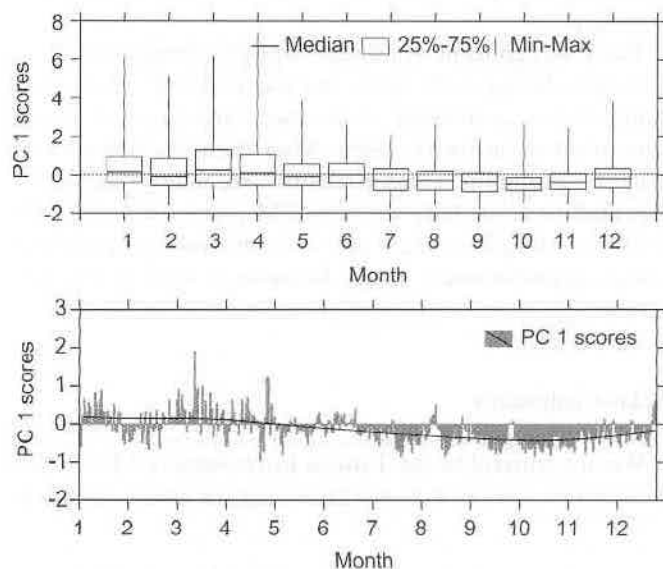


Fig. 5. Seasonal variations of the first component (PC1) scores over a 10-year period (a) box plots of daily distribution in a like month. (b) the time series and the median PC1 scores for a like day from 1994 to 2003. The bold line represents the distance-weighted least-squares smoothing curve of PC1.

During the summer months; however, the strong south-western wind or thermal convection provides favorable conditions for pollutants to disperse, resulting in relatively lower pollutant concentrations. When fall begins, the wind speed picks up, and the meteorological conditions become favorable for dispersing pollutants such that they have difficulty accumulating locally.

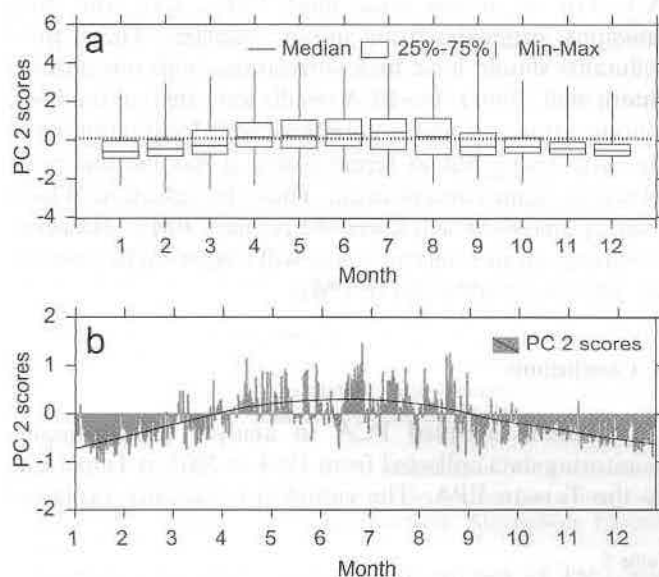


Fig. 6. Seasonal variations of the second component (PC2) scores over a 10-year period (a) box plots of daily distribution in a like month. (b) the time series and the median PC2 scores for a like day from 1994 to 2003. The bold line represents the distance-weighted least-squares smoothing curve of PC2.

Fig. 6(a) indicates that there are larger variations of PC2 scores from April to August, while the median of PC2 scores shows positive scores at the same time period, as shown in Fig. 6(b). After April, the temperature rises, and high temperatures favor the formation of ozone. According to the statistics published by the Central Weather Bureau (CWB <http://www.cwb.gov.tw/>), the lowest averages of

monthly temperatures in Taiwan take place in January and February at 13.3 and 13.6 °C, respectively. The highest average of monthly temperature can reach 25.9 °C, creating meteorological conditions favorable to ozone formation. In the summer, from May to August, both the solar insolation and temperature are high, causing high ozone concentrations and positive PC2 scores. It is worth noting that the weakening northeastern monsoon in April causes positive scores for both the PC1 and the PC2.

3.5. Source contribution shifting in PM_{10}

Previous analyses indicate that the primary pollutant concentration in Taipei City is much improved, but ozone pollution is becoming worse. In order to understand the variations in the relationships between PM_{10} concentration and other pollutants after the decrease of primary pollutants, this study compiled yearly PC1 and PC2 loadings from 1994 to 2003 in Table 5. From 1994 to 1998, the PC1 loadings of PM_{10} range from 0.63 to 0.74, which are similar to the average of 0.70 for the decade. In contrast, the PC2 loadings of PM_{10} are between 0.46 and 0.64. It is worth noting that starting in 1999, all the yearly loadings of the PC1 of PM_{10} are less than 0.6, and the loadings of the PC2 of PM_{10} become higher than that in PC1. For an urban area, most PM_{10} , CO, and NO_x emissions originate from motor vehicles. These three pollutants should have high correlations with one another (Smith et al., 2001). The PCA results indicate that the PM_{10} concentration was highly correlated with local pollution in the early years, but in recent years, it has become more related to ozone concentration. Thus, the reduction of local primary emissions will lower the primary PM_{10} . However, increasing ozone concentrations will trigger higher secondary aerosol contribution to PM_{10} .

4. Conclusions

This study adopted PCA to analyze the air-quality monitoring data collected from 1994 to 2003 in Taipei City by the Taiwan EPA. The cumulative variance explained

from the two retained PCs amounts to 82.73% of the total variance. PC1, which represents primary pollutants such as CO, NO_x , and SO_2 , has shown a decreasing trend for the last 10 years. A reduction of primary pollutant concentrations found in this study demonstrates that the control measures implemented by the authorities are effective. However, PC2, that represents secondary pollutants, e.g., ozone, shows an annually increasing trend after the reduction of local pollutant emissions.

The resolved two PCs are further used for sorting out cases with PC1 scores lower than 1.5 and O_3 max concentrations less than 60 ppb, which were classified as originating from transboundary transport such as from Asian yellow sand. After removing these individual cases, the results of the long-term tendencies or seasonal patterns will better represent the source contributions from local activities in Taipei City.

Air pollution concentrations clearly show monthly variations in Taipei City. Most of the PC1 scores are positive from January to April averaged in 10 years because of poor dispersion resulting in pollutant accumulation. The PC2 scores are positive from April to August, indicating high photochemical pollution. After September, the wind speed increases, and the temperature drops. Both the scores of the PC1 and the PC2 become negative, indicating low pollutant concentrations from local source activities.

The PM_{10} concentration was strongly correlated to local pollution during early years, but has gradually changed to being more correlated with the variations of ozone concentration in recent years. After the reduction of local primary pollutant concentrations, secondary aerosol is expected to contribute more to PM_{10} concentration. The study on ozone formation and aerosol characterization for source apportionment hence becomes a need in the near future.

Acknowledgments

We are grateful to the Taiwan Environmental Protection Administration (TEPA) for providing the monitoring data

Table 5
The first two component loadings resulted from PCA after Varimax rotation and explained variance for each year of the five major air pollutants collected in Taipei City

	1994		1995		1996		1997		1998		1999		2000		2001		2002		2003	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
SO_2	0.49	0.65	0.53	0.69	0.77	0.39	0.51	0.71	0.71	0.53	0.57	0.54	0.48	0.65	0.39	0.71	0.42	0.70	0.29	0.76
CO	0.94	0.23	0.95	0.21	0.92	0.23	0.93	0.22	0.94	0.17	0.92	0.26	0.95	0.17	0.94	0.17	0.95	0.23	0.93	0.28
O_3 max	0.08	0.93	0.03	0.93	0.18	0.96	0.10	0.92	0.11	0.95	0.13	0.91	0.00	0.91	0.04	0.87	0.08	0.89	0.11	0.80
PM_{10}	0.72	0.54	0.63	0.64	0.74	0.46	0.63	0.61	0.70	0.48	0.55	0.57	0.53	0.65	0.53	0.60	0.50	0.71	0.42	0.72
NO_x	0.97	0.06	0.97	0.10	0.95	0.12	0.96	0.17	0.96	0.08	0.96	0.11	0.97	0.10	0.96	0.13	0.96	0.16	0.96	0.13
Eigenvalue	2.59	1.62	2.52	1.81	2.93	1.36	2.46	1.79	2.81	1.46	2.42	1.54	2.35	1.71	2.25	1.67	2.26	1.86	2.06	1.83
Variance explained (%)	51.85	32.47	50.49	36.12	58.58	27.19	49.14	35.80	56.12	29.30	48.50	30.71	47.00	34.23	44.93	33.38	45.28	37.23	41.13	36.65
Cumulative variance (%)	51.85	84.32	50.49	86.61	58.58	85.77	49.14	84.94	56.12	85.42	48.50	79.21	47.00	81.23	44.93	78.31	45.28	82.51	41.13	77.78

used in this study. Although, the data were taken from the TEPA official monitoring network, the results of this paper were not peer-reviewed by TEPA, and the mention of instrument trade names does not connote the endorsement of TEPA. We would also like to show our gratitude for the contributions from two anonymous reviewers for their constructive comments.

References

- Ashbaugh, L.L., Myrup, L.O., Flocchin, R.G., 1984. A principal component analysis of sulfur concentrations in the western United States. *Atmospheric Environment* 18, 783–791.
- Buhr, M.P., Trainer, M., Parrish, D.D., Sievers, R.E., Fehsenfeld, F.C., 1992. Assessment of pollutant emission inventories by principal component analysis of ambient air measurements. *Geophysical Research Letters* 19, 1009–1012.
- Buhr, M.P., Parrish, D., Elliot, J., Holloway, J., Carpenter, J., Goldan, P., Kuster, W., Trainer, M., Montzka, S., McKeen, S., Fehsenfeld, F., 1995. Evaluation of ozone precursor source types using principal component analysis of ambient air measurements in rural Alabama. *Journal of Geophysical Research* 100, 22853–22860.
- Chang, K.H., Jeng, F.T., Tsai, Y.L., Lin, P.L., 2000. Modeling of long-range transport on Taiwan's acid deposition under different weather conditions. *Atmospheric Environment* 34, 3281–3295.
- Chang, S.C., Lee, C.T., 2006a. Evaluation of the trend of air quality in Taipei, Taiwan from 1994 to 2003. *Environmental Monitoring and Assessment*, in press, doi:10.1007/s10661-006-9262-1.
- Chang, S.C., Lee, C.T., 2006b. Ozone variations through vehicle emissions reductions based on air quality monitoring data in Taipei City, Taiwan, from 1994 to 2003. *Atmospheric Environment* 40, 3513–3526.
- Chou, C.K., Huang, S.H., Chen, T.K., Lin, C.Y., Wang, L.C., 2005. Size-segregated characterization of atmospheric aerosols in Taipei during Asian outflow episodes. *Atmospheric Research* 75, 89–109.
- Eder, B.K., 1989. A principal component analysis of SO_4^{2-} -precipitation concentrations over the eastern United States. *Atmospheric Environment* 23, 2739–2750.
- Eder, B.K., Davis, J.M., Bloomfield, P., 1993. A characterization of the spatiotemporal variability of non-urban ozone concentrations over the eastern United States. *Atmospheric Environment* 27A, 2645–2668.
- Graham, B., Falkovich, A.H., Rudich, Y., Maenhaut, W., Guyon, P., Andreae, M.O., 2004. Local and regional contributions to the atmospheric aerosol over Tel Aviv, Israel: a case study using elemental, ionic and organic tracers. *Atmospheric Environment* 38, 1593–1604.
- Henry, R.C., Hidy, G.M., 1979. Multivariate analysis of particulate sulfate and other air quality variables by principal components—Part I: annual data from Los Angeles and New York. *Atmospheric Environment* 13, 1581–1596.
- Henry, R.C., Hidy, G.M., 1982. Multivariate analysis of particulate sulfate and other air quality variables by principal components—Part II: Salt Lake City, Utah, and St. Louis, Missouri. *Atmospheric Environment* 16, 929–943.
- Ho, K.F., Lee, S.C., Chan, C.K., Yu, J.C., Chow, J.C., Yao, X.H., 2003. Characterization of chemical species in $\text{PM}_{2.5}$ and PM_{10} aerosols in Hong Kong. *Atmospheric Environment* 37, 31–39.
- Hooper, R.P., Peters, N.E., 1989. Use of multivariate analysis for determining sources of solutes found in wet atmospheric deposition in the United States. *Environmental Science and Technology* 23, 1263–1268.
- Hosiokangas, J., Rhuskanen, J., Pekkanen, J., 1999. Effects of soil dust episodes and mixed fuel sources on source apportionment of PM_{10} particles in Kuopio, Finland. *Atmospheric Environment* 33, 3821–3829.
- Lehman, J., Swinton, K., Bortnick, S., Hamilton, C., Baldrige, E., Eder, B., Cox, B., 2004. Spatio-temporal characterization of tropospheric ozone across the eastern United States. *Atmospheric Environment* 38, 4357–4369.
- Lin, C.Y., Liu, S.C., Chou, C.K., Huang, S.J., Liu, C.M., Kuo, C.H., Young, C.Y., 2005. Long-range transport of aerosols and their impact on the air quality of Taiwan. *Atmospheric Environment* 39, 6066–6076.
- Lin, T.H., 2001. Long-range transport of yellow sand to Taiwan in Spring 2000: observed evidence and simulation. *Atmospheric Environment* 35, 5873–5882.
- Liu, C.M., Huang, C.Y., Shieh, S.L., Wu, C.C., 1994. Important meteorological parameters for ozone episodes experienced in the Taipei Basin. *Atmospheric Environment* 28, 159–173.
- Marcazzan, G.M., Vaccaro, S., Valli, G., Vecchi, R., 2001. Characterization of $\text{PM}_{2.5}$ and PM_{10} particulate matter in the ambient air of Milan (Italy). *Atmospheric Environment* 35, 4639–4650.
- Rodriguez, S., Querol, X., Alastuey, A., Mantilla, E., 2002. Origin of high summer PM_{10} and TSP concentrations at rural site in Eastern Spain. *Atmospheric Environment* 36, 3101–3112.
- Smith, S., Stribley, F.T., Milligan, P., Barratt, B., 2001. Factors influencing measurements of PM_{10} during 1995–1997 in London. *Atmospheric Environment* 35, 4651–4662.
- StatSoft, Inc., 2001. STATISTICA (data analysis software system), version 6.0 <www.statsoft.com>.
- Taiwan EPA, 2005. Environmental database. <http://edb.epa.gov.tw/>.
- Tao, Z., Larson, S.M., Williams, A., Caughey, M., Wuebbles, D.J., 2005. Area, mobile, and point source contributions to ground level ozone: a summer simulation across the continental USA. *Atmospheric Environment* 39, 1869–1877.
- van der Wal, J.T., Janssen, L.H.J.M., 2000. Analysis of spatial and temporal variations of PM_{10} concentrations in the Netherlands using Kalman filtering. *Atmospheric Environment* 34, 3675–3687.
- Vukovich, F.M., 1994. Boundary layer ozone variations in the eastern United States and their association with meteorological variations: long-term variations. *Journal of Geophysical Research* 99 (D3), 16839–16850.
- Vukovich, F.M., Sherwell, J., 2003. An examination of the relationship between certain meteorological parameters and surface ozone variations in the Baltimore-Washington corridor. *Atmospheric Environment* 37, 971–981.
- Yang, K.L., 2002. Spatial and seasonal variation of PM_{10} mass concentrations in Taiwan. *Atmospheric Environment* 36, 3403–3411.
- Yu, T.Y., Chang, L.F., 2001. Delineation of air quality basins utilizing multivariate statistical methods in Taiwan. *Atmospheric Environment* 35, 3155–3166.

Impacts on industry of Europe's emerging chemicals policy REACH

Gerhard Angerer^{a,*}, Ralf Nordbeck^b, Christian Sartorius^a

^aFraunhofer Institute for Systems and Innovation Research (ISI), Breslauer Straße 48, D-76139 Karlsruhe, Germany

^bUniversity of Natural Resources and Applied Life Sciences (BOKU), Feistmantelstraße 4, A-1180 Vienna, Austria

Received 30 June 2006; received in revised form 8 December 2006; accepted 12 December 2006

Available online 23 February 2007

Abstract

For Europe, a new regime in chemicals regulation is about to start. After the proposal of the European Commission concerning the Registration, Evaluation and Authorization of Chemicals (REACH) passed its readings in the European Parliament and some differences with the European Council of Ministers were resolved, the regulation will come into force in June 2007. This paper is focused on the question how serious the cost burdens for industry induced by REACH will be, and whether the New European Member States (NMS) which joined the European Union in May 2004 will be able to cope with the regulation. This evaluation has been done by assessing the legislative, administrative and economic framework in New Member States and by analysing real business cases in companies. The empirical showcase business impact studies are at the same time of interest for companies of EU-15 states, other European countries who may implement the regulation, and even for exporters of raw materials and chemicals outside Europe, who will also have to comply with REACH if they market in the European Community. The results give no indications that REACH adoption will bring significant drawbacks to companies in the NMS. The emerging regulation will bring challenges for individual companies, especially for small and medium-sized ones, but for the European chemical industry as a whole, there is no question that it will be able to cope with REACH burdens without losing its global competitiveness.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: European chemicals policy; REACH; Substance registration costs; Acquis communautaire; Policy impact analysis; Business impact study; Chemical industry; Specialty chemicals sector; Industry's vulnerability; Cost burdens of industry; EU New Member States

1. Introduction

When adopted, the emerging European chemicals regulation of the European Parliament and the Council concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) will be equally applied in all 25 Member States (CEC, 2003a). The impacts of REACH on the chemical industry in the New European Member States (NMS) could, however, be different from those in EU-15, because of different specific environments in the economy and administration. Little knowledge is yet available as to how these different starting positions may affect the performance of industry in the individual countries.

To fill this gap, the European Commission and the European industry associations UNICE and CEFIC decided to analyse the potential impacts of REACH on the NMS (Fraunhofer ISI et al., 2005a, b). The objective of the project was to explore the ability of the industry in the NMS, especially the chemical industry, to cope with the implementation of REACH. The focus of the study was intended to be placed on the specialty chemicals sector, which is considered as being particularly vulnerable to the new chemicals policy. The empirical impact analysis was done in business case studies performed in the Czech Republic, Poland and Estonia. The impact assessment is based on the European Commission's regulation draft issued on October 29, 2003. After substantial amendments made by the European Parliament and the European Council it is clear that the regulation coming into force in 2007 will differ from this early draft. But the key cost figures used for the impact calculation in this analysis will not change significantly.

*Corresponding author. Tel.: +49 721 6809 117; fax: +49 721 689 152.

E-mail addresses: g.angerer@isi.fraunhofer.de (G. Angerer), ralf.nordbeck@boku.ac.at (R. Nordbeck), c.sartorius@isi.fraunhofer.de (C. Sartorius).

This paper provides an assessment of the actual state of affairs in the European Union's New Member States (Section 2), the vulnerability of their chemical industries (Section 3) and cost burdens arising from compliance with REACH (Section 4).

2. The actual state of affairs in the EU New Member States

The amendment of the national chemicals regulations started some time before the accession of the NMS in May 2004, leading to the situation that the NMS were dealing with the implementation of the *Acquis Communautaire*, including the "Chemicals Acquis", into national legislation, while European chemical legislation was already on the way to being reformed. This situation is an important distinguishing factor between EU-15 and the NMS. This section provides information on the completeness of transposition, implementation and enforcement of the chemicals acquis in the NMS and on that basis assesses the potential implications for the implementation of REACH.

From the very beginning, it was recognized that transposition and implementation of the EU's environmental laws presents a challenge to the accession countries. This related to both the sheer scale of past environmental liabilities and the gap in the level of environmental protection in central and eastern Europe compared with the situation in EU-15. It was recognized that full compliance with the environmental acquis will probably only be achievable in the long run (CEC, 1998, p. 1).

The EU Commission (CEC, 1998, p. 2) identified three key challenges in the area of approximation of environmental legislation: (1) transposition and implementation of the environmental acquis as a *legislative challenge*; (2) a need to strengthen the administrative structure necessary for monitoring and enforcement as an *institutional challenge*; and (3) financing the investment costs of meeting the environmental acquis as a *financial challenge*.

The following assessment of the degree to which the relevant parts of the chemicals acquis have been applied and implemented in the NMS will give an indication of the

preparedness of these countries for the implementation of REACH.

2.1. The legislative challenge

The existing legislation—transposed in the NMS as "chemicals acquis"—contains 16 pieces of legislation, including 12 directives and 4 regulations (Fig. 1). It is divided into four different areas due to their different nature: testing requirements, transport of dangerous goods, notification of chemical substances including biocides and finally, chemical product regulations. With accession of the NMS to the EU in 2004, almost all directives of the chemicals acquis had been transposed into national legislation.

A first assessment of the compliance levels with EU environmental legislation in 10 CEE countries was done by the Regional Environmental Centre (REC) in 1996. In the legislative field of chemicals, industrial risks and biotechnology, the study showed an average compliance level of 27%, lagging significantly behind the average compliance level of 46% for all legislative fields. The 27% of compliance as an average indicated only the beginning of drafting or some very early attempts to regulate (REC 1996, p. 9).

Much work has been carried out since then by the accession countries to transpose the EU chemicals directives into national legislation. Between 1998 and 2002, the candidate countries steadily made progress in the task of drafting laws and administrative regulations to transpose the EU obligations. In the Czech Republic, Estonia, Hungary and Slovenia, substantial parts of chemicals legislation were already in place in 1999 (Andrijewski, 2001; Bláha, 2001; Grabner, 2001; Lang, 2002). During the EU screening process in the same year, no major problems were identified in the field of chemicals legislation for most of the candidate countries. Full compliance with the chemical acquis was foreseen in all countries by the end of 2002 and no transition periods were requested in the negotiations with the Commission (Joint Commission/CEFIC Working Group on Enlargement, 2000; Lang, 2002).

TESTING	TRANSPORT	MARKET NOTIFICATION	CHEMICAL PRODUCTS
Animal Testing (86/609/EEC)	Dangerous Goods by Road (94/55/EEC)	Classification, Packaging and Labelling of Dangerous Substances (67/548/EEC)	Evaluation and control of the risks of existing substances (EEC/793/93)
Good Laboratory Prac- tice (87/18/EEC)		Dangerous Preparations (1999/45/EEC)	Restrictions on Marketing and Use (76/769/EEC)
GLP Inspection (88/320/EEC)		Risk Assessment of New Substances (93/67/EEC)	Asbestos (87/217/EEC)
		Biocidal Products (98/8/EC)	Detergents (73/404/EEC)
		Import & Export of Danger- ous Substances (EEC/2455/92)	Ozone-Depleting Substances (EC/2037/2000)
			Safety Data Sheet (91/155/EEC)

Fig. 1. EU chemicals acquis communautaire.

In November 2003, the EU Commission concluded in its Comprehensive Monitoring Report that the legislation concerning chemicals was in line with the *acquis* in all candidate countries, except for asbestos in Hungary and biocides in Estonia, Lithuania, and Poland (CEC, 2003a). In mid-2004, transposition of the chemical *acquis* was finalized in all New Member States.

But it would be wrong to assume that the chemical legislation as a whole is now in force and being implemented in the New EU Member States. With transposition essentially complete in 2004, the debate switched away from transposition of the EU chemical *acquis* and instead focused on its implementation and enforcement. Gaps remain in implementation and enforcement and will continue to remain for some time, as will be shown in the following sections.

2.2. The administrative challenge

Having in place the necessary administrative structures for implementation and enforcement was the second element of approximation. However, establishment of an adequate administrative capacity to implement and enforce the legislation is more difficult, especially for chemicals, since various institutions are involved in chemicals management: ministries, agencies like national chemicals bureaus, and enforcement inspectorates. In most of the NMS, coordination between responsible ministries and agencies obviously remained the biggest problem in terms of approximation towards EU chemicals legislation.

The new laws on chemical substances and preparations adopted in the NMS during the last years assign tasks to different state administrations at the policy, implementation, and enforcement level (Witzani and Andrijewski, 2004; Stocker and Anwander, 2004). At the policy level, the responsibilities for chemicals legislation have been divided between two or more ministries, in most cases including the Ministry of Economics, the Ministry of Environment, the Ministry of Health, the Ministry of Labour and Social Policy, and the Ministry of Agriculture. In almost all countries the leading role for chemicals legislation has been assigned either to the Ministry of Environment or the Ministry of Health. At the implementation level, several New Member States have established a National Chemicals Bureau (NCB) as central administrative coordination unit. At the enforcement level, in almost all countries responsibilities for enforcement have been divided between different inspectorates, including the Environmental Inspectorate, the Health or Sanitary Inspectorate, the Labour Inspectorate, and the Customs Authority.

Institutions to implement and enforce the EU chemicals legislation have been set up in all NMS and significant progress was made with regard to the necessary administrative capacities. Nevertheless, some problems still remain in this area. Probably the biggest problem is the low number of staff at the implementation level in some countries. For instance, the Polish Bureau for Chemical

Substances and Preparations with 20 employees is currently clearly understaffed. A PHARE twinning project recommended in March 2004 that the number of risk assessors and risk managers at the Bureau should be doubled and the reinforcement should be completed within the next 2 yr. Otherwise the Bureau would face difficulties fulfilling its role as competent authority (Witzani and Andrijewski, 2004). Similar problems were reported in Slovenia, where the NCB staff currently numbers 25. The PHARE twinning project on chemicals safety in Slovenia recommended focusing only on EU-oriented priorities, with the present number of staff at the NCB and the Health Inspectorate. Similarly, it is reported from the Baltic States that responsible authorities are seriously understaffed. In some cases, one or two persons in the ministries engaged with REACH have to work on different and complex policy areas at the same time.

The second problem with regard to administrative capacity has been the low level of coordination among the responsible authorities. Although this problem is only of a temporary nature, it might seriously hamper the implementation of REACH.

2.3. The financial challenge

The amount of investment required to comply with the environmental *acquis* has been discussed in the last 10 yr. A first study in 1997 carried out for DG Environment estimated the amount of environmental investments required in the candidate countries to reach compliance at €120 billion (EDC and EPE, 1997). More recent calculations produced lower figures, estimating the costs of compliance at between €80 and €110 billion (DANCEE, 2002, p. 37). Nevertheless, financing the infrastructure necessary for compliance with EU environmental legislation poses a major financial challenge to the New EU Member States. The figure of 2–3% of GDP per annum is mentioned as the level of spending required by the NMS to ensure full implementation (CEC, 2001, p. 4). Current annual environmental expenditures in 2000 ranged between 0.5% and 1.6% of GDP (CEC, 2002, p. 13).

There are only few studies available which make any cost estimates for the implementation of the chemicals *acquis* (Table 1). This is for two reasons. First, the investment costs for implementing the chemical *acquis* in the four countries where studies have been carried out are roughly between 0.5% and 1.5% of the total investments required

Table 1
Estimated costs of compliance with the chemicals *acquis* 1998–2005 in €1000

	Public sector	Private sector
Czech Republic	3157–3298 (per year)	3737–28,420 (per year)
Lithuania	493+870 (per year)	12,000 k€
Poland	n/a	60,000–101,250

Source: CEFIC (1998), Soil and Water (1999), Bucknall et al. (2000).

for compliance with EU environmental legislation. Compared to the investment needs for the heavy investment directives, this is almost a negligible factor. Second, detailed cost estimates and financial plans had to be prepared only for those directives where the candidate countries requested transition periods. This was not the case for any directive of the chemical acquis in the ten candidate countries. The lack of requests for transition periods can thus be seen as an indication of the relatively modest financial efforts needed to comply with the chemicals acquis.

In Lithuania, the initial investment for the public sector was estimated at €493,000 (Soil and Water, 1999, p. 65). These expenditures are due to legal transposition of the EU requirements in Lithuanian chemicals legislation, setting up a new Chemicals Supervisory Committee, increasing technical capacities and expertise by building new laboratories, and staff training. Operational costs in the public sector for implementation of requirements, administrative procedures and enforcement were estimated at €870,000/yr. Enforcement of the requirements in the directives and regulations at the regional level accounted for more than half of the total operational costs. Another large part of the annual operational costs is caused by the administrative procedures on new and existing substances, and bans and restrictions on dangerous substances.

In the Czech Republic, the average annual recurrent costs for the state administration were estimated at about €3 million. Annual costs for the private sector range between €3.7 and €28 million. These overall annual expenditures for implementing the chemical acquis represent only 0.6–1.9% of the overall annual compliance costs for the whole environmental acquis, the total annual cost being estimated as €1.15–1.69 billion (Bucknall et al., 2000, p. 15).

The compliance costs in Poland have been estimated only very roughly. A CEFIC/PHARE project estimated the annual costs of compliance for the Polish chemical industry in 1998 as 2% of the total turnover, which then equalled €180 million. The adaptation costs, as lump sum induced by EU approximation, were estimated to constitute 300–450% of the routine annual costs (CEFIC, 1998, p. 96). This amounts to estimated costs of compliance ranging between €540 and €810 million in total, or annually €60–101 million over 8 yr. These compliance costs reflect the fact that the Polish chemical sector is by far the largest one, compared with those of other NMS.

3. The vulnerability of the chemicals industries for REACH

Compliance with REACH represents a burden for the economies of the NMS, which however will not affect all countries in the same manner. In order to assess their actual impacts, we will have a closer look at the specialty chemicals sector in different countries. Due to the large number of small-volume chemicals used and produced in that sector, the cost of compliance with REACH is expected to be higher

there. So the larger the share of the specialty chemicals is, the more a country's economy will be affected. On the other hand, this negative effect could be offset by the (potentially) surpassing growth or productivity of the specialty sector. Finally, the burden of REACH will crucially depend on the quantity of (specialty) chemicals imported from, or exported to, countries outside the EU. This is important, since for imported chemicals it is questionable whether registration will be required and the import can continue, whereas exported chemical may lose their competitiveness because of the cost burden of substances registration.

In the following, we describe the specialty chemical sector in three New Member States: the Czech Republic, Poland and Estonia. They represent a wide range with regard to country size, importance of chemical industry and position of the specialty chemical sector.

3.1. The specialty chemical sector in the Czech Republic

The chemical industry¹ contributed 2.3% or €3.8 billion to the turnover and 2.15% or €1 billion to the GDP of the whole Czech economy in 2002 (CMIT, 2004). Over time, the turnover of the chemical industry (in constant 1995 prices) increased from less than €2 billion in 1992 to about €3.2 billion in 2000, which corresponds to an average annual increase by 6%, and remained more or less constant since then (CEFIC, 2005). Relative to the development of the entire manufacturing sector, the chemical industry maintained a more or less constant share (of turnover) of slightly more than 6% of total manufacturing until 2000, but lost ground since then, yielding a share of slightly less than 5% in 2003 (Czech Republic Statistical Office). By contrast, the specialty chemicals sector² grew significantly faster than the chemical sector in general in terms of turnover. While in the period from 2000 to 2003 the latter grew on average by less than 1% annually, the average rate of increase for specialty chemicals was almost 3%. The development in the specialty chemicals sector looks even more favourable, if the value added is considered. Similar to the turnover, the contribution of the specialty chemical sector to the added value of the total chemical sector grew between 2000 and 2003 from 20.0% to 23.8%. This time, however, the respective average growth rates were 10.4% and 4.2% before and after 2000, respectively (EUROSTAT, 2005). So, while the chemical industry in general is facing stagnation, the specialty chemicals sector still shows considerable growth.

The Czech Republic, like all New Member States, shows a large deficit in foreign trade and the major part of this

¹In this report, the chemical industry is defined rather narrowly as comprising NACE category 24 (manufacture of chemicals and chemical products).

²If not stated otherwise, the specialty chemicals sector is specified by the aggregate of NACE categories 24.3 (manufacture of paints, varnishes and similar coatings, printing ink and mastics), 24.5 (manufacture of soap and detergents, cleaning and polishing preparations, perfumes and toilet preparations) and 24.6 (manufacture of other chemical products).

deficit is due to the chemical industry and its products. 65% of the chemical industry's sales go into export, whereas chemical products worth 135% of total sales are imported. To make things even worse, the deficit grew by more than 10% annually during the past 4 yr. While specialty chemicals contributed 27% to both the export and import side of the deficit of chemical industry in general in 2003, the growth of the deficit is exclusively due to mass chemicals and pharmaceuticals. In contrast, the trade deficit for specialty chemicals remained almost constant since 2000, with the exports growing much faster (7% p.a.) than imports (3% p.a.) as they started from a much lower basis (SCHP, 2003). This again is a clear indication of the competitiveness of the Czech specialty chemical sector.

The proportion of specialty chemicals traded with countries outside the EU-27³ in 2003 was 25.5% in export and 12.2% in import on total trade with specialty chemicals (SCHP, 2003). This implies that the Czech Republic is to some extent affected by a loss of competitiveness with its specialty chemicals in non-EU markets, but much less by a lack of availability of chemicals coming from these markets.

3.2. The specialty chemical sector in Poland

The chemical industry had a turnover of about €8.5 billion and contributed 1.3% or €2.6 billion to the GDP of the whole Polish economy (EUROSTAT, 2005). Over time, the turnover of the chemical industry increased from €2 billion in 1992 to almost €8 billion in 2001 (CEFIC, 2005), which corresponds to an average annual increase by 25% until 1995, by 10% until 2001 and only a slight increase (by 1% p.a.) since then. Compared with increases of about 11% annually between 1995 and 2001 and a 1–2% range thereafter in the whole manufacturing sector, these figures indicate that the chemical industry lost ground slowly, with its share declining from 7.7% in 1995 to less than 7% in 2002 (EUROSTAT, 2005). Despite the rather fragmentary character of statistical data from Poland (PIPC, 2003; EUROSTAT, 2005) with only NACE categories 24.3 and 24.5 specified explicitly, our estimates of contributions of the specialty chemical sector of 28.1% (value added), 48.9% (number of firms) and 26.0% (employees) to the chemical industry indicate that the relative importance of the specialty chemical sector (in 1999, the latest year available) is significantly higher in Poland than in the Czech Republic (in 2002). The productivity of labour in the specialty chemical sector in Poland seems to be similar to that in the Czech Republic.

The stronger position of the specialty chemical sector in Poland is confirmed by turnover data from EUROSTAT (2005) and PIPC (2003). It shows that in Poland, the specialty chemical sector grew significantly faster (by an average 11% annually) than the chemical sector in general

(6% annually). The development in the specialty chemicals sector looks even more favourable in terms of value added, where the contribution of the specialty chemical sector grew from 28.6% in 1998 to about 37% in 2002, showing an average annual growth rate of 14.8% (7.6% in chemical industry). The fact that the latter values were significantly higher than those of employment (0.4% and –6.2%, respectively) indicates that the productivity of labour (expressed as value added per full-time equivalent employee) must have undergone a substantial increase. With an average 14% growth annually, not only is the increase in productivity in Poland much higher than in the Czech Republic (with hardly 5%); but also the absolute productivity achieved in 2002 in Poland is 60% higher in the specialty chemicals sector and 6% higher in the entire chemical industry.

Like the Czech Republic, Poland shows a large deficit in foreign trade, the major part of which is due to the chemical industry and its products. To make things worse, the deficit grew strongly, by more than 15% annually from 1995 to 2001, but slowed down in recent years (CEFIC, 2005). While specialty chemicals contributed to the export and import side of the deficit, exports grew so much more strongly than imports that specialty chemicals contributed only less than 10% to the deficit increase taking place in 2002/2003. This again is a clear indication of the competitiveness of the Polish specialty chemicals sector.

Data specifying foreign trade in specialty chemicals were not available. Alternatively, data for the entire chemicals sector showed that with 22.3% and 30.7%, the respective proportions of export and import with countries outside EU-27 from total export and import of chemicals were not much different from figures known for the Czech Republic, where it is 20.5% and 18.8%, respectively (SCHP, 2003; PIPC, 2004a, b).

Summarizing these facts, the position of the specialty chemical sector in Poland appears to be similarly favourable for the implementation of REACH as that of the Czech Republic.

3.3. The specialty chemical sector in Estonia

With 1.4 million inhabitants, Estonia has little more than one-thirtieth of the population of Poland. Accordingly, its economic power was reflected in a GDP of just €7.5 billion in 2002. The chemical industry had a turnover of about €212 million and contributed 0.6% or €44 million to the GDP of the whole Estonian economy (EUROSTAT, 2005)—much less than in Poland (1.3%) or even the Czech Republic (2.1%). In contrast to the manufacturing sector which grew by an average annual 13% between 1996 and 2002, the development of the chemical industry was characterized by stagnation and the value added grew only by less than an annual 2.5% on average (EUROSTAT, 2005). Compared with the total chemical sector, the combined shares of the specialty chemical sector are 73% (value added), 60% (number of firms) and 57%

³This is EU-25 plus Bulgaria and Romania.

(employees). So, although the degree of industrialization in Estonia as measured by the contribution of the manufacturing sector to the country's GDP is relatively low, the relative importance of the specialty chemical sector within the Estonian chemical industry appears to be higher than in Poland and much higher than in the Czech Republic.

On assessing the competitiveness of the Estonian specialty chemical sector in the international context, it was found that the productivity expressed as value added per full-time equivalent employee was considerably lower in the Estonian specialty chemicals sector (€11,900 per employee). It is also significantly lower than in the entire chemical industry of Estonia (€15,900) (EUROSTAT, 2005). All this must be seen in the context of a specialty chemicals sector assuming a high, constant share of the production and employment of the chemical industry, which is in turn unable to participate in the steady growth of the manufacturing sector in general. It remains to be seen whether this finding is confirmed by the analysis of the foreign markets.

At 15%, the relative contribution of the chemical industry to the total Estonian trade deficit (€2066 million) (Statistical Office of Estonia, 2005) is much smaller than in Poland or even the Czech Republic. This is mainly because large quantities of chemical products are exported immediately after import by Estonian chemical trading companies. In this context, the strong and growing export position of the specialty chemicals sector would give rise to a positive assessment of the competitiveness, if it were accompanied by a similarly positive assessment for the chemical industry as a whole, but this unfortunately is not the case. Even worse, in a future perspective, the strong position of Estonia with regard to foreign trade will mean REACH will have a strong impact on the country. Due to the very large share of Estonian exports to non-EU countries (82.4%), competitive disadvantages arising from the necessity of registering such exports in accordance with REACH are stronger than in Poland or the Czech

Republic. By contrast, the proportion of chemicals (36.8%) and even specialty chemicals (17.4%) to Estonia from outside EU-27 is basically comparable with that of Poland and the Czech Republic, so that the risk of non-availability of many substances after the implementation of REACH can be considered as not serious (Statistical Office of Estonia, 2005).

3.4. Country comparison

The data characterizing the relevance and performance of the specialty chemicals sector in three NMS are compared in Table 2. From these data it is evident that the chemical industry in Estonia will be affected more under REACH than those of the other countries because (1) the specialty chemicals sector represents a much larger part of it and (2) the Estonian specialty chemical industry is engaged extensively in export relations with non-EU countries. All this has to be seen against the backdrop that, on the one hand, the productivity in the Estonian specialty chemicals sector is by far the lowest of all three countries, while on the other hand the contribution of Estonia's chemical industry to its GDP is rather low anyway. The chemical industries of Poland and the Czech Republic, by contrast, will be affected to a much lesser extent because they are stronger, the contribution of the specialty chemicals sector is smaller, its productivity higher and relations to non-EU countries less important.

4. Cost burdens arising from REACH

4.1. Methodology of the cost impact analysis on company level

The REACH regulation will enforce a registration of each chemical substance produced in or imported to the EU market. All manufacturers or importers who market substances in quantities above 1 t/yr are responsible for

Table 2
Relevance of the chemical sectors in the national economy (2002)

	Czech Republic	Poland ^a	Estonia
GDP (million €)	46,500	22,500	7500
Value added of chemical industry (million €)	1000	2600	44
Turnover of chemical industry (million €)	3800	8500	212
Turnover with specialty chemicals (million €)	1033	3531	144
Productivity of the specialty chemicals sector (€/employee)	18,000	29,200	11,900
Growth rate of the chemical industry 2000–2003 (% p.a.)	1	2	7
Growth rate of specialty chemicals sales 2000–2003 (% p.a.)	3	11	5
Share of exported chemicals in produced chemicals (%)	65	40	121
Share of specialty chemicals export in produced specialty chemicals (%)	64	31	67
Share of specialty chemicals export to outside EU-27 from total export (%)	25.5	22.3 ^b	82.4
Share of imported chemicals in produced chemicals (%)	135	88	267
Share of specialty chemicals import in produced specialty chemicals (%)	123	57	166
Share of specialty chemicals import from outside EU-27 from total import (%)	12.2	30.7 ^b	17.4

^aSpecialty chemicals figures in Poland do not include NACE category 24.6.

^bThe Polish shares of export/import to/from outside EU-27 are for all chemicals.

registration. It is expected that about 30,000 substances will be subject to registration by the new European Chemicals Agency in Helsinki. Part of the registration dossier will be information about all uses of the substance. For substances produced in quantities above 10 t/yr, the manufacturer must submit a chemical safety report which assesses human and environmental health hazards. For dangerous substances, an exposure assessment and a risk characterization of all uses is required.

Phasing out production of certain substances as a consequence of financial expenditures for testing and registration is one of the most frequently discussed impacts of the proposed REACH regulation. Especially low-volume and low-price chemicals are endangered or, in other words, vulnerable. Vulnerability of a substance means that the substance supplier is in doubt whether or not to invest in registration to maintain production of this substance under REACH. Accordingly, the decision of a company to phase out production is considered an investment problem. Employing the net present value (NPV) approach, investment in registration is commercially attractive if the NPV is positive

$$K_N = \frac{wPA}{f_N} - I_0, \quad (1)$$

$$f_N = \frac{(1+p)^N p}{(1+p)^N - 1}, \quad (2)$$

where K_N is the NPV; w the profit margin for the substance ($0 < w < 1$); P the price of substance (€/kg); A the yearly production of the substance (kg/a); f_N the annuity factor ($1/a$); p the discount rate ($0 < p < 1$) ($1/a$); N the payback time (years); I_0 the cost of registration for the substance (€).

From Eq. (1) a criterion can be derived to prove whether registration of a certain substance is profitable or not:

$$\frac{I_0}{A} \leq \frac{w}{f_N} P. \quad (3)$$

For a quick proof, one could assume to have a sales margin w of 8%, a discount rate of 10% and a requested payback time of 5 yr. This gives the handy criterion:

$$\frac{I_0}{A} \leq 0.3P. \quad (4)$$

It says investment in registration of a substance is profitable at the first glance if the registration costs divided by its annual production do not exceed 30% of substance's price.

For registration costs I_0 the figures of Table 3 were applied (Pedersen et al., 2003). The costs refer to the average cost scenario of that source (EC DG JRC IHCP) and are based on a rate of €1000 per person and day.

The other crucial issue is the possibility to pass on registration costs to customers down the supply chain. As all manufacturers and importers in the European Union are affected by the REACH regulation, passing on costs along the value chain should be possible, in principle. But to which extent cost pass-on will be enforceable for a specific company depends on its production volume. A small manufacturer has to bear the same registration cost as a company with a large production volume of the same substance. This means the small company has to bear a higher sales price burden. But in a competitive market, he can only earn if his substance price does not exceed the lower price of his larger competitor. Therefore, the small manufacturer will not be able to pass on registration costs fully to the market.

Eq. (5) is used to calculate a full pass-on of registration costs on substance price. Achieving this price increase on the market leaves the sales margin in €/kg unchanged after investment in registration

$$\Delta P_k = P_k - P_{0k} = \frac{I_{0k} f_N}{A_k}, \quad (5)$$

where P_{0k} is the price of substance k prior to REACH (€/kg); P_k the price of substance k under REACH (€/kg); I_{0k} the registration cost for substance k (€); A_k the yearly

Table 3
Applied costs for substance registration in € per substance depending on manufacturer's production volume

Cost elements	1–10 t/a	10–100 t/a	100–1000 t/a	> 1000 t/a
Hazard assessment: Art. 13, 3 a–d		1500	8700	8700
Chemical safety report: Art. 13, 1		1000	2000	2000
Administration for the preparation of the technical dossier	5000	5000	10,000	10,000
Agency fees	400	400	8000	8000
Testing (IHCP “average test needs”) to fulfil provisions of Annex V–VIII	7700	73,100	163,000	208,000
Registration costs non-dangerous substance	13,100	81,000	191,700	236,700
Robust study summary: Art. 9 a vii; Annex I, 1.1.4, 3.1.5 (summary of full study reports from the application of Annexes V–IX)			500	1000
Exposure assessment: Art. 13, 4 a for dangerous substances		2700	7200	19,500
Liaison with downstream users: Art. 13, 4. a for exposure assessment of dangerous substances		3500	12,000	15,000
Risk characterization: Art 13, 4 b for dangerous substances		800	3500	3500
Registration costs dangerous for dangerous substances listed in Directive 67/548/EEC	13,100	88,000	214,900	275,700

production of the substance k by the supplier (kg/a); and f_N the annuity factor (1/a).

The cost impact on a preparation due to the price increase of the contained substances is given by

$$C - C_0 = \sum_{k=1}^K c_k (P_k - P_{0k}), \quad (6)$$

where C_0 is the preparation cost prior to REACH (€/kg); C the preparation cost under REACH (€/kg); c_k the concentration of substance k in the preparation ($0 < c_k < 1$); and K the number of substances in the preparation.

4.2. Approach of business data collection

In order to assess cost impacts of the emerging new chemicals policy REACH, a total of 14 companies were selected for business case studies with the help of the national chemical industry associations in Estonia, the Czech Republic and Poland. The country portfolio was intended to include a small, medium and large national economy from the New Member States in the sample. Collection of information and quantitative figures was done in a two-step approach, first by a written questionnaire followed by face-to-face interviews. Members of company's top management and technical levels participated in the interviews. All basic figures used for the business cases calculations were confirmed by the companies. In spite of this effort for quality assurance, the analyses cannot claim to be representative for the whole chemicals sector. No such analysis could claim this, because assessing a representative sample of the 25,000 European chemical industry companies is definitely impossible. The interesting feature of such case studies is the spotlight-like insight into the real daily business of chemical companies.

The results reported here cover four companies. Among them are two substance suppliers and two preparation formulators. Of their product portfolio, 29 substances and 3 preparations were analysed in detail. One of the preparation formulators produces only paints and varnishes. The other has a mixed portfolio covering washing and cleaning agents for households, hospitals and industry, as well as accumulator masses. None of the four companies fulfil the criteria for SME, but the varnish manufacturer is close to the SME definition of 250 employees.

Most companies in the NMS are winners in the EU accession process. Due to low wages, which are between €2.65 and 4.50/h in the company sample, they gain a considerable competitive advantage over EU-15 industries. For the four companies, this is reflected in high annual growth rates of between 5.5% and 10.5%. The gross profit margin, reflecting the economic strength and competitive environment, varies between 4% and 14% of turnover.

On the product markets, competition is primarily enforced by global player companies located in EU-15. In

contrast to that, companies from outside the EU do not play any role in product markets. Due to low wages, raw material input is the dominant cost factor. In the weighted average 43% of total costs are spent to purchase raw materials for production. The substance suppliers spend 5% and the preparation makers 18% on personnel costs.

4.3. Vulnerability of substance manufacturing

As described above, phasing-out of substance production as a consequence of cost burdens for testing and registration is expected to be one impact of REACH. For this reason, this issue was paid special attention in the analyses. Substance suppliers stated that phasing-out of substances for economic or technical reasons occurred frequently in the past and is still part of daily business. The number of substances phased out of production in the near past was given as close to 10% per year.

In the four companies' sample, 29 substances were analysed for vulnerability to phasing out as a consequence of REACH registration cost burdens. According to the methodological approach described previously, a company's decision to mobilize capital for testing and substance registration is considered an investment problem. In this view the net present value (NPV) indicates whether registration of a certain substance is commercially attractive or not, when applying company's profitability requirements. The analysis shows that four substances came out with a negative NPV. One is a bulk resin, produced by one company in a relatively small volume of below 2000 t/a. Within the company's required payback time of 3.5 yr, 75% of the investment have been earned at a discount rate of 6% p.a. The other three are specialty chemicals; two are pigments, the third graphite of high purity. For one pigment 86% of registration costs are refinanced at the end of the required payback time. Those substances with a negative NPV are endangered by phase-out if in addition it is not possible for the manufacturer to pass on registration costs to customers partially or totally. Cost pass-on may become a problem for the bulk resin because the company is in an unfavourable market position with its low production volume, but could be easier for the three specialties.

4.4. Cost impacts on chemical preparations and products

Three preparations were traced for cost impacts as a consequence of substance registration costs passed down the supply chain. These preparations were suggested by the companies for a cost impact analysis, because they are considered as key and vulnerable chemicals. "Key" means the product is of commercial or strategic importance for the company. "Vulnerable" means that severe impacts of REACH are expected.

The first preparation is an agent for surface treatment of steel. It is more expensive than substitutes, but extracts less steel material off the processed surface. Therefore, the

agent is predominantly applied for the treatment of (expensive) stainless steel. Sales price of the preparation is 1.54 €/kg.

The preparation consists of three high volume chemicals. With a payback time of 4.4 yr and a discount rate of 6.6% p.a., price increases of the contained substances stay between 0.3% and 3.5%. This yields a preparation price increase of 0.0125 €/kg, corresponding to 0.8% of sales price.

Application of the preparation in the downstream user industries is for surface treatment of stainless steel. Shaped flat stainless steel is sold for a price of 2200 €/t. Applying 7 kg of the de-scaling agent per ton of steel yields a cost increase of 0.009 € per metric ton of steel. This causes a mark-up of the semi-finished steel product of 0.004%, which can certainly be considered negligible.

The *second preparation* is an electrode mass for rechargeable nickel–cadmium accumulators. It is predominately used for portable appliances, like power tools, camcorders, flash lights, etc. The sales price of the preparation is 7.90 €/kg. The suppliers' production volumes of the contained substances range from 60 to 30,000 t/a. Again full cost pass-on of the registration costs is assumed, applying a discount rate of 6.6% and a pay-back time of 4.4 yr. Price increases of contained substances in this case stay between 0.03% and 14.5%, yielding a preparation cost increase of 0.16 €/kg. This is 2% of the preparation's present sales price.

A widespread product using this electrode mass is for AA type (mignon) accumulators. Such an accumulator weighs 23 g, of which 30% are due to the electrode mass. This yields a price increase of 0.11 euro cent per piece. With a sales price of such an accumulator of €2, the price charge of 0.06% is certainly a negligible cost burden for consumers.

The *third preparation* is a varnish applied as top-coating for metal surfaces. The sales price is 4.10 €/kg. The varnish contains 43 substances. The binder as well as some solvents and pigments are high volume chemicals. The other chemicals are produced by the suppliers in the tonnage band of 100–1000 t/yr. The previously applied profitability requirements for substance registration yield substance registration-induced price increases between 0.17% and 13.3%.

The substance mark-ups induce a cost increase of the varnish of 0.045 €/kg, that is 1.1% of present product price. To get an idea of the cost impact for industrial downstream users, coated 0.5 mm flat steel was considered as application. With a steel price of 1.50 €/kg, the semi-finished flat steel product costs around 6.50 €/m². Coating with 0.20 kg varnish per m² steel yields a cost increase of 0.009 €/m², which is 0.14% of the present coated semi-finished steel product price.

With a negative NPV, the varnish binder has to be considered vulnerable, endangered by production phase-out. If this happens, the need for a reformulation arises. The binder in the varnish is an essential functional

component and its substitution requires a basic reformulation (redesign) of the preparation. For this, companies reported expenditures between €40,000 and €50,000 per preparation. This would increase the price of the paint by another 0.02 €/kg, resulting in a total cost increase of the varnish by 1.6%.

These examples exhibit the heavy dilution of substance registration cost impacts down the supply chain, from substance manufacturing over preparation formulation to the production of articles. Cost impacts on the end users outside the chemical industry turned out to be not serious, even if more than one chemical is involved in the production of a certain article (product). Table 4 summarizes the results of the above discussed cost impact assessment on the chemicals supply chain.

4.5. Cost burdens for companies

If cost impacts on products are not as serious as often worried about, the question arises how seriously do REACH registration cost burdens impact company's economic performance? The cost impact analysis of the two substance suppliers in the company sample exhibits quite a moderate cost burden, neither endangering competitiveness nor causing harm to business. Total registration costs are related to turnover and profit cumulated over 11 yr, because this is the time horizon foreseen for substance registration under REACH.

For one substance supplier, cumulated turnover and gross profit margin are eroded by 1% and 9.6%, respectively, for the other by 0.3% and 4%. These impacts have to be borne by the two companies if market conditions do not allow any cost pass-on to downstream users of the chemicals. This, obviously, is a rather pessimistic assumption. A similarly insignificant cost impact occurs for one of the preparation formulators. His turnover and gross profit margin are lowered by 0.6% and 4.3%, respectively, if his substance suppliers fully pass on substance registration costs to him and, at the same time, he is not able to pass these costs on to his own markets at all.

In contrast, a challenge becomes visible for the second preparation formulator. This company handles the huge number of 142 purchased substances. Full pass-on of substance registration costs to him would increase the costs for substance purchasing by €358,975. The cost impact analysis is shown in Table 5. There the applied registration costs in column 4 are taken from Table 3 and take the mix of dangerous and non-dangerous substances in the company's portfolio into consideration.

This is the cost burden for the company if the substance supplier fully passes on his registration costs to the varnish manufacturer. If, at the same time, the varnish manufacturer is not able to pass on these costs to his markets, the cost increase will cut his profit. Again, this is a worst case assumption. If it becomes reality, substance registration cost impacts would cut company's gross profit by 58%.

Table 4
Substance registration cost impacts along the supply chain

Chain stage	Preparation code	First	Second	Third	
	Kind of preparation	Steel de-scaling agent	Accumulator mass	Varnish	
Substance manufacturer	Number of contained substances	3	6	43	
	Substances prices	0.30–11.00	2.40–9.10	0.44–12.98	€/kg
	Substances mark-up	0.0049–0.077	0.0025–0.364	0.0025–0.58	€/kg
		0.3–3.5	0.03–14.5	0.17–13.3	%
Formulator	Preparation price	1.54	7.90	4.10	€/kg
	Preparation mark-up	0.0125	0.16	0.0451	€/kg
		0.8	2.0	1.1	%
Articles producer	End product	Shaped flat stainless steel	AA type accumulator	Coated 0.5 mm flat steel	
	End product price	2.20 €/kg	2.00 €/piece	1.50 €/kg	
	End product mark-up	0.00009 €/kg	0.0011 €/piece	0.0090 €/m ²	
		0.004	0.06	0.14	%

Table 5
Registration costs impact to a varnish producer

Tonnage band (t/a)	Number of substances	Purchased amount (t/a)	Registration costs (€/substance)	Production volume of substance suppliers (t/a)	Purchased substances price rise (€)
<1	63	15	0	1	0
1–10	44	184	13.100	10	64.897
10–100	23	680	81.910	100	149.962
100–1000	10	2.644	194.716	1.000	138.611
>1000	2	2.537	241.770	30.000	5.505
Total	142	6.060			358.975

bringing the profit margin down to 1.7% (Table 6), which must be considered insufficient. Furthermore, an unknown additional cost burden arises with cost increases of purchased preparations.

These business examples show that cost impacts of the new chemicals regulation REACH are quite diverse. A tailor-made case analysis is indispensable to estimate individual cost burdens.

5. Conclusion and outlook

Implementation and enforcement of the chemical Acquis Communautaire placed companies in the EU New Member States in a favourable starting position for future REACH adoption. Due to this, safety data sheets, the central information tool under REACH for information exchange along the supply chain, is well established in the companies. On the other hand, investments in equipment and plants to comply with the EU air, water and waste standards are still under way when REACH comes into force. Therefore companies in the New Member States will be stressed by two cost burdens.

The raw materials supply from non-EU countries, which will be affected by the REACH registration scheme, was

Table 6
Worst case profit cut of a varnish producer due to substance registration cost burdens

	Without registration costs impact	With registration costs impact
Turnover	15,500,000 €	15,500,000 €
Gross profit	620,000 €	261,025 €
Gross profit margin	4%	1.7%

revealed as not as serious an obstacle for REACH implementation as was thought initially. It may force some individual companies to adapt their past chemicals supply scheme, but no appreciable impact on the chemical industry sector altogether is expected. Especially affected are importers in the NMS for whom import of cheap raw materials from non-EU eastern countries was a lucrative business in the past.

Almost all companies see a need to improve information exchange about chemicals handled along the supply chain. The tradition of such cooperation as part of supply chain management is restricted to operational business needs and interests. A typical example is support and service for preparation formulators in the application of chemicals by

substance manufacturers. Such competences are excellent starting points for the enforcement of the information exchange provisions required by REACH. But the challenge is to build an information network which covers all levels of the supply chain, where substance manufacturers or importers, preparation formulators, and downstream users participate. This will become one of the most crucial lessons for companies to learn under REACH. Actually, there is only a poor understanding of this future need. None of the interviewees was able to draw even a rough picture of his company's approach to identify substance uses and to perform exposure assessments and risk characterizations for downstream users of own chemicals.

The appreciation of REACH by the companies is rather adverse. Positive effects like enhanced supply chain management for chemicals are admitted, but not taken into consideration in the overall rating of REACH. Bureaucracy, high cost burdens, and gain of competitive advantages of EU-15 companies are the prominent fears.

One crucial difference to EU-15 companies is the companies' low R&D budgets. Responsible management recognizes the importance of R&D, but so far it does not translate this knowledge into operational business. The adverse effects on REACH are obvious.

On-going privatization, restructuring and modernization of production, efforts to comply with the acquis, and the tougher competition with EU-15 companies after accession absorb much of the companies' management capacity. Coping with all these changes within a relatively short time span is a typical challenge for companies in the eastern NMS, not present in EU-15. This may be one reason why knowledge about REACH still lags behind that of the EU-15 companies.

One lesson learned from the case studies is that the health, safety and environment (HSE) management in the interviewed NMS companies is better equipped with personnel resources than in EU-15 (Ostertag et al., 2004). This may have to do with lower wages. The well equipped HSE staff favors adoption of REACH because HSE management plays a key role in this process, e. g. for substance testing, exposure assessment and risk characterization.

Although the NMS adoption of EU legislation has brought the starting point for REACH to a common level in EU-15 and the NMS, we saw that different company environments are present which may cause the one or other additional challenge compared to EU-15. Overall, the results of the case studies which are admittedly a limited but well-founded experience of daily business, give no indications that REACH adoption will bring significant drawbacks to companies in the NMS.

In this paper individual challenges for companies and industrial sectors have been discussed, but for the European chemical industry as a whole, there is no question that this industry will be able to cope with REACH burdens. Within 11 yr, which is the time given for substance registration, the European chemical industry's

turnover will accumulate to €5800 billion, and its profit at least to €170 billion. The total cost burden of REACH, as assessed by the European Commission, approaches €7.5 billion for the chemical and other industries (CEC, 2003b). Exclusively charging the chemical industry with these costs would erode its 11 yr turnover by 0.13% and its 11 yr profit by 4.4%. Even if one adds 50% to the Commission's estimate of REACH's direct implementation costs, as a US consultant has done (Ackermann and Massay, 2004), the cost burden remains far below critical thresholds and will not harm the strong international competitive position of Europe's chemical industry, nor trigger relocation of downstream industries outside Europe. On the other hand, better knowledge of chemical properties and risks acquired through REACH, and—most important—the communication of this information along the supply chain is an inestimable advantage compared to the present situation. It reduces commercial risks for companies working under legal liabilities. Moreover, better control and safe management of chemicals is occupational and public health protection, a benefit the European Commission assesses to be in the order of €50 billion over a period of 30 yr (CEC, 2003b).

Acknowledgements

The authors are grateful to DG Environment and DG Enterprise of the European Commission which made it possible to undertake this study. The authors would like to thank Mr. Luis Delgado and Oliver Wolf of the Sustainability in Industry, Energy and Transport Unit of European Commission's DG JRC IPTS in Seville for sponsoring the project.

References

- Ackermann, F., Massay, R., 2004. The true costs of REACH. A study performed for the Nordic Council of Ministers, Global Development and Environment Institute, Tufts University, Medford, MA.
- Andrijewski, M., 2001. Current state in transposing EU Chemical Legislation in Poland. Presentation at the Workshop on Application of EU Chemicals Legislation in the Candidate Countries, Istanbul.
- Bláha, K., 2001. Chemicals Legislation in the Czech Republic: Two Years of Implementation. Presentation at the Workshop on Application of EU Chemicals Legislation in the Candidate Countries, Istanbul.
- Bucknall, J., Cestti, R., Hughes, G., 2000. The Czech Republic—complying with the European Environmental Directives. World Bank, Washington, DC.
- CEC, 1998. Meeting the challenge of enlargement with the candidate countries in Central and Eastern Europe. COM (98) 294, Brussels.
- CEC, 2001. Communication from the commission: the challenge of environmental financing in the candidate countries. COM (2001) 304 final, Brussels.
- CEC, 2002. Environmental Protection Expenditure in Accession Countries. Office for Official Publications of the European Communities, Luxembourg.
- CEC, 2003a. Proposal for a Regulation of the European Parliament and the Council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency and amending Directive 1999/45/EC and Regulation (EC) (on Persistent Organic Pollutants), Commission of the

- European Communities, COM(2003) 644 final, 29.10.2003, vol. I–VI, Brussels.
- CEC, 2003b. Extended Impact Assessment. Regulation of the European Parliament and of the Council concerning REACH. Commission staff working paper, SEC (2003) 1171/3, Brussels.
- CEFIC, 1998. Impacts of the Commission's White Paper on the Chemical Industry in the CEE Countries—Final Reports, Brussels, CEFIC.
- CEFIC, 2005. European Chemical Industry Council, Economic Time Series, <http://www.cefic.org/factsandfigures/level02/append_index.html>.
- CMIT, 2004. Czech Ministry of Industry and Trade, Panorama of the Czech Industry 2003, Prague, <http://www.mpo.cz/xqw/webdav/UTF8/dms_mpo/getFileinternet/18767/19274/panorama_en.zip>.
- DANCEE, 2002. The Environmental Challenge of EU Enlargement in Central and Eastern Europe, Copenhagen.
- EDC and EPE, 1997. Compliance Costing for Approximation of EU Environmental Legislation in CEEC.
- EUROSTAT, 2005. Statistical Office of the European Community, 'Industrie, Handel und Dienstleistungen', <<http://europa.eu.int/comm/eurostat/>>.
- Fraunhofer ISI, PBF, Nordbeck R, IPTS, BEF, Oekopol, 2005a. Implementation of REACH in the New Member States, Part One: Overview of the chemical and specialty chemicals sector in the New Member States, European Commission, DG Joint Research Centre, Institute for Prospective Technological Studies, Seville, Spain.
- Fraunhofer ISI, PBF, Nordbeck R, IPTS, BEF, Oekopol, 2005b. Implementation of REACH in the New Member States, Part Two: Business case studies in selected New Member States, European Commission, DG Joint Research Centre, Institute for Prospective Technological Studies, Seville, Spain.
- Grabner, A., 2001. Slovene Chemicals Legislation. Presentation at the Workshop on Application of EU Chemicals Legislation in the Candidate Countries, Istanbul.
- Joint Commission/CEFIC Working Group on Enlargement, 2000. Transposition and Implementation of Chemicals Acquis by the Candidate Countries of Central and Eastern Europe—Recommendations for Development, Brussels.
- Lang, K., 2002. Summary of the Transposed Chemicals Acquis in the Candidate Countries in Central and Eastern Europe. Presentation at the Conference "Making the Enlarged Internal Market for Chemicals a Reality", November 2002.
- Ostertag, K., Marscheider-Weidemann, F., Angerer, G., et al., 2004. Analysis of the Costs and Benefits of the New EU Chemicals Policy: An Examination Based On Selected Sectors Taking into Account Effects on Competitiveness, Innovation, Environment and Health, Research Report of the Federal Environmental Agency (Umweltbundesamt), Berlin, Germany.
- Pedersen, F., de Bruijn, J., Minn, S., van Leeuwen, K., 2003. Assessment of Additional Testing Need Under REACH, Effects of (Q)SARS. Risk Based Testing and Voluntary Industry Initiatives, European Commission, DG JRC, Institute for Health and Consumer Protection (IHCP), Ispra, Italy.
- PIPC (Polish Chamber of Chemical Industry), 2003. Production Sold of Chemical Industry According to Branches in 2002, memo.
- PIPC (Polish Chamber of Chemical Industry), 2004a. Chemical industry in Poland. Annual Report 2003, downloads: <http://www.pipc.org.pl/raport_2004%20cz-1.pdf and <http://www.pipc.org.pl/rap.pdf>>.
- PIPC (Polish Chamber of Chemical Industry), 2004b. Characterisation of the Chemical Industry in Poland. Interview with PIPC members, 16 September, 2004.
- REC, 1996. Regional Environmental Centre, Approximation of European Union Environmental Legislation: case studies of Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, the Slovak Republic and Slovenia, Budapest, REC.
- SCHP, 2003. Association of Chemical Industry of the Czech Republic, Annual Report 2003 on the development of chemical industry in Czech Republic, Prague, download: <http://www.schp.cz/schp/zpravy/Annual_Report_2003.zip>.
- Soil and Water Ltd., 1999. Development of Action Programme for Implementation of EU Legislation on Chemicals in Lithuania, Final Report DISAE Project LIT-109.
- Statistical Office of Estonia, 2005. General Export and Import by Commodity Groups, <http://pub.stat.ee/px-web.2001/I_Databas/Economy/Economy.asp>.
- Stocker, E., Anwender, E., 2004. Assessment of Institutional Capacities: Implementation of European Legislation in the Area of Chemical Substances in the Czech Republic. Final Report Twinning Project CZ/2002/IB/EN/01, Prague.
- Witzani, H., Andrijewski, M., 2004. Strengthening of the Administrative Capacity for Risk Assessment and Chemicals Control. 3rd Quarterly Report Twinning Project PL/IB/2002/OT/04.

Rural industries and water pollution in China

Mark Wang*, Michael Webber, Brian Finlayson, Jon Barnett

School of Social and Environmental Enquiry, The University of Melbourne, Victoria 3010, Australia

Received 8 May 2006; received in revised form 7 December 2006; accepted 12 December 2006

Available online 26 February 2007

Abstract

Water pollution from small rural industries is a serious problem throughout China. Over half of all river sections monitored for water quality are rated as being unsafe for human contact, and this pollution is estimated to cost several per cent of GDP. While China has some of the toughest environmental protection laws in the world, the implementation of these laws in rural areas is not effective. This paper explains the reasons for this implementation gap. It argues that the factors that have underpinned the economic success of rural industry are precisely the same factors that cause water pollution from rural industry to remain such a serious problem in China. This means that the control of rural water pollution is not simply a technical problem of designing a more appropriate governance system, or finding better policy instruments or more funding. Instead, solutions lie in changes in the model that underpins rural development in China.

© 2007 Published by Elsevier Ltd.

Keywords: China; Water; Pollution; Rural industry; Transition

The house is new, the money is enough, but the water is foul, and life is short.

(A popular saying in coastal China, from Schmidt, 2002)

1. Introduction

Water supply and quality are fundamental issues in China. A few years ago, the debate about who will feed China emphasised scarcity of farmland and the food crisis (Brown, 1995). Yet the most critical resource in China is not land or food, but water (as Brown later (2001) came to recognise). Not only are per capita water resources limited (Niu and Harris, 1996) and the spatial distribution of water resources extremely uneven, there is also significant waste of water. This waste is related to inefficient irrigation practices, leaking water pipes, and water pollution. Growing municipal and industrial waste discharges, coupled with limited wastewater treatment capacity, are the principal drivers of water pollution. About two-thirds of the total waste discharge into rivers, lakes and the sea derives from industry, and about 80% of that is untreated.

Most of the untreated discharge comes from rural industries.

Rural industries stand out as one of the most spectacular respondents to China's 1978 economic reform. They represent a middle ground between private and state ownership and have not developed in any other country on such a large scale and at such a rapid rate. They have become the driving force behind China's economic growth and a significant engine of China's transition, with double-digit growth rates since the late 1970s. To a large degree, this growth of rural industry was neither planned nor anticipated (Bruton et al., 2000).

However, the environmental cost of China's rural industrialisation is enormous. Rural industry consumes massive quantities of water and pollutes a large proportion of rural water bodies (Anid and Tschirley, 1998; Wheeler et al., 2000). While a few large rural enterprises have advanced technology and sophisticated wastewater treatment facilities, rural enterprises are characterised by their small scale, outmoded technology, obsolete equipment, poor management and heavy consumption of water resources (Qu and Li, 1994). As a result, water pollution is a serious problem wherever there are rural industries. Over 80% of China's rivers have some degree of contamination (Qi et al., 1999). China's 2002 *State of the*

*Corresponding author. Tel./fax: +61 3 9349 4218.

E-mail address: myw@unimelb.edu.au (M. Wang).

Environment Report shows that 70% of the 741 river sections monitored were unfit for human contact (pollutions levels at or above Grade IV standard) (SEPA, 2002). The most polluted was the Hai River, where 86% of monitored sections were unfit for human contact.

China's environmental policy is widely considered as a comparatively success in urban areas, with marked declines in urban pollution (Florig et al., 1995; Zhang et al., 1997; Abigail, 1997). However, these efforts rarely reach the rural areas, where environmental policy is not effective (Florig et al., 1995; Zhang et al., 1997; Abigail, 1997). One of the names for this phenomenon is a policy 'implementation gap' (Chan et al., 1995). It is the aim of this paper to understand the reasons for this implementation gap. Previous studies have attributed it to a combination of factors, including legislative shortcomings, poorly designed policy instruments, an unsupportive work environment for environmental regulators, and a pro-growth political and social environment (Chan et al., 1995; Wong and Hon, 1994; Ross, 1992). These are important; however, we argue that the most fundamental factors causing rural water pollution are the very same factors that have underpinned the economic success of rural industry. The problem of water pollution, we argue, is therefore unlikely to be remedied by discrete institutional changes, and instead requires a transformation of the models associated with rural development.

We present this argument in the following sections. We begin by explaining the characteristics of rural industry and the reasons for its success. This will be followed by a summary of the water crisis as it relates to rural industry. In Sections 4–6, we explain our interpretation of the reasons for water pollution problems in rural China. Section 4 is focussed on how the characteristics of rural

industry contribute to rural water pollution; in Section 5, we review the institutional arrangements that make it difficult to control water pollution by rural industry; and in Section 6, we emphasise how the on-going transition has made water pollution control more difficult. The conclusion integrates this evidence and explains our interpretation of it.

2. Rural industrialisation

The rise of rural industry is one of the outcomes of China's transition. The rural household responsibility system introduced in the late 1970s released hundreds of millions of peasants from the farming sector. To accommodate increasing under- and unemployment in China's countryside, the central government allowed industrial development in rural areas (Lin, 1997; Oi, 1995; Lieberthal, 1995). Peasants were encouraged "to leave the land but not the village" (*litu bu lixiang* in Chinese) (Tan, 1993; Wang, 1997). As a result, over 120 million peasants abandoned agriculture to work in the emerging rural industrial sector.

By the end of 2003, China had nearly 22 million rural enterprises, producing output valued at ¥191 billion, which is 15 times that of 1988 (National Bureau of Statistics of China (NBSC), 2004). The total added value contributed by rural enterprises accounted for over 30% of China's GDP, compared with only 12.3% in 1988. In 2003, rural industry employed over 135 million from the rural labour pool without need for state investment (Table 1). Many 'made-in-China' products are manufactured by rural enterprises. From 1995 to 2000, exports by TVEs grew at an average annual rate of 10% (NBSC, 2004; Fu and Balasubramanyam, 2004). In 2004, TVE exports accounted for one-third of China's total exports (NBSC, 2004). So,

Table 1
TVEs size: number of enterprises and employment by category of ownership: China, 2003

Enterprise category (registered as)	Number of enterprises		Total employees		Employees/enterprise
	(000)	%	(000)	%	Persons
Urban ^a	197.22	100	25 6390	100	1331
SOEs ^b	34.28	17.4	68 760	26.8	2005
Others	162.94	82.6	187 630	73.2	1151
TVEs total	21 850.8	100	135 729.3	100	6
Domestic	21 796.1	99.7	128 433.2	94.6	6
Collective	292.1	1.3	12 359.7	9.1	42
Share-holding cooperative	84.8	0.4	3670.2	2.7	43
Joint operation	26.7	0.1	666.5	0.5	25
Limited liability	202.8	0.9	10 272.8	7.6	51
Share holding	32.4	0.1	1819.4	1.3	56
Private	2542.4	11.6	38 708.6	28.5	15
Individual	8940.1	40.9	29 916.9	22.9	3
Others	9674.8	44.3	31 019.1	22.9	3
Hong Kong/Macau/Taiwan	38.0	0.2	4928.6	3.6	130
Foreign	16.7	0.1	2367.6	1.7	142

Source: NBSC (2004).

^aIncluding all SOEs and these non-state enterprises with annual sales over 5 million Yuan. Almost all of them are located in either cities or towns.

^bSOEs include all SOEs and all those enterprises with State as the major share holder.

although TVEs have simple production methods and have obtained technical expertise from staff formerly employed in state-owned enterprises (SOEs) (Peng et al., 1997), they have clearly outperformed SOEs in terms of growth, job creation and profitability (Chen and Jefferson, 1999).

