

Journal of
Environmental
Management

58.408
LS

ol.88 No.3

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



27188

Journal of environmental management.



ELSEVIER

27188
000002287

Orders, claims, and journal enquiries: please contact the Regional Sales Office nearest you:

Orlando: Elsevier, Customer Service Department, 6277 Sea Harbor Drive, Orlando, FL 32887-4800, USA; phone: (877) 8397126 [toll free within the USA]; (+1) (407) 5636022 [outside the USA]; fax: (+1) (407) 3631354; e-mail: JournalCustomerService-usa@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: JournalsCustomerServiceEMEA@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: JournalsCustomerServiceJapan@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: JournalsCustomerServiceAPAC@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). 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>

A comparative study for the sorption of Cd(II) by K-feldspar and sepiolite as soil components, and the recovery of Cd(II) using rhamnolipid biosurfactant

Y. Aşçı^a, M. Nurbaşı^a, Y. Sağ Açık^{b,*}

^aDepartment of Chemical Engineering, Eskişehir Osmangazi University, 26480 Batı Meşelik, Eskişehir

^bDepartment of Chemical Engineering, Hacettepe University, 06800 Beytepe, Ankara, Turkey

Received 12 July 2006; received in revised form 30 January 2007; accepted 10 March 2007

Available online 25 April 2007

Abstract

This study investigated the sorption characteristics and recovery of selected heavy metal Cd(II) from K-feldspar and sepiolite, representative soil components, using rhamnolipid biosurfactant. Although the proposed technique was classified as a soil bioremediation process, it can also be applied to treatment of waste waters containing Cd(II) ions with minor modifications. The effect of initial Cd(II) concentration on sorption capacity was characterized by determining the sorption isotherms. Of the four models examined, the Freundlich model showed the best fit for the sorption of Cd(II) on K-feldspar, whereas the Langmuir-model was used successfully to characterize the sorption of Cd(II) on sepiolite. Although a high Cd(II) uptake of 7.49 mmol/kg by K-feldspar was obtained, sepiolite was a superior Cd(II) accumulator, with a maximum Cd(II) uptake of 24.66 mmol Cd(II)/kg. The presence of Cd(II) in the sepiolite or K-feldspar prior to addition of the rhamnolipid generally resulted in less rhamnolipid sorption to sepiolite or K-feldspar. The maximum Cd(II) desorption efficiency by rhamnolipid from K-feldspar was substantially higher than that of sepiolite and determined to be 96% of the sorbed Cd(II), whereas only 10.1% of the sorbed Cd(II) from sepiolite was recovered by rhamnolipid solution.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Soil remediation; K-feldspar; Sepiolite; Langmuir model; Freundlich model; Desorption; Rhamnolipid biosurfactant

1. Introduction

Due to its acute toxicity, cadmium together with lead and mercury forms the “big three” of heavy metals with the greatest potential hazard to humans and the environment. The industrial uses of cadmium are increasing in electroplating, paint pigments, plastics, silver–cadmium and nickel–cadmium battery industry and smelter operations, often affecting the environment at distances of a 100 km order of magnitude (Volesky, 1990; Farquhar et al., 1997). Anthropogenic activities such as industrial waste disposals and sewage sludge disposals on land have led to accumulation of Cd in soil (Vig et al., 2003). In addition,

intensive agriculture in developed countries has shown up the use of important quantities of intrans, i.e. fertilisers and amendments like activated sludges or manure (Lebeau et al., 2002). The relevant EU Directive, as well as the US EPA, and the FAO/WHO have set the maximum contaminant level Cd(II) cations in domestic water and in soil to be $5 \mu\text{g l}^{-1}$ and $70 \mu\text{g day}^{-1}$, respectively (Vig et al., 2003; Zouboulis et al., 2004).

Remediation of soil contaminated with toxic metal cations has conventionally involved the excavation and disposal in landfills or capping of the contaminated site. However, these traditional technologies do not actually remediate the soils, and very expensive (Mulligan et al., 2001a). The search for novel technologies has recently been directed to the application of bioremediation. Technologies using microorganisms and biosorbents to remove heavy

*Corresponding author. Tel.: +90 312 2992124; fax: +90 312 2977444.
E-mail address: yesims@hacettepe.edu.tr (Y. Sağ Açık).

metals from wastewaters have been extensively studied (Bruins et al., 2000; Sağ et al., 2003a,b). Biosorption involves a combination of active and passive transport mechanisms, starting with the diffusion of metal ions to the surface of microbial biomass. Metal ion uptake by passive transport mechanisms may involve the contribution of adsorption, chelation, complexation, coordination or micro-precipitation mechanisms (Yılmaz, 2003; Lin and Lin, 2005). Sorption by any of these mechanisms effectively limits the availability of metals for removal by flushing. Soil pores vary greatly in size ranging from less than 2 µm, whereas the size of a bacterial cell is as large as 0.2 mm in diameter. The smallest pores can act as a filter for metal-binding microorganisms and transport of the metal through the soil is prevented (Miller, 1995). As an alternative technology, biosurfactants produced by bacteria and yeasts could be used for the environmental remediation of heavy metals from soils, surface and subsurface waters (Christofi and Ivshina, 2002). With respect to size, biosurfactants have a distinct advantage over wholecells and exopolymers. In the pH range of 6.0–7.0, rhamnolipid structures are predominantly small vesicles, <50 nm in diameter, and micelles, approximately 5 nm in diameter (Tan et al., 1994; Herman et al., 1995). Surfactants have both hydrophilic (polar), usually referred to as the head, and hydrophobic/lipophilic (nonpolar), known as the tail, portions in the molecule (Zhang and Miller, 1992; Christofi and Ivshina, 2002; Zouboulis et al., 2003). The heads of the surfactant molecules point toward the surrounding aqueous solution while the tails face inward toward the hydrophobic center of the structure. The anionic biosurfactant carries a negative charge, so when the molecule encounters a cationic metal, an ionic bond is formed (Frazer, 2000; Kim et al., 2000). If this bond is stronger than the metal's bond with the soil, it will be enough to pump water through the soil to remove the contaminant. The biosurfactant used in this study, a rhamnolipid, was from the glycolipid group and was produced by *Pseudomonas aeruginosa*, a common soil microorganism (Sandrin et al., 2000; Mulligan and Wang, 2006).

Addition to type of biosurfactant, type of soil, pH of soil, cation exchange capacity, particle size, permeabilities and contaminants all affect metal removal efficiencies (Frazer, 2000; Alumaa et al., 2001; Mulligan et al., 2001b; Markiewicz-Patkowska et al., 2005; Shawabkeh, 2005; Pu and Cutright, 2006). When the type of soil changes, the efficiency of metal removal process from soil using biosurfactant will also be change. For that reason the metal desorption efficiency using rhamnolipid biosurfactant must be searched in various components of soil rather than in various soils. Otherwise, increasing number of experiments performed on various types of soils will not enough to obtain expected results for simulation of real heavy metal-polluted environment. Feldspars are the name of an important group of rock-forming minerals which make up perhaps as much as 60% of the Earth's crust,

commonly found in rocks, sediments, and soils. K-feldspar (KAlSi_3O_8 ; microcline or orthoclase) and Na-feldspar (albite; $\text{NaAlSi}_3\text{O}_8$) in a significant proportion of feldspar ores exist in the same matrix usually in quantities of about 3–5% Na_2O and K_2O (Farquhar et al., 1997; Demir et al., 2003). On the other hand, sepiolite [$\text{Mg}_4\text{Si}_6\text{O}_{15}(\text{OH}) \cdot 6\text{H}_2\text{O}$], a zeolite like clay mineral, is a hydrous magnesium silicate characterized by its fibrous morphology and intracrystalline channels. Sepiolite contains a continuous two-dimensional tetrahedral sheet of T_2O_5 ($\text{T} = \text{Si, Al, Be, ...}$) but has no continuous octahedral sheets. Sepiolite has molecular sized channels ($3.6 \times 10^6 \text{ \AA}$) and large specific surface area with more than $200 \text{ m}^2 \text{ g}^{-1}$ (Garcia-Sanchez et al., 1999; Sabah et al., 2002; Vico, 2003; Kocaoba and Akyuz, 2005). Because of this excellent fibrous structure, organic and inorganic ions can penetrate into the structure of sepiolite, and this makes it an important industrial raw material in sorptive, rheological and catalytic applications. Most of the world's sepiolite reserves are found in Turkey.

The objective of this study was to determine the sorption of Cd(II) ions onto K-feldspar and sepiolite as well as to research rhamnolipid-mediated desorption of Cd(II) ions from K-feldspar and sepiolite. There is little literature published about the sorption/desorption of heavy metal ions onto soil components such as K-feldspar, sepiolite, kaolin (Aşçı et al., 2007). This technique with small changes can also be applied to the removal and recovery of heavy metal ions from wastewaters using K-feldspar and sepiolite as sorbent and rhamnolipid as complexation or chelating agent. In addition, the sorption of rhamnolipid to K-feldspar and sepiolite was investigated. Little is currently known about the sorption of biosurfactant monomers or aggregate structures by soil components.

2. Materials and methods

2.1. Biosurfactant

The rhamnolipid used (JBR 425) was obtained from Jeneil Biosurfactant Co., LLC, Saukville, WI. The rhamnolipid used is a mixture of two major rhamnolipids, monorhamnolipid, α -L-rhamnopyranosyl- β -hydroxydecanoate (R1: $\text{C}_{26}\text{H}_{48}\text{O}_9$) and dirhamnolipid, 2-O- α -L-rhamnopyranosyl- α -L-rhamnopyranosyl- β -hydroxydecanoate (R2: $\text{C}_{32}\text{H}_{58}\text{O}_{13}$). Molecular weights of R1 and R2 are 504 and 650, respectively. The rhamnolipid aggregate morphology undergoes significant changes in the pH region of 6.5–7.5. At pH 5.0, rhamnolipid begins to visibly precipitate out of solution. The rhamnolipid is unstable at extreme pH due to hydrolysis of the glycosidic linkage between sugar and lipid. The biosurfactant is stable to 121 °C for at least 1 h and at room temperature. The rhamnolipid used has a critical micelle concentration (CMC) of 50 mg/L ($\approx 0.1 \text{ mM}$) and a surface tension of 29 mN/m (Jeneil Biosurfactant Co., LLC, 2001).