In this paper, rural industry is taken to mean all businesses located in rural areas and involved in non-agricultural activities. Rural industry can be broadly classified into two categories of ownership: township (*xiang*) and village (*cun*) enterprises as collective enterprises (strictly speaking, TVEs), and private rural enterprises. This latter category includes a heterogeneous group of enterprises of different sizes, types of business, organisational structure, and ownership. However, there are many different rural industrialisation models in China. Based on the sources of initial investment, structure of ownership, industrial orientation, income distribution and management systems, Chinese economists (Ding et al., 2004; Zuo, 2001) classified TVEs into three generic models: (1) the Sunan model (Southern Jiangsu model); (2) the Wenzhou model; and (3) the Pearl River Delta model. They are different because of their firm size, technical sophistication and their initiative drivers. The Sunan model is characterised by a dominant initial investment by township and village government; the Wenzhou model by private investors; and the Pearl River Delta model by foreign direct investment and export-oriented manufacturing. These models are neither static nor exclusive.

The boundary between TVEs and private enterprises has blurred over time (Weitzman and Xu, 1994; Naughton, 1994; Che and Qian, 1998). Some TVEs have been transformed into joint stock co-operatives, others partially privatised (with local governments continuing to hold a stake) and others even wholly privatised by the sale of the entire stake to managers, employees and/or community residents (Xu and Tan, 2002). Many private investors have become shareholders of TVEs in Sunan and many private enterprises in Wenzhou have been merged to form a collectively owned corporation. The ownership of rural enterprises can be private or collective or a combination of shareholders in local communities, village and township governments, and local and foreign private owners (Table 1).

Many explanations have been offered for the remarkable success of rural industry in the Chinese economy (e.g., Byrd and Lin, 1990; Nee, 1992; Chang and Wang, 1994; Ho, 1994; Weitzman and Xu, 1994; Naughton, 1994). Success has been seen to be due to the shortage of major product markets in the 1980s in China, cheap rural labour, tax concessions from local governments, and the problems of state industries (Byrd and Lin, 1990; Ho, 1994; Naughton, 1994). Other contributions to their success have come from their unique institutional structure that facilitates co-operation through implicit contracts among community members (Weitzman and Xu, 1994); the inter-organisational relationship between TVEs and local governments (Chen and Jefferson, 1999; Fan, 1997) and

TVEs' capacity to adapt and configure their strategy in response to the external competitive environment (Luo et al., 1998, 1999). However, two of the success factors are especially important for the pollution performance of rural industry.

One of the most important of these is the fact that many rural enterprises face hard budget constraints. For private enterprises, family saving is the major source of capital: they have limited opportunity to access bank loans or attract government investment. TVEs are in a slightly better position because they receive local township or village government sponsorship, though the limited tax base means that such financial support is insufficient. TVEs also have limited access to financing through the banking sector. Even though local governments often act as guarantor and supporter for their TVEs' applications for bank loans, banks are hesitant to lend to what they consider to be risky enterprises, and banks have been encouraged in the past by the central government to support the ailing SOE sector instead. Therefore, rural industries have very strong incentives to ensure that their operations are profitable. Such hard budget constraints force rural industrial operators to focus mainly on generating profits for their survival, and they are therefore extremely reluctant to incur costs to conserve water use or control pollution.

Another common driving force for the success of rural industry is the strength of local government networks. The existing research has described the advantages stemming from the TVEs' peculiar 'internal institutional form' that facilitates cooperation through implicit contracts among community members locked into an ongoing relationship (Nee, 1992; Weitzman and Xu, 1994). The success of TVEs is mainly due to the interaction of TVEs with the whole community through a "set of interlocking financial, administrative, personnel, and other ties" (Byrd, 1990, p. 74) and the existence of a strong 'cooperative culture' (Weitzman and Xu, 1994).

The local township or village governments give birth to TVEs, and they appoint TVE managers, and act as financial intermediaries by financing TVEs and using their connections to find alternative sources of financing. In return, local governments share the profits generated by their TVEs. With an inefficient legal system, the local government is the only body that can settle disputes arising from the operation of TVEs. As we will argue, such a 'parenting' relationship is one of the key factors to the 'implementation gap' identified in China's environmental protection system.

For rural private enterprises, keeping good *guanxi* (in the Chinese business world, *guanxi* is understood as the network of relationships among various parties that cooperate together and support one another) with local government is extremely important for survival. Few private enterprises depend on local government for financial assistance or marketing their products, but they keep close *guanxi* with local officials, including local

environmental monitors and inspectors. Such close *guanxi* can include bribery in various forms. As our fieldwork interviews (with peasants and village heads in four villages in Jinan and Zhengzhou and with their municipal/provincial government water officials in October 2005) show, managers are sometimes ‘informed’ before environmental inspections are carried out so that they can temporarily shut down their polluting functions, only to return to ‘normal’ operation after the inspection is over.

There are, then, a variety of reasons for the emergence of this remarkable growth engine of the Chinese economy. Some of these reasons are directly related to the polluting tendencies of that engine. Before we examine these reasons more deeply, we briefly identify the main characteristics of the water pollution problems that are attributable to rural industry.

3. Rural industry and the water crisis

Rural industries have been widely criticised for their waste of natural resources, including water resources, due to substandard equipment and basic technology (Shen et al., 2005; Edmonds, 1994; Wong, 1999). Most rural industries have no wastewater or hazardous waste treatment facilities. Almost all wastewater is directly discharged into local river systems. According to the State Environmental Protection Agency (SEPA, 2002), the total discharge of sewage and wastewater in China was 62 billion tons in 2000, of which only 24% was treated up to the national standard; the rest was not treated or treated but not up to the national standard before being discharged or used for irrigation. It has been conservatively estimated that TVEs alone discharge over 10 billion metric tons of wastewater per year, which is half of the industrial wastewater discharge in China. According to Zhao and Wong (2002), 76% of wastewater created serious pollution while only five per cent of wastewater was discharged at the required standard. As the data from SEPA demonstrates, this untreated wastewater has polluted most Chinese rivers, and 3/4 of China’s lakes have significant levels of pollution (SEPA, 2002).

Since the year 2000 there has been a decline in the number of environmental accidents in China (Table 2).

Whereas the number of environmental accidents in China grew rapidly between 1997 and 2000, the number fell between 2000 and 2003. Of all the classes of environmental accidents, water pollution accidents are the most frequent, and they have become more frequent between 1997 and 2003. Of course, some of these are from SOEs and large urban industries, but the enforcement of environmental regulations in cities, anecdotal evidence and media reports suggest that the majority of these accidents are caused by rural enterprises. There is probably also significant under-reporting of water pollution accidents from rural industries due to the close connections between environmental regulators and rural entrepreneurs (discussed later).

Water pollution and consumption by rural industry are related to the type of industrial activity. The major water polluters include an array of industries such as paper and pulp milling, chemical manufacturing, metal casting, and brick making that produce large quantities of wastewater, adding nitrogen, phosphates, phenols, cyanide, lead, cadmium, mercury, and other pollutants to the water near rural residential areas—the same water that is used for drinking. The worst water polluters are those rural industries related to papermaking, cement and bricks (Xu, 1999). For example, TVE paper mills in Henan Province consume 3.6 times more water per unit of paper production than do state-owned paper mills (according to interviews in 2005). In China, these three sectors have increased their production dramatically: between 1978 and 1993 the volume of paper produced from rural industries increased from 0.4 to 10.3 million tons, the volume of cement from 3.3 to 127.6 million tons, and the number of bricks from 73 to 494.8 billion (NBSC, 1995). With these large increases in production have come concomitantly large increases in emissions of pollutants into waterways.

Water pollution and consumption by rural industry are also of course related to the density of production sites. Xu et al. (2001) compare the density of TVEs in the provinces of China (using TVEs output value per land area) and the ratio of wastewater to runoff (using TVEs discharged wastewater per cubic metre of runoff). Their results show a clear positive correlation between the economic density of TVEs and the runoff load that they generate ($r = 0.93$). Provinces in which TVEs produce a lot of output per unit

Table 2
Number of environmental accidents in China (1997, 2000, 2003)

Type of accident	1997		2000		2003	
	Number of accidents	%	Number of accidents	%	Number of accidents	%
Total accidents	1992	100.0	2411	100.0	1843	100.0
Water pollution	986	49.5	1138	47.2	1042	56.5
Air pollution	752	37.7	864	35.8	654	35.5
Solid waste pollution	55	2.8	103	4.3	56	3.0
Noise pollution	119	6.0	266	11.0	50	2.7
Other pollution	80	4.0	40	1.7	41	2.2

Source: NBSC (2004).

area have a high ratio of wastewater to runoff; these are principally the provinces along the coast and in the south. There are also some provinces in the drier north and west (such as Tianjin, Hebei, Henan, and Ningxia) that have high levels of rural industrialisation and associated high levels of wastewater to runoff.

As a result of this pollution, water is now a public health risk in China—arguably even more so than air pollution (Schmidt, 2002). Over half of China's 1.3 billion people drink water contaminated with chemical and biological wastes such as petroleum, ammonia nitrogen, volatile phenols, and mercury (SEPA, 2002). The health impacts of water pollution have been costed at US\$3.9 billion annually (World Bank, 1997), and there is evidence that the increasing incidence of cancers is associated with pollution of drinking water (Banister, 1998; UNDP, 2002). Maurer et al. (1998) list a number of studies that link wastewater from TVEs with health effects such as elevated cancer rates and abnormal pregnancy outcomes (for example, spontaneous abortions, premature births, birth defects). Xu (1992) finds that the incidence of disease differs between more and less polluted regions; in the most polluted areas, chronic disease is up to three times more common than in the nation as a whole. Xu (1992) also finds that average mortality in polluted areas is 4.7 per 1000, higher than the average of 3.6 in less polluted areas.

Environmental degradation costs China dearly, though that cost has been estimated using various techniques with varying outcomes (Smil and Mao, 1998; World Bank, 1997). The aggregate annual cost of environmental damage has been estimated to be between 4.5% and 18% of GDP: water pollution is estimated to cost between 0.6% and 4.5% of GDP, with estimates centring on 1.7%, about the same as the cost of air pollution.

A number of analysts argue that poor regulation and enforcement of environmental measures allows many small rural enterprises to operate without waste treatment facilities (Jahiel, 1997; Lin, 1997; Qu and Li, 1994). The Chinese government has imposed a pollution levy on industrial polluters since 1979 and introduced a new incentive-based pollution control programme in which the environmental performance of firms is rated from best to worst using five colours—green, blue, yellow, red and black—and the ratings are disseminated to the public through the media (Xie and Florig, 1997; Wang et al., 2004). This system is arguably one of the most complete pollution levy systems in developing countries. However, the levy has been criticised as being too low to give polluters incentives to reduce their emissions, for the water pollution fees are small relative to the marginal costs of pollution control (Wang, 2000; Wang and Wheeler, 2000; Vermeer, 1999; Krupnick, 1992; World Bank, 1992; Sinkule, 1993). Thus, many enterprises simply choose to pay the effluent or emissions fee rather than incur the costs of pollution control. Further, because effluent and emissions fees and fines can be lower than even the operating and maintenance costs of pollution control equipment,

many enterprises that install pollution control equipment have little incentive to operate it once installed. For example, Sinkule (1993) reveals a case in which the operating costs for wastewater treatment were more than eight times the fee imposed for not operating the equipment.

In addition to the pollution levy, since the mid-1990s the Chinese government has taken several important steps towards pollution control. The government will punish whoever is in violation of State environmental regulations and causes serious pollution of land, water or the atmosphere. The penalty is up to 3 years imprisonment and a fine. Up to 7 years imprisonment can be imposed in the most serious cases. (The most extreme measures include capital punishment under the codification of environmental crimes introduced in 1997.) All relevant ministries of the central government—including SEPA, the Ministry of Agriculture, the State Planning Commission and the State Economic and Trade Commission—issued new regulations to curb pollution by TVEs in 1997. The new regulations hold the county magistrates and township mayors responsible for environmental regulation enforcement. In addition, the central government will cut off all state funding for rural enterprises that are considered environmental hazards. Governments has warned that it may revoke business licenses, cut off power supplies and state funding or detain managers on criminal charges to bring polluting TVEs into line (*South China Morning Post* (Hong Kong), 17 December 1996).

The central government has set up tough regulations and targeted certain geographical areas as priorities for pollution control. In 1997, it started a so-called '33211' programme targeting priority pollution control projects in geographical areas which were experiencing major environmental problems: three rivers (Huai, Hai, and Liao), three lakes (Tai, Chao and Dianchi), two control regions (which were major sources of SO₂ and acid rain), one sea (Bohai), and one city (Beijing—in consideration of the 2008 Olympic Games).

The Huai River valley is an area of particular concern to the central government. The director of the SEPA has himself inspected the Huai River valley regularly and ordered the closure of thousands of polluting factories. The water pollution control campaign in the Huai River valley is the first river valley based, large-scale water pollution control program conducted in China. In these programs, China has targeted 15 categories of small rural enterprises, such as paper pulp mills, textile mills, dyeing mills, small chemical plants, small breweries and small currying mills—the '15 small' enterprises. There has been closure of more than 40 000 of these 15 small heavy pollution factories on the Huai River and 70 000 throughout China (Xi and Xu, 2002).

The experience of the Huai River valley in water pollution shows that the political will of the government, especially at the national level, is critical to the success of water pollution control. However, the introduction of

tough environmental legislation and closure of many polluting factories in rural areas have not solved the water pollution problem. This is mainly because the legislation and implementation as well as the campaigns have failed to consider the unique nature of rural enterprises and the difficulties faced by environmental law enforcement in rural areas. In the remainder of this paper we will discuss these factors, under three headings—the nature of rural industry (its size, locational characteristics and relation to local officials), the institutional framework that prevents agencies from effectively enforcing environmental policies and regulatory mechanisms, and the on-going economic transition.

4. The characteristics of rural industry

4.1. Small is not beautiful

One of the common characteristics of rural industry is small enterprises. The mushrooming Chinese economy allows a few small rural enterprises to become large and some successful rural enterprises have grown to employ well over 1000 workers. Likewise, those rural enterprises registered as foreign or Hong Kong/Macau/Taiwan owned employed on average over 130 workers (Table 1). Nevertheless, the average size of all rural enterprises was only six employees in 2003. In fact, more than 93% of all the rural enterprises are run by rural households (MoA, 1995). Most rural enterprises are relatively small; indeed, most private enterprises in rural China are family run businesses, many using outdated methods and primitive production technology with low energy and resource use efficiency and high pollution (Liu, 1992).

There is, though, some geographical variation in the size and productivity of rural enterprises across China

(Table 3). Along the coast, rural industries tend to employ more workers and to have higher values of output per worker than in inland areas. Tibet is a notable exception—here rural industries are large but the value of their output per worker is low. Overall, the smallness and relatively low value of output per worker means that these industries lack the resources to manage their waste streams effectively (Lin, 1997).

4.2. Dispersion of rural industries

In addition to their smallness, China's rural enterprises are also spatially dispersed. SOEs and other large non-state enterprises with annual output of more than ¥5 million are mostly located in the 669 major cities of China, but rural enterprises are located almost in every corner of China's territory. The more than 21 million rural enterprises are spread widely in 43 735 towns and 734 715 villages. This dispersal is a significant point of difference from western economies. Since the Industrial Revolution, urban areas have been the sites for most industries in the west, whereas rural industrial development on a massive scale has never occurred (Fothergill, 1985; Fulton, 1974).

Such a pattern of enterprises scattered in villages and townships makes environmental monitoring difficult. Not only is it difficult to monitor rural enterprises because of their dispersed locations, it is also the case that China's environmental monitoring stations are under staffed. In China, there were 46 000 environmental monitoring staff and inspectors in 2003 most of whom are located in urban administrative centres. Even if all environmental monitoring staff were assigned to monitor water pollution from rural enterprises, each staff member would have to be responsible for 111 rural enterprises. Because environmental monitoring staff are insufficient to mount frequent

Table 3
China's rural industry: average size and output value per worker by province (in 2003)

Region	Employees per enterprise (persons)	Output value per worker (000 RMB)	Region	Employees per enterprise (persons)	Output value per worker (000 RMB)
China total	6.3	111	Anhui	6.3	60
Shanghai	52.1	264	Shandong	6.2	137
Chongqing	24.4	94	Heilongjiang	6.2	106
Tibet	19.8	50	Jiangxi	5.1	63
Jiangsu	10.2	184	Hunan	5	70
Tianjin	10.2	173	Guizhou	4.8	59
Zhejiang	10	204	Shaanxi	4.6	52
Guangdong	9	112	Liaoning	4.6	184
Shanxi	8.5	87	Inner M.	4.5	91
Hebei	8.4	126	Sichuan	4.5	88
Beijing	8	141	Qinghai	4.3	52
Fujian	7.7	125	Guangxi	4	53
Gansu	7.1	67	Yunnan	4	63
Henan	7	89	Ningxia	3.7	44
Hainan	6.8	61	Jilin	3.4	87
Hubei	6.6	93	Xinjiang	2.5	46

Source: NBSC (2004).

Note: an Approximate exchange rate is US\$1 = RMB 8.

unannounced inspections, rural enterprises are not properly inspected (Sinkule, 1993). For example, 2000 employees in Zhejiang EPA have to monitor over one million TVEs, which means that each staff member on average has to monitor 500 rural enterprises (in addition to urban enterprises).

Another reason for the omission of many rural enterprises from the environmental inspection system is related to the current policy of levy charges. According to the regulation, 80% of discharge fees are returned to industries for investment in waste treatment facilities; the remaining 20% goes to the Environmental Protection Bureau (EPB). This money collected from industries is crucial to the EPB's operation: to pay for employees' salaries and bonuses, research funding, environmental supervision, and environmental education campaigns. However, in the current pollution levy system, fines imposed on small enterprises for pollution violations are very small and, as we have seen, most rural enterprises are very small indeed. Thus, the revenue from pollution fee collection could drop significantly if local EPBs allocated most of their limited resources to monitoring small, scattered, rural enterprises. Such revenue is essential in the development and maintenance of local waste treatment facilities so the local EPB has to give priority to the larger enterprises and more serious polluters, leaving small rural enterprises immune. For example, Xu (1999) found that in 1998 over 90% of Jiangsu's rural enterprises were not in compliance and only a small number of the polluting enterprises were inspected by the local environmental monitoring station (which was based in the county town). Only three of these polluters received warning notices from the station, and they were all located in the town. Factories in rural areas producing similar types but greater amounts of wastewater had never been inspected by the EPB (Xu, 1999).

4.3. Administrative constraints

Local governments' close financial links to rural industry also limit their enforcement of environmental regulations. As the Chinese economy becomes increasingly decentralised, local governments are required to carry more responsibility for environmental performance. However, rural industries are the major sector of employment of local labour and also the major financial source for local government. This is due to the fact that China's system of public finance is highly centralised. Local governments are only allowed to retain a small amount of tax income to offset expenditure, normally sufficient to provide only limited support for the basic operation of a town government. However, so-called extra-budgetary revenues are not subject to budgetary supervision by higher levels of government. These revenues can be derived from supplementary agricultural, industrial and commercial taxes. Before 1980, extra budgetary revenues were largely drawn from supplementary agricultural taxes. Following the

spread of rural industries and central government's restrictions on local supplementary taxes, such 'extras' now come mainly from local enterprises. In many parts of the coastal region, the majority of local governments' extra-budgetary revenues now come from these enterprises, which have thus become a big income generator for town governments (Song and Du, 1990). Che and Qian (1998) have shown that local governments receive direct profit from local rural enterprises. Lieberthal (1995) suggests that rural industries can account for up to 80% of the community's total revenues. In addition, town and village governments protect rural industries because they are often the guarantors for rural enterprises' loans. If the rural industries were to face economic losses because of the enforcement of environmental regulations, the town and village governments would also face economic losses.

Local governments are also integrated with the management of TVEs. Village heads, other political cadres or their relatives manage many village enterprises. Such a double role becomes problematic when local environmental staff inspect pollution control activities in these industries. Conflicts of interest make the local officials reluctant to penalise what are virtually their own enterprises for waste discharges. Bribery and corruption at small and local scales are the extreme forms of such linkages.

Many local government officials and rural enterprise managers complain that pollution abatement funds do not go to rural enterprises but to urban and state owned enterprises (Wang and Lu, 1997). Local officials are also concerned about the negative impact on the competitiveness of local products in the market if strict pollution control is imposed. They believe that forcing a marginally profitable enterprise to set up an expensive waste treatment facility would be equivalent to ordering it to shut down, since the costs would have to be transferred to its products and the rising prices of its products would force it out of the market, causing rising unemployment and falling township or village revenues. In addition, local officials would not want to force their local enterprises to set up waste treatment facilities unless their competitors in other villages and towns were forced to do the same upgrading to the same standards and at the same time. They do not want their local products to lose price competitiveness.

Therefore, the multiple roles of the village and town governments, together with the uneven implementation of environmental laws, make it difficult for local authorities to penalise enterprises for environmental pollution. It is not surprising that Chan et al. (1995) found that over a quarter of local officials responsible for administering the pollution levy in the cities of Guangzhou, Zhengzhou and Nanjing disagreed with the polluter pays principle (Chan et al., 1995). Smil's comment reflects the local officials' concern: "If you are a local official, you don't want to interrupt TVEs or burden them with environmental controls. If you do that, they will just move to the next county." (quotation from Schmidt, 2002, p. 515).

In addition, the promotion criteria for local government officials are mainly related to the economic growth rate of their locality. Environmental protection is not listed as one of the criteria except if serious environmental disasters are made public and raise concern. Such a system indirectly contributes to failure in pollution control. Many local governments become more lenient towards the polluters when they are under pressure to meet their investment goals (Zhang and Ferris, 1998).

So, the financial success of a township or village government and the social success of its officials depend on the continued economic success of its enterprises—creating strong countervailing forces against environmental regulation. An immediate consequence of this dependence of local governments on rural industry is that rural enterprise managers often resist or even scorn local environmental inspectors when they come to collect discharge fees. Some studies show the existence of local resistance to the pollution levy system (Florig et al., 1995; Wang and Lu, 1997). Since local government officials support them, rural enterprise managers tend not to treat environmental monitoring staff seriously. For example, in a study of rural enterprises in Zhejiang, Xu (1999) found that only 3.2% of rural enterprises fully complied with the environmental inspectors' orders for waste treatment, nearly two fifths ignored the order, and another 16% partially complied. Many enterprises that were issued orders simply terminated production—which is good for the environment but justifies the concerns of local officials about the impacts of environmental regulation on economic growth. The limited scale of capital investment facilitates this cut-and-run behaviour.

These conditions lead to a levy collection system with two notable characteristics. The first is collection by negotiation. The amount that the EPB finally collects is the result of a negotiation between the two sides rather than based on officially set fees. Sometimes, local government officials give instructions for the fee to be collected. However, the EPBs are typically weak agencies within the local bureaucracy, so they often end up on the losing side of such a negotiation. The second feature is collection by relationship. At the local level, the levy is often collected on the basis of personal relations between the local bureaucracy and enterprises. If EPB officials have a personal relationship with the heads of enterprises, small fees are levied. In the absence of personal ties, enterprises are charged larger fees.

However, the factors that give rise to this effect are precisely the factors that have been important in encouraging rural enterprises to grow so fast and to make such a contribution to the Chinese economy and the rural workforce. The development of the Household Responsibility System gave farmers the incentive to economise on farm labour and to find alternative sources of off-farm work. These farmers were short of capital, distributed all over the country, and (usually) lacking technical skill; their principal asset was abundant, cheap labour. Furthermore,

township and village governments themselves in the 1980s often established these enterprises and one of the reasons for their success—then, as now—was their close ties to those governments. Such enterprises as these farmers set up were bound to have the characteristics that we have identified in this section as factors leading to a gap in the implementation of pollution regulation. The growth, emergence and success of TVEs are inextricably bound up with their water polluting characteristics.

5. Institutional framework

The factors identified in the previous section only partially explain the problem of environmental enforcement. The national environmental protection system is also a significant factor. When the reform program was introduced in 1978, the central Chinese government also introduced a legal framework for environmental protection and, as mentioned above, tough punishments are stipulated in the laws for violating environmental regulations. Since 1978, many water pollution laws have been issued including the environmental protection law (1979), a water pollution prevention and control law (1984), a water resources law (1988), a water and soil conservation law (1991), an improved environmental protection law (1989), and a revised national water law that came into force in late 2002 (Jahiel, 1997; Ross, 1992; Yuan and Chen, 2005). The central government has also released a series of management guidelines, regulations, and standards for environmental protection (Environmental Protection Commission, 1986, 1987, 1990, 1991).

To facilitate the enforcement of these laws and regulations, the central government set up eight policy implementation mechanisms: environmental impact assessment, the three synchronisations (pollution controls are to be incorporated into the design, construction, and operation phases of new projects), pollutant discharge fees, the discharge permit system, the environmental responsibility system, annual assessment of urban environmental quality, centralised pollution control, and limited time treatment (Jahiel, 1997; Sinkule and Ortolano, 1995). China now has a comprehensive legal framework and a nationwide organisational structure for implementing environmental protection measures in both urban and rural areas (Jahiel, 1997).

However, the existing institutional structure is problematic. At the top level, the chief organ of water administration is the Ministry of Water Resources but the State (sometimes called National) Environmental Protection Agency (SEPA) is the main body responsible for water pollution control. SEPA and the Ministry of Water Resources are complemented in their roles by at least five other ministries which are also responsible for water use and pollution protection: agriculture, land and resources; urban and rural construction; forestry; transportation; and the State Development Planning Commission (Gu and Sheate, 2005). For example, TVEs are

regulated by the Ministry of Agriculture; urban water supply systems are the responsibility of the Ministry of Urban and Rural Construction; water is supplied by water supply corporations; while sewage and wastewater treatment is managed by the EPA.

The structure of environmental organisations in China is a grid of vertical and horizontal linkages. In the vertical hierarchy, each level of government below SEPA has an environmental bureau; and there are provincial, prefecture, municipal, county and township EPB. As subsidiaries of SEPA, these bureaus have numerous responsibilities including environmental impact assessment, monitoring, discharge fee collection, and environmental education (Sinkule and Ortolano, 1995). Other departments with sectoral environmental protection responsibilities at the same level of government manage pollution or resource issues in their respective sectors and collaborate with the EPB in environmental supervision and management. The relationship between EPBs and other government authorities is structured in vertical and horizontal dimensions (Jahiel, 1998; Sinkule and Ortolano, 1995; Mao and Zhang, 2003). An EPB belongs to two distinct government units. Vertically, a local bureau is part of the environmental protection functional line from the national level (SEPA) through provincial, municipal, and district/county EPBs and as such receives policy mandates and program direction from the upper-level EPB. Horizontally, it is also one of the departments in a local government and relies heavily on that government for financial support. The head of a local government has the authority to appoint and remove the director of the EPB within his/her jurisdiction (Gu and Sheate, 2005).

Not surprisingly, the vertical and horizontal dual institutional structure for environmental protection sometimes functions poorly. EPBs at various levels of government are still in a relatively weak position in the government political hierarchy: local EPBs or other environmental units have relatively low bureaucratic status (Jahiel, 1994). The vertical functional line does not work when a local EPB is pressured by the local government. In the horizontal bureaucratic hierarchy at the same level of government, the EPB is often challenged by other authorities with much longer histories and more powerful influence over regulatory enforcement and decision making. Furthermore, other local government departments at the same level of government (a township or village, say) cannot be required by an EPB at the same level to act to protect the environment: only a higher level EPB (in the county government, say) can require such action. This is important because the administration of TVEs falls under the jurisdiction of local agriculture departments.

Guanxi also heavily influences environmental policy implementation. Good *guanxi* between government departments is important and EPBs need to keep good *guanxi* with other government departments. In addition to levying fees on enterprises that exceed pollution discharge standards, environmental protection bodies can issue orders to

shut down enterprises that repeatedly fail to meet national standards. Closing a plant, however, requires the support of other departments, such as the planning, construction and other powerful industrial bureaus committed to economic development. Not surprisingly, these other bureaus often fail to support an EPB's decision. In addition, an EPB's cooperative *guanxi* will ensure effective monitoring of the polluting enterprises. For example, the industrial and commercial department approves industrial operation licenses and without good *guanxi* with this department, EPBs remain uninformed about such fundamental matters as where enterprises are located and what they are producing. Although, by law, all new enterprises that produce pollution should be approved by the EPB, they often are not. The reason is poor coordination and perhaps rivalry between departments.

Since the administrative structure places the local EPBs under the dual supervision of both local government and the upper level EPBs, local EPBs find it difficult to carry out their mandates. In particular, the township environment coordinators have to monitor pollution and other environmental problems at the direction of the upper-level EPBs; they have to satisfy local government officials; and they have to deal with water polluters directly. As indicated, local government officials often protect TVEs and other rural enterprises from the prescribed consequences of their pollution. The central government tries to remedy such problems by using the mass media, environmental organisations, environmental education programs, and environmental students' movements to raise public awareness of environmental issues (Hamburger, 1998).

These factors are the product of institutional design failures. Many of the current environmental orders are the result of deals between local environmental protection agencies, SEPA, other ministries, local governmental bodies and the polluting enterprises themselves. They are therefore often ineffective, and their creation is inefficient. The degree of actual compliance and enforcement depends on the region concerned and the personalities of the different players involved. However, as we will demonstrate in the following section, these weaknesses are themselves directly related to the character of the transition in China from a command to a more market-oriented economy.

6. Administrative transition

We have already claimed that the factors that underpin water pollution among TVEs are also the factors that have helped TVEs become so important: in that respect, water pollution is integral to the model of TVE growth. But Section 5 demonstrated some institutional weaknesses that prevent the complete regulation of rural industry by the state's environmental protection system. To a large degree, these weaknesses reflect the character of the transition in China.

In a political sense, transition in China has involved a decentralisation of power from the central to the local levels of government, empowering the local levels of government and permitting distinct, regional models of development to emerge (Webber et al., 2002). Local elites now have substantial flexibility to pursue economic goals for themselves and their localities. In many accounts, this local flexibility is one of the key factors that have underpinned the rapid growth of China's economy over the past quarter of a century (Whiting, 2001; Li, 2005; Horowitz and Marsh, 2002).

However, it is this same local flexibility and power that underpins the matrix government structure and which makes economic development exclude environmental considerations. The financial and other associations with local enterprises make local governments unable and unwilling to implement water pollution regulations. The case of the Huai River basin indicates that only when the SEPA director himself is personally in charge of pollution control in targeted areas do local officials have to cooperate with the national regulations. However, such an approach cannot fix the pollution problems in the far more numerous rural areas.

Local flexibility and power contributes to the water pollution problems of rural areas in another way. The clean up of cities in China has put urban enterprises under increasing pressure to modify the environmental impacts of their operations. Polluters have two options when they are targeted in this way. One is to move their operation to rural areas where the environmental regulations are less strictly enforced and environmental standards are lower. This movement has led to a large scale relocation of pollution. For example, according to Xu (1999), over 700 industrial enterprises in central Shanghai were classified as 'serious polluters' and relocated to outlying suburbs as part of a campaign to reduce pollution within the city (Wu and Wang, 2007; China Environmental Reporter, 1997). Some enterprises moved to rural areas in the neighbouring provinces of Jiangsu, Anhui and Zhejiang. The second option is to upgrade their production equipment to meet the environmental standard and to sell the disused equipment, often sold to rural enterprises. Xu's (1999) survey in Qinsan in Zhejiang province indicates that village enterprises are particularly frequent users of second-hand equipment from urban areas—that had been banned from cities by strict pollution discharge regulations. This, too, has led to an increase in the amount of pollutant discharge in villages.

7. Conclusion

Rural industrial growth in China has occurred almost outside central environmental management systems. Despite a variety of new laws, regulations and guidelines, implementation gaps still exist. The current water pollution control system relies on a top-down approach to monitoring, control and supervision. While this may work in cities

where industries are spatially concentrated and pollution-monitoring systems are well developed, it does not work in the rural areas where water polluting industries are dispersed in villages and have close associations with local government officials. Water pollution control regulations can be applied to rural industries only to the degree that local government officials are able and willing to implement environmental standards and exercise authority.

The problems of controlling rural water pollution derive from the same characteristics of Chinese development that have proved so successful in generating rapid economic growth. The limited capital investment, small labour forces and dispersal over the countryside that made the TVE model so adaptable to the conditions of the Chinese countryside are also the conditions that make rural enterprises so polluting and hard to monitor. The close ties to local government that have underpinned the competitive success of rural enterprises against the more sophisticated urban enterprises are also the ties that make it so difficult to enforce environmental regulations in rural areas. The power and flexibility of local governments to set their own development agendas has encouraged governments to adopt a variety of models of development suitable to their regional conditions, but has also worked to reduce the power of the central government to set uniform standards of practical regulation.

In other words, it is not possible to separate the problem of controlling rural water pollution from the model of development that the Chinese state has been following. Rural water pollution is an integral, if only implicit, effect of that model. It follows that, to control water pollution from rural industry more effectively will be to modify the model of (rural) development. It is not possible to expect to control rural water pollution simply by encouraging more effective local regulation; it is the development model that has to change. There are two obvious contending possibilities of change. One is to reverse the empowerment of local governments. This would give the central government more direct control, but would mean that local individual paths of development would probably be discouraged, and it would impose a huge monitoring task on SEPA. The other is to alter the relative salience of economic growth and environmental quality within local societies through raising environmental awareness, providing media with more freedom to report on environmental conditions, and giving citizens more voice in local government affairs. This requires the government to allow these kinds of inherently social changes, even though they may in turn modify the nature of governance in rural areas. Indeed, there have been some changes that are facilitating some such democratisation of local environmental governance, which is appropriate given that rural water pollution is primarily a product of—and impacts on—local rural societies.

References

- Abigail, R.J., 1997. The contradictory impact of reform on environmental protection in China. *China Quarterly* 149, 81–103.

- Anid, P.J., Tschirley, J., 1998. Environmental Monitoring in China's Hubei Province". UN FAO Paper posted July 1998, available at <http://www.fao.org/sd/EPdirect/EPan0011.htm>.
- Banister, J., 1998. Population, public health and the environment in China. *The China Quarterly* 156 (Special Issue), 986–1015.
- Brown, L.R., 1995. *Who Will Feed China? Wake-up Call for a Small Planet*. W.W. Norton & Co., New York.
- Brown, L.R., 2001. How water scarcity will shape the new century. *Water Science and Technology* 43, 17–22.
- Bruton, G.D., Lan, H.L., Lu, Y., 2000. China's township and village enterprises: Kelon's competitive edge/executive commentary. *The Executive—ADA Then Briefcliff* 14 (1), 19–30.
- Byrd, W.A., 1990. Rural industrialization and ownership in China. *Comparative Economic Studies* 32 (1), 73–107.
- Byrd, W.A., Lin, Q. (Eds.), 1990. *China's Rural Industry: Structure, Development, and Reform*. A World Bank Publication.
- Chan, H.S., Wong, K.K., Chueng, K.C., Lo, J.M., 1995. The implementation gap in environmental management in China: the case of Guangzhou, Zhengzhou, and Nanjing. *Public Administration Review* 55 (4), 333–340.
- Chang, C., Wang, Y.J., 1994. The nature of the township–village enterprise. *Journal of Comparative Economics* 19, 434–452.
- Che, J.H., Qian, Y.Y., 1998. Insecure property rights and government ownership of firms. *Quarterly Journal of Economics* 2, 467–496.
- Chen, J.B., Jefferson, G.H., 1999. Development of the Hengdian township enterprise group: A case study. In: Jefferson, G.H., Singh, Inderjit (Eds.), *Enterprise Reform in China: Ownership, Transition, and Performance*. Oxford, pp. 279–287.
- China Environmental Reporter, 1997. Food contamination is threatening. 2 August 1997.
- Ding, D.Z., Ge, G., Warner, M., 2004. Evolution of organizational governance and human resource management in China's township and village enterprises. *International Journal of Human Resource Management* 15 (4–5), 836–852.
- Edmonds, R.L., 1994. *Patterns of China's Lost Harmony: A Survey of the Country's Environmental Degradation and Protection*. Routledge, London.
- Environmental Protection Commission (PRC), 1986. Management guidelines on environmental protection for construction projects of the People's Republic of China, Beijing.
- Environmental Protection Commission (PRC), 1987. Regulations for engineering design on environmental protection of construction projects, unofficial translation. National Environmental Protection Agency of China, Beijing.
- Environmental Protection Commission (PRC), 1990. Management procedures for environmental protection of construction projects. National Environmental Protection Agency of China, Beijing.
- Environmental Protection Commission (PRC), 1991. Full-scale EIA Including Measurement, Modelling and Prediction Techniques. Department of Development and Supervision, National Environmental Protection Agency of China, Beijing.
- Fan, G., 1997. Zhongguo nongcun qiye zuzhi de bianhua ji qi lilun qishi—you guan "Hengdian moshi" de yixie chubu sikao [Changes in the Organization of Chinese TVEs and Their Theoretical Implications—First Thoughts on the "Hengdian Model"]. *Jingji Yanjiu (Economic Research)* 5, 43–45.
- Florig, H.K., Spofford Jr., W.O., Ma, X., Ma, Z., 1995. China strives to make the polluter pay. *Environmental Science and Technology* 29, 268–273.
- Fothergill, S., 1985. Rural industrialization trends and causes. In: Healey, M.J., Ilberg, B.W. (Eds.), *The Industrialization of the Countryside*. Geo Books, Norwich, UK, pp. 147–160.
- Fu, X.L., Balasubramanyam, V.N., 2004. Exports, FDI, growth of small rural enterprises and employment in China. Working Paper No. 286, ESRC Centre for Business Research, University of Cambridge.
- Fulton, M., 1974. Industry's view of rural area. In: Whiting, L.R. (Ed.), *Rural Industrialization: Problems and Potentials*. The Iowa State University Press, Ames, USA, pp. 68–78.
- Gu, L.X., Sheate, W.R., 2005. Institutional challenges for EIA implementation in China: a case study of development versus environmental protection. *Environmental Management* 36 (1), 125–142.
- Hamburger, J., 1998. Western influence on environmental management in China: dreams of markets and democracy. *Sinosphere* 1 (1), 10–13.
- Ho, S., 1994. *Rural China in transition: non-agricultural development in rural Jiangsu, 1978–1990*. Oxford.
- Horowitz, S., Marsh, C., 2002. Explaining regional economic policies in China: interest groups, institutions, and identities. *Communist and Post-Communist Studies* 35, 115–132.
- Jahiel, A., 1994. Policy implementation under "socialist reform": The case of water pollution management in the People's Republic of China. Ph.D. Dissertation, University of Michigan.
- Jahiel, A., 1997. The contradictory impact of reform on environmental protection in China. *The China Quarterly* 149, 81–103.
- Jahiel, A., 1998. The organisation of environmental protection in China. *The China Quarterly* 156, 757–787.
- Krupnick, A., 1992. Incentive Policies for Industrial Pollution Control in China. A conference paper presented at the annual meeting of the American Economic Association, New Orleans, LA, 2–5 January 1992.
- Li, P.P., 2005. The puzzle of China's township–village enterprises: the paradox of local corporatism in a dual-track economic transition. *Management and Organization Review* 1, 197–224.
- Lieberthal, K., 1995. *Governing China: From Revolution Through Reform*. W.W. Norton, New York.
- Lin, G., 1997. Transformation of a rural economy in the Zhujiang delta. *The China Quarterly* 149, 56–80.
- Liu, Y.L., 1992. Reform from below: the private economy and local politics in the rural industrialization of Wenzhou. *The China Quarterly* 130, 293–316.
- Luo, Y., Tan, J.J., Shenkar, O., 1998. Strategic responses to competitive pressure: the case of township and village enterprises in China. *Asia Pacific Journal of Management* 15, 33–50.
- Luo, Y., Tan, J.J., Shenkar, O., 1999. Township and village enterprises in China: Strategy and environment, chapter 1. In: Kelley, M., Luo, Y. (Eds.), *China 2000. Emerging Business Issues*. Sage, London.
- Mao, W., Zhang, S., 2003. Impacts of the economic transition on environmental regulation in China. *Journal of Environmental Assessment Policy and Management* 5 (2), 183–204.
- Maurer, C., Wu, C., Wang, Y., Xue, Y., 1998. Water pollution and human health in China. *China Environment Series (Woodrow Wilson Centre)* 2, 28–38.
- MoA (Ministry of Agriculture, PRC), 1995. *Concise statistical Materials of China's Agriculture*. China Agriculture Press, Beijing.
- Naughton, B., 1994. Chinese institutional innovation and privatization from below. *American Economic Review* 84 (2), 266–270.
- National Bureau of Statistics of China (NBSC), 1995. *Rural Statistical Yearbook of China*. China Statistics Press, Beijing.
- National Bureau of Statistics of China (NBSC), (various years). *China Statistical Yearbook*. China Statistics Press, Beijing.
- Nee, V., 1992. Organizational dynamics of market transition: hybrid forms, property rights, and mixed economy in China. *Administrative Science Quarterly* 37, 1–27.
- Niu, W.Y., Harris, W.M., 1996. China: the forecast of its environmental situation in the 21st Century. *Journal of Environmental Management* 47 (2), 101–114.
- Oi, J., 1995. The role of the local state in China's transitional economy. *The China Quarterly* 144, 1132–1149.
- Peng, Y.S., Zucker, L.G., Darby, M.R., 1997. Chinese rural industrial productivity and urban spillovers. NBER Working Paper no. 6202, National Bureau of Economic Research, Cambridge, MA.
- Qi, F., Cheng, G.D., Masao, M., 1999. Water resources in China: problems and countermeasures. *Ambio* 28 (2), 202–203.
- Qu, G., Li, J., 1994. *Population and the Environment in China*. Lynne Rienner Publications, Boulder, CO.
- Ross, L., 1992. The politics of environmental policy in the People's Republic of China. *Policy Studies Journal* 20, 628–642.

- Schmidt, C.W., 2002. Economy and environment: China seeks a balance. *Environmental Health Perspective* 110 (9), 517–522.
- State Environmental Protection Administration (SEPA), 2002. Report on the State of the Environment in China 2002. Environmental Information Centre (SEPA), Beijing.
- Shen, G.Q., Lu, Y.T., Wang, M.N., Sun, Y.Q., 2005. Status and fuzzy comprehensive assessment of combined heavy metal and organochlorine pesticide pollution in the Taihu Lake region of China. *Journal of Environmental Management* 76 (4), 355–362.
- Sinkule, B., 1993. Implementation of industrial water pollution control policies in the Pearl River Delta Region of China. Ph.D. Dissertation, Department of Civil Engineering, Stanford University.
- Sinkule, B., Ortolano, L., 1995. Implementing Environmental Policy in China. Praeger, Westport, CT.
- Smil, V., Mao, Y.S., 1998. The Economic Costs of China's Environmental Degradation. American Academy of Arts and Sciences, Cambridge.
- Song, L., Du, H., 1990. The role of township governments in rural industrial development. In: Byrd, W.A., Lin, Q. (Eds.), *China's Rural Industry: Structure, Development, and Reform*. Oxford University Press and the World Bank, pp. 342–357.
- Tan, K.C., 1993. Changing rural–urban relations. In: Taubmann, W. (Ed.), *Urban Problems and Urban Development in China*. Mitteilungen des Instituts für Asienkunde Hamburg, Hamburg, pp. 128–142.
- United Nations Development Program (UNDP), 2002. China human development report 2002: Making green development a choice. Produced by the Stockholm Environment Institute in collaboration with the UNDP. Oxford University Press, New York.
- Vermeer, E.B., 1999. Shareholding cooperatives: a property rights analysis. In: Oi, J.C. and Walder, A.G. (Hg.) (Eds.), *Property Rights and Economic Reform in China*, Stanford, pp. 123–144.
- Wang, H., 2000. Pollution charges, community pressure, and abatement cost of industrial pollution in China. Policy Research Working Paper #2337, World Bank, Washington DC.
- Wang, H., Wheeler, D., 2000. Endogenous enforcement and effectiveness of China's pollution levy system. Policy Research Working Paper #2336, World Bank, Washington, DC.
- Wang, H., Bi, J., Wheeler, D., Wang, J.D., Cao, D., Lu, G.F., Wang, Y., 2004. Environmental performance rating and disclosure: China's GreenWatch program. *Journal of Environmental Management* 71 (2), 123–133.
- Wang, J., Lu, X., 1997. Economic policies for environmental protection in China: practice and perspectives. In: *Applying Market-based Instruments to Environmental Policies in China and OECD Countries*. Organization for Economic Co-operation and Development (Eds.), OECD, Paris, pp. 15–29.
- Wang, M.Y., 1997. Rural labour transition in China. *Asian Studies Review* 21 (2/3), 198–211.
- Webber, M., Wang, M.Y., Zhu, Y. (Eds.), 2002. *China's Transition to a Global Economy*. Palgrave MacMillan, Hampshire.
- Weitzman, M.L., Xu, C.G., 1994. Chinese township–village enterprises as vaguely defined cooperatives. *Journal of Comparative Economics* 18, 121–145.
- Wheeler, D., Wang, H., Dasgupta, S., 2000. Can China Grow and Safeguard Its Environment? The Case of Industrial Pollution. Working Paper No. 68, Centre for Research on Economic Development and Policy Reform, Stanford University.
- Whiting, S.H., 2001. *Power and Wealth in Rural China*. Cambridge University Press.
- Wong, K.K., 1999. Challenge of sustainable development. In: Teather, D.C.B., Yee, H.S. (Eds.), *China in Transition: Issues and Politics*. Macmillan Press, London, pp. 171–193.
- Wong, K.K., Hon, S.C., 1994. The development of environmental management system in the People's Republic of China. In: Bosseau, M., Lo, C.K. (Eds.), *China Review*. Chinese University Press, Hong Kong, pp. 1–21.
- World Bank, 1992. *China: Long-term Issues and Options in the Health Transition*. The World Bank, Washington, DC.
- World Bank, 1997. *Clear Water, Blue Skies: China's Environment in the New Century*. World Bank, Washington, DC.
- Wu, J.P., Wang, M.Y., 2007. Manipulating the margins: the case of Shanghai. In: McGee, T.G., Lin, C.G., Wang, M.Y. (Eds.), *Constructing Urban China: Repositioning China's Urban Space*. Routledge (forthcoming).
- Xi, W., Xu, Z.X., 2002. Legal control of water pollution in Huai river, China: a case study. In: Conference Paper for Sixth International Conference on Environmental Compliance and enforcement, San Jose, Costa Rica, April 15–19.
- Xie, J., Florig, H.K., 1997. Incentive to reduce pollution emissions: a quantitative study of pollution levies and subsidies in China. In: Xu, D., et al. (Eds.), *China Tax System Reform*. Chinese Economics Press, Beijing, pp. 1–31.
- Xu, F., 1992. Analysis of the economic costs of TVI pollution impacts on human health. *Wei Sheng Yan Jiu. Journal of Hygiene Research* 21 (Suppl.), 1–23.
- Xu, H.G., Liu, D.Z., Wu, H.Z., 2001. Environmental regionalization for the management of township and village enterprises in China. *Journal of Environmental Management* 63, 203–210.
- Xu, H.Q., 1999. Environmental policy and rural industrial development in China. *Research in Human Ecology* 6 (2), 72–80.
- Xu, W., Tan, K.C., 2002. Impact of reform and economic restructuring on rural systems in China: a case study of Yuhang, Zhejiang. *Journal of Rural Studies* 18 (1), 65–81.
- Yuan, D.L., Chen, L.J., 2005. Yellow River water right transfer system and practice. In: YRCC (Ed.), *Proceedings of the Second International Yellow River Forum*, Vol. V. Yellow River Conservancy Press, Zhengzhou, pp. 323–331.
- Zhang, H.J., Ferris Jr., R., 1998. Shaping an environmental protection regime for the new century: China's environmental legal framework. *Sinosphere* 1 (1), 6–9.
- Zhang, T. Z., Wang, D.Q., Wu, G.Y., Wang, J.N., 1997. Water pollution charges in China: Practice and prospects. In: *Organization for Economic Co-operation and Development* (Ed.), *Applying Market based Instruments to Environmental Policies in China and OECD Countries*. OECD, Paris, pp. 239–257.
- Zhao, S.X.B., Wong, K.K., 2002. The sustainability dilemma of China's township and village enterprises: an analysis from spatial and functional perspectives. *Journal of Rural Studies* 18 (3), 257–273.
- Zuo, Z., 2001. *Zhujiang Sanjiaozhou Moshi de Zongti Tezhen yu Chengyin* (Characteristics and Origin of "Pearl River Delta Model"). *Economic Theory and Management*, 71–75 (in Chinese).

Determination of granule size distribution in a UASB reactor

Apostolos Vlyssides, Elli Maria Barampouti, Sofia Mai*

School of Chemical Engineering Athens, National Technical University of Athens, 9 Heroon Polytechniou Street, Zographou 15700, Greece

Received 4 May 2006; received in revised form 6 December 2006; accepted 12 December 2006

Available online 5 February 2007

Abstract

Granular sludge is the key factor for an efficient operation of an upflow anaerobic sludge blanket (UASB) reactor. In order to monitor the granularity of anaerobic sludge, the determination of the granule size distribution is of vital importance. For this reason, several techniques have been proposed; however, they are either tedious, imprecise or expensive and hardly applicable in full-scale treatment plants. There was then the need for a simple and low-cost technique. This technique involves the determination of the settling velocities of a sludge sample and of extrapolating the corresponding diameters using a mathematical algorithm. In the proposed algorithm, the granules density was calculated, the flow regime was examined and finally the granule size distribution was obtained. Two very important correlations were suggested by the experimental results. The granule density and diameter were strongly correlated with the VSS/TSS ratio.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Granule density; Granule diameter; Granule size distribution; UASB; VSS/TSS ratio

1. Introduction

The last 20 years, many full-scale upflow anaerobic sludge blanket (UASB) reactors have been constructed to treat industrial wastewater. The limitations of UASB reactors are related to the wash out of biomass. So, the granulation process is a pre-requisite of a UASB system start-up and operation. The treatment capacity of a UASB system depends on whether granulation is accomplished with particular operating conditions, e.g. wastewater composition, organic loading rate, temperature, etc. (Jeison and Chamy, 1998; Xu and Tay, 2001). The quality and stability of sludge directly dictates the behaviour of the entire treatment system. One of the most important characteristics is their size distribution. This size with their density defines the settling properties of the sludge, an attribute that is fundamental from the operational point of view (Jeison and Chamy, 1998).

To date, three techniques are routinely used to determine the particle size distribution of UASB sludge. Two of them consist in the direct size measurement with a microscope of at least 100 sludge granules immobilized in a Petri dish. This can be done either manually with a portion graticule or automatically using image analysis and computerized data processing (Hulshoff Pol, 1989; Dudley et al., 1993). The last method involves particle size analysis by laser and has been introduced in the late 1990s (Yan and Tay, 1997). Unfortunately, all these procedures present several disadvantages. The graticule one is very tedious, while the image and laser analysis are expensive (at least \$US 10,000 for image analysis and between 30,000 and 60,000 for the laser one). Finally, none of the previous techniques, because of their complexity or cost can be easily set up in treatment plants, and as a consequence are almost confined to laboratory studies (Laguna et al., 1999).

It is therefore of great importance to have a technique that allows the evaluation, in a reliable and reproducible way, of the size distribution of anaerobic granular sludge. This paper presents a way to evaluate the sludge size distribution using settling velocities of a sludge sample and extrapolating the corresponding diameters.

*Corresponding author. Tel.: +30 210 7723268; fax: +30 210 7723269.

E-mail addresses: avlys@tee.gr (A. Vlyssides), belli@central.ntua.gr (E.M. Barampouti), mai@central.ntua.gr (S. Mai).

2. Materials and methods

2.1. Source of granular sludge

In the present investigations, the physical properties of anaerobic granular sludges cultivated on different types of wastewater and coming from four different UASB reactors were determined and compared. The granular sludge samples tested were:

- granular sludge cultivated on wastewater from a potato processing industry;
- granular sludge cultivated on dairy wastewater with iron addition;
- granular sludge cultivated on dairy wastewater with calcium addition;
- granular sludge cultivated on dairy wastewater.

The choice of the sludges ensured a wide range of sludge composition and physical characteristics.

2.2. Granule density measurement

The granule density was determined with a picnometer according to the method of Mahling (1965). This density measurement was performed at 30 °C using 25 ml flasks with glass stoppers with capillars. The granule density (ρ_p) was calculated according to (Hulshoff Pol, 1989):

$$\rho_p = (\rho_w - \rho_a) \times \frac{m_3 - m_1}{(m_2 - m_1) - (m_4 - m_3)} + \rho_a \quad (1)$$

in which ρ_a is the density of the air (kg m^{-3}), which is 1.165 at 30 °C, and a P_{atm} of 760 mmHg; m_1 the weight of the empty picnometer (kg); m_2 the weight of the picnometer filled with water (kg); m_3 the weight of the picnometer filled with sludge (kg); m_4 the weight of the picnometer filled with water and sludge (kg); ρ_w the density of water (kg m^{-3}), which is = 995.85 at 30 °C.

2.3. Upflow velocity test

The settling properties of granules were evaluated by the fractions of granules exited under certain upflow velocities in a fractionating device (Fig. 1) (Andras et al., 1989). About 8 ml of granules from the reactor were separated into ten fractions under upflow liquid velocities of 1.8, 2.2, 3.4, 5.1, 10.7, 16.8, 21.1, 30.9, 46.6, 99.6 m h^{-1} , respectively, and at each velocity for 5 min. Each fraction from the fractionating device was collected in a Whatman (1 PS) filter paper. TSS and VSS in each fraction were determined using the method described by Andras et al. (1989).

The analysis for TSS and VSS were conducted in accordance with the standard methods.

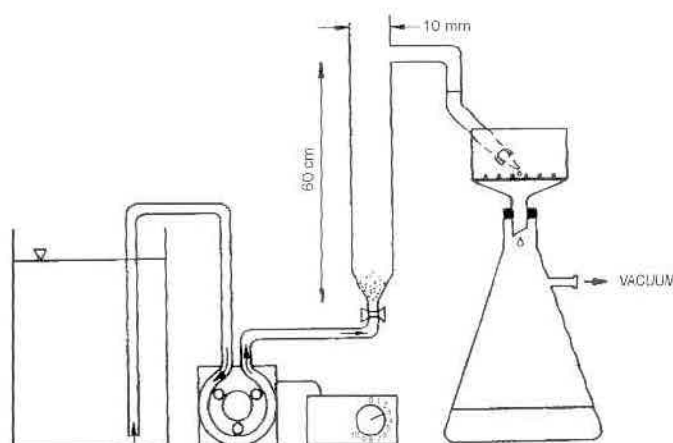


Fig. 1. Upflow velocity test apparatus.

3. Results

3.1. Granule density

A total of 400 sludge samples from the four UASB reactors operating under different conditions (e.g. loading rate, metal addition, temperature) were collected for a period of 2 months. For these samples, the total and volatile suspended solids concentrations as well as their density were determined. Each measurement was repeated three times and the results presented are the mean values. The standard deviations of total and volatile suspended solids concentrations were 5% and 6%, respectively. Regarding the standard deviation of density, it was 10%.

A significant correlation ($R^2 = 0.968$) between the granule density and the ratio VSS/TSS of the sludge arose (Eq. (2)):

$$\rho_p = 1387 - 3.77 \frac{\text{VSS}}{\text{TSS}}, \quad (2)$$

in which VSS/TSS is the % of volatile suspended solids.

Fig. 2 presents the densities measured experimentally for the 400 different sludge samples and the respective densities predicted by Eq. (2).

3.2. Settling properties of granular sludge

For 100 sludge samples, the upflow velocity test was performed, and VSS, TSS concentrations of every fraction were determined. A settling velocity profile of each particular sample can be depicted by plotting upflow velocity versus per cent of TSS lost from the glass test device (Fig. 3a). A plot of cumulative per cent of TSS lost from the glass test device versus upflow velocity is more revealing and an easier way to interpret the results (Fig. 3b). Figures such as Figs. 3(a and b) were plotted for each of the 100 samples examined.

A mathematical algorithm was developed in order to estimate the size distribution of a sludge sample from the upflow velocity and the VSS/TSS ratio of each fraction, assuming spherical granules. Given the upflow velocities, the diameter of each fraction, may be estimated, after a

looping procedure, from the following Eqs. (3)–(5) (Metcalf and Eddy, 1991) taking into account the flow regime:

For laminar flow ($Re < 0.5$)
$$d = \left[\frac{18uv}{g \left(\frac{\rho_p}{\rho_w} - 1 \right)} \right]^{0.5}, \tag{3}$$

For transient flow ($0.5 < Re < 10^5$)
$$d = \frac{3u^2 C_D}{4g \left(\frac{\rho_p}{\rho_w} - 1 \right)}, \tag{4}$$

For turbulent flow ($Re > 10^5$)
$$d = \frac{0.3u^2}{g \left(\frac{\rho_p}{\rho_w} - 1 \right)}, \tag{5}$$

where d is the granule diameter (m); u the upflow velocity (m s^{-1}); ν the kinematic viscosity ($\text{m}^2 \text{s}^{-1}$), which is 8.01×10^{-7} at 30°C ; g the gravitational acceleration (9.81 m s^{-2}); and Re the Reynolds number.

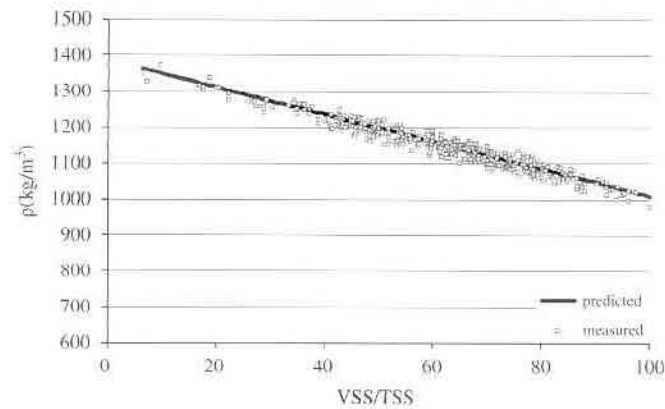


Fig. 2. Correlation of densities with VSS/TSS ratio.

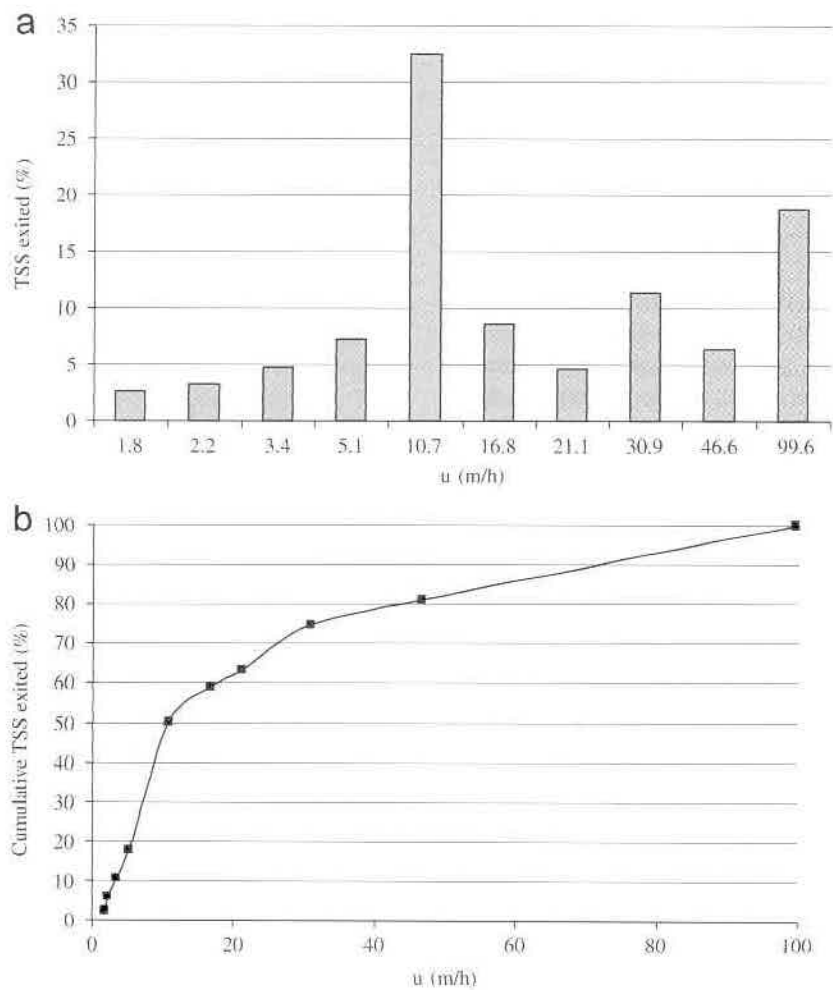


Fig. 3. Upflow velocities profile (a) and cumulative solids loss plot (b) for a typical sludge sample.

C_D = Drag coefficient, which can be calculated by

$$C_D = \frac{24}{Re} + \frac{3}{Re^{0.5}} + 0.34, \quad (6)$$

for transient flow.

For the diameter calculation, the value of granule density (ρ_p) is necessary, so Eq. (2) was indispensable.

For the typical sludge sample presented in Fig. 3, the corresponding size distribution is depicted in Fig. 4. It is evident that by a simple settleability test, the size distribution of a sludge sample may be obtained avoiding tedious, imprecise and expensive tests, such as microscope sizing, image and laser analysis.

After processing the experimental results of the upflow velocity test, the size distribution of all samples came up. This revealed a significant correlation between the granule diameter and the ratio VSS/TSS. Eq. (7) depicts this correlation ($R^2 = 0.975$):

$$d = -8 \times 10^{-4} \times \log \left(1 - \frac{VSS/TSS}{97.27} \right). \quad (7)$$

Fig. 5 presents the granule diameters estimated from experimental data for the different sludge samples and the respective diameters predicted by Eq. (7).

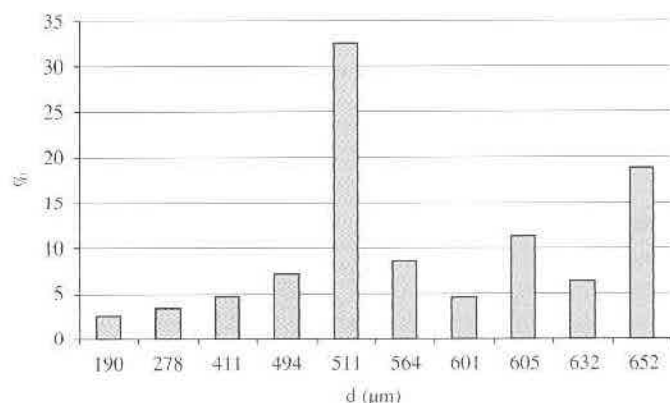


Fig. 4. Particle size distribution for a sludge sample.

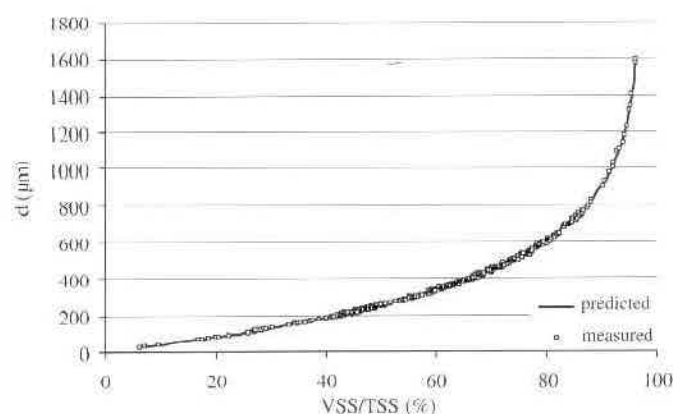


Fig. 5. Correlation of granule diameter with VSS/TSS ratio.

4. Conclusions

Among all the different types of anaerobic digesters applied at full scale, UASB reactors present the best commercial acceptance. The success of these reactors is related to their capacity for biomass accumulation by settling without the need of a carrier. Good settling properties are obtained through the flocculation of the biomass in the form of dense granules with diameters up to several millimetres. The evolution of size distribution is of vital importance for the monitoring of UASB reactors. In this paper, a simple technique for the determination of size distribution was presented. No sophisticated equipment is needed and results can be obtained within a couple of hours. The analyses are completely reproducible, with good precision and exactitude. Small samples of sludge are needed, which is especially important in the case of laboratory scale reactors, where large samples are not available. The proposed technique consists of the determination of the settling velocities of a sludge sample, by the test proposed by Andras et al. and also used by a number of researchers (Hulshoff Pol, 1989; Grotenhuis et al., 1991; Laguna et al., 1999) and of extrapolating the corresponding diameters using a mathematical algorithm. In the proposed algorithm, the granules were assumed spherical, their density was calculated, the flow regime was examined and finally the granule size distribution was obtained.

Two very important correlations were suggested by the experimental results. First, a good correlation between the density and the VSS/TSS ratio of granules was established. The diameter of granules was also correlated with the VSS/TSS ratio. This correlation is a very useful tool for an easy and quick estimation of the granule diameter in a sludge sample. Thus, by the simple and routine, even in industrial scale, analysis of TSS and VSS, the characterization of a sludge bed as far as its granulation is concerned can be easily achieved.

References

- Andras, E., Kennedy, K.J., Richardson, D.A., 1989. Test for characterizing settleability of anaerobic sludge. *Env. Technol. Lett.* 10 (5), 463–470.
- Dudley, B.T., Howgrave-Graham, A.R., Bruton, A.G., Wallis, F.M., 1993. Image analysis to quantify and measure UASB digester granules. *Biotechnol. Bioeng.* 42 (3), 279–283.
- Grotenhuis, J.T.C., Kissel, J.C., Plugge, C.M., Stams, A.J.M., Zehnder, A.J.B., 1991. Role of substrate concentration in particle size distribution of methanogenic granular sludge in UASB reactors. *Water Res.* 25 (1), 21–27.
- Hulshoff Pol, L.W., The Phenomenon of Granulation of Anaerobic Sludge. Ph.D. Thesis, Wageningen Agricultural University, The Netherlands, 1989.
- Jeison, D., Chamy, R., 1998. Novel technique for measuring the size distribution of granules from anaerobic reactors for wastewater treatment. *Biotechnol. Technol.* 12 (9), 659–662.
- Laguna, A., Ouattara, A., González, R.O., Baron, O., Fama, G., El Mamouni, R., Guiot, S., Monroy, O., Macarie, H., 1999. A simple

- and low cost technique for determining the granulometry of upflow anaerobic sludge blanket reactor sludge. *Water Sci. Technol.* 40 (8), 1–8.
- Mähling, A.M., 1965. Die Dichte. In: Schormueller, J. (Ed.), *Handbuch der Lebensmittel*, Band II, Teil 1. Analytik der Lebensmittel. Springer, Berlin, BRD.
- Metcalf, A., Eddy, A., 1991. *Wastewater Engineering—Treatment, Disposal and Reuse*, third ed. McGrawHill International Editions, pp. 1298.
- Xu, H.L., Tay, J.H., 2001. Preserved granular sludge for inoculation of new UASB reactors. *J. Env. Sci. Health A* 36 (9), 1747–1756.
- Yan, Y.G., Tay, J.H., 1997. Characterization of the granulation process during UASB start-up. *Water Res.* 31 (7), 1573–1580.

A preliminary process design and economic assessment of a catalyst rejuvenation process for waste disposal of refinery spent catalysts

Meena Marafi, Antony Stanislaus, Ezra Kam*

Petroleum Refining Department, Petroleum Research and Studies Center, Kuwait Institute for Scientific Research, P.O. Box 24885, Safat 13109, Kuwait

Received 23 April 2006; received in revised form 10 December 2006; accepted 12 December 2006

Available online 20 February 2007

Abstract

Spent hydroprocessing catalysts from refineries have been classified as hazardous solid waste by the United States Environmental Protection Agency (USEPA), refiners must find a viable but economical solution to solve this serious environmental issue. Catalyst rejuvenation is an attractive option for minimizing the environmental problems associated with spent catalysts. In this study, a preliminary design for such a process and the corresponding economic analysis are performed to assess the proposed catalyst rejuvenation process for metal-fouled spent catalysts generated in residue hydroprocessing units. The scenarios used in the economic assessment are based on three options of process synthesis and two operator modes. It is found that the option of rejuvenating medium and lightly fouled spent catalyst produced by the refinery will be the best solution for refiners, both environmentally and economically. © 2007 Elsevier Ltd. All rights reserved.

Keywords: Hazardous waste; Hydroprocessing; Spent catalyst recycling; Rejuvenation; Economic assessment; Waste utilisation

1. Introduction

Spent catalysts form a major source of solid wastes in the petroleum refining industries (Habermehl, 1988; Trimm, 1990; Furimsky, 1996; Chang, 1998). The quantity of spent catalysts discharged from different processing units depends largely on the amount of fresh catalysts used, their life and the deposits formed on them during use in the reactors. In most refineries, a major portion of the spent catalyst wastes come from the hydroprocessing units because of the use of large quantities of catalysts in the hydrotreating process for the purification and upgrading of various petroleum streams and residues. The volume of spent hydroprocessing catalysts discarded as solid waste has increased significantly in recent years due to the following reasons:

- rapid growth in the distillates hydrotreating capacity to meet the increasing demand for ultra-low sulphur transportation fuels;
- a steady increase in the processing of heavier feedstocks containing higher sulphur and metals content, and

- rapid deactivation of and unavailability of a reactivation process for resid hydroprocessing catalysts.

In Kuwait's refineries, over 250 000 barrels of residues are upgraded and converted to high-quality products by catalytic hydroprocessing, bringing substantial economic returns to the country. These operations generate a substantial amount of spent catalysts as solid waste every year. Currently, about 6000 tons of spent catalysts are discarded as solid wastes from Kuwait's refineries annually. This will increase further and exceed 10 000 tons/yr when a fourth refinery is built to process heavy crudes and residues.

Environmental regulations concerning spent catalyst disposal have become more severe in recent years (Rhodes, 1996). Spent hydroprocessing catalysts have been classified as hazardous wastes by the United States Environmental Protection Agency (Rapaport, 2000; USEPA, 2003). The most important hazardous characteristic of spent hydroprocessing catalysts is their toxicity. Due to the hazardous nature of spent catalysts, refiners are being pressured by environmental authorities to ensure the safe handling of spent catalysts. Disposal in landfill is environmentally

*Corresponding author. Tel.: +965 3980499; fax: +965 3980445.
E-mail address: ekam@prsc.kisr.edu.kw (E. Kam).

restricted because chemicals, such as vanadium (V), molybdenum (Mo), nickel (Ni), and cobalt (Co), present in the catalysts can be leached by water after disposal and subsequently pollute the environment (Furimsky, 1996). In the USA, the disposal and treatment of spent refinery catalysts is governed by the Resource Conservation and Recovery Act (RCRA), which holds not only the approved dump-site owner liable, but also the owner of the buried waste. This environmental responsibility continues for the life of the dump-site. The current RCRA regulations require landfills to be built with double liners as well as with leachate collection and groundwater monitoring facilities. Thus, the landfill option is becoming expensive today. In addition, it carries with it a continuing environmental liability. Treatment prior to landfilling may be necessary in some cases, further increasing the cost.

In recent years, increasing emphasis has been placed on the development of processes for recycling of the waste catalyst materials as much as possible (Marafi and Stanislaus, 2003b). Recovery of metals and other components from the spent catalysts is possible, particularly for catalysts that contain high concentrations of valuable metals (Berrebi et al., 1994; Case et al., 1995; Lianos and Deering, 1997; Kar et al., 2004; Marafi and Furimsky, 2005; Chen et al., 2006). However, fluctuations in the market prices of the recovered metals and their purity, together with the high costs of shipping significantly influence the economics of the metal reclamation process making it less attractive for spent catalysts that contain low metals concentrations.

An attractive approach to minimize the environmental problem associated with spent catalyst is to reactivate and reuse them. A survey of literature revealed that the technology for reactivating spent catalysts from residue hydroprocessing operations is not yet well developed (Furimsky and Massoth, 1993; Furimsky, 1996; Trimm, 2001; Marafi and Stanislaus, 2003a). Catalysts are deactivated by pore blockage and fouling of the active surface with coke and metal (V and Ni) deposits originating from the heavy feedstock (Furimsky and Massoth, 1999; Kam et al., 2005; Al-Dalama and Stanislaus, 2006; Juraidan et al., 2006). Regeneration by conventional procedures using nitrogen-air or steam-air under controlled conditions does not result in complete reactivation of the catalysts. While carbon deposits are removed completely, metallic impurities remain on the catalysts. The foulant metals are usually concentrated near the outer surface of the pellet, blocking pore mouths and markedly reducing the active surface area available within the inner pores of the catalysts (Quann et al., 1988; Al-Dalama and Stanislaus, 2006). If the contaminant metals can be removed selectively by chemical treatment without significantly affecting the chemical and physical characteristics of the original catalyst, then the spent catalyst could be rejuvenated and reactivated. Therefore, considerable effort was devoted to this issue in the Laboratories of the Kuwait Institute for Scientific Research (KISR) as part of a research program

on the handling and utilization of spent catalyst (Marafi et al., 1998). Factors influencing the selectivity of major metal foulant (V) removal from the deactivated catalyst were investigated, and improvements in the key catalyst properties (e.g. surface area and pore volume) and hydrodesulfurization (HDS) activity after metal leaching were examined in our studies (Stanislaus et al., 1993, 1996; Marafi et al., 1994; Marafi and Stanislaus, 2003b) and a bench-scale process to rejuvenate the metal-fouled spent residue hydroprocessing catalyst was developed based on the results.

In this paper, a detailed description of the process and an economic assessment of the overall rejuvenation process are presented. Scenarios used in the economic assessment are based on three options of process synthesis according to the amount of spent catalyst to be rejuvenated, the severity of catalyst deactivation (light, medium and heavy fouling), and two developer modes, i.e., building and operation either by a local refinery or by an independent company.

1.1. Process scheme

Fig. 1 shows a schematic diagram of the rejuvenation process, which consists of several operations, such as de-oiling, sieving of catalyst fines, separation of lightly fouled (low-density) spent catalyst from medium and severely fouled portions by jugging, metal leaching, and decoking to produce the rejuvenated catalyst. The reagents and reaction conditions in each step in the rejuvenation study are described in Section 2 on Process Design. The physical and chemical characteristics of the rejuvenated catalyst compared with those of fresh and spent catalysts are shown in Table 1. It is seen that the spent catalyst is fouled with carbon (20 wt%) and vanadium (6.5 wt%) deposition, and its surface area and pore volume are, respectively, 62% and 86% lower than that of the fresh catalyst. Substantial improvements in the chemical and physical characteristics have occurred in the rejuvenated process. The rejuvenated catalyst contains no coke, and its V content is 80% lower than that of the spent catalyst. V distribution profile measurement by electronic microprobe analysis showed that the small amount of V remaining in the rejuvenated catalyst was not concentrated near the outer surface of the catalyst pellets, but evenly in the pores (Stanislaus et al., 1996). The surface area increased from 88 to 240 m²/g. A similar increase in catalyst pore volume is also noticed. HDS activity measurements showed that over 92% of the activity of fresh catalyst was recovered by rejuvenation.

1.2. Process architecture

Similar to the bench-scale processing scheme, the spent catalyst rejuvenation process is basically a combination of several types of unit operations in chemical engineering, such as solvent extraction, separation, and chemical reaction. Based on the amount and severity of deactivation

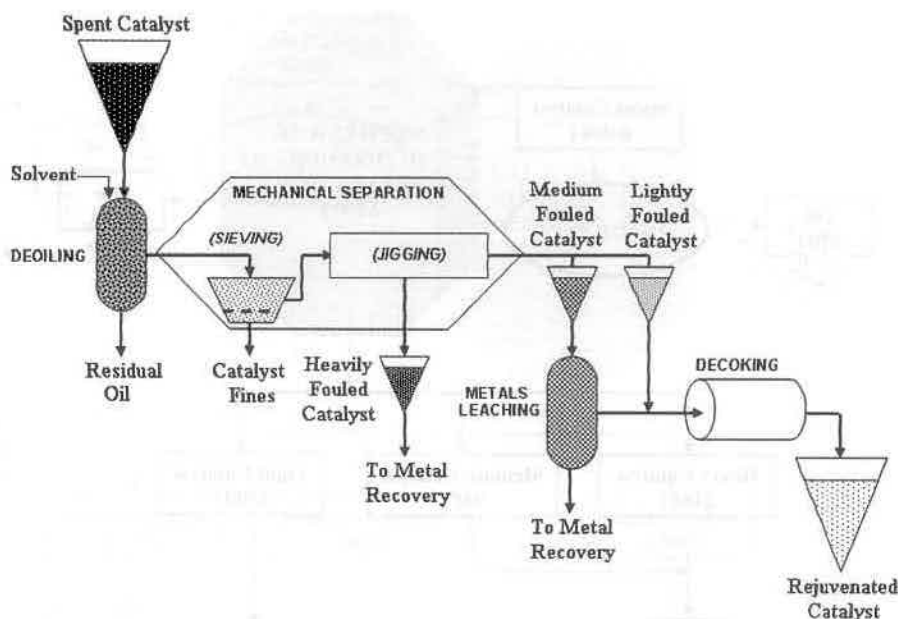


Fig. 1. A bench-scale spent catalyst rejuvenation processing scheme.

Table 1
Characteristics of spent, fresh and rejuvenated catalysts

Catalyst characteristic	Spent	Fresh	Rejuvenated (via option)	
			I and II	III
Chemical composition (wt%)				
MoO ₃	11.20	12.00	10.10	11.80
CoO	3.20	4.20	3.05	4.01
Ni	2.80	—	0.94	0.60
V	6.50	—	1.30	1.60
C	20.00	—	—	—
S	6.30	—	—	—
Fe	0.19	0.10	0.06	0.05
Na	0.50	0.20	0.35	0.27
Physical properties				
Surface area (m ² /g)	98.0	257.00	240.00	221.00
Bulk density (g/ml)	1.03	0.70	0.79	0.78
Side crushing strength (N/mm)	7.30	7.50	7.10	7.15
Pore volume (ml/g)	0.05	0.55	0.48	0.49
Relative HDS activity	0.28	1.00	0.92	0.93

of the spent catalysts to be rejuvenated, three process configuration options can be used.

Option I (Fig. 2): spent catalyst soaked with residual oil is de-oiled first by washing with kerosene. Then the oil-free spent catalyst is mechanically separated to catalyst fines (<0.5 nm particles), and lightly, medium, and heavily fouled catalyst portions. The medium dense and heavy portions are subjected to chemical treatment to remove foulant metals by leaching. Finally, the resulting leached catalyst and lightly fouled catalyst portions are mixed together and decoked by controlled combustion to yield rejuvenated catalysts that are free of coke and most metal

foulants as the finished product. The numbers shown in Fig. 2 are the mass balances for the rejuvenation of 6000 tons of spent catalyst.

Option II (Fig. 3): de-oiling, sieving and jigging are as in Option I. After separating the spent catalyst mechanically into three groups, the heavy fouled catalysts are sold as de-oiled spent catalyst to metal recovery plants. Only the medium densely fouled catalyst portion is subjected to chemical treatment for foulant metals removal. The resulting leached catalyst and lightly fouled catalyst are decoked to produce rejuvenated catalyst.

Option III (Fig. 4): Similar to the Options I and II, the spent catalysts are subjected to de-oiling and grouping. The heavy and medium-dense catalysts groups are sold as de-oiled spent catalyst without any further processing. No metal leaching operation is required. The lightly fouled catalyst portions are decoked under controlled conditions to produce the rejuvenated product.

Overall and individual inputs from the three process configuration options to each subprocess for spent catalyst are summarized in Table 2. The sizes of the de-oiling and mechanical separation processes for the three options are identical, as are those of the metals leaching plant for Options I and II. The size of the decoking unit declines from Option I to Option II and to Option III. The overall mass balances for the raw materials, various catalyst products and other by-products for each option are shown in Table 3.

2. Process designs

To design a plant, it is important to know the process capacity, operating conditions, and special processing or operational precautions needed due to corrosion. However, it is equally important to consider carefully the toxicity

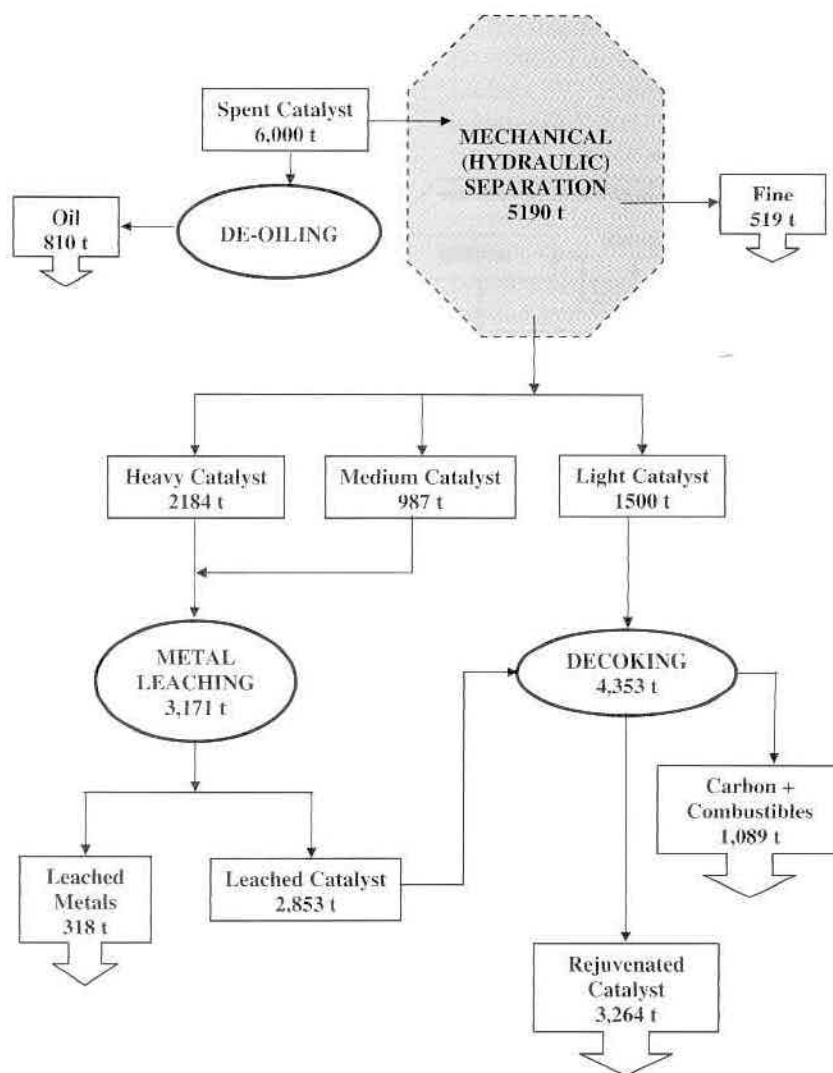


Fig. 2. Rejuvenation process—Option I.

and hazards of processing chemicals, materials needed for construction, and means to comply environmental regulations.

2.1. De-oiling process

Before attempting to leach the poisonous metals from spent catalyst, it is essential to remove all of the residual oil inside or covering the catalyst particles. This enables the leaching agent to reach the compound metal deposits without contamination or restriction. The main environmental consideration with this process is the release of volatile organic compounds (VOCs) as an additional effluent in air emissions from vapourized solvent, which must be kept under tight control.

Since this is the first process in a sequence and is applied to all three options, the plant capacity is expected to be 6000 tons/yr. The solvent quantity and process temperature depend on its solubility and boiling point, respectively, as shown in Table 4. The residence time varies with the API of

residual oils, the type of solvents, and the effective diffusivity of the solvent. Since the removal of residual oil is achieved in two stages—oil on the catalyst surface and oil in catalyst pores—the latter is usually the controlling step under normal operation at atmospheric pressure.

Vapor from any of the solvents—naphtha, kerosene, diesel or gas-oil—is toxic to humans and constitutes a fire hazard. Thus, and a closed system for solvent flow is recommended for better air-emissions control. In terms of corrosion, the solvents have a minor effect on carbon steel, and hence, 304 stainless steel is recommended.

From the bench- and pilot-plant scale experiments, in which kerosene was used as the solvent (Marafi et al., 1998), flow rates, and solvent, residual oil, catalyst and oil-in-solvent ratio in the feed or discharge stage are known; thus the number of stages can be evaluated. Since de-oiled catalyst rather than residual oil is the final product, the design factors depend heavily on the solids mass balance. From the overall solids residence time, the time catalyst

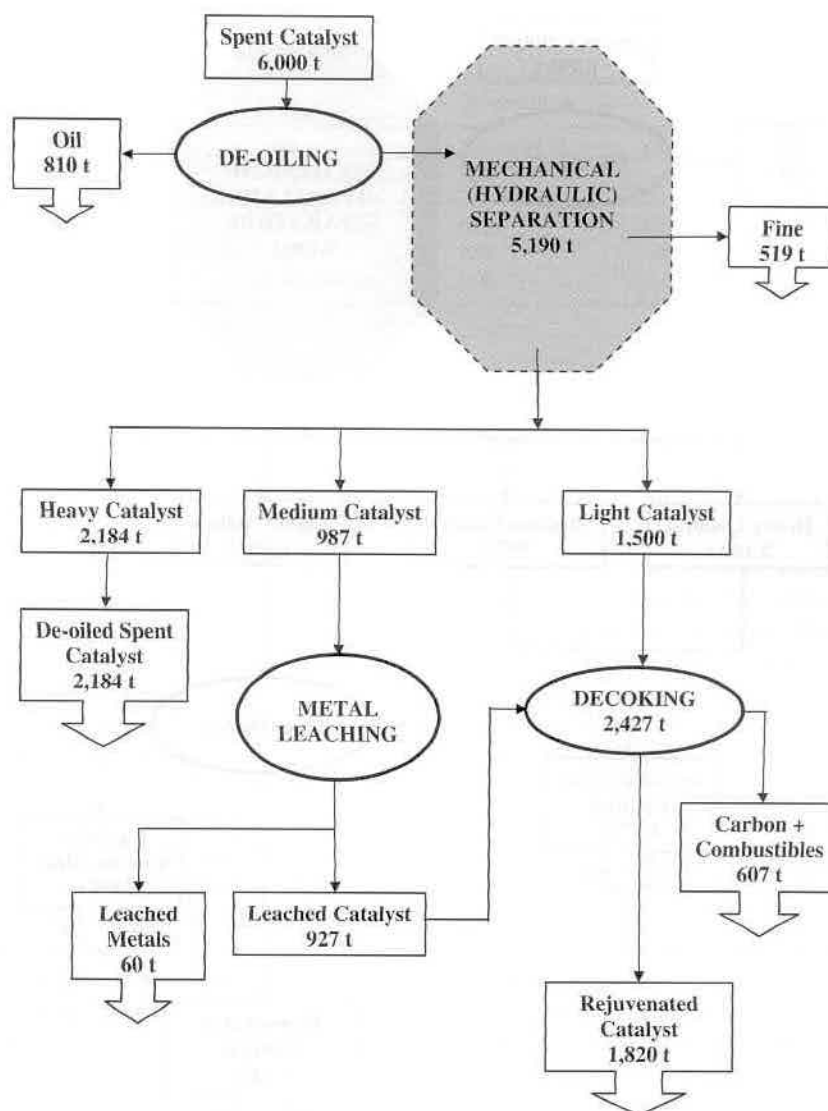


Fig. 3. Rejuvenation process—Option II.

particles stay in each stage can be calculated. Subsequently, the speed of the screw feeder and rotating scraper arm of the rotating-blade leaching column can be determined. Since the recommended space between two consecutive stages is 0.3 m, and the column diameter can be any size between 0.3 and 2.2 m. (Foust et al., 1960; Treybal, 1980), the overall column size is based on the daily catalyst throughput.

De-oiling is a physical process that involves no chemical reactions. Residence time depends on the physical and transport properties of the residual oils and solvents, such as the solubility, diffusivity and viscosity, which are functions of process temperature, pressure and flow rate. The main unit is a contacting column for co- or counter-current flow of the liquid and solid phase. The method used is based on a fresh solvent and solids feed that is mixed and separated in a counter-current leaching mode using an 'ideal stage concept' (Grosberg, 1950; Dahlstrom et al., 1997). This implies that the concentration of the solution

adhering to the catalysts in the underflow is assumed to be the same as that in the overflow at any stage. It also implies constant underflow rates or solvent-to-catalyst ratio between stages.

2.2. Mechanical separation process

After de-oiling, the amount of catalyst is reduced to 5190 tons/yr on a dry basis. Two processes are employed in the bench-scale separation—catalyst fines sieving and catalyst particle grouping according to density or jigging (Fig. 1). It would be advantageous to combine the two operations in commercial-scale operations to avoid breakage and attrition in fines sieving in the rest of the spent catalysts. This would also minimize particulate concentrations in air emissions. Furthermore, process operations and control could be improved. However, the particulate matter from spent catalyst fines is still a major

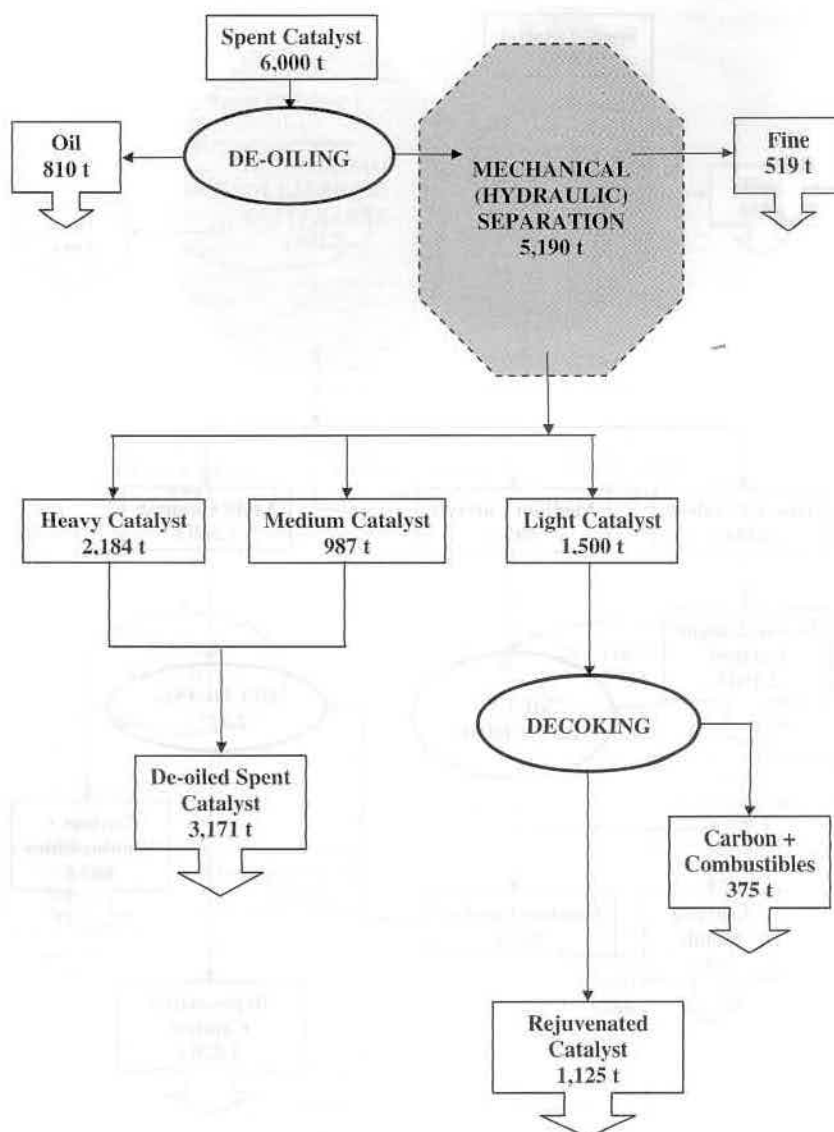


Fig. 4. Rejuvenation process—Option III.

Table 2
Mass input (tons/year) to sub-processes for the three options

Option	De-oiling	Mechanical (hydraulic) separation	Metal leaching	Decoking
I	6000	5190	3173	4353
II	6000	5190	987	2427
III	6000	5190	0	1500

environmental concern, which must be considered very carefully in the process design.

No drying is required for de-oiled catalyst before separation, saving energy. The wet catalyst, in the form of slurry after mixing with additional solvent, is fed to the tank at a predetermined rate. Since the same solvent as was used in the de-oiling is employed, the expected hazards and materials for construction of the unit are the same as for

the previous process. Also, environmental precautions must be applied in the process design.

The unit is primarily a phase separator employing the fluid mechanics. A surface-velocity classifier, such as a gravity settling vessel (Fig. 5) is used. It processes 5190 tons/yr of de-oiled catalyst at 298–313 K under atmospheric pressure. Separation depends on the particle terminal velocity, which is a function of the physical and transport properties of the solvents and catalyst particles. By adjusting the liquid flow rate from each compartment, the catalyst particles with different densities are separated. For the steady flow of a fluid past a moving solid, boundary layers are established. The force exerted on the solid is a combination of boundary layer drag, form drag, external force and buoyant force. The rate of separation of phases settling in a gravitational field is usually limited by the rate of fall of the smallest particles with residence time, which is important in determining the size of the unit.

Table 3
Mass output (tons/year) for the rejuvenated spent catalyst process

Option	Catalyst products			By-products			Total
	Fines	De-oiled spent catalyst	Rejuvenated spent catalyst	Residual oil	Leached metals	Carbons and combustibles	
I	519	0	3264	810	318	1089	6000
II	519	2184	1820	810	60	607	6000
III	519	3171	1125	810	0	375	6000

Table 4
Solvent properties and process temperatures

Solvent	Boiling point (K)	Solubility	Process temperature (K)
Naphtha	443	Excellent	400
Kerosene	503	Excellent	460
Diesel	600	Excellent	555
Gas oil	625	Very good	580

2.3. Integration of de-oiling and separation processes

Since de-oiling and mechanical separation are needed in all three process options, it will be advantageous to integrate the two processes to a single process. Additional benefits could be realized when the same liquid is used as both the solvent and the separation medium. This combined-process unit is designated as Unit 01. A schematic of the integrated de-oiling and mechanical separation process is shown in Fig. 6. The dotted lines show the gas flows, the thin solid lines show the liquid routes, and the thicker solid lines show the solids paths.

The spent catalyst is carried to a holding tank on top of the de-oiling column (01V01) via the spent catalyst escalator (01ES01). The solvent feed is pumped by the solvent recycle pump (01P04). A small amount of solvent is added to the catalyst to form a slurry feed. Most solvent feed, heated in the solvent feed heater (01H01) enters from the bottom of the de-oiling column (01V01). It moves upward to contact spent catalyst particles and remove the attached residual oil, while the catalyst moves downward. The de-oiled spent catalyst leaves 01V01 and is carried by the de-oiled catalyst escalator (01ES02) to the gravity settler (01V03) for separation according to density.

The effluent leaves from the top 01V01 and enters the distillation tower (01V02) for residual oil and solvent separation. The recovered solvent, as the top product of 01V02 passes through the solvent distillate heat exchanger (01E01) before joining the makeup solvent stream. The combined stream is pumped by the gravity settler unit feed pump (01P01A/B) to facilitate the separation of the de-oiled spent catalyst in 01V03. The separation depends on the solvent flow rate, which is controlled by the appropriate valves. After separating the de-oiled catalyst into light,

medium and heavy categories, it is dried in the dryer, 01V06. Each category of catalyst is transported individually on its own conveyer. The spent solvent from 01V03 is pumped by the filter press feed pump (01P02A/B) to the catalyst fines filter (01F01) to remove catalyst fines before joining the makeup solvent stream. The combined solvent stream is pumped by the solvent recycle pump (01P04) back to 01V01 for de-oiling. The recovered residual oil is passed to the residual oil raffinate heat exchanger (01E02) for waste heat recovery. It is then pumped by the heavy fuel fuel pump (01P03) to fuel the solvent feed heater (01H01), the drying air feed heater (01H02), and the decoker furnace of the decoking unit. The remainder is transferred to the fuel dump. The catalyst fines from 01F01, after removing any traces of solvent are dried and stored for sale.

The air supply is split into two streams upon exit from the air compressor (01C01). One air stream, at ambient temperature, is used to control the air feed temperature to 01V06. The other air stream, after recovering the waste heat from the three heat exchangers (01E01, 01E02 and 01E03), is heated in the drying air feed heater (01H02). Part of the heated air is transported to decoking unit for combustion. The remaining hot air mixes with the cool air stream in the drying air mixer (01V05) to obtain the desired temperature before entering 01V06 to dry the de-oiled catalyst. The spent air contains vapour from the solvent, which is stripped out of the air in the solvent stripper (01V04). The stripped air from the top of 01V04 joins the makeup air stream after heat recovering in 01E03, while the condensed solvent, from the bottom of 01V04, combines with the makeup solvent stream for recycling.

2.4. Metals leaching process

A fixed-bed reactor was used in the bench-scale metals leaching experiments. When it was scaled-up to pilot-plant experimentation, both fixed-bed reactors and ebullated-bed reactors (EBRs) were employed. The mechanically separated catalysts must be free of solvent before leaching can commence. The leaching agents used in the process are considered corrosive hazardous waste if discharged. In the process design developed herein, the leaching agents are regenerated and reused after removal of the leached metals, which are sold as low-grade metals without further processing.

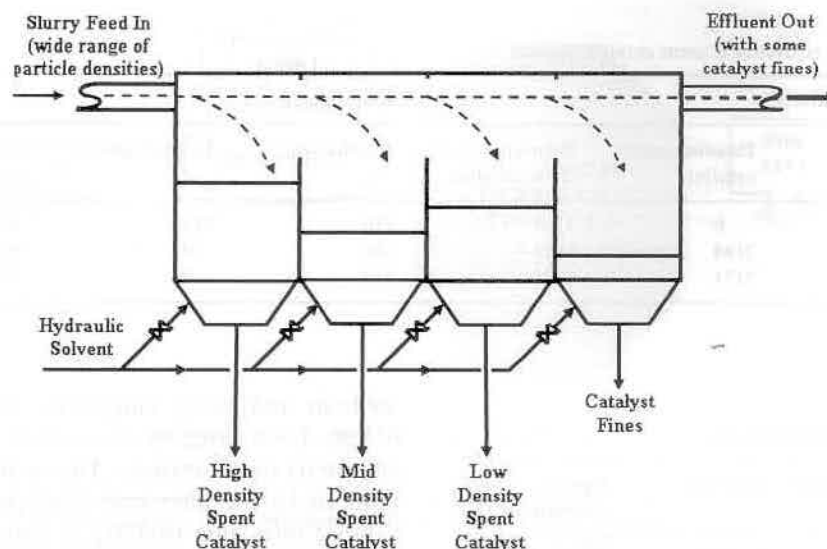


Fig. 5. A gravity settling unit for catalyst separation.

2.4.1. Design basis

The amount of spent catalyst and severity of metal contamination differ in the two process configuration options. Tables 2 and 3 show the details of a yearly mass balance. In Option I, 3171 tons of the heavy and medium portions of de-oiled catalysts are processed to yield 318 tons of leached liquid metals and 2853 tons of leached catalysts. However, only 987 tons of medium fouled catalysts are processed to produce 60 and 927 tons of leached metals and catalysts, respectively, in Option II. A solution of 6% of oxalic acid, 8% ferric nitrate in water is used as the leaching agent, and the optimal process temperature has been found to be 313 K. The results from the respective kinetic studies have been reported in the previous sections. A reagent regeneration unit is required for economic and environmental reasons.

In metals leaching, the solid undergoes a chemical reaction but is essentially insoluble in the fluid phase. It can be represented schematically as follows:

A (fluid) + B (solid) → soluble and insoluble product.

The reaction scheme consists of the following steps:

- diffusion of species A through liquid film adjoining solid particles;
- diffusion of species A through a layer of porous solids to the reaction surface; and
- chemical reaction of reactant A and solid species B.

The rate of any one of the three steps can be rate controlling.

From our previous study (Marafi et al., 1996), the leaching kinetics are found to be a dual controlling mechanism in which the initial leaching is under diffusional control while the latter stage is under chemical reaction control. A kinetic model covering these aspects has been

developed (Marafi et al., 1996). Briefly, the leaching process is considered to involve two operations:

- removal of metal foulants along the main mass transfer channels (macropores) connected to micropores until the pore structure begins to re-approach that of the fresh catalyst, and
- removal of metal foulants from the pore structure.

In the first stage, metal deposits restrict the movement of leaching agents and effluents in the pore network, and therefore, the intraparticle mass transfer becomes the rate-limiting step.

2.4.2. Process description

The fixed-bed reactor or EBR is the heart of this process and must be carefully considered in the design. An additional major item in the EBR unit is the ebullition pump, which must be capable of ebullating the spent catalysts continuously throughout the leaching operation.

The unit is designated as Unit 02 in which two EBRs (02V01A/B) are employed. This is a semi-continuous process in which the spent catalyst particles stay in one of the EBRs for the required residence time to yield optimum leaching, while the other is on standby. The process flow diagram is shown in Fig. 7. The leaching agent is prepared and stored in the leaching agent feed tank (02TK04). It is pumped by the leaching agent feed pump (02P05) to the leaching agent buffer tank (02TK01). The high- and medium-density de-oiled spent catalyst is carried by the processed catalyst escalator, 02ES01A/B, to the top of the EBR (02V01A/B) and the leaching agent is pumped from 02TK01 to 02V01A/B by the ebullition pump (02P02). Leaching reaction takes place while the catalyst particles are ebullated. When leaching is complete, the leached catalyst is transported to the leached

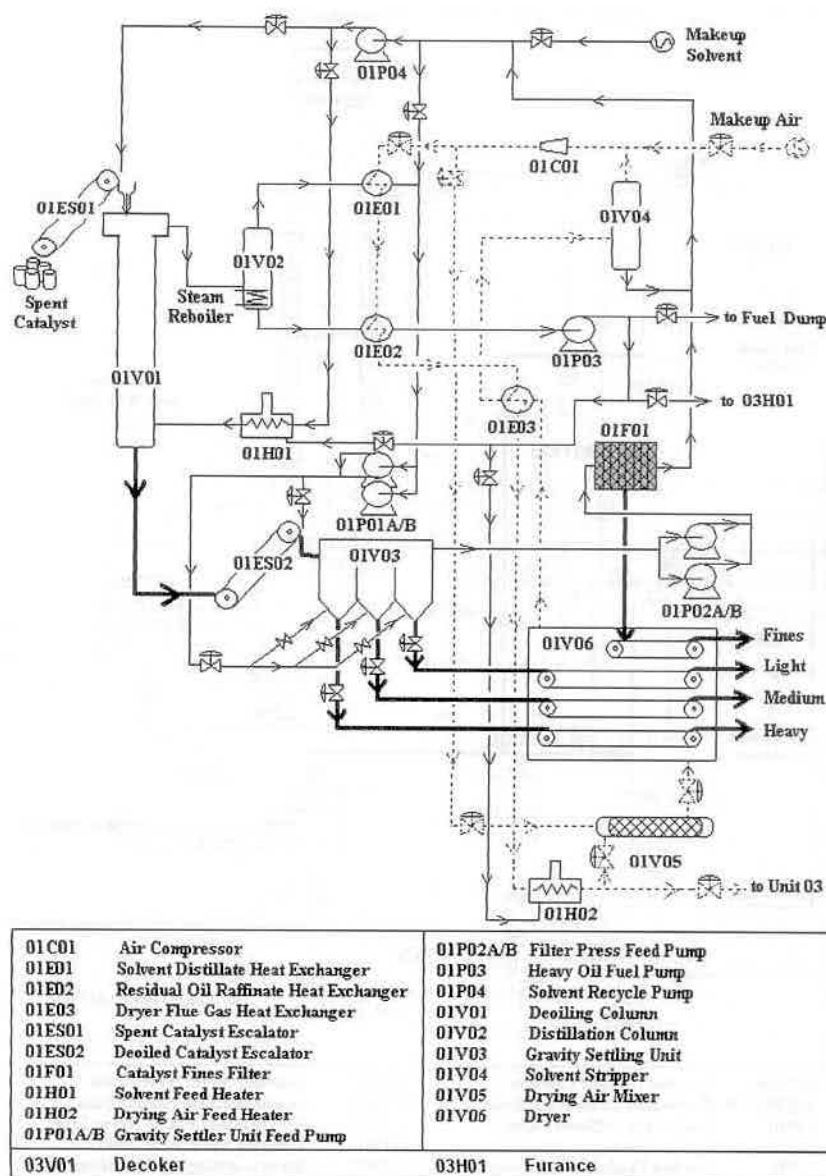


Fig. 6. Integrated de-oiling and mechanical separation process flow diagram.

catalyst-holding tank (02TK02) by fluidization using the leaching agent as the fluidizing media. Until all of the catalysts are unloaded in 02TK02, the leaching agent continues to be sent to the spent leaching agent holding tank (02TK05) for regeneration. The leaching catalyst is washed *in situ* in 02TK02 to remove any traces of the leaching agent. The leached catalyst wash pump, 02P01 circulates the water. Once leaching is completed, the catalyst is sent to Unit 03 for decoking through the leached catalyst conveyer (02CY01), and the fouled water is sent to the water treatment plant.

The regeneration of spent leaching agent takes place in any one of the three leaching agent regenerators (02V02A/B/C). The regenerated leaching agent is stored in the regenerated leaching agent tank (02TK03) from where it is pumped to 02TK04 for mixing with makeup leaching agent for recycling. The packing materials used to regenerate the

leaching agent in 02V02A/B/C also require reactivation. The washing agent is stored in the regenerator wash tank (02TK06) and circulated by the regenerator wash pump (02P06). It is then sent to the effluent treatment plant for further purification.

2.5. Decoking process

For most of the commercial catalytic processes in which catalysts are deactivated by coke deposition, a decoking unit is included in the catalyst regeneration process. This is accomplished mainly by burning off the coke under controlled temperature and oxygen concentration. Consequently, greenhouse gases and trace of gaseous sulphur compounds are produced as pollutants in air emissions, which must be minimized in the process design.

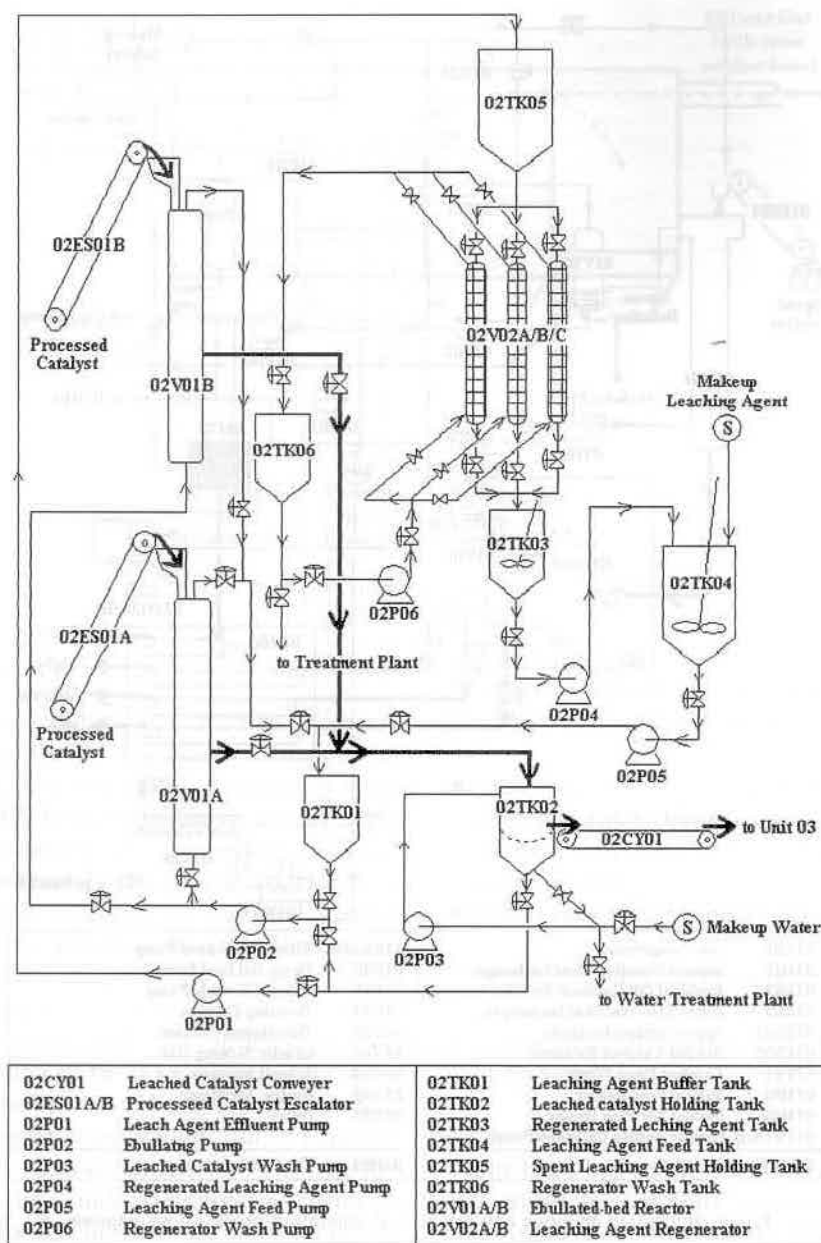


Fig. 7. Leaching process flow diagram.

2.5.1. Design basis

The amount of de-oiled and/or leached spent catalyst to be processed varies with process option, as shown in Table 2. The plant capacities are 4353 tons/yr for Option I, 2427 tons/yr for Option II, and 1500 tons/yr for Option III. Two reactor configurations, fixed-bed and moving-bed are attempted. The process temperature, reagent concentration and flow rate have been optimized in bench-scale in pilot-plant studies. The solids residence time is an additional parameter to be optimized in moving-bed reactor mode.

This unit is used to burn out the coke from catalysts. The effectiveness of coke removal in a fixed-bed or moving-bed reactor dictates the size of the contacting unit in processing the required throughput. The process temperature and reactant gas concentration, which are the most important

parameters, must be precisely controlled to produce a final product that meets specifications.

2.5.2. Process description

The decoking process (Fig. 8), is designated as Unit 03, which is capable of handling de-oiled, light-density spent catalyst or leached spent catalyst. The processed catalyst is carried up to the hopper of the decoker (03V03) by a processed spent catalyst escalator (03ES01). The decoker is a counter-current moving-bed reactor in which catalyst particles move down from the top while combustion air enters from the bottom. The reactor temperature is carefully maintained by tightly controlling the temperature in the decoker furnace (03H01) to obtain optimal properties in the rejuvenated catalyst. The combustion air supply

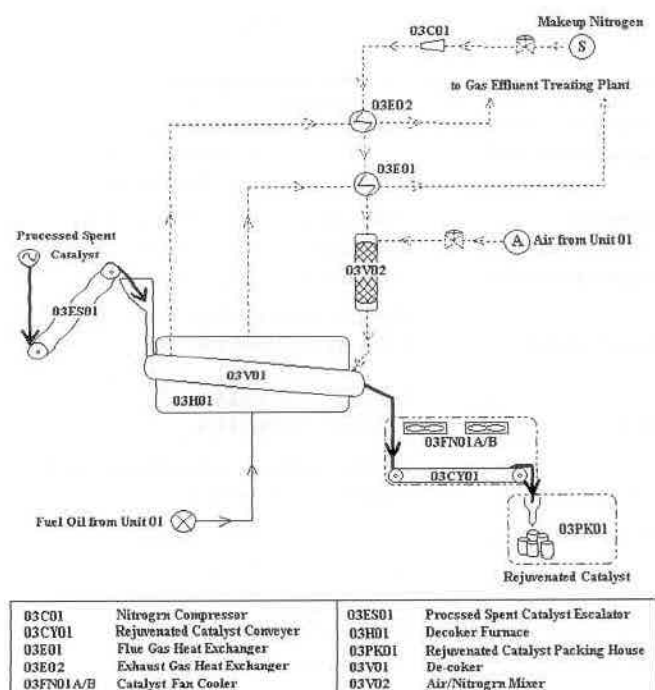


Fig. 8. Decoking process flow diagram.

to the decoker (03V01) must maintain a ratio of 1:20 for oxygen to nitrogen. The nitrogen streams are heated by the exhaust gas heat exchanger (03E02) and flue gas heater (03E01), from 03V01 and 03H01 where the heat is recovered, respectively. The heated air, coming from Unit 01, mixes with nitrogen gas at a specified ratio, in the air–nitrogen mixer before entering 03V01. The fuel oil for 03H01 also comes from Unit 01. The exhaust gases from 03E01 and 03E02 are sent to an effluent treatment plant to remove greenhouse gases and other harmful materials.

The decoked catalyst particles are cooled by two rejuvenated catalyst cooling fans (03FN01A/B), while they are transported on the rejuvenated catalyst conveyer (03ES01) to the rejuvenated catalyst packing house (03PK01) for packing as the final product.

2.6. Other auxiliary process

If the rejuvenation process is not constructed in a refinery complex, additional processing units are required. For instance, a medium-pressure (3.06 MPa) steam-raising plant is necessary for the heat duty in reboilers of any distillation columns or in strippers. Other utilities such as general water systems, fuel systems and power-generation plants are also required. In addition, treatment plants for gas and liquid effluents have to be installed to remove harmful materials before releasing them into the environment. Table 5 lists the additional process requirements for stand-alone operation. Moreover, site development, industrial buildings, and offsite facilities have to be considered. Therefore, the capital costs will be substantially increased.

3. Costs estimation

3.1. Capital cost estimation

To initiate the capital cost estimation, the following economical assumptions as well as operating and technical conditions are taken into considerations as follows:

- a plant capacity is 6000 tons of spent catalyst/yr;
- a plant useful life is 15 yr;
- an assumed on-stream factor of 0.905, for 330 days/yr with three 8-h shifts/day;
- direct capital costs covering free on board (FOB) cost of equipment, auxiliary materials needed up to commissioning, field labour, contingencies, and building and construction;
- indirect capital costs including engineering and supervision, design, freight, insurance, and taxes;
- working capital requirements, consisting of raw materials, finished product in stock, accounts receivable, cash in hand, accounts payable and taxes payable, which are assumed to be 25% of the total capital costs and for a 1-month payable period (Peters and Timmerhaus, 1991);
- three process configuration options are available, i.e., Options I, II, and III;
- direct and indirect capital cost estimations for each of the three processing units, Units 01–03, individually, with the total cost being the sum of the units required for each option configuration;
- depreciation based on the fixed assets only (direct capital investment) over a project life span of 15 yr applying straight-line depreciation model;
- amortization assumed to be 20% of the indirect capital investment using a straight-line method for the non-fixed assets only;
- interest charge estimated as 10% of the working capital cost;
- administration overhead assumed to be 20% of the factory costs;
- sales and promotion taken to be 15% of the actual sales revenue of the rejuvenated catalysts for each of the three options;
- a production capacity assumed to be 50% in the first year with 25% annual growth rate to reach full capacity in year 3;
- escalated cost factors taken as 3% for materials; 2% for labour and 2.2% for indirect costs;
- further costs incurred for the land, site preparation and other facilities required for chemical process operations, as well as extra manpower for the administration, management and operation of the additional units, if the project is not adopted by the refinery.

To estimate the direct and indirect capital costs, a commercial software package, CAPCOS Version 6 (Che-

Table 5
Additional process requirements for stand-alone operation

General type	Specific function	Specifications
Fuel system	Fuel oil, pumps, storage, piping control and distribution	
General water systems	Treated water, filtered and softened	55 000 gal/d
	Distilled water	5000 gal/d
	Drinking and service water, general facilities	10 000 gal/d
Power generation and distribution	Generation facilities	Capacity = 10 000 kVA
	Electrical power distribution for general purposes	Capacity = 10 000 kVA
	Main transformer station	Capacity = 10 000 kVA
	Secondary transformer station	Capacity = 10 000 kVA
Steam generation and distribution	Packaged boiler unit	Pressure = 500 psi Temperature = 500 F
	Steam distribution for general purposes	Capacity = 1 50 000 lb/h
Gas treatment plant	CO ₂ , CO, SO ₂ , solvent vapour	
Liquid effluent treatment plant	Acid, solvent	
Water treatment plant		
Chemical analytical laboratory	Daily chemical analysis on gas and liquid effluents	
	Quality control on reagents, water and solvent	
	Housing for all electronic process-control hardware, cables and accessories	
Process control room	Air-conditioning	

Table 6
Comparison of total capital cost for all options operated in a refinery and independently

Option	FOB (US \$ million)	Auxiliary materials (US \$ million)	Field labour (US \$ million)	Indirect cost (US \$ million)	Total (US \$ million)
Refinery operation					
I	5.0581	1.3232	1.4478	1.3191	9.1482
II	2.5375	0.8989	0.9550	0.9278	5.3192
III	2.0563	0.7779	0.8053	0.8014	9.1582
Independent operation					
I	6.2728	1.1045	2.2368	1.7772	11.3913
II	5.4507	0.8989	1.9882	1.5717	9.9095
III	4.9695	0.7779	1.8385	1.4453	9.0312

mEng Software, 1996) is used. The estimation is performed for each unit according to the respective process design.

The total capital costs reported previously for each option operating in a refinery are shown in Table 6. The four main cost elements of, i.e., FOB, auxiliary materials, field labour, and indirect costs are also given. The highest capital investment would be US \$6.80 million for Option I. This is followed by the others, in descending order: of Options III (US \$5.32 million) and Option II (US \$4.44 million). If the project is implemented independently, additional capitals in the amount of US \$4.6 million would be required (Table 6). The additional capital costs in each option are more than those of the catalyst rejuvenation process alone.

3.2. Production cost

Production costs are analysed for the three options either built in the refinery or independently over a project life of 15 yr based on the following assumptions:

- *Raw materials*: quantities and costs of raw materials differ for each option. Since spent catalyst has zero value, it is not counted in the raw materials costs.
- *Labour*: the three options assume operation in three 8-h shifts/day for 330 days/yr. Extra staff are needed for the project to be done independently.
- *Maintenance and repairs*: 4% of the direct capital investment costs.

- **Factory overhead:** 20% of the total labour costs for each option.
- **Operating supplies:** 15% of the maintenance and repair costs.
- **Insurance:** assumed to be 2% of the total direct capital investment costs.
- **Electricity and utilities:** electricity and water are used in all the units. Gas supply and extra water for cooling and washing are required for Unit 03. Utilities requirements vary for each option and, hence, the costs.
- **Depreciation:** for fixed assets over a project life span of 15 yr, applying a straight-line depreciation model.
- **Amortization:** 20% of the indirect capital investment costs using a straight-line method over the first 5 yr of the project's life span. The amortization is for non-fixed assets only.
- **Interest charge:** 10% of the working capital cost.
- **Administration overhead:** assumed 20% of the factory costs.
- **Sales and promotion:** 15% of the actual sale revenue of the rejuvenated catalysts only.

Basically there are two categories of production costs: operating and non-operating costs. Operating costs are subdivided into factory and non-factory costs. Together with the above assumptions, the total production costs are estimated and presented in Table 7 for the proposed spent catalyst rejuvenating process, operated in a refinery or independently.

Generally, the operating cost increases as the capacity of the plant increases. If the project is carried out independently, the operating costs will be higher than if it is carried out in the refinery. In this study, Option I is designed to rejuvenate all of the spent catalysts. However, the leaching of heavily fouled spent catalysts is more difficult to reach 90% metals removal. Consequently, the operating costs for Unit 02 in Option I of the spent catalyst rejuvenation process are far higher than those for Unit 02 in Options II and III. From the cost estimation, Option I operated independently has the highest yearly operating costs of US

\$17.23 million. The yearly operating costs decrease to US \$15.11 million if Option I is operated in a refinery. The yearly operating costs for Option II would be at US \$7.18 million if operated independently and US \$5.24 million if operated in a refinery, and for Option III, they would be US \$5.22 million if operated independently and US \$3.45 million if operated in a refinery.

3.3. Sale revenue estimation

To estimate the revenue to be derived from sale of various products, such as rejuvenated and de-oiled spent catalysts, catalysts fines, and low-grade fuel oil from the proposed process, the sale prices, after consultation with the local refineries, are taken to be as follows:

- **Sales price of rejuvenated catalyst:** The price of fresh catalyst, on average, is US \$6000/tons. For year 1, because the product would be new, the sales price is assumed to be 17% lower than that of fresh catalyst, or US \$5000/tons. The sales price is assumed to increase to 92% of the first year fresh catalyst price (US \$5500/tons) in the second year and to 96% (US \$5750/tons) in year 4. This assumption applies till the end of the project life.
- **Sales price of the de-oiled spent catalyst:** The sales price of de-oiled catalyst is assumed to be 7% of the fresh catalyst price of US \$5000/tons. The sales price will increase in accordance with the market price of fresh catalyst of US \$5500/tons in year 2 and US \$5750/tons in year 4. This assumption applies till the end of the project life.
- **Sales price of fuel oil:** The sales price of fuel oil is assumed to be 75% below the market price of US \$300/tons because the quality of the product is lower than that of the product available in the market.
- **Sales price of fines:** Fines will be sold at US \$100/tons throughout the project life span since their metals contents can be recovered as valuable products.

With the above price assumptions and based on the mass balance of the products of the proposed process, as shown in Table 3, a summary of the sale revenues expected for the three options is presented in Table 8.

4. Economic assessment

Economic assessment is a standard but necessary procedure to determine the economic feasibility of a project's profitability in terms of a number of economic measures, such as the net cash flow (NCF), discounted cash flow rate of return (DCFRR), net present value (NPV), internal rate of return (IRR), break-even point (BEP), and payback period (PBP), among others. Additional assumptions are made to assess the economics of commercializing this catalyst rejuvenation process as follows:

- The local refinery will buy back all the rejuvenated catalyst at market price.

Table 7
Production costs for operation in a refinery and independently

Cost (US \$ million)	Option I	Option II	Option III
Refinery operation			
Operation cost			
Factory cost	10.6311	2.8620	1.7701
Non-factory cost	3.7243	2.0523	1.4106
Non-operating cost	0.7547	0.3246	0.2695
Total production cost	15.1101	5.2389	3.4502
Independent operation			
Operation cost			
Factory cost	12.0610	3.9245	2.6558
Non-factory cost	4.3069	2.6349	1.9932
Non-operating cost	0.8677	0.6221	0.5670
Total production cost	17.2356	7.1815	5.2160

Table 8
Comparison of the expected first year sales revenues for the three options

Million US \$	Low-grade fuel oil	Catalyst fines	Leached metals	De-oil spent catalyst	Rejuvenated catalyst	Total
Option I	0.0459	0.0519	0.4144	0.0000	16.3200	16.8322
Option II	0.0459	0.0519	0.0782	0.1897	9.1000	9.4657
Option III	0.0459	0.0519	0.0000	0.2755	5.6250	5.9983

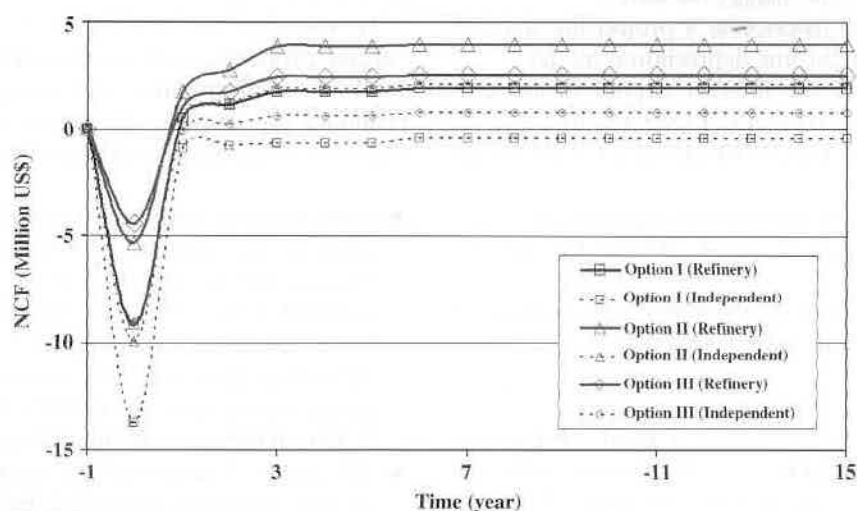


Fig. 9. NCF of the six basic scenarios at 0% DCFRR.

- The final product meets the local refinery's specifications and international standards (Oxenham Technology Associates, Inc., 1985).
- The technology will be exploited by the refinery group or a company independent from the refinery.
- Payment of the initial investment starts during the construction period.
- The minimum attractive IRR is 15% to implement the rejuvenation project.

4.1. Net cash flow (NCF)

NCF is fundamentally a measure of a company's financial health. It is defined as a measure of cash receipts less cash payments over a given period of time; or equivalently, net profit less amounts for depreciation, depletion, and amortization. There are six base-case scenarios, which cover three process-configuration options and two operator modes, i.e., each option operated by either a refinery or independently. Fig. 9 presents the NCF for each option, either in a refinery (solid lines) or independently (dotted lines). The greater the area above the baseline at NCF = 0, the more profitable is the process. The highest NCF appears for Option II operated in a refinery (solid line with triangle symbols). This is followed by Option III operated in a refinery (solid line with diamond symbols). The second option operated indepen-

dently (dotted line with triangle symbols) has the third highest NCF, and it is better than the first option operated in a refinery (solid line with square symbols). Option III operated independently (dotted line with diamond symbols) has the lowest positive NCF. However, the first option operated independently (dotted line with square symbols) has a negative NCF throughout the 15-yr project life and is the least profitable.

4.2. Discounted cash flow rate of return (DCFRR)

The DCF is a mechanical valuation method used to estimate the attractiveness of an investment opportunity. DCF analysis uses future free cash flow projections and discounts them to arrive at a present value, which is used to evaluate the potential for investment. If the DCF value obtained is higher than the current cost of the investment, the opportunity may be a good one. To project the cash flows by DCFRR analysis in the current study, a terminal value approach of a maximum of 15 yr is used because it will be harder to make realistic estimate of cash flows as time goes on.

Fig. 10 shows the accumulative NCF at 0% DCFRR for the six baseline scenarios for the three options (I, II, and III) and their respective operating modes (in a refinery or independently). Option I operated in a refinery has the lowest negative cash flow among the scenarios when the project starts, and it becomes worse as time goes on. In

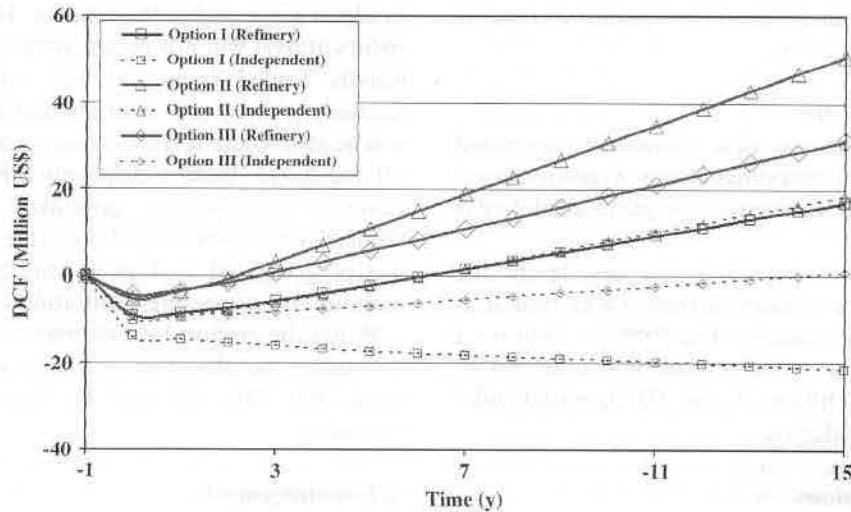


Fig. 10. Accumulative NCF of the six basic scenarios at 0% DCFRR.

Table 9
Summaries of NPV, IRR, BEP, and PBP

Economic analysis		Option I	Option II	Option III
Refinery operation				
NPV	Million US \$ @ 15% DCFRR	−0.03	15.12	5.57
IRR	%	3.40	54.70	16.91
BEP	% of full capacity	86.15	54.54	54.37
PBP	y	9.50	3.80	5.40
Independent operation				
NPV	Million US \$ @ 15% DCFRR	−17.22	−0.02	−5.92
IRR	%	—	14.95	—
BEP	% of full capacity	98.72	72.86	80.57
PBP	y	—	10.50	14.00

contrast, Option II operated in a refinery, which has a slightly deeper negative cash flow than Option III operated in a refinery, shows the best cash flow among the six baseline scenarios. Option II operated independently has higher returns than Option III operated in a refinery after 7 yr. Based on the DCFRR analysis, other economic indicators can be readily performed.

4.2.1. Net present value (NPV)

The NPV depends directly on the fractional interest rate, such as the DCFRR. Since it has been assumed that the commercial viability of a project should have a minimum 15% IRR, a comparison of the NPV of the six baseline scenarios is referenced at this DCFRR value, as presented in Table 9. The NPV from Option II operated in a refinery is the highest at US \$15.12 million, and it is followed by Option III also operated in a refinery (US \$5.57 million). The others have negative NPVs.

4.2.2. Interest rate of return (IRR)

The IRR has traditionally been defined as the discount rate at which the NPV is equal to zero. The strength of the IRR is in comparing project cost streams directly. The IRR

values determined for this study are shown in Table 9 for the six basic scenarios. Options II and III operated in a refinery are above the 15% acceptable IRR level for investment. The highest IRR, 54.7%, is obtained from Option II. This is followed by Option III at 16.9%. Option I operated in a refinery can only achieve 3.4% IRR. When the project is undertaken independently, Option II has 14.95% IRR, and Option III has only a 1.42% IRR, which is well below the acceptable level. Option I operated independently has a negative IRR.

4.2.3. Break-even point (BEP)

The BEP is defined as the production capacity required for each baseline scenario, in which total costs excluding depreciation equal total sales realization at full capacity production. Table 9 also presents the values for all base cases, which show that all options can breakeven, but at different capacities. Option II operated in a refinery has the lowest BEP at 54.54% of the full capacity. The highest BEP, 98.72%, is from Option I operated independently.

A lower BEP value is more profitable. It indicates that the spent catalyst rejuvenation process is not required to operate at full capacity all of the time. This reduces the

energy consumption requirement of the operating costs, as well as associated air emissions.

4.2.4. Payback period (PBP)

The BEP is an indicator that shows the expected financial return from an investment over a given period of time. Thus, the shorter the period, the more attractive is the project.

The period required to pay back the investment for Option II operated in a refinery is only 3.8 yr, but it is 10.5 yr if operated independently. The PBP for Options I and III operated in a refinery are 9.5 and 5.4 yr, respectively. However, Options I and III operated independently will never breakeven.

5. Summary and conclusions

Ever since spent catalysts were first classified as hazardous refinery waste, the disposal of the substantial amount generated yearly has become a major concern in industrial waste management, in terms of the environment, health, and safety. The ideal solution is to have a means, which should be based on the principles of a waste recycling to reactivate and reuse the waste, which can minimize the risk of non-compliance by the waste originators, and maintain the profitability of refining operations.

In this study, a rejuvenation process for spent catalyst capable of removing most of the hazardous contaminants, such as heavy metals, carbonaceous materials and residue oils is proposed. Three options are possible, depending on the amounts and types of spent catalyst to be rejuvenated. Laboratory testing on the rejuvenated catalyst show that their activity level can match the performance of fresh catalyst at a lower price. Careful consideration has been paid to minimizing the generation of further waste products and air emissions through the proposed rejuvenation process, and further improvement can be realized when a state-of-the-art control system is incorporated into the detailed engineering design for emissions control and waste minimization. Furthermore, the environment-oriented process design should be more acceptable to an environmental impact assessment (EIA), which is mandatory before granting an approval to build a new spent catalyst rejuvenation plant.

All the economic indicators in the economic assessment of the proposed spent catalyst rejuvenation process point to Option II operated in a refinery as the best scenario. It can achieve an IRR of 54.7%, reach the PBP in 3.80 yr, and reach the BEP at 54.54% of the full capacity. Thus, the refiners need not to pay for a third party to dispose of spent catalyst and risk potential environmental non-compliance in the future. Instead, the refinery will receive handsome profits from investing in the spent catalyst rejuvenation process.

Option II of the rejuvenation process is designed to process the quantities and types of spent hydroprocessing

catalysts generated in the refinery. However, heavily fouled spent catalyst will not be rejuvenated. Instead, the de-oiled heavily fouled spent catalyst can be sold for metals recovery. All of the medium and lightly fouled catalysts will be rejuvenated. Hence, this is an attractive approach to solving the problem of disposing the spent hydroprocessing catalysts continuously generated in the refinery. The refiners' objectives of reducing the amount of hazardous wastes generated and producing a cleaner environment without compromising profitability are clearly possible.

When the carbon tax becomes mandatory in Kuwait, it will easily be absorbed in the overall refinery operating costs, but this will not be the case for independent operators.

Acknowledgements

The authors would like to acknowledge the financial support of the Kuwait Institute for Scientific Research (KISR) and the Kuwait Foundation for Advancement of Sciences (KFAS). The authors would like also to acknowledge the support of the Kuwait National Petroleum Company (KNPC) in supplying the required information, as well as for their useful discussion of the results.

References

- Al-Dalama, K., Stanislaus, A., 2006. Comparison between deactivation patterns of catalyst in fixed bed and ebullating bed residue hydroprocessing units. *Chemical Engineering Journal* 120, 33–42.
- Berrebí, G., Dufresne, P., Jacquier, Y., 1994. Recycling spent hydroprocessing catalysts: Eurecat Technology. *Resources, Conservation and Recycling* 10, 1–9.
- Case, A., Garretson, G., Wiewiorowski, E., 1995. Ten years of catalyst recycling: a step to the future. In: Presented at the Third International Symposium on Recycling of Metals. Point Clear, November 12–15, Alabama, USA.
- Chang, T., 1998. Spent catalyst options. *Oil and Gas Journal* October 19, 79–84.
- ChemEng Software, 1996. *Process Engineers' Toolbox for Windows*, vol. 2. ChemEng Software and Services Ltd., Dorset, England.
- Chen, Y., Feng, Q., Shao, Y., Zhang, G., Ou, L., Lu, Y., 2006. Research on the recycling of valuable metals in spent Al_2O_3 -based catalyst. *Material Engineering* 19, 94–97.
- Dahlstrom, D.A., Bennett, R.C., Emmett, R.C., Harriot, P., Laros, T., Leung, W., McCleary, C., Miller, S.A., Morey, B., Oldshue, J.Y., Priday, G.P., Silberblatt, C.E., Slottee, J.S., Smith, J.C., Todd, D.B., 1997. Liquid–solid operations and equipment. In: Perry, R.H., Green, D.W., Maloney, J.O. (Eds.), *Perry's Chemical Engineers' Handbook*, seventh ed. McGraw-Hill, New York, pp. 18.1–18.133.
- Foust, A.S., Wenzel, L.A., Clump, C.W., Maus, L., Andersen, L.B., 1960. *Principles of Unit Operations*. Toppan Co, Tokyo.
- Furimsky, E., 1996. Spent refinery catalysts: environment safety and utilization catal. *Today* 30, 223–286.
- Furimsky, E., Massoth, F.E., 1993. Regeneration of hydroprocessing catalysts. *Catalysis Today* 17, 537–668.
- Furimsky, E., Massoth, F.E., 1999. Deactivation of hydroprocessing catalysts. *Catalysis Today* 52, 381–495.
- Habermehl, R., 1988. Safe handling and disposal of spent catalysts. *Chemical Engineering Progress* February, 16–19.
- Grosberg, A.B., 1950. Extension of Baker's method for variable underflow in liquid–solid extraction unit design. *Industrial and Engineering Chemistry* 42, 154–161.

- Juraiddan, M., Al-Shamali, M., Qabazard, H., Kam, E.K.T., 2006. A refined hydroprocessing catalysts deactivation and reactors performance model—pilot plant accelerated test applications. *Energy and Fuels* 20, 1354–1364.
- Kam, E.K.T., Al-Shamali, M., Juraiddan, M., Qabazard, H., 2005. A hydroprocessing multicatalyst deactivation and reactor performance model—pilot-plant life test applications. *Energy and Fuels* 19, 753–764.
- Kar, B.B., Datta, P., Misra, V.N., 2004. Spent catalyst: secondary source of molybdenum recovery. *Hydrometallurgy* 72, 87–92.
- Lianos, Z.R., Deering, W.G., 1997. GCMC's integrated process for recovery of metals from spent catalysts. In: Presented at the Air and Waste Management Association's 90th Annual Meeting, June 8–13, Toronto, Canada.
- Marafi, M., Furimsky, E., 2005. Selection of organic agents for reclamation of metals from spent hydroprocessing catalysts. *Erdöl Erdgas Kohle* 121, 93–96.
- Marafi, M., Stanislaus, A., 2003a. Options and processes for spent catalyst handling and utilization. *Journal of Hazardous Materials B* 101, 123–132.
- Marafi, M., Stanislaus, A., 2003b. Studies on rejuvenation of spent hydroprocessing catalysts by leaching of metal foulants. *Journal of Molecular Catalysis A: Chemical* 202, 117–125.
- Marafi, M., Stanislaus, A., Absi-Halabi, M., 1994. Heavy oil hydrotreating catalyst rejuvenation by leaching of foulant metals with ferric nitrate-organic acid mixed reagents. *Applied Catalysis B* 4, 19–27.
- Marafi, M., Kam, E.K.T., Stanislaus, A., Absi-Halabi, M., 1996. Rejuvenation of residual oil hydrotreating catalysts by leaching of foulant metals: modelling of the metal leaching process. *Applied Catalysis A* 147, 35–46.
- Marafi, M., Stanislaus, A., Kam, E.K.T., Jassem, F., Abu.Seedo, F., 1998. Optimization of a Continuous Process for the Rejuvenation of Spent Hydroprocessing Catalysts. Kuwait Institute for Scientific Research, KISR 5259, Kuwait.
- Oxenham Technology Associates, Inc., 1985. *Hydrotreating Catalysts: Market and Technology: 1984–1989*, Western Europe, Japan and Middle East.
- Peters, M.S., Timmerhaus, K.D., 1991. *Chemical Plants: Design and Construction*, fourth ed. McGraw-Hill, New York.
- Quann, R.J., Ware, R.A., Hung, C.W., Wei, J., 1988. Catalytic hydrodemetallation of petroleum. *Advances in Chemical Engineering* (14), 95–259.
- Rapaport, D., 2000. Spent hydroprocessing catalysts listed as hazardous wastes. *Hydrocarbon Processing* 79, 11–22.
- Rhodes, A.K., 1996. Catalyst handling and disposal become more important in environment era. *Oil and Gas Journal* March 18, 64–66.
- Stanislaus, A., Marafi, M., Absi-Halabi, M., 1993. Studies on the rejuvenation of spent catalysts: effectiveness and selectivity in the removal of foulant metals from spent hydroprocessing catalysts in coked and decoked forms. *Applied Catalysis A* 105, 195–203.
- Stanislaus, A., Marafi, M., Absi-Halabi, M., 1996. Residual oil hydrotreating catalyst rejuvenation by leaching of foulant metals: effect of metal leaching on catalyst characteristics and performance. *Chemical Industry* 67, 327–336.
- Treybal, R.E., 1980. *Mass-Transfer Operations*, third ed. McGraw-Hill, New York.
- Trimm, D.L., 1990. Deactivation, regeneration and disposal of hydroprocessing catalysts. *Studies in Surface Science and Catalysis* 53, 41–60.
- Trimm, D.L., 2001. The regeneration or disposal of deactivated heterogeneous catalysts. *Applied Catalysis A* 212, 153–160.
- USEPA, 2003. Hazardous waste management system. United States Environmental Protection Agency Federal Register, vol. 68(202), pp. 59935–59940.

A process integration approach to industrial water conservation: A case study for a Chinese steel plant

J.R. Tian^{a,b}, P.J. Zhou^{a,*}, B. Lv^c

^aCollege of Resources and Environmental Sciences, Wuhan University, No. 129 Luoyu Road, Wuhan, Wuhan, 430079, PR China

^bInstitute of Environmental Protection Science in Qingdao, No. 15 Jingshan Road, Qingdao 266003, PR China

^cSchool of Environment & Urban Construction, Wuhan University of Science and Engineering, Wuhan, 430079, PR China

Received 28 March 2006; received in revised form 27 November 2006; accepted 12 December 2006

Available online 7 May 2007

Abstract

A systematic approach to optimizing water network has traditionally been utilized to exam and plan water conservation in industrial processes. In the present case study, water-pinch technology was used to analyze and optimize the water network of a steel plant near China's Zhangjiakou city. A system design was developed and a limiting constraint (Cl⁻ concentration) was identified based on investigations of water quality then the minimum freshwater and wastewater targets were determined without considering water losses. The analysis was then extended by calculating the additional input of freshwater required to balance the actual water losses. A nearest-neighbor algorithm (NNA) was used to distribute the freshwater and recycled water among each of the plant's operations. The results showed that with some reconstruction of the water network, the flow rates of freshwater and wastewater could be decreased by 57.5% and 81.9%, respectively.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Process integration; Water conservation; Steel plant; Water network; Limiting constraints; Mass load; Water-pinch analysis

1. Introduction

Water is a vital component of many industrial operations, and is utilized for a wide range of purposes in industrial processes. With an increasing requirement for water as a result of a reduction in the available water resources, it is becoming increasingly necessary to implement water conservation measures in industrial processes so as to guarantee sustainable industrial development. In recent years, there has been considerable development of systematic methods to minimize the use of freshwater and the production of wastewater by industry in response to increasingly stringent environmental regulations, economic driving forces and public concern over the quality of the environment (Goldblatt et al., 1993; Hallale, 2002). The process integration approach is one method that is

often used to implement water recycling: in this approach, a flow of recycled water is used in an operation that produces a high level of contaminant (Wenzel et al., 2002).

Two main groups of methods are used for the systematic design of water-recovery networks: rigorous graphical targeting approaches (Wang and Smith, 1994; El-Halwagi, 1997; Prakash and Shenoy, 2005; Aly et al., 2005) and mathematical-based optimization techniques (Keckler and Allen, 1999). A systematic approach to designing optimal water recycling in an industrial plant comprises two main tasks (Dunn and Wenzel, 2001). The first task is to define the maximum amount of recycling that is possible, and the second task is to design alternative recycling networks that will allow the operations to meet the water recycling target. Wang and Smith (1995) studied operations with a fixed flow rate and suggested the use of local recycling and the splitting of water-using operations to meet the flow rate constraints. El-Halwagi et al. (2003) presented a rigorous graphical targeting approach that

*Corresponding author. Tel.: +86 27 87152823; fax: +86 27 68778893.
E-mail address: zhoupj@whu.edu.cn (P.J. Zhou).

minimized the use of fresh resources by means of segregation, mixing, and direct recycling and reuse strategies. Dhole et al. (1996) also proposed a targeting methodology for networks involving operations with a fixed flow rate. Polley and Polley (2000) improved the work by Dhole et al. (1996) and demonstrated that to minimize consumption of freshwater, it is necessary that the outlet streams leave at their maximum allowable concentration. Prakash and Shenoy (2005) used the nearest-neighbor principle to design minimum freshwater networks for fixed contaminant load problems and fixed flow rate problems. Manan et al. (2004) estimated the minimum water target using the water cascade analysis (WCA) technique, which is a numerical alternative to the graphical water targeting technique and can quickly yield an accurate estimate of the minimum water target, the pinch-point locations, and the water allocation target for maximum water recovery.

Takama et al. (1980) used mathematical programming to solve a refinery example problem. Similarly, a multi-level problem formulation and an interactive algorithm were presented for multi-objective optimization of large-scale systems associated with the hierarchical structure of decision-making (Takama and Loucks, 1981). Alva-Argáez et al. (1998) presented a solution approach for multi-component systems based on mathematical programming. In addition, Alva-Argáez et al. (1999) used certain simplifying assumptions to propose a transshipment model. Savelski and Bagajewicz (2000a, b) have shown that the outlet concentrations must equal their maximum allowable values in an optimal water network. They identified which processes need to be fed first using freshwater and what processes must receive each volume of wastewater by introducing the concept of monotonicity in the process-to-process connections. In addition, Gómez et al. (2001) provided an algorithmic procedure to address this matter. Bagajewicz and Savelski (2001) proposed a linear programming formulation of the problem based on the necessary conditions of optimality. Jacob et al. (2002) analyzed the water network of a pulp and paper processes using linear programming. A non-iterative algorithmic procedure was used to design water utilization networks in refineries and process plants (Savelski and Bagajewicz, 2001). Yang et al. (2000) proposed a mathematical optimization approach to design a wastewater reuse network for a cleaning and rinsing operation of an electroplating process, a papermaking process, and a semiconductor manufacturing process.

The process integration method has been broadly used in many countries. A company in Britain applied the analytical method in a sewage treatment plant and reduced freshwater consumption by 30%, decreased the investment from US\$15 to US\$3.5 million, and decreased the operating expenses and raw material costs by US\$1 million (Mann and Liu, 1999). Wenzel et al. (2002) used the process integration method to design a new water network

in a Danish factory that dyes polyester yarn by focusing on water savings and wastewater reduction. Thevendiraraj et al. (2003) improved the water network of a citrus plant to minimize water consumption and wastewater generation using this method, showing a reduction in freshwater consumption and wastewater generation of up to 30%. Sun et al. (2003) applied the method to hard chromium plating in China, and reduced freshwater consumption by 30%.

In the steel industry of developed countries, an average of 4.0 m^3 of freshwater has been required to produce 1 t of steel and the recycling rate averaged 97% (Qi and Li, 2003). In China, 15 m^3 is generally required to produce 1 t of steel (Zheng and Shan, 2005). In addition, there is wide variation between steel plants in their water-use efficiency. As a result, considerable potential for water conservation exists in steel plants in China, and this conservation could be achieved by reusing wastewater. This paper presents a case study of the steel plant located near China's Zhangjiakou city. The plant not only squanders valuable water resources but also pollutes the local environment. This inefficiency is caused by the plant's obsolete water network.

In this paper, we used a process integration approach to define a water conservation target for the steel plant. The aims of our study were to minimize the consumption of freshwater and the production of wastewater and to design optimal targets for water use so that the plant's outdated water network could be updated while accounting for current economic and technological restrictions.

2. Investigation of the water network

The main operations at the steel plant include several processes: ore dressing, power plant, hot air furnace, blast furnace, rinsing residue, and shower. Based on the production requirement of each process, the water flows and water quality parameters were provided by the managing engineer responsible for the steel mill and the results are shown in Table 1.

In total, the plant used 8370 t d^{-1} of freshwater and discharged 5910 t d^{-1} of wastewater. Rinsing residue consumed 1500 t d^{-1} of water. During this stage, 600 t d^{-1} of water was lost and the remaining 900 t d^{-1} of wastewater was recycled, so a total of 600 t d^{-1} of freshwater was required to replace the lost water. The other operations all lost water at rates that can be calculated from the data in Table 1 by subtracting the outlet value from the inlet value and the wastewater that is discharged without reuse. The water quality does not need to be considered in ore dressing and rinsing residue operations because water quality parameters do not affect the efficiency of these operations. However, water quality is important in all other phases because of the technology limitations. Shower is used for 3 h each day, whereas the remaining operations all run day and night.