2.2. K-feldspar and sepiolite

The component of soil, K-feldspar, is used as raw material in ceramic factories. It was ground and sieved to 0.224–0.425 mm size. Chemical analyses of the K-feldspar and sepiolite samples were carried out on X-ray fluorescence spectrophotometer. Typical chemical properties of the K-feldspar used for sorption/desorption experiments are presented in Table 1. The analysis to this material revealed that it contain mainly SiO_2 , Al_2O_3 and K_2O , which they account for 96.2% of the total weight, respectively.

Sepiolite was obtained from sepiolite mines of Eskişehir region of Turkey. The chemical composition of sepiolite is also shown in Table 1. It contains mainly SiO_2 , MgO , which they constitute 76.7% of the total weight. The sepiolite samples were ground and sieved to 0.140–0.425 mm size.

2.3. Rhamnolipid sorption by K-feldspar and sepiolite

The K-feldspar and sepiolite pellets were autoclaved to prevent biodegradation of rhamnolipid. To measure rhamnolipid sorption, 1.5 g of K-feldspar or sepiolite pellets were suspended in 10 mL of varying concentrations of rhamnolipid (0–100 mM) in 50-mL centrifuge tubes. Control solutions contained the rhamnolipid solutions at the same concentration without added K-feldspar or sepiolite. The tubes were agitated on a shaker at 25 °C for 72 h, which is sufficiently long for sorption equilibrium. The samples were centrifuged at 19,000g for 20 min. Rhamnolipid concentration in the supernatant was estimated by surface tension measurement using a surface tensiometer (KRÜSS Digital Tensiometer K9, KRÜSS GMBH, Wissenschaftliche Laborgeräte, Borsteler Chaussee 85-99a, D-22453 Hamburg), which employs the Du Nouy ring method. Rhamnolipid concentration was

determined using a calibration curve relating surface tension (mN/m) to rhamnolipid concentration (mM). Calibration curves were prepared under conditions identical to those for each experiment.

2.4. Cd(II) removal by K-feldspar and sepiolite

To investigate Cd(II) recovery by rhamnolipids from the K-feldspar and sepiolite, firstly Cd(II) was loaded to the K-feldspar and sepiolite. Cd(II) solutions were prepared by diluting 8.9 mM stock solution of cadmium(II), obtained by dissolving $\text{CdN}_2\text{O}_6 \cdot 4\text{H}_2\text{O}$ in distilled water. The range of concentrations of prepared Cd(II) solutions varied from 0.25 to 5.32 mM. A 1.5 g of K-feldspar or sepiolite placed into 50-mL centrifuge tubes was treated in 10 mL of metal ion solutions. Control experiments were performed with the same metal ion solutions without added K-feldspar or sepiolite. Batch kinetic sorption experiments were conducted at pH 6.8, a constant agitation rate of 150 rpm and temperature of 25 °C. Batch type contact studies were continued for 72 h. After reaching ultimate equilibrium, the tubes were centrifuged at 19,000g for 20 min and the supernatant was acidified to pH 2.0 with 1% HNO_3 for atomic absorption analysis. The concentration of unadsorbed Cd(II) in the sample supernatant was determined using an atomic absorption spectrophotometer (ATI-UNICAM 929 England) with an air-acetylene flame. Sorbed Cd(II) concentration by K-feldspar or sepiolite was calculated from the difference between initial Cd(II) concentration in solution and Cd(II) concentration remained in the supernatant after sorption.

2.5. Recovery of Cd(II) by rhamnolipid biosurfactant from K-feldspar and sepiolite

The Cd(II)-sorbed K-feldspar or sepiolite was treated with 10 mL of increasing concentrations of rhamnolipid solutions in the range 0–100 mM. The control solution was used to determine whether distilled water alone recovered K-feldspar or sepiolite-bound Cd(II) or not. The centrifuge tubes containing K-feldspar or sepiolite mixed with rhamnolipid solution were placed on a shaker at 25 °C and 150 rpm for 72 h and then centrifuged at 19,000g for 20 min. The supernatant was analyzed for Cd(II) concentration by atomic absorption analysis. The pH of supernatant samples were then adjusted with concentrated HNO_3 to <2 to precipitate the rhamnolipid. To recover the rhamnolipid pellets, each sample was centrifuged and then the supernatant was diluted in 1% HNO_3 for atomic absorption analysis.

3. Results and discussion

3.1. Sorption of rhamnolipid by K-feldspar and sepiolite

To determine the sorption characteristics of the rhamnolipid onto K-feldspar and sepiolite in the absence of

Table 1
Chemical properties of K-feldspar and sepiolite used for sorption/desorption experiments

K-feldspar composition	Content (wt%)	Sepiolite composition	Content (wt%)
Na_2O	2.56	CO_2	21.2
MgO	0.0745	Na_2O	0.0427
Al_2O_3	18.4	MgO	24.6
SiO_2	64.3	Al_2O_3	0.195
SO_3	0.0124	SiO_2	52.1
K_2O	13.5	P_2O_5	0.0149
CaO	0.135	SO_3	0.0291
Cr_2O_3	0.163	K_2O	0.0338
Fe_2O_3	0.784	CaO	1.46
NiO	0.0625	Fe_2O_3	0.153
CuO	0.0184	NiO	0.0866
Rb_2O	0.0220	CuO	0.0153
SrO	0.0127		

Table 2

Amount of rhamnolipid in liquid phase (%) in the absence and presence of 1 mM Cd(II) when K-feldspar or sepiolite was used as sorbent

$C_{\text{Rhamnolipid con. (mM)}}^a$	Sepiolite		K-feldspar	
	No metal	Cd(II)	No metal	Cd(II)
12.5	0	8.5	81.8	85.1
25	0	10.2	90.1	91.8
50	13.5	21.8	86.8	96.8
80	25.1	42.8	89.6	96.9
100	31.8	50.1	95.1	100

^aConcentration of rhamnolipid added to K-feldspar or sepiolite.

Cd(II) ions, the concentrations of the rhamnolipid solutions were varied between 12.5 and 100 mM. Rhamnolipid sorption onto sepiolite was considerably higher than that onto K-feldspar (Table 2). In the absence of Cd(II) ions, at the lower rhamnolipid concentrations tested (12.5–50 mM), rhamnolipid sorption to sepiolite was high (between 100% and 86.5% sorbed). Liquid phase rhamnolipid concentration increased as the amount of rhamnolipid added was increased. At higher rhamnolipid concentrations (80 and 100 mM), sorption was reduced to between 74.9% and 68.2%. Although the general change of rhamnolipid sorption onto K-feldspar with increasing rhamnolipid concentration was the same with sepiolite, rhamnolipid sorption efficiency remained in the range of 18.2–4.9%. To investigate the effect of Cd(II) ions on the rhamnolipid sorption to K-feldspar and sepiolite, the concentrations of rhamnolipid solutions were increased from 12.5 to 100 mM while the initial Cd(II) concentrations in each sorption medium were kept constant at 1 mM. The presence of Cd(II) in the sepiolite prior to addition of the rhamnolipid was observed to affect the sorption of rhamnolipid. The presence of sepiolite-bound Cd(II) resulted in less rhamnolipid sorption to sepiolite. Sorption of rhamnolipid onto sepiolite was decreased from 91.5% to 49.9% in the presence of 1 mM Cd(II) with a larger amount of Cd(II) sorbed (100% or 6.58 mmol Cd(II) kg⁻¹ sepiolite). Due to the weak sorption of rhamnolipid by K-feldspar, aqueous phase rhamnolipid concentrations was increased less in the presence of 1 mM Cd(II) in the solution (or 1.87 mmol Cd(II) kg⁻¹ K-feldspar). Sorption of rhamnolipid onto K-feldspar decreased from 14.9% to 0%. There are two possible mechanisms for rhamnolipid sorption explained in literature. The first mechanism is cation bridging between the anionic polar head group and sorbed cations on soil-component. Hydrophobic interactions between the nonpolar tails and hydrophobic regions in the soil-component organic matter also play important role in rhamnolipid sorption (Herman et al., 1995; Torrens et al., 1998; Ochoa-Loza et al., 2007). However, the mechanism of rhamnolipid sorption and the role of cation bridging in rhamnolipid sorption have not been investigated in detail yet.

3.2. Removal of Cd(II) ions by K-feldspar and sepiolite

3.2.1. Effect of pH

Effect of pH was studied in the pH range 5.0–7.6 with optimized 1.5 g K-feldspar or sepiolite dose, 0.9 mM initial Cd(II) concentration and at temperature 25 °C. Fig. 1 depicts the effect of pH on sorption efficiency of Cd(II) onto K-feldspar. The sorption efficiency for K-feldspar increased from 44.7% to 50.3% with an increase in solution pH from 5 to 6, then remained approximately constant at 51.4%. The sorption efficiency of Cd(II) onto sepiolite seems to be independent of pH with in studied pH range and initial Cd(II) concentration. When sepiolite was treated with 0.9 mM Cd(II) solution, the sorption efficiency was determined as 100% at all pH values tested.

3.2.2. Effect of initial Cd(II) ion concentration

The concentration of Cd(II) was varied from 0.25 to 5.29 mM. The pH was kept as 6.8 while the amount of K-feldspar or sepiolite added was 1.5 g. Although sorbed Cd(II) quantity per unit weight of K-feldspar or sepiolite increased with increasing initial Cd(II) concentration, the sorption efficiencies of Cd(II) for both sorbents decreased (Fig. 2). The relative number of binding sites available to the metal ion would reduce as the concentration of the metal ion was increased. The K-feldspar or sepiolite became nearly saturated at high initial concentrations. The Cd(II) sorption capacity of sepiolite was higher substantially than that of K-feldspar. A maximum Cd(II) uptake of 100% by sepiolite was reached at initial Cd(II) concentrations in the range of 0.25–1.0 mM while K-feldspar exhibited maximum Cd(II) uptake of 63.2% at initial Cd(II) concentration of 0.45 mM.

3.2.3. Sorption isotherms of Cd(II) onto K-feldspar and sepiolite

To investigate the sorption equilibrium of Cd(II) onto K-feldspar and sepiolite, the sorption equilibrium results were fitted to the Langmuir, Freundlich, Redlich–Peterson,

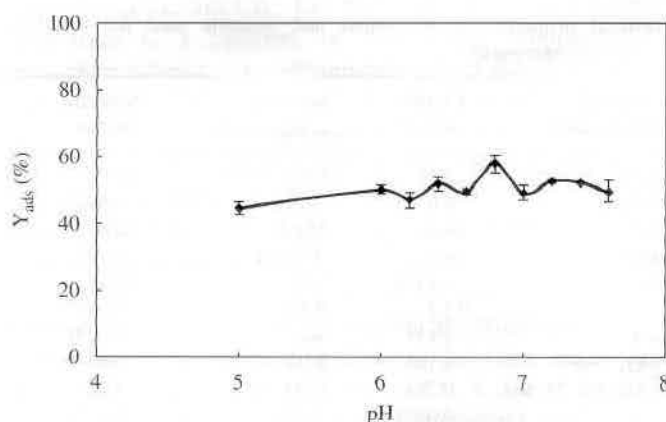


Fig. 1. Effect of pH on Cd(II) sorption efficiencies by K-feldspar ($C_{\text{Cd},i}$, 0.90 mM; temperature, 25 °C; amount of K-feldspar, 1.5 g; stirring rate, 150 rpm).

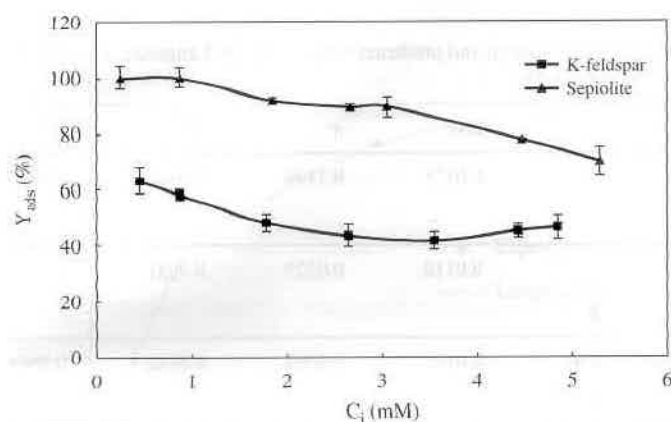


Fig. 2. Effect of initial Cd(II) concentration on Cd(II) sorption efficiencies by K-feldspar and sepiolite (pH, 6.8; temperature, 25 °C; amount of K-feldspar or sepiolite, 1.5 g; stirring rate, 150 rpm).

and Koble–Corrigan sorption models to obtain the characteristic parameters of each model.

The most widely used isotherm equation for modeling of the sorption equilibrium data is the Langmuir equation (Langmuir 1916):

$$q_{eq} = \frac{Q^0 K C_{eq}}{1 + K C_{eq}} \quad (1)$$

where C_{eq} is the metal concentration in solution, Q^0 is the maximum amount of sorbed material required to give a complete monolayer on the surface, and K is a measurement of relative sorption affinity. The Langmuir model assumes that each site accepts only one molecule, that sorbed molecules are organized as a monolayer, that all sites are energetically equivalent and that there is no interaction between sorbed molecules. The Langmuir equation obeys Henry's law at low concentrations.

The Freundlich expression is an empirical equation based on sorption on a heterogeneous surface. The Freundlich equation is commonly presented as (Freundlich, 1907)

$$q_{eq} = K_F C_{eq}^{1/n} \quad (2)$$

The Freundlich constant K_F is an indication of the sorption capacity of the sorbent; $1/n$ indicates the effect of concentration on the sorption capacity and represents the sorption intensity.

A further empirical isotherm has been developed by Redlich and Peterson (RP), incorporating three parameters (Redlich and Peterson, 1959):

$$q_{eq} = \frac{K_R C_{eq}}{1 + a_R C_{eq}^\beta} \quad (3)$$

when $\beta = 1$, the Redlich–Peterson equation reduces to the Langmuir equation. If $a_R C_{eq}^\beta$ is much greater than 1, then it becomes the Freundlich equation. If $a_R C_{eq}^\beta$ is much less than 1, which occurs at low concentrations, then it becomes a linear isotherm equation.

Koble–Corrigan model is another three-parameter empirical model for the representing equilibrium sorption data. It is a combination of the Langmuir and Freundlich isotherm-type models and is given by (Koble and Corrigan, 1952)

$$q_{eq} = \frac{A C_{eq}^b}{1 + B C_{eq}^b} \quad (4)$$

where A , B and b are the Koble–Corrigan parameters, respectively. When $b = 1$, the Koble–Corrigan equation reduces to the Langmuir equation. If $B C_{eq}^b$ is much less than 1, sorption is very low, then it becomes the Freundlich equation. If $B C_{eq}^b$ is much greater than 1, sorption is very high, the sorbed sorbate quantity per unit weight of sorbent at equilibrium remains constant and is defined by the following equation:

$$q_{eq} = \frac{A}{B} \quad (5)$$

Generally, the Langmuir isotherm model is applied to experimental sorption data in order to investigate the maximum monolayer sorption capacity of sorbent while Freundlich isotherm model described better the sorption on heterogeneous surfaces. Redlich–Peterson and Koble–Corrigan isotherm models are applied to improve the fit for a wide range of initial sorbate concentration.

The sorption model constants were estimated from the sorption data of Cd(II) ions on K-feldspar and sepiolite by using a DataFit (trial version) computer program. The Langmuir, Freundlich, Redlich–Peterson and Koble–Corrigan constants and the residual sum of squares (SSE), standard error of the estimate (σ), coefficient of multiple determination (R^2) and adjusted coefficient of multiple determination (R_a^2) between the experimental values and the predicted values using the models are given in Table 3 for K-feldspar and in Table 4 for sepiolite. Langmuir and Freundlich type sorption models are used to model sorption from dilute aqueous solutions. All sorption models fit the sorption equilibrium data of Cd(II) ions onto K-feldspar appearing excellent the coefficients of multiple determination R^2 within the metal concentrations studied (0.45–2.64 mM). However, the Redlich–Peterson model seems to be best represent the measured biosorption data. The goodness of fit of sorption models is also dependent on the number of adjustable parameters taking into account the reduction in degrees of freedom. In fact the adjusted coefficient of multiple determination (R_a^2), which takes into account the number of model parameters is also needed to calculate and compare. The value of R_a^2 for Freundlich model was the highest of the four models examined, in this case, Freundlich model with two adjustable parameters must be preferred over the Redlich–Peterson model with three parameters. The Freundlich sorption isotherm obtained at pH 6.8 for the sorption of Cd(II) ions onto K-feldspar is shown in Fig. 3. In all the figures where metal uptake equilibrium data are shown, the

Table 3

Comparison of the sorption isotherm coefficients and SSE, σ , R^2 and R_a^2 values between experimental and predicted values using by Langmuir, Freundlich, Redlich–Peterson and Koble–Corrigan models for K-feldspar

Langmuir model	Q^0 (mmol kg ⁻¹)	K (L mmol ⁻¹)		SSE	σ	R^2	R_a^2
	6.66	0.86		0.0575	0.1199	0.9900	0.9875
Freundlich model	K_F (mmol ⁿ kg ⁻¹ L ⁿ)	n					
	2.99	1.63		0.0110	0.0525	0.9981	0.9976
Redlich–Peterson model	K_R (L kg ⁻¹)	a_R (L ^{β} mmol ^{-β})	β				
	32.71	9.90	0.43	0.0105	0.0593	0.9982	0.9969
Koble–Corrigan model	A (mmol ^{1-b} kg ⁻¹ L ^{b})	B (L ^{b} mmol ^{-b})	b				
	3.14	0.05	0.64	0.0108	0.0600	0.9981	0.9969

Table 4

Comparison of the sorption isotherm coefficients and SSE, σ , R^2 and R_a^2 values between experimental and predicted values using by Langmuir, Freundlich, Redlich–Peterson and Koble–Corrigan models for sepiolite

Langmuir model	Q^0 (mmol kg ⁻¹)	K (L mmol ⁻¹)		SSE	σ	R^2	R_a^2
	28.17	4.64		0.4012	0.3657	0.9966	0.9955
Freundlich model	K_F (mmol ⁿ kg ⁻¹ L ⁿ)	n					
	22.41	3.48		6.6276	1.4863	0.9439	0.9252
Redlich–Peterson model	K_R (L kg ⁻¹)	a_R (L ^{β} mmol ^{-β})	β				
	112.72	3.82	1.07	0.1819	0.3016	0.9985	0.9969
Koble–Corrigan model	A (mmol ^{1-b} kg ⁻¹ L ^{b})	B (L ^{b} mmol ^{-b})	b				
	177.07	6.61	1.17	0.1603	0.2831	0.9986	0.9973

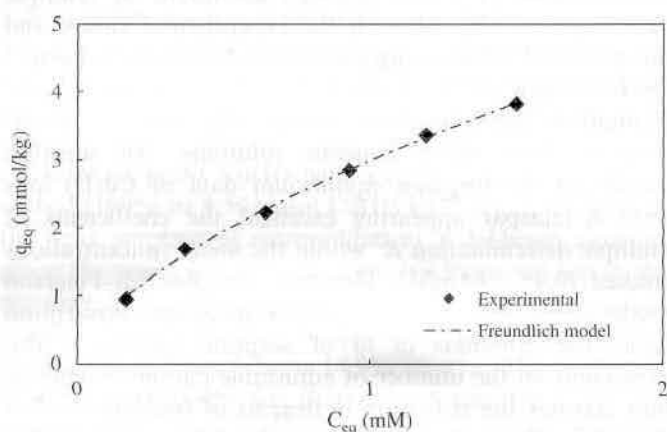


Fig. 3. Freundlich sorption isotherm for Cd(II) sorption onto K-feldspar (pH: 6.8; temperature, 25 °C; amount of K-feldspar, 1.5 g; stirring rate, 150 rpm).

model profiles are presented as dashed lines and the experimentally obtained values are given by symbols.