Table 1
Process data for the steel plant's water system

Operation	Inlet and outlet	Flow rate (t d ⁻¹)	Suspended solids (mg L ⁻¹)	Alkalinity (CaO; mg L ⁻¹)	Rigidity (CaCO ₃ ; mg L ⁻¹)	pH	Cl ⁻ (mg L ⁻¹)
Ore dressing	Inlet	2400	0.02	91.39	20	6.7	20
	Outlet	2100	2.86	66.67	24	6.2	31
Rinsing residue	Inlet	1500	0.08	42.74	98	7.5	13
	Outlet	900	1.68	86.13	95	7.0	16
Hot air furnace	Inlet	1680	0.02	40	20	—	90
	Outlet	960	0.03	120	72	—	400
Blast furnace	Inlet	1920	0.02	52	20	—	75
	Outlet	1440	0.025	138	87	—	100
Shower	Inlet	90	0.015	91.39	20	6.7	20
	Outlet	90	0.02	138.07	23	7.0	45
Power plant	Inlet	1680	—	66	35	6.5	40
	Outlet	1320	—	90	60	6.5	45

3. Definition of design system and limiting constraint

Table 1 shows that the wastewater qualities in the ore dressing and rinsing residue operations are good. However, the ore dressing operation is far from the other operations and the wastewater it produces was considered to be recycled during this operation. As a result, ore dressing was excluded from the design of water network to avoid the need to construct long-distance pipelines. The rinsing residue operation was also excluded from the water network because the existing water recycling did not require any investment in reconstruction. Because of the smaller quantity of freshwater (about 1% of the total) and its high alkalinity in the shower operation, this operation was also omitted from the water network. However, relatively large quantities of freshwater were required by the power plant, hot air furnace and blast furnace, and water quality in these phases was constrained by the technology limitations, so the three operations were regarded as key components of the water network design.

As can be seen in Table 1, the concentration of Cl⁻ changes greatly between the input and output phases of these operations, except for the power plant phase. Furthermore, the concentration of Cl⁻ is more difficult to remove from wastewater than is the case for remediation of the other parameters. On this basis, the parameter of Cl⁻ concentration was chosen as the limiting constraint in the water network. The allowable inlet and outlet concentrations of Cl⁻ in Table 1 were used as the limiting concentrations, represented by $c_{i,in}^{lim}$ and $c_{i,out}^{lim}$, respectively. Essential data for the process integration study and the water-pinch analysis are presented in Table 2. In Table 2, each operation is divided into two components (Mann and Liu, 1999): one without water loss, and one with water loss. The operations Hot air furnace 1, Blast furnace 1 and Power plant 1 represent processes with no water loss, and Hot air furnace 2, Blast furnace 2 and Power plant 2 represent processes with the water losses of 720, 480 and 360 t d⁻¹, respectively. The value for water

Table 2
Essential process data for the water network design

Operation	Inlet flow rate (t d ⁻¹)	Outlet flow rate (t d ⁻¹)	$c_{i,in}^{lim}$ (Cl ⁻ ; mg L ⁻¹)	$c_{i,out}^{lim}$ (Cl ⁻ ; mg L ⁻¹)
Hot air furnace 1	960	960	90	400
Hot air furnace 2	720	0	90	—
Blast furnace 1	1440	1440	75	100
Blast furnace 2	480	0	75	—
Power plant 1	1320	1320	40	45
Power plant 2	360	0	40	—

with no loss in each operation equals the outlet flow rate of the operation.

4. Minimization of freshwater consumption and wastewater production

The target for minimum freshwater use (i.e., the apparent minimum freshwater; Mann and Liu, 1999) without considering water loss was calculated first. The limiting flow rate F_i (t d⁻¹) is determined by

$$F_i = 1000 \Delta m_{i,tot} / (c_i^* - c_0^*), \quad (1)$$

where c_0^* is the concentration of Cl⁻ in freshwater (which equaled 20 mg L⁻¹), c_i^* is concentration of Cl⁻ in the i th concentration interval (defined in Table 3) and $\Delta m_{i,tot}$ is the total mass load of Cl⁻ (kg d⁻¹) in the i th concentration interval = $\Delta m_1 + \Delta m_2 + \dots + \Delta m_i$. Δm_i (the mass load of Cl⁻ in the i th concentration interval) is determined by

$$\Delta m_i = \sum f_i \times (c_{i,out}^{lim} - c_{i,in}^{lim}) / 1000, \quad (2)$$

where $\sum f_i$ is the total flow rate of operations for the i th concentration interval, $c_{i,in}^{lim}$ and $c_{i,out}^{lim}$ represent the values in Table 2. The results of these calculations appear in Table 3. Arrows represent the range of observed Cl⁻ concentrations within a process (Table 1).

Table 3
Calculation of the limiting flow rate without considering water loss

Concentration interval i	Concentration (Cl^- , mg L^{-1})	Hot airfurnace (960 t d^{-1})	Blast furnace (1440 t d^{-1})	Power plant (1320 t d^{-1})	Δm_i (kg d^{-1})	$\Delta m_{i,\text{tot}}$ (kg d^{-1})	F_i (t d^{-1})
1	20				0	0	0
2	40				0	0	0
3	45				6.6	6.6	264
4	75				21.6		
5	90				24	28.2	402.86
6	100					52.2	652.5
7	400				288	340.2	895.26

In Table 3, the three operations are arranged in increasing order of contaminant concentration. The maximum F_i is the apparent minimum amount of freshwater required: with no loss of water from the system, the apparent minimum freshwater and wastewater both equal 895.26 t d^{-1} , so the apparent pinch point occurs at the last concentration interval, at a Cl^- concentration of 400 mg L^{-1} and a mass flow of $340.2 \text{ kg Cl}^- \text{ d}^{-1}$. The pinch means that freshwater must be used for mass transfers that take place from the lowest concentration to a concentration of $400 \text{ mg Cl}^- \text{ L}^{-1}$. No freshwater is needed in operations taking place at concentrations above the pinch point.

Additional freshwater must be added in all three operations in order to replace the water loss. The additional flow rate of freshwater required to replace this loss was determined as follows:

$$f = f_{\text{loss}} \times (c_{\text{pinch}} - c_{\text{in}}^{\text{lim}}) / (c_{\text{pinch}} - c_0^*), \quad (3)$$

where f (t d^{-1}) is the additional flow rate required to replace the loss, f_{loss} (t d^{-1}) is the actual loss of water in the operation, and c_{pinch} is the concentration at the apparent pinch point. The flow rate of water loss was calculated based on the principle of conservation of mass (Mann and Liu, 1999) that is, the mass load produced by the additional flow rate is equal to that produced by the lost water. Based on Eq. (3), the additional flow rates of freshwater required by the hot air furnace, blast furnace and power plant were 587.37 , 410.53 and 341.05 t d^{-1} , respectively. The minimum flow rate of freshwater consists of the apparent minimum flow rate (895.26) plus the additional flow rates, and thus equals 2234.21 t d^{-1} .

5. Water network synthesis

A nearest-neighbor algorithm (NNA; Prakash and Shenoy, 2005) was applied to optimization of the water

network to meet the minimum freshwater target of 2234.21 t d^{-1} . The outlet flow rates of these operations required to fulfill the inlet (in such a manner that the inlet concentration is equal to the maximum allowable value) are determined by simultaneously solving the overall flow rate balance and the contaminant material balance equations (Prakash and Shenoy, 2005) given below

$$f_{Rk,Sp} + f_{Rm,Sp} = f_{Sp} - \sum f_{Ri,Sp}, \quad (4)$$

$$f_{Rk,Sp} c_{Rk} + f_{Rm,Sp} c_{Rm} = f_{Sp} c_{Sp} - \sum f_{Ri,Sp} c_{Ri}, \quad (5)$$

where Rk (Rm) and Sp are the numbers of outlets and inlets, respectively, in order of increasing Cl^- concentration; $f_{Rk,Sp}$ ($f_{Rm,Sp}$) and c_{Rk} (c_{Rm}) are the flow rates (t d^{-1}) and Cl^- (mg L^{-1}) concentrations, respectively, of Rk (Rm) flowing into Sp ; f_{Sp} and c_{Sp} are the flow rate (t d^{-1}) and Cl^- (mg L^{-1}) concentration of Sp , respectively; $i = k + 1$ to $i = m - 1$. Freshwater is $R0$. The power plant inlet ($S1$) has a Cl^- concentration of 40 mg L^{-1} , and based on $R0$ and the power plant inlet ($R1$), the equation can be solved to obtain the $S1$ flow rate (1680 t d^{-1}). However, the calculated flow rate for $R1$ (1344 t d^{-1}) is more than the available amount (1320 t d^{-1}). Thus, water from the blast furnace ($R2$) must be used. According to Eqs. (4) and (5), $R1$ (1320 t d^{-1}) was used completely and 352.5 t d^{-1} of $R0$ and 7.5 t d^{-1} of $R2$ were needed to satisfy the constraints for $S1$. Based on the NNA, the entire demand for the blast furnace ($S2$) was supplied by $R0$ and $R2$. The hot air furnace ($S3$) has a Cl^- concentration of 90 mg L^{-1} , which is less than that of $R2$ and $R3$. $R2$ was considered first and was used completely due to its smaller amount, so $R3$ was also needed. After repeating the calculation in Eqs. (4) and (5), 1281.71 t d^{-1} of freshwater, 112.5 t d^{-1} of $R2$ and 285.79 t d^{-1} of $R3$ could provide $S3 = 1680 \text{ t d}^{-1}$. The remaining 674.21 t d^{-1} of wastewater could be discharged

into the drainage system. The final network that results from these calculations is shown in Fig. 1.

In order to maximize water conservation in the steel plant, the ore dressing and rinsing residue operations should be reconstructed. For the ore dressing wastewater, water losses occur during filtration and a 20% loss was planned (Qi and Li, 2003). Therefore 1680 t.d^{-1} of wastewater ($2100 \times (100\% - 20\%) = 1680$) could be provided for recycling and another 720 t.d^{-1} of freshwater ($2400 - 1680$; Fig. 2) would be needed in the ore dressing operation. Rinsing residue had an inlet concentration of $13 \text{ mg Cl}^{-1} \text{ L}^{-1}$ before reconstruction, but the inlet water quality of the operation was not a severe constraint. As a

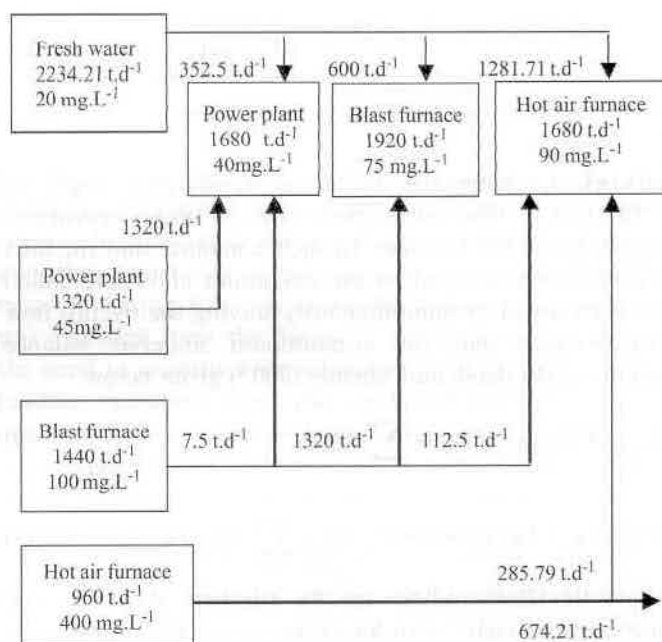


Fig. 1. The minimum freshwater network for the steel plant determined using the NNA.

result, 600 t.d^{-1} of wastewater from the hot air furnace can be used in the rinsing residue operation and the remaining 74.21 t.d^{-1} discharged into the drainage system. Fig. 2 shows the original and reconstructed pipelines for five of the operations and the corresponding freshwater flow rates.

6. Discussion

After improving the water network for three operations (the hot air furnace, blast furnace and power plant) based on the water-pinch analysis, the total flow rate of freshwater was reduced from 5280 to 2234.21 t.d^{-1} , which represents a 57.5% freshwater savings. The flow rate of wastewater was reduced from 3720 to 674.21 t.d^{-1} , representing a reduction of 81.9%. If the plant runs 340 days per year, the freshwater conservation and wastewater reduction would both equal $1.04 \times 10^6 \text{ t.a}^{-1}$. Reconstructing the water network for the five operations in Fig. 2 would reduce the total freshwater flow rate by 64.3% and reduce wastewater discharge by 92.6%. The conservation of freshwater and reduction of wastewater would then be equal to 1.81×10^6 and $2.17 \times 10^6 \text{ t.a}^{-1}$, respectively. Based on local water price, more than US\$0.5 million would be saved every year versus an investment of US\$0.15 million to pay for the construction. The estimated cost of freshwater saving is enough to pay for the cost of reconstruction and operation in less than 1 year. When the process integration approach is used to design or improve the water network in an industrial process, a detailed survey of the streams that supply water to the plant must be carried out and the limiting constraint must be defined.

7. Conclusion

The process integration approach described in this paper provided a powerful tool for planning water conservation

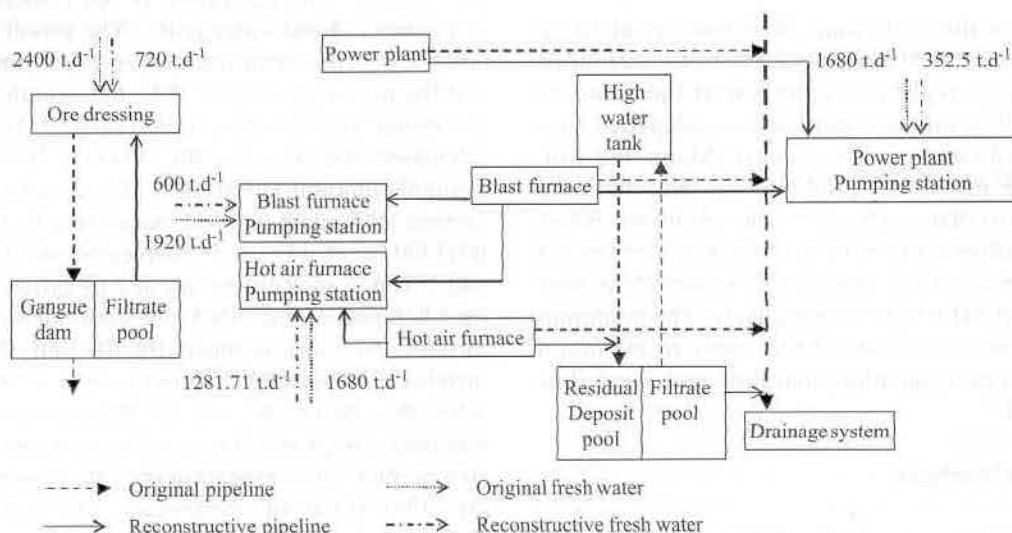


Fig. 2. Drainage flow sheet for the major operations in the steel plant.

in the steel plant described in this case study. After improving three operations (the hot air furnace, blast furnace and power plant) using the water-pinch approach, the total flow rates of freshwater and wastewater were reduced by 57.5% and 81.9%, respectively. After reconstruction of the water network for the five main operations, the total freshwater flow rate could be reduced by 64.3% and the wastewater flow rate could be reduced by 92.6% resulting in savings of more than US\$0.5 million every year. Greater economic and environmental efficiency can both be achieved by means of the proposed improvements in the water network.

The efficiency of the improved network is obvious when the flow rate is large and the limiting constraints are relatively few. The actual water network that results from the analysis in this study should be adjusted to account for the tradeoff between capital and operating costs as well as the complexity and operability of the resulting processes.

Acknowledgment

We thank the managers of the Zhangjiakou steel plant for their cooperation and financial support of our study.

References

- Alva-Argáez, A., Kokossis, A.C., Smith, R., 1998. Wastewater minimisation of industrial systems using an integrated approach. *Computers and Chemical Engineering* 22 (Suppl.), S741–S744.
- Alva-Argáez, A., Vallianatos, A., Kokossis, A.C., 1999. A multicontaminant transshipment model for mass exchange networks and wastewater minimisation problems. *Computers and Chemical Engineering* 23 (10), 1439–1453.
- Aly, S., Abeer, S., Awad, M., 2005. A new systematic approach for water network design. *Clean Technologies and Environmental Policy* 7 (3), 154–161.
- Bagajewicz, M.J., Savelski, M.J., 2001. On the use of linear models for the design of water utilization systems in process plants with a single contaminant. *Chemical Engineering Research and Design (Part A)* 79, 600–610.
- Dhole, V.R., Ramchandani, N., Tainsh, R.A., Wasilewski, M., 1996. Make your process water pay for itself. *Chemical Engineering* 103 (1), 100–103.
- Dunn, R.F., Wenzel, H., 2001. Process integration design methods for water conservation and wastewater reduction in industry. Part 1: Design for single contaminants. *Clean Products and Processes* 3 (3), 307–318.
- El-Halwagi, M.M., 1997. *Pollution Prevention Through Process Integration: Systematic Design Tools*. Academic Press, San Diego.
- El-Halwagi, M.M., Gabriel, F., Harell, D., 2003. Rigorous graphical targeting for resource conservation via material recycle/reuse networks. *Industrial & Engineering Chemical Research* 42 (19), 4319–4328.
- Goldblatt, M.E., Eble, K.S., Feathers, J.E., 1993. Zero discharge: what, why and how. *Chemical Engineering Progress* 89 (4), 22–27.
- Gómez, J., Savelski, M., Bagajewicz, M., 2001. On a systematic design procedure for water utilization systems in refineries and process plants. *Chemical Engineering Communications* 186, 183–203.
- Hallale, N., 2002. A new graphical targeting method for water minimization. *Advances in Environmental Research* 6 (3), 377–390.
- Jacob, J., Kaipe, H., Coudere, F., Paris, J., 2002. Water network analysis in pulp and paper processes by pinch and linear programming techniques. *Chemical Engineering Communications* 189 (2), 184–206.
- Keckler, S.E., Allen, D.T., 1999. Material reuse modeling: a case study of water reuse in an industrial park. *Journal of Industrial Ecology* 2 (4), 79–92.
- Manan, Z.A., Tan, Y.L., Foo, C.Y., 2004. Targeting the minimum water flowrate using water cascade analysis technique. *AIChE Journal* 50 (12), 3169–3183.
- Mann, J.G., Liu, Y., 1999. *Industrial Water Reuse and Wastewater Minimization*. Petrochemical Press, Beijing, China, pp. 158–160.
- Polley, G.T., Polley, H.L., 2000. Design better water networks. *Chemical Engineering Progress* 96 (2), 47–52.
- Prakash, R., Shenoy, U.V., 2005. Targeting and design of water networks for fixed flowrate and fixed contaminant load operations. *Chemical Engineering Science* 60 (1), 255–268.
- Qi, L.L., Li, Y.C., 2003. Question and Answer of Industrial Water and Water Conservation Management. China Petrochemical Press, Beijing, China, p. 87 (in Chinese).
- Savelski, M.J., Bagajewicz, M.J., 2000a. On the optimality conditions of water utilization systems in process plants with single contaminants. *Chemical Engineering Science* 55 (21), 5035–5048.
- Savelski, M., Bagajewicz, M., 2000b. Design of water utilization systems in process plants with a single contaminant. *Waste Management* 20 (8), 659–664.
- Savelski, M.J., Bagajewicz, M.J., 2001. Algorithmic procedure to design water utilization systems featuring a single contaminant in process plants. *Chemical Engineering Science* 56 (5), 1897–1911.
- Sun, H., Feng, L.M., Feng, L.J., 2003. Water-pinch technology applied in the techniques of plating and rinsing of hard chromium. *Materials Protection* 36 (11), 60–62 (in Chinese).
- Takama, N., Loucks, D.P., 1981. Multi-level optimization for multi-objective problems. *Applied Mathematical Modelling* 5 (3), 173–178.
- Takama, N., Kuriyama, T., Shiroko, K., Umeda, T., 1980. Optimal water allocation in a petroleum refinery. *Computers and Chemical Engineering* 4 (4), 251–258.
- Thevendiraraj, S., Klemesš, J., Paz, D., Aso, G., Cardenas, G.J., 2003. Water and wastewater minimisation study of a citrus plant. *Resources Conservation and Recycling* 37 (3), 227–250.
- Wang, Y.P., Smith, R., 1994. Wastewater minimisation. *Chemical Engineering Science* 49 (7), 981–1006.
- Wang, Y.P., Smith, R., 1995. Wastewater minimisation with flowrate constraints. *Transactions of the Institution of Chemical Engineers—Part A* 73, 889–904.
- Wenzel, H., Dunn, R.F., Gottrup, L., Kringelum, J., 2002. Process integration design methods for water conservation and wastewater reduction in industry. Part 3: Experience of industrial application. *Clean Technologies and Environmental Policy* 4 (1), 16–25.
- Yang, Y.H., Lou, H.H., Huang, Y.L., 2000. Synthesis of an optimal wastewater reuse network. *Waste Management* 20 (4), 311–319.
- Zheng, X.P., Shan, W.C., 2005. Industrial water saving measures of Jinan iron and steel group. *Metallurgical power* 110 (4), 57–58 (in Chinese).

Urban sprawl and air quality in large US cities

Brian Stone Jr.*

City and Regional Planning Program, College of Architecture, Georgia Institute of Technology, Atlanta, GA 30332-0155, USA

Received 3 April 2006; received in revised form 20 November 2006; accepted 14 December 2006

Available online 26 March 2007

Abstract

This study presents the results of a paper of urban spatial structure and exceedances of the 8-h national ambient air quality standard for ozone in 45 large US metropolitan regions. Through the integration of a published index of sprawl with metropolitan level data on annual ozone exceedances, precursor emissions, and regional climate over a 13-year period, the association between the extent of urban decentralization and the average number of ozone exceedances per year, while controlling for precursor emissions and temperature, is measured. The results of this analysis support the hypothesis that large metropolitan regions ranking highly on a quantitative index of sprawl experience a greater number of ozone exceedances than more spatially compact metropolitan regions. Importantly, this relationship was found to hold when controlling for population size, average ozone season temperatures, and regional emissions of nitrogen oxides and volatile organic compounds, suggesting that urban spatial structure may have effects on ozone formation that are independent of its effects on precursor emissions from transportation, industry, and power generation facilities.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Urban form; Ozone formation; Urban planning; Land use; Air quality; Urban sprawl

1. Introduction

Identified in the popular media as the source of numerous social, environmental, and economic ills, the phenomenon of urban sprawl is increasingly the subject of more rigorous inquiry in the peer-reviewed literatures on economics, urban planning, and public health. Generally defined as decentralized land use patterns characterized by low population densities and auto-oriented design schemes, urban sprawl has been demonstrated to greatly elevate the cost of urban services by increasing the distance between new development and the established infrastructure of roads, sewer lines, and transit systems (Burchell et al., 2002; Real Estate Research Corporation, 1974). Additional work has associated sprawling urban development patterns with increased vehicle travel and congestion (Ewing et al., 2003; Downs, 1992), increased volumes of stormwater runoff (Stone and Bullen, 2006), loss of prime agricultural lands (Heimlich and Anderson, 2001), and, perhaps most

provocatively, increased rates of obesity in children and adult populations (Frumkin et al., 2004).

In this study, I explore the implications of urban form for air quality within the largest metropolitan regions of the United States. Through the integration of data on land use attributes and air quality trends recorded in 45 of the 50 largest US metropolitan regions, a quantitative index of urban sprawl is associated with the emissions of ozone precursors and the annual number of high ozone days in each region between 1990 and 2002. While a handful of studies has demonstrated an association between various attributes of urban development and vehicle emissions in individual cities (Frank et al., 2000; Johnston et al., 2000), this work is the first to assess the implications of sprawl for an ambient measure of regional air quality in multiple cities while controlling for population, precursor emissions, and meteorological attributes significant to ozone formation.

Specifically, this study addresses two principal research questions. First, are metropolitan regions characterized by high levels of sprawl associated with a greater quantity of ozone precursor emissions from vehicles and industry than more compact regions? And, second, do metropolitan regions characterized by high levels of sprawl experience a

*Tel.: +1 404 894 6488.

E-mail address: stone@gatech.edu.

greater number of annual ozone exceedances than more compact regions? The significance of these questions lies in their potential to inform land-use strategies to combat air pollution in large cities.

The results of this study indicate that for the 45 US metropolitan regions surveyed urban form is significantly associated with both ozone precursor emissions and ozone exceedances during a 13-year study period. Significantly, a positive association between sprawl and ozone exceedances was found to hold true when controlling for average ozone season (May through September) temperatures and annual emissions of chemical precursors to ozone formation, suggesting that the well-established linkage between decentralized development patterns and auto use may be only one of multiple mechanisms through which sprawl influences air quality. Overall, the most sprawling cities were found to experience over 60% more high ozone days than the most compact cities.

2. Background

The literature on land use and air quality has explored two principal mechanisms through which the physical dimensions of urban form can influence regional air quality. The first and most widely investigated set of interactions concerns the linkages between land use, vehicle travel, and vehicle tailpipe emissions. The second set of interactions, posited in a small but growing number of studies, concerns the linkages between land use, local meteorology, and regional air quality. A brief overview of these two areas of research is provided in the following paragraphs to establish a basis for understanding the potential significance of urban sprawl to ozone formation.

2.1. Urban form, vehicle travel, and emissions

A significant relationship between land use and vehicle travel has been widely documented (Transportation Research Board, 1995; Apogee, 1998). Perhaps the most compelling evidence of this relationship is provided by the handful of studies that has examined readily available measures of land use and travel within a large number of cities. In one of the most widely cited of these studies, Newman and Kenworthy (1989) documented a strong and significant negative relationship between population density and per capita fuel usage within 63 large metropolitan regions around the world ($R^2 = 0.86$).¹ Similar significant relationships have been found to exist between population density and vehicle ownership, vehicle trip generation, and vehicle miles traveled (VMT) in American cities and abroad (Pucher and Lefevre, 1996).

¹It should be noted that the Newman and Kenworthy study has been challenged on methodological grounds pertaining to the authors' failure to control for demographic influences and the use of inherently related compound variables. For a more thorough discussion of these issues, see Brindle (1994) and Gordon and Richardson (1989).

In addition to these widely cited multi-city studies, projects employing travel simulation models or travel survey data from a single metropolitan region have also tended to illustrate a weak to moderate but statistically significant relationship between urban form and travel behavior. Of 18 studies of this nature reviewed by the author, 15 documented a statistically significant relationship between the density of development, land use mix, and/or the configuration of the street network (i.e., street connectivity) and some measures of vehicle travel (e.g., Frank and Pivo, 1994; Cervero and Gorham, 1995). For example, in a study of land use and vehicle travel in Sacramento, California, Johnston et al. (2000) found the combined impacts of transit improvements and land use strategies to reduce VMT between 4% and 7% over 20 years. In one of the most widely cited studies of this nature, the Land Use, Transportation, and Air Quality (LUTRAQ) analysis conducted in Portland, Oregon, a compact development scenario emphasizing transit-oriented development, pedestrian infrastructure improvements, and transportation demand management policies was found to reduce daily VMT by 8% relative to the business as usual scenario of highway expansion (Cambridge Systematics, Inc. and Parsons Brinkerhoff Quade and Douglas, Inc., 1992).

While the relationship between land use and vehicle travel generally has been assumed to hold implications for tailpipe emissions and air quality, only a handful of studies has sought to statistically link urban form to vehicle emissions directly. In addition to establishing an association between transit-oriented development and daily VMT, the LUTRAQ study cited above also found the emissions of nitrogen oxides (NO_x) and carbon monoxide (CO) to be reduced by 6% and 3%, respectively, in response to various urban development scenarios (Cambridge Systematics, Inc. and Parsons Brinkerhoff Quade and Douglas, Inc., 1992). Through the integration of household travel survey data obtained in Seattle, Washington with a mobile source emissions model, Frank et al. (2000) found statistically significant relationships between household density, employment density, street connectivity, and tailpipe emissions. Specifically, vehicle emissions of CO, NO_x , and volatile organic compounds (VOC) were found to exhibit a significant negative relationship with household and employment density when controlling for household size, income, and vehicle ownership. While limited in number, these studies provide support for the hypothesis that lower density patterns of development are associated with a greater magnitude of vehicle emissions, with potential implications for regional air quality.

2.2. Urban form and regional climate

In addition to vehicle travel and emissions, the spatial structure of metropolitan regions has been associated with meteorological phenomena that are important to regional air quality. The most widely recognized meteorological

effects of urbanization result from the urban “heat island effect” (Oke, 1987). Defined as a differential in the air temperatures of urban centers relative to adjacent rural areas, the urban heat island effect is driven by the displacement of natural vegetation by the impervious surfaces of roads and buildings, as well as by the emission of vast quantities of waste heat from buildings, industry, and automobiles. In combination, these properties of urbanization can serve to raise by several degrees the average air temperature of large cities. Because regionalized air pollutants such as ozone and fine particulate matter are sensitive to temperature, the resulting urban heat island holds important implications for both climate and air quality.

Numerous authors have illustrated a relationship between the intensity of urban heat islands and weather events downwind of large cities. Diem and Mote (2005) found heavy rainfall events to increase by more than 40% in one location in suburban Atlanta over a 50-year period and associated this phenomenon with the region’s rapid pace of development and growing heat island. Likewise, Shepherd et al. (2002) found an approximate 20% increase in rainfall events downwind of the metropolitan Atlanta region relative to an upwind control region during the period of 1998–2000. Other studies have shown similar urban rainfall anomalies in New York, New York (Bornstein and LeRoy, 1990) and St. Louis, Missouri (Rozoff et al., 2003).

Of perhaps greater significance to public health in cities is the potential influence of the urban heat island effect on ground-level ozone formation. Ozone is a colorless gas that is formed in the presence of sunlight through the chemical interaction of NO_x and VOC, two classes of pollutants produced from the combustion of fossil fuels. Ambient heat influences ground-level ozone formation through its effects on the rate of emission of ozone precursors from both human and natural sources. Urban heat islands can intensify ozone formation through at least two mechanisms. First, enhanced temperatures during the warm season elevate the need for air conditioning, increasing the emission of ozone precursors from regional power plants (Rosenfeld et al., 1998). Second, the rate of evaporative VOC emissions from vehicle engines, as well as natural forms of VOCs emitted by some species of trees, increase in response to rising ambient temperatures (Cardelino and Chameides, 1990). Thus, heat island formation can exacerbate regional air quality problems by increasing the rate at which ozone precursors are emitted from both human and natural sources.

The influence of urban temperatures on regional ozone formation is well documented (Rao et al., 1995, 1992; Kelly et al., 1986). In a study of temperature trends and ozone formation in large US cities, Stone (2005) found a strong positive correlation between mean temperature and the average number of high ozone days per year ($r = 0.86$; $p < 0.000$). More directly focused on the role of land use in ozone formation, Rosenfeld et al. (1998) developed a

model to assess the impact of various heat island management strategies on ozone formation in the Los Angeles, California metropolitan region. The cooling benefits of a region-wide program designed to increase the surface reflectivity of development and the extent of tree canopy cover were found to reduce the number of annual ozone exceedances by about 12% (Rosenfeld et al., 1998).

In light of the established potential for urban spatial structure to influence ozone formation through two distinct mechanisms—the quantity of precursor emissions from vehicle travel and the meteorological effects of urban heat island formation—land use planning strategies designed to reduce vehicle travel and mitigate heat island formation may prove successful in improving air quality. Little work, however, has sought to establish a relationship between urban form and air quality measured at the metropolitan scale, and land use is only rarely employed as a control measure in state air quality plans. To assess the potential significance of land use to metropolitan air quality, this paper presents a methodology to measure the influence of urban form on annual ozone exceedances, reports the results of my analysis, and concludes with a discussion of the elements of urban form found to be most closely associated with air quality.

3. Measuring urban sprawl

While sprawling, decentralized land use patterns have long been associated with a number of social and environmental problems, it is only in recent years that researchers have attempted to quantify sprawl in a meaningful way. Most of these efforts have focused only on gross measures of population density or the proportion of the urban population living a minimum radius from the central business district. For example, in seeking to gauge the influence of sprawl on affordable housing and the distribution of urban population by race, Kahn (2001) quantified sprawl through measuring the proportion of employment situated more than 10 miles from the downtown district of large US cities. In a study of sprawl, land use values, and metropolitan governance, Pendall (1999) operationalized sprawl as the change in a city’s population density over time. As argued by Ewing et al. (2003), however, sprawl is a complex phenomenon that can only be effectively measured through quantifying several dimensions of urban form. In light of this observation, Ewing et al. developed a composite index of sprawl based on four measures of urban form: centeredness, connectivity, density, and land use mix. As this index has been found to be a useful metric for measuring sprawl and a strong predictor of travel behavior, this present study employs the Ewing et al. data for 45 major US cities to assess the influence of sprawl on annual ozone exceedances.

The literature on sprawl has identified numerous spatial symptoms of the phenomenon. Four spatial dimensions commonly associated with sprawl include low density development, segregation of distinct land use types, growth

in the absence of definable centers, and a lack of physical connectivity between new areas of growth (Burchell et al., 2005). Unable to capture each of these spatial dimensions within a single indicator of urban form, Ewing et al. (2003) employed data from the US Census Bureau, the American Housing Survey, and US Department of Agriculture, among other sources, to develop an integrated measure of urban sprawl through principal components analysis. Each of the four dimensions measured and the methods employed to do so are briefly outlined in Table 1.

Through the methods outlined in Table 1, a composite sprawl score was derived by Ewing et al. (2003) for 83 major US metropolitan regions and was found to exhibit statistically significant relationship with four measures of vehicle travel behavior. These measures included person hours of travel, person miles of travel, hours of delay, and number of annual vehicle fatalities. In addition, it should be noted that the sprawl index was also found to exhibit a significant positive relationship with a measure of metropolitan air quality, the fourth highest maximum 8-h ozone level measured in 2000. While supportive of the hypothesis explored herein—that decentralized development is conducive to ozone formation—the Ewing et al. analysis of urban form and air quality is deficient on a number of grounds. A brief discussion of these limitations follows to clearly differentiate this present study from previous work.

As discussed above, due to the fact that ozone formation is highly sensitive to meteorological variables, there is a critical need to control for regional climate when comparing ozone levels recorded in different regions of the country. Because a disproportionate number of cities ranking highly in terms of the Ewing et al. (2003) index

are situated in southern latitudes of the country, the association found between the sprawl rankings and ozone levels may be a product of local meteorology, which was not incorporated into the analysis. On a similar note, the sensitivity of ozone to annual fluctuations in meteorology requires that any analysis of air quality trends focus on more than a single year of observations to avoid analysis of an anomalous year in the meteorological record. That is, ozone must be measured as a trend rather than as an occurrence. And, finally, the fourth highest 8-h ozone level does not capture the frequency of high ozone days during the course of a year, which is of more relevance to overall air quality than the magnitude of a single event.

In light of these observations, there remains a need to investigate further the relationship between urban land use patterns and regional air quality in large US cities. To do so, this study integrates data on annual ozone exceedances, precursor emissions, regional climate, and regional population size with the urban form measures developed by Ewing et al. (2003) for 45 of the 50 most populous metropolitan regions in the US. What follows is a discussion of a methodology to isolate the influence of urban sprawl on ozone formation while controlling for theoretically important meteorological and population factors and the presentation of study results and conclusions.

4. Research approach

To investigate the hypothesized influence of regional development patterns on annual ozone formation, data were obtained from several sources. The sprawl index and constituent land use data published by Ewing et al. (2002,

Table 1
Development of a sprawl index (Ewing et al., 2003)

Variable ^a	Derivation
Centeredness	The centeredness variable is a measure of the degree of mono or polycentrism within a metropolitan region and is based on three indicators: a density gradient, the percentage of the metropolitan population within a fixed radius of the central business district, and the number of population centers as defined by proximity of census tracts to regional density maxima. The composite measure is derived through principal components analysis with data obtained from the 2000 Census and the Claritas Corporation
Connectivity	The connectivity variable is a measure of the density of the street network and was based on the average block size and the percentage of blocks less than approximately 500 ft on a side (consistent with the dimension of a traditional urban block). As block size increases, the number of street intersections per unit of area decreases, which serves as an indicator of street network density. This measure was computed through principal components analysis with data obtained from US Census TIGER files
Density	A composite measure of population density was computed from four data sources, including the 2000 Census, the USDA Natural Resource Inventory, the American Housing Survey, and the Claritas Corporation. A single density factor was derived through principal components analysis incorporating measures of gross population density, the proportion of metropolitan populations living at very low or very high densities, and the proximity of census tracts to urban centers
Land use mix	Three elements of land use mix were integrated into a single, composite measure through principal components analysis. These elements include the ratio of jobs to population, the diversity of land uses, and the accessibility of residential uses to non-residential uses at the level of the transportation analysis zone and within a one mile radius. The data used to develop the mix factor were obtained from the American Household Survey and the Census Transportation Planning Package
Sprawl index	A composite measure of urban compactness or sprawl was developed through an integration of these four urban form factors through principal components analysis

^a Each urban form attribute is reported on a scale with a mean value of 100 and a standard deviation of 25 (across the 83 regions included in the Ewing et al. study (2003)). Higher values of the scores for centeredness, connectivity, density, and land use mix reflect higher intensities of these attributes. Note that the Ewing et al. sprawl index scales in a negative direction (i.e., higher scores denote lower levels of sprawl) and has been modified in this study to scale in a positive direction (i.e., higher scores denote higher levels of sprawl) for ease of interpretation.

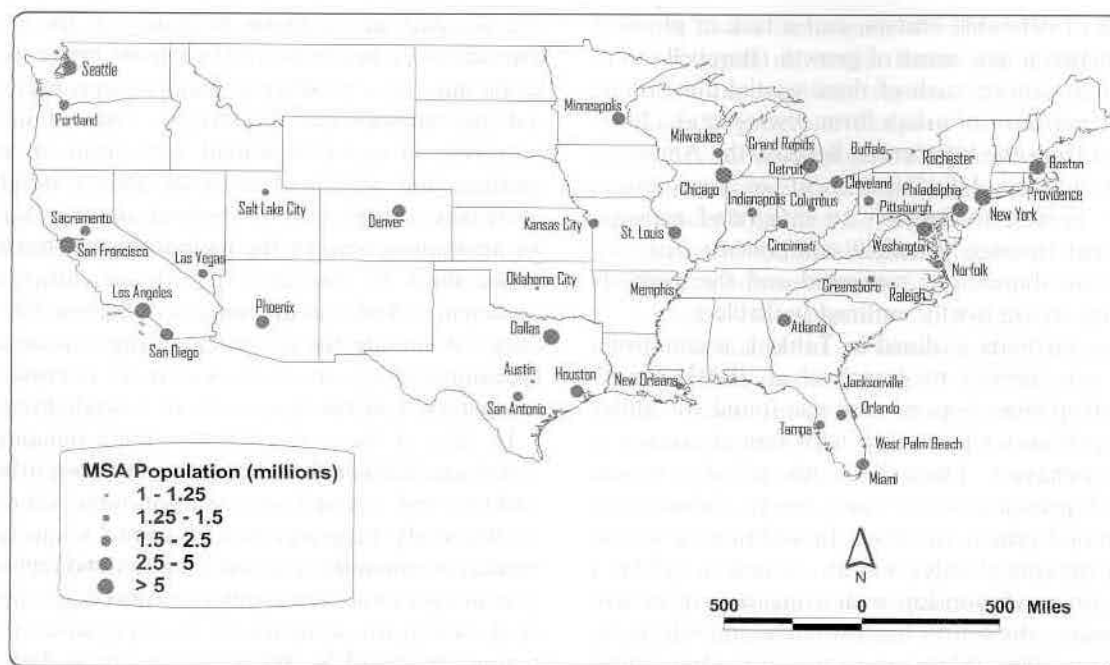


Fig. 1. Metropolitan statistical areas (only primary cities are listed).

2003) were used as the basis for quantifying urban spatial structure within major US metropolitan areas. As the published data includes statistics for 45 of the largest 50 metropolitan statistical areas (MSAs) by population, these regions, presented in Fig. 1, serve as the study areas for this analysis.

In addition to measures of urban spatial structure, data are needed on regional ozone precursor emissions and meteorological conditions for each year of the study. Through the National Emissions Inventory (NEI), the US Environmental Protection Agency (EPA) compiles data on aggregate annual emissions of all nationally regulated air pollutants at the county level. For this study, data on annual emissions of NO_x and VOC were obtained for all counties comprising each of the 45 MSAs included in the study and averaged between 1990 and 2002, which is the most recent year for which complete emissions data were available from the NEI.² The total average annual tonnage of NO_x and VOC emissions at the MSA level were summed to create a single measure of regional ozone precursor emissions.

As noted above, of most relevance to urban public health is the number of exceedances of the national standard for ozone rather than the maximum ambient concentration experienced during the course of a year. In its annual *Air Trends* report on national air quality, the EPA reports the number of days per year for which the national ambient air quality standard (NAAQS) for ozone was exceeded (referred to as an ozone “exceedance”) for the 100 largest

metropolitan regions in the country.³ For this study, the number of annual ozone exceedances per region was compiled and averaged over the 13-year period between 1990 and 2002.

Because ozone formation is highly sensitive to regional meteorology, metropolitan data on annual temperature trends was obtained from the National Oceanic and Aeronautical Administration’s National Climatic Data Center for each year of the study.⁴ Used as a control variable, the average ozone season temperature for each city between 1990 and 2002 was included in the analysis to account for the effects of temperature on ozone formation rates in different climatic regions of the country. For the purposes of this analysis, the ozone season was defined as the months of May through September.

In addition to climatic variations, differences in population size among the sampled regions can be expected to influence ozone formation through its effects on precursor emissions. Naturally, as the number of vehicles and housing units increases, emissions of ozone precursors from tailpipes and regional powerplants are also expected to increase, potentially enhancing ozone formation. As a result, population size must be controlled in the analysis to effectively isolate the influence of urban form on air quality, independent of population. To account for differences in urban populations, data on MSA populations was obtained

²This study makes use of the most commonly cited standard for ozone, the 8-h NAAQS.

⁴Mean temperature data for the months of May through September were obtained from airport weather stations in each of the 45 MSAs and averaged over the period of 1990–2002.

²1990 was selected as the initial year in this study due to its concurrence with passage of the most recent amendments to the Clean Air Act.

Table 2
Descriptive statistics

Variable	Minimum	Maximum	Mean	Standard deviation
Centeredness	52	145	98.0	21.1
Connectivity	37	155	102.8	25.2
Density	74	243	103.7	27.4
Land use mix	40	141	96.9	22.2
Sprawl index	22	153	98.2	24.7
MSA population in 2000 (millions)	1.0	21.2	3.3	3.8
Mean ozone season temperature (°C)	16.6	32	22.6	3.7
Mean annual emissions of NO _x and VOC (tons)	110,662	1,733,117	392,823	331,971
Mean annual ozone exceedances (days per year)	0.5	67.1	15.1	15.6

from the 2000 US Census and incorporated into the analysis as a control variable.

Table 2 presents descriptive statistics on each of the variables incorporated into the analysis.⁵

Once constructed, the resulting database was analyzed to test the following hypotheses:

- (1) Urbanized regions characterized by high levels of sprawl are associated with a greater quantity of annual ozone precursor emissions than urbanized regions characterized by low levels of sprawl when controlling for average ozone season temperature and population size.
- (2) Urbanized regions characterized by high levels of sprawl are associated with a higher number of average annual ozone exceedances than urbanized regions characterized by low levels of sprawl when controlling for average ozone season temperature and population size.
- (3) Urbanized regions characterized by high levels of sprawl are associated with a higher number of average annual ozone exceedances than urbanized regions characterized by low levels of sprawl when controlling for average ozone season temperature and average annual emissions of NO_x and VOC.

In testing the first hypothesis, my goal is to confirm the existence of a statistical association between urban form and ozone precursor emissions. While a handful of studies has documented an association—between land use and precursor emissions within individual cities, this analysis is the first to consider a link between land use and ozone precursor emissions across multiple MSAs and over a multiyear period. Because the emission of these pollutants is expected to vary by the size and climate of metropolitan regions—variables that are assumed to be unrelated to urban form—population size and average annual temperatures will be included in each model as control variables.

⁵Statistics are reported for all regions included in Fig. 1, with the exception of San Antonio, Texas, which was found to be an influential outlier with respect to the number of annual ozone exceedances during the study period, and thus was not included in the analysis.

Separate models testing the influence of the sprawl index and each of the four constituent urban form measures—centeredness, connectivity, density, and land use mix—on emissions are specified to test the combined and independent effects of specific urban form attributes on regional air quality.

The second hypothesis addresses the direct association between urban form and ozone exceedances. Because ozone formation is sensitive to both precursor emissions and regional meteorology, a finding of a significant association between urban form and precursor emissions is not sufficient to conclude that land use has a significant influence on ozone exceedances. Population size and mean temperatures are controlled in these models to account for regional variations in travel and energy demand and in climate. In the interest of testing the hypothesis that urban form influences ozone exceedances through both emissions and non-emissions-related phenomena, mean annual emissions of NO_x and VOC are not controlled in this set of models.

A final set of models is specified to test the hypothesis that urban form influences ozone exceedances through a mechanism that is independent of its effects on ozone precursor emissions.

To do so, mean annual precursor emissions will be incorporated into these models, along with mean ozone season temperature, as a control variable.⁶ When controlling for precursor emissions, the finding of a significant coefficient on one or more of the urban form attributes will support the hypothesis that the spatial structure of cities influences ozone exceedances through a non-emissions-related mechanism, such as urban heat island formation.

5. Results

To investigate the hypothesized relationship between urban form and ozone precursor emissions, ordinary least squares was employed to derive model coefficients and measures of goodness of fit. In each of five models, a

⁶Because population size and ozone precursor emissions are highly correlated at the MSA level, population will not be included in these models to avoid multicollinearity.

composite (sprawl index) or individual measure of urban form is combined with population size and mean temperatures and regressed against average regional precursor emissions. The results of this modeling process are presented in Table 3.

The model results show only one of the five urban form measures to be significantly related to regional emissions of NO_x and VOC. As indicated by the standardized coefficient on the density variable, for every one standard deviation increase in density, the mean annual tonnage of ozone precursors is reduced by approximately 74,700 tons, or by about 19% of the mean level of emissions. Overall, each of the models exhibited a high level of predictive power, with R^2 values exceeding 0.90. However, much of the predictive power is a product of the population size variable, which was found to have a standardized regression coefficient five times larger than that for the density measure. The composite sprawl index was not found to have a significant association with metropolitan precursor emissions over the 13-year study period.

In sum, these findings suggest a significant and moderately strong association between one urban form attribute—population density—and ozone precursor emissions in large US metros, providing support for the study's first hypothesis. However, due to the fact that ozone is sensitive to both precursor emissions and regional meteorology, any association found between urban form and emissions may not be found to exist between urban form and ozone exceedances. As a result, a second set of models was specified to assess the relationship between the spatial

structure of large metros and ozone exceedances directly. The results of this modeling process are reported in Table 4.

As indicated by the model results, several attributes of urban form, including the composite sprawl index, were found to have a statistically significant association with mean annual ozone exceedances when controlling for MSA population size and mean ozone season temperature. With a standardized regression coefficient of -1.04 , density was again found to be the most powerful urban form predictor, indicating that each standard deviation increase in density is associated with a reduction of approximately 16 high ozone days per year, on average, in large US metros. Also found to be a significant and strong predictor of ozone exceedances was the connectivity measure. With a standardized regression coefficient of -0.35 , each standard deviation increase in connectivity is associated with a reduction of approximately 5.5 high ozone days per year.

In combination, the four urban form attributes were found to influence ozone exceedances in the form of the composite sprawl index. With a significant standardized coefficient of 0.42 , each standard deviation increase in the sprawl index is associated with an increase of approximately 6.6 high ozone days in large US metros. Overall, the predictive power of the models for density, connectivity, and the sprawl index were found to be moderate to strong, with adjusted R^2 ranging from 0.47 to 0.63 .

These results suggest that the spatial structure of large US metros has a strong association with the average number of annual ozone exceedances, one that appears to

Table 3
Regression results for urban form and mean annual emissions of NO_x and VOC (tons)^a

Variable	B-coefficient	Standardized coefficient	Significance	Model adj. R^2
Centeredness	-708.0	-0.045	0.370	0.91
Connectivity	-463.1	-0.035	0.508	0.91
Density	-2,732.3	-0.225	0.020**	0.92
Land use mix	222.4	0.015	0.778	0.91
Sprawl index	735.1	0.055	0.300	0.91

^aMSA population size and mean ozone season temperatures are incorporated into each model as control variables. Mean annual precursor emissions are computed by summing the average emissions of NO_x and VOC at the MSA level between 1990 and 2002.

**Significant at the $p < 0.05$ level.

Table 4
Regression results for urban form and mean annual ozone exceedances^a

Variable	B-coefficient	Standardized coefficient	Significance	Model adj. R^2
Centeredness	-0.166	-0.225	0.080*	0.41
Connectivity	-0.216	-0.350	0.008**	0.47
Density	-0.592	-1.04	0.000**	0.63
Land use mix	-0.109	-0.156	0.249	0.39
Sprawl index	0.266	0.422	0.001**	0.52

^aMSA population size and mean ozone season temperature are incorporated into each model as control variables.

*Significant at the $p < 0.10$ level.

**Significant at the $p < 0.05$ level.

Table 5

Regression results for urban form and mean annual ozone exceedances (emissions controlled)^a

Variable	B-coefficient	Standardized coefficient	Significance	Model adj. R^2
Centeredness	−0.144	−0.195	0.123	0.42
Connectivity	−0.187	−0.302	0.017**	0.47
Density	−0.286	−0.504	0.007**	0.49
Land use mix	−0.112	−0.160	0.227	0.41
Sprawl index	0.226	0.359	0.004**	0.50

^aMean annual emissions of NO_x and VOC and mean ozone season temperatures are included in each model as control variables.**Significant at the $p < 0.05$ level.

be stronger than the association between urban form and precursor emissions. Taken together, these first two sets of models support the hypothesis that urban form drives ozone formation through both an “emissions-based” and a “non-emissions-based” mechanism, and further raises the possibility that the non-emissions-based mechanism or set of mechanisms is the more powerful influence on ozone formation. To confirm the existence of a non-emissions-based effect of urban form on ozone formation, a final set of models was specified to regress the urban form attributes against mean annual ozone exceedances while controlling for mean regional temperature and precursor emissions. The results of these models are presented in Table 5.

The significant model coefficients reported in Table 5 for several of the urban form attributes confirms the existence of an association between land use and regional air quality that is independent of the influence of land use on precursor emissions. Having controlled for mean annual emissions of NO_x and VOC and mean ozone season temperatures, the density, connectivity, and sprawl index measures were each found to be statistically significant at the $p < 0.05$ level, supporting the hypothesis of a non-emissions-based effect of urban form on ozone formation. As indicated by the standardized coefficient of 0.36, each standard deviation increase in the sprawl index is associated with an average increase of 5.6 high ozone days per year. A one standard deviation increase in density is associated with a reduction of about 8 high ozone days, and a one standard deviation increase in connectivity is associated with a reduction of 4.7 high ozone days per year. While the density measure was found to have the greatest magnitude of effect on ozone exceedances, the sprawl index model exhibited the most predictive power, with an adjusted R^2 of 0.50.

6. Discussion and conclusions

The results of this analysis demonstrate a significant association between regional land use patterns and air quality in the country's largest metropolitan areas. While urban form was found to have only moderate effects on precursor emissions measured at the MSA level, the spatial attributes of density and connectivity were found to have a

statistically significant ($p < 0.05$) and reasonably strong association with ozone exceedances when controlling for mean ozone season temperature, population size, and precursor emissions. The composite measure of sprawl also was found to have a significant and strong association with mean annual ozone exceedances when controlling for mean ozone season temperature, population size, and precursor emissions.

The potential for land use strategies to reduce ozone exceedances is most clearly illustrated through a comparison of the most sprawling and most compact cities included in the study. Fig. 2 presents the average urban form attributes and expected mean ozone exceedances between 1990 and 2002 for the metropolitan regions ranking in the 75th percentile and greater (sprawling) of the sprawl index to those ranking in the 25th percentile and lower (compact). The expected ozone exceedances are reflective of the modeled number of annual exceedances for each set of regions, controlling for population size and average ozone season temperature. For sprawling cities, levels of centeredness, connectivity, density, and land use mix are, on average, about one-third lower than in compact cities, and the average number of high ozone days per year is about 62% higher.

The results of this study support the hypothesis that urban form influences regional ozone formation through both an emissions-based and non-emissions-based mechanism. The finding of a significant association between population density and emissions of NO_x and VOC suggests that urban spatial structure plays a role in ozone formation through its effects on ozone precursor emissions. However, the influence of urban form on precursor emissions may account for only one of several mechanisms through which land use impacts regional air quality. While insufficient data are available through this study to identify the precise non-emissions-related mechanisms through which urban form influences ozone formation, at least two additional phenomena could explain the results presented herein. These include enhancement of the urban heat island effect and the geographic distribution of ozone monitors in highly decentralized cities.

As discussed in the preceding review of the literature on urban form and regional climate, urban heat island formation can significantly elevate surface and air

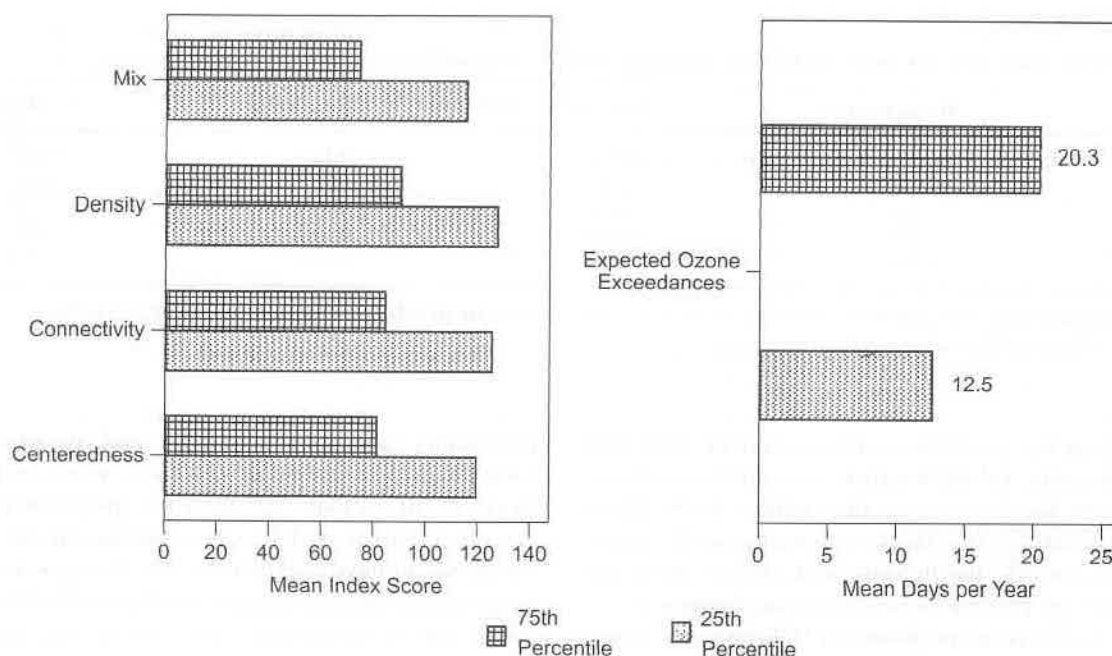


Fig. 2. Expected ozone exceedances for metropolitan regions ranking in the highest and lowest quartiles of sprawl index. *Note:* Expected ozone exceedances are based on model results reported in Table 4.

temperatures in large cities, with direct implications for air quality. Previous work on ozone formation in large US cities has found mean regional temperatures to exhibit a stronger association with ozone exceedances than the emission of ozone precursors (Stone, 2005). In addition, analyses of the effects of city size and neighborhood scale design on heat island formation have supported the hypothesis that lower density, decentralized patterns of urbanization generate more surface heat than higher density forms (Stone and Rodgers, 2001; Price, 1979). If true within the regions included in this analysis, cities characterized by a high degree of sprawl may experience a greater number of ozone exceedances than more compact cities due to enhanced heat island formation. Future work will investigate this hypothesis further through the measurement of heat island formation within the metropolitan regions included in this study.

The finding of a positive association between sprawl and ozone exceedances may also be an artifact of the geographic distribution of air quality monitors in large, decentralized cities. For cities characterized by a large geographic extent, the average distance between an ozone monitor and the region's central business district may be greater than that within the most compact cities. If true, advecting plumes of ozone precursor emissions may travel farther before reaching the most remote regional air quality monitors, potentially enabling higher ozone concentrations to form and be detected than within more spatially compact monitoring networks. While a potential artifact of the monitoring network, such a phenomenon would also be reflective of the fact that a greater proportion of the metro population in decentralized regions lives in these

downwind locales and is thus subject to higher ozone exposures.

While the precise linkages between urban form and enhanced ozone formation will require further study, the finding of a strong association between the spatial structure of cities and regional air quality raises the question of whether environmental managers should give more serious consideration to land use strategies in developing air quality control plans. Two elements of urban form in particular—the density and connectivity of development—were found to be significantly related to precursor emissions and ozone formation and are directly responsive to land development policies instituted at the local or regional levels of government. In particular, “urban growth boundaries” and “form-based codes” are two planning strategies increasingly adopted in US cities that are worthy of consideration as long-term, regional scale approaches to reducing total emissions and ozone formation through urban form.

While long used outside of the United States as a means of protecting natural areas and limiting the peripheral spread of urban zones, urban growth boundaries have only in recent decades been instituted by a handful of state and municipal governments as a means to control growth. Growth boundaries establish circumferential zones beyond which urban population and employment densities are not permitted. Two sets of planning policies are typically employed to establish such zones. First, large minimum lot sizes (e.g., 40 acres) and use restrictions are established through zoning ordinances beyond the zone in which urban development densities are to be allowed. Second, limitations on the extension of public infrastructure and services,

such as sewer lines and fire protection, are often coterminous with urban growth boundaries, serving to further discourage urban development beyond the designated zone. Today, Oregon, Washington, and Tennessee require cities to establish growth boundaries, and much of the limited evidence available suggests that such boundaries are effective in limiting peripheral growth and promoting density (Song and Knaap, 2004; Nelson, 1994), with potential benefits for reduced vehicle travel, greenspace preservation, and air quality.

A second land use planning strategy that may prove effective over the long term in improving air quality is the replacement of traditional zoning ordinances with “form-based” codes. Form-based codes are designed to place an emphasis on the physical dimensions of development, such as the size, placement, and integration of buildings and streets, rather than on the specific uses and densities permitted in a particular zone. The rationale for form-based codes is rooted in the observation that, while the use of buildings and parcels tends to change over time, the form of the built environment changes only very slowly, and it is the physical form of development that is of most significance to community interests, including environmental quality. One advantage of form-based codes related to this present analysis is the ability for municipalities to improve vehicle and pedestrian connectivity within and between development zones. While traditional zoning often permits developers to design and organize streets within defined block sizes, form-based codes more strictly control street, sidewalk, and pedestrian network designs, while permitting builders more flexibility in the uses and densities of the buildings that are constructed within the network. As higher levels of street connectivity have been associated with lower levels of vehicle travel and emissions, form-based codes designed to emphasize network connectivity could yield benefits to regional air quality.

In closing, it is important to note that more compact development patterns alone will not be sufficient to fully address our urban air quality problems. Technological advances in emissions controls have proven highly effective in reducing tailpipe emissions over the last several decades, and emerging technologies, such as hybrid vehicles and alternative fuels, are expected to continue these reductions. The significance of land use-oriented approaches to air quality management lies in the potential for these strategies to limit the dramatic annual growth in national vehicle miles of travel, which has greatly undermined the benefits of per mile tailpipe reductions since the 1960s, as well as in addressing the local-scale meteorological drivers of air pollution. Undoubtedly, the most effective approach to air quality management would seek to achieve a balance between stack and tailpipe emissions reductions with reductions in emissions-producing activities. To this end, land use strategies designed to promote more compact urban forms are needed to complement technological emissions controls as a means of addressing the root bases of urban air quality problems.

References

- Apogee Research, Inc., 1998. The effects of urban form on travel and emissions: a review and synthesis of the literature. Draft report prepared for the United States Environmental Protection Agency, Washington, DC, HBIX Reference C611-005.
- Bornstein, R., LeRoy, M., 1990. Urban barrier effects on convective and frontal thunderstorms. Extended abstracts, Fourth Conference on Mesoscale Processes, American Meteorological Society, Boulder, CO, pp. 120–121.
- Brindle, R., 1994. Lies, damned lies and ‘automobile dependence’—some hyperbolic reflections. In: Proceedings of the 1994 Australian Transport Research Forum, Melbourne, Australia, pp. 117–131.
- Burchell, R., Downs, A., Mukherji, S., McCann, B., 2005. *Sprawl Costs: Economic Impacts of Unchecked Development*. Island Press, Washington, DC.
- Burchell, R., Lowenstein, G., Dophin, W., et al., 2002. *Costs of Sprawl—2000*. Transportation Research Board, National Academy Press, Washington, DC.
- Cambridge Systematics, Inc., Parsons Brinkerhoff Quade and Douglas, Inc., 1992. Making the Land Use, Transportation, Air Quality Connection: Analysis of Alternatives. 1000 Friends of Oregon, Portland, OR.
- Cardelino, C., Chameides, W., 1990. Natural hydrocarbons, urbanization, and urban ozone. *Journal of Geophysical Research* 95 (D9), 13971–13979.
- Cervero, R., Goham, R., 1995. Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association* 61 (2), 210–225.
- Diem, J., Mote, T., 2005. Interdecadal changes in summer precipitation in the southeastern United States: evidence of possible urban effects near Atlanta, Georgia. *Journal of Applied Meteorology* 44, 717–730.
- Downs, A., 1992. *Stuck in Traffic: Coping with Peak-hour Traffic Congestion*. The Brookings Institution, Washington, DC.
- Ewing, R., Pendall, R., Chen, D., 2002. *Measuring Sprawl and its Impact*. Smart Growth America, Washington, DC.
- Ewing, R., Pendall, R., Chen, D., 2003. Measuring sprawl and its transportation impacts. *Transportation Research Record* 1831, 175–183.
- Frank, L., Pivo, G., 1994. Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking. *Transportation Research Record* 1466, 44–52.
- Frank, L., Stone Jr., B., Bachman, W., 2000. Linking land use with household vehicle emissions in the central Puget Sound: methodological framework and findings. *Transportation Research D* 5, 173–196.
- Frumkin, H., Frank, L., Jackson, R., 2004. *Urban Sprawl and Public Health*. Island Press, Washington DC.
- Gordon, P., Richardson, H., 1989. Gasoline consumption and cities: a reply. *Journal of the American Planning Association* 55 (3), 342–345.
- Heimlich, R., Anderson, W., 2001. Development at the urban fringe and beyond: Impacts on agriculture and rural land. *Agriculture Economic Report (AER803)*, United States Department of Agriculture.
- Johnston, R., Rodier, C., Choy, M., Abraham, J., 2000. Air quality impacts of regional land use policies. Prepared for US Environmental Protection Agency, Urban and Economic Development Division, Washington, DC.
- Kahn, M., 2001. Does sprawl reduce the black/white housing consumption gap? *Housing Policy Debate* 12 (1), 77–86.
- Kelly, N., Ferman, M., Wolff, G., 1986. The chemical and meteorological conditions associated with high and low ozone concentrations in southeastern Michigan and nearby areas of Ontario. *Journal of the Air Pollution Control Association* 36, 150–158.
- Nelson, A.C., 1994. Oregon’s urban growth boundary policy as a landmark planning tool. In: Abbott, C., Howe, D., Adler, S. (Eds.), *Planning the Oregon Way: A Twenty Year Evaluation*. OSU Press, Corvallis, OR, pp. 25–47.
- Newman, P., Kenworthy, J., 1989. Gasoline consumption and cities: a comparison of US cities with a global survey and its implication. *Journal of the American Planning Association* 55 (1), 24–37.

- Oke, T., 1987. *Boundary Layer Climates*. Routledge, New York.
- Pendall, R., 1999. Do land use controls cause sprawl? *Environment and Planning B* 26 (4), 555–571.
- Price, J., 1979. Notes and correspondence: assessment of the urban heat island effect through the use of satellite data. *Monthly Weather Review* 107, 1554–1557.
- Pucher, J., Lefevre, C., 1996. *The Urban Transport Crisis in Europe and North America*. MacMillan Press Limited, London.
- Rao, T., Sistia, G., Henry, R., 1992. Statistical analysis of trends in urban ozone air quality. *Journal of the Air and Waste Management Association* 42, 1204–1211.
- Rao, T., Zalewsky, E., Zurbenko, I., 1995. Determining temporal and spatial variations in ozone air quality. *Journal of the Air and Waste Management Association* 45, 57–61.
- Real Estate Research Corporation, 1974. *The Costs of Sprawl: Environmental and Economic Costs of Alternative Residential Development Patterns at the Urban Fringe*. US Government Printing Office, Washington, DC.
- Rosenfeld, A., Akbari, H., Romm, J., Pomerantz, M., 1998. Cool communities: strategies for heat island mitigation and smog reduction. *Energy and Buildings* 28, 51–62.
- Rozoff, C., Cotton, W., Adegoke, J., 2003. Simulation of St. Louis, Missouri, land use impacts on thunderstorms. *Journal of Applied Meteorology* 42, 716–738.
- Shepherd, J., Pierce, H., Negri, A., 2002. Rainfall modification by major urban areas: observations from spaceborne rain radar on the TRMM satellite. *Journal of Applied Meteorology* 41, 689–701.
- Song, Y., Knaap, G., 2004. Measuring urban form: is Portland winning the war on sprawl? *Journal of the American Planning Association* 70 (2), 210–225.
- Stone, B., 2005. Urban heat and air pollution: an emerging role for planners in the climate change debate. *Journal of the American Planning Association* 71 (1), 13–25.
- Stone, B., Bullen, J., 2006. Urban form and watershed management: how zoning influences residential stormwater volumes. *Environment and Planning B* 33, 21–37.
- Stone, B., Rodgers, M., 2001. Urban form and thermal efficiency: how the design of cities influences the urban heat island effect. *Journal of the American Planning Association* 67 (2), 186–198.
- Transportation Research Board, 1995. *Expanding Metropolitan Highways: Implications for Air Quality and Energy Use*. Special report 245. Author, Washington, DC.



ELSEVIER

Journal of Environmental Management 86 (2008) 699–711

Journal of
**Environmental
Management**

www.elsevier.com/locate/jenvman

Assessing the US Clean Water Act 303(d) listing process for determining impairment of a waterbody

Arturo A. Keller*, Lindsey Cavallaro

Bren School, University of California, Santa Barbara, CA, USA

Received 9 February 2006; received in revised form 13 September 2006; accepted 14 December 2006

Available online 31 January 2007

Abstract

This study evaluated the US Clean Water Act (CWA) Section 303(d) listing and delisting processes, based on historical and current federal and state guidelines, to determine whether there are regional differences in water quality assessment criteria used by various states to determine impairment of a waterbody for inclusion in the 303(d) list. A review of almost 50 total maximum daily load (TMDL) and delisting documents revealed that the basis for listing or delisting a waterbody varies considerably and that, in many cases, determination of impairment was based on insufficient water quality information. Historical USEPA guidance on the 303(d) listing and delisting processes has been generally broad, resulting in wide interpretation of the assessment criteria by various states. This has led to unclear or conflicting listing methodologies among states, leading to inconsistencies in impairment determination. Common problems include inconsistent data quality and quantity, differences in frequency of monitoring, variable interpretation of narrative water quality standards, and differences in specificity of implementation and monitoring plans, resulting in significant difference in the basis for listing and delisting waterbodies. In response, several states have taken the initiative to provide much more specific guidance for their internal agencies. Listing and delisting criteria are generally clearer at the state level, but the development of differing state guidance documents has resulted in diversity in the development of the 303(d) lists and in the process of delisting a waterbody. While state guidelines are better able to address local considerations, such as variations in climate, landuse, and water quality objectives, as well as social and economic preferences, the variation in listing criteria has led to inconsistencies across state boundaries in the levels of attainment of national water quality objectives. For stakeholders that participate in the 303(d) listing process within a particular state, these types of discrepancies may not have a significant impact. However, these inconsistencies can lead to confusion for some stakeholders who participate in the process in multiple states, and must deal with differing and sometimes conflicting requirements depending on the location of their facilities.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Water quality; Water policy; TMDL; Impairment; Listing; Delisting; 303(d)

1. Introduction

Since the implementation of the Clean Water Act (CWA) in 1972, the water quality of US waterbodies has improved significantly. However, the CWA Section 303(d) list of impaired waters for 2002 included 59,783 impairments for the 34,225 listed US waterbodies. Over half of these impairments are caused by one of five major pollutants: pathogens, metals, excess nutrients, excessive sediment, or organic enrichment. According to data collected by the US Environmental Protection Agency

(USEPA, 2006a), the spatial distribution of impaired waterbodies covers most of the US, with the exception of some of the very dry areas in the southwestern portion of the country and some sparsely populated areas in northwestern US. The distribution of impairments across the country demonstrates that poor water quality is endemic in many of the nation's waters.

Assessment of impaired waterbodies for Section 303(d) listing is primarily carried out by states and tribal nations, with oversight and final approval by USEPA. Since the initial 303(d) listing, guidance provided to the states by USEPA was very general with no specific parameters for assessing impairment, a number of waterbodies were included in the 303(d) list with little data to support the

*Corresponding author. Tel.: +1 805 453 1822.

E-mail address: keller@bren.ucsb.edu (A.A. Keller).

assessment. Additionally, listing decisions were based on criteria that varied widely across the nation. Once a waterbody is listed as impaired, the state is required to develop a total maximum daily load (TMDL) to address the impairment and implement measures to restore the waterbody to its designated use. The development of TMDLs is extremely costly, requiring a considerable amount of time and resources on the part of state agencies and many stakeholders. Therefore, the large number of waterbodies listed as impaired represents a potentially substantial national investment in the TMDL process. There have been some studies that have looked in general at the implementation issues related to the TMDL program (e.g. Freedman et al., 2004; NRC, 2001; Boyd, 2000); although to date the listing and delisting processes have not been evaluated in detail.

Although, in principle, a water quality assessment through a comparison of water quality observations to water quality objectives seems relatively straightforward, in practice there are many challenges. For example, many water quality objectives are based on surrogate variables, such as nutrients concentration or chlorophyll *a* to indicate excessive algal growth. However, because the objectives may be narrative, there is opportunity for inconsistency in the application of surrogate water quality criteria used to determine impairment. Water quality criteria and numerical targets often differ among various regulatory agencies (e.g. federal, state, local, tribal), and may even be different from one river reach to the next, complicating the assessment even further. In addition, the frequency of monitoring (i.e. annual, semi-annual, quarterly, monthly, daily, arbitrary or strategic) may play a significant role in the determination of impairment, yet is often unaddressed by regulatory agencies. Discrepancies in assessment criteria can not only make it difficult to justify a listing decision, but also make it difficult to support delisting of a waterbody once it has been placed on the 303(d) list.

While USEPA guidance on the 303(d) listing process has improved since the initiation of the CWA, the guidelines are generally broad, leaving it up to state and local agencies to resolve many facets of the assessment, such as frequency of data collection, data quality, statistical method of analysis, and interpretation of water quality criteria. Given the significant resources required to prepare TMDLs for listed waterbodies, in recent years several states have developed their own listing and delisting methodologies with additional guidance for their staff to facilitate and improve listing and delisting decisions. These differing methodologies incorporate a wide range of water quality assessment approaches and interpretations of water quality standards (WQS), and often emphasize different aspects of the listing and delisting processes, depending on the objectives of different state agencies. Additionally, a consistent listing methodology is not always applied throughout a given state. For stakeholders that participate in the 303(d) listing process within a particular state, these types of discrepancies may not have a significant impact.

However, these inconsistencies can lead to confusion for stakeholders that participate in the process in multiple states, and must deal with differing and sometimes conflicting requirements depending on the location of their facilities.

This study evaluated the Section 303(d) listing and delisting processes, based on federal and state guidelines, to establish regional differences in assessment criteria used by various states. We first reviewed the historical development of USEPA guidance documents related to the 303(d) listing and TMDL process. This was followed by a review of 41 TMDL documents, to determine whether different water quality assessment methodologies were used by states in listing decisions. Finally, we evaluated and compared various state guidance documents that go beyond USEPA guidance, to establish similarities and differences in listing and delisting water quality criteria used by states throughout the US.

2. Background

When the United States Congress passes a law, environmental or otherwise, federal agencies are required to develop rules and regulations that are consistent with the enacted law. However, state and local regulatory agencies are often responsible for the implementation of federal regulations, with guidance and oversight from federal agencies. In addition to rules imposed by federal regulations, states develop their own rules, typically in the form of permits, which are consistent with the federal regulations, if not more stringent. For instance, when the Clean Air Act Amendments (CAAA) were passed in 1990, Congress did not specify the criteria for attainment of clean air standards. Instead, USEPA had the responsibility of writing the regulations to uphold the CAAA. Each State then had to develop an individual State Implementation Plan (SIP) to ensure that regional air quality would meet the standards set by USEPA. While USEPA has final approval over SIPs, the states are ultimately responsible for implementation of the regulations. States are then left to provide guidance to their staff for implementing federal regulations. Such is the case for other environmental regulations, including those mandated by the CWA.