On the other hand, of the four models examined, the Koble–Corrigan and Redlich–Peterson models showed the best fit for the Cd(II) sorption data onto sepiolite, whereas

the Freundlich-type model failed to describe the sorption results. However, the Redlich–Peterson constant β and the Koble–Corrigan constant n approach 1. This means that the sorption isotherms reduced the Langmuir model within the metal concentrations studied (1.85–5.29 mM) (Fig. 4). If the adjusted coefficients of multiple determination of all models examined are also compared, it is seen that these coefficients in case of sepiolite did not show a significant change for two and three-parameters models, and order of fit did not change.

Sorption isotherms can also be used to compare sorptive capacity of soil components for Cd(II) ions. The Langmuir constant, Q^0 , represents the sorption capacity when the surface is saturated with metal ions. It is roughly an indicator of the ultimate sorption capacity. The Langmuir constant, K is the ratio of adsorption/desorption rates and reflects the affinity between the sorbents and Cd(II) ions. A large value of K implies strong and irreversible bonding of metal ions. The highest K and Q^0 values were obtained for the sorption of Cd(II) ions onto sepiolite. As the values of Freundlich sorption constant, n , for K-feldspar and sepiolite are found to be higher than 1, these values indicate favorable sorption. Substantially

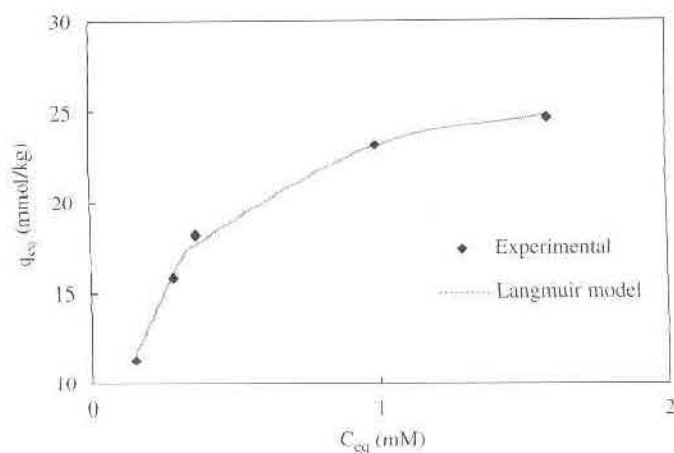


Fig. 4. Langmuir sorption isotherm for Cd(II) sorption onto sepiolite (pH, 6.8; temperature, 25 °C; amount of sepiolite, 1.5 g; stirring rate, 150 rpm).

higher K_F and n values obtained for the sorption of Cd(II) ions onto sepiolite verify higher Cd(II) sorption capacity and intensity of sepiolite versus K-feldspar. The values of Redlich–Peterson and Koble–Corrigan constants for sepiolite were also considerably higher than those for K-feldspar, and were a sign of high Cd(II) sorption capacity, affinity, and intensity.

Because of the crystal-chemical features, heavy-metal uptake by the sepiolite occur by adsorption and/or cation exchange reactions. Adsorption takes places on the oxygen ions of the tetrahedral sheets, on the water molecules at the edges of the octahedral sheet and on Si–OH groups along the direction of fibers. The adsorption of H^+ and OH^- at the oxide–water interface is related to the protolysis of surface OH groups $\blacksquare SOH$. Bivalent metal ions such as Cd(II) are sorbed on surface $\blacksquare SOH$ sites on the broken edges of the sepiolite with simultaneous release of protons. The next step of this reaction is the incorporation of sorbed cation to octahedral sheet. Ion exchange occurs by substituting the solvated cations inside the channels and/or inside the octahedra at the edges of the channels. In the process of ion exchange, bivalent metal ions like Cd(II) replace Na(I) and/or Mg(II) in the edges of octahedral sheet (Brigatti et al., 2000; Kara et al., 2003; Vico, 2003). On the other hand, studies on heavy metal sorption mechanisms with feldspars are very scarce in literature. Cd(II) was reported to only be physisorbed by perthitic feldspar and outer-sphere complexation played an important role in the Cd(II) removal mechanisms of perthitic feldspar (Farquhar et al., 1997).

3.3. Recovery of Cd(II) ions from K-feldspar and sepiolite using rhamnolipid biosurfactant

3.3.1. Effect of pH

The biosurfactant-containing solution was adjusted to various pHs in the range 5.0–7.6 at 25 °C and allowed to stand for 72 h, following which Cd(II) concentrations were

measured. This pH range was chosen for several reasons. In aqueous solution, surfactants reduce surface tension by accumulating at interfaces and facilitating the formation of emulsions between liquids of different polarities. The surface tension of rhamnolipid solutions is quite sensitive to pH. The surface activity of the rhamnolipid is highest between pH 7.0 and 7.5. As the pH is increased above 7.5, there is a slight decrease in surface activity that results in an increase in surface tension. As the pH is decreased from 7.0 to 5.0, surface activity decreases significantly, resulting in a considerable increase in surface tension (Zhang and Miller, 1992). Rhamnolipid precipitation is observed when the pH is decreased below 5.0. At low concentration, surfactants are present as individual molecules. However, as the concentration of the surfactant is increased, a concentration is reached where no further change in interfacial properties takes place. The amount of surfactant needed to achieve the lowest possible surface tension is defined as the CMC. At the CMC, surfactant molecules aggregate to form structures such as bilayers, vesicles, or micelles. The type and size of aggregate formed depends on the surfactant structure and on the solution pH. At a low pH (<6.0), rhamnolipids form liposome-like vesicles. Between pH 6.0 and 6.6, rhamnolipids form either lamella-like structures or lipid aggregates. The rhamnosyl moiety is negatively charged above pH 6.8 and micelles form. Micelles are the smallest basic structure formed, generally less than 5 nm in diameter, and main effective form for metal immobilization. Vesicle structures are next in size then micelles and range from 10 nm to more than 500 nm in diameter (Miller, 1995). In these technology, the size of the metal–ligand complex is critical to prevent filtration of metal complexes by small soil pores. The best recovery efficiency, 86.1% of the sorbed Cd(II) ($1.56 \text{ mmol kg}^{-1}$) by K-feldspar and 10.12% of the sorbed Cd(II) ($7.71 \text{ mmol kg}^{-1}$) by sepiolite, was achieved by adjusting the medium pH value to 6.8 (Fig. 5).

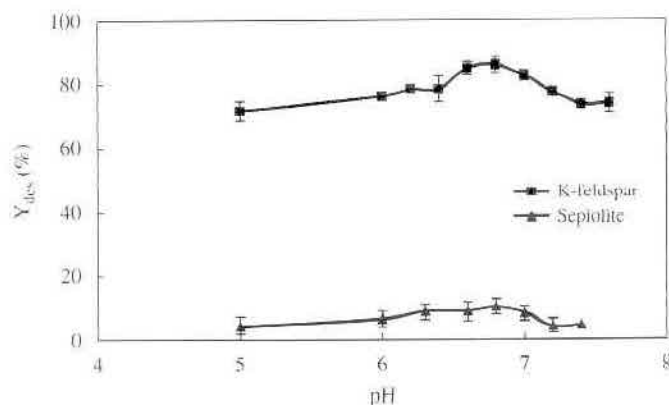


Fig. 5. Effect of pH on Cd(II) desorption efficiencies from K-feldspar and sepiolite using rhamnolipid biosurfactant ($C_{\text{Rhamnolipid conc.}}$, 25 mM for K-feldspar and 50 mM for sepiolite; $q_{\text{K-feldspar}}$, $1.56 \text{ mmol kg}^{-1}$; $q_{\text{sepiolite}}$, $7.71 \text{ mmol kg}^{-1}$; temperature, 25 °C; amount of sorbent, 1.5 g; stirring rate, 150 rpm).

3.3.2. Effect of rhamnolipid concentration

The effect of rhamnolipid concentration for Cd(II) desorption from K-feldspar and sepiolite, at pH 6.8 is shown in Fig. 6. In the case of desorption from K-feldspar, the recovery efficiency was approximately 96.0% of the sorbed Cd(II) ($1.87 \text{ mmol kg}^{-1}$) for the rhamnolipid concentrations in the range 50–80 mM. At rhamnolipid concentrations lower than 50 mM, Cd(II) desorption efficiency slightly decreased, weak rhamnolipid sorption to K-feldspar appeared to a little limit the ability of rhamnolipid to mobilize K-feldspar-bound Cd(II). To prevent excess rhamnolipid treatment, the rhamnolipid concentration should be kept at a minimum-optimum level of 50 mM. On the other hand, 50 mM rhamnolipid recovered only 7.81% of the sorbed Cd(II) ($6.58 \text{ mmol kg}^{-1}$) from sepiolite, and increasing concentrations of rhamnolipid upto 50 mM decreased the Cd(II) desorption efficiency from 7.81% to 4.54%. There may be two reasons of this behavior. Firstly Cd(II) is sorbed more strongly and slowly reversible to sepiolite than K-feldspar and form more stable complexes with ligands of sepiolite. Secondly rhamnolipids are also strongly sorbed to sepiolite. To mobilize sepiolite-bound metals, rhamnolipid sorption must be minimized.

3.3.3. Effect of amount of Cd(II) loaded to K-feldspar and sepiolite

After treatment of the K-feldspar to Cd(II) solutions in the range of concentrations varying from 0.45 to 5.32 mM, the amounts of desorbed Cd(II) by 25 mM rhamnolipid solution were determined at pH 6.8. Although the Cd(II) recovery efficiency was the highest (88.8%) at the minimum sorbed Cd(II) quantity per unit weight of K-feldspar ($0.95 \text{ mmol kg}^{-1}$), it decreased with increasing amount of Cd(II) loaded to K-feldspar (Fig. 7). To investigate the effect of amount of Cd(II) loaded to sepiolite on the Cd(II) desorption efficiency, the sepiolite samples were contacted with Cd(II) solutions at the concentrations varying from 0.25 to 5.29 mM. A slight increase in the Cd(II) desorption

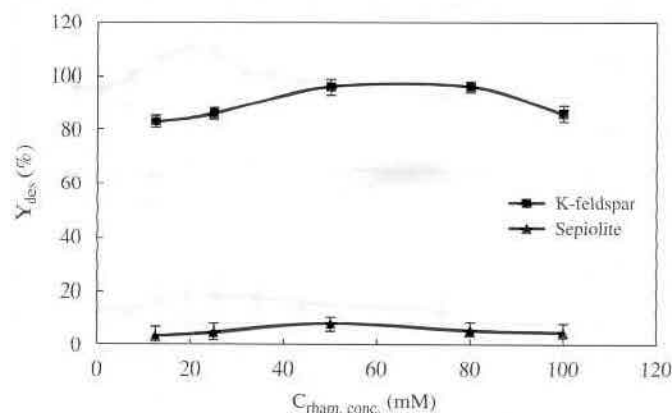


Fig. 6. Effect of rhamnolipid concentration on Cd(II) desorption efficiencies from K-feldspar and sepiolite (pH, 6.8; $q_{\text{K-feldspar}}$, $1.87 \text{ mmol kg}^{-1}$; $q_{\text{sepiolite}}$, $6.58 \text{ mmol kg}^{-1}$; temperature, 25°C ; amount of sorbent, 1.5 g; stirring rate, 150 rpm).

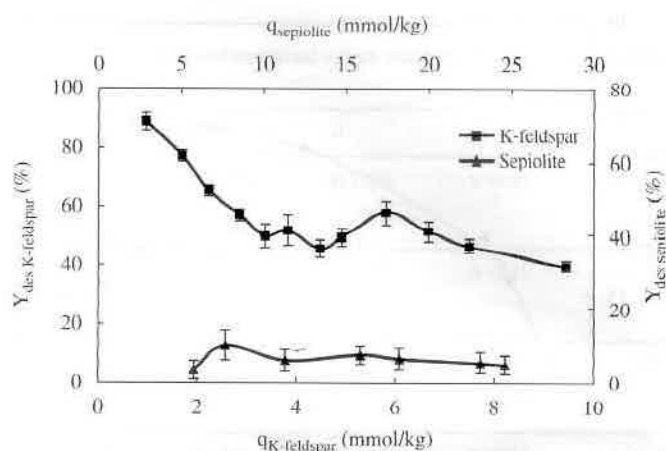


Fig. 7. Effect of amount of Cd(II) loaded to K-feldspar and sepiolite on Cd(II) desorption efficiencies (pH, 6.8; $C_{\text{Rhamnolipid conc.}}$, 25 mM for K-feldspar and 50 mM for sepiolite; temperature, 25°C ; amount of sorbent, 1.5 g; stirring rate, 150 rpm).

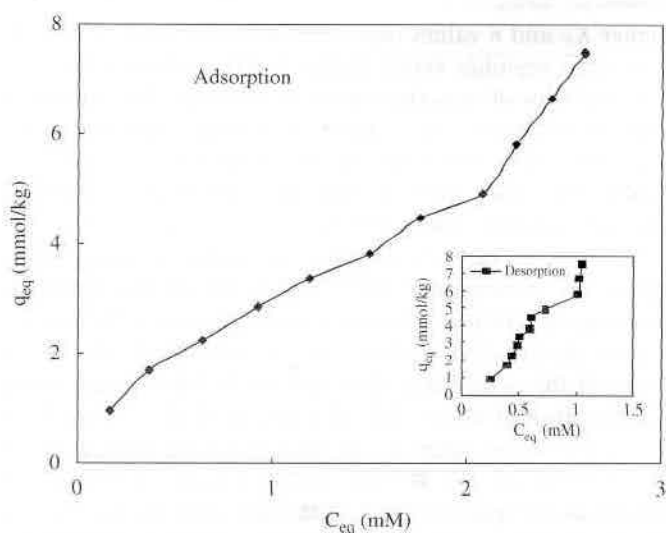


Fig. 8. Sorption/desorption isotherms of Cd(II) onto/from K-feldspar (pH, 6.8; $C_{\text{Rhamnolipid conc.}}$, 25 mM; temperature, 25°C ; amount of K-feldspar, 1.5 g; stirring rate, 150 rpm).

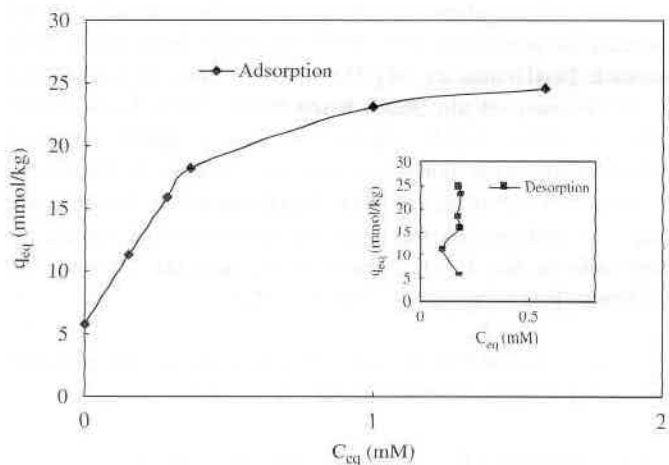


Fig. 9. Sorption/desorption isotherms of Cd(II) onto/from sepiolite (pH, 6.8; $C_{\text{Rhamnolipid conc.}}$, 50 mM; temperature, 25°C ; amount of sepiolite, 1.5 g; stirring rate, 150 rpm).

efficiency was observed with increasing amount of Cd(II) loaded to sepiolite upto $7.71 \text{ mmol kg}^{-1}$, and 10.12% of loaded Cd(II) was recovered by the 50 mM rhamnolipid solution (Fig. 7). Cd(II) desorption efficiencies from K-feldspar using the rhamnolipid biosurfactant were considerably higher than those from sepiolite. Although the rhamnolipid biosurfactant was recovered Cd(II) ions from K-feldspar successfully, it remained ineffective in the case of sepiolite, a superior Cd(II) sorbent.

The resulting sorption–desorption isotherms show that the desorption of Cd(II) from K-feldspar was more extensive than that from sepiolite (Figs. 8 and 9). The amounts of Cd(II) desorbed from sepiolite were very small. The limited desorption of Cd(II) from sepiolite suggests that the sorption reactions taking place are only slowly reversible or a high activation energy is required for desorption.

4. Conclusions

In this study, the sorption of Cd(II) by K-feldspar and sepiolite as representative soil components was studied as a function of pH and initial metal ion concentrations. Although both the soil components sorbed the substantial amounts of Cd(II), sorption capacity of sepiolite was considerably higher than that of K-feldspar. For K-feldspar a maximum Cd(II) uptake was measured as 7.49 mmol/kg . On the other hand, sepiolite reached a maximum uptake value of $24.66 \text{ mmol Cd(II)/kg}$. Sepiolite, due to its structure and morphology, has a great number of terminal silica tetrahedrons present in its external surfaces and consistently also a great number of silanol groups, which are effective in the uptake of bivalent metal cations. The Freundlich model represented the sorption process of K-feldspar better than the Langmuir-type ones. On the other hand, sorption data of sepiolite were best modeled by the Langmuir model.

To determine the rhamnolipid availability for metal recovery, rhamnolipid concentrations in K-feldspar and sepiolite were determined in the absence and presence of Cd(II). Rhamnolipid was shown to be strongly sorbed to sepiolite where as the rhamnolipid sorption capacity of K-feldspar was very low. The presence of Cd(II) in the sepiolite prior to addition of the rhamnolipid decreased the sorption of rhamnolipid from 86.5% to 78.2% of the rhamnolipid added. Because of the poor rhamnolipid sorption capacity of K-feldspar, the presence of Cd(II) in the K-feldspar had a little effect on the liquid phase rhamnolipid concentration. This study demonstrated that rhamnolipid washing technology can be used successfully in the bioremediation of Cd(II)-contaminated soil containing K-feldspar as main component. The best recovery efficiency from K-feldspar, approximately 96% of the sorbed Cd(II) ($1.87 \text{ mmol kg}^{-1}$) was achieved by adjusting the pH to 6.8 and using the rhamnolipid concentrations in the range 50–80 mM. On the other hand, the Cd(II) recovery efficiency remained 10.12% of the sorbed Cd(II)

($7.71 \text{ mmol kg}^{-1}$) by the 50 mM rhamnolipid solution. Another interesting result obtained in this study total Cd(II) and rhamnolipid sorption capacity of sepiolite was considerably higher than that of K-feldspar. There are many binding sites on sepiolite to bind Cd(II) and rhamnolipid partially separate.

Acknowledgments

The authors wish to thank Eskişehir Osmangazi University Scientific Research Projects Commission for the partial financial support of this study (Project no. 200315055). We also wish to thank Jeneil Biosurfactant Co. (Saukville, WI) for supplying the rhamnolipid biosurfactant.