Federal versus state governance is a subject of much debate regarding the implementation of federal laws for a wide range of issues, such as gas distribution (Moring, 1999), illegal immigration (Rhymer, 2005), welfare (Rogers-Dillon, 1999), health care (Parmet, 1993) and education, as well many environmental regulations, including the CAAA and CWA. Because the federal government authorizes states with the power to implement and enforce federal regulations, states have a significant role in the effectiveness of environmental regulation. Advocates of local environmental regulation assert that local policymakers are better able to set policy that reflects local concerns, while opponents fear that decentralized regulation will lead to a “race to the bottom” in environmental

quality.” (Boyd, 2000). Some argue that environmental decentralization is beneficial because states are actually more stringent with environmental regulations (Sigman, 2003). Others suggest that the stringency in which environmental regulations are enforced, such as the CWA, are determined by other factors, such as budgetary and political concerns (Helland, 1998).

In cases where the Federal government has been slow to implement environmental regulation, or when guidance is too general or unclear, some states have taken the initiative to implement and enforcement federal environmental rules. For instance, the Western Governors Association (WGA), an organization representing the western states, addresses environmental policy and governance in the West in an effort to implement better methods of achieving environmental goals (WGA, 2006). The WGA has identified problems associated with implementation of federal environmental laws and attempts to strengthen the states’ role in the federal system by improving the effectiveness and efficiency of environmental regulation. For example, some of the problems related to the development of SIPs that the WGA outlines are that the CAAA emphasize process rather than results, and that overly prescriptive requirements could be barriers to environmental innovation. In addition, the excessive length of development of the SIPs and the lack of effective guidance and technical tools has lead to delays and uncertainties in the implementation process, and an inefficient use of resources (Air Quality Initiative Steering Committee, 1997). While these issues uncovered by the WGA relate to the SIP development process, they are also relevant in implementation of the CWA at a state level, particularly with regards to the 303(d) listing and TMDL process. USEPA cooperates with WGA’s efforts and provides grants to assist with the implementation and enforcement of federal environmental regulations. For example, WGA has received grant money from USEPA to assist the western states in implementing the TMDL program.

Recently, the federal government has been working with the states to develop common guidance documents, for example the consolidated assessment and listing methodology (CALM) that requires that each state establish a consistent process for listing and delisting waterbodies (USEPA, 2006b).

Anecdotal experiences with 303(d) listing and TMDL development indicated that insufficient guidance on the listing and delisting process had resulted in wide differences in the criteria applied around the nation. With this in mind, we sought to review the existing guidance documents and then using a sample of listing and delisting decisions determine the main issues that states and stakeholders face.

3. Approach

A review of historical USEPA documents on water quality-based listing was conducted to determine the specificity of guidance provided to the states regarding

the 303(d) listing process. The original Section 303(d) of the CWA asserts that states, territories and authorized tribes are required to develop a priority ranked list of water quality limited segments. States develop WQS and criteria for each waterbody, and if a state identifies an exceedance of a WQS, the waterbody must be added to the list of impaired waters. In addition, the Act indicates that the lists (denominated the “303(d) lists”) would be used to prioritize each waterbody for TMDL development. Each state is required to establish a TMDL to address impairment of waterbodies on the 303(d) list, which must be submitted to USEPA for final approval.

While the concept of the 303(d) process appears inherently simple, the initial USEPA guidelines did not provide states with specific assessment criteria to be used as a basis for listing waterbodies. This led to confusions among state agencies and discrepancies in listing methodologies. Additionally, no guidelines were provided on the criteria required to remove a waterbody from the 303(d) list. As a result, USEPA has released additional guidance documents, generally every listing cycle, in an attempt to clarify the 303(d) listing and delisting process. The following USEPA guidance documents were reviewed for this study:

- Guidance for water quality-based decisions: The TMDL process (USEPA, 1991).
- Guidance for 1994 Section 303(d) lists (USEPA, 1993).
- National clarifying guidance for 1998 State and Territory Section 303(d) listing decisions (USEPA, 1997).
- Integrated water quality monitoring and assessment report guidance (USEPA, 2001).
- EPA Review of 2002 Section 303(d) lists and guidelines for reviewing TMDLs under existing regulations issued in 1992 (USEPA, 2002a).
- Recommended framework for EPA Approval decisions on 2002 State Section 303(d) list submissions (USEPA, 2002b).
- Guidelines for reviewing TMDLs under existing regulations issued in 1992. (USEPA, 2002c).
- Guidance for 2004 assessment, listing and reporting requirements pursuant to Sections 303(d) and 305(b) of the CWA (USEPA, 2003).
- Guidance for 2006 assessment, listing, and reporting requirements pursuant to Sections 303(d), 305(b), and 314 of the CWA (USEPA, 2006c).

These guidance documents typically explain the programmatic elements and requirements of the TMDL process as it relates to the 303(d) list. However, the documents do not provide states with a specific methodology for assessing the water quality of a waterbody, particularly with regards to water quality data collection, interpretation and determination of impairment. States have resorted to developing their own methodology for collecting and evaluating water quality data to determine if

WQS are exceeded; in several cases the procedures and criteria have in the past varied even within a state agency. A more detailed review of each document is available in Keller et al. (2005).

As can be seen, the number of USEPA guidance documents is substantial, and a new guidance document has been developed for every recent listing cycle. The lengthy effort spent by USEPA in the late 1990s seeking to develop a major new “Watershed Rule” (NRC, 2001), which essentially hit a stone wall, is a major reason USEPA is still having to work through guidance to sharpen ideas on improved methods and procedures to achieve greater consistency, transparency, and quality in the way waters are listed under the 303(d)/TMDL process.

To evaluate the different methodologies used by states to determine impairment of a waterbody, 50 USEPA approved TMDL or delisting documents were reviewed from the following states: Alabama, Alaska, California, Colorado, Georgia, Iowa, Kansas, Maine, Michigan, Mississippi, New Mexico, New York, Pennsylvania, Rhode Island, South Dakota, Utah, Vermont, Washington, and West Virginia (Table 1). A more detailed summary and review of each document is available in Keller et al. (2005). These states were chosen as case studies, in part, because each has developed a sufficient number of publicly available USEPA approved TMDL documents, and, in part, to represent a cross-section of states spanning the country. Each TMDL document was reviewed to determine the listing rationale used by the state and evaluate the effectiveness of the TMDL in addressing impairment of the waterbody. From this analysis, differences within a state and between states were assessed, with regards to the application of listing criteria to support an impairment decision.

Finally, an extensive review of the most recent USEPA and state guidance documents was conducted to compare and contrast the specificity of guidance between USEPA and a number of states that have taken a proactive role in developing documents to assist their staff in the listing and delisting processes. Guidance documents were evaluated from California (California SWRCB, 2004), Colorado (Colorado DPHE, 2005), Iowa (Iowa DNR, 2003), Kansas (Kansas DHE, 2003), Minnesota (Minnesota PCA, 2004), New York (New York DEC, 2002), and Washington (Washington DE, 2006). In many cases, the state guidance documents detail the specific aspects of the 303(d) listing process that USEPA has not sufficiently addressed.

4. Results and discussion

Although guidance regarding water quality assessment and impairment determination is generally more specific at the state level, evaluation of the TMDL case studies demonstrate that, while many states provide more detailed guidance than USEPA, the application of particular criteria to justify impairment decisions are highly inconsistent throughout the country. Not only did listing

rationale differ significantly among states, but many states did not apply an internally consistent methodology for determining impairment of waterbodies. Most states did not utilize a uniform approach for data collection, sampling, analysis, or quality assurance, resulting in discrepancies in impairment determination. Frequently, listing decisions were based on insufficient data, and several TMDL documents lacked adequate implementation plans to achieve water quality goals for the waterbody. In several cases, the criteria for delisting were not explicit in the TMDL document. –

4.1. Water quality data collection and analysis

Each state is required to develop a 303(d) listing methodology, with guidance and final approval from USEPA, which details how water quality data will be used to determine impairment of a waterbody. The methodology must disclose specific details on data sources, quality control assurance, sampling methods and analysis, and justification of analytical approach. The guidelines for water quality data collection and quality, criteria for credible sources, and requirements for data age vary significantly across states, which has led to many inconsistencies in the listing and delisting process.

Clearly, obtaining sufficient data to make a decision regarding impairment of a waterbody is one of the most important elements of the listing or delisting processes. If a state can prevent unnecessary listing by using current and accurate water quality data, it will avoid the costly and time-consuming development of a TMDL. Many of the problems with associated with inadequate water quality data collection, quality assurance, and analyses are symptomatic of insufficient guidance. Recent guidelines provided by USEPA and the states attempts to clarify some of the major issues associated with water quality data, to varying degrees.

The 2004 USEPA guidance document (USEPA, 2003) suggests that states consider data from a wide range of sources, including but not limited to: prior listing reports, recent CWA 319 non-point source assessment, ambient water quality data, dilution calculations or predictive models, water quality management plans, Superfund Records of Decision, Safe Drinking Water Act source water assessments, and volunteer monitoring networks. While USEPA suggests that states consider data from a range of sources, it requires that states only use credible data for water quality assessment, which is any data consistent with a quality assurance or quality control program. The states generally accept data from a variety of sources, but each data has a different method of data quality assurance and quality control to ensure that water quality data is credible.

The TMDL case studies, however, demonstrate that although each state requires water quality data to meet quality control requirements, significant disparities exist in the way states address data quality assurance. Data quality

Table 1
TMDL or delisting documents reviewed^a

	State	Year	Waterbody	Impairment
1	Alabama	2002	Big Nance Creek	Fecal coliform
2	Alabama	2002	Big Nance Creek	Low dissolved oxygen/organic loading ammonia
3	Alaska	1995	Eagle River	Metals
4	Alaska	2000	Duck Creek, Mendenhall Valley	Fecal coliform
5	Alaska	2004	Fish Creek in Anchorage	Fecal coliform
6	California	1998	Garcia River	Sediment
7	California	1998	San Diego Creek and Newport Bay	Nutrients
8	California	2000	Navarro River	Temperature and sediment
9	California	2003	Santa Clara River	Nitrogen compounds
10	California	2004	Waddell Creek, Santa Cruz County	Nutrients delisting
11	Colorado	2000	Big Thompson River, Little Thompson River; South Platte Basin	Fecal coliform, determination of inapplicability
12	Colorado	2000	Cripple Creek, Arequa Gulch/Squaw Creek	Determination of TMDL inapplicability
13	Colorado	2000	Big Springs Creek, Rio Grande River Basin	Determination of TMDL inapplicability
14	Colorado	2000	Unnamed tributary to Willow Creek	Ammonia
15	Colorado	2000	Straight Creek, Summit County	Sediments
16	Colorado	2002	Fruitgrowers Reservoir	Determination of TMDL inapplicability
17	Colorado	2002	Upper Animas River Basin	Metals (Al, Cd, Cu, Fe, Mn, Pb, Zn) and pH
18	Colorado	2002	Little James Creek, Boulder County	Metals (Cd, Fe, Mn, Zn) and pH
19	Colorado	2003	N Fork Cache La Poudre River, Larimer City	Sediment
20	Georgia	2001	Middle/Lower Savannah River	Total mercury
21	Georgia	2000	Stekoa Creek Watershed	Sediment
22	Indiana	2004	West Fork White River	<i>E. coli</i>
23	Iowa	2000	Rock Creek, Clinton County	Ammonia and NO _x
24	Iowa	2000	Corydon Reservoir, Wayne County	Atrazine
25	Kansas	2004	Neosho Basin, Tar Creek	Lead, cadmium, and zinc
26	Kansas	2004	Neosho Basin, Olpe City Lake	Eutrophication
27	Maine	2003	Goosefare Brook	Metals (Cd, Cr, Cu, Fe, Ni, Pb, Zn)
28	Maine	2004	Annabessacook Lake	Phosphorous
29	Michigan	2002	Carrier Creek, Eaton County	Biota
30	Michigan	2003	Grand River, Jackson County	Dissolved oxygen
31	Michigan	2003	Grand River, Jackson County	Biota
32	Michigan	2003	Bean Creek, Lenawee and Hillsdale Counties	<i>Escherichia coli</i>
33	Mississippi	2000	Pearl River	Fecal coliform
34	New Mexico	2001	Silver Creek	Conductivity delisting rationale
35	New Mexico	2001	Whitewater Creek	Metals delisting rationale
36	Pennsylvania	1999	Cheat River, Fayette County	Chlordane
37	Pennsylvania	1999	Susquehanna River	PCBs
38	South Dakota	1999	Hiddenwood Watershed	Sediment and total phosphorous
39	Utah	2003	Beaver River Watershed	Total phosphorus, noxious aquatic plants, riparian habitat alteration, dissolved oxygen, and temperature
40	Utah	2002	Ken's Lake	Temperature
41	Utah	2003	Mill Creek, Grand County	Temperature, TDS
42	Utah	2003	Browne Lake	Dissolved oxygen and phosphorus
43	Vermont/NY	2002	Lake Champlain	Phosphorus
44	Washington	2000	Lower Nooksack River Basin	Bacteria
45	Washington	2001	Granger Drain	Fecal coliform bacteria
46	Washington	2002	Wind River Watershed	Temperature
47	Washington	2003	South Prairie Creek	Bacteria and temperature
48	Washington	2004	Upper Chehalis River	Fecal coliform bacteria
49	West Virginia	2002	Ten Mile Creek	Al, Fe
50	Rhode Island	2004	Chickasheen Brook, Barber and Yawgoo Ponds	Total phosphorous

^a All these documents are available at <http://www.epa.gov/owow/tmdl/>.

is not typically assessed within the TMDL documents, making it difficult to establish the validity of an impairment determination. For instance, some states, such as Alaska, Maine, and Utah, have in some instances relied on water quality data provided by volunteer groups without sufficient assurance of data quality. While water quality

data collected by volunteer groups may be valuable, the validity of a listing decision can be called into question if the data does not meet a quality control standard. Some states, such as Minnesota, have clearly defined guidelines for data collected by volunteer groups (Minnesota PCA, 2003).

Another common problem evident in the TMDL case studies is that listing decisions were often based on little or sometimes no water quality data. For example, Alabama listed Big Nance Creek in 1992 for exceeding the fecal coliform WQS based on a single exceedance in the water quality data (Alabama DEM, 2002). Duck Creek in Alaska was also listed for non-attainment the State's fecal coliform WQS, a standard based on exceedance of the geometric mean over a 30-day period (USEPA Region 10, 2000). Duck Creek was determined to be impaired, even though the TMDL document states that sufficient data to determine the frequency of exceedance or geometric mean was unavailable. Although many TMDL documents indicated the lack of sufficient data to justify listing a waterbody, most did not address the need for improved monitoring or collection of water quality data.

A few TMDL documents, however, did demonstrate listing decisions that were supported by a sufficient quantity of water quality data. For example, Colorado listed Willow Creek for exceeding the nutrient WQS based on a large sampling effort of 229 samples over a 6-year period (Colorado DPHE, 2000). Rhode Island listed Chickasheen Brook, Barber Pond, and Yawgoo Pond for total phosphorus, based on water quality data from monthly sampling efforts between the months of May and November from 1988 through 2001 (Rhode Island DEM, 2004). However, even when water quality assessments were supported by large quantities of data, many TMDL documents made no reference to frequency of exceedance of water quality samples, even though many WQS require a specific exceedance frequency.

While sample collection and analysis are also rarely addressed in the TMDL documents, a number of states did apply a consistent methodology of data sampling and analysis to determine impairment and justify listing decisions. The state of Michigan developed TMDLs for Grand River and Carrie Creek to address biota impairment using a standard biological assessment method, called "Procedure 51" (Michigan DEQ, 2002, 2003). By having a standardized water quality assessment and listing methodology, Michigan ensured consistency in impairment determination and listing decisions.

In addition to these different criteria for data sources and quality control, data age and collection timeframe is also quite relevant in the listing process. USEPA indicates that data should not be excluded from consideration solely based on age, but says that states may choose to exclude data that is no longer representative of current waterbody conditions. While many states do have specific requirements regarding the age of water quality data, the topic is not consistently handled.

Colorado recommends that data be no older than 5 years, unless an explanation is provided proving that the data continues to reflect current conditions (Colorado DPHE, 2005). California directs its staff to consider all data and information, regardless of age, to determine which data should be used in the listing assessments

(California SWRCB, 2004). However, California also recognizes that older data may not represent current water quality conditions or may reflect the result of less precise laboratory analytical procedures, so the age of the data used to support listing must be disclosed.

In Minnesota, the Minnesota Pollution Control Agency (MPCA) uses data collected over the most recent 10-year period for all water quality assessments, except in the case of fish contaminated with mercury, where there is no age limit for the use of mercury contaminated fish tissue data (Minnesota PCA, 2004). Similarly, in Washington data used in the assessment should have been collected less than 10 years before the "call for data" period, whenever possible. If no water quality data is provided or too little data is available for an adequate assessment, then data more than 10 years old will be used in the assessment (Washington DE, 2006). New York and Iowa generally only consider data and information from the preceding 5 years for water quality assessments used for Section 303(d) listing (New York DEC, 2002; Iowa DNR, 2003).

Since some states do not place an age limit on water quality data or do not require that past water quality data be assessed for accuracy, this can result in listing decisions based on outdated data that may not reflect the current condition of the waterbody. For instance, Alaska listed Fish Creek as impaired for fecal coliform in 1998 using data samples taken in 1990 (Alaska DEC, 2004). Likewise, Vermont listed Lake Champlain as impaired for phosphorous in 2000 using water quality data from 1990 (Vermont ANR, 2002). Some of these listing decisions were based on older data, without reassessment of more recent water quality to ensure that the data still reflected the current conditions of the waterbodies.

USEPA indicates that states should also specify a sampling approach and identify target sample sizes for impairment decisions in the listing methodology, but should not exclude from consideration data sets that do not meet a target sample size and should provide for a further assessment of data. Given this somewhat ambiguous approach, states have developed much more specific sampling requirements. For example, California guidelines state that samples collected within 200 m of each other should be considered samples from the same location/station, unless justified as independent (California SWRCB, 2004). Samples collected during a single day or short-term natural event shall not be used as the primary data set to support listing. Timing of sampling should include the critical season for the pollutant and applicable WQS.

In Washington, only one sample per day per segment will be used in water quality assessments (Washington DE, 2006). If more than one sample exists, replicates are averaged. For most pollutants, Washington requires a total of at least three samples for an assessment. Colorado guidelines indicate that a waterbody determined to be impaired based on three or fewer samples will be placed on the Monitoring and Evaluation (M&E) list for further evaluation (Colorado DPHE, 2005). A waterbody

determined to be impaired based on four to 10 samples or biological or physical evidence will be listed on the 303(d) list, unless it is determined that the water quality data is not representative (i.e. the data doesn't account for spatial or temporal variation).

Colorado and Iowa also have specific guidelines for water quality sampling in lakes. Colorado requires 2 years of sample results with a minimum of 12 events, representative of seasonal and diel variation, or where acute conditions result in overwhelming evidence (Colorado DPHE, 2005). In Iowa, lakes are sampled three times during the summer at the deepest point in the lake basin (Iowa DNR, 2003).

In New York, the state water monitoring program conducts rotational drainage basin monitoring and sampling (New York DEC, 2002). The collected data is used to determine if a waterbody is achieving its beneficial use. Minnesota also conducts rotational basin wide monitoring, and has specific data requirements for evaluating impairment, including a minimum number of values and data treatment for the following indicators: total phosphorous, chlorophyll a, Secchi disk depth, IBI scores, and fish tissue concentration of PCBs or mercury (Minnesota PCA, 2004). Exceedance thresholds for these indicators are also provided.

With regards to data analysis and analytical approach, USEPA recommends a probability based monitoring and sampling design and encourages states to consider published statistical methods for data analysis. The methodology should provide a justification of why the state has chosen a specific analytical approach.

California's guidelines, for example, are very prescriptive and recommend a weight-of-evidence approach to sample analysis (California SWRCB, 2004). Rationale must be provided for the selection of sampling sites, water quality parameters, sampling frequency and methods that assure the samples are spatially and temporally representative of the surface water and representative of conditions within the targeted sampling timeframe. To be considered temporally independent, samples collected during the averaging period shall be combined and considered one sampling event. For data that is not temporally independent, the measurements shall be combined and represented by a single resultant value. If the averaging period is not stated for the standard, objective, criterion, or evaluation guideline, then samples collected less than 7 days apart shall be averaged.

4.2. Water quality data interpretation and impairment determination

Once water quality data has been collected and analyzed for a particular waterbody, it must then be evaluated and interpreted to determine if a WQS is exceeded. Data is interpreted differently depending on whether water quality criteria are numeric or narrative, as well as whether the data is chemical, biological or physical. USEPA guidance

with regards to data interpretation is generally quite vague, and simply requires states to describe methods of interpretation in the documented listing methodology. While the states have attempted to provide more detailed guidance for internal staff, requirements for data interpretation vary greatly among different states agencies.

Although USEPA guidelines do not specifically address the interpretation of chemical data, most states do provide specific guidance on this issue. For example, the following are some of Colorado's criteria that specifically apply to attainment of chemical standards (Colorado DPHE, 2005):

- chronic chemical standards: 85th percentile;
- hardness-based metals: 85th percentile and mean hardness or paired hardness and concentration data;
- total recoverable metals: 85th percentile;
- dissolved oxygen (DO): 15th percentile;
- minimum pH: 15th percentile;
- maximum pH: 85th percentile;
- coliform: geometric mean.

In California, staff must also follow detailed guidelines pertaining to data interpretation. For toxins, and all conventional and other pollutants, the binomial distribution must be evaluated to determine if exceedances supports rejection of the null hypothesis (California SWRCB, 2004). For DO, if measurements taken over the day show low concentrations in the morning and sufficient concentrations in the afternoon, then it shall be assumed that nutrients are responsible for the DO concentrations if riparian cover, substrate composition or other pertinent factors can be ruled out as controlling DO fluctuations. When continuous monitoring data is available, the 7-day average of daily minimum measurements shall be assessed. In the absence of diel measurements, concurrently collected measurements of nutrient concentration shall be assessed using applicable water quality objectives or acceptable evaluation guidelines and using the binomial distribution. For bacteria, if site-specific exceedance frequency data is unavailable, a binomial distribution should be used. If frequency data is available, it may be used instead. For example, for bacteria data from coastal beaches collected from April 1 to October 31, a 4% exceedance percentage must be applied.

In Kansas state guidelines follow USEPA recommendation of a 10% raw score based on binomial analysis to screen for conventional pollutant data from monitoring sites (Kansas DHE, 2003). Those sites that fail the raw score test (greater than 10% exceedance) are subject to a binomial test. If the binomial test indicates impairment, then the waterbody is placed on the 303(d) list. If the binomial test indicates full support, those sites will be subject to a final screening, which includes a check for evidence of recent excursions in the sample data. If the number of excursions is within the critical number of excursions needed to list a waterbody, and any one of those

excursions occurred in the most recent year of sampling, then the waterbody will be listed as impaired.

Lakes and reservoirs present specific water quality considerations, and the interpretation of chemical data in lakes and reservoirs is addressed in detail by three of the states considered in this review. In Colorado, data interpretation must account for seasonal and diurnal variation (Colorado DPHE, 2005). For DO and temperature, a vertical profile must be developed. DO for the entire water column must be assessed for a mixed waterbody, and individual DO samples are compared to the water quality criterion. For all other chemical parameters, samples are compared to numeric standards, using volumetrically weighted measurements when available.

Kansas guidelines provide the following specific cases in which a lake or wetland should be included on the 2004 303(d) list (Kansas DHE, 2003):

- The lake or wetland assessment unit appeared in the 2002 Section 303(d) list and has not had a TMDL developed for its specified impairment(s).
- The lake has a designated use of primary contact recreation and chlorophyll *a* average concentration greater than 12 ppb.
- The lake has a designated use of secondary contact recreation and chlorophyll *a* average concentration greater than 20 ppb.
- The lake has average total phosphorus concentration greater than 50 ppb over the last 5 years.
- The wetland has average total phosphorus concentration greater than 100 ppb over the last 5 years.
- The lake or wetland, for any other parameter, exceeded WQS or regional norms for more than 1 year in the last 5 years.

In Iowa, the interpretation of fecal coliform in lakes (Iowa DNR, 2003) is typically based on USEPA guidance which requires a comparison of: (1) the geometric mean of at least five samples collected over a 30-day period to state water quality criteria for indicator bacteria (fecal coliforms, *E. coli*, and/or enterococci); and (2) the percentage of samples that exceed a single-sample maximum value. In cases where the geometric mean exceeds the state water quality criterion or more than 10% of the samples exceed the single-sample maximum value, primary contact uses should be assessed as impaired. However, the frequency of sampling at lakes in Iowa with a designated use of primary contact recreation does not provide an adequate number of data to use USEPA method, and thus modified techniques are used to interpret data from these locations.

USEPA guidelines indicate that states using biological assessment to make impairment determinations should also consider other types of information, such as chemical or physical data. Threshold values for impairment determination should be addressed in the state's individual listing methodology.

California, Colorado, Iowa and Washington, in general, have consistent guidelines for biological data. Data can be used to support non-attainment of numeric standards, or narrative standards/designated use classifications. An assessment is typically compared to expected reference condition, such as a Benthic Macroinvertebrate Index of Biotic Integrity (BM-IBI) or a Fish Index of Biotic Integrity (F-IBI). These indexes use metrics to provide a broad assessment of stream biological conditions, which are based on qualitative scoring. In addition, California recommends that visual assessment may be used as a secondary line of evidence (California SWRCB, 2004). Whenever possible, a binomial distribution must be used in the assessment. In Colorado, sampling of multiple sites to determine reference conditions is recommended but not required (Colorado DPHE, 2005).

USEPA and most states do not specifically address the interpretation of physical data, although some use the same guidelines for biological data. California does require evaluation of physical habitat data and other water quality data, when available, to support conclusions about the status of the water segment (California SWRCB, 2004). Kansas also has specific cases in which streams or lakes and wetlands should be listed based on biological data (Kansas DHE, 2003):

- The stream biological impairment appears on the 2002 Section 303 (d) list and has not had a TMDL developed.
- One or more of the biological metrics indicate partial or non-support for biological monitoring stations with three or more samples over the latest 5 years.
- Fish tissue samples show excessive amounts of bio-accumulative pollutants (PCB, chlordane, mercury, etc.) for 3 or more years over the latest 5 years.
- For lakes and wetlands, one or more of the biological metrics indicate partial or non-support for biological monitoring stations with three or more samples over the latest 5 years.

Another concern related to data interpretation is that states often use narrative standards to determine impairments, but usually have no consistent method of interpreting narrative standards across states or even within a particular state. For instance, many states have narrative WQS that apply for nutrient levels and eutrophication in waterbodies. Some states use a Tropic State Index (TSI) to determine impairment, while others use dissolved oxygen levels or other concentration based water quality data. States have also justified listing waterbodies that lack numeric WQS by using Best Professional Judgment (BPJ). However, the use of BPJ can lead to inconsistencies in impairment determination. For instance, Rock Creek in Iowa was listed in 1998 as impaired for ammonia and NO_x ($\text{NO}_2^-/\text{NO}_3^-$) based on BPJ, but was later determined to be not impaired for NO_x during the development of the TMDL, again based on BPJ (Iowa DNR, 2000). This

indicates a potential for inconsistency when using BPJ as a basis for determining impairment of a waterbody.

USEPA provides little specific guidance for determining impairment based on narrative standards, but the states need to address this issue. USEPA does have a program to support development of numeric nutrient criteria and target concentration ranges by ecoregion, which could overcome the interpretation of narrative standards. Freedman et al. (2004) and others have also worked on developing a method for addressing narrative standards, since their interpretation can result in wide differences in listing criteria.

However, in the meantime states must deal with the evaluation of existing narrative standards. For example, Colorado guidelines indicate that determination of impairment based on narrative standards may be supported by chemical data and/or biological/physical assessments that provide clear and convincing evidence, or when physical/habitat data or biological community conditions are significantly less than reference conditions (Colorado DPHE, 2005).

In Washington, listing based on narrative standards must include all of the following (Washington DE, 2006):

- Documentation of environmental alteration related to deleterious chemical or physical alterations, such as nutrients or sediment deposition, as measured by indices of resource condition or resource characteristic or other appropriate measure.
- The alteration must be measured and documented using a generally accepted method based on site-specific information, with literature thresholds appropriate to the situation or with reference sites.
- Documentation of impairment of an existing or designated use related to the environmental alteration on the same waterbody segment.
- Identification of a human contribution to the environmental alteration.

For impairment decisions in California based on narrative and qualitative standards, the data submittal must (California SWRCB, 2004): (1) describe events or conditions that indicate impacts on water quality; (2) provide linkage between the measurement endpoint (e.g. a study that may have been performed for some other purpose) and the WQS of interest; (3) be scientifically defensible; provide analyst's credentials and training; and (4) be verifiable by the State Water Resources Control Board (SWRCB) or the Regional Water Quality Control Board (RWQCB).

In New York, narrative assessments resulting in impaired status results from one of the following (New York DEC, 2002):

- Occasional water quality, or quantity, conditions and/or habitat characteristics periodically prevent specific uses of the waterbody.

- Waterbody uses are not precluded, but some aspects of the use are limited or restricted.
- Waterbody uses are not precluded, but frequent/persistent water quality, or quantity, conditions and/or associated habitat degradation discourage the use of the waterbody.
- Support of the waterbody use requires additional/advanced measures or treatment.

In Iowa, numerical standards have a preference over narrative standards (Iowa DNR, 2003). A narrative standard shall not constitute the basis for determining impairment unless the department identifies specific factors as to why a numeric standard is not sufficient to assure adequate water quality.

Impairment from unknown pollutants is a far more challenging issue. USEPA guidelines, which follow the CWA directly, indicate that: "Waterbodies should be placed on the 303(d) list if the impairment or threat is caused by a pollutant, even if the pollutant is unknown". In Colorado, if it is unclear if impairment is caused by pollution, it is placed on the M&E list (Colorado DPHE, 2005). If impairment is caused by an unidentified pollutant, then the waterbody is listed. However, discharges to the waterbody will not be regulated until the pollutant is identified. In Washington, if the source of impairment is unidentified by suspected to a pollutant, the segment is placed on the 303(d) list (Washington DE, 2006). In Iowa, if a pollutant causing impairment is unknown, the waterbody may be listed (Iowa DNR, 2003). However, the department shall continue to monitor the waterbody to determine the cause of impairment before a TMDL is developed and the listing will have low priority, unless otherwise determined. Finally, in New York waters known to be impaired, but by causes/pollutants that have not been identified, will not be included on the Section 303(d) list (New York DEC, 2002). Some states seem to follow USEPA guidance regarding unknown pollutants, while others either do not address the issue or take an approach contrary to what USEPA suggests.

In a number of instances, states have relied on fish consumption advisories as a basis for listing decisions, even though the advisory may not actually indicate a WQS is exceeded. Georgia listed the Middle and Lower Savannah River in 2000 for exceeding mercury WQS, but the decision was based on a fish consumption advisory, which did not provide indication of exceedances of the numeric WQS (USEPA Region 4, 2001). States sometimes also use fish consumption advisories to determine impairment of WQS, but then set TMDL targets based on other criteria. For example, Pennsylvania developed a TMDL to address Chlordane impairment in the Cheat River, which was listed after issuance of a fish consumption advisory (Pennsylvania DEP, 1999). However, the TMDL was established based on human health water quality criterion, not on the fish advisory action level. It does not always make sense to use these criteria interchangeably, and these types of

inconsistencies can be confusing to stakeholders who are required to adhere to load reductions in the TMDL. In part the confusion arises from the fact that in some cases USEPA asserts that Food and Drug Administration (FDA) action levels do not provide an adequate level of protection, leading to inconsistent values between the fish advisory based on FDA levels and human health criteria.

For fish consumption advisories, USEPA has specific guidance (e.g. USEPA, 2000). For example, USEPA considers impairment based on fish or health advisory if the following conditions are met: (1) the advisory is based on fish/shellfish tissue data, (2) a lower than approved national shellfish sanitation program classification is based on water column/shellfish data, (3) data is collected from the specific segment, and (4) risk assessment parameters are equal or less protective than state WQS.

Many states go further in providing guidance to their staff on the use of fish consumption advisories for determining impairment. In Minnesota, an advisory to limit fish consumption to less than one meal per week for any member of the population is an indication of impairment (Minnesota PCA, 2004). For Washington, impairment is determined if the average of three single-fish samples with the highest concentration or one composite sample made up of at least five fish exceeds the criteria for human health impacts based on EPA's bio-concentration factors and water column criteria (National Toxic Rule) (Washington DE, 2006).

In California, impairment is determined if tissue pollutant levels exceed a pollutant-specific evaluation guideline using the binomial distribution (California SWRCB, 2004). The frequency of exceedance is based on sample size. Colorado guidelines for impairment based on fish advisories are currently being re-evaluated. For now, they are analyzed on a case-by-case basis taking into account: scientific basis for advisory; any existing, relevant sediment or water column data; associated risk level; other relevant technical info; consideration of EPA guidance on fish and shellfish consumption advisories (Colorado DPHE, 2005). For acute standards, impairment is determined by an exceedance frequency of greater than once in 3 years.

In Iowa, the existence of a fish consumption advisory indicates impairment. Generally, impairment is determined if two samples are collected in consecutive years that show contaminant levels exceed an FDA action level (Iowa DNR, 2003). An advisory is rescinded if two consecutive samples show levels are below the action level. A single fish kill during the most recent 3-year period, caused by a pollutant of unknown origin, determines impairment. The maximum frequency of violation is once in 3 years.

4.3. Unnecessary development of TMDLs

Another issue in the TMDL development process is that some states establish TMDLs for waterbodies where WQS either are already attained or cannot be attained. Some

TMDL documents are developed without reassessing the water quality of a waterbody to determine if the basis for listing is still valid. For example, in 2002 Utah developed a TMDL for Browne Lake to address non-attainment of WQS for DO (Utah DEQ, 2003b). However, the document states that "analysis of the most recent dissolved oxygen data (1993–2001) indicates attainment of the appropriate dissolved oxygen criterion," which demonstrates that Browne Lake was in attainment of the dissolved oxygen WQS at the time the TMDL was established. Utah also developed a TMDL for Ken's Lake, which was listed as impaired for temperature (Utah DEQ, 2002). However, a technical analysis was conducted for the TMDL that determined solar energy to be the only thermal input into the lake. Based on the results of the analysis, the document asserts that "Ken's Lake is unable to meet the current state WQS for temperature." A use attainability analysis would have been more appropriate.

Additionally, TMDLs are not required for waterbodies that are expected to attain WQS through other programs. However, some states still develop TMDLs for these waters. The Middle and Lower Savannah River Watershed in Georgia was listed for total mercury based on fish consumption advisories (USEPA Region 4, 2001). It was estimated that approximately 99% of the mercury loading to the watershed is from atmospheric deposition, a non-point source. The TMDL established specific load allocations for atmospheric deposition, even though the EPA expected most of the reduction to be achieved through implementation of the current Clean Air Act (CAA) requirements in the local airshed. Additionally, the TMDL does not specify the sources of atmospheric deposition, how to reduce loading, or how load reductions could be assessed. The TMDL also established a point source waste load allocation, even though the point source polluters were already regulated by NPDES permits, and the amount of point loading to the watershed was negligible compared to atmospheric deposition. States that develop such TMDLs may be unnecessarily using resources needed for addressing waterbodies in real need of a TMDL. However, it must be noted that this TMDL has since been withdrawn by Georgia.

4.4. TMDL implementation and monitoring

Many load and waste load allocations established in TMDLs do not identify specific loading allocations for every pollutant source, and are usually based on reductions that do not consider the feasibility of the source(s) achieving the required reductions. The lack of specificity in load allocations is generally a result of inadequacy in the data used in the listing and subsequent TMDL development process. Load reductions in TMDLs that address impairment from non-point source pollution often do not include methods of ensuring that reductions by the non-point sources will achieve WQS. For instance, Alabama's Big Nance Creek TMDL for fecal coliform relies on

voluntary, incentive-based mechanisms to reduce pollutant loading from non-point sources, with no mechanism for determining if targeted reductions are met (Alabama DEM, 2002). The TMDLs that address nutrient impairment for Annabessacook Lake (Maine DEP, 2004) and Beaver River, (Utah DEQ, 2003a), were developed with limited understanding of pollutant loading from non-point sources due to lack of sufficient data, yet the TMDLs required non-point source load reductions. The TMDLs did not adequately establish effective methods of reducing pollutant loading because the sources were unknown.

Ultimately, a TMDL will not be effective in addressing water quality if it is not properly implemented. A clear and concise implementation strategy is an essential component of the TMDL process. A monitoring schedule is also necessary to ensure that the TMDL effectively reduces pollutant loading and WQS are attained. The majority of the reviewed TMDL documents did not include an implementation or monitoring program, or if they did, many are vague. Sometimes a TMDL will offer several alternatives for implementation, but will not specify which alternative to implement. Additionally, many TMDLs do not include schedule or timeframe for attainment of WQS.

However, a few TMDLs do provide a framework for implementation and targets for long term monitoring. The Santa Clara River, CA TMDL includes a comprehensive monitoring program, including monitoring for publicly owned treatment works, municipal separate storm water sewer systems (MS4s), and non-point sources (California LARWQCB, 2003). An implementation schedule is provided with specific dates for completion of various components of the implementation plan. In addition, a cost analysis of implementation was conducted. In addition, California, as well as other states, will sometimes include TMDL implementation and monitoring programs in their regional basin plans. The TMDLs with prescriptive implementation plans and scheduled monitoring programs are more likely to be effective in attaining desired water quality goals.

4.5. Delisting issues

The purpose of establishing a TMDL for a waterbody is to attain WQS for designated uses, resulting in removal of the waterbody from the Section 303(d) list. Many states have refrained from delisting waterbodies due to lack of clear guidance from USEPA. The requirements provided by USEPA for removing a waterbody from the 303(d) list generally follow the criteria for the listing process, but also include some additional considerations. USEPA (2003) guidelines indicate that a waterbody may be delisted with demonstration of "good cause," such as:

- more recent or more accurate data or modeling demonstrate that the applicable WQS(s) is being met;
- demonstration that the waterbody was incorrectly listed;
- development of a new listing methodology concluding that WQS are in attainment;

- demonstration that more stringent effluent limitations or other pollution control measures will result in the attainment of WQS;
- approval or establishment by EPA of a TMDL since the last 303(d) list.

To successfully remove a waterbody from the 303(d) list, a state must apply the same methodology used to justify initially listing the waterbody, or a methodology that is consistent with the state's WQS and deemed statistically reasonable by the EPA. In most cases, delisting follows USEPA's biannual review process, to coincide with the listing cycle.

While the delisting process is arguably just as complicated, if not more so, than the listing process, a small handful of states, including Colorado and New Mexico, have documented rationale for delisting decisions. In 1998, Colorado listed Fruitgrowers Reservoir as impaired for fecal coliform and ammonia, but it was unclear if listing was based on a quantitative assessment of water quality data (Colorado DPHE, 2002). In 2001, water quality data was re-assessed, including 90 samples for fecal coliform and 67 samples for ammonia, which indicated that Fruitgrowers Reservoir was in attainment of the fecal coliform and ammonia WQS. Colorado delisted Fruitgrowers Reservoir based on USEPA guidance that states a waterbody may be removed from the 303(d) list at such time as review of the original listing decision is found to be in error.

In 2000, Colorado successfully removed Big Thompson River and Little Thompson River, previously listed for non-attainment of WQS for fecal coliform, from the 303(d) list (Colorado DPHE, 2000). An analysis of 50 samples from Big Thompson and 14 samples from Little Thompson indicated that each waterbody was in attainment of the designated fecal coliform WQS. Colorado's decision to delist these waterbodies was supported by USEPA guidance asserting that a waterbody may be removed from the list at such times as the waterbody is meeting all applicable WQS. Similarly, a segment of Whitewater Creek in New Mexico was listed on the 2000–2002 303(d) lists for non-attainment of WQS for metals based on an inaccurate data assessment (New Mexico ED, 2001). Re-evaluation of the data did not show any exceedances of the WQS. Therefore, the segment of Whitewater Creek was removed from the next 303(d) list.

5. Conclusions

The process of determining impairment in a waterbody has become much more formalized in the past 15 years. Earlier guidance provided by USEPA resulted in wide differences in criteria used for including waterbodies on the 303(d) list. While it is clear that USEPA guidance has improved from the early 1990s, several states decided in the early 2000s to develop additional guidance for their own agencies for the listing and delisting processes, given the

significant resources required to prepare TMDLs. This has resulted in significant disparities among states in emphasis on different aspects of the listing and delisting processes.

The review of numerous TMDL documents from more than 10 states throughout the country illustrates the range of the issues associated with impairment determination and TMDL development, most notably the general lack of consistency in impairment determination. Common problems, including unclear or conflicting listing methodologies, inconsistent interpretation of narrative and numeric WQS, arbitrary and unmeasured pollutant loading allocations and reduction requirements, varying data quality and quantity, and vague and inadequate implementation and monitoring, can result in TMDLs that do not adequately address impairment. The results of this analysis demonstrate the need for adequate guidance in the 303(d) listing and TMDL development process.

In general, while states have taken the initiative to address these pertinent issues, it is clear from the review of the TMDL documents that even within a state there are significant differences in application of particular criteria. States needed to address specific aspects of the listing process, such as consideration of data sources and data quality or interpretation of narrative water quality criteria, because USEPA did not provide clear guidance on these issues. Sampling frequency, number of exceedances, duration of exceedance, age of data, timeframe to be considered, and many other such issues had been left to the interpretation of local staff, but are now being addressed more specifically by the States.

Although listing and delisting criteria is generally clearer at the state level, the development of differing state guidance documents has resulted in diversity in the development of the 303(d) lists and in the process of delisting. For many stakeholders, this has no significant impact, since they operate within a particular state. However, this can lead to confusion for some stakeholders that participate in the 303(d) process in several states, and thus must deal with differing and sometimes conflicting requirements depending on the location of their facilities. Several more iterations of the listing and delisting process may be necessary before common national criteria for the determination of impairment of a waterbody are established.

Acknowledgments

The authors wish to acknowledge funding from the Electric Power Research Institute (Grant no.W06085-02). The views expressed here are those of the authors and may not reflect those of the funding agency.

References

- Air Quality Initiative Steering Committee, 1997. Improving the State Implementation Plan Process. Western Governors Association <<http://www.westgov.org/wga/publicat/sippap3.htm>>. Last visited on 9/5/2006.
- Alabama DEM, 2002. USEPA Office of Water. Final Big Nance Creek TMDL Development for Fecal Coliform <<http://www.epa.gov/waters/tmdl/docs/Big%20Nance%20Creek%20Pathogens.pdf>>. Alabama Department of Environmental Management. Last visited on 09/05/2006.
- Alaska DEC, 2004. Total Maximum Daily Load (TMDL) for Fecal Coliform in the Waters of Fish Creek in Anchorage, Alaska <<http://www.dec.state.ak.us/water/tmdl/pdfs/fishcreekanchorage.pdf>>. Alaska Department of Environmental Conservation. Last visited on 09/05/2006.
- Boyd, J., 2000. The new face of the Clean Water Act: A critical review of the EPA's proposed TMDL rules. Resources for the Future, Washington, DC. RFF-DP-00-12 <<http://www.rff.org/Documents/RFF-DP-00-12.pdf>>. Last visited on 9/5/2006.
- California LARWQCB, 2003. Santa Clara River Total Maximum Daily Loads for Nitrogen Compounds <http://www.swrcb.ca.gov/rwqcb4/html/meetings/tmdl/tmdl_ws_santa_clara.html>. Los Angeles Regional Water Quality Control Board. Last visited on 09/05/2006.
- California SWRCB, 2004. Final Federal Equivalent Document Appendix A: Proposed Listing Policy. California State Water Resources Control Board <http://www.swrcb.ca.gov/tmdl/docs/ffed_appxa093004.pdf>. Last visited on 9/5/2006.
- Colorado DPHE, 2000. Determination of TMDL Inapplicability: Big Thompson River/ COSPBT05, Little Thompson River/COSPBT09; South Platte Basin <http://iaspub.epa.gov/tmdl/enviro.control?p_list_id=COSPBT05&p_cycle=1998>. Colorado Department of Public Health and Environment (DPHE), Water Quality Control Division. Last visited on 09/05/2006.
- Colorado DPHE, 2002. Determination of TMDL Inapplicability: Fruitgrowers Reservoir <http://iaspub.epa.gov/tmdl/enviro.control?p_list_id=COGULG09&p_cycle=1998>. Colorado Department of Public Health and Environment, Water Quality Control Division. Last visited on 09/05/2006.
- Colorado DPHE, 2005. Section 303(d) Listing Methodology: 2006 Listing Cycle. Colorado Department of Public Health and Environment <[http://www.cdph.state.co.us/op/wqcc/SpecialTopics/303\(d\)/revised303\(d\)listmethproposal.pdf](http://www.cdph.state.co.us/op/wqcc/SpecialTopics/303(d)/revised303(d)listmethproposal.pdf)>. Last visited on 09/05/2006.
- Freedman, P.L., Larson, W.M., Dilks, D.W., Schechter, D., Nemura, A., Naperal, T., DePinto, J.V., Prothro, M.G., Boese, G.W., Dettelbach, A., Nothman, L., Thornton, K., Ford, D., Massirer, P., Soerens, T., Stevens, K.B., Sobrinho, J.A.H., 2004. Navigating the TMDL process: method development for addressing narrative criteria. WERF Report: (Project 00-WSM-1). ISBN:1843396742. ISSN Online:1476-1777. IWA Publishing. Available at <<http://www.iwapublishing.com/>>.
- Helland, E., 1998. The revealed preferences of state EPAs: stringency, enforcement, and substitution. *Journal of Environmental Economics and Management* 35, 242–261.
- Iowa DNR, 2000. TMDL for Ammonia and NO_x, Rock Creek, Clinton County, Iowa <http://www.epa.gov/owow/tmdl/examples/nutrients/ia_rockcreek.pdf>. Iowa Department of Natural Resources. Last visited on 09/05/2006.
- Iowa DNR, 2003. Methodology for developing Iowa's 2002 Section 303(d) list of impaired waters. Iowa Department of Natural Resources <http://wqm.igsb.uiowa.edu/WQA/303d/2002/2002_303d_methodology.pdf>. Last visited on 9/5/2006.
- Kansas DHE, 2003. Methodology for the Evaluation and Development of the 2004 Section 303(d) List of Impaired Water Bodies for Kansas. Department of Health and Environment <<http://www.kdhe.state.ks.us/tmdl/2004Kansas303dMethod.pdf>>. Last visited on 9/5/2006.
- Keller, A.A., Cavallaro, L., Ryals, C., 2005. Determination of Impairment of a Waterbody. EPRI Report, EPRI, Palo Alto, CA. Available at <<http://www2.bren.ucsb.edu/~keller/publications.htm>>. Last visited on 9/5/2006.
- Maine DEP, 2004. Total Maximum Daily Load for Phosphorus in Annabessacook Lake <<http://epa.gov/region1/eco/tmdl/assets/pdfs/me/annabessacooklake.pdf>>. Maine Department of Environmental Protection. Last visited on 09/05/2006.
- Michigan DEQ, 2002. Total Maximum Daily Load for Biota of Carrier Creek, Eaton County <<http://www.deq.state.mi.us/documents/>>.

- deq-swq-gleas-tmdlcarrier.pdf). Michigan Department of Environmental Quality. Last visited on 09/05/2006.
- Michigan DEQ, 2003. Total Maximum Daily Load for Biota for the Grand River, Jackson County <<http://www.deq.state.mi.us/documents/deq-wb-sw-as-tmdl-biota-bass.pdf>>. Michigan Department of Environmental Quality. Last visited on 09/05/2006.
- Minnesota PCA, 2003. Volunteer Surface Water Monitoring Guide. Minnesota Pollution Control Agency. Minnesota Pollution Control Agency <<http://www.pca.state.mn.us/water/monitoring-guide.html>>. Last visited on 9/5/2006.
- Minnesota PCA, 2004. Guidance Manual For Assessing the Quality of Minnesota Surface Waters for the Determination of Impairment, 305(b) Report and 303(d) List. Minnesota Pollution Control Agency <<http://www.pca.state.mn.us/publications/manuals/tmdl-guidancemanual04.pdf>>. Last visited on 9/5/2006.
- Moring, F., 1999. State-Federal issues in LDC residential unbundling operations. *Natural Gas* 15 (11).
- New Mexico ED, 2001. Metals Assessment and Delisting Rationale for Whitewater Creek <http://www.nmenv.state.nm.us/swqb/Metals_Assessment_and_Delisting_Rationale_for_Whitewater_Creek_11-05-01.pdf>. New Mexico Environmental Department: State Water Quality Bureau. Last visited on 09/05/2006.
- New York DEC, 2002. New York State Section 305(b) and 303(d) Consolidated Assessment and Listing Strategy. New York State Department of Environmental Conservation <<http://www.dec.state.ny.us/website/dow/calhome.html>>. Last visited on 9/5/2006.
- NRC, 2001. Assessing the TMDL Approach to Water Quality Management. National Research Council, Commission on Geosciences, Environment and Resources (CGER). National Academy of Sciences, 500 Fifth St. N.W., Washington, DC, 20001. Available at <<http://www.nap.edu/books/0309075793/html/>>.
- Parinet, W.E., 1993. Regulation and federalism—legal impediments to state health-care reform. *American Journal of Law and Medicine* 19 (1–2), 121–144.
- Pennsylvania DEP, 1999. Final Total Maximum Daily Load for Cheat River, Fayette County, Chlordane <http://www.epa.gov/owow/tmdl/examples/pesticides/pa_cheat.html>. Pennsylvania Department of Environmental Protection. Last visited on 09/05/2006.
- Rhode Island DEM, 2004. Chickashen Brook, Barber Pond, and Yawgoo Pond, Total Phosphorous <<http://www.dem.ri.gov/programs/benvirion/water/quality/rest/reports.htm>>. Rhode Island Department of Environmental Management. Last visited on 09/05/2006.
- Rhymer, R.N., 2005. Taking back the power: federal vs. state regulation on postsecondary education benefits for illegal immigrants. *Walburn Law Journal* 44, 16–18.
- Rogers-Dillon, R.H., 1999. Federal constraints and state innovation: lessons from Florida's family transition program. *Journal of Policy Analysis and Management* 18 (2), 327–332.
- Sigman, H., 2003. Letting states do the dirty work: state responsibility for federal environmental regulation. *National Tax Journal* 56 (1), 107.
- USEPA, 1991. Guidance for Water Quality-Based Decisions: The TMDL Process. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/decisions/dec1c.html>>. Last visited on 09/05/2006.
- USEPA, 1993. Guidance for 1994 Section 303(d) Lists. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/1994guid.html>>. Last visited on 09/05/2006.
- USEPA, 1997. National Clarifying Guidance For 1998 State and Territory Section 303(d) Listing Decisions. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/lisgid.html>>. Last visited on 09/05/2006.
- USEPA, 2000. Guidance: Use of Fish and Shellfish Advisories and Classifications in 303(d) and 305(b) Listing Decisions. USEPA Office of Science and Technology <<http://www.epa.gov/waterscience/library/wqstandards/shellfish.pdf>>. Last visited on 09/05/2006.
- USEPA, 2001. 2002 Integrated Water Quality Monitoring and Assessment Report Guidance. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/2002wqma.html>>. Last visited on 09/05/2006.
- USEPA, 2002a. EPA Review of 2002 Section 303(d) Lists and Guidelines for Reviewing TMDLs under Existing Regulations issued in 1992. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/guidance/csmemo.html>>. Last visited on 09/05/2006.
- USEPA, 2002b. Recommended Framework for EPA Approval Decisions on 2002 State Section 303(d) List Submissions. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/guidance/listapproval.pdf>>. Last visited on 09/05/2006.
- USEPA, 2002c. Guidelines for Reviewing TMDLs under Existing Regulations issued in 1992. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/guidance/final52002.pdf>>. Last visited on 09/05/2006.
- USEPA, 2003. Guidance for 2004 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d) and 305(b) of the Clean Water Act, TMDL -01-03. USEPA Office of Water <http://www.epa.gov/owow/tmdl/tmdl0103/2004rpt_guidance.pdf>. Last visited on 09/05/2006.
- USEPA, 2006a. Total Maximum Daily Load Reports. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/>>. Last visited on 09/05/2006.
- USEPA, 2006b. Consolidated Assessment and Listing Methodology—Toward a Compendium of Best Practices. USEPA Office of Water <<http://www.epa.gov/owow/monitoring/calhome.html>>. Last visited on 09/05/2006.
- USEPA, 2006c. 2006 Integrated Report Guidance. USEPA Office of Water <<http://www.epa.gov/owow/tmdl/2006IRG/>>. Last visited on 09/05/2006.
- USEPA Region 10, 2000. Total Maximum Daily Load (TMDL) for Fecal Coliform Bacteria in the Waters of Duck Creek in Mendenhall Valley, Alaska <<http://www.dec.state.ak.us/water/tmdl/pdfs/duckcreekfecaltmdl.pdf>>. USEPA Region 10. Last visited on 09/05/2006.
- USEPA Region 4, 2001. TMDL for Total Mercury in the Middle/Lower Savannah River, Georgia <http://www.epa.gov/owow/tmdl/examples/mercury/ga_savfinal.pdf>. Last visited on 09/05/2006.
- Utah DEQ, 2002. TMDL For Temperature in Ken's Lake, Utah <http://www.waterquality.utah.gov/TMDL/Kens_Lake_TMDL.pdf>. Utah Department of Environmental Quality, Division of Water Quality TMDL Section. Last visited on 09/05/2006.
- Utah DEQ, 2003a. Beaver River Watershed TMDL <http://www.waterquality.utah.gov/TMDL/Beaver_Watershed_TMDL.pdf>. Utah Department of Environmental Quality, Division of Water Quality TMDL Section. Last visited on 09/05/2006.
- Utah DEQ, 2003b. Browne Lake Dissolved Oxygen and Phosphorus TMDLs <http://www.waterquality.utah.gov/TMDL/Browne_Lake_TMDL.pdf>. Utah Department of Environmental Quality, Division of Water Quality TMDL Section. Last visited on 09/05/2006.
- Vermont ANR, 2002. Lake Champlain Phosphorus TMDL. Vermont Agency of Natural Resources, Department of Environmental Conservation and New York State Department of Environmental Conservation <http://www.anr.state.vt.us/dec/waterq/lakes/does/lp_lctmdl-report.pdf>. Last visited on 09/05/2006.
- Washington DE, 2006. Assessment of Water Quality for the Clean Water Act Sections 303(d) and 305(b) Integrated Report. Washington Department of Ecology <http://www.ecy.wa.gov/programs/wq/qa/wq_assessment_policy0506-draft.pdf>. Last visited on 09/05/2006.
- WGA, 2006. Western Governors Association. Mission, Priorities, and Strategies <http://www.westgov.org/wga_mission.htm>. Last visited on 09/05/2006.

Environmental regulatory failure and metal contamination at the Giap Lai pyrite mine, Northern Vietnam

N. Håkan Tarras-Wahlberg^a, Lan T. Nguyen^{b,*}

^aSwedish Geological, P.O. Box 19090, 104 32 Stockholm, Sweden

^bDepartment of Water and Environmental Studies, Linköping University, 581 83 Linköping, Sweden

Received 13 March 2006; received in revised form 26 September 2006; accepted 14 December 2006

Available online 6 February 2007

Abstract

The causes for the failure in enforcement of environmental regulations at the Giap Lai pyrite mine in northern Vietnam are considered and the environmental impacts that are associated with this mine are evaluated. It is shown that sulphide-rich tailings and waste rock in the mining area represent significant sources of acid rock drainage (ARD). The ARD is causing elevated metal levels in downstream water bodies, which in turn, represent a threat to both human health and to aquatic ecosystems. Metal concentrations in impacted surface waters have increased after mine closure, suggesting that impacts are becoming progressively more serious. No post-closure, remediation measures have been applied at the mine, in spite of the existence of environmental legislation and both central and regional institutions charged with environmental supervision and control. The research presented here provides further emphasis for the recommendation that, while government institutions may need to be strengthened, and environmental regulations need to be in place, true on the ground improvement in environmental quality in Vietnam and in many other developing countries require an increased focus on promoting public awareness of industrial environmental issues.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Metal contamination; Environmental policy; Acid rock drainage; Vietnam

1. Introduction

This paper examines the causes for the failure in enforcement of environmental regulations at the Giap Lai pyrite mine in northern Vietnam (Fig. 1), and evaluates the environmental impacts that are associated with this mine. The research was conducted as part of a development cooperation project, supported by the Swedish International Development Agency (Sida), and the Vietnamese Ministry of Industry.

Baker (1990) suggested that the two main reasons for environmental policy failure are: (i) regulatory, when no legal mandate exists for environmental protection and (ii) related to enforcement, when the agency charged with environmental supervision and control lacks the infrastructure and/or power to enforce regulations. Efforts to encourage improved environmental management in devel-

oping countries have generally focused on the development of environmental regulations, and the setting up of state environmental agencies charged with the implementation of these regulations. Important efforts have also been made in introducing and promoting environmental awareness and the related concepts of cleaner production (CP) and environmental management systems (EMS) within mining companies (Hilson, 2000; Hilson and Nayee, 2001). Internationally, this has led to significant environmental improvements at many mine sites. In developing countries, there have even been cases where international mining companies have assisted developing country regulators in developing new environmental regulations to help ensure responsible mining (e.g. see Hilson and Murck, 2000).

In Vietnam, the legislative framework for environmental management is laid out in the National Law on Environmental Protection (NLEP) from 1993. The NLEP calls for the parties responsible for environmental pollution to pay compensation (the polluter pays principle), and provides a basis for the elaboration of environmental standards, and

*Corresponding author. Tel.: +46 13 282 959; fax: +46 13 133 630.

E-mail address: LanNg@tema.liu.se (L.T. Nguyen).

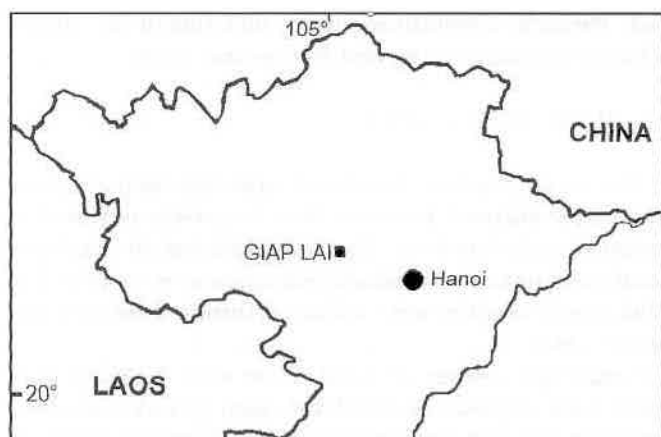


Fig. 1. Map showing location of the Giap Lai mine in Northern Vietnam.

the requirements for Environmental Assessments on both new and existing facilities. The main agency responsible for environmental supervision and control is the National Environment Agency (NEA), within the recently created Ministry of Natural Resources and Environment (MONRE). At the local level, the government has established Departments of Science, Technology and Environment (DOSTEs) in each province, charged with implementing environmental laws for all but the largest firms. So-called Peoples Committees also play an important role in implementing government policies at the local level. People's Committees exist at a range of levels, from village to commune (a group of villages). In this paper, the Giap Lai mining operations serve as an example of to what degree the regulatory regime outlined above is being implemented.

Metal mine sites are often significant sources of metal pollution in water bodies and soils, with associated effects on local ecosystems (Lottemoser et al., 1999). World-wide, there have been many instances of environmental degradation caused by metal contamination from abandoned mining operations (e.g. see Environment Australia, 1997; Bervoets et al., 1998; Azcuc, 1999; Apodaca et al., 2000; Martin et al., 2001). The release of metals from such sites is often caused through so-called acid rock drainage (ARD). ARD is generated when sulphide minerals oxidise, releasing sulphuric acid and hence metal ions into solution (Loxham, 1988; Salomons, 1995). Typically, ARD is formed in abandoned mine adits or in tailings impoundments and waste rock deposits. It has been shown that once the formation of ARD has been initiated, it is extremely difficult to mitigate and/or manage. As ARD formation typically takes place over very long time spans, this process may lead to the creation of significant and long-term environmental liabilities (Harries and Ritchie, 1988; Loxham, 1988). The Giap Lai pyrite mine may be such an environmental liability.

2. Study area

The Giap Lai mining area is located in the province of Phu Tho, about 80 km west of Hanoi (Fig. 1). The area is

situated in the Doidong valley, at an altitude of 70 m and nestled in among low hills, which reach 200–400 m in height. Currently, about 500 people live in the immediate vicinity of the mine site, and some 5000 live in small villages situated within a 10 km radius.

Mining of pyrite by the Giap Lai pyrite company occurred during the period 1975–1999, after which these operations ceased. The company was then renamed to the Phu Tho Mining and Mineral servicing company and it is now operating a number of smaller kaolin, feldspar and other industrial mineral mines in the area.

Until recently, all mine ventures in Vietnam belonged to the state. Thus, the Giap Lai pyrite company belonged to the Vietnam Mineral Corporation (VMC), which in turn, is positioned within the Ministry of Industry. At the height of operations, activities employed some 1100 people, and 3–4000 people, largely dependent on the mine, were living in areas adjacent to the operations. Most of the employees and beneficiaries were people from the nearby districts, with only some 20% of the work force being drawn from other provinces. At present, the company employs some 150 people.

The mining area is located on a regional water divide, and is drained by two small streams (Fig. 2). These are the Thanh Son, which empties into the Bua River (6.5 km to the NW), and the Thanh Thuy, which empties into the Da River (7 km to the SE). The Thanh Son's discharge varies between 10 and 100 l s⁻¹, whereas the Thanh Thuy's discharge varies between 5 and 350 l s⁻¹. The Bua River carries 7–50 m³ s⁻¹, whereas the Da River is a very large river with flows of several hundred cubic metres per second (SGAB-CIE, 2002).

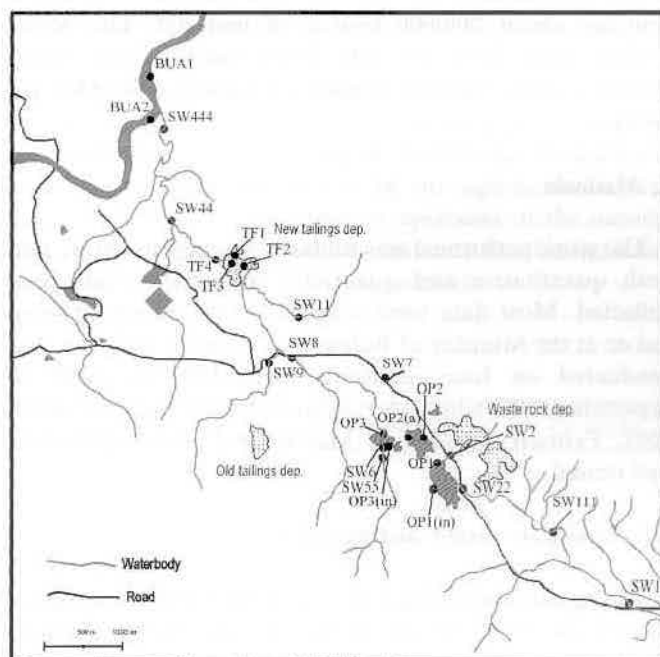


Fig. 2. Map showing the location of all surface water and sediment sampling sites in the Giap Lai mining area.

The climate is monsoonal, with an annual rainfall average of 1600 mm, and with high humidity (80–90%) throughout the year. There is one rainy season, from April to September, during which the monthly rainfall exceeds 200 mm whereas in the remainder of the year, it ranges from 30 to 80 mm. Yearly mean temperature is 23 °C.

During pyrite mining, the exploited ore consisted of three massive ore bodies, with a total length of about 2 km and a grade of pyrite in the range 26–34% (SGAB-CIE, 2002). Mine records suggest that initially, the annual production figures were in the range 10–50 thousand tonnes of pyrite ore. The mine was expanded substantially in the mid-1980s when production initially peaked at over 200 thousand tonnes of ore, then gradually decreasing to below 30 thousand tonnes in the late 1990s (SGAB-CIE, 2002). Mining was performed in three pits. Open Pit 1 (13 ha) was mined from 1975 to 1995, Open Pit 2 (15 ha) from 1992 to 1996 and Open Pit 3 (2 ha) was mined from 1996 to 1999 (Fig. 2). The mining operation terminated in 1999, and all three pits are now abandoned and filled with water. Open pits 1 and 3 have surface water outflows, whereas the second pit is hydrologically closed (SGAB-CIE, 2002).

Mine records suggest that a total of more than 5 million m³ of waste rock was removed from the three pits during mining. About one million m³ of this waste rock is stored in two deposits situated to the north of the open pits, and it is also believed that a substantial part of waste rock has been used to backfill the three open pits (SGAB-CIE, 2002). The ore was treated in a central concentration plant, and mine records suggest that a total of 1.2 million tonnes of pyrite ore was produced. Tailings from the process were deposited in two tailings facilities (Fig. 2). A first tailings dam was used up until the late 1980, and it contains about 200,000 tonnes of material. The newer facility, used from the late 1980s until mining ended contains some 880,000 tonnes of tailings (SGAB-CIE, 2002).

3. Methods

The work performed was multidisciplinary in nature and both quantitative and qualitative information has been collected. Most data were collected either on site in Giap Lai or at the Ministry of Industry in Hanoi. Fieldwork was conducted on four occasions from 1997 to 2002: in September 1997 when operations were still ongoing; April 2001; February 2002; and March 2002 when operations had ceased.

3.1. Literature review and interviews

The research included a review of the Vietnamese policy framework, the relevant legislation and the institutions involved in environmental management issues. Interviews with mine management and staff at the Ministry of Industry as well as relevant local institutions (DOSTE

and People's committees) were undertaken to obtain detailed information relevant to this case study.

3.2. Waste characterisation

The waste products produced from the mining operations were analysed to assess their long-term potential to produce contamination. This was achieved by analysing metal content and by conducting leaching experiments. The solid waste samples were collected during February and March 2002.

Composite samples of solid waste were sampled from waste rock deposits north of the open pits (10 samples), and from new and old tailings facilities (3 samples each). At each site, composite samples comprising five different 0.5 kg subsamples collected from areas of about 10 m² were collected. The samples were taken both at the surface and at depths of as much as 0.5 m. The composite samples were thoroughly mixed and then crushed in a mortar, whereafter they were dissolved in aqua regia before the contents of Al, Fe, Cu, Ni, Pb and Zn were determined by Atomic Absorption at the NIMM laboratories in Hanoi.

Four source materials for metal contamination were selected for leaching experiments using humidity cells: (i) waste rock; (ii) fresh material from the new tailings facility; (iii) older and partly oxidised material from the new tailings facility and (iv) material from the old tailings facility. The procedures and designs of the leaching tests followed Sobek et al. (1978), and the tests were run for a period of 28 weeks. The objective of such tests is to accelerate the processes of leaching and buffering, and in this way ascertain if the material will, in the longer term, generate ARD. Water from the new tailings impoundment was used to moisturise these tested samples before the oxidation process to ensure the presence of bacteria during oxidation. The leachates collected each week were analysed for pH, electrical conductivity, and sulphate concentration. The leachate analysis was carried out by using handheld instruments, through titration and by using a spectrophotometer, respectively, at the NIMM laboratories in Hanoi.

3.3. Assessment of environmental impacts

The severity of environmental impacts was investigated by collecting samples of surface waters, and stream sediments at selected sites in and around the mining area (Fig. 2). Surface water was taken during all sampling campaigns. Stream sediment sampling was collected twice (in September 1997 and February 2002).

3.3.1. Surface water

Physico-chemical parameters were determined in the field by standard handheld instruments and complimentary wet chemistry analysis at National Institute of Mining and Metallurgy's (NIMM) laboratories in Hanoi. The water samples for analysis of metals were passed through a

0.45 µm filter and thereafter acidified to pH 2 in acid-washed polyethylene bottles. Metal concentrations were analysed by ICP-AES/MS at the Analytica laboratories in Luleå Sweden.

3.3.2. Stream sediment

Stream sediment composite samples, comprising three water-saturated subsamples, were collected with a spatula at each site, with each subsample weighing about 250 g. The subsamples were thoroughly mixed together to make up the composite samples, and these were collected in plastic bags and then kept in a cooler. The samples were submitted to the Analytica laboratories in Luleå, Sweden for analysis. At the laboratory, the samples were wet sieved through <63 µm mesh and dried before they were bomb digested in a mix of concentrated nitric acid and hydrogen peroxide (30%). The samples were then analysed by ICP-AES/MS for all metals, except mercury, which was analysed by Cold Vapour Atomic Fluorescence.

3.3.3. Method of interpretation

The evaluation of impacts is based upon comparisons with: (i) estimations of world wide average concentrations of metals in water and sediment (Wedepohl, 1991; Förstner and Wittman, 1983); (ii) WHO guidelines for drinking water quality (WHO, 2004); (iii) US EPA water quality standards for the protection of aquatic life (US EPA, 2005a); (iv) US EPA sediment quality guidelines for the protection of aquatic ecosystems (US EPA, 2005b) and (v) conditions at sites that have not been affected by mining and processing, or other industrial activities—so-called background values.

4. Results

4.1. Waste characterisation

The sources of metal contamination in the mining area mainly consist of: (i) waste rock, and (ii) process tailings. Table 1 summarises the metal content of samples taken from the two tailings impoundments and the waste rock deposits. It can be seen that most of the samples contain

appreciable amounts of metals. Further, the levels of toxic trace metals are similar in all three types of material.

The results of the leaching tests are illustrated in time series in Fig. 3. The progressive leaching of a fresh sample from the new tailings impoundment (new tails) shows that this material is initially well buffered, as the pH remains high until week 6. However, after week 6 the pH falls dramatically to stabilise at between 2 and 3. Simultaneously, there is an increase in the conductivity of the leachate coupled with a concomitant increase in sulphate levels. This provides good evidence that, in spite of initially being well buffered, this material will eventually generate ARD.

The other three samples, representing waste rock, material from the old tailings impoundment, and already oxidised parts of the new tailings impoundment have similar leaching characteristics. Unlike the new tails, pH values are low already from the onset, equilibrated at around 3. The pH remained stable through the duration of the experiment. Conductivity readings and sulphate concentrations gradually decrease with time, showing how the easily leachable sulphide minerals become progressively scarcer. Thus, it is suggested that waste rock and tailings material will generate considerable amounts of ARD, and that this may occur over a long periods of time. These conclusions are corroborated by the results of the field sampling and analysis (see below).

4.2. Environmental impacts

Three main areas of surface water contamination were examined: (i) areas downstream of the new tailing impoundment; (ii) water in the open pits and (iii) water downstream of the waste rock deposits. Further, six water samples were collected to represent local background conditions. The median background values were thus calculated. The results are summarised in Table 2.

The background conditions encountered are overall in line with existing estimation of worldwide background levels. Thus, there appears to be no significant natural elevation of metal levels in water upstream of the mining area. Conversely, the contrast between the background conditions and impacted areas is marked.

Table 1
Median metal content in waste rock, and material from the old and new tailings facilities

		Al (%)	Fe (%)	Cu (mg kg ⁻¹)	Ni (mg kg ⁻¹)	Pb (mg kg ⁻¹)	Zn (mg kg ⁻¹)
Waste rock (n = 11)	Median	8.1	16.7	76	92	26	88
	Std. dev	2.7	24	106	63	24	57
Old tailings dam (n = 3)	Median	3.4	23.9	97	55	19	223
	Std. dev	0.69	9.2	23	5	0.58	68
New tailings dam (n = 6)	Median	5.9	4.9	125	58	22	69
	Std. dev	1.7	4.1	111	56	9.2	39
US EPA (ERM)		—	—	270	52	218	410

US EPA sediment quality guidelines are used for comparison.

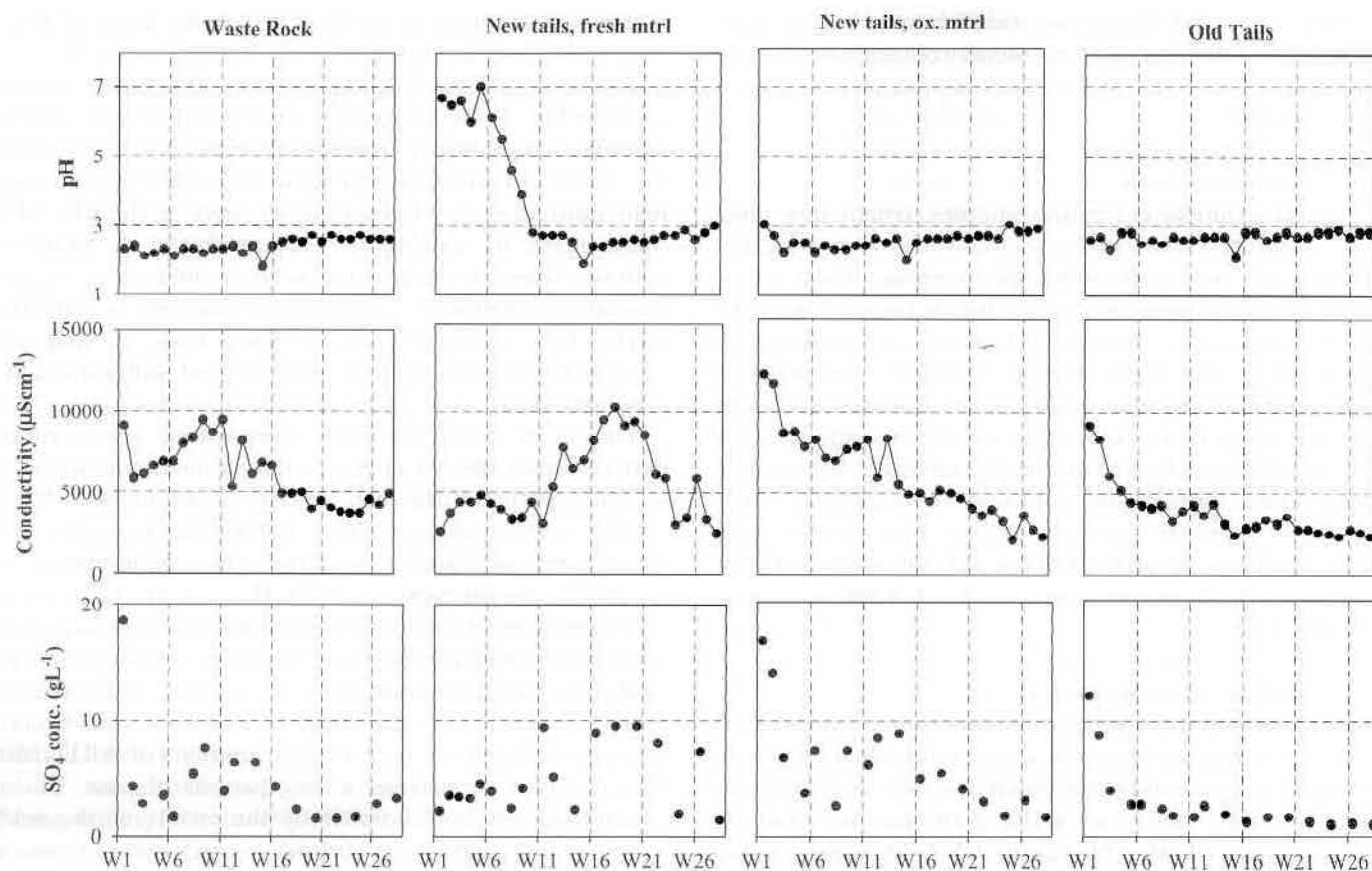


Fig. 3. Graphs showing results of the leaching experiments.