References

- Alumaa, P., Steinnes, E., Kirso, U., Petersell, V., 2001. Heavy metal sorption by different Estonian soil types at low equilibrium solution concentrations. *Proceedings of the Estonian Academy of Science—Chemistry* 50, 104–115.
- Aşçı, Y., Nurbas, M., Sağ Açık, 2007. Sorption of Cd(II) onto kaolin as a soil component and desorption of Cd(II) from kaolin using rhamnolipid biosurfactant. *Journal of Hazardous Materials B* 139, 50–56.
- Brigatti, M.F., Lugli, C., Poppi, L., 2000. Kinetics of heavy-metal removal and recovery in sepiolite. *Applied Clay Science* 16, 45–57.
- Bruins, M.R., Kapil, S., Oehme, F.W., 2000. Microbial resistance to metals in the environment. *Ecotoxicology and Environmental Safety, Environmental Research Section B* 45, 198–207.
- Christofi, N., Ivshina, I.B., 2002. Microbial surfactants and their use in field studies of soil remediation: a review. *Journal of Applied Microbiology* 93, 915–929.
- Demir, C., Benti, I., Gölgönül, I., Çelik, M.S., 2003. Effects of bivalent salts on the flotation separation of Na-feldspar from K-feldspar. *Mining Engineering* 16, 551–554.
- Farquhar, M.L., Vaughan, D.J., Hughes, C.R., Charnock, J.M., England, K.E.R., 1997. Experimental studies of the interaction of aqueous metal cations with mineral substrates: lead, cadmium, and copper with perthitic feldspar, muscovite, and biotite. *Geochimica et Cosmochimica Acta* 61, 3051–3064.
- Frazer, L., 2000. Innovations. Lipid lather removes metals. *Environmental Health Perspectives* 108, A320–A323.
- Freundlich, H., 1907. Ueber die adsorption in Loesungen. *Zeitschrift für Physikalische Chemie* 57A, 385–470.
- Garcia-Sanchez, A., Alastuey, A., Querol, X., 1999. Heavy metal adsorption by different minerals: application to the remediation of polluted soils. *Science of the Total Environment* 242, 179–188.
- Herman, D.C., Artioli, J.F., Miller, R.M., 1995. Removal of cadmium, lead, and zinc from soil by a rhamnolipid biosurfactant. *Environmental Science and Technology* 29, 2280–2285.
- Jeneil Biosurfactant Co., LLC, 2001. Material safety data sheet for JBR 425. <<http://www.biosurfactant.com/downloads/jbr425msds.pdf>>.
- Kara, M., Yuzer, H., Sabah, E., Celik, M.S., 2003. Adsorption of cobalt from aqueous solutions onto sepiolite. *Water Research* 37, 224–232.
- Kim, S.H., Lim, E.J., Lee, S.O., Lee, J.D., Lee, T.H., 2000. Purification and characterization of biosurfactants from *Nocardia* sp. L-417. *Biotechnology and Applied Biochemistry* 31, 249–253.
- Koble, R.A., Corrigan, T.E., 1952. Adsorption isotherms for pure hydrocarbons. *Industrial and Engineering Chemistry* 44, 383–387.
- Kocaoba, S., Akyuz, T., 2005. Effects of conditioning of sepiolite prior to cobalt and nickel removal. *Desalination* 181, 313–318.

- Langmuir, L., 1916. The constitution and fundamental properties of solids and liquids. *Journal of the American Chemical Society* 38, 2221–2295.
- Lebeau, T., Bagot, D., Jézéquel, K., Fabre, B., 2002. Cadmium biosorption by free and immobilised microorganisms cultivated in a liquid soil extract medium: effects of Cd, pH and techniques of culture. *Science of the Total Environment* 291, 73–83.
- Lin, C.-C., Lin, H.-L., 2005. Remediation of soil-contaminated with the heavy metal. *Journal of Hazardous Materials A* 122, 7–15.
- Markiewicz-Patkowska, J., Hursthouse, A., Przybyla-Kij, H., 2005. The interaction of heavy metals with urban soils: sorption behaviour of Cd, Cu, Cr, Pb and Zn with a typical mixed brownfield deposit. *Environment International* 31, 513–521.
- Miller, R.M., 1995. Biosurfactant-facilitated remediation of metal-contaminated soils. *Environmental Health Perspectives* 103, 59–62.
- Mulligan, C.N., Wang, S., 2006. Remediation of a heavy-metal contaminated soil by a rhamnolipid foam. *Engineering Geology* 85, 75–81.
- Mulligan, C.N., Yong, R.N., Gibbs, B.F., 2001a. Remediation technologies for metal-contaminated soils and groundwater: an evaluation. *Engineering Geology* 60, 193–207.
- Mulligan, C.N., Yong, R.N., Gibbs, B.F., 2001b. Surfactant-enhanced remediation of contaminated soil: a review. *Engineering Geology* 60, 371–380.
- Ochoa-Loza, F.J., Noordman, W.H., Janssen, D.B., Brusseau, M.L., Maier, R.M., 2007. Effect of clays, metal oxides, and organic matter on rhamnolipid biosurfactant sorption by soil. *Chemosphere* 66, 1634–1642.
- Pu, X., Cutright, T.J., 2006. Sorption-desorption behavior of PCP on soil organic matter and clay minerals. *Chemosphere* 64, 972–983.
- Redlich, O., Peterson, D.L., 1959. A useful adsorption isotherm. *Journal Physical Chemistry* 63, 1024–1029.
- Sabah, E., Turan, M., Çelik, M.S., 2002. Adsorption mechanism of cationic surfactants onto acid- and heat-activated sepiolites. *Water Research* 36, 3957–3964.
- Sağ, Y., Tatar, B., Kutsal, T., 2003a. Biosorption of Pb(II) and Cu(II) by activated sludge in batch and continuous-flow stirred reactors. *Bioresearch Technology* 87, 27–33.
- Sağ, Y., Akçael, B., Kutsal, T., 2003b. Application of multi-component adsorption models to the biosorption of Cr(VI), Cu(II) and Cd(II) ions on *Rhizopus arrhizus* from ternary metal mixtures. *Chemical Engineering Communication* 190, 797–812.
- Sandrin, T.R., Chech, A.M., Maier, R.M., 2000. A rhamnolipid biosurfactant reduces cadmium toxicity during naphthalene biodegradation. *Applied Environmental Microbiology* 66, 4585–4588.
- Shawabkeh, R.A., 2005. Solidification and stabilization of cadmium ions in sand-cement-clay mixture. *Journal of Hazardous Materials B* 125, 237–243.
- Tan, H., Champion, J.T., Artiola, J.F., Brusseau, M.L., Miller, R.M., 1994. Complexation of cadmium by a rhamnolipid biosurfactant. *Environmental Science and Technology* 28, 2402–2406.
- Torrens, J.L., Herman, D.C., Miller-Maier, R.M., 1998. Biosurfactant (rhamnolipid) sorption and the impact on rhamnolipid-facilitated removal of cadmium from various soils under saturated flow conditions. *Environmental Science and Technology* 32, 776–781.
- Vico, L.I., 2003. Acid-base behaviour and Cu^{2+} and Zn^{2+} complexation properties of the sepiolite/water interface. *Chemical Geology* 198, 213–222.
- Vig, K., Megharaj, M., Sethunathan, N., Naidu, R., 2003. Bioavailability and toxicity of cadmium to microorganisms and their activities in soil: a review. *Advances in Environmental Research* 8, 121–135.
- Volesky, B., 1990. *Biosorption of Heavy Metals*. CRC Press, Boca Raton, pp. 13–14.
- Yılmaz, E.I., 2003. Metal tolerance and biosorption capacity of *Bacillus circulans* strain EB1. *Research in Microbiology* 154, 409–415.
- Zhang, Y., Miller, R.M., 1992. Enhanced octadecane dispersion and biodegradation by a *Pseudomonas* rhamnolipid surfactant (biosurfactant). *Applied Environmental Microbiology* 58, 3276–3282.
- Zouboulis, A.I., Matis, K.A., Lazaridis, N.K., Golyshin, P.N., 2003. The use of biosurfactants in flotation: application for the removal of metal ions. *Mining Engineering* 16, 1231–1236.
- Zouboulis, A.I., Loukidou, M.X., Matis, K.A., 2004. Biosorption of toxic metals from aqueous solutions by bacteria strains isolated from metal-polluted soils. *Process Biochemistry* 39, 909–916.

Effects of soil data resolution on SWAT model stream flow and water quality predictions

Mengistu Geza^{*,1}, John E. McCray²

Environmental Science and Engineering Division, Colorado School of Mines, 1500 Illinois St., Golden CO 80401, USA

Received 8 July 2006; received in revised form 4 February 2007; accepted 10 March 2007

Available online 1 May 2007

Abstract

The prediction accuracy of agricultural nonpoint source pollution models such as Soil and Water Assessment Tool (SWAT) depends on how well model input spatial parameters describe the characteristics of the watershed. The objective of this study was to assess the effects of different soil data resolutions on stream flow, sediment and nutrient predictions when used as input for SWAT. SWAT model predictions were compared for the two US Department of Agriculture soil databases with different resolution, namely the State Soil Geographic database (STATSGO) and the Soil Survey Geographic database (SSURGO). Same number of sub-basins was used in the watershed delineation. However, the number of HRUs generated when STATSGO and SSURGO soil data were used is 261 and 1301, respectively. SSURGO, with the highest spatial resolution, has 51 unique soil types in the watershed distributed in 1301 HRUs, while STATSGO has only three distributed in 261 HRUs. As a result of low resolution STATSGO assigns a single classification to areas that may have different soil types if SSURGO were used. SSURGO included Hydrologic Response Units (HRUs) with soil types that were generalized to one soil group in STATSGO. The difference in the number and size of HRUs also has an effect on sediment yield parameters (slope and slope length). Thus, as a result of the discrepancies in soil type and size of HRUs stream flow predicted was higher when SSURGO was used compared to STATSGO. SSURGO predicted less stream loading than STATSGO in terms of sediment and sediment-attached nutrients components, and vice versa for dissolved nutrients. When compared to mean daily measured flow, STATSGO performed better relative to SSURGO before calibration. SSURGO provided better results after calibration as evaluated by R^2 value (0.74 compared to 0.61 for STATSGO) and the Nash-Sutcliffe coefficient of Efficiency (NSE) values (0.70 and 0.61 for SSURGO and STATSGO, respectively) although both are in the same satisfactory range. Modelers need to weigh the benefits before selecting the type of data resolution they are going to use depending on the watershed size and level of accuracy required because more effort is required to prepare and calibrate the model when a fine resolution soil data is used.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: SWAT model; STATSGO; SSURGO; Water quality; Stream flow; Sediment; Nutrients

1. Introduction

Soil physicochemical information is one of the crucial inputs needed to assess the impacts of existing and alternative agricultural management practices on water quality. Assuming that there are no significant impacts from other factors that control soil hydrodynamics such as

crusting in sandy soils and micro-aggregation in clay soils, sandy soils allow a high rate of water infiltration and produce less runoff while soils consisting of poorly drained clay soils allow a low infiltration and produce more runoff (Haan et al., 1994). Soil properties also influence sediment and nutrient loading to streams. Thus, it is important to evaluate the effects of the spatial scale at which the soil database is developed. In the United States, State Soil Geographic database (STATSGO) and Soil Survey Geographic database (SSURGO) are the most commonly available soil databases. Because soil data sets may yield different results in water-quality predictions, it is important to consider this important issue when developing Total

^{*}Corresponding author. Tel.: +1 720 982 5359; fax: +1 303 273 3413.

E-mail addresses: mgezanis@mines.edu (M. Geza), jmcrcay@mines.edu (J.E. McCray).

¹Post doctoral researcher.

²Associate Professor.

Maximum Daily Loads (TMDLs) for impaired watersheds where conflicting interests of stakeholders may opt for different soil data sets. STATSGO and SSURGO developed by the Natural Resources Conservation Agency (NRCS) are digital soil databases that are used in most water-quality simulation models to derive soil input data. The STATSGO maps are compiled by the NRCS through generalizing more detailed soil survey maps, often county level soil maps that are at the scale of SSURGO (USDA-NRCS, 1995). Where SSURGO-level maps are not available, data on geology, topography, vegetation, climate, and Landsat images of the area are used to generate STATSGO maps.

The mapping scale for STATSGO is 1:250,000 (USDA, 1994). The base map used is the US Geological Survey 1:250,000 topographic quadrangles. The number of soil polygons per quadrangle map is between 100 and 400. SSURGO is the most detailed level of soil mapping. Maps are made at scales ranging from 1:15,840 to 1:31,680 (Kroner and Cozzie, 1999). The smallest soil map unit represented in STATSGO is about 6.25 km² (625 ha), whereas it is about 0.02 km² (2 ha) in the SSURGO database. In this study, the two spatially different soil databases (STATSGO and SSURGO) have been used to determine the effect of scale on stream flow and water quality.

STATSGO soil data is available throughout the United States. However, not all areas have SSURGO data in digital format. NRCS expects completion of the SSURGO data digitizing by 2008 (USDA, 2006). For this reason, very few studies have considered the effect of spatial scale on water quality when soil input is derived at different spatial resolution. Besides, not all hydrology and water-quality models are compatible with SSURGO data. For instance, the SSURGO in its standard format is incompatible for use within Soil and Water Assessment Tool (SWAT) especially with ArcView SWAT (AVSWAT) version 2000 (Peschel et al., 2003). A pre-processor extension for the AVSWAT model (SSURGO SWAT 2.0) was developed by Peschel et al. (2003). SSURGO database is included in the current version of SWAT (AVSWATX 2003) although digital SSURGO data is limited especially in western United States. In addition, because there may be thousands of SSURGO units in a watershed, using SSURGO data that is not digitized is not usually practical for watershed modeling studies.

Soil surveys use map unit boundaries across which observed differences are significant and within which the soil is relatively homogeneous. However, there is heterogeneity within soil map units. Thus, map unit in all soil surveys includes other soil components that are not identified in the map unit either because they are too small to be delineated separately at a given soil survey scale or deliberately included in delineations of another map unit to avoid excessive detail in the map (Soil Survey Division Staff, 1993). These inclusions reduce the homogeneity or purity of map units and often affect interpretation or

modeling. Lin et al. (2005) assessed soil spatial variability at multiple scales. They demonstrated that soil spatial variability is a function of map scale, spatial location, and specific soil property. They used a measure of spatial of soil variability within a map unit termed map unit purity. In general, area weighted map unit purity decreased with decreasing map scale.

Modeling studies have been conducted to evaluate the effect of soil data resolutions on stream flow and water-quality predictions. Chaplot (2005) evaluated the impact of DEM mesh size and soil map scale on SWAT runoff, sediment, and NO₃ predictions. The study indicated that there might be no significant increase in the accuracy of models, as a result of more precise topographic or soil information, which increases the input data collection and preparation. The study used different map scales (1/25,000; 1/250,000, and 1/500,000 scale) within the SWAT to simulate runoff, sediment, and NO₃ load. For runoff, few differences existed between the map scales. In contrast, the scale of the soil input maps greatly affected nitrogen as well as sediment loads. The 1/250,000 and 1/500,000 maps significantly decreased the mean monthly N load. For all DEM mesh sizes, the finest soil information enhanced the prediction quality for runoff, nitrogen and sediment loads.

Another study on DEM grid resolution by Zhang and Montgomery (1994) showed that DEM grid size significantly affects topographic parameters. The finest resolution DEM (10 m) provided better results than a 30 m and 90 m data. Brown et al. (1993) investigated response of Areal Non-Point Source Watershed Environment Response Simulation (ANSWERS) to variations in soil, terrain, and land cover data aggregation levels. Soils, land cover, slope angle data were generalized by assigning a class value occupying the majority of the area within an aggregation unit to all cells within the unit. The study concluded that the magnitude of change in the model outputs due to aggregation depends on the spatial dependence of the input variables. The degree of spatial dependence within the input variables had an influence in the results.

Levick et al. (2004) compared Food and Agricultural Organization (FAO) soils with STATSGO and SSURGO at the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) Walnut Gulch Experimental Watershed in Arizona using the Kinematic Runoff and Erosion Model (KINEROS2). The study showed that runoff using STATSGO soils were generally higher than with SSURGO. In most cases, FAO soils data produced less runoff than STATSGO soils, although it produced more runoff than SSURGO soils about half the time. The differences in runoff and soil properties were attributed to the difference in data resolution.

Peschel et al. (2003) developed and applied SSURGO SWAT 2.0 extension to Leon Creek Watershed in Texas to compare the effect of soil data resolution on stream flow. In this study, unlike the finding by Levick et al. (2004) total flow obtained using SSURGO soil data was higher than flow obtained using STATSGO data. It was determined

that transmission losses were lower when SSURGO was used. The study did not draw any firm conclusions, however, the increased surface flow was attributed to higher resolution of SSURGO because other inputs were the same.

Wang and Melesse (2006) assessed the effects of using STATSGO versus SSURGO as an input for the SWAT model simulation of the stream flows in the upper 45% of the Elm River watershed, located in eastern North Dakota. The results indicated that SSURGO provided an overall better prediction of the stream discharges than STATSGO; although both did a good and comparable job on predicting the high stream flows. However, STATSGO predicted the low stream flows more accurately and had a slightly better performance during the validation period.

Prasanna and Mulla (2005) compared flow, sediment, nitrate and phosphorus losses for High Island Creek, an agricultural watershed located in south-central Minnesota using a field scale water table management model (ADAPT). Statistical comparison of calibration results with measured data indicated excellent agreement for both soil databases. However, evaluation of alternative management practices indicated that STATSGO based predicted annual nitrate losses are consistently higher than that for SSURGO data, and vice versa for predicted phosphorus losses.

These few studies on the effect of soil data resolution on stream flow and water quality have yielded different results depending on soil type and climate. Most have acknowledged the importance of data input resolution on model predictions and the dependency of the results on specific soil and climate conditions. Thus, it is important to look at the effect of the soil data resolution on many different watersheds of a variety of hydrologic and soil conditions in hopes that generalities about the relative appropriateness of using the two databases can be assessed. For this study, we have selected a mountainous watershed in Jefferson County, Colorado with steep topography and with low infiltration soils accounting for more than 50% of the total drainage area and semi-arid climate. In addition to soil-related parameters, soil data resolution, when used in watershed models, affects other physical parameters. The fact that the high resolution SSURGO has more HRUs in a sub-basin will affect slope and slope length computations and sediment yield, the effect of which will be more prevalent in mountainous watersheds compared to flat topography. To our knowledge, the soil database comparison has not been completed in a mountainous watershed and we believe the paper can contribute to the general understanding of advantages and disadvantages of using soils with different resolution in modeling. In addition to topography and soil type discussed above, the effect of soil data resolution on SWAT model prediction will vary depending on the climate. SWAT uses the SCS curve number method for predicting storm runoff volume (Arnold et al., 1998). Watershed wetness prior to the storm

event of interest (antecedent moisture condition, AMC) is accounted in the curve number method. AMC depends on initial soil water content, which is affected by the watershed climate including amount of precipitation and evaporation. Thus, the results from different soil data resolution will vary depending on the climate of the region whether it is arid, semi-arid or humid. The study does not intend to solve computational issues. It only attempts to evaluate the sensitivity of the SWAT model to soil data resolution as it applies to a mountainous watershed in a semi-arid region. Thus, the objective of this study was to assess the effects of different soil data resolution using STATSGO and SSURGO when used as input for the SWAT model on stream flow, sediment and nutrients predictions in the Turkey Creek watershed.

2. Methodology

2.1. SWAT model

Several water-quality models are available for watershed analysis. Examples are Hydrological Simulation Program—FORTRAN (HSPF), SWAT, and Watershed Analysis Risk Management Framework (WARMF). In WARMF, soils information cannot be directly input from GIS-based soil databases. The soil properties must be manually input by the user. WARMF is relatively new, however, SWAT and HSPF have many more users. SWAT is better at simulating land-based processes and more convenient to input data. Thus, SWAT has been selected to do the comparison.

A detailed description of SWAT is given in Neitsch et al. (2002a,b) and Arnold et al. (1998). An overview is given here. SWAT is a river basin or watershed scale model developed by the USDA-ARS to predict the impact of land-management practices on water, sediment and agricultural chemical yields in large complex watersheds with varying soils, land use and management conditions over long periods of time. SWAT model operates on daily time step.