Samples collected downstream of the new tailings impoundment have low pH and highly elevated concentrations of Al and of the trace metals Cd, Cu, Ni, Pb and Zn. Further, concentrations of the less toxic Zn and Mn are very highly elevated downstream of this facility. Seepage appearing at the foot of the tailings impoundment (TF4) contains especially high concentrations of metals, and sampling at this site in 1997, 2001 and 2002 suggests that metal leaching is becoming progressively more severe with time (Fig. 4). The pH of this seepage has decreased dramatically during this period, from just below neutral in 1997 when operations were still ongoing (6.3), to strongly acidic in 2001 and 2002 (5.4 and 3.0, respectively). In the 2002 investigation, the seepage water contained high concentrations of metals that are toxic to both humans and to aquatic life. The concentrations are, however, relatively rapidly diluted downstream (Fig. 5). Thus, the metal concentrations at SW444, a bit more than 1 km downstream of the new tailings impoundment, although still noticeably elevated, no longer exceed US EPA water quality standards. Further, the effect on the Bua River (cf. samples Bua 1 and Bua 2) appears to be negligible (Fig. 5). Similarly, water samples collected downstream of the waste rock deposits have low pH and elevated concentrations of Cd, Cu, Ni, Pb and Zn (Table 2). However, low metal concentrations in samples taken some distance downstream (e.g. SW111 and SW1) suggest that significant impacts are

restricted to an area less than a few kilometres downstream of the waste rock deposits.

Water samples collected from the three open pits show that water quality varies significantly between the pits (Table 2). Water in open pits 1 and 3 is of good quality, with near neutral pH, and only modestly elevated levels of metals (although with a fairly high level of Cd being recorded in Pit 1 in 1997). In contrast, the water in open pit 2 has levels of Cd, Cu and Ni that exceed water quality criteria for protection of aquatic life. Further, it is clear that water quality of Pit 2 deteriorated from 1997 to 2002. The pH was 6.7 in 1997, whereas it was 3.1 in 2002. The concentration of Al, Cr, Cu, Mn and Pb increased dramatically during the same period (Table 2). The reason for the poor water quality found in Pit 2 is likely to be a result of the hydrological regime of the pit in combination with the presence of large deposits of sulphidic waste rock on the pit's northern and eastern shores. As Pit 2 has no surface water outflow, the ARD drainage that is formed in these deposits cannot be drained out of the pit. Thus, once the buffering capacity of the water and the sediment in the pit were exhausted, the pH dropped and metals entered into solution. As a result, Pit 2 is devoid of life, whereas the other two pits support aquatic fauna (SGAB-CIE, 2002).

Stream sediment quality data are shown in Table 3. Site SW55 was selected as a background site. However, it can be seen that the levels of As, Cd, Cu, Hg and Zn

Table 2
Analysis results for surface water samples collected in the Giap Lai mining area

Sample	Date	pH	Al ($\mu\text{g l}^{-1}$)	As ($\mu\text{g l}^{-1}$)	Cd ($\mu\text{g l}^{-1}$)	Cr ($\mu\text{g l}^{-1}$)	Cu ($\mu\text{g l}^{-1}$)	Mn ($\mu\text{g l}^{-1}$)	Ni ($\mu\text{g l}^{-1}$)	Pb ($\mu\text{g l}^{-1}$)	Zn ($\mu\text{g l}^{-1}$)
<i>Samples affected by drainage from the new tailings deposits</i>											
TF1	Feb-02	7.0	4.8	0.18	0.023	0.05	1.1	230	1.5	0.078	0.88
TF2	Feb-02	3.1	1700	0.093	0.98	2.4	81	1700	72	3.9	69
TF3	Sep-97	6.8	15	1.1	10	0.89	54	3400	180	<0.20	220
TF4	Sep-97	6.3	390	<4.3	3.7	1.6	35	3300	73	1.5	74
	Apr-01	5.4	1960	2.4	3.9	0.7	290	6700	240	0.97	130
	Feb-02	3.0	26,000	19	3.4	66	720	3000	280	8.1	300
SW44	Apr-01	6.6	190	0.91	2.1	0.027	61	4000	130	<0.030	55
	Feb-02	3.8	4700	0.13	0.73	5	160	980	53	0.28	60
SW444	Apr-01	7.2	17	0.66	0.18	0.072	2.7	730	11	0.21	3.2
<i>Samples affected by drainage from the waste rock deposits</i>											
SW1	Sep-97	7.2	170	0.65	0.077	0.84	0.86	1400	7.1	0.34	7
SW111	Apr-01	6.7	6.3	0.5	0.11	0.026	0.51	5300	10.7	<0.030	4.9
SW2	Sep-97	6.8	880	0.47	0.24	1.5	16	1600	14	0.94	18
	Apr-01	6.5	28	0.32	0.6	0.019	2.2	6300	24	<0.030	65
SW22	Mar-02	6.4	39	0.72	0.18	<0.05	<0.5	7000	18	<0.050	26
SW3	Sep-97	7.2	16,000	<4.8	1.2	12	73	9600	170	9.1	220
SW4	Sep-97	4.6	6700	<4.3	1.9	5.4	160	8300	99	0.57	140
SW8	Sep-97	7.3	1000	1.4	0.12	5.6	14	390	5.3	8.4	19
<i>Samples from the open pits</i>											
OP1	Sep-97	6.7	69	<1.0	1.3	0.69	1.7	1100	31	<0.20	220
	Apr-01	8.0	66	0.38	0.21	0.039	1.4	500	9.1	0.18	12
	Mar-02	7.6	20	0.45	0.15	0.034	0.75	550	6.6	0.074	9.6
OP2	Sep-97	6.7	116	1.2	7.6	0.89	36	6100	250	<0.20	1400
OP2 (a)	Mar-02	3.1	9400	<0.10	3.1	5.2	79	8200	220	4.4	350
OP3	Apr-01	7.7	28	0.56	0.0089	0.12	2.1	100	3.1	0.058	0.7
OP3 (out)	Nov-01	6.9	5.2	1.9	0.0025	0.076	1.2	530	2.6	0.033	1.4
<i>Samples from the Bua River</i>											
BUA1	Apr-01	7.4	6.4	0.23	0.0083	0.19	0.64	15	0.36	0.093	0.7
BUA2	Apr-01	7.4	7.6	0.24	0.014	0.2	0.76	20	0.42	0.14	1.3
<i>Background sites</i>											
OP1 (in)	Nov-01	6.7	190	0.7	0.1	0.36	2	110	1.5	0.94	9.4
OP3 (in)	Nov-01	6.9	112	1.4	<0.002	0.15	3.6	100	0.56	0.26	2.2
SW11	Sep-97	7.0	160	<0.30	<0.020	1.4	0.94	68	1.3	0.68	1.5
SW55	Apr-01	7.0	7.5	0.49	<0.005	0.26	0.43	650	1.2	0.05	11
SW6	Sep-97	7.4	97	<0.30	<0.020	1.2	0.4	82	1.3	0.24	2.1
SW9	Sep-97	7.4	200	0.73	<0.020	1.2	0.84	150	1.7	1.7	2.3
Median background		7.7	112	0.7	0.02	0.36	0.84	110	1.3	0.26	2.3
Background (Förstner & Wittman, 1983)					0.07	—	1.8	—	—	0.2	0.5
US EPA CCC ^a			87	150	0.25	74/11 ^b	9	—	52	2.5	120
WHO			100	10	3	50	2000	400	20	10	3000

Data in bold exceed the chronic level of US EPA quality standards for the protection of aquatic life, whereas data in italics exceed WHO guidelines for drinking water quality. Data in italic bolds exceed both standards.

^aDissolved concentration in water hardness of 100 mg l⁻¹ CaCO₃.

^bCr (III)/(VI).

encountered at this site are high in comparison to the world average for metal concentrations in sediments (Förstner and Wittman, 1983). Downstream of the contamination sources, the metal levels are elevated compared to background. The elevations are, however, modest, and there are only a small number of cases where metal levels exceed US EPA standards. Those highest concentrations are found at

the sites TF4 and SW44 which are both situated downstream of the new tailings facility. The impacts on sediment quality rapidly become less serious further downstream. This is illustrated by two samples taken from the Bua river where the sample taken upstream of the influence of the Giap Lai mining area is near identical to the one taken further downstream (Fig. 6).

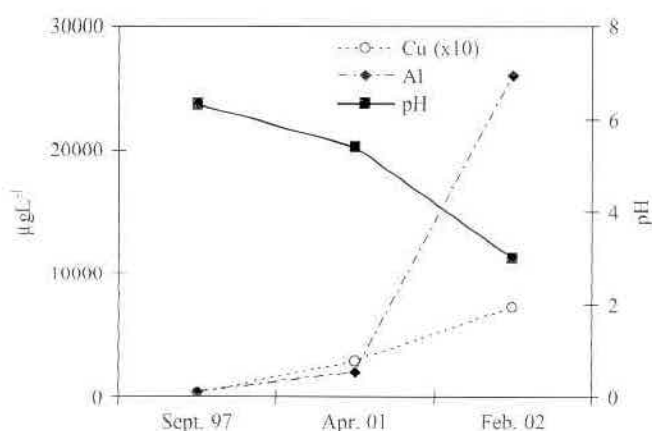


Fig. 4. Graph showing pH and the concentrations of Cu and Al at a site (TF4) downstream of the new tailings deposit.

4.3. Policy discussion

It is shown that sulphide-rich tailings and waste rock in the Giap Lai mining area represent significant sources of ARD, causing elevated metal levels in downstream water bodies, which in turn, represent a threat to both human health and to aquatic ecosystems. The mine site now represents a considerable environmental liability.

Furthermore, as a result of the failure to undertake any efforts to remediate this liability, the impacts are becoming progressively more severe.

In addition to the NLEP, the Mineral Law of 1996 contains environmental regulations that specifically target the mining sector. In this law, it is stated that activities in the sector must use technologies that are environmentally friendly, and that after the completion of mining the affected land and surroundings must be rehabilitated and/or reclaimed. The Mineral Law also states that any organisation or individual involved in a mining venture must deposit funds in an authorised bank account in order to assure that funds for rehabilitation will exist at the end of the mine's life. Additionally, the Vietnamese

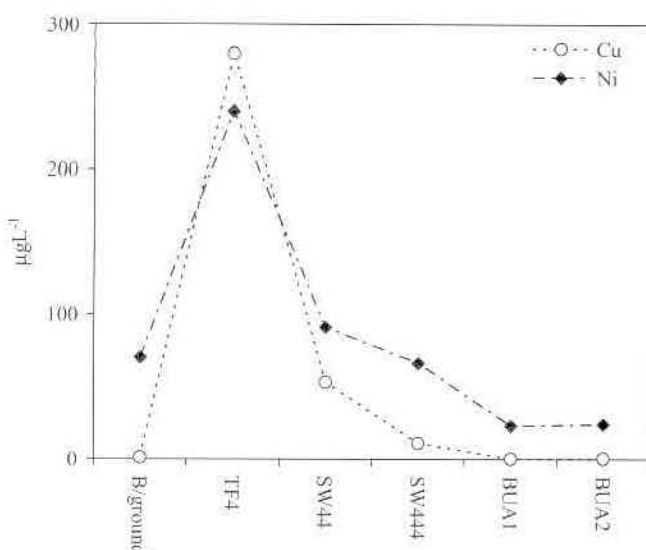


Fig. 5. Cu and Ni concentrations in surface water affected by drainage from the new tailings facility. The sampling locations are plotted in downstream direction, left to right. Samples taken in April 2001.

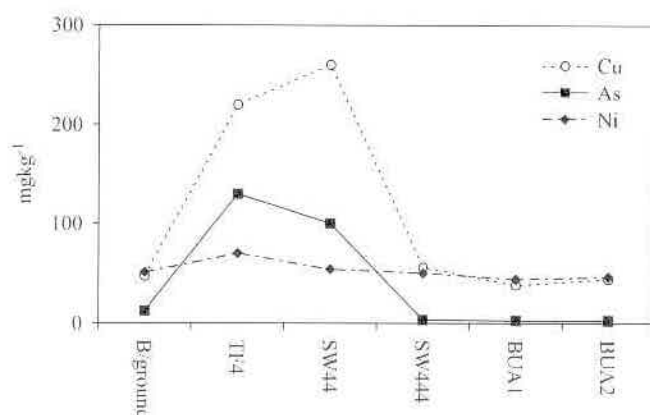


Fig. 6. Cu, Ni and As concentrations in sediments downstream the new tailings facility. Sampling locations are plotted in downstream direction, left to right.

Table 3
The content of metals and arsenic in sediment samples

Sample	As (mg kg ⁻¹)	Cd (mg kg ⁻¹)	Co (mg kg ⁻¹)	Cr (mg kg ⁻¹)	Cu (mg kg ⁻¹)	Hg (mg kg ⁻¹)	Ni (mg kg ⁻¹)	Pb (mg kg ⁻¹)	Zn (mg kg ⁻¹)
TF4	130	0.72	12	99	220	0.042	70	12	120
SW44	100	0.45	11	100	260	0.057	54	21	87
SW444	3.2	0.33	27	100	56	0.044	50	24	98
BUA1	2.4	0.21	23	103	38	0.081	44	23	85
BUA2	2.4	0.32	22	100	44	0.075	46	24	93
OP1	36	0.53	8.9	34	21	0.043	40	12	420
SW2	110	0.67	18	110	110	0.090	54	27	300
SW55	12	0.99	42	69	47	0.12	51	24	220
World average	6	0.35	—	—	30	0.06	—	35	90
US EPA ERM	70	9.6	—	370	270	0.71	52	218	410

World average and US EPA sediment quality guidelines are included for comparison. Sample SW55 represent background conditions. Data in bold exceed US EPA sediment quality guidelines.

Government Circular from 1999 provides further guidelines for how such funds should be set up, and stipulates requirements for mine closure. Further, according to Ministry of Industry staff, these laws and regulations were communicated to all Vietnamese mining operations. It is clear, thus, that a legal mandate exists for environmental protection in the Vietnamese mining industry.

In many countries, informal community pressures have been shown to increase the frequency and effectiveness of environmental supervision and control (Dasgupta et al., 2000; O'Rourke, 2004). Also in the Giap Lai area, during active mining, there were cases where the local community lodged environmentally related complaints to the company; including in these complaints were issues such as contamination of water resources, and agricultural land. In line with Vietnamese policy, these complaints were lodged at the People's Committee, as well as with the Phu Tho DOSTE. According to mine management, the complaints led compensation being paid out for damage caused in cases where drainage from the mining operations found their way into farming areas and domestic wells.

Before closure there were activities undertaken by the Ministry of Industry, which suggest that the need for environmental measures were acknowledged by the mine owners. Investigations were undertaken and needs for mitigation measures identified (SGAB-CIE, 2002). The waste materials were identified as significant environmental liabilities, which needed to be addressed. Importantly, the significance of not allowing the new tailings facility to oxidise and thereby cause ARD was identified at a relatively early stage (SGAB-CIE, 2002). However, no effective measures were subsequently taken. Instead, the new tailings facility which was water covered during operations, was subsequently allowed to drain and partially dry out and, as predicted by leaching experiments, thereby exposing the tailings to extensive oxidation and associated mobilisation of metals. This, in turn, is leading to a progressive worsening of downstream impacts (cf. above). Mine management argued that the lack of appropriate environmental measures was a consequence of "the financial support for addressing these impacts not being forthcoming, neither from the VMC nor the Ministry of Industry...". This statement, in turn, implies that the Giap Lai pyrite company was not prepared to, on its own accord, take on the environmental responsibility that is stipulated in Vietnamese law. The end effect was that no effective post-closure, remediation measures were applied. Possible reasons for this regulatory failure are discussed below.

Tarras-Wahlberg (2002) suggests a number of reasons behind the failure to apply existing environmental regulations in a mining district, some of which may be relevant in Giap Lai area. These include a lack of financial resources on part of the mining companies; lack of awareness of the legislation on part of the mining companies and/or the supervising authorities; inadequate capacity of the governmental and provincial authorities charged with

environmental supervision and control and lack of public participation and pressure in issues related to environmental management. O'Rourke (2004) reports from Vietnam and identifies public involvement in industrial environmental management issues, not including mining, as the key driver in cases where environmental regulations have been successfully applied in this country. The type of public participation identified by the author is, however, not official participation in the regulatory process (which is, in any case restricted by law) but rather a more spontaneous and unofficial form. O'Rourke (2004) identifies a number of further characteristics that are important in these cases where public participation has succeeded in influencing the regulatory process, including community cohesion and mobilisation, and external political linkages (mainly to local government authorities through personal connections).

Even in the Giap Lai case, it is clear that the community has been active and lodged complaints against the company, and as a result some limited compensation was paid out. However, with regards to issues of major environmental importance, that is the adequate and long-term management of the mine's waste products, this public pressure was not sufficient to prompt the company to take appropriate action. Neither was funds deposited for post-closure rehabilitation, in spite of the existence of such a regulatory requirement. O'Rourke (2004) suggests that when a Vietnamese firm has a direct linkage to the state through ownership, this may increase its ability to resist community pressure and the firm thereby not be forced into compliance. Further, the lack of international investments in the Vietnamese mining sector, has meant that the local industry has not become exposed to, and learned from, the significant improvements of environmental performance in mining operations that have been achieved in recent years (c.f. Warhurst and Bridge, 1996; Hilson and Murck, 2000). It is suggested that these facts, coupled with the general finding that mining companies that are not forced to plan for closure, usually do not have sufficient funds available for remediation once a mine is due for closure, are the important reasons for the lack of enforcement of environmental regulations at the Giap Lai mine.

5. Conclusion

Sulphide-rich tailings and waste rock in the Giap Lai mining area represent significant sources of ARD. The ARD thus formed is causing elevated metal levels in downstream water bodies, which, in turn, constitute a threat to both human health and aquatic ecosystems. It is shown that metal concentrations in impacted surface waters have increased after mine closure, suggesting that impacts are becoming progressively more serious.

No post-closure, remediation measures have been applied at the Giap Lai mine, in spite of the existence of environmental legislation and both central and regional

institutions charged with supervision and control. The research presented here provides further emphasis for the recommendation that, while government institutions may need to be strengthened, and environmental regulations need to be in place, true on the ground improvement in environmental quality in Vietnam and in many other developing countries require an increased focus on promoting public awareness and participation in industrial environmental issues.

Acknowledgements

The Swedish International Development Agency (Sida) and the Vietnamese Ministry of Industry are acknowledged for their funding of the project upon which this study is based.

References

- Apodaca, L.E., Driver, N.E., Bails, J.B., 2000. Occurrence, transport, and fate of trace elements, Blue River basin, Summit County, Colorado: an integrated approach. *Environmental Geology* 39 (8), 901–913.
- Azcue, J.M. (Ed.), 1999. *Environmental Impacts of Mining Activities: Emphasis on Mitigation and Remedial Measures*. Springer, Berlin, 300pp.
- Baker, R., 1990. Institutional innovation, development and environmental management: an “administrative trap” revisited. *Public Administration and Development* 9, 29–47.
- Bervoets, L., Solis, D., Romero, A.M., Damme, P.A.V., Ollevier, F., 1998. Trace metal levels in chironomid larvae and sediments from a Bolivian river: impacts of mining activities. *Ecotoxicology and Environmental Safety* 41, 257–283.
- Dasgupta, S., Laplante, B., Maminga, N., Wang, H., 2000. Local enforcement of environmental regulation in China: the role of inspection and complaint. *World Bank Policy Research Working Paper*, World Bank, Washington, DC.
- Environment Australia, 1997. *Managing Sulphidic Mine Wastes and Acid Drainage*, in a Series on Best Practice Environmental Management in Mining. Commonwealth of Australia, 84pp.
- Förstner, U., Wittman, G.T.W., 1983. *Metal Pollution and the Aquatic Environment*. Springer, Berlin.
- Harries, J.R., Ritchie, A.I.M., 1988. Rehabilitation measures at the rum jungle mine site. In: Salomons, W., Förstner, U. (Eds.), *Environmental Management of Solid Waste: Dregred Material and Mine Tailings*. Springer, Berlin, pp. 131–151.
- Hilson, G., 2000. Barriers to implementing cleaner technologies and cleaner production (CP) practices in the mining industry: a case study of the Americas. *Minerals Engineering* 13, 699–717.
- Hilson, G., Murck, B., 2000. Sustainable development in the mining industry: clarifying the corporate perspective. *Resources Policy* 26, 227–238.
- Hilson, G., Nayee, V., 2001. Environmental management system implementation in the mining industry: a key to achieving cleaner production. *International Journal of Mineral Processing* 64, 19–41.
- Lottermoser, B.G., Ashley, P.M., Lawie, D.C., 1999. Environmental geochemistry of the Gulf Creek copper mine area, north-eastern New South Wales, Australia. *Environmental Geology* 39 (1), 61–74.
- Loxham, M., 1988. The predictive assessment of the mitigation of leachate in the subsoils surrounding mine tailing and dredging spoil sites. In: Salomons, W., Förstner, U. (Eds.), *Environmental Management of Solid Waste: Dregred Material and Mine Tailings*. Springer, Berlin, pp. 3–23.
- Martin, A.J., McNee, J.J., Pedersen, T.F., 2001. The reactivity of sediments impacted by metal-mining in Lago Junin, Peru. *Journal of Geochemical Exploration* 74, 175–187.
- O'Rourke, D., 2004. *Community Driven Regulation: Balancing Development and the Environment in Vietnam*. F. MIT Press, Cambridge, 300pp.
- Salomons, W., 1995. Environmental impact of metals derived from mining activities: processes, predictions, prevention. *Journal of Geochemical Exploration* 52, 5–23.
- SGAB-CIE, 2002. *Closure Plan for the Giap Lai Pyrite Mine, Phu Tho Province, Vietnam. Project on Strengthening the Capacity of Environmental Protection in the Mining Sector and Related Activities in Vietnam, 1997–2002*. National Institute of Mining and Metallurgy, Hanoi, 80pp.
- Sobek, A.A., Schuller, W.A., Freeman, J.R., Smith, R.M., 1978. Field and laboratory methods applicable to overburden and mine soils. Report EPA-600/2-78-054, US Environmental Protection Agency, 204pp.
- 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.
- US EPA, 2005a. *Water Quality Criteria*. US Environmental Protection Agency. Available from <<http://www.epa.gov/waterscience/criteria/aqlife.html>>. Retrieved: 07 April 2005.
- US EPA, 2005b. *Sediment Quality Guidelines*. US Environmental Protection Agency. Available from <<http://www.epa.gov/waterscience/es/guidelines.htm>>. Retrieved: 07 April 2005.
- Warhurst, A., Bridge, G., 1996. Improving environmental performance through innovation: recent trends in the mining industry. *Minerals Engineering* 9, 907–921.
- Wedepohl, K.H., 1991. Chemical composition and fractionation of the continental crust. *Geologische-Rundschau* 80 (2), 207–223.
- WHO, 2004. *Guidelines for Drinking-Water Quality*, third ed. World Health Organization, Geneva, Switzerland.

Estimated nationwide effects of pesticide spray drift on terrestrial habitats in the Netherlands

Frank M.W. de Jong^{a,*}, Geert R. de Snoo^b, Jan C. van de Zande^c

^aNational Institute for Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

^bInstitute of Environmental Sciences (CML), Leiden University, P.O. Box 9518, 2300 RA Leiden, The Netherlands

^cPlant Research International, Wageningen UR (WUR-PI), PO-BOX 16, 6700 AA Wageningen, The Netherlands

Received 19 February 2006; received in revised form 14 September 2006; accepted 14 December 2006

Available online 5 February 2007

Abstract

This study estimated the potential effects of pesticide drift on terrestrial ecosystems outside target areas, for the Dutch situation. A series of field trials was conducted to estimate the effects of drift on different species groups at different distances from a treated plot for different categories of pesticide: herbicides, fungicides and insecticides. Measurements of the pesticide drift deposition resulting from standard agricultural practice were used to model deposition outside the treated area. These data were then combined with national statistics on cropland and pesticide use to assess the ecological effects of pesticide drift for the Netherlands as a whole. Three scenarios were considered: the recent past (1998), the present (2005) and an optimised scenario based on 'best available practice' (2010). In the recent past the impact of herbicide drift on sensitive life stages non-target vascular plants is estimated to have exceeded the 50% effect level on 59% of adjacent linear landscape elements such as ditch banks and hedgerows. For the impact of insecticides and fungicides on non-target insects and fungi this 50% effect figure was 29% and 28% of linear elements, respectively. In the present situation, with (narrow) unsprayed buffer zones and other measures in place, these percentages are down to 41% for herbicides, 21% for insecticides and 14% for fungicides. In the optimised scenario, with a greater buffer width of 2.25 m for potatoes (compared to 1.50 m in 2005) and 1 m for other crops (compared to 0.25 and 0.5 m in 2005) and 'best available practice', these percentages can be cut to zero. In natural areas located within farming regions the 10% effect level can be reduced from 31% of such areas (1998) to 0% under conditions of 'best available practice'.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Pesticide; Herbicide; Fungicide; Insecticide; Drift; Side-effect; Non-target area; The Netherlands; Terrestrial habitats

1. Introduction

In the Netherlands, annual agricultural pesticide usage in 2003 stands at about 9.6×10^6 kg active substance (a.s.) (CBS (Statistics Netherlands), 2005a), or an average of 11.8 kg a.s./yr/ha of agricultural land (pastureland excluded). Compared to other European countries this figure is very high: in 2002 mean European usage stood at about 3.5 kg a.s./ha/y (as derived from ECPA, 2003 and FAO data). This high per hectare usage is due to the high

intensity of land use in the Netherlands in combination with climatic conditions, which necessitate intensive fungicide use in potatoes and fruit, and to the considerable area devoted to non-food crops such as flower bulbs and greenhouse horticultural produce, likewise associated with relatively high pesticide usage (De Jong et al., 2001).

A certain fraction of the pesticides used on farmland is dispersed to the surrounding environment (atmosphere, surface waters, soil and ground water). Calculations indicate that in the case of arable land this fraction may be up to 9.5% (De Nie, 2002). The other 90% lands on the crop or on the soil within the cropped area. A major proportion (on average 95% of 9.5%) of the airborne pesticides enters the atmosphere as a result of drift or volatilisation. As a result of this dispersion pesticides are

*Corresponding author. Tel.: +31 30 2743939; fax: +31 30 2744401.

E-mail addresses: frank.de.jong@rivm.nl (F.M.W. de Jong),
snoo@cml.leidenuniv.nl (G.R. de Snoo), jan.vandezande@wur.nl
(J.C. van de Zande).

found in *all* environmental compartments outside treated areas. Concentration measurements in surface waters throughout the Netherlands indicate that, for all categories of pesticide, ecological standards (Maximum Permissible Concentration), defined as the level at which the No Observable Effect Concentration (NOEC) is not exceeded for 95% of species; method by Van Straalen and Denne-
man (1989) adapted by Aldenberg and Slob (1991) are exceeded at over 40% of sampling points (CIW, 2002, www.pesticidesatlas.nl, see: De Snoo et al., 2005a). In agricultural areas, many pesticides are regularly observed in rainwater (Van Dijk and Guicherit, 1999). Pesticides have also infiltrated into the ground water (Notenboom et al., 1999). For the soil itself no quantitative information is available.

Given that pesticides are dispersed to non-target areas in excess of ecological standards, the question arises: what are the effects on biodiversity adjacent to treated lands? In a number of studies effects of pesticide drift on single species have been observed in bioassays (Boatman et al., 1999; Cooke, 1993; Freemark and Boutin, 1995; Marrs and Frost, 1997). Indication of the situation in the Netherlands has already been gained from semi-field and bioassay studies (Brock et al., 2000a,b; De Jong and Bergema, 1994; De Jong, 2001), which yielded clear proof of effects on individual species under experimental field conditions. A second indication is provided by the case of a particular herbicide used to protect wheat against dicotyledonous weeds, where the number of dicotyledonous species on the ditch banks bordering the treated plot was significantly lower than adjacent to the same field with an intervening unsprayed buffer zone 3–6 m wide (De Snoo and Van der Poll, 1999). This kind of information is based on specific case studies, however, and a comprehensive review of the full range and magnitude of side-effects is therefore lacking.

In 1991, a pesticide policy plan was adopted in the Netherlands (MJP-G, 1991), the main aim of which was to reduce pesticide use by 50% in the year 2000 compared to average usage over 1984–1988 (21×10^6 kg a.s., 23 kg a.s./ha). With emission abatement measures also being taken, emissions to the environment were in fact expected to decline by substantially more than 50%. In 2000, this reduction in usage had indeed been achieved, mainly through major cuts in the use of fumigant soil disinfectants. As these latter compounds yield no droplet drift, though, overall pesticide droplet drift has hardly diminished. In 2000, a supplementary policy was therefore instigated aiming to reduce drift by means of such measures as compulsory adoption of unsprayed, crop-free field margins and use of low-drift nozzles (Lozingenbesluit, 2000). In 2004, aims have been defined in the Policy Note “Sustainable Crop Protection”. The main environmental aim of this note is a reduction of the environmental impact of 95% in 2010 compared to 1998.

Dutch policy in this field focuses mainly on the potential effects of pesticide drift on the aquatic environment, i.e.

ditches and groundwater. In the most intensively cropped area of the Netherlands (defined as the area with >50% arable land/km) the total length of ditches is about 33,000 km, i.e. 5 km of ditch per km² (LKN, 1997). There is far less focus on the terrestrial environment, however, and there is still no clear picture of the actual side-effects of pesticide drift on terrestrial organisms. Ditch banks, with their often relatively high biodiversity, clearly extend twice the length of the ditches, giving 66,000 km of ditch banks in all, and a significant number of these border sprayed cropland. In addition, there are about 1000 km of hedge-rows in the same intensively cropped area (LKN, 1997, combined with ‘Top10’ vector file data from the Dutch Topographic Service).

In Europe the main emphasis is also on the aquatic environment. For instance the FOCUS working group on Landscape and Mitigation Factors in Ecological Risk Assessment provides guidance for mitigation measures in the aquatic environment.

This article seeks to provide an estimate of the type and extent of potential side-effects of pesticide drift on terrestrial habitats in the Netherlands. The starting point is field data on pesticide drift, combined with field data on the effects of drift on biodiversity obtained using bioassays. These data are subsequently combined and extrapolated for the whole of the Netherlands for three scenarios: ‘normal agricultural practice’ for the recent past (1998), the present (2005) and the near future (2010). An estimate of effects in the near future has been included to provide an indication of the kind of improvements to be achieved through use of ‘best available techniques’.

2. Methods

The research was essentially conducted in four steps: (i) estimation of drift deposition adjacent to a treated field; (ii) estimation of effects of drift deposition on biota; (iii) modelling of the distance–effect relationship; and (iv) extrapolation of the results to the Netherlands as a whole.

2.1. Drift scenarios

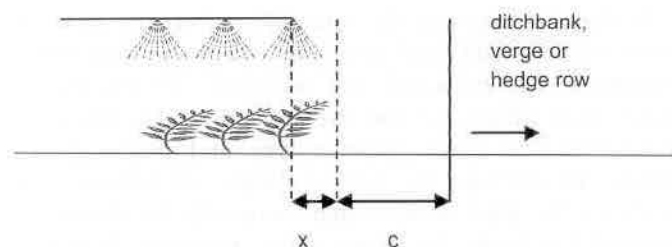
Three drift scenarios were analysed: (i) the recent past (1998); (ii) the present (2005) and the near future (2010). The equipment used and the different buffer zones are summarised in Table 1. In all cases the ‘standard’ conditions were: wind speed 3–4 m/s, boom height 50 cm above crop, nozzle type XR 110.04, ‘medium’ spray quality (Southcombe et al., 1997) and spray volume 300 l/ha.

Using measured pesticide deposition data for the Netherlands as reported by WUR-PRI (Porskamp et al., 1995; Van de Zande et al., 2000a,b) average drift deposition of both vapour and droplets was calculated at various distances from the treated crop. These data are based on tests, performed in a potato crop, with a bare soil surface or a low (grassy or herbaceous) vegetation adjacent to the plot. A model has been constructed for the

Table 1

Specification of the three scenarios (Sources: Lozingenbesluit, 2000; DLV Agricultural Advice Service, 2000)

	Equipment	Buffer zone potatoes	Buffer zone wheat	Buffer zone other crops
1998	Nozzle type XR 110.04 Medium spray quality Southcombe et al. (1997)	75 cm crop free	0 cm crop free	0 cm crop free
		62.5 cm between last nozzle and field edge	25 cm between nozzle and field edge	50 cm between nozzle and field edge
2005	Nozzle type DG 110.04 Coarse spray quality Southcombe et al. (1997)	150 cm crop free	25 cm crop free	50 cm crop free
		137.5 cm between last nozzle and field edge	50 cm between last nozzle and field edge	75 cm between last nozzle and field edge
2010	Nozzle type DG 110.04 Coarse spray quality Southcombe et al. (1997) Air-assistant spraying (Hardy Twin)	225 cm crop free	100 cm crop free	100 cm crop free
		212.5 cm between last nozzle and field edge	125 cm between last nozzle and field edge	125 cm between last nozzle and field edge

Fig. 1. Distance between crop edge and nozzle (x) and crop free buffer zone (c).

relationship between distance from treated fields and deposition of pesticides. To this end we used a power model of the form $y = a(x+c)^{-b}$ (Ganzelmeier et al., 1995), where y is the deposition as a percentage of maximum field dose, x is the distance from the last sprayer nozzle, c is the distance between the last nozzle and the field edge, and a and b are constants (see Fig. 1). For these two constants a best-fit method was adopted. With this equation, at the field edge $x = 0$ m and $c =$ the buffer zone width.

The 2005 scenario is based on the Lozingenbesluit (2000), and for the optimised scenario (2010) we took a buffer zone of 2.25 m (three rows) in potatoes and 1.00 m in other crops, combined with low-drift nozzles and air-assisted spray equipment. This combination of spraying equipment and buffer zone was included to show what can in principle be achieved using the current 'best available techniques'.

In considering the various linear landscape elements bordering arable land (see Section 2.3), there was a need to distinguish between road verges and hedgerows on the one hand and ditch banks on the other. In the case of ditch banks only the terrestrial part was considered (above mean water level). As banks may be rather steep, however, the associated pattern of deposition may be quite

different to that on horizontal habitats such as verges or hedgerows, where the same amount of deposition falls on a greater surface area. For this reason two different models were used: for ditch banks and for other linear elements. To estimate spray drift deposition on ditch banks, data were taken from Porskamp et al. (1995), who measured the drift deposition occurring across a ditch, on both the banks and the water surface. The fraction of spray drift deposited on ditch banks and water compared with horizontal ground at the same distances from the crop was calculated for a 'medium' spray quality nozzle and a sprayer boom height of 0.7 m. These fractions were used to calculate spray drift deposition in ditches from the 1998, 2005 and 2010 drift curves. From these data again a model was constructed for the relationship between distance from treated plot and pesticide deposition.

2.2. Effects measured in bioassays

In a number of field experiments on an experimental plot the effects of pesticide drift on biota were measured using bioassays (see Table 2). The design and results of these tests are described in detail in De Jong and Van der Nagel (1994); De Jong (1998); De Jong and Udo de Haes (2001) and De Jong (2001). Using the average results per pesticide category (herbicides, insecticides and fungicides), models were developed describing the relationship between deposition and effect. During these studies droplet drift was measured. However, using water-sensitive paper, effects were sometimes found at distances at which no droplet deposition was measured, rendering it plausible that parts of the drift (very small droplets and/or vapour) were missed by measuring only droplet deposition. Deposition data derived from the Plant Research International, Wageningen University Research (WUR-PRI) were therefore used for the correlation between deposition and effect. In this

Table 2
Review of experimental data used to establish the drift–effect relationship

Pesticide	Organism	No. of experiments	Effect parameter
<i>Herbicides</i>			
Glyphosate	<i>Brassica napus</i> , <i>Poa annua</i>	2	Growth inhibition of 14-day-old Plants over a 3-week period
Bentazone	<i>Brassica napus</i>	1	
Diquat	<i>Brassica napus</i>	1	
<i>Insecticides</i>			
Diffubenzuron	<i>Pieris brassicae</i>	2	Mortality of 48-h-old caterpillars up to emergence
<i>Fungicides</i>			
Captan	Micro-organisms	1	Number of colonies on agar over a 3-day period

case the standard situation for potatoes in 1998 was used. As comparable spray techniques were used in both experiments, the spray drift deposition and biological efficacy relation could be matched.

2.3. Prediction of effects

The described models were combined, allowing us to predict the effects occurring at different distances from the treated areas both with and without drift-reducing measures. Two types of terrestrial habitat were considered: (i) the most common linear landscape elements bordering directly on arable fields: ditch banks, verges and hedgerows, and (ii) natural areas, bordering directly on arable fields.

Sub (i) In the Netherlands ditch banks and road verges are mowed once or twice yearly, to keep vegetation low. Today, management is geared increasingly towards improving species richness, by removing the mown swath and preventing exposure to fertilisers, among other means. Hedgerows vary considerably in composition and often contain tree species as well as shrubs and lower herbaceous vegetation (Schmitz, 1993). With respect to dimensions, a standard ditch bank of 2.1 m and a ditch of 1 m width were assumed, leading to a total length of 5.2 m along the ditch profile and a total horizontal width of 4 m (cf. Huijsmans et al., 1997). In arable farming areas fields are generally bordered by small roads with relatively narrow verges about 2 m wide. The mean width of these roads was estimated to be 5 m. Hedgerows have an estimated width of 3 m (cf. LKN, 1997) and are assumed to be situated directly adjacent to the sprayed fields.

Sub (ii) To obtain an idea of pesticide impacts on natural areas at greater distances from treated plots national pesticide emission data were used (De Nie, 2002), giving a figure of 0.19 for the proportion of arable farmland bordering on natural areas. As a rough approximation, then, we can assume that $123,000 \times 0.19 = 23,370$ km of arable field margins border on a natural area (see below for calculation of length of arable field borders). In practice, however, there will at least be a ditch between

the field and the natural area, giving a greater distance, here taken to be 4 m.

2.4. Extrapolation to the national level

In the previous step the relationship between distance from treated field and pesticide effects on non-target organisms was estimated and modelled. To extrapolate these local effects to the Netherlands as a whole, a distinction was made between type and magnitude of effects. To estimate the *type* of effects, an estimate was made of the kind of ecosystems generally bordering on sprayed farmland. To estimate the aggregate length of these borders, calculations ran as follows. Ignoring pastureland, almost 90% of Dutch farmland is devoted to arable farming (excluding e.g. green-houses). The size of the average field (apart from maize) on these arable farms is 7.5 ha (CBS (Statistics Netherlands), 1997); assuming average dimensions of 250×300 m, each field then has a perimeter of 1100 m. Since maize is the crop grown on the greatest proportion of Dutch farmland (224,486 ha in 2004; CBS, 2005a), it was considered separately. In contrast to other arable crops, maize is grown mainly for silage fed to livestock. On these farms the average field area is smaller, an average maize field measuring 5.4 ha (CBS, 1997); taking dimensions of 200×270 m gives a perimeter of 940 m. For the country as a whole this means a total length of 123,000 km arable field borders.

CBS statistics show that in 2000 insecticides are used on 43% of arable fields, fungicides on 56% and herbicides on 92% (CBS (Statistics Netherlands), 2005b). This means that herbicide dispersal may potentially affect plants in almost the entire farming area of the Netherlands, with dispersal of insecticides and fungicides affecting invertebrates and fungi in about half that area. Combining these data with the distances at which effects were predicted, the total area affected was then estimated for the three scenarios.

The scenarios were compared for the pesticides use in 1998. This means that the outcomes only show the effects of the measures considered in the scenarios. The actual

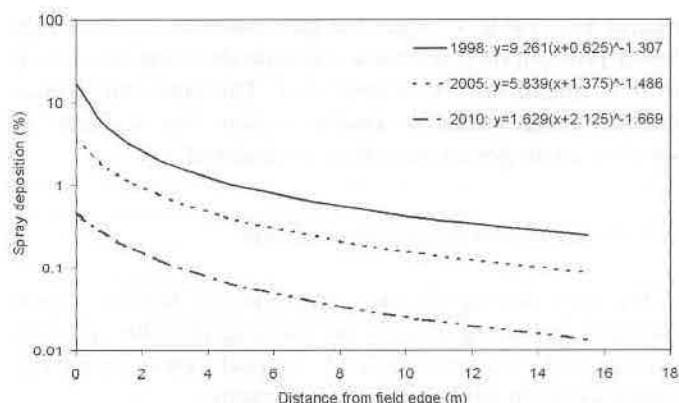


Fig. 2. Results of spray drift deposition measurements and additional modelling, for potatoes, without a ditch (sources: Porskamp et al., 1995; Van de Zande et al., 2000a, b).

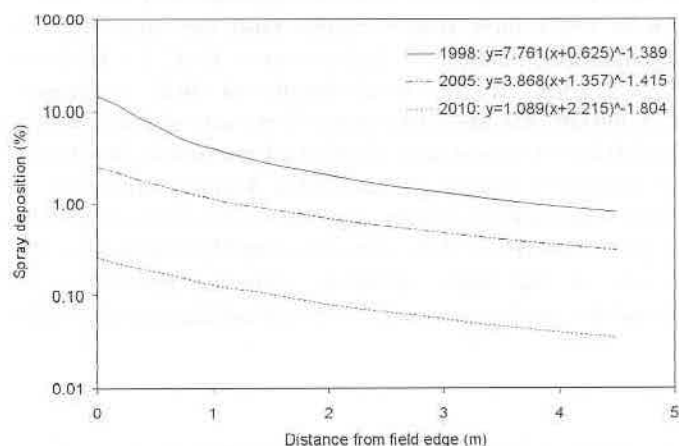


Fig. 3. Drift deposition scenarios, based on measurements and additional modelling, with a ditch; the lines describe deposition on ditch and ditch banks only (sources: Porskamp et al., 1995; Van de Zande et al., 2000a, b).

extent of the effects can deviate from the predictions in the scenarios due to changes in pesticide.

3. Results

3.1. Drift scenarios

Figs. 2 and 3 show the empirical relationship between pesticide deposition and distance from field edge for a potato crop, with and without an adjacent ditch. The formulae describing the relationship are given in Table 3.

For deposition along a sprayed field in the situation without a ditch (Fig. 2) data were available for both the standard (1998) and the 2005 and 2010 scenarios and these data were modelled. The formulae thus derived are representative for potatoes. For other crops similar formulae were derived by substituting smaller buffer zones in the formulae (see Table 3).

The models for drift deposition on the ditch banks, derived from the data of Porskamp et al. (1995), are shown in Fig. 3. As different buffer zones are prescribed for

Table 3

Models used to simulate drift deposition in the situation with and without a ditch

		Without ditch	Deposition on ditch banks
Potatoes	1998	$y = 9.26(x + 0.625)^{-1.31}$	$y = 7.76(x + 0.625)^{-1.39}$
	2005	$y = 5.84(x + 1.375)^{-1.49}$	$y = 3.87(x + 1.375)^{-1.42}$
	2010	$y = 1.63(x + 2.125)^{-1.67}$	$y = 1.09(x + 2.125)^{-1.80}$
Wheat	1998	$y = 9.26(x + 0.25)^{-1.31}$	$y = 7.76(x + 0.25)^{-1.39}$
	2005	$y = 5.84(x + 0.50)^{-1.49}$	$y = 3.87(x + 0.50)^{-1.42}$
	2010	$y = 1.63(x + 1.25)^{-1.67}$	$y = 1.09(x + 1.25)^{-1.80}$
Other crops	1998	$y = 9.26(x + 0.50)^{-1.31}$	$y = 7.76(x + 0.50)^{-1.39}$
	2005	$y = 5.84(x + 0.75)^{-1.49}$	$y = 3.87(x + 0.75)^{-1.42}$
	2010	$y = 1.63(x + 1.25)^{-1.67}$	$y = 1.09(x + 1.25)^{-1.80}$

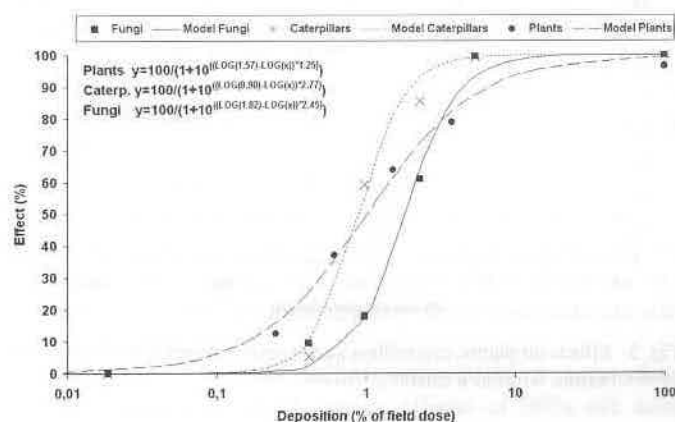


Fig. 4. Relationship between deposition of herbicides, insecticides and fungicides and effects on, respectively, plants, caterpillars and fungi (sources: De Jong and Van der Nagel, 1994; De Jong, 1998; De Jong, 2001; De Jong and Udo de Haes, 2001).

potatoes, wheat and other crops, three different models were used, as summarised in Table 3.

3.2. Effects measured in bioassays

Fig. 4 summarises the results of bioassay experiments. As described in the section method, the models of Table 3 were used for the calculation of the deposition at different distances.

In Fig. 4 the effects on plants are percentage growth inhibition, effects on fungi are reduction in number of colonies on agar and effects on caterpillars are mortality. These differences in effect parameter partly explain the differences between the plots. Effects on plant growth are relatively higher at low concentrations and relatively lower at high concentrations than in the case of caterpillar mortality. For all three types of effect, EC_{50} is between 0.9% and 1.5% of the maximum field dose.

3.3. Prediction of effects

In the next step the correlation between distance and deposition was substituted into the deposition model.

Fig. 5 shows the situation for 1998 without a ditch, for example. Similar figures were prepared for the two other scenarios (2005 and 2010) and the other crops.

From Fig. 5 (and the formulae) the distances at which given effects occur can now be derived. These results are summarised in Table 4.

As mentioned, the buffer zones and therefore models for potatoes, wheat and other crops differ. Separate area figures were therefore needed for the respective crops; from national farming statistics we took a split of 20% potatoes and 24% wheat and 56% other crops (CBS, 2005a). The

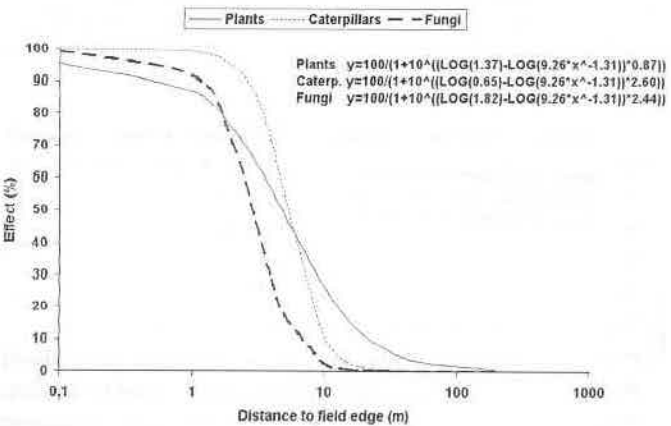


Fig. 5. Effects on plants, caterpillars and fungi combined with drift for the 1998 situation without a ditch.

Table 4
Calculated distances (Edist.) from field edge at which there is a 90%, 50% and 10% effect in the three scenarios for plants, caterpillars and fungi, with and without a ditch; — = no effect found adjacent to treated field

Scenario	Plants			Caterpillars			Fungi		
	Edist90 (m)	Edist50 (m)	Edist10 (m)	Edist90 (m)	Edist50 (m)	Edist10 (m)	Edist90 (m)	Edist50 (m)	Edist10 (m)
<i>Without ditch</i>									
1998	0.8	5.0	22.3	2.8	5.5	10.4	1.3	3.0	6.4
2005	0.2	2.4	10.6	1.2	2.7	5.2	0.4	1.4	3.2
2010	—	0.1	2.7	—	0.1	0.9	—	—	0.3
<i>With ditch (estimate of Edist. on ditch banks)</i>									
1998	0.7	3.9	> 5.2	2.2	4.2	> 5.2	1.0	2.4	5.0
2005	—	1.8	> 5.2	0.8	2.0	4.2	0.2	0.9	2.4
2010	—	—	1.5	—	—	0.4	—	—	—

Table 5
Length (in km) of linear elements potentially affected by pesticide use (herbicides, insecticides and fungicides); only elements directly adjacent to fields were included (thus ignoring the ditch bank or verge on the other side of the ditch or road)

	Total length alongside arable fields in the Netherlands (10 ³ km)	Part of total length of elements in the NL (%)	Length affected (10 ³ km)		
			Herbicides	Insecticides	Fungicides
Ditch banks	81.2	24	74.7	34.9	45.5
Verges	40.6	25	37.3	17.4	22.7
Hedgerows	1.2	7	1.1	0.5	0.7

figures and Table 4 reflect the fact that there is relatively less deposition on ditch banks because the same amount of drift is spread over a sloping area. The table shows that drift-reducing measures clearly reduce the distance at which a given percentage effect is observed.

3.4. Extrapolation to the national level

We here distinguish two categories of habitat: linear elements bordering directly on the treated fields (ditches, verges and hedgerows) and natural areas generally separated from fields by a linear element.

3.4.1. Linear elements

Since a field is a geographically separate unit, each field is bordered by ditch banks, verges or hedgerows. Data from LKN (a database with information on a km² basis, LKN, 1997) show that in arable land the ratio between ditch banks, verges and hedgerows is 66:33:1. Applying these ratios to the total length of field perimeter: 123,000 km (see Methods) gives 81,000 km of ditch banks, 40,500 km of verges and 1200 km of hedgerows bordering on intensively cropped fields. Table 5 shows the length of linear elements potentially exposed to the different groups of pesticides. These data were calculated by multiplying the length of the linear elements with the percentage of pesticides used, e.g. on 92% of the adjacent arable field

of the 81.2×10^3 km of ditch banks result in a potentially 74.7×10^3 km of ditch banks affected.

The mean dimensions of ditches and ditch banks were assumed to be 1 and 2.1 m, respectively. In arable farming areas fields are generally bordered by small roads with relatively narrow verges about 2 m wide. The mean width of these roads was estimated to be 5 m. Hedgerows have an estimated width of 3 m (cf. LKN, 1997).

Multiplying the pesticide effect distances (Table 4) by the length of adjacent habitats (Table 5) yields the total pesticide-affected area for the three different effect levels; the results are shown in Table 6. As an example, 74.7×10^3 km ditch banks, potentially affected by herbicides (Table 5) were multiplied by the 0.7 m effect distance for 90% effect for herbicides in 1998 (Table 4), resulting in the 4975 ha of ditchbanks affected (differences due to rounded figures in Tables 4 and 5). When the distance at which effects were found exceeded the width of the element in question, the element width was used to calculate the affected area. In the case of insecticides and hedgerows, for instance, in 1998 the 10% and 50% effect ranges were in excess of 3 m, the maximum width of the hedgerows. Therefore, in this case the maximum possible area of 158 ha

was within effect range, leading to the same value for the two effect levels.

For the percentage of areas of linear elements affected we related the areas affected to the linear elements in those square kilometre grid cells with over 10% arable land. This selection meant that a total of 17,000 km² were accounted for (~40% of the Netherlands). For the Netherlands as a whole the percentages are lower and of course cannot be higher than the figures cited in Table 5.

The main overall conclusion from Table 6 is that herbicides have the greatest ecological impact. This was indeed to be expected, as herbicides are used on 92% of arable fields. In the 1998 scenario the EC₅₀ for herbicides is exceeded in about half the ditch banks and verges, while for hedgerows the figure is even higher: over 90%. The figures are lower for insecticides and fungicides, mainly because they are used relatively less, and are estimated at 33% and 28% of ditch banks and 21% and 28% of verges, respectively, for the two groups of compound. For hedgerows these figures are higher: 43% for insecticides and 56% for fungicides.

In the present situation (2005) the results of the drift-reducing measures already in place are plain to see, but pesticide effects are still clearly present. This can be explained by the fact that although effect distances are reduced (see Table 4) most adjacent linear elements are still within the effect range.

With 'best available techniques' (2010) a very clear improvement was found; severe effects of 90% are now absent altogether, while the 50% effect level is exceeded only for insecticides, on a limited area (up to 2%). The 10% effect level is exceeded for herbicides over a considerable area (up to 33% in the case of ditch banks). To provide a compact overview, the data for the different linear elements are summed in Table 7, aggregating the different linear elements. As can be seen, although pesticide effects on adjacent habitats are down in the 2005 scenario,

Table 6
Area and percentage of linear elements in farming regions with >10% arable land potentially affected by pesticide use; 10%, 50% and 90% indicate the percentage effect estimated

Ditch banks		Herbicides		Insecticides		Fungicides	
		ha	%	ha	%	ha	%
1998	90%	4975	15	7330	21	4656	14
	50%	21,658	64	11,170	33	9548	28
	10%	31,368	92	14,661	43	18,168	53
2005	90%	72	0	2731	8	831	2
	50%	13,118	38	6926	20	4058	12
	10%	31,368	92	11,730	33	9548	28
2010	90%	0	0	0	0	0	0
	50%	0	0	0	0	0	0
	10%	11,219	33	1326	4	0	0
<i>Verges</i>							
1998	90%	3134	19	3491	21	2914	18
	50%	7468	46	3491	21	4546	28
	10%	14,937	92	6981	43	4546	28
2005	90%	741	5	2175	13	951	6
	50%	7468	46	3491	21	3130	19
	10%	14,937	92	3491	21	4546	28
2010	90%	0	0	0	0	0	0
	50%	0	0	245	2	0	0
	10%	0	0	1517	9	642	4
<i>Hedgerows</i>							
1998	90%	95	26	146	40	88	24
	50%	339	92	158	43	206	56
	10%	339	92	158	43	206	56
2005	90%	22	6	66	18	29	8
	50%	275	74	143	39	95	26
	10%	339	92	158	43	206	56
2010	90%	0	0	0	0	0	0
	50%	0	0	7	2	0	0
	10%	0	0	46	12	19	5

Table 7
Summarised effects of pesticide use on adjacent linear landscape elements, expressed as percentage of linear element area in farming regions for the three scenarios

Scenario	% of linear elements affected		
	90% effect	50% effect	10% effect
<i>Herbicides</i>			
1998	16	59	92
2005	2	41	92
2010	0	0	22
<i>Insecticides</i>			
1998	22	29	43
2005	10	21	29
2010	0	0	6
<i>Fungicides</i>			
1998	15	28	45
2005	4	14	28
2010	0	0	1

Table 8
Natural area affected (10% effect) by pesticide use, in hectares and as a percentage relative to the total extent of natural area in farming regions

	Herbicides		Insecticides		Fungicides	
	ha	%	ha	%	ha	%
1998	39,346	31	6431	5	3141	2
2005	14,190	11	1206	1	0	0
2010	0	0	0	0	0	0

they are still present, while the optimised scenario for 2010 shows that considerable additional reductions can be achieved using best available techniques.

3.4.2. Natural areas

As shown in Section 2.3, 23,370 km of arable field margins border on a natural area. On this basis, Table 8 shows the extent of natural area estimated to be affected by pesticide drift. In this table only the 10% effect level is shown, on the reasoning that in natural areas any ecological effects are undesirable. The percentages shown are relative to the total extent of natural area in regions with over 10% arable land (128,000 ha, LKN, 1997). As an example, for the herbicides in 1998, the effect distance for 10% effect is 22.3 m (Table 4). Assuming that a 4 m ditch will be present between the treated parcels and the nature area, the distance affected within the nature area is 22.3–4 is 18.3 m. This figure is multiplied with the length of the border with the nature areas (23,370) and the percentage at which herbicides are used (92%), resulting in 39,346 ha potentially affected. For an indication of the percentage affected this area was related to the surface of the nature areas adjacent to the treated parcels, for this example resulting in $39,346/128,000 = 31\%$.

In the 1998 scenario 10% effects are found in natural areas at levels of up to 31% for herbicides. Now the 2005 scenario is already characterised by markedly less impact, because a distance of 4 m between field and natural area was assumed and here the distance at which effects are observed is generally less than 4 m. In the optimised scenario for 2010 the 10% effect level is no longer exceeded anywhere.

4. Discussion

The scenarios presented in this paper compare the effects of drift-reducing measures on the pesticides used in 1998. In the paper it was not accounted for changes in use, available pesticides, partly because these data are not available for the 2010 scenario, and partly because we aimed to show the effects of the measures apart from other changes. This means that the absolute data can deviate from the data presented in this paper.

The aim of this research was to estimate the extent of severe impacts on sensitive life-stages of organisms in

habitats close to pesticide-treated arable land, effects confirmed in earlier bioassay studies (Cooke, 1993). Other studies have shown that leaving the outer part of the crop unsprayed increases the number of plant species found in the boundary vegetation (Fischer and Milberg, 1997; De Snoo and Van der Poll, 1999). In several European countries an overall decline in natural values in agricultural areas has been reported. A number of factors have been implicated in this decline, including nitrification, desiccation due to lowered water tables and habitat destruction, but the role of pesticides has never been quantified. Perhaps remarkably, though, there is very little field evidence of effects actually occurring. In broad terms, this could be for two reasons: the effects are there but good field studies are lacking; or the effects are absent because models are overestimating them.

One possible explanation is that ecosystems close to arable fields contain relatively few species, which are well adapted, robust and have a high capacity for recovery. Other activities like ditch cleaning, mowing, ploughing and manure application will certainly also play a role. In the case of plants, a field study has shown that clear short-term effects are observed at very low pesticide dosages, but that these effects do not have any long-term impact at community level (De Snoo et al., 2005b).

Another possible explanation is that effects were overestimated in our study, based as it was on modelling and a number of assumptions, the principal of which will now be discussed.

The first set of assumptions concern estimated pesticide drift. Here, reported British and Dutch data are of the same order of magnitude (Cooke, 1993). The Dutch data have been validated several times and are specific to the Dutch situation, moreover. Our calculations of mean drift took into account average weather conditions, spray volume, tractor driving speed and distance between spray boom and crop. Although these factors may vary considerably at the local level, use of the mean will not lead to overestimation of overall effects. An important remaining source of variation is sprayer nozzle type. Certain nozzle designs cause from 2.5 less to 10 times more drift than the standard nozzle assumed for the year 2005 (derived from Van de Zande et al., 2000a). In practice, then, local effects may be considerably more or less severe than those presented here. However, at the national level we assume the approach adopted here is more reliable. Variation can also be caused by label instructions, concerning measures for specific applications. For different compounds different label instructions concerning drift-reducing measures can occur for different applications.

Another important source of variation are the effect parameters chosen. In the present study we focused on effects on sensitive life-stages of sensitive organisms for the pesticide in question. For each pesticide a wide range of organisms will be potentially affected and these will exhibit negligible up to extreme sensitivity to the compound in question (cf. Davis et al., 1991; Fletcher et al., 1990).

Bearing in mind the many different types of pesticides in use (in the Netherlands 224 different active ingredients in 2006), some of them sometimes applied in combination as well as repeatedly within a given season (in the case of certain fungicides, 10–15 times) and the wide variety of non-target organisms, we can assume that the situation presented here is on the precautionary side but is still representative for a significant number of species. It is also noteworthy that the EC_{50} values found for the specific species considered lie within a range of 0.9–1.5% of the maximum field dose, or between 4 and 8 m from the treated plot (1998 scenario, no ditch). In terms of the distances at which effects are observed and predicted, these results are in line with other, bioassay studies that measured the effects of pesticide drift (e.g. Cooke, 1993; Marrs and Frost, 1997).

A third source of uncertainty is the estimation of areas of linear and natural elements. In the former case an average figure was taken for element width, in the full knowledge that local differences may be substantial. However, these differences work both ways. In some cases the road or ditch will be narrower than the average value taken and the opposite ditch bank or verge will lie in the effect range, but the opposite may also hold. Again this source of variation may be of major influence at the local level, but at the national level it will not lead to any overall over- or underestimation. Another important factor is that only part of the adjacent landscape elements or natural areas will in fact be downwind of the field at the time of spraying, and so our estimates of the area affected are probably too high. On the other hand, multiple treatments were omitted from the calculations, and wind direction may well vary from treatment to treatment and multiple treatments may lead to repeated exposure.

Further validation is required on all these aspects and there is certainly due scope for refining our models. Also an uncertainty analyses of a number of data as discussed above, is recommended in a next stage of this research. However, the aim of this study has been to integrate drift data with effect data, to provide a first-pass impression of the type and extent of the effects concerned.

Although the assumptions and sources of uncertainty discussed above are certainly no reason to reject the general outcome of this study, due care should still be taken in using the results, which still need to be validated in the field.

5. Conclusions

In the recent past (1998) herbicide use exceeded the 50% effect level on sensitive life stages of non-target plants on 59% of the area occupied by linear landscape elements adjacent to Dutch farmland. For insecticides the percentage area affected was 29% (14,819 ha) and for fungicides 28% (14,300 ha). As the various categories of pesticide are not necessarily applied on the same fields, the total percentage area affected may well have exceeded 59%. In

the case of insecticides there was even 90% effect on 22% of these linear elements (10,967 ha). In previous decades these percentages were probably higher still, given heavier pesticide use and less concern to reduce drift.

In 2005 the 50% effect level was still being exceeded on 41% of these adjacent linear elements in the case of herbicides. Although drift had been reduced through use of low-drift nozzles, adjacent hedgerows, ditches and verges were still within effect range. For insecticides the percentage affected was down to 21% for this 50% effect level (10,560 ha), but 10% of the linear elements were still within the 90% effect range (4972 ha).

If low-drift nozzles and air-assisted equipment are combined with broader unsprayed buffer zones, as assumed for the near future, pesticide impacts will be substantially reduced below the 1% effect level.

In 1998, 31% of natural areas lying in farming regions were affected by herbicides to at least the 10% effect level (39,346 ha). Policies now in place have reduced this figure to 11% (14,190 ha) and in the optimised scenario for 2010 the 10% effect level is no longer exceeded for any of the compounds.

Introduction of unsprayed buffer zones in the Netherlands in 2005 has already clearly reduced pesticide impacts on linear landscape elements adjacent to cropland. The main benefits are short range, though, and measures to reduce the amount of drift itself (low-drift nozzles, air-assisted spraying) will be beneficial over the longer range, too.

Even with the drift-reducing policies recently introduced in the Netherlands, it is still anticipated that linear landscape elements and natural areas will be adversely affected. Only with stricter measures in place, as scheduled for the near future, will the effects of pesticide drift outside sprayed areas be significantly reduced.

References

- Aldenberg, T., Slob, W., 1991. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. Report 719102002, RIVM, Bilthoven, Netherlands.
- Boatman, N.D., Davies, D.H.K., Chaney, K., Feber, R., De Snoo, G.R., Sparks, T.H. (Eds.), 1999. Field Margins and Buffer Zones: Ecology, Management and Policy. *Asp. Appl. Biol.* 54, AAP, Wellesbourne.
- Brock, T.C.M., Lahr, J., Van den Brink, P.J., 2000a. Ecological risks of pesticides in freshwater ecosystems: part 1: herbicides. Report 088, Alterra, Wageningen, Netherlands.
- Brock, T.C.M., Van Wijngaarden, R.P.A., Van Geest, G.J., 2000b. Ecological risks of pesticides in freshwater ecosystems: part 2: insecticides. Report 089, Alterra, Wageningen, Netherlands.
- CBS (Statistics Netherlands), 1997. *De Landbouwtelling 1997*. Elsevier, Doetinchem.
- CBS (Statistics Netherlands), 2005a. *Land- en tuinbouwcijfers 2005*. LEI, Wageningen, CBS, Voorburg, Netherlands.
- CBS (Statistics Netherlands), 2005b. CBS website <www.cbs.nl>, 20-12-2005.
- CIW, 2002. *Bestrijdingsmiddelenrapportage 2002. Het voorkomen van bestrijdingsmiddelen in het Nederlandse oppervlaktewater in de periode 1999 en 2000*. Werkgroep V van de Commissie Integraal Waterbeheer, The Hague, Netherlands.

- Cooke, A.S., 1993. The Environmental Effects of Pesticide Drift. English Nature, Peterborough, UK.
- Davis, B.N.K., Lakhani, K.H., Yates, T.J., 1991. The hazards of insecticides to butterflies of field margins. *Agriculture Ecosystems and Environment* 36, 151–161.
- De Jong, F.M.W., 1998. Development of a field bioassay for the side-effects of pesticides on decomposition. *Ecotoxicology and Environment Safety* 40, 103–114.
- De Jong, F.M.W., 2001. Field trials for side-effects of pesticides. Thesis, Leiden University, the Netherlands.
- De Jong, F.M.W., Bergema, W., 1994. Field bioassays for side-effects of pesticides. Report 112, CML, Leiden, Netherlands.
- De Jong, F.M.W., Udo de Haes, H.A., 2001. Development of a field bioassay for the side-effects of herbicides on vascular plants using *Brassica napus* and *Poa annua*. *Environmental Toxicology* 16, 397–407.
- De Jong, F.M.W., Van der Nagel, M.C., 1994. A field bioassay for side-effects of pesticides on *Pieris brassicae*. *Mededelingen Faculteit Landbouwwetenschappen Universiteit Gent* 59/2a, 347–356.
- De Jong, F.M.W., De Snoo, G.R., Looij, T.P.J., 2001. Trends of pesticide use in the Netherlands. *Mededelingen Faculteit Landbouwwetenschappen Universiteit Gent* 66/2b, 823–834.
- De Nie D.S., 2002. Emissie-evaluatie MJP-G 2000: achtergronden en berekeningen van emissies van gewasbeschermingsmiddelen. Report no. 716601004/2002, RIVM, Bilthoven, Netherlands.
- De Snoo, G.R., Van der Poll, R.J., 1999. Effect of herbicide drift on adjacent boundary vegetation. *Agriculture Ecosystems and Environment* 73, 1–6.
- De Snoo, G.R., Tamis, W.L.M., Van't Zelfde, M., 2005a. Dutch pesticides atlas for surface water. *Pesticides News* 68, 8.
- De Snoo, G.R., Tamis, W.L.M., Van de Brink, P.J., 2005b. Non target plant field study: effects of glufosinate-ammonium on off crop vegetation. Report no. 161, CML, Leiden, Netherlands.
- DLV [Agricultural Advice Service], 2000. Gewasbescherming in 2000 in de akkerbouw en veehouderij. DLV, Assen, Netherlands.
- ECPA, 2003. ECPA Statistical Review 2002. ECPA, Brussels.
- Fischer, A., Milberg, P., 1997. Effects on the flora of extensified use of field margins. *Swedish Journal of Agricultural Research* 27, 105–111.
- Fletcher, J.S., Johnson, F.L., McFarlane, J.C., 1990. Influence of greenhouse versus field testing and taxonomic differences on plant sensitivity to chemical treatment. *Environmental Toxicology and Chemistry* 9, 769–776.
- Freemark, K., Boutin, C., 1995. Impacts of agricultural herbicide use on terrestrial wildlife in temperate landscapes: a review with special reference to North America. *Agriculture Ecosystems and Environment* 52, 67–91.
- Ganzelmeier, H., Rautmann D., Spangenberg R., Streloke M., Hermann M., Wenzelburger H.J., Walter, H.F., 1995. Untersuchungen zur Abtrift von Pflanzenschutzmitteln. Ergebnisse eines bundesweiten Versuchsprogrammes. Mitteilungen aus der Biologische Bundesanstalt für Land- und Forstwirtschaft, Heft 304, Berlin.
- Huijsmans J.F.M., Porskamp, H.A.J., Van de Zande, J.C., 1997. Spray drift reduction in crop protection application technology. Evaluation of spray drift in orchards, field crops and nursery tree crops spraying (state-of-the-art December 1996). Institute of Agricultural and Environmental Engineering, IMAG-DLO Report 97-04, Wageningen, Netherlands (in Dutch with English summary).
- LKN, 1997. Landschaps Ecologische Atlas van Nederland. CD-ROM, PUDDOC-DLO, Wageningen, Netherlands.
- Lozingenbesluit, 2000. Lozingenbesluit open teelt en veehouderij. Staatsblad van het Koninkrijk der Nederlanden 43, 1–114.
- Marrs, R.H., Frost, A.J., 1997. A microcosm approach to the detection of the effects of herbicide spray drift in plant communities. *Journal of Environmental Management* 50, 369–388.
- MJP-G, 1991. Meerjarenplan Gewasbescherming. Parliamentary Proceedings, Netherlands Second chamber, session 1990–1991, 21677, 3–4, SDU, The Hague, Netherlands.
- Notenboom, J., Verschoor, A., Van der Linden, A., Van der Plassche, E., Reuther, C., 1999. Pesticides in ground water: occurrence and ecological impacts. Report 601506002, RIVM, Bilthoven, Netherlands.
- Porskamp, H.A.J., Michielsen J.M.G.P., Huijsmans J.F.M., Van de Zande, J.C., 1995. Emission-reducing pesticide application in potato growing. The effects of air assistance, nozzle type and spray-free zone on the drift deposition outside the field. Institute of Agricultural and Environmental Engineering, IMAG-DLO Report 95–19, Wageningen, Netherlands (in Dutch, with English summary).
- Schmitz, H., 1993. Houtwallen, heggen en singels. Stichting Landelijk Overleg Natuur en Landschapsbeheer, Utrecht, Netherlands.
- Southcombe, E.S.E., Miller, P.C.H., Ganzelmeier, H., Van de Zande, J.C., Miralles, A., Hewitt, A.J., 1997. The international (BCPC) spray classification system including a drift potential factor. In: *Proceedings of the Brighton Crop Protection Conference—Weeds*, 1997. November 1997. Brighton, UK, pp. 371–380.
- Van de Zande, J.C., Porskamp, H.A.J., Michielsen, J.M.G.P., Holterman, H.J., Huijsmans, J.F.M., 2000a. Classification of spray applications for driftability, to protect surface water. *Aspects of Applied Biology* 57, 57–65.
- Van de Zande, J.C., Michielsen, J.M.G.P., Stallinga, H., De Jong, A. de, 2000b. Ranking of low-drift nozzles and air assistance for spray drift. Paper 00-PM-066 presented at AgEng 2000, Warwick, UK, 2000, 9pp.
- Van Dijk, H.F.G., Guicherit, R., 1999. Atmospheric dispersion of current-use pesticides: a review of the evidence from monitoring studies. *Water, Air, and Soil Pollution* 115, 21–70.
- Van Straalen, N.M., Denneman, C.A.J., 1989. Ecotoxicological evaluation of soil quality criteria. *Ecotoxicology and Environmental Safety* 18, 241–251.

Modelling the response of surface water quality to the urbanization in Xi'an, China

Hongming He^{a,b,*}, Jie Zhou^a, Yongjiao Wu^{b,c}, Wanchang Zhang^d, Xiuping Xie^{a,b}

^aState Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, P.O. Box 17, Xi'an 710075, China

^bGraduate School of Chinese Academy of Sciences, China

^cInstitute of Salt Lake, Chinese Academy of Sciences, Xining, Qinghai 810008, China

^dInternational Institute for Earth System Science, Nanjing University, Nanjing 210093, China

Received 1 July 2005; received in revised form 14 December 2006; accepted 30 December 2006

Available online 6 March 2007

Abstract

The study investigated the response of surface water quality to urbanization in Xi'an, China. We qualitatively described the change in urban land use from 1996 to 2003, analyzed the status of the surface water environment, and constructed a model of urban expansion to simulate the water environment's response to urbanization. Our results revealed that patterns of land use changed dramatically, the rate of economic growth exceeded that of urbanization during the study period, and increasing urban land use was correlated with fluctuations in water quality. The simulated results suggested that urbanization had reached the environmental carrying capacity based on the average land utility and the marginal costs of pollution.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Urbanization; Xi'an; Land-use change; Surface water quality

1. Introduction

The most intensive interactions between human beings and the environment take place in cities and surrounding areas. Land use and changes in land cover affect the functions and processes of the ecological landscape (Hietel et al., 2004). Urbanization changes the land use, while at the same time bringing about increasing environmental loads created by expanding use of energy and resources and increasing impacts on human health and the ecosystem. Urban sprawl is a key environmental issue that must be monitored and predicted if we pursue sustainable development. Study of the causes, processes, and consequences of land-use change is a major research topic in landscape ecology (Wu and Hobbs, 2002). An ecology perspective for

urban ecosystems is both appropriate and imperative (Sheryl and Wu, 2004). Studies on the impacts of urbanization on environmental systems typically correlate changes in environmental systems with simple aggregate measures of urbanization (Alberti, 2003). Study of water quality in urban areas revealed different behaviours (Yoshimura et al., 2005; Chang, 2005; Mvungi et al., 2003). For example, although Japanese rivers are still heavily impacted by canalization, loss of most dynamic flood plains, flow regulation, invasion by exotic species, and intensive urbanization, river water quality has improved remarkably during the past decades through major restoration initiatives to improve ecological integrity of river ecosystems (Yoshimura et al., 2005). In South Korea, water quality dramatically declined in the middle section of the Han River where the river receives inputs from polluted tributaries (Chang, 2005). In Harare City, Zimbabwe, water quality in a tributary of the Marimba River, Harare deteriorated due to home industries and led to urgent demands for urban water management (Mvungi et al., 2003).

In recent years, rapid economic growth in the city of Xi'an, China has been accompanied by rapid urbanization.

*Corresponding author. State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, P.O. Box 17, Xi'an 710075, China. Tel.: +86 29 88322651; fax: +86 29 88320456.

E-mail address: hehoming@yahoo.com (H. He).

The gross domestic product (GDP) of Xi'an was 8.95 billion USD (American dollars) in 2001, which is equivalent to 1297 USD per capita, and the GDP is expected to grow to 14.3 billion USD by 2005. Since 1996, the urban area has expanded by 90.05 km² (at a rate of 6.4% per year), from 166.4 km² in 1996 to 256.45 km² in 2003. Simultaneously, the urban population increased from 2.3 million (37.3% of the local total) in 1990 to 3.12 million (43.6%) in 2003 (Xi'an Statistical Yearbook, 2003). However, the urban and rural ecological environments have both been deteriorating. During this rapid urbanization, surface water pollution has become a serious issue. It is important to analyze the impacts of urban expansion on surface water quality and changes in land use in Xi'an so we can propose strategies that will permit sustainable regional development and protection of the ecological environment. This present study represents a preliminary investigation of the relationships between urbanization and surface water quality in Xi'an. We began with an examination of the rate of urbanization and the changes in land use within Xi'an between 1996 and 2003. We then analyzed the relationship between surface water quality and urbanization. Finally, we constructed a model of urban expansion to simulate and predict the environmental carrying capacity (the maximum or optimum amount required to support a healthy ecological environment and socioeconomic community) in Xi'an.

2. Study area

Xi'an is the capital of Shaanxi Province and is one of northwestern China's key cities from the perspective of politics, the economy, and culture. It lies in the centre of the Guanzhong Plain (107°40'E–109°49'E, and 33°39'N–34°44'N, within the larger Weihe Plain), bounded by the Qinling Mountains in the south and the Loess Plateau in the north. Xi'an has 14 administrative districts and counties. The most urbanized regions have a total area of 363.66 km². Baqiao, Beilin, the hi-tech zone, Lianhu, Weiyang, Xincheng, and Yanta represent the main part of the urban area. Around these urban areas lie Changan and Lintong, two regions that are vigorously developing their potential and showing strong urban-rural interactions. The Gaoling, Huxian, Lantian, Yanliang, Changan, and Zhouzhi areas surround Xi'an and have predominantly rural characteristics (Fig. 1).

The river networks of Xi'an consist of nine main rivers, including the Weihe River, Jinghe River, Tianyu River, Chanhe River, Bahe River, Fenghe River, Yuhe River, Zhaohe River, Heihe River (Fig. 1). These rivers provide the industrial and residential water supply for Xi'an. All rivers belong to the lower Weihe River drainage system, but most riverchannels are hundreds of kilometres long and run through forested hilly land to a considerable distance from the city. This creates certain difficulties in evaluating the surface water quality of the total watershed from the viewpoint of urbanization. To study the response of surface water quality to urbanization, we selected 21

observation sites that covered these rivers from upstream to downstream within the area of Xi'an City.

3. Materials and methods

3.1. Data collection

To identify dynamic changes in the landscape that occurred between 1996 and 2003, we determined changes in land use from remote-sensing satellite images. We chose the two LANDSAT5 TM satellite images with the best quality during this period (26 May 1996 and 29 May 2003; orbit 127/36), and combined this information with a map of the area's land types at a scale of 1:550 000 and a 30-m spatial resolution digital elevation model (DEM, serving as a superimposed layer with satellite images for land use cover change interpretation) provided by the Institute of Remote Sensing Application of the Chinese Academy of Sciences.

Data from our 21 river observation sites covered the abovementioned nine rivers. Sampling sites were selected along the middle of the Weihe River (sites 2, 5, 7, and 8; sites 66, 67, and 68), the Chanhe River (sites 11, 12, and 17), the Bahe River (sites 20, 21, 24, and 29), the Fenghe River (sites 31, 34, and 36), the Yuhe River (sites 41, 44, and 47), and the Zhaohe River (sites 53, 57, and 59). We obtained water quality data for parameters such as chemical oxygen demand (COD), biological oxygen demand (BOD₅), KMnO₄, volatile hydroxybenzenes, and total phosphorus which were collected twice monthly during 1996 and 2003 from the Xi'an Municipality Bureau of Environment Protection (Xi'an Municipality Bureau of Environment Protection, 2003). Water quality samples were measured by the standard analytical method for surface water quality provided by the China State Environmental Protection Administration (Environmental Standards for Surface Water, 1988). Water samplings were compared to the values of the environmental quality standards for surface water (GB3838-88) issued by the China State Environmental Protection Administration (Environmental Standards for Surface Water, 1988). We used monthly mean precipitation data provided by Xi'an Municipality Bureau of Meteorology to address the relationships between natural environmental change and surface water quality in Xi'an. We obtained statistical socioeconomic data for Xi'an from the Xi'an Yearbook (Xi'an Statistical Yearbook, 2003) to allow us to discuss the roles of human activities in the process of urban expansion.

3.2. Changes in land use and GDP

Using the two abovementioned LANDSAT5 TM satellite images, we mapped land use in 1996 and 2003 using the Normalized Difference of Barren Index (NDBI) and an artificial neural network (ANN) (Rumelhart and McClelland, 1986). After pre-processing the image data (orthographic correction and atmospheric correction), we processed the spatial data on urban expansion as follows.

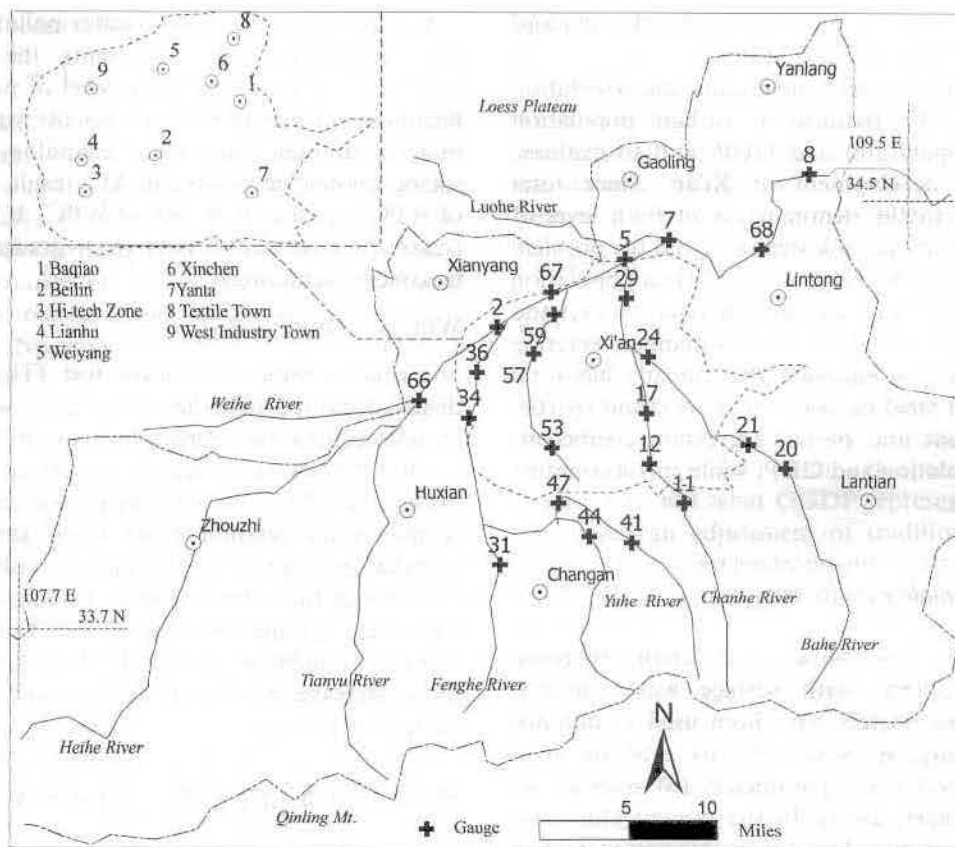


Fig. 1. Sketch map of study area: river network and water quality sampling sites in and around Xi'an.

First, we used NDBI (Zha et al., 2003) to identify the urban boundaries at different times. NDBI is calculated as $(\text{band } 5 - \text{band } 4) / (\text{band } 5 + \text{band } 4)$. We then mapped land use in 1996 and 2003 using an ANN to adjust the weight of the multiple-layer perception in the Land Use Cover Change (LUCC) classification of satellite images (Rumelhart and McClelland, 1986; Wu and Martin, 2002). By mapping land use in 1996 and 2003, we provided the inputs for the ANN used in this study based on the six bands of the TM image. As a result, there are six nodes in the input layer and we only employed one hidden layer in the neural network. Since land use was classified into six types (such as urban, rural, farmland, vegetation, water, and uncultivated), there were also six output layer nodes. Based on these parameters, we processed the 1996 and 2003 images using the BP neural network method (Rumelhart and McClelland, 1986). Spatially explicit models of land-use change focus on the rate of change between two or more classes, the locations of changes to one or more classes, or the rates and locations of changes between two or more classes (Turner, 1987; Lopez et al., 2001). As a result, a Markov Chain transition matrix can be used to analyze the relationships between these changes. The Markov Chain methods give the probability that a process at time t will be in state k , given that at time $t-1$, the process was in state h . That is

$$\text{Prob}(X_t = a_k | X_{t-1} = a_h). \quad (1)$$

This model predicts the estimated change in land use from non-urban to urban, which typically represents an impervious surface that cannot absorb water (Cathlin, 1999).

It is interesting to study the economic parameter referred to as "land utility" during the process of urbanization (Brown, 2003). Land utility is usually defined as usefulness of land quality in land-use type evaluation. Land capacity analysis (Manasquan River Watershed Management Group, 1999) focuses mostly on the attributes of the land and its potential uses. In the present paper, we have defined land capacity as the optimum amount of land to support a healthy ecological and socio-economic community, defined land utility as its usefulness for ecological and human activities, and calculated the usefulness as a dollar value in market. We introduced a land utility index (LUI) to represent the value of land for different combinations of land types. LUI is calculated as follows:

$$\text{LUI} = \sum_{i=1}^n P_i X_i. \quad (2)$$

Here, X_i (USD) is the utility of land type i in any year, and P_i is the weight of land type i . We assume that the maximum value of LUI corresponds to the optimized value of land utility. We completed the overall data processing using the analytic hierarchy process (AHP). This analysis

used the expert-consulting graded approach (Mustafa and Al-Bahar, 1991) to calculate land utility.

We used a regression model to investigate the association between the level of urbanization (urban population divided by total population) and GDP, and to evaluate the socioeconomic development of Xi'an. Since total population appears in the denominator of both level of urbanization and GDP, we took steps to avoid the problem of spurious correlation. Namely, we used urban population as the response variable (Y), and total population (x_1) and GDP (x_2) as predictor variables. In this manner, entering GDP into the regression equation that already has total population in it will yield measures of association (partial correlation coefficient and partial regression coefficient) between urban population and GDP, while still accounting for total population.

3.3. Surface water quality evaluation

Many case studies have shown that urban expansion often leads to problems with surface water quality. Although various parameters have been used to monitor surface water quality, it is difficult to evaluate river pollution using only specific parameters (Mvungi et al., 2003; Nafu and Geiger, 2004). In the present study, we combined our monitoring data for many parameters to produce a Comprehensive Pollution Index (CPI) and Pollution Load (PL). CPI is defined as follows:

Pollutant load:

$$\left\{ \begin{array}{l} PI_j = \frac{1}{n} \sum_{i=1}^n PI_{ij} \\ K_j = \left(PI_j / \sum_{j=1}^m PI_j \right) 100\% \Rightarrow K_j = \left(\frac{1}{n} \sum_{i=1}^n PI_{ij} / \sum_{j=1}^m PI_j \right) 100\%, \end{array} \right. \quad (3)$$

$$PI_{ij} = \begin{cases} \frac{C_i}{C_{im}} & \text{if } C_{im} = C_{\text{lower limit}}, \\ \frac{|C_i - C_{\text{median}}|}{|C_{\text{upper limit}} - C_{\text{lower limit}}|} & \text{if } C_{\text{lower limit}} \leq C_{im} \leq C_{\text{upper limit}}, \\ \frac{C_{i\text{max}} - C_i}{C_{i\text{max}} - C_{im}} & \text{if } C_{im} = C_{\text{upper limit}}, \end{cases} \quad (4)$$

where i is the evaluated parameter, j is the gauge code, n is the total number of parameters evaluated, C_i is the mean value of item i , C_{im} is the value of evaluated item i specified in the water quality standard, $C_{i\text{max}}$ is the maximum value of item i , PI_{ij} is pollution index of item i in gauge j , PI_j is pollution index of all evaluated items in gauge j , and K_j is pollution load of gauge j .

The pollutant complex pollution index (PCPI) is defined as the comprehensive pollution index of one pollutant at one sample site or in one river, and the river comprehensive pollution index (RCPI) is the comprehensive pollution index of all pollutants at one sample site or in one river.

We calculated the cost of water pollution (WPC) under two situations. WPC represents the water pollution treatment costs under a fixed level of pollution treatment technology, whereas WPC' represents water pollution costs under a gradually improving technology that results from socioeconomic development. As a result, the rate of growth of WPC is greater than that of WPC'. If T_2 is the pollution treatment cost that results from gradual innovations in treatment technology:

$$WPC' = f(WPC + T_2). \quad (5)$$

Spearman's rank correlation test (Thomas, 2001) is a distribution-free test that determines whether there is a monotonic relation between two variables (X , Y) and offers a good way to study trends in surface water quality for urban rivers based on the CPI parameter described above. A monotonic relation exists when any increase in one variable is invariably associated with an increase or decrease in the other variable. In equation form, for the pairs (X_1 , Y_1) and (X_2 , Y_2): if $X_2 > X_1$ then $Y_2 \geq Y_1$ for a monotonic increase; if $X_2 < X_1$ then $Y_2 \leq Y_1$ for a monotonic decrease. Spearman's rank correlation coefficient is defined as follows:

$$R_s = 1 - \left[6 \sum_{i=1}^N d_i^2 \right] / [N_3 - N], d_i = X_i - Y_i, \quad (6)$$

where, d_i represents the algebraic difference between variables X and Y , X_i represents the increase in pollutant concentration, Y_i represents the time sequence, and W_p represents the critical value and should be indexed from a statistical table of Spearman rank correlation coefficients. If $R_s \geq W_p$, then R_s represents a distinctive shift, and when R_s has a negative value, this means a decreasing trend.

3.4. Model construction: average land utility vs. marginal water pollution costs

Urban models are often developed to explain and simulate urban dynamics (Allen et al., 1984; Batty, 1971; Johnston and Barra, 2000; Robert and Tomas, 2000; Stephan et al., 2005; Irene et al., 2003; Pacione, 2003; Brown, 2003; Wegener, 1994). These models focus separately on the socioeconomic and environmental behaviors during urban expansion, and fail to provide an integrated ecological view of urban expansion. Therefore, we developed our own model of urban environmental change for the Xi'an City Development Scheme. In the present study, we constructed a mathematical model to quantify the responses of surface water quality to urbanization.

According to social welfare maximization theory (Negishi, 1960), "social welfare" refers to the overall utilitarian state of society, and is defined as the sum of the welfare of each individual in the society. Welfare can be adequately measured in monetary terms. The social welfare function is a function of individual welfare, and can be defined as the expression of

the objectives of a society. For example:

$$W = Y_1 + Y_2 + \dots + Y_n,$$

where W is the social welfare and Y_n the income of the n th individual in a society. In this example, maximizing the social welfare function means maximizing the total income of the people in the society, without regard to how incomes are distributed in society (Negishi, 1960). However, according to social welfare maximization theory, welfare is maximized when those society members who had the least utility reach the greatest utility. No economic activity will increase social welfare unless it improves the position of the poorest society member. Average utility (AV) is the average total income of all individuals. The marginal cost (MC) defines the increase in cost per each unit increase in consumption, and is expressed as the first derivative of the input equation, and represents the consumption of input required to generate additional income. The basic welfare economics problem is to find the point of intersection (also called the “critical point”) between average utility and marginal cost, and to define the theoretical maximum for the social welfare function subject to various constraints such as the state of production technology and the available natural resources.

Accordingly, we defined the maximum social welfare as the point where the average land utility (AV (LU)), the average total USD of a combination of different land-use types equals the marginal water pollution cost (M (WPC), the increase in USD cost of water pollution per unit GDP increase), and equilibrium is reached. As M (WPC) is an increasing function, and AV (LU) is a decreasing function, then AV (LU) should exceed M (WPC) cost before the two reach their point of intersection, and environmental quality should remain healthy. After this critical point, urbanization will create increasingly negative values of social welfare and will deteriorate surface water quality. We should stop urban expansion before this critical point is reached, but to do so, we must first try to locate the balance point. In theory, we could rationalize the critical point according to optimized schemes for Xi'an urban development. We could then compare the real value at any point in time with the value at the critical point. The first step in such an analysis is to establish a mathematical relationship between land utility (LU) and the land utility index (LUI) and another relationship between GDP loss (GDPL) and M (WPC). Here, LUI/LU is regarded as the average utility. The derivative of GDPL with respect to water pollution cost (WPC) equals the marginal cost.

Likewise, we can calculate LU under two conditions. In the first condition, LU represents the land utility under the assumption of fixed technologies for a given level of GDP whereas in the second, LU' represents the land utility under the assumption of gradual improvements in technologies for a given level of GDP that result from socioeconomic development. Therefore, the values of LU are lower than those of LU'. If T_1 represents the land utility effect that results from improvements in technologies for a given level

of GDP, then:

$$LU' = f(LU + T_1). \quad (7)$$

As described above, we assume that LU is a function of different land types (LUCC) and that WPC, which reflects the status of surface water quality, is a function of the treatment costs for the various river pollutants. Thus:

$$LU = f(LUC_1, LUC_2, \dots, LUC_n), \quad (8)$$

$$WPC = f(Pollu_1, Pollu_2, \dots, Pollu_n). \quad (9)$$

In order to evaluate the social welfare provided by urbanization from the perspective of environmentally sustainable development, we introduce the concept of Green GDP (GGDP). Although specialists hold different opinions on just what GGDP represents, it can simply be considered an adjustment of traditional GDP that subtracts resource and environmental costs from the sum of all economic activities that are included in GDP (World Bank, 1997). Indeterminate factors such as pollution may be aggravated or ameliorated by natural influences and can complicate the calculation of GGDP. We recognize that improvement of the surface water environment is achieved not only by water treatment technology and technologies for a given level of GDP, but also by policy and public awareness. We integrated these factors, as well as potential unexpected pollution events and other factors related to water quality improvement, in our calculation of GDPL. We have tried to quantify these factors as accurately as possible to permit the use of GGDP in our study (Hotelling, 1932). Suppose that GGDP can be expressed as follows:

$$GGDP = GDP - H - E_p - C_{pd},$$

where H represents Hotelling rents for depleted resources, E_p is expenditures on pollution controls, and C_{pd} is the (external) costs from pollution and environmental degradation.

For the convenience of model construction, we have defined GDP as the total of GGDP and GDPL, where GDPL represents the treatment costs for surface water pollution in the analysis described in the present study. Thus:

$$\begin{cases} GGDP = GDP - GDPL, \\ GDPL = f(WPC), \end{cases} \quad (10)$$

$$\begin{cases} LU = f(LUI), \\ WPC = f(LUI). \end{cases} \quad (11)$$

The equilibrium between the land utility and the marginal WPC occurs at the intersection between these two functions (i.e., at the critical point):

$$\begin{cases} \frac{f(LUI)}{LUI} = \frac{\partial f(WPC)}{\partial WPC} \times \frac{\partial WPC}{\partial LUI}, \\ WPC = f(LUI). \end{cases} \quad (12)$$

4. Results

4.1. Land use changed dramatically and land capacity improved slowly during urbanization

One of the most impressive characteristics of urbanization is the resulting change in spatial patterns (Charbonneau et al., 1993; Treitz et al., 1992). To ensure the authenticity of the urban boundary, we used NDBI to extract the urban boundary of Xi'an using the LANDSAT images. We coded regions with NDBI values between -0.44 and 0.10 (i.e., non-urban areas) in black, and coded regions with values between -0.10 and 0.16 (i.e., urban

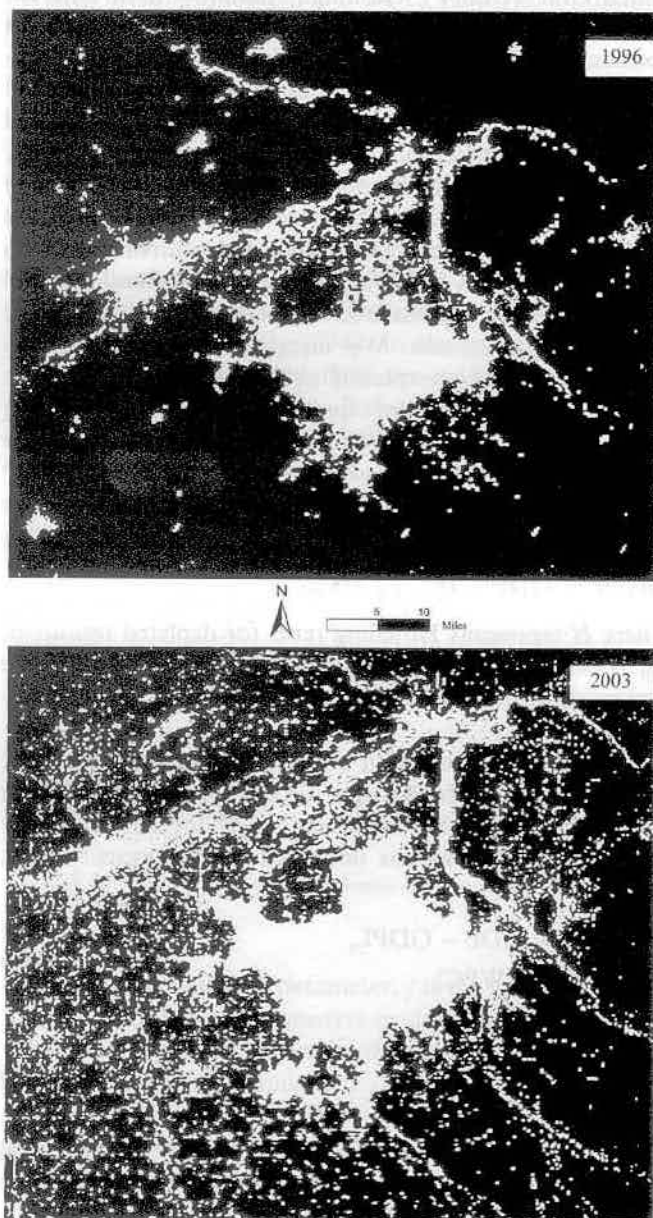


Fig. 2. Results of the NDBI classification of urban and non-urban area in 1996 and 2003. Large blocks of white pixels represent urban areas, black pixels represent non-urban areas, and scattered white pixels represent urbanizing areas that exhibit rural-urban interactions.

areas) in white. The resulting land-use maps processed using the BP neural network method are shown in Fig. 2, which depicts the urban and non-urban areas in the Xi'an municipal region in 1996 and 2003. In these images, large white blocks represent the main urban areas, and smaller white spots represent urbanizing areas. Conversely, large black blocks represent non-urban areas, most of which were found in rural and mountainous regions. By analyzing the urban boundaries of Xi'an in 1996 and 2003, it is clear that Xi'an has undergone rapid urban expansion. To reflect the urban-rural interaction process based on LUCC in Xi'an, we classified land-use types as six types, such as urban, rural, farmland, vegetation, water, and uncultivated. Table 1 and Fig. 3 show the results of the classification. It can be seen that the areas classified as urban increased greatly, whereas the area of uncultivated and rural land decreased rapidly. The construction of urban infrastructure increased by 90.05 km^2 , while vegetation, farmland, water, uncultivated and rural decreased by 4.26 , 18.45 , 0.05 , 40.14 , and 27.21 km^2 , respectively. From 1996 to 2003, the urban area expansion from 166.4 to 256.45 km^2 represented an average growth rate of 6.4% per year.

Table 2 presents the actual Markov transition matrix for land-use changes based on the classification results in Table 1. From 1996 to 2003, subtracting the area of unchanged land (the sum of diagonal values in the table, 212.74 km^2) from the total area (363.66 km^2) gives a total area of 150.92 km^2 that transitioned from one land use to another. During 1996–2003, urban areas had the largest area (153.62 km^2 , the first number in Table 2), and had the smallest possibility of transition ($1.00 - 0.92 = 0.08$). This result suggests that urban land was the dominant type of land use during the urbanization process. The Markov transition matrix displayed interactive transition behaviours (that is, the transition possibility between one land use and another was affected by the transition probability to a second land use) among the types of land use. As expected, the model showed both increases in some land-use types and decreases in others during the process of urbanization of the urban area, 102.37 km^2 were created by transitions from rural (38.49 km^2), uncultivated (29.24 km^2), farmland

Table 1

Land-use change classification of satellite images of Xi'an in 1996 and 2003

Land use type	1996		2003	
	Area (km^2)	Percent of the total	Area (km^2)	Percent of the total
Urban	166.40	46	256.45	70
Vegetation	41.08	11	36.82	10
Farmland	42.21	12	23.76	7
Water	2.72	1	2.67	1
Uncultivated	48.42	13	8.28	2
Rural	63.13	17	35.92	10
Total	363.96	100	363.96	100

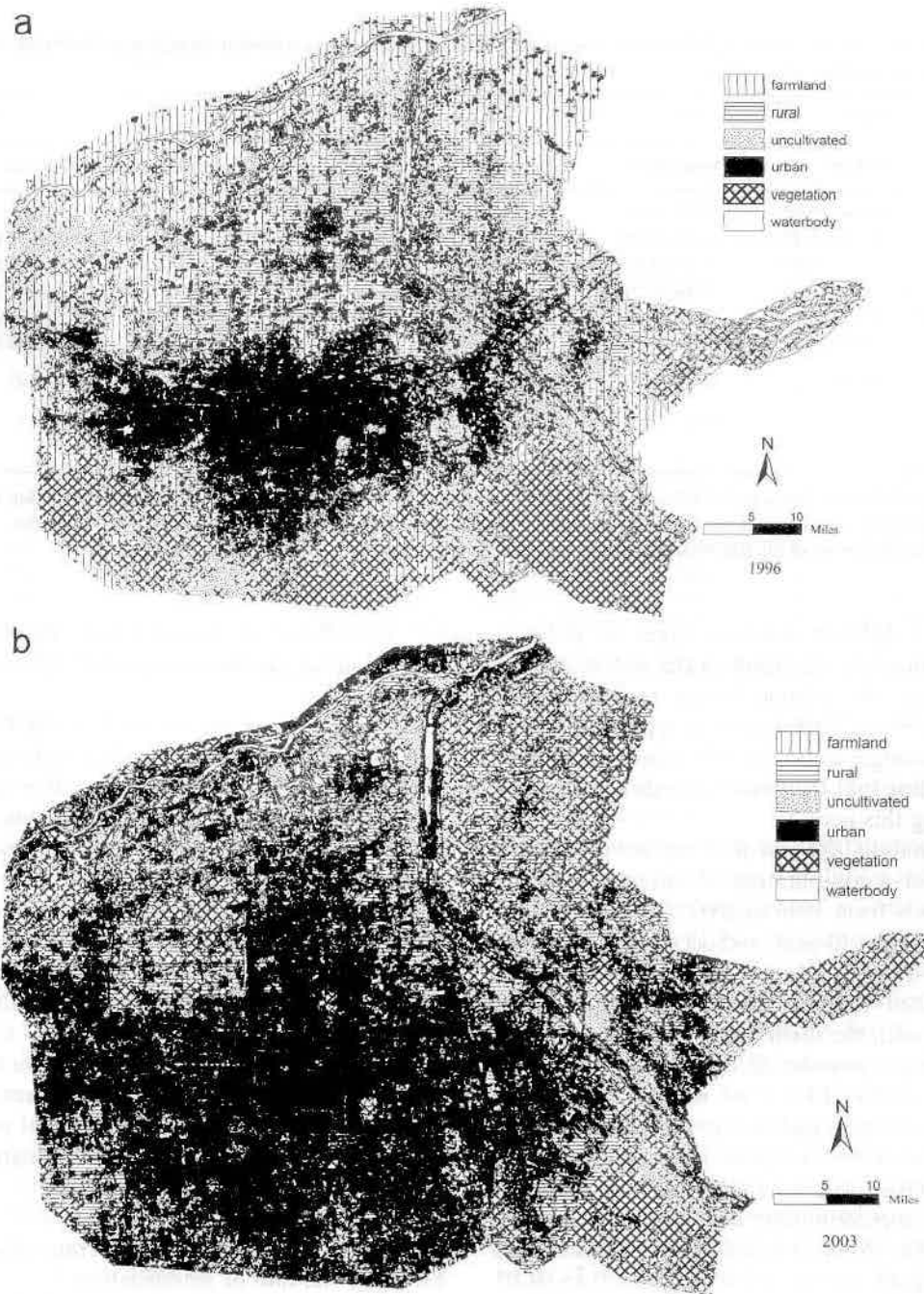


Fig. 3. Results of classification of satellite images of land use in the Xi'an area in 1996 and 2003.

(23.77 km²), vegetation (10.38 km²), and water (0.49 km²), and 12.3 km² were transitioned into the five other land-use types). The most conspicuous expansions of urbanization occurred in the northern part of the city (Weiyang District) and the southern part of the city (Yanta District and Changan District). By comparison, there was relatively little change in the eastern and western parts of the city. The urban expansion in the northern and southern parts of Xi'an resulted from the creation of the Xi'an High Technology Industry Zone and the Xi'an Economy & Technology Development Zone, respectively, in these two regions. Infrastructure associated with large-scale construction per-

mitted rapid urban expansion in these development zones. Cultivated land decreased rapidly in extent, and was largely replaced by this urban infrastructure. Some urban areas transitioned into areas of vegetation (4.21 km²), farmland (3.81 km²), water (0.22 km²), and cultivated (0.2 km²) and rural (3.86 km²) land. Urban areas planted new vegetation and created new bodies of water as city lakes, and were replaced by rural land, farmland, and uncultivated land along rivers in flooded areas. Theoretical Markov transition matrix (matrix based on Markov transition theory) represents relationships between land-use changes. These values were statistically independent, and integrated the transition

Table 2

Markov transition matrix for land-use change in Xi'an from 1996 to 2003; values in parentheses represent transition probabilities from the land use in the 1996 column to the land use in 2003 (units: km²)

		2003						1996
		Urban	Vegetation	Farmland	Water	Uncultivated	Rural	
1996	Urban	153.62 (0.92)	4.21 (0.03)	3.81 (0.02)	0.22 (0.001)	0.2 (0.001)	3.86 (0.02)	165.92 (1)
	Vegetation	10.38 (0.25)	24.61 (0.60)	2.18 (0.05)	0.32 (0.008)	0.13 (0.003)	3.73 (0.09)	41.35 (1)
	Farmland	23.77 (0.56)	1.27 (0.03)	12.17 (0.29)	0.26 (0.006)	0.16 (0.004)	4.58 (0.11)	42.21 (1)
	Water	0.49 (0.18)	0.64 (0.24)	0.06 (0.02)	1.3 (0.48)	0.08 (0.03)	0.15 (0.06)	2.72 (1)
	Uncultivated	29.24 (0.6)	3.88 (0.08)	1.73 (0.04)	0.45 (0.009)	5.76 (0.12)	7.85 (0.16)	48.91 (1)
	Rural	38.49 (0.61)	1.49 (0.02)	4.45 (0.07)	0.37 (0.006)	2.47 (0.04)	15.28 (0.24)	62.55 (1)
2003		256.45	36.82	23.76	2.67	8.28	35.92	
Theoretical transition probability		0.62	0.04	0.08	0.03	0.09	0.14	363.66

The diagonal transition probabilities less than 1.0 indicate that an area of one land-use type in 1996 was not 100% likely to remain in that land-use type in 2003; the remaining transition probabilities in that row represent the probabilities of transition into other land-use types. Theoretical transition probabilities represent the likelihood of change to the specified land-use type.

probabilities among different land-use types in different years. This matrix presents the trend in the distribution of land-use types under the driving forces responsible for urbanization from 1996 to 2003. Land-use types with higher values represent a stronger trend towards transition to these land types. It was clear that the trend towards urbanization was strongest during this period.

Land capacity analysis showed that the environmental carrying capacity for a combination of different land use types improved slowly from 1996 to 2003. According to the municipal government's 20-year socioeconomic development plan (2000–2020), Xi'an plans to become an ecologically urbanized tourism city. Urban construction will reach 780 km², with the main urban area expanding to 660 km² and attaining a population of 7.6 million by 2021. To calculate land utility (LU), we invited five experts (three university researchers and two government officials) to grade land use in Xi'an at 3-year intervals from 1991 to 2021 on the basis of the land's potential carrying capacity under the socioeconomic conditions expected to occur at that time. The experts graded the X_{ij} values of each land type and assigned weights of 0.40, 0.20, 0.15, 0.10, 0.10, and 0.05 to urban, plantation, orchard, woodland, water, and uncultivated land use, respectively. From this process, we obtained LUI values for each year. These values fell between 0 and 10. We assumed that the five average LUI values (one per expert) represented the final results. Table 3 lists these values. The experts clearly agreed that land utility was low during the 1990s, and that land utility developed towards moderate levels in the 2000s. However, the rate of increase in land utility would not stop after 2015 under the current 20-year plan, and by the end of the 2020s, it would reach a relatively high level. The increasing LUI values reflect the belief that land utility would gradually increase from low levels toward high levels during the process of ecologically inspired urbanization.

4.2. GGDP growth increased more slowly than GDP growth: urban expansion negatively effected the environment

Growth rates of factors such as GDP, of the structures and economic products of various industrial sectors, and of the urban and rural populations offer significant evidence of urbanization. The GDP of Xi'an was 8.95 billion USD (1297 USD per capita) in 2001, and is expected to grow to 14.3 billion US Dollars in 2005. Table 3 presents the changes in overall GDP, in the economic products for three industrial sectors (agriculture, industry, and services), and the investment in fixed assets (FAI) from 1994 to 2003. The urban population increased from 2.3 million (37.3% of the local total) in 1990 to 3.12 million (43.6%) in 2003 (Xi'an Statistical Yearbook, 2003). Regression analysis revealed a positive relationship ($R^2 = 0.89$) between urban population as the response variable, and the total population (x_1) and GDP (x_2) as the two predictors of Xi'an:

$$y = 99.4x_1 + 660.36x_2 - 1849.6. \quad (13)$$

This equation shows that the rate of economic growth exceeded the rate of urbanization.

The use of GGDP statistics, which include not only capital produced by humans but also the costs imposed by environmental and ecological factors, could paint a more comprehensive and realistic picture of the social economy of an area that depends on natural capital, since environmental pollution and ecological degradation can both decrease GDP. However, there are different methods to calculate GGDP. The World Bank reported in 1997 that China's environmental losses had reached 3–8% of GDP, which means that GGDP accounted for 92–97% of GDP (World Bank, 1997). According to the Xi'an City Development Plan, total GDP is expected to reach 69 billion USD by 2021 and this would occur with a population of 10.3 million in the city, which would cover an area of 780 km² at that time. Based on this information and the

Table 3

Expert graduated LUI and urban development scheme expected GDP, GDPL, agriculture products, industry products, service products, and fixed assets investments for Xi'an from 1991 to 2021

	LUI A	LUI B	LUI C	LUI D	LUI E	LUI	GDP	GDPL	GDPL + T	AP	IP	SP	FAI
1991	2.9	3.3	3.4	3.6	3.6	3.4	2.56	0.22	0.17	—	—	—	—
1994	3.9	4.3	4.4	4.6	4.5	4.3	3.54	0.43	0.2	0.39	1.56	1.59	1.1
1995	—	—	—	—	—	—	—	—	—	0.5	1.65	1.88	1.2
1996	4.3	4.6	4.5	4.8	5.2	4.7	—	—	—	0.57	2.02	2.37	1.3
1997	4.6	4.9	5.0	5.6	5.5	5.1	5.96	0.76	0.28	0.62	2.41	2.93	1.5
1998	—	—	—	—	—	—	—	—	—	0.63	2.63	3.15	1.8
1999	—	—	—	—	—	—	—	—	—	0.56	3	3.48	2.4
2000	5.1	5.7	5.6	5.9	6	5.7	7.84	0.93	0.34	0.55	3.45	3.84	2.8
2001	—	—	—	—	—	—	—	—	—	0.56	3.98	4.41	3.5
2002	—	—	—	—	—	—	—	—	—	0.6	4.5	4.8	4.1
2003	5.4	6.0	6.2	6.3	6.4	6.1	11.46	1.3	0.49	0.6	5.45	5.41	5.9
2006	6.1	6.6	6.9	7.0	7.0	6.7	16.62	1.83	0.68	—	—	—	—
2009	6.6	7.4	7.3	7.7	7.8	7.4	25.39	2.66	0.99	—	—	—	—
2012	7.5	7.7	7.8	8.2	8.3	7.9	36.09	3.61	1.3	—	—	—	—
2015	8.2	8.5	8.6	8.7	8.7	8.5	49.11	4.83	1.55	—	—	—	—
2018	8.4	9.1	9	9.2	9.1	9.0	61.29	5.93	1.68	—	—	—	—
2021	8.9	9.2	9.1	9.7	9.5	9.3	73.29	7.05	1.62	—	—	—	—

(Units of GDP, GDPL, A, I, S, FAI: 10^9 USD).

LUI (A, B, C, D, and E) represent the land utility grades assigned by Experts A, B, C, D and E. AP, IP, SP, and FAI represent agriculture product, industry product, services product, and fixed assets investment, respectively.

environmental costs for pollution treatment, potential pollution events, and unexpected environmental hazards, we calculated the GDPL and GGDP for Xi'an from 1991 to 2021 (Table 3). Our statistics show that GDPL was 0.22 billion USD in 1991 but would reach 6.64 billion USD by 2021, and that GDPL increased more slowly than GDP. This means that rapid urban expansion has negative effects on the urban ecological environment.

4.3. Water quality fluctuated temporally, and varied from upstream to downstream

Considering the influence of the natural environment on water pollution, we calculated correlation coefficients between water pollutants and the amount of precipitation and plotted pollution levels as a function of the level of precipitation (Fig. 4). From 1996 to 2003, the correlation between BOD5 and precipitation reached a maximum of -0.97 , and that between volatile hydroxybenzenes and precipitation reached -0.96 . This suggests that pollutant concentrations measured in terms of BOD5 and volatile hydroxybenzenes exhibited seasonal variation, and were negatively correlated with precipitation (i.e., levels of these pollutants were highest in the dry seasons). The correlations between precipitation and the concentrations of other pollutants were low, precipitation had little influence on these concentrations. Precipitation varied seasonally, but wastewater discharge did not, therefore surface water quality deteriorated during the dry seasons. More and more wastewater has been discharged into nearby rivers, and sediments accumulated in rivers and channels during the dry seasons; therefore, when the rainy season arrives, a

larger quantity of pollutants is flushed into the rivers, degrading water quality.

Based on our pollutant comprehensive pollution index (PCPI, the comprehensive pollution index of one pollutant at one sample site or in one river) and river comprehensive pollution index (RCPI, the comprehensive pollution index of all pollutants at one sample site or in one river), we were able to use Spearman's rank correlation test to evaluate the status of surface water quality for Xi'an during 1996 and 2003 in comparison with the national standards (Table 4). R_s is the Spearman rank correlation coefficient. W_p is the critical value in testing R_s against zero. Evaluation of trend is based upon the strength of statistical significance of R_s . A negative value shows improved water quality trend and a positive value shows deteriorated water quality trend, and a higher absolute value of R_s shows a more distinct change trend. Table 5 shows that from 1996 to 2003, discharges of petroleum and lead pollutants decreased, volatile hydroxybenzenes increased, and the trends for other pollutants showed slight decreases, and that pollution levels in the Xinh River improved, that pollution levels in the Weihe River, Chanhe River, Bahe River, Yuhe River, Zhaohe River, Heihe River, Fenghe River and Laohe River improved slightly, but that levels in Linhe River deteriorated. From the table, we can see that the pollution situation has generally improved since 1996.

Although the PCPI of two specific pollutants (KMnO and petroleum) increased from 1996 to 2002 (then decreased in 2003 as the government took tougher measures to control paper factories and car washing), other pollutants decreased steadily (Fig. 5). We ranked PCPI values from low to high as: 2003 < 2001 < 2002 < 2000 < 1996 < 1997 < 1998 < 1999.

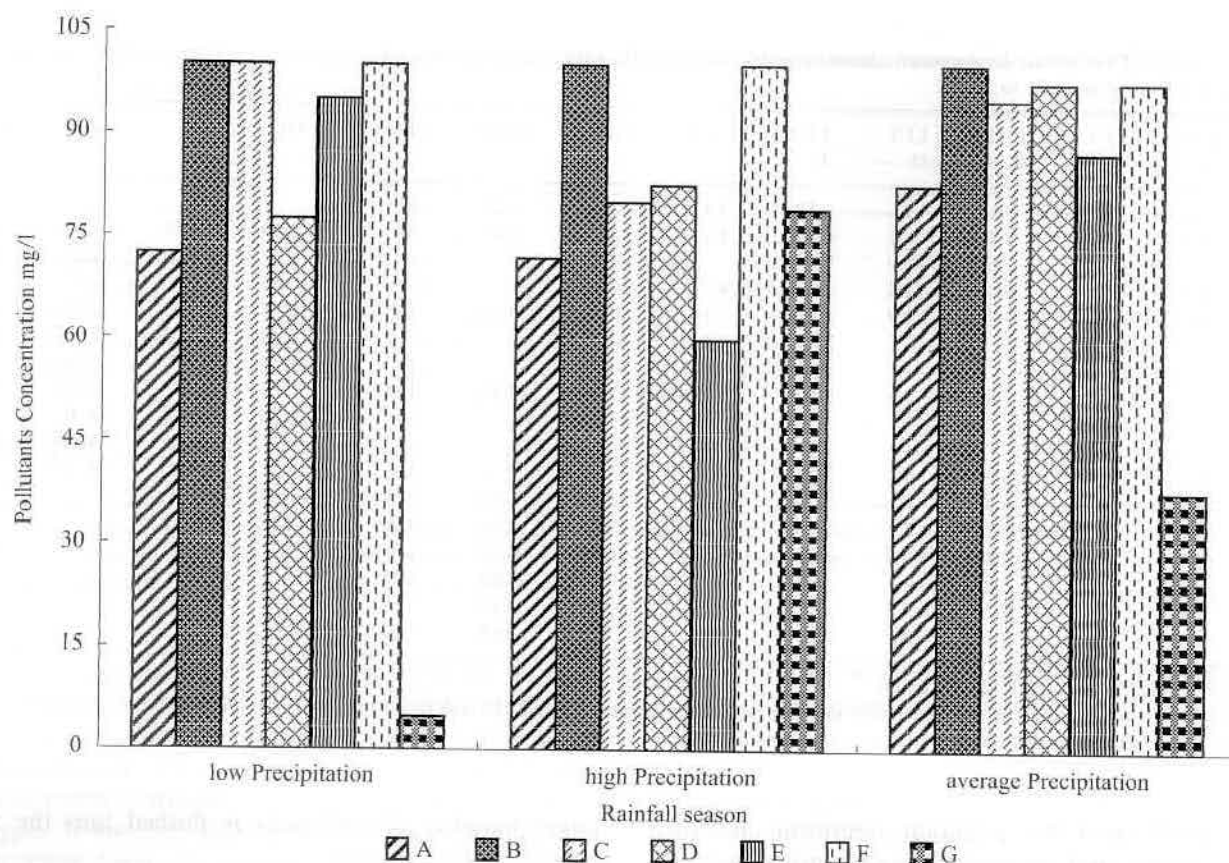


Fig. 4. Relationship between pollutant levels and rainfall. A, dissolved oxygen; B, KMnO; C, BOD5; D, non-ionic ammonia; E, volatile hydroxybenzenes; F, petroleum; G, precipitation. October, November, March and April are the months with average levels of precipitation; May, June, July, August, and September are months with high levels of precipitation; and December, January, and February are months with low levels of precipitation.

Table 4
Environmental quality standard for surface water (GB3838-88, GHZB1-1999, issued by China's State Environmental Protection Administration) (units: mg/L)

Category	I	II	III	IV	V
DO	≥ Saturated rate 90% (or 7.5)	6	5	3	2
KMnO	≤ 2	4	6	10	15
COD	≤ 15	15	20	30	40
BOD5	≤ 3	3	4	6	10
NH ₃ -N	≤ 0.15	0.5	1	1.5	2
TP	≤ 0.02	0.1	0.2	0.3	0.4
TN	≤ 0.2	0.5	1	1.5	2
As	≤ 0.05	0.05	0.05	0.1	0.1
Hg	≤ 0.00005	0.00005	0.0001	0.001	0.001
Cd	≤ 0.001	0.005	0.005	0.005	0.01
Cr ⁶⁺	≤ 0.01	0.05	0.05	0.05	0.1
Pb	≤ 0.01	0.01	0.05	0.05	0.1
Cyanide	≤ 0.005	0.05	0.02	0.2	0.2
VH	≤ 0.002	0.002	0.005	0.01	0.1
Petroleum	≤ 0.05	0.05	0.05	0.5	1

Codes I, II, III, IV, and V represent very good, good, slightly polluted, polluted, and seriously polluted; DO, TP, TN and VH represent dissolved oxygen, total phosphorus, total nitrogen, and volatile hydroxybenzenes, respectively.

Though PCPI values fluctuated between 1996 and 2003, the general trend was towards improved water quality after 1999 (Fig. 5).

Table 5
Computed R_s values of PCPI and RCPI for the rivers in Xi'an shown in Fig. 2 from 1996 to 2003

PCPI	R_s	PCPI	R_s
TSP	-0.42	Cd	-0.24
TH	0.54	Petroleum	-1
KMnO	0.38	RCPI	R_s
BOD5	-0.48	WR	-0.68
NIA	-0.72	CR	-0.72
NO ₃ ⁻	0.45	BR	-0.67
NO ₂ ⁻	0.38	FR	-0.13
VH	1	YR	-0.52
Cyanide	-0.48	ZR	-0.83
Ar	-0.83	HR	-0.66
TM	0.15	LaoR	-0.38
Cr ⁶⁺	-0.33	XR	-0.9
Lead	-0.9	LinR	0.19

TSP is total suspended particles, TH is total hardness, NIA is non-ionic ammonia, NO₃⁻ is Nitrate nitrogen, NO₂⁻ is Nitrite nitrogen, VH is volatile hydroxybenzenes, and TM is total mercury. WR, CR, BR, FR, YR, ZR, HR, LaoR, XR, and LinR represent Weihe River, Chanhe River, Bahe River, Fenghe River, Yuhe River, Zhaohe River, Heihe River, Laohe River, Xinghe River, Linhe River, respectively.

A comparison of Fig. 6 with the sampling locations in Fig. 1 shows that water quality in most rivers degraded moving from upstream to downstream. Some other

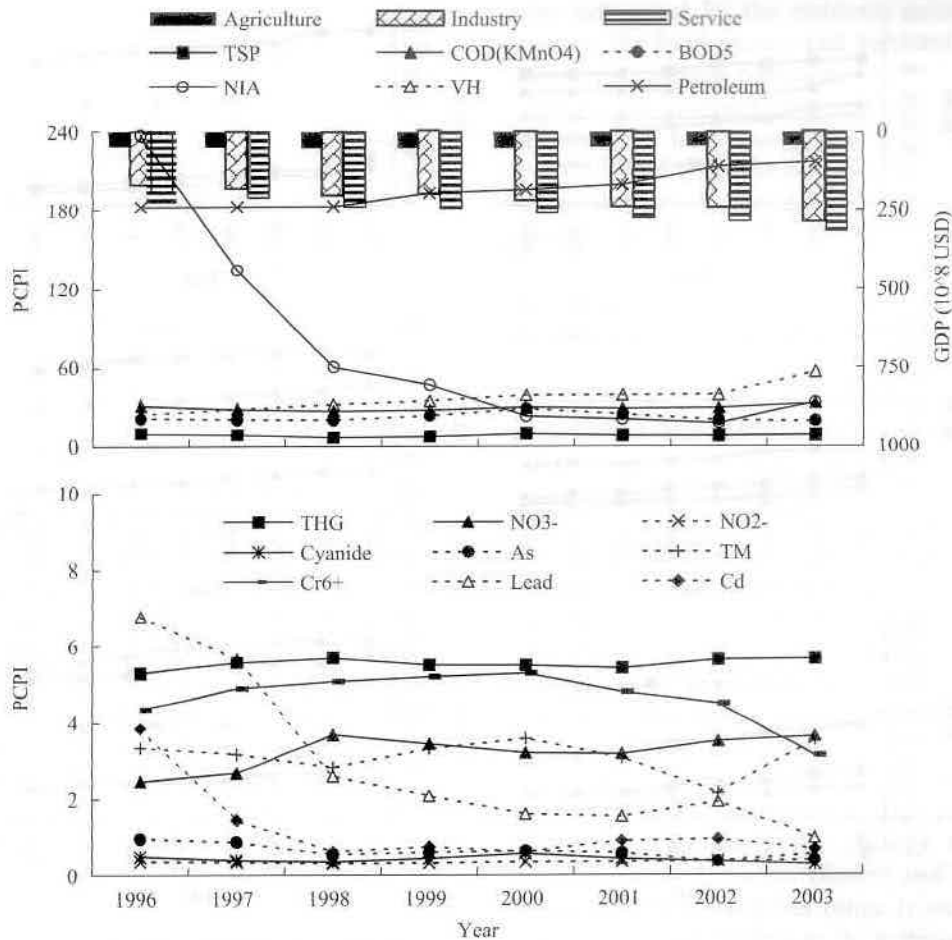


Fig. 5. The relationship between economic growth and PCPI of the main pollutants of rivers in Xi'an between 1996 and 2003.

relevant studies also showed similar findings, i.e., that water quality in other rivers degraded moving from upstream to downstream, such as Han River in South Korea (Chang, 2005) and Marimba River of Harare City, Zimbabwe (Mvungi et al., 2003). Of the 21 sampling sites, the most seriously polluted sections of the river were sections 66, 68, 57, 67, 5, 17, and 59. The most seriously polluted river was the Zhaohe River, which has become an effluent channel, followed by the Weihe River, Fenghe River, Chanhe River, Bahe River, and Yuhe River. Fig. 6 shows that RCPI for the Bahe River (sites 20, 21, 24, and 29) and the Yuhe River (sites 41, 44, and 47) were low, but that pollution levels increased in these rivers moving downstream; this was also the case for the Chanhe River (sites 11, 12, and 17) and the Zhaohe River (sites 53, 57, and 59). Along with the Weihe River and the Fenghe River, the middle Weihe River (sites 2, 5, 7, and 8) and the middle Fenghe River (sites 31, 34, and 36) maintained high RCPI values. The Weihe River had the highest mean RCPI value in general, which meant that it was the most severely polluted of these six rivers during the study period.

4.4. Improper human activities contributed to water pollution, but the water environment improved gradually

Table 2 shows the theoretical values for the Markov transition probabilities for changes in land use in Xi'an from 1996 to 2003. The trends in this table may partially explain Xi'an surface water quality problems. In particular, the rapid decline in land use classified as uncultivated, rural, and vegetation, the high proportion of industrial land, and the lack of green space may have contributed to the rapidly deteriorating surface water quality during this period. For example, as vegetation declined, it was replaced by industrial or residential lands, which generally have a much greater potential for degrading surface water quality. Although farmland decreased at a rate of $3 \times 10^5 \text{ m}^2 \text{ yr}^{-1}$ (from $3.51 \times 10^7 \text{ m}^2$ in 1980 to $2.88 \times 10^7 \text{ m}^2$ in 2001), the agricultural production remained steady. To obtain equal production from less land, more fertilizer and pesticides were used, and the fertilizer application rate increased 12.53% from 1996 to 2003 (Xi'an Municipality Bureau of Environment Protection, 2003). Hence, surface water pollution from the agricultural sector increased, and this probably had

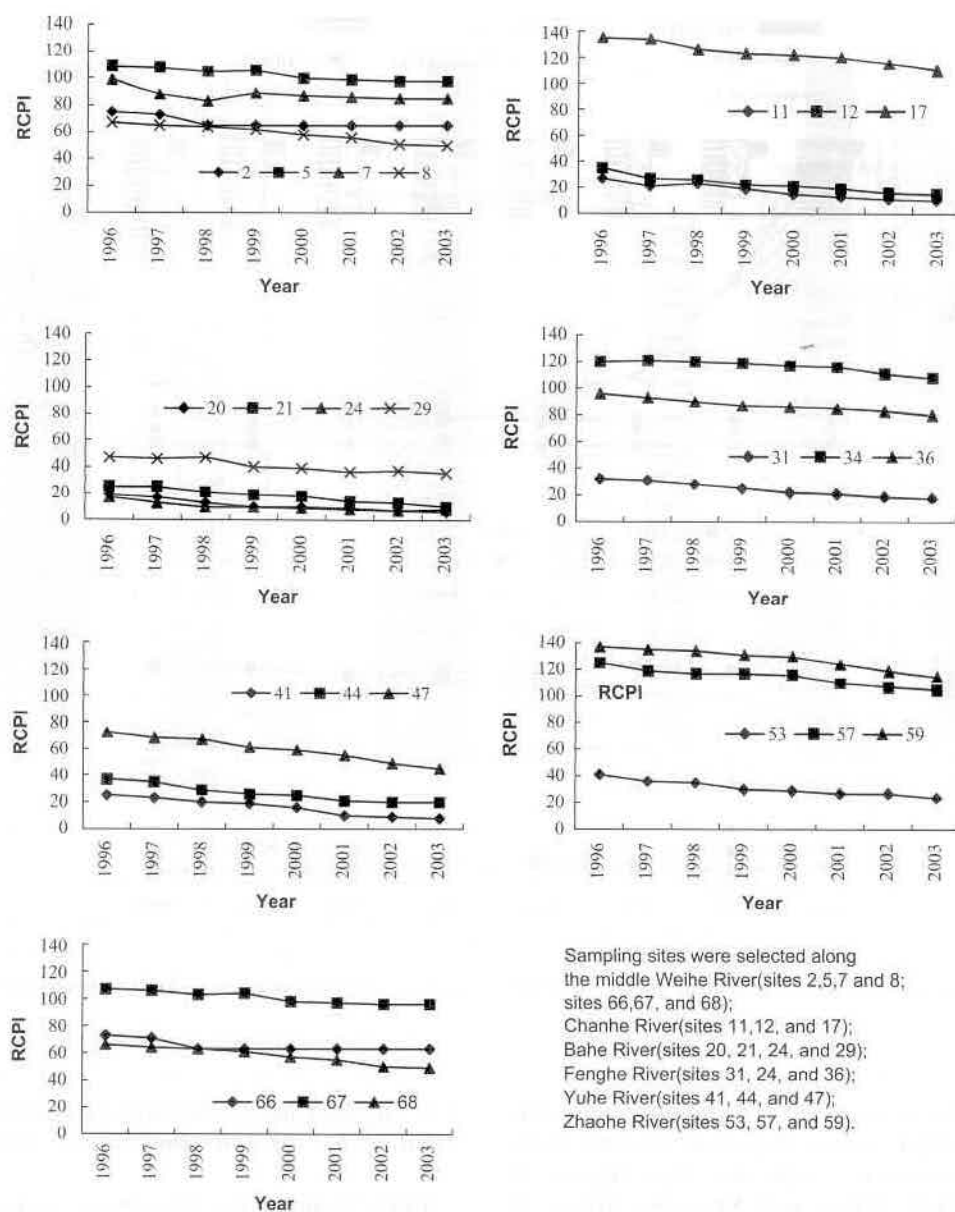


Fig. 6. Variation in RCPI values for the main rivers that were sampled from 2001 to 2003.

negative effects on surface water quality, as was the case for urban and residential land uses.

The distribution of industrial factories (such as Textile Town and West Industry Town, Fig. 1) has had a large impact on the quality of river water. Most of these factories lie in the eastern and western parts of Xi'an. The Chanhe River (in the eastern part of the city) and the Zhaohe River (in the west) are the two most seriously polluted rivers, with organic pollutants dominating the wastewater. However, water quality in the Chanhe River reached Category II ("good" according to the GB 3838-88 standard, Government of China 1988; Table 4) before it flowed north through the Textile Town part of Xi'an (in the northern part of Baqiao District), but deteriorated to Category IV ("highly polluted") by the time it had flowed out of the Textile Town. Water pollutants were mainly sulfide,

ammonia nitrogen, nitrosamine, and volatile hydroxybenzenes. When the Bahe River flowed into the Chanhe River, the water quality in the merged rivers improved rapidly and reached Category I ("very good"). Water quality in the Zhaohe River degraded from Category III ("slightly polluted") into Category V ("seriously polluted") before and after it flowed through the West Industry Town (in the northern part of Weiyang District). In the upstream portion of the Fenghe River, wastewater discharged by the food service industry led to high levels of pollution in terms of BOD₅. However, by the middle and lower reaches of this river, COD and Mn pollution were most serious due to excessive discharge of wastewater by the paper industry.

The increase in economic products from the industrial and service sectors was accompanied by increased organic pollution (Fig. 5), though wastewater discharge has generally

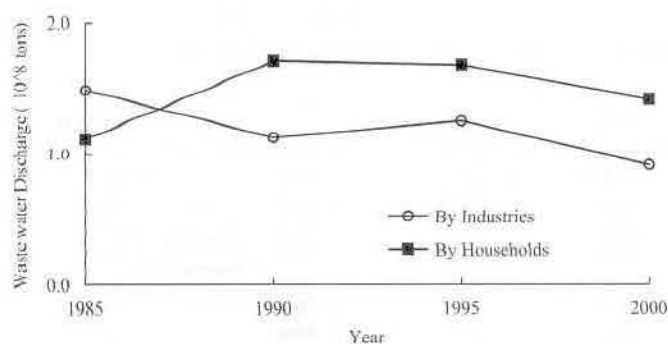


Fig. 7. Annual wastewater discharge in Xi'an (Sources data: Xi'an Statistical Yearbook).

remained constant or has decreased during the study period (Fig. 7). Investigation showed that the COD and Cr components accounted for $54,000 \text{ t yr}^{-1}$ and were produced by about 100 paper factories along the Laohe River, Xinhe River, Zhaohe River, Fenghe River, Yuhé River, Heihe River, and Gaohe River during 1996–2002 (Xi'an Statistical Yearbook, 2003). The food service sector increased from 15,295 restaurants in 1997 to 20,554 in 1999, and the number of employees increased from 71,204 in 1997 to 103,780 in 1999. Organic wastewater produced by these restaurants was emitted directly into effluent channels, where it flowed into rivers without further treatment. Car washing was another important non-point source of pollution (petroleum), since about 400,000 L of water are consumed in washing cars in Xi'an every day. Given the ongoing threat to economic growth and human health posed by pollution of surface water, the municipal government undertook tougher efforts to control this pollution after 1999, and issued a series of laws to protect the city's surface water environment. Industries that propose to discharge waste directly into the rivers without further treatment must first apply for a waste control permit. The government also promotes the development of treatment plants by each industry. Most plants followed waste control measures, but some have been improperly managed. This may explain why discharges of KMnO_4 (discharged by paper factories) and petroleum (discharged by car washing) continued increasing from 1996 to 2002 and why wastewater was still discharged into nearby rivers before being properly treated.

We collected surface water pollution costs from yearbooks and statistical data for Xi'an from 1991 to 2003. Based on the future socioeconomic development plan for Xi'an from 2005 to 2021, we calculated the costs of surface water pollution (Table 6). Rapid economic growth clearly contributed to water pollution. As discharges from both domestic and industrial effluents increased, clean water became increasingly scarce. Constantly increasing socioeconomic activity clearly influenced the natural environment. The development of urbanization, industry, and agriculture has been accompanied by the production of large amounts of wastewater and pollution, both of which determine the quality of rivers and lakes. Bodies of water

were influenced by the nutrients released from domestic sources, the food sector, and light industry, as well as by diffuse agricultural pollution.

From this analysis, we conclude that the accelerated urbanization has clearly had a negative effect on the ecological environment in Xi'an for a long time. However, the quality of the surface water is improving, and will possibly reach equilibrium in the near future.

4.5. Model simulation: improved technologies will improve the maximum environmental carrying capacity

Xi'an is continuing to urbanize rapidly, and the city government has decided to expand the city to cover 780 km^2 in the coming 15 years; moreover, the city may continue to expand after that period. Therefore, this urbanization will have even greater impacts on the ecological environment. It is unlikely that Xi'an can expand its urban scale indefinitely, so it is necessary to evaluate the sustainability of this urbanization process. We will discuss the impacts of urbanization on the carrying capacity of Xi'an surface water environment in the next section by discussing our urbanization model based on social welfare maximization theory.

Urban growth models have been developed to explain and simulate urban dynamics, and play an instrumental role in understanding the ecology of cities and their spatiotemporal dynamics (Sheryl and Wu, 2004; Wu and Marceau, 2002). Methods range from traditional macro-simulation urban models to the rapidly developing field of cellular automata (Clarke et al., 1996; Couclelis, 1985, 1997; White et al., 1997) and emerging agent-based models (Franklin and Graesser, 1996). While models cannot fully capture all of the intricacies of urban dynamics, they can provide significant insights into general urban growth trends. Despite their limitations, models can still provide significant insights. We constructed a model to simulate the carrying capacity of the ecological environment during the urbanization process. We have discussed aspects of the model-construction process in the text that introduces and describes Eqs. (2) through (5) and (7) through (12); the data in Tables 3 and 6 can be used to construct a model that simulates the response of water quality to urbanization.

We used data in Tables 3 and 6 to construct a quadratic model for GDPL as a function of WPC, and for WPC as a function of LUI.

$$\text{LU/LUI} = f(\text{LUI}) \rightarrow y_1 = -2.4x_1^2 + 43.9x_1 + 12.7, \quad (14)$$

$$\text{GDPL} = f(\text{WPC}) \rightarrow y_2 = x_2^2 - 4.8x_2 - 66.8. \quad (15)$$

Therefore:

$$\begin{cases} \frac{f(\text{LUI})}{\text{LUI}} = \frac{\partial f(\text{WPC})}{\partial \text{WPC}} \times \frac{\partial \text{WPC}}{\partial \text{LUI}} \rightarrow \begin{cases} \frac{y_1}{x_1} = -2.4x_1 + \frac{12.7}{x_1} + 43.9, \\ \frac{y_2}{x_2} = 5.5x_1 - 4.8. \end{cases} \\ \text{WPC} = f(\text{LUI}) \end{cases} \quad (16)$$

Table 6
LU, WPC, AV(LU), and M(WPC) for Xi'an under the assumption of fixed technology (t_1) and developing technology (t_2) from 1981 to 2021 at 3-year intervals from 1991 to 2021

Year	LU	F(LU+ t_1)	WPC	F(WPC+ t_2)	AV(LU)	M(WPC)	AV(LU+ t_1)	M(WPC+ t_2)
1991	0.74	180	0.12	3.9	38.4	22.7	80.4	7.7
1994	1.06	260	0.16	4.0	36.6	29.3	78.6	14.3
1997	1.34	291	0.17	5.6	35.3	35.0	77.3	20.0
2000	1.55	314	0.19	7.1	34.6	39.4	76.6	24.4
2003	1.52	316	0.2	8.1	32.0	42.3	74.0	27.3
2006	1.46	317	0.21	10.3	28.3	46.6	70.3	31.6
2009	1.39	334	0.24	10.7	27.0	51.7	69.0	36.7
2012	1.7	348	0.25	11.8	26.4	55.3	68.4	40.3
2015	1.82	364	0.27	12.7	25.8	59.7	67.8	44.7
2018	1.85	372	0.31	13.8	24.7	63.3	66.7	48.3
2021	1.76	367	0.34	13.3	23.0	65.5	65.0	50.5

(Units of LU and WPC are 10^9 USD).
LU, WPC, AV(LU), and M(WPC) represent land utility, water pollution cost, average land utility, and marginal water pollution costs.

On this basis, the critical point value is a LUI value of 4.7; this means that land utility and the cost of water pollution reach a balance state at this value for constant values of technologies related to the domestic product and water pollution treatment technology.

We can carry our analyses further as follows. First, we must emphasize that this result is based on the assumption that technological factors can be “reduced” to simplify our model by assuming that they are external variables in this model. These technologies can also be considered to be improving as a result of innovations for a given level of GDP and water pollution treatment technologies. Without considering improvements in technologies for a given level of GDP, average land utility would increase as the land area expanded and different land-use types recombined. Under these conditions, average land utility would remain low because land values grow only in response to growth in urban land. In contrast, if technologies for a given level of GDP are considered to improve, average land utility will reach a higher level (Figs. 8, 10, 12, and 13). However, the situation for M(WPC) is different. With innovation in pollution treatment technologies, industries and households would be able to employ new technologies to reduce pollution and wastewater before these are discharged into the rivers. Therefore, M(WPC) is higher with fixed pollution treatment technologies than under improving technologies (Figs. 9, 10, 12, and 13).

In Figs. 11–13, we can see how the relationships between the various factors discussed above will determine the balance point. As we discussed above, the marginal pollution cost M(WPC) is defined as the derivative of GDPL with respect to WPC, so average land utility equals LUI/LU. Here, we have assumed that M(WPC) minus AV(LU) equals $|\Delta(M-AV)|$ (Fig. 11).

Suppose that:

$X_1 = X_0 - \Delta X, X_2 = X_0 + \Delta X.$ (17)

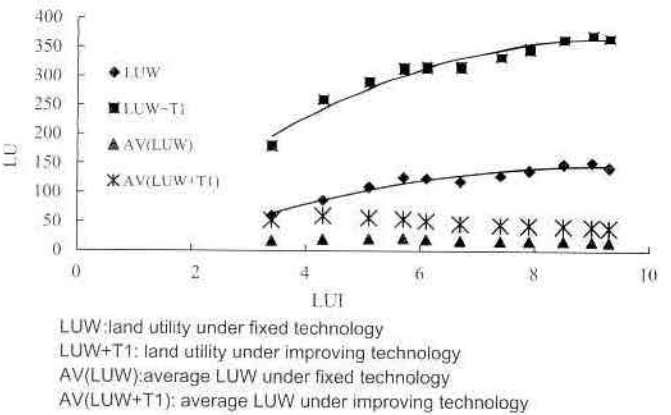


Fig. 8. Relationship between LUI and LU, and between LUI and AV (LU).

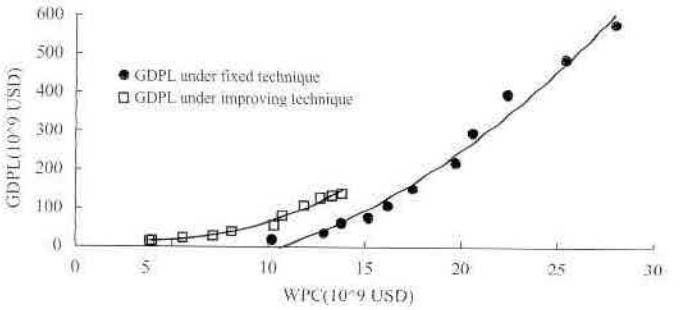


Fig. 9. Relationship between GDPL and WPC.

Therefore:

$$\begin{cases} M(X_1) = M(X_0 - \Delta X) \\ M(X_2) = M(X_0 + \Delta X) \\ AV(X_1) = AV(X_0 - \Delta X) \\ AV(X_2) = AV(X_0 + \Delta X) \end{cases}$$

$$\rightarrow \begin{cases} |\Delta(M - AV)|_1 = |M(X_1) - AV(X_1)| \\ |\Delta(M - AV)|_2 = |M(X_2) - AV(X_2)| \end{cases} \quad (18)$$

From X_1 to X_0 , the difference between $M(WPC)$ and $AV(LU)$ equals the social welfare. That is, $|\Delta(M - AV)|_1 = \Delta(M - AV)_1 > 0$. This means that $AV(LU)$ exceeds $M(WPC)$, but that the growth rate of $AV(LU)$ decreases. Thus, $\Delta(M - AV)_1$ decreased during urbanization and might lead to deteriorated water quality. At the critical point X_0 , $M(WPC)$ and $AV(LU)$ reach equilibrium and social welfare reaches its maximum value, and $M(WPC)$ and $AV(LU)$ reach an equilibrium state at the maximum environmental carrying capacity. After this point, urban expansion no longer increases LU . When urbanization proceeds to point X_2 , $|\Delta(M - AV)|_2 = \Delta(M - AV)_2 < 0$, social welfare decreases (becomes negative) and $M(WPC)$ exceeds LU . Therefore, urban expansion has negative effects in general for the relationship between LU and $M(WPC)$, and this leads to deterioration of water quality. After the critical point, urbanization leads to degradation of the ecological environment. We should stop urban expansion before it reaches this point in the urbanization process. Our simulated results show that urbanization in Xi'an reaches equilibrium when LUI equals 4.7 (Fig. 13). Does this mean that urbanization in Xi'an should have stopped before 1997? The answer is no,

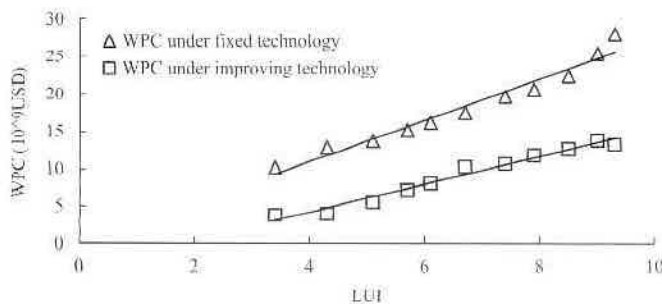


Fig. 10. Relationship between LUI and WPC .

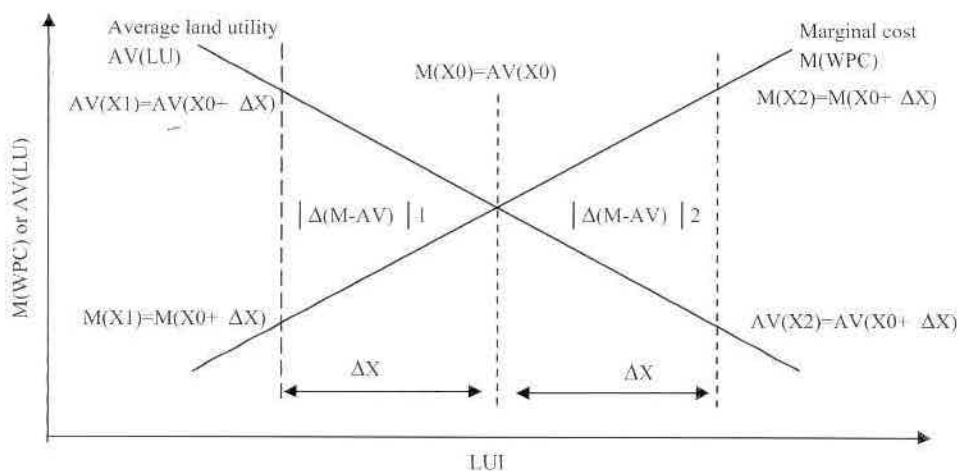


Fig. 11. Analysis of social welfare as a function of the balance between $M(WPC)$ and $AV(LU)$ as a function of LUI .

because we have not included the external variable of technological change in this model. Doing so produces a different conclusion.

Technology is not constant, and we should include technological change in our model of the urbanization process. As we know from the previous discussion, technologies for a given level of GDP (T_1) and water

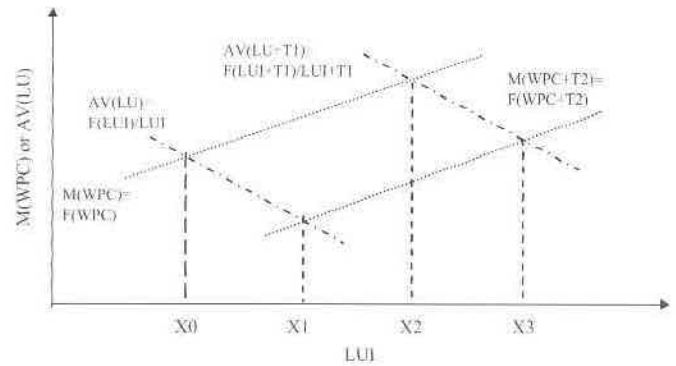


Fig. 12. Analysis of technology-induced equilibrium between $M(WPC)$ and $AV(LU)$, as a function of LUI . T_1 , technologies for a given level of GDP included; T_2 , water pollution treatment technologies included.

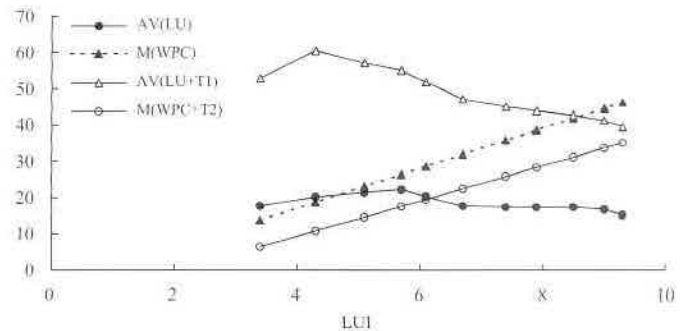


Fig. 13. $AV(LU)$ and $M(WPC)$ values for Xi'an as a function of LUI from 1991 to 2021. T_1 , technologies related to domestic product; T_2 , water pollution treatment technologies.

treatment technologies (T_2) are external variables:

$$WPC' = f(WPC + T_2). \quad (19)$$

$$LU' = f(LU + T_1). \quad (20)$$

Therefore:

$$\begin{cases} \frac{f(LU + T_1)}{LU + T_1} = \frac{f(WPC + T_2)}{\partial(WPC + T_2)} \times \frac{\partial(WPC + T_2)}{\partial LU}, \\ WPC + T_2 = f(LU + T_1). \end{cases} \quad (21)$$

Notice that equilibrium occurs at X_0 , X_1 , X_2 , and X_3 in Fig. 12.

Calculation of the maximum environmental carrying capacity for Xi'an mostly depends on the external variables T_1 and T_2 . These variables determine at what LUI value the balance state is occurring. Data used for T_1 and T_2 were derived from GDP growth regressions. We further discussed different conditions through models for GDPL as a function of WPC, and for WPC as a function of LUI, and data in Tables 3 and 6 were used to calculate the critical points in the same ways as Functions 14–16.

Without considering external technological variables $T_1 = T_2 = 0$, $AV(LU)$ and $M(WPC)$ reach equilibrium at the critical point X_0 :

$$\begin{cases} AV(LU) = F(LUI)/LUI \\ M(WPC) = F(WPC)' \end{cases} \rightarrow \frac{F(X_0)}{X_0} = \frac{\partial f(WPC)}{\partial WPC} \times \frac{\partial WPC}{\partial X_0}. \quad (22)$$

The maximum environmental carrying capacity for Xi'an occurred at an LUI of 4.7 around 1996 (Fig. 13).

However, if we allow improvements in technologies for a given level of GDP and ignore water pollution treatment technologies $T_1 = 0$, $T_2 > 0$, then the critical point is at X_1 :

$$\begin{cases} AV(LU) = F(LUI)/LUI \\ M(WPC) = F(WPC)' \end{cases} \rightarrow \frac{F(X_1)}{X_1} = \frac{\partial f(WPC)}{\partial WPC} \times \frac{\partial WPC}{\partial X_1}. \quad (23)$$

On this basis, the maximum environmental carrying capacity for Xi'an will occur at an LUI of 8.4 in around 2015.

If we consider improving water pollution treatment technologies but ignore technologies for a given level of GDP, $T_1 > 0$, $T_2 = 0$, then the critical point is at X_2 :

$$\begin{cases} AV(LU) = F(LUI)/LUI \\ M(WPC) = F(WPC)' \end{cases} \rightarrow \frac{F(X_2)}{X_2} = \frac{\partial f(WPC)}{\partial WPC} \times \frac{\partial WPC}{\partial X_2}. \quad (24)$$

On this basis, the maximum environmental carrying capacity for Xi'an would have occurred at an LUI value of 6.3 around 2004.

If we include improvements in both technologies, $T_1 > 0$, $T_2 > 0$, then the critical point is at X_3 :

$$\begin{cases} AV(LU) = F(LUI)/LUI \\ M(WPC) = F(WPC)' \end{cases} \rightarrow \frac{F(X_3)}{X_3} = \frac{\partial f(WPC)}{\partial WPC} \times \frac{\partial WPC}{\partial X_3}. \quad (25)$$

On this basis, the maximum environmental carrying capacity for Xi'an will occur at an LUI value of 9.8 some time after 2021.

Obviously, improving technology can increase the $AV(LU)$ and reduce $M(WPC)$ from X_0 to X_1 , thereby enhancing the environmental carrying capacities and postponing the point at which urbanization must stop. Technological innovation can promote the urbanization process. However, as is the case at a low level of technology, urbanization should proceed sufficiently slowly to maintain a healthy environment. In a rapidly developing society, urban sprawl is difficult to control, but urbanization must still stop before reaching a critical point if we are to maintain an environmental balance that permits sustainable development.

Note that when $T_1 = 0$, $T_2 > 0$ the critical point is at X_1 and when $T_1 > 0$, $T_2 = 0$, the critical point occurs at X_2 . This shows that improved water pollution treatment technologies will improve the maximum environmental carrying capacity more than improvements in technologies for a given level of GDP. Therefore, the crucial task for Xi'an is to improve water pollution treatment during urbanization rather than just expanding the urban area. These results suggest that Xi'an could theoretically complete its urbanization process after 2021 rather than in 1996.

5. Conclusion and discussion

To evaluate the impacts of urbanization on land use and surface water quality in Xi'an from 1996 to 2003, we studied both factors using statistical methods and geographical information system (GIS) techniques. Our results showed a strong relationship between land use and surface water quality. This relationship is evident in both the statistical and the spatial analysis. The development of a Markov transition matrix revealed the patterns of change between different land uses. The urban periphery of Xi'an is the region where land use changed most. The primary feature of this change was that a large amount of farmland was replaced by urban structures. In addition, the development strategy for western China and the policy for reclamation of woodland from farmland converted much farmland into economic and noneconomic woodland. The Markov transition probabilities could describe the land use changes in Xi'an from 1996 to 2003.

The trends in the proportions of different land uses explain some of the reasons for Xi'an surface water quality problems. In particular, the rapid decline in agricultural land, the increased and high proportion of industrial land, and the lack of green space may all be contributing to the

rapidly deteriorating surface water quality. Surface water pollution in the rivers of Xi'an showed a strong organic component. Rapid economic growth also clearly contributed to surface water pollution. As discharges of both domestic and industrial effluents have increased, clean water has become increasingly scarce. The distribution of factories has also had large impacts on river water quality. The increasing industrial and service sector products were both accompanied by increased organic pollution, and this caused deterioration of surface water quality. The calculation of water quality indices (RCPI, PCPI) is a traditional method for comprehensive analysis of water quality, and both indices proved capable of modelling surface water quality with an acceptable degree of accuracy. All of the simulated values were very close to the actual monitored values.

Land use change in Xi'an has been driven by many forces. However, human activity was the most important one. Rapid economic development and urbanization were the fundamental reasons for land use change from 1996 to 2003. Investment in social assets, development of the urban economy, construction of infrastructure such as transportation facilities, and internal adjustments in agricultural land use also played important roles in the process of land use change.

Our urban expansion model suggests that urbanization does not inevitably lead to degradation of the ecological environment provided that technological improvements keep pace with the rate of urbanization. However, urban expansion must stop before certain critical points are reached. We found that marginal water pollution costs were higher with fixed pollution treatment technologies than with improving pollution treatment technologies. Urban expansion in Xi'an should have reached equilibrium when the land use index (LUI) equalled 4.7 in 1996 if we ignore the impacts of technological factors, but the critical value of the land-use index could potentially increase to 9.8 at some time after 2021 if we assume technological innovation.

It is true that treatment technologies will possibly fail to maintain adequate water quality at very high levels of urbanization because of increases in diffuse or unpredictable sources of pollution such as urban runoff and accidental releases. We accounted for these unexpected pollution events in our GDPL calculation. Our simulation results showed that focusing on water pollution treatment technologies would improve the maximum environmental carrying capacity more than improvements in technologies for a given level of GDP. Therefore, it is crucial to deal with water pollution treatment during urbanization rather than just expanding the urban area. Our results indicate that Xi'an could theoretically delay the completion of its urbanization process until after 2021 based on these improvements, rather than reaching an estimated limit in 1996 in the absence of these improvements.

As an abstraction of reality, modelling will always have shortcomings. Nevertheless, we are reliant on simulations to plan for the future implications of current decisions. As researchers, modelling is also necessary to test theories

relating to urbanization. In this sense, models are crucially important for the management of urban systems and for understanding their functioning (Wegener, 1994; <http://www.sprawlsim.org>). From the present results, it is clear that our model of urbanization for Xi'an could be improved. For example, in order to facilitate the development of our model, we simplified the processing of data for variables such as GGDP, GDPL, LUW, and WPC, which are complex to calculate. In addition, external variables such as technological factors should be discussed in more detail. At each stage of urban expansion (i.e., for certain values of LUI), urbanization should include optimal technological factors, which must be treated as dynamic variables. Our model fails to reflect these variations, since it has been simplified for the sake of discussion. We integrated these unexpected pollution events and water quality improvement related factors into our GDPL calculation. Improvement of the water environment must be achieved not only by water treatment technology and technologies for a given level of GDP, but also by policy decisions, enforcement of these decisions, and public awareness. We plan to improve the model through future studies that acquire relevant data. Despite the model's deficiencies, we have confidence in the basic results of our simulation because the predictions are generally reasonable, and often do a good job of predicting actual measured values.

Acknowledgments

The study was supported by parts of grants of Chinese Academy of Sciences (KZCX3-SW-146), EU-SINO (ICA4-CT-2002-10004), State Key Laboratory of Loess and Quaternary Geology (SKLLQG0417). We gratefully acknowledge the invaluable suggestions to three anonymous reviewers and to Dr. Heqing Huang and Dr. Zhongping Lai of Oxford University.

Appendix. Nomenclatures

ANN	artificial neural network
AV	average utility, which is the average total income of all individuals.
AV (LU)	average land utility, the average total USD of combination of different land use types
AV(LU + t_1)	average land utility under the assumption of fixed technology (t_1)
CPI	comprehensive pollution index
Environmental carrying capacity	the maximum or optimum amount required to support a healthy ecological environment and socioeconomic community
FAI	fixed assessment Investment
$F(LU + t_1)$	land utility in the condition of fixed technology (t_1)

$F(WPC + t_2)$	water pollution cost in the condition of developing technology (t_2)
GGDP	Green GDP, which is represented as $GGDP = GDP - H - Ep - Cpd$
GDPL	GDP loss
Land capacity	land optimum amount to support healthy ecological and socio-economic community, defined land utility as its usefulness for ecological and human activities, and calculated the usefulness as a dollar value in market
LU	land utility, under the assumption of fixed technologies for a given level of GDP, see Function 8
LU'	land utility, under the assumption of gradual improvements in technologies for a given level of GDP that result from socio-economic development, see Function 7
LUCC	land use cover change
LUI	land utility index, to represent the value of land for different combinations of land types, see Function 2
MC	marginal cost, defines the increase in cost per each unit increase in consumption, and is expressed as the first derivative of the input equation, and represents the consumption of input required to generate additional income
M (WPC)	marginal water pollution cost, the increase in USD cost of water pollution in per unit in GDP increase
$M(WPC + t_2)$	marginal water pollution costs under the assumption of developing technology (t_2)
NDBI	normalized difference of barren index
PCPI	pollutant complex pollution index, and was defined as the comprehensive pollution index of one pollutant at one sample site or in one river
PL	pollution load(see Functions 3–4)
RCPI	river comprehensive pollution Index, and was defined as the comprehensive pollution index of all pollutants at one sample site or in one river
WPC	represents the water pollution treatment costs under a fixed level of pollution treatment technology, see Function 9
WPC'	represents water pollution costs under a gradually improving technology that results from socioeconomic development

References

- Alberti, M., 2003. Determinants of urban stream degradation. http://www.cfr.washington.edu/research/urbaneco/student_info/classes/Aut2003/UE_10_14_03.pdf.
- Allen, P.M., Boon, F., Engelen, G., et al., 1984. Modelling evolving spatial choice patterns. *Applied Mathematics and Computation* 1, 97–129.
- Batty, M., 1971. Modelling cities as dynamic systems. *Nature* 231, 425–428.
- Brown, A.L., 2003. Increasing the utility of urban environmental quality information. *Landscape and Urban Planning* 1–2, 85–93.
- Cathlin, M., 1999. Projecting Land Use in Austin. <http://www.crrwr.utexas.edu/gis/gishydro00/Class/trmproj/McCullough/TermProject.htm>
- Chang, H., 2005. Spatial and temporal variations of water quality in the Han River and its tributaries, Seoul, Korea, 1993–2002. *Water, Air, and Soil Pollution* 161, 267–284.
- Charbonneau, L., Morin, D., Royer, A., 1993. Analysis of different methods for monitoring the urbanization process. *Geocarto International* 8, 17–25.
- Clarke, K.C., Hoppen, S., Perez, S., 1996. Clarke urban growth model. <http://www.geo.arc.nasa.gov/usgs/clarke/hilt.html>.
- Couclelis, H., 1985. Cellular worlds: a framework for modelling micro-macro dynamics. *Environment and Planning A* 17, 585–596.
- Couclelis, H., 1997. From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B: Planning and Design* 24, 165–174.
- Franklin, S., Graesser, A., 1996. Is it an agent, or just a program? A taxonomy for autonomous agents. In: *Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages*: 132–145. <http://www.mscl.memphis.edu/~franklin/AgentProg.html>.
- GB 3838-88. Environmental Standards for Surface Water. The National Standards of the People's Republic of China, 1988. <http://www.zhb.gov.cn/english/SOE/soechina1997/water/standard.htm>
- Hietel, E., Walldardt, R., Otte, A., 2004. Analysing land-cover changes in relation to environmental variables in Hesse, Germany. *Landscape Ecology* 19, 473–489.
- Hotelling, H., 1932. Edgeworth's taxation paradox and the nature of demand and supply function. *Political Economy* 40, 577–616.
- Irene, V.K., Kees, L., Gooitske, M., et al., 2003. Urban environmental quality and human well-being: towards a conceptual framework and demarcation of concepts; a literature study. *Landscape and Urban Planning* 1–2, 5–18.
- Lopez, E., Bocco, G., Mendoza, M., 2001. Predicting land-cover change in the urban fringe. *Landscape and Urban Planning* 55, 271–285.
- Manasquan River Watershed Management Group. 1999. Manasquan River Watershed: <http://www.visitmonmouth.com/mosquito/mwmg/landdev.pdf>
- Mustafa, M.A., Al-Bahar, J.F., 1991. Project risk assessment using the analytic hierarchy process. *IEEE Transactions on Engineering Management* 38, 46–52.
- Mvungi, A., Hranova, R.K., Love, D.J., 2003. Impact of home industries on water quality in a tributary of the Marimba River, Harare: implications for urban water management. *Physics and Chemistry of the Earth* 28, 1131–1137.
- Nafo, I.I., Geiger, J., 2004. A method for the evaluation of pollution loads from urban areas at river basin scale. *Physics and Chemistry of the Earth* 29, 831–837.
- Negishi, T., 1960. Welfare economics and the existence of equilibrium for a competitive economy. *Metroeconomica* 12, 92–97.
- Pacione, M., 2003. Urban environmental quality and human well-being—a social geographical perspective. *Landscape and Urban Planning* 1–2, 19–30.
- Robert, A., Tomas, B., 2000. Comprehensive regional modelling for long-range planning: linking integrated urban models and geographic information systems. *Transportation Research Part A: Policy and Practice* 2, 125–136.
- Rumelhart, D., McClelland, J. (Eds.), 1986. *Parallel Data Processing*, vol. 1. The MIT Press, Cambridge, MA, pp. 318–362 (Chapter 8).
- Sheryl, B.W., Wu, J., 2004. Modelling urban landscape dynamics: a case study in Phoenix, USA. *Urban Ecosystems* 7, 215–240.

- Stephan, P., Roland, E., Yvonne, G., 2005. Modelling the environmental impacts of urban land use and land cover change—a study in Merseyside. *Landscape and Urban Planning* 2, 295–310.
- Thomas, D.G., 2001. Detecting trends using Spearman's rank correlation. *Environmental Forensics* 2, 359–362.
- Treitz, P.M., Howarth, P.J., Gong, P., 1992. Application of satellite and GIS technologies for land-cover and land-use mapping at the rural-urban fringe: a case study. *Photogrammetric Engineering and Remote Sensing* 58, 439–448.
- Turner, M.G., 1987. Spatial simulation of landscape changes in Georgia: a comparison of 3 transition models. *Landscape Ecology* 1, 29–36.
- Wegener, M., 1994. Operational urban models. *Journal of the American Planning Association* 60, 17–29.
- White, R., Engelen, G., Uljee, I., 1997. The use of constraint cellular automata for high resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design* 24, 323–343.
- World Bank, 1997. <<http://www.ias.unu.edu/binaries/BaliPrepComReport4.pdf>>
- Wu, F.L., Martin, D., 2002. Urban expansion simulation of Southeast England using population surface modelling and cellular automata. *Environment and Planning A* 10, 1855–1876.
- Wu, J., Hobbs, R., 2002. Key issues and research priorities in landscape ecology: an Idiosyncratic synthesis. *Landscape Ecology* 17, 355–365.
- Wu, J., Marceau, D., 2002. Modelling complex ecological systems. *Ecological Modelling* 153, 1–6.
- Xi'an Statistical Yearbook, 2003. China Statistics Press, pp. 1–201.
- Yoshimura, C.O.T., Furumai, H., Tockner, K., 2005. Present state of rivers and streams in Japan. *River Research and Applications* 21 (2–3), 93–112.
- Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* 24, 583–594.

Application of a long-path differential optical absorption spectrometer (LP-DOAS) on the measurements of NO₂, SO₂, O₃, and HNO₂ in Gwangju, Korea

Jeongsoon Lee^{a,*}, Ki-Hyun Kim^b, Young J. Kim^c, Jaihoon Lee^c

^aEnvironment measurement Group, Korea Research Institute of Standards and Science (KRISS), 1 Doryong Dong, Yuseong Gu, Daejeon 305-600, Republic of Korea

^bDepartment of Earth & Environmental Sciences, Sejong University, 98 Gwon-Ja Dong, Gwang-Jin Gu, Seoul 143-747, Republic of Korea

^cAdvanced Environmental Monitoring Research Center, Department of Environmental Science & Engineering, Gwangju Institute of Science and Technology (GIST), 1 Oryong Dong, Buk Gu, Gwangju 500-712, Republic of Korea

Received 5 June 2005; received in revised form 14 December 2006; accepted 30 December 2006

Available online 1 March 2007

Abstract

A differential optical absorption spectrometer (DOAS) technique has been applied to monitor airborne trace pollutants including NO₂, SO₂, O₃, and HNO₂ in the ultraviolet (UV) region (290–350 nm) over a 1.5 km beam path (two ways) during an intensive measurement campaign held at Gwangju, Korea (March 2002). Their mean mixing ratios (and standard deviations) were computed as 11.3 (8.8), 1.9 (1.7), 17.1 (19.3), and 0.5 (0.4) ppbv, respectively.

As a means to evaluate the performance of the long-path DOAS (LP-DOAS) system with conventional point monitoring systems (PMS), correlation analysis was conducted between the two data sets. These data sets were then inspected to account for the influence of the environmental conditions on the correlation strength between the two systems, especially with respect to light level and wind speed. To facilitate the comparison, correlation analyses were conducted after dividing the data sets for those parameters into several classes. The strength of the correlations between DOAS and meteorological parameters was also examined to evaluate their effects on the DOAS performance. It was found that, among the four pollutant species, O₃ is the most sensitive to changes in meteorological conditions in relation with atmospheric mixing conditions. The overall results of our study indicate that open-path monitoring techniques can be used to effectively diagnose air quality and be substituted for the conventional point monitoring methods with the proper consideration of those parameters affecting the DOAS sensitivity (e.g., light level and wind speed).

© 2007 Elsevier Ltd. All rights reserved.

Keywords: UV spectrometer; Light level; Wind speed; DOAS; Diurnal variation; Ozone

1. Introduction

As one of the open-path monitoring techniques, long-path differential optical absorption spectroscopy (LP-DOAS) is promising and powerful in that it can optically measure multiple atmospheric trace gases simultaneously. As DOAS allows measurements without pre- and/or post-processing for an open path over several kilometers, it can overcome the technical and/or spatial limitations which the

traditional point monitoring systems (PMS) tend to suffer from. The LP-DOAS system has been deployed to measure intact the concentrations of a multitude of trace gas species which include: NO₂, NO, NH₃, CH₂O, SO₂, CS₂, O₃, NO₃, HNO₂, and volatile organic compounds (VOC) (e.g., Febo et al., 1996; Kourtidis et al., 2002; Perner and Platt, 1979; Plane and Chia-Fu Nien, 1992; Platt, 1994; Platt and Stutz, 2006).

Although many advantages of the DOAS system have long been recognized, its application is yet uncommon relative to the conventional PMS under many situations. Recently, however, there has been a growing tendency to report the application of the DOAS technique under

*Corresponding author.

E-mail addresses: leejs@kriss.re.kr (J. Lee),
kkim@empal.com (K.-H. Kim).

various circumstances (Chiu et al., 2005; Febo et al., 1996; Kim and Kim, 2001; Kourtidis et al., 2002; Lee et al., 2002, 2005; Kim, 2004; Mathew et al., 2001; Virkkula, 1997). Those attempts have focused mainly on the examination of the DOAS feasibility for the quantification of criteria pollutants or aromatic VOCs. In this work, such efforts have been extended further to assess the performance of an LP-DOAS against the PMS techniques in relation to meteorological parameters. For the purpose of this comparative analysis, the mixing ratios of several pollutants were measured continuously by DOAS at a suburban monitoring station in Korea during the spring of 2002. In the course of this study, the potential influence of several basic meteorological parameters (wind speed, light level (visibility), etc.) was also assessed by examining the relationship between DOAS and PMS. The results of this study are thus expected to help understand the basic assets of DOAS.

2. Experimental

2.1. Description of DOAS system

An LP-DOAS system, built by the Gwangju Institute of Science and Technology (GIST) group, consists of a photodiode-array detector (PDA, Hamamatsu S3904) coupled with a spectrometer, a coaxial telescope, a set of seven retroreflectors, and an electronic control system, as described in Lee et al. (2002). A spectrometer, a conventional czerney-turner type, with a holographically grooved grating, blazed at 330 nm with 1200 grates/mm, was used to measure the trace gas species (NO_2 , SO_2 , O_3 , and HNO_2) in each respective ultraviolet (UV) region. The instrument was periodically calibrated using the Hg emission spectrum. The spectral resolution was ~ 0.3 nm (FWHM) at around 334 nm. To prepare the corresponding reference absorption cross-sections, the literature-based absorption spectra were convoluted and interpolated with this instrumental function.

The concentration data of each gas species in the wavelength range 290–350 nm were obtained sequentially by taking logarithms, filtered, and then followed with the non-linear fitting of the absorption spectra for a set of measurements. From the differential optical depth ($D'_{\text{meas}}(\lambda_k) = \ln(I_0(\lambda)/I(\lambda))'$) determined by two spectra, the respective values (c_i , α_j) of i th species' concentration and a coefficient of the polynomial of degree j were acquired by the following fitting process (Platt, 1994), which complies with the Beer–Lambert law:

$$\left| \sum_k \left\{ D'_{\text{meas}}(\lambda_k) - L \left(\sum_i c_i \sigma'_i(\lambda_k) - \sum_j \alpha_j \lambda_k^j \right) \right\} \right| = \text{Residual.} \quad (1)$$

Here $\sigma'_i(\lambda_k)$ denotes the differential cross-section of the i th trace gas at the wavelength of the k th pixel of the photodiode. The residual term shown in Eq. (1) is left after

fitting to atmospheric spectra; these residual term values then tend to vary within a few per mil (‰). The residuals may stem from the combined effects of the following factors: the atmospheric absorption of unidentified species, the irremovable instrumental spectral patterns, the detector state, the statistical fit errors of the least square (LSQ) function, and the erroneous matches of the measurement and reference spectra. Therefore, its structure has an important role in determining a lower detection limit (DL).

2.2. Trace gas species measurements

All DOAS measurements were carried out on top of a three-story environmental monitoring research center building at the GIST during the period from 11 through 22 March 2002. The PMS was also operated for an intercomparison purpose from the 14th through the 21 of March 2002. The measurement site in the GIST is located on the northern border area of Gwangju city, Korea (35.10N, 126.53E) which is surrounded by a large rice paddy area (Fig. 1). Gwangju is the fifth largest city in Korea with a population of 1.4 million.

The two spectral regions used for the evaluation were set to cover: (1) the 293–312 nm band for SO_2 , O_3 , and HCHO and (2) the 325–342 nm band for NO_2 and HNO_2 . The length of the light path was set at 1500 m (750 m in both ways) with a height of 6–15 m above ground. The ambient air spectrum was acquired with an integration time of 120 s (3 s exposure time and 40 scans for each). For an intensive intercomparison analysis, all measurement data were averaged at hourly intervals. Three types of spectra (ambient air, background, and lamp) measured automatically were put together for the derivation of the concentration of each individual gas species. Prior to such calculations, each spectrum was processed to remove the electronic offset and the dark current. Because the background scattering spectra were superposed on the receiving light spectra (during daytime), the background spectrum was subtracted from the atmospheric spectrum (Lee et al., 2005). The background spectrum corresponds to the atmospheric spectrum measured after the light source (here a Xenon lamp) was blocked. A shutter in front of the light source in our instrument (Fig. 1(1) of Lee et al., 2005) was used to measure the diffusive background light, which was scattered into the telescope. The light level (LL) in this study was hence used to express the practically measurable LL compared to the fully obtainable LL as the barometer of LL. The spectrum for the LL of 100% was determined as the one measured at the time when the received light intensity reached a maximum during the measurement period.

Information concerning the total uncertainty and the DL for each gas is presented in Table 1(a) (Lee, 2004). To begin with, the DL can be calculated as the minimum detectable optical density from a residual structure depending on its magnitude. In this study, the DL was actually determined as three times (Ingle and Crouch, 1988) the standard

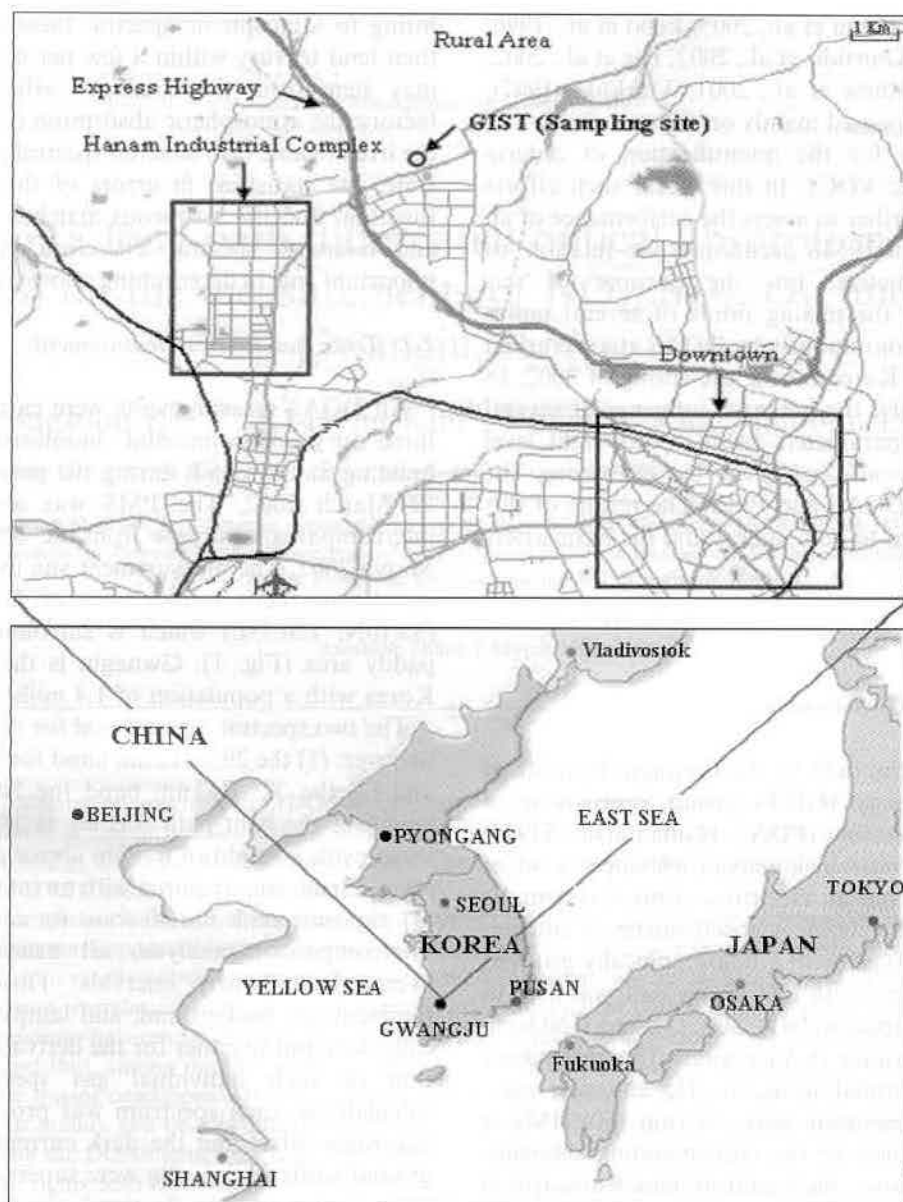


Fig. 1. A geographical location of the air quality monitoring site in Gwangju city, Korea.

deviation of fit results obtained after evaluating 30 randomly shuffled residuals (i.e. noise) at clean daytime conditions (visibility ~ 20 km). Briefly, the uncertainty was calculated from the square root of the sum of squares of the standard deviations (1σ) of factors, such as light source stability, light path length, absorption, absorption cross-section spectra from reference and evaluation process (fitting process).

The intensive intercomparison study introduced above was conducted to directly evaluate the relative performance of a line-integrating monitoring technique (DOAS) and a conventional technique (PMS). The PMS consisted of three separate gas analyzers (manufactured by environment sa). The concentration of O_3 was measured by a UV photometry ozone analyzer (O341M) at 253.7 nm and by the UV fluorescence method (AF21M) at 213.9 nm, respectively.

The concentrations of NO_2 and HNO_2 were also monitored by a chemo-luminescent nitrogen-oxide analyzer (AC31M) with UV fluorescence and by a channel glass-coil sampling method with a chemical treatment (Park et al., 2004). The DL and the drift information of all PMS are also presented in Table 1(b). As shown in Table 1, the LP-DOAS used for the analysis of O_3 was not as sensitive as the PMS (O341M). However, where the detectability of SO_2 is concerned, the LP-DOAS can be superior to the PMS method (AF21M).

2.3. Meteorological conditions

A multitude of meteorological parameters (including solar radiation, visibility, and precipitation) were monitored at hourly intervals from weather stations at a

Table 1
The basic analytical properties of LP-DOAS and PMS investigated in this study

(a) LP-DOAS^a

Species	NO ₂	SO ₂	O ₃	HNO ₂	HCHO
DL (ppbv)	0.50	0.20	4.26	0.17	1.52
Relative uncertainty (%)	8.0	4.5	13.5	8.0	8.0

(b) PMS

Species	NO ₂ (AC31M)	SO ₂ (AF21M)	O ₃ (O341M)	HNO ₂ channel glass coil sampling
DL (ppbv)	0.35	1.0	0.4	0.2
Noise (ppbv)	0.17	0.5	0.2	—
Span drift	<1%/day	<1%/day	<1%/day	—
Zero drift	<1 ppb/week	<1 ppb/week	<1 ppb/week	—

^aThe detection limits of DOAS system are determined by the atmospheric spectra measured over a 1500 m path length at 30% light level of the maximum light level (VR ~20 km).

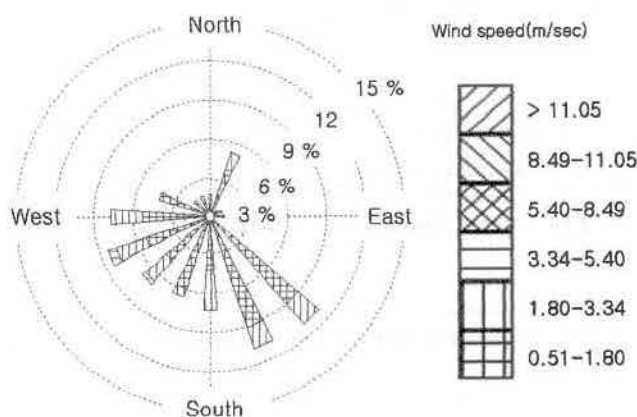


Fig. 2. A diagram of the windrose pattern at the study site. The wind direction data were obtained at hourly intervals during the entire study period.

Gwangju meteorological office (GMO). In addition, a number of basic meteorological parameters (wind speed, wind direction, temperature, relative humidity (RH), and particulate mass) were also measured from the GIST monitoring site. The meteorological conditions during the whole study period can be summarized as follows. The prevailing wind direction was southerly, as shown in Fig. 2. The mean wind speed was computed to be 2.6 m s^{-1} . The mean temperature was measured as 10.1°C with a mean RH of 55%. The observed results thus represent typical weather conditions of a suburban area in the spring season of Korea. It was also noted that three Asian dust (AD) storm events occurred during our study period on the 17, 19, and 21 March (Park, 2002; Park and In, 2003). On 21 March, it was severe enough to impair visibility to less than 500 m distance.

In order to check for the possible existence of emission sources near the measurement site, the observed concentration data were initially analyzed in relation with the windrose pattern. It indicated that there was no distin-

guishable emission source around the measurement station; such a prediction is realistic, if one considers that the pollutant distribution patterns were very similar to the wind rose pattern in Fig. 2. As can be seen from the wind rose pattern (Fig. 2), most winds came from the downtown area and the industrial complex located in the southern part of the measurement site (refer to the map in Fig. 1). This situation may influence the distribution characteristics of each pollutant, as there are no identifiable sources around the measurement site. Hence, the sources of air pollution in the study area are expected to be controlled by a number of factors: transport from the southern area (by wind), photochemical oxidation phenomenon (for O_3), or direct emissions of traffic sources from the surrounding suburban area, similar to a previous study made in Taiwan (Chiu et al., 2005).

3. Results and discussion

3.1. The overall patterns of DOAS monitoring

Fig. 3 shows the time series plots of the trace gas species and the LL data for the entire study period. It has been reported that the reduction of the light intensity can impair the performance of DOAS to a certain extent (Virkkula, 1997). In this study, a large portion of DOAS data was also lost by the frequent reduction in LL. The prevalence of heavy fog (in most mornings) and/or dust storm events principally contributed to the attenuation of light which then caused the loss of data.

A statistical summary of our pollutant data acquired by both DOAS and PMS is given along with the LL data observed by the DOAS system (Table 2). According to this analysis, the following was generally found: (1) the average concentrations of the DOAS are generally smaller than their PMS counterpart; and (2) the data loss rate for DOAS was comparable to (or lower than) the PMS depending on the trace species. The cause of the former

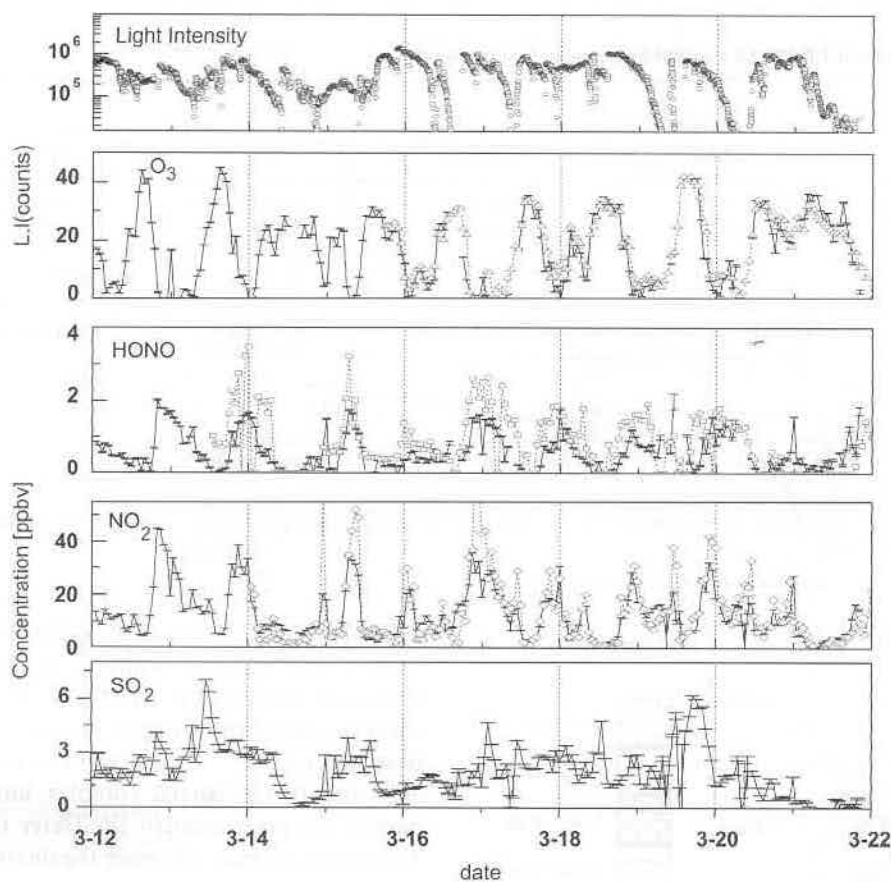


Fig. 3. Comparison of the pollutant concentrations measured by the DOAS system (solid lines) with those of the PMS (open triangles for O_3 , open squares for $HONO$, and open diamonds for NO_2). The vertical bars indicate the standard deviation of the hourly data of DOAS data. Light intensity is shown here as an open circle.

may be somewhat natural, because the path integration monitoring technique (like DOAS) yields the average mixing ratio over a given path length. However, the differences in the average values measured by each system seem to fall in a moderately small range, if one considers the uncertainty of each individual system (see Table 1). Hence, one may conclude that the two separate instrumental systems maintained a fairly good compatibility during the intercomparison period.

In Table 2, the data loss rates for both systems are evaluated and compared to each other. The results indicate that the loss rates of DOAS for each trace gas (NO_2 , $HONO$, SO_2 , and O_3) correspond to 22%, 23%, 20%, and 21%, respectively, while their PMS counterparts correspond to 21%, 28%, 24%, and 24%. The main reasons for such data loss can be ascribed to the low pollution level in comparison to the poor DLs under the severe environmental conditions (frequent fog, precipitation (rain), dust storm, etc.). However, during our study period, the loss rate of DOAS-based data was even smaller than that of PMS; the gradual deterioration of the PMS inlet by the precipitation and dust storm may have acted as an important factor to affect the data acquisition rate for the traditional method.

Ozone typically showed a diurnal variation pattern with a daytime maximum (afternoon) and a nighttime minimum (around midnight) (see Fig. 3). However, as it is well known, the temporal distributions of NO_2 and $HONO$ are opposite to that of O_3 (Seinfeld and Pandis, 1998; Finlayson-Pitts and Pitts, 1999). In general, O_3 as a secondary photochemical pollutant is formed in the atmosphere via the reaction of the photodissociation of NO_2 . O_3 in the surface layer is hence expected to decrease after sunset, while NO_2 behaves contrary to O_3 . The SO_2 data measured during the study period did not show clear diurnal variation, but the pattern seemed to be affected rather sensitively by the traffic conditions over the daily cycle. In Fig. 3, the concentrations of some pollutants (SO_2 , NO_2 , and $HONO$) are commonly seen to approach their DL during most of the study period. Such phenomena were more significant, especially during the rainy and heavy dust storm periods (e.g., 14, 17, 20, and 21 March 2002).

Interestingly, we were able to find a significant but unusual increase in O_3 mixing ratio during some nighttime periods (e.g., 14, 18, and 20th), contrary to the general pattern with the daytime maximum (due to photochemical reactions). Such an elevated O_3 concentration level seemed

Table 2

A statistical summary of the DOAS and PMS hourly data measured and the meteorological parameters measured concurrently during the study period (all mixing ratios in ppb unit)

(a) Pollutants

	Light level (%)	DOAS system (DOAS)				Point monitoring system (PMS)			
		NO ₂	HNO ₂	SO ₂	O ₃	NO ₂	HNO ₂	SO ₂	O ₃
Mean	30.7	11.3	0.5	1.9	17.1	13.9	0.8	2.8	18.7
Median	27.6	8.8	0.4	1.7	19.3	10.0	0.8	2.0	21.0
SD	21.9	8.4	0.5	1.2	12.1	8.4	0.7	2.3	11.3
Min	0.5	0.2	-0.2	-0.0	-0.03	0	0	0	0
Max	100.0	35.8	1.8	6.2	41.5	61	2.6	9.0	42
¹ N _{eff}	166	147	145	150	149	149	136	142	112
N	188	188	188	188	188	188	188	188	147

(b) Met. data

	Wind speed (m s ⁻¹)	T (°C)	RH (%)	SR (MJ m ⁻²)
Mean	2.9	10.5 ±	57.3	1.2
Median	2.6	10.2	56.0	0.8
SD	2.0	4.3	22.5	0.9
Min	0	0.4	10	0
Max	8.6	22.4	92	3.0
N	188	188	166	104

The number of data has been fixed as 188 during the comparison study (from 14 to 21 March 2002) between DOAS and PMS. In addition, N_{eff} is also given to represent the effective number of the concentration data measured at light level above 0.5%. However, if the total number of data is concerned for the entire measurement period (from 11 to 22 March 2002), it goes up to 266 in the case of some meteorological parameters.

to occur under the conditions of low wind speed at high temperature. It can thus be speculated whether this is due to the transport of external ozone or to the formation of a second O₃ maxima at nighttime (nocturnal ozone) (Evanagelisti et al., 1995; Salmond and Mckendry, 2002).

3.2. Correlation analysis between DOAS and PMS

As a means to evaluate the relative performance of the two measurement techniques, a correlation analysis was conducted. As explained by many previous studies conducted for intercomparative purposes, the correlation coefficient (r^2) can be used as a simple barometer to judge the compatibility between the DOAS and PMS techniques (Stevens and Conner, 1991; Martini et al., 1994; Virkkula, 1997). There has been a good consensus that the strength of correlation is affected most effectively by such factors as (1) instrumental calibration errors, (2) atmospheric homogeneity, (3) visibility, and (4) proportion of below detection limit (BDL) values (Stevens and Conner, 1991; Martini et al., 1994; Virkkula, 1997). In the course of the present study, the effects of (1) and (4), affected by the instrumental conditions, have already been examined in Sections 2.1 and 3.1, respectively. Hence, the effects of meteorological parameters (2) and (3) are investigated more intensively below.

The results of the correlation analysis are summarized in Table 3. If the extent of compatibility between the two methods is assessed by the magnitude of the correlation

Table 3

The results of linear regression analysis between DOAS and PMS

	r^2	Correlation equation (unit in ppb)
NO ₂	0.89 (147) ^a	Y(PMS) = 1.30 (0.06) × (DOAS) - 0.72 (0.79)
HNO ₂	0.46 (133)	Y(PMS) = 0.66 (0.11) × (DOAS) + 0.52 (0.08)
SO ₂	0.37 (114)	Y(PMS) = 0.64 (0.13) × (DOAS) + 1.37 (0.30)
O ₃	0.94 (110)	Y(PMS) = 0.83 (0.03) × (DOAS) + 4.91 (0.61)

^aNo. of data is given in the parenthesis.

coefficient, it is generally ranked in the order of O₃, NO₂, HNO₂, and SO₂. In spite of the complicated meteorological conditions and the generally low pollutant concentrations in the study area, the observed compatibility for NO₂ and O₃ is fairly high, in compliance with the previous studies (Lee et al., 2005; Febo et al., 1996; Martini et al., 1994; Stevens and Conner, 1991). However, as the correlation coefficients derived for SO₂ and HNO₂ were fairly poor, the discussion of compatibility is directed mainly to NO₂ and O₃ in this study. Probably, our study site may have been influenced by some sources of SO₂ and HNO₂ on an irregular basis during the entire study period (Table 2(a) and Fig. 2).

3.3. The influence of LL and wind speed on the inter-instrumental compatibility

The results of correlation analysis in Table 4 (and Fig. 4) are compared as a function of two major parameters such

Table 4

Results of correlation analysis between DOAS and PMS obtained for the two major meteorological parameters: (a) light level (LL) and (b) wind speed (WS)

(a) Light level

LL (%)	HNO ₂	NO ₂	O ₃	SO ₂
0–5	—	—	—	—
6–20	0.57 ¹	1.41	0.72	0.09
	0.46 ²	–2.33	5.10	1.96
	0.26 ³	0.81	0.92	—
21–40	0.68	1.18	0.84	0.88
	0.55	0.82	5.04	0.75
	0.22	0.84	0.89	0.16
41–60	0.88	1.47	0.93	1.05
	0.36	–1.33	2.82	1.47
	0.25	0.73	0.9/	0.31
61–100	1.00	1.07	0.78	1.54
	0.61	0.09	6.52	0.83
	0.12	0.67	0.8	0.21

(b) Wind speed

WS (m s ^{–1})	HNO ₂	SO ₂	NO ₂	O ₃
0–0.9	0.90 ¹	–0.29	1.64	0.80
	0.46 ²	2.93	–6.07	3.51
	0.3 ³	0.01	0.63	0.73
1.0–1.9	0.59	0.10	1.17	0.83
	0.65	2.36	–0.06	4.28
	0.19	0	0.81	0.81
2.0–2.9	0.62	0.53	1.23	0.79
	0.61	1.36	1.53	5.46
	0.18	0.07	0.77	0.91
3.0–4.9	0.62	0.57	1.46	0.64
	0.51	1.52	–3.10	10.9
	0.13	0.15	0.84	0.73
5.0–8.6	0.02	1.12	1.46	0.75
	0.39	0.89	0.18	7.85
	0	0.52	0.79	0.8

Superscripts 1, 2, and 3 denote slope, offset, and correlation coefficient, respectively.

as LL and wind speed (WS). To examine the direct influence of the LL on the DOAS performance, a correlation analysis was conducted after dividing the LL data sets into five separate groups (Table 4(a)). At the LL values below 5% (assuming the fully received LL as 100%), the correlation was insignificant. However, the correlation strength improved significantly above 5%. For O₃ and NO₂, the correlation coefficient increased rapidly at LL values beyond 5%. The level of correlation with LL, however, seems to remain nearly constant in that range, as shown in Fig. 4(a). The correlation coefficients of SO₂ and HNO₂ also exhibited a trend similar to those of the above two gases. It thus suggests that changes in LL are not efficient enough to affect the instrumental compatibility beyond a certain level.

To further elucidate the effects of environmental factors on the DOAS performance, the wind speed data sets were examined in a manner similar to LL. The correlation patterns derived based on the grouping of wind speed data

sets, however, resulted in a very unique pattern (Fig. 4(b)). When the wind speed was less than 1 m s^{–1}, the correlations became the weakest for most pollutants, except for HNO₂. In the meantime, the correlation coefficients of SO₂ tended to grow monotonously with wind speed. The correlation coefficients for O₃ and NO₂ were, however, found to peak in the intermediate wind speed ranges of 2–3 and 3–5 m s^{–1}, respectively. In compliance with the general findings of previous studies (Kim, 2004; Kim and Kim, 2001; Martini et al., 1994), the results of this study indicate the complicated relationship between the wind speed criterion and each individual target analyte. In the case of Martini et al. (1994), two data sets were analyzed after being sorted by the wind speed criterion of 2 m s^{–1} (obtained from the ozone correlation analysis). These authors observed a notable enhancement of correlation values (r^2) for both SO₂ and NO₂ data sets. In addition, there was also an attempt to analyze wind speed criterion in relation with the absolute concentration differences between the two systems; for instance, it was found that differences in NO₂ concentrations between the two systems tended to be reduced with increasing wind speed (Kim and Kim, 2001). The concentration differences between the two systems also tended to converge to zero with the rise of wind speed (in the case of O₃, SO₂, and NO₂ in this study). Therefore, it clearly indicates that the wind speed data can be used effectively to diagnose the compatibility between the two systems, because an increase of wind speed can generally enhance the extent of atmospheric mixing.

3.4. Diurnal variation patterns during the intercomparison study

Knowing that most trace species and the meteorological parameters exhibited distinct diurnal variations (Fig. 3), their temporal patterns were examined as another means to evaluate the DOAS performance (Fig. 5). As shown in Fig. 5, many meteorological parameters (such as temperature, wind speed, and RH) exhibited clear diurnal variation patterns, while being significantly correlated with each other ($r^2 > 0.87$). The meteorological conditions are in fact distinguished between daytime and nighttime during the entire period. During the daytime ($N = 143$), the winds came dominantly from the south at a mean wind speed of 3.2 m s^{–1}. The mean temperature was measured as 11.8 °C with a RH of 51%. In the nighttime ($N = 123$), the winds were southerly or northerly at a mean wind speed of 2.0 m s^{–1}. During the night, the mean temperature was observed to be 8.0 °C at 60% RH.

The results obtained for NO₂ and O₃ generally exhibited a strong compatibility between the two systems over the diurnal scale. Their correlation coefficients (r^2), when derived using diurnal data sets, were 0.89 and 0.94, respectively. In contrast, the compatibility of SO₂ and HNO₂ are not so good in that they show very poor correlations (with r^2 values of 0.04 and 0.30, respectively). (An extended study between the two systems may be

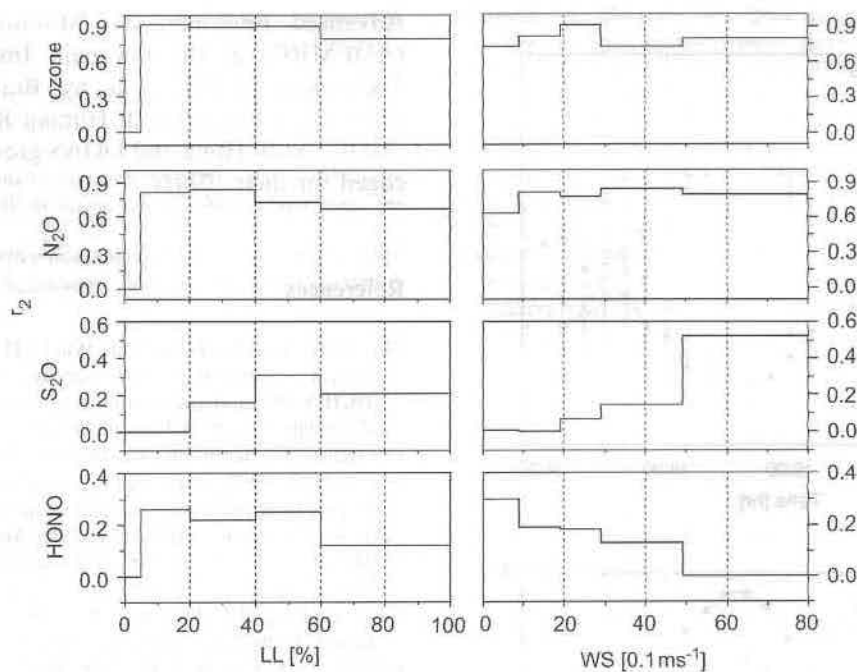


Fig. 4. The compatibility between DOAS and PMS is expressed in terms of correlation coefficients (r^2). All results are compared as a function of (a) light level (LL) and (b) wind speed (WS).

needed to explain the weak compatibility of those compounds in the future.) An examination of the diurnal variation patterns measured by both systems also indicates that both O_3 and NO_2 can exhibit considerably high correlation coefficients (r^2) with temperature, solar irradiation, and wind speed: (1) in the case of O_3 , the correlation results between DOAS-derived O_3 and those meteorological parameters were: 0.73, 0.66, and 0.66, while those of PMS 0.81, 0.49, and 0.74, respectively. (2) NO_2 data for DOAS showed values of 0.38, 0.59, and 0.34 and for PMS 0.21, 0.27, and 0.23, respectively. It was thus found that ozone, out of the four pollutants investigated in this study, exhibited the strongest correlation with temperature, RH, wind speed, LL, and solar irradiation for both measurement techniques. These results appear to be quite compatible with those of previous studies (Chiu et al., 2005; Martini et al., 1994; Evangelisti et al., 1995; Virkkula, 1997). Based on this comparative analysis, it can be concluded that O_3 is a very sensitive indicator to assess the homogeneity of air masses (Martini et al., 1994).

4. Conclusions

In this study, we performed an intercomparative analysis of two different measurement techniques (DOAS and PMS) using the data sets for four trace gas species (NO_2 , O_3 , SO_2 , and $HONO$). The measurements were carried out from a suburban area (GIST station) in Gwangju city, South Korea during a 10-day study period. The results obtained by the two independent systems were examined to evaluate the compatibilities, mainly in terms of correlation analysis. The influence of the LL and the wind speed was

examined with a primary focus on evaluating the degree of the compatibility between each pollutant data set obtained by the two systems.

The results of our analysis indicated that wind speed can be used as one of the key variables to explain the instrumental compatibility of DOAS with others. Such compatibility can in fact be optimized under certain meteorological conditions (e.g., good vertical and horizontal mixing). At present, the data sets for $HONO$ and SO_2 were not examined fully because of relatively poor compatibility, and the comparison for $HCHO$ was not made because of the absence of matching data pairs from PMS. In a future study, it will be desirable to examine the compatibility of other gases and their dependence on the environmental conditions.

The overall results of our study can be summarized as follows:

- (1) If the magnitudes of all target analytes were compared each other, the DOAS-values were generally lower than those of PMS.
- (2) The correlation coefficients (r^2) of NO_2 and O_3 , when computed for the two measurement systems, were 0.89 and 0.94, respectively. The strengths of correlation between DOAS and PMS were very poor at LLs below 5%. However, the correlation values increased significantly at LLs above 5% and remained almost constantly so after that.
- (3) The correlation results of O_3 and NO_2 were found to be the highest in the wind speed range $2.0\text{--}3.0\text{ m s}^{-1}$ ($r^2 \sim 0.81$) and $2.0\text{--}5.0\text{ m s}^{-1}$ ($r^2 \sim 0.84$), respectively. However, other than those ranges, their correlation

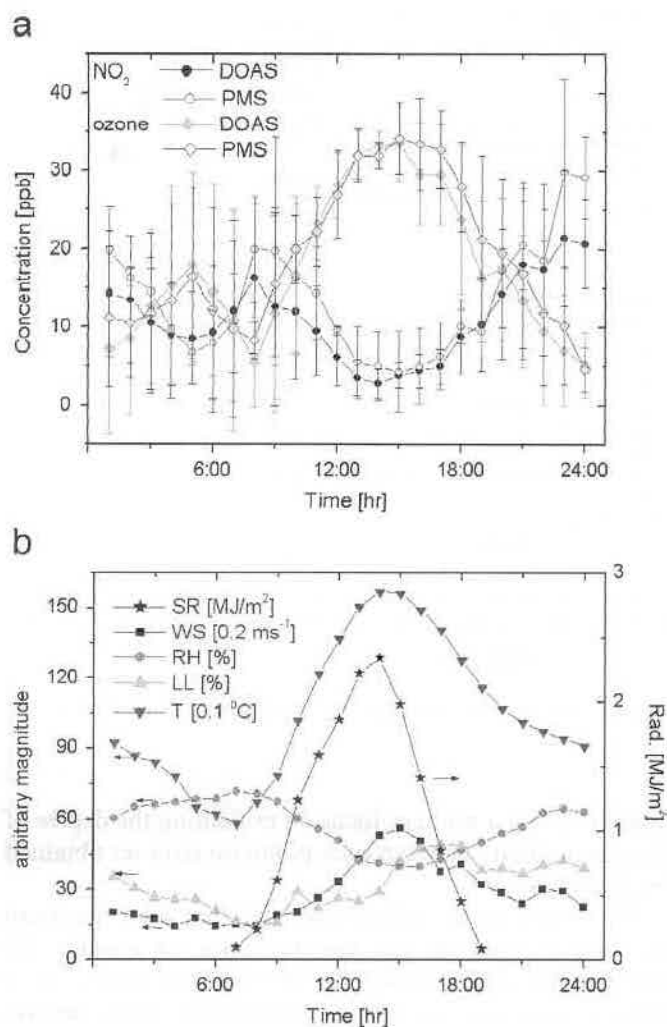


Fig. 5. Comparison of diurnal variation patterns for various pollutants and meteorological parameters: (a) NO₂ and O₃ concentrations (with SD shown as vertical bar) measured by the two systems and (b) meteorological parameters (wind speed (WS), relative humidity (RH), light level (LL), temperature (T), and solar irradiation (SR)) during the study period.

strengths were reduced to $r^2 \sim 0.73$ and ~ 0.63 , respectively. The correlation strength (r^2) of SO₂ increased (from 0.0 to 0.52) monotonously with the growth of wind speed, while the reverse pattern was apparent for HNO₂ (from 0.3 to 0.0).

When all the measurement data were examined, O₃ was affected most sensitively by the changes in meteorological conditions. For O₃ measurement, PMS was found to respond as sensitively as DOAS to effectively reflect the changes in environmental conditions (e.g., the variation of wind speed, RH, LL, and temperature).

Acknowledgments

The authors are pleased to acknowledge the constructive criticism made by two anonymous reviewers.

This experiment was supported by the Korea Science and Engineering Foundation (KOSEF) through the

Advanced Environmental Monitoring Research Center (ADEMRC) at the Gwangju Institute of Science and Technology (GIST) and by Brain Korea 21 project, Ministry of Education & Human Resources Development (MOE). S.-B. Hong and DOAS group are greatly acknowledged for their efforts.

References

- Chiu, K.H., Sree, U., Tseng, S.H., Wu, C.-H., Lo, J.-G., 2005. Differential optical absorption spectrometer measurement of NO₂, SO₂, O₃, HCHO and aromatic volatile organics in ambient air of Kaohsiung Petroleum refinery in Taiwan. *Atmospheric Environment* 39, 941–955.
- Evangelisti, F., Baroncelli, A., Bonasoni, P., Giovannelli, G., Ravegnani, F., 1995. Differential optical spectrometer for measurement of tropospheric pollutants. *Applied Optics* 34 (15), 2737–2744.
- Febbo, A., Perrino, C., Allegrini, I., 1996. Measurement of nitrous acid in Milan, Italy, by DOAS and diffusion denuders. *Atmospheric Environment* 30, 3599–3609.
- Finlayson-Pitts, B.J., Pitts Jr., J.N., 1999. *Chemistry of the Upper and Lower Atmosphere*. Academic Press.
- Ingle Jr., J.D., Crouch, S.R., 1988. *Spectrochemical Analysis*. Prentice-Hall, Englewood Cliffs, NJ.
- Kim, K.-H., 2004. Comparison of BTX measurements using a commercial differential optical absorption spectroscopy and an on-line gas chromatography system. *Environmental Engineering Science* 21 (2), 181–194.
- Kim, K.-H., Kim, M.-Y., 2001. Comparison of an open path differential optical absorption spectroscopy system and a conventional in situ monitoring system on the basis of long term measurements of SO₂, NO₂, and O₃. *Atmospheric Environment* 35, 4059–4072.
- Kourtidis, K.A., Ziomas, I., Zerefos, C., Kosmidis, E., Symeonidis, P., Christophilopoulos, E., Karathanassis, S., Mploutsos, A., 2002. Benzene, toluene, ozone, NO₂ and SO₂ measurements in an urban street canyon in Thessaloniki, Greece. *Atmospheric Environment* 36, 5355–5364.
- Lee, J.S., 2004. Simultaneous monitoring of atmospheric trace gases and atmospheric spectral extinction using a LP-DOAS system. Doctoral Dissertation, Gwangju Institute of Science and Technology (GIST), Korea.
- Lee, J.S., Kuk, B.J., Kim, Y.J., 2002. Development of differential optical absorption spectroscopy (DOAS) system for detection of atmospheric species: NO₂, SO₂ and O₃. *Journal of Korean Physics Society* 41 (5), 693–698.
- Lee, J.S., Kuk, B.J., Kim, Y.J., 2005. Simultaneous measurements of atmospheric pollutants and visibility with a long path DOAS system in urban areas. *Environmental Monitoring Assessment* 104, 281–293.
- Martini, L., Sladkoic, R., Slmr, F., Werle, P., 1994. Monitoring of air pollutants: long-term intercomparison of DOAS with conventional techniques. In: *Proceedings of the 87th Annual Meeting of Air and Waste Management Association*, Cincinnati, USA, June 1994, pp. 19–24.
- Mathew, L., Tai, W.R., Lo, J.-G., 2001. Measurements of sulfur dioxide and formaldehyde in Taipei using a differential optical absorption spectrometer. *Journal of Air & Waste Management Association* 51, 94–101.
- Park, S.-U., 2002. Field survey of Yellow Sand source regions. In: *Proceedings of Workshop of Asian Dust*, 22 March 2002, Korea Meteorological Administration, Korea.
- Park, S.-U., In, H.-J., 2003. Parameterization of dust emission for the simulation of the Yellow Sand (Asian dust) observed in March 2002 in Korea. *Journal of Geophysical Research* 108 (D19), 4618.
- Park, S.S., Hong, S.B., Jung, Y.G., Lee, J.H., 2004. Measurements of PM10 aerosol and gas-phase nitrous acid during fall season in a semi-urban atmosphere. *Atmospheric Environment* 38, 293–304.

- Perner, D., Platt, U., 1979. Detection of nitrous acid in the atmosphere by differential optical absorption. *Geophysical Research Letters* 6, 917–920.
- Plane, J.M.C., Nien, C.-F., 1992. Differential optical absorption spectrometer for measuring atmospheric trace gases. *Review of Scientific Instruments* 63 (3), 1867–1876.
- Platt, U., 1994. Air monitoring by spectroscopic techniques. In: Markus, W.S. (Ed.), *Chemical Analysis Series*, vol. 127. Wiley, New York, pp. 27–84.
- Platt, U., Stutz, J., 2006. *Differential Optical Absorption Spectroscopy (DOAS). Principle and Application*. Springer, Heidelberg. ISBN:3-340-21193-4.
- Salmond, J.A., Mckendry, I.G., 2002. Secondary ozone maxima in a very stable nocturnal boundary layer: observations from the lower Fraser Valley. *Atmospheric Environment* 36, 5771–5782.
- Seinfeld, J.H., Pandis, S.N., 1998. *Atmospheric Chemistry and Physics*. Wiley-Interscience, New York, pp. 235–336.
- Stevens, R.K., Conner, T.L., 1991. Air quality monitoring with the differential optical absorption spectrometer. In: *Proceedings of the Remote Sensing of Atmospheric Chemistry*, Los Angeles, USA, 1991, SPIE 1491, pp. 56–67.
- Virkkula, A., 1997. A performance of a differential optical absorption spectrometer for surface O₃ measurements in the Finish arctic. *Atmospheric Environment* 31, 545–555.

Erratum

Erratum to “Habitat inventory at a regional scale: A comparison of estimates of terrestrial Broad Habitat cover from stratified sample field survey and full census field survey for Wales, UK” [J. Environ. Manage. 85 (2007) 224–231]

T.H. Blackstock, C.R. Burrows, E.A. Howe, D.P. Stevens, J.P. Stevens*

Countryside Council for Wales, Maes-y-Ffynnon, Penrhosgarnedd, Bangor, Gwynedd LL57 2DW, UK

Available online 10 January 2008

The Publisher regrets that a number of errors were introduced to the above article published in the Journal of Environmental Management after the receipt of proof corrections. In eight places, the term ‘1 km square’ should have read ‘1 km squares’. Additionally, ‘2.2 Countryside survey’ should have read ‘2.2 Countryside Survey’, and ‘country side survey’ should have read ‘Countryside Survey’.

We apologise for any inconvenience caused to the authors and readers of this paper.

DOI of original article: 10.1016/j.jenvman.2006.09.008

*Corresponding author. Tel.: +441248 385523; fax: +441248 385510.

E-mail address: j.stevens@ccw.gov.uk (J.P. Stevens).

A woman dances with her son

**because of brain surgery that
reduced her epileptic seizures**

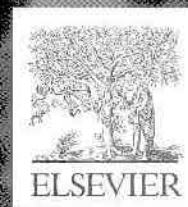
**performed by a neurosurgeon who was
able to pinpoint the foci of the seizure**

**due to breakthroughs in the
mapping of the human brain**

**advanced by physicians, mathematicians
and computer engineers around the world**

**inspired by new discoveries
in imaging technologies**

**reported in medical and
scientific publications.**



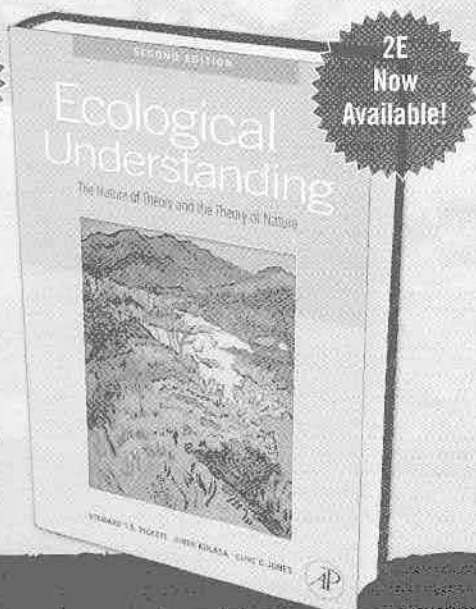
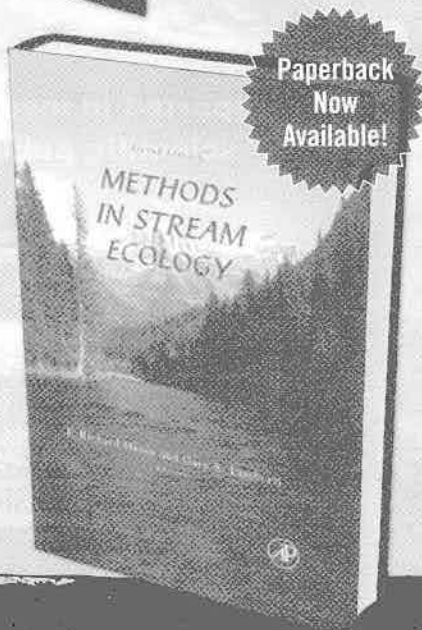
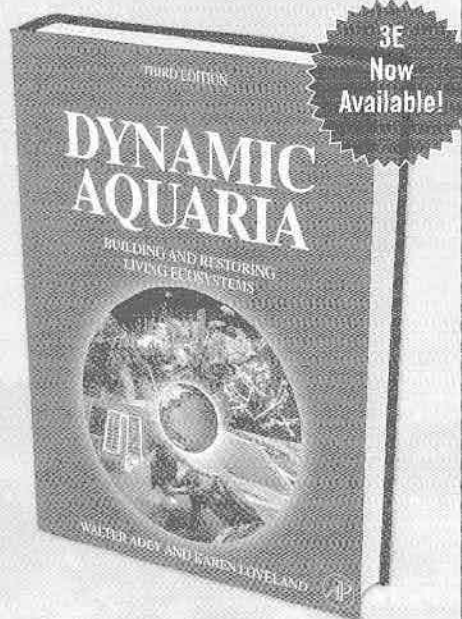
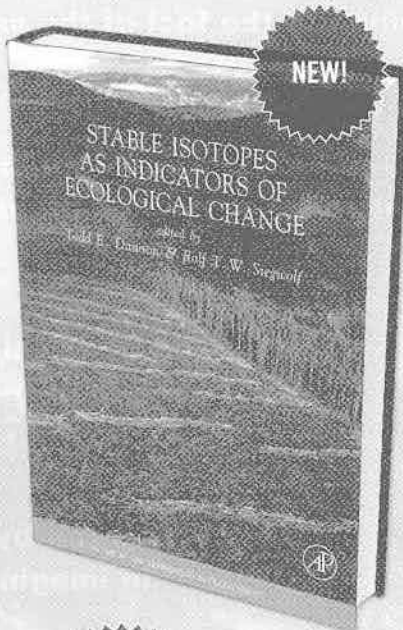
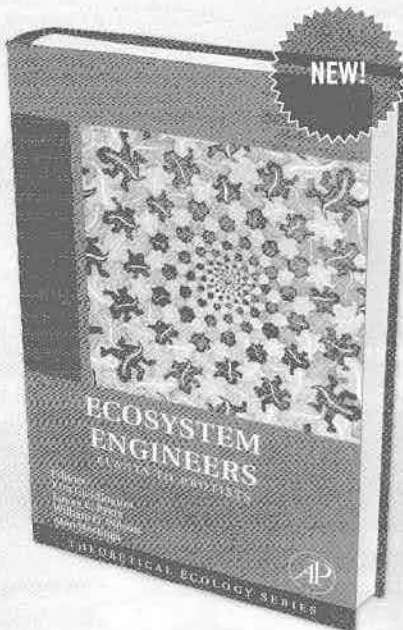
Building Insights. Breaking Boundaries.™

The creation of knowledge depends on a continuous line of connections between people and ideas. That's why, for over a century, Elsevier has linked researchers and practitioners worldwide to the best and newest thinking in their fields. As a professional publisher of scientific and medical information, Elsevier facilitates the review, selection, dissemination and preservation of information, publishing thousands of journals and books and providing access to a world of information via the desktop or PDA. After all, we believe that getting the right information into the right hands can make a difference. elsevier.com

NEW!

ECOLOGY RESEARCH TOOLS at your fingertips...

...from Academic Press



Visit books.elsevier.com/agbio today
for more information on these outstanding resources!



ACADEMIC PRESS
An imprint of Elsevier

H. He, J. Zhou, Y. Wu, W. Zhang and X. Xie

J. Lee, K.-H. Kim, Y.J. Kim and J. Lee

T.H. Blackstock, C.R. Burrows, E.A. Howe,
D.P. Stevens and J.P. Stevens

Modelling the response of surface water quality to the urbanization in Xi'an, China

Application of a long-path differential optical absorption spectrometer (LP-DOAS)
on the measurements of NO_2 , SO_2 , O_3 , and HNO_2 in Gwangju, Korea

Erratum to "Habitat inventory at a regional scale: A comparison of estimates of
terrestrial Broad Habitat cover from stratified sample field survey and full census field
survey for Wales, UK" [J. Environ. Manage. 85 (2007) 224–231]

731

750

760

Journal of
**Environmental
Management**

March 2008 Volume 86 Number 4

Contents

O.D. Yay, O. Alagha and G. Tuncel	Multivariate statistics to investigate metal contamination in surface soil	581
P.R. Kannel, S. Lee and Y.-S. Lee	Assessment of spatial-temporal patterns of surface and ground water qualities and factors influencing management strategy of groundwater system in an urban river corridor of Nepal	595
S. ud din, A. Al Dousari and P. Literathy	Evidence of hydrocarbon contamination from the Burgan oil field, Kuwait—Interpretations from thermal remote sensing data	605
R. Ranjan	Environmental restoration of invaded ecosystems: How much versus how often?	616
S.-C. Chang and C.-T. Lee	Evaluation of the temporal variations of air quality in Taipei City, Taiwan, from 1994 to 2003	627
G. Angerer, R. Nordbeck and C. Sartorius	Impacts on industry of Europe's emerging chemicals policy REACH	636
M. Wang, M. Webber, B. Finlayson and J. Barnett	Rural industries and water pollution in China	648
A. Vlyssides, E.M. Barampouti and S. Mai	Determination of granule size distribution in a UASB reactor	660
M. Marafi, A. Stanislaus and E. Kam	A preliminary process design and economic assessment of a catalyst rejuvenation process for waste disposal of refinery spent catalysts	665
J.R. Tian, P.J. Zhou and B. Lv	A process integration approach to industrial water conservation: A case study for a Chinese steel plant	682
B. Stone Jr.	Urban sprawl and air quality in large US cities	688
A.A. Keller and L. Cavallaro	Assessing the US Clean Water Act 303(d) listing process for determining impairment of a waterbody	699
N. Håkan Tarras-Wahlberg and L.T. Nguyen	Environmental regulatory failure and metal contamination at the Giap Lai pyrite mine, Northern Vietnam	712
F.M.W. de Jong, G.R. de Snoo and J.C. van de Zande	Estimated nationwide effects of pesticide spray drift on terrestrial habitats in the Netherlands	721

Continued on inside back cover

Abstracted/indexed in Gale Group Business A.R.T.S., Periodical Abs-Plustext, Zoological Record Online®, Science Citation Index, Science Citation Index Expanded, CC/Agriculture, Biology & Environmental Sciences, BIOSIS Previews, Biological Abstracts, GEOBASE®, Pascal, Embase, Environmental Sciences, CAB Abstracts, Social Sciences Abstracts, Elsevier BIOBASE, CSA Life Sciences Abstracts, Uncover, Medline®, Pollution Abstracts, Agricola, Biosis Toxline®, CSA Technology Research Database, ToxFile. Also 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(200803)86:4;1-R

Available online at



ScienceDirect

www.sciencedirect.com