The first level of sub-division in SWAT is a sub-basin. A sub-basin contains at least one hydrologic response unit (HRU), a tributary channel and a main channel or reach. HRUs possess unique land-use, management and soil attributes (Neitsch et al., 2002b). One reach or main channel is associated with each sub-basin. Loads from the sub-basin enter the channel network in the associated reach segment. SWAT simulates surface runoff volumes for each HRU using a modification of the SCS curve number method and peak runoff rates using a modified rational method. Using small and relatively uniform HRUs reduces the error due to lumping effects. Another advantage of using SWAT is that it is a continuous simulation model. The initial conditions for each day are determined based on the conditions on the previous day. SWAT has a plant-growth model component based on heat units. The plant-growth model

is used to assess removal of water and nutrients from the root zone, transpiration and biomass yield. Sediment yield is estimated for each HRU with the Modified Universal Soil Loss Equation (MUSLE) (Williams, 1975). Using MUSLE single-storm sediment yield can be estimated. The hydrology model supplies estimates of runoff volume and peak runoff rate, which are used to calculate the runoff erosive energy variable. Other factors of the erosion equation are evaluated as described by Wischmeier and Smith (1978).

SWAT tracks the movement and transformation of several forms of nitrogen (N). Nutrients may be introduced to the main channel and transported downstream through surface runoff and lateral flow. The three major forms of nitrogen in mineral soils are organic nitrogen associated with humus, mineral forms of nitrogen held by soil colloids, and mineral forms of nitrogen in solution. Plant use of nitrogen is estimated using the supply and-demand approach. In addition to plant use, nitrate and organic N may be removed from the soil via mass flow of water. Amounts of NO_3 contained in runoff, lateral flow and percolation are estimated as products of the volume of water and the average concentration of nitrate in the layer. Organic N transport with sediment is calculated with a load function developed by McElroy et al. (1976) and modified by Williams and Hann (1972) for individual runoff events. The load function estimates the daily organic N runoff loss based on the concentration of organic N in the topsoil layer, the sediment yield and the enrichment ratio. Organic nitrogen levels are assigned based on C: N ratio for humic material. The enrichment ratio is the concentration of organic N in the sediment divided by that in the soil (Neitsch et al., 2002a).

SWAT simulates the movement and transformation of several forms of phosphorous (P). The three major forms of phosphorus in mineral soils are organic phosphorus associated with humus, insoluble forms of mineral phosphorus and phosphorus in soil solution. The concentration of solution phosphorus in all layers is initially set to 5 mg/kg soil, representative of unmanaged land under native vegetation and 25 mg/kg soil for a cropland. The amount of soluble P removed in runoff is predicted using labile P concentration in the top 10 mm of soil, the runoff volume and a partitioning factor. Phosphorus in soil is mostly associated with the sediment phase. Organic and mineral P attached to soil particles are transported by surface runoff to the main channel.

SWAT allows the user to define management practices taking place in every HRU that may influence simulation results. The user may define the beginning and the ending of the growing season; specify timing and amounts of fertilizer, pesticide and irrigation applications, timing of tillage operations and other agricultural practices. Once SWAT determines the loads of water, sediment, nutrients and pesticides to the main channel, the loads are routed through the stream network of the watershed.

2.2. Description of study area

The 126 km² Turkey Creek watershed is located in Jefferson County, southwest of Denver, Colorado. The topography is mostly steep with elevations ranging from about 3200 to 1800 m. Based on the United States Geological Survey (USGS) Land Use Land Cover data, the land use classification in Turkey Creek Watershed consists of open water, residential, commercial, deciduous forest, evergreen forest, shrubland, grassland, pasture-hay, row crops and urban and recreational grasses. More than 80% of the watershed classified as forested or shrubland. Resistant parent materials (primarily igneous and metamorphic rocks) and steep slopes have limited soil development in the watershed (Murray, 2003). The hydrologic group classification reported in the SSURGO database show that most part of the watershed is classified in hydrologic soil group D. The USGS gage 06710995 at Turkey Creek at the mouth of the canyon near Morrison, Colorado is used in this study for hydrologic calibration.

2.3. Model input

The SWAT model requires weather data input and three GIS data layers, namely digital elevation model (DEM), soils, and land use. The 30 m seamless DEM used in this study was taken from the USGS database, which offers a seamless DEM coverage of the entire United States. The model calculates sub-basin parameters such as slope and slope length as well as the definition of the stream network using the DEM. The DEM for Turkey Creek Watershed is shown in Fig. 1. The resulting stream network was used to delineate the entire watershed into a reasonable number of sub-basins. Before the DEM was used for modeling, it was projected to Universal Transverse Mercator (UTM) under appropriate zones. Most part of Colorado belongs zone 13; a small portion of the state on the west belongs to zone 12. Turkey Creek watershed is in zone 13.

The land use data was derived from the 1992 National Land Cover Dataset (NLCD)—a 21-class land cover classification scheme applied over the United States. It is derived from the early to mid-1990 Landsat Thematic Mapper Satellite data. In this study, there are six major types of land cover namely: water, residential, agricultural, range, pasture, and forest as shown Fig. 2. Recent estimates of land cover indicate that the forest canopy cover was estimated to be 60–70% (Bossong et al., 2003). The 1992 NLCD data also shows that forest cover accounts for about 67%. There is no considerable land use change in the watershed. Because agricultural land accounts for only 0.35%, the influence from crop rotation is negligible. Thus, The land cover data from 1992 is acceptable for a comparative study of the effect of soil data resolution.

SWAT was originally developed for STATSGO data to define soil attributes. The GIS soil data must have either Soils5id for USDA soil series (S5ID) or State Map Unit ID

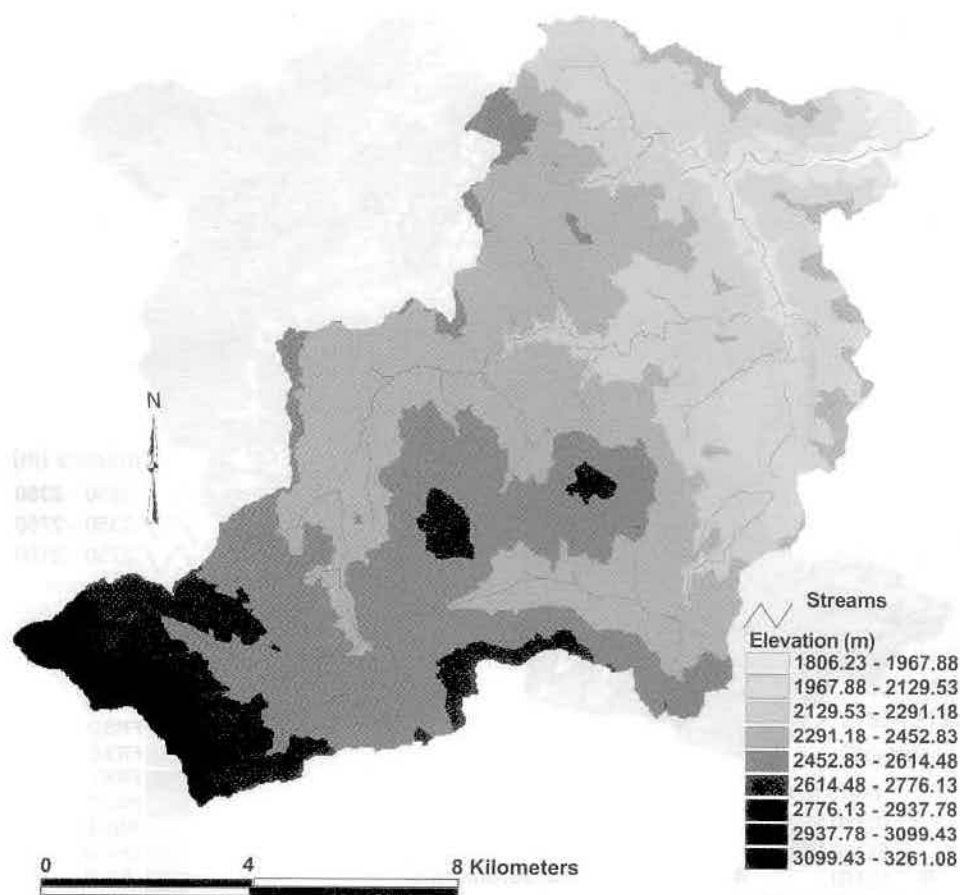


Fig. 1. Digital elevation model for Turkey Creek Watershed, Colorado.

(STMUID) to link an area anywhere in the watershed to the STATSGO database. In this study, the STMUID was used. The STATSGO and SSURGO soil maps for Turkey Creek Watershed are given in Figs. 3 and 4.

The soils data required by SWAT and extracted from both soil databases include several physical and chemical characteristics. These data include soil hydrologic group, maximum rooting depth, soil profile depth from soil surface to bottom of a layer, moist bulk density, available water capacity of the soil layer, saturated hydraulic conductivity, and soil texture data (% clay, sand, silt, and rock fragment content) that are required in stream flow computations and others such as USLE_K required to compute sediment yield. Chemical properties of soil such as the fraction of porosity (void space) from which anions are excluded, and organic carbon content (% soil weight) and initial concentrations of chemicals in the soil are also required in SWAT to determine transformation and transport of chemicals.

SWAT requires daily values of weather data as an input. These data are precipitation, maximum and minimum temperature, solar radiation, relative humidity and wind speed. To run SWAT one can either prepare a file that contains observed data, or use daily values simulated by the SWAT model weather generator from monthly average data summarized over a number of years (Neitsch et al.,

2002a). Observed weather values from USGS Cooperative Observer Program (COOP) were utilized to create files for daily precipitation, minimum and maximum daily temperatures. The rest of the weather parameters such as solar radiation, wind speed and relative humidity are simulated by weather generator module of SWAT.

2.4. Sub-basin delineation

In SWAT the watershed is divided into sub-basins and each sub-basin contains at least one HRU. Each HRU has uniform land cover and soil type. The smaller the area assigned to each sub-basin, the more the number of sub-basins and the more detailed the drainage network (Di Luzio, et al., 2002). A less detailed watershed can be delineated by choosing few but larger size sub-basins. In this study, the watershed was divided to 61 sub-basins. The total number of HRUs was 261 when STATSGO data was used and 1301 when the more detailed SSURGO data was used. Bingner et al. (1997) evaluated sub-watershed size dependency of the Soil and Water Assessment Tool (SWAT) erosion model to adequately simulate annual runoff and fine sediment from the 21.3 km² Goodwin Creek Watershed (GCW). The results showed that runoff volume is not appreciably affected by the number and size of sub-watersheds. They determined that, an upper limit to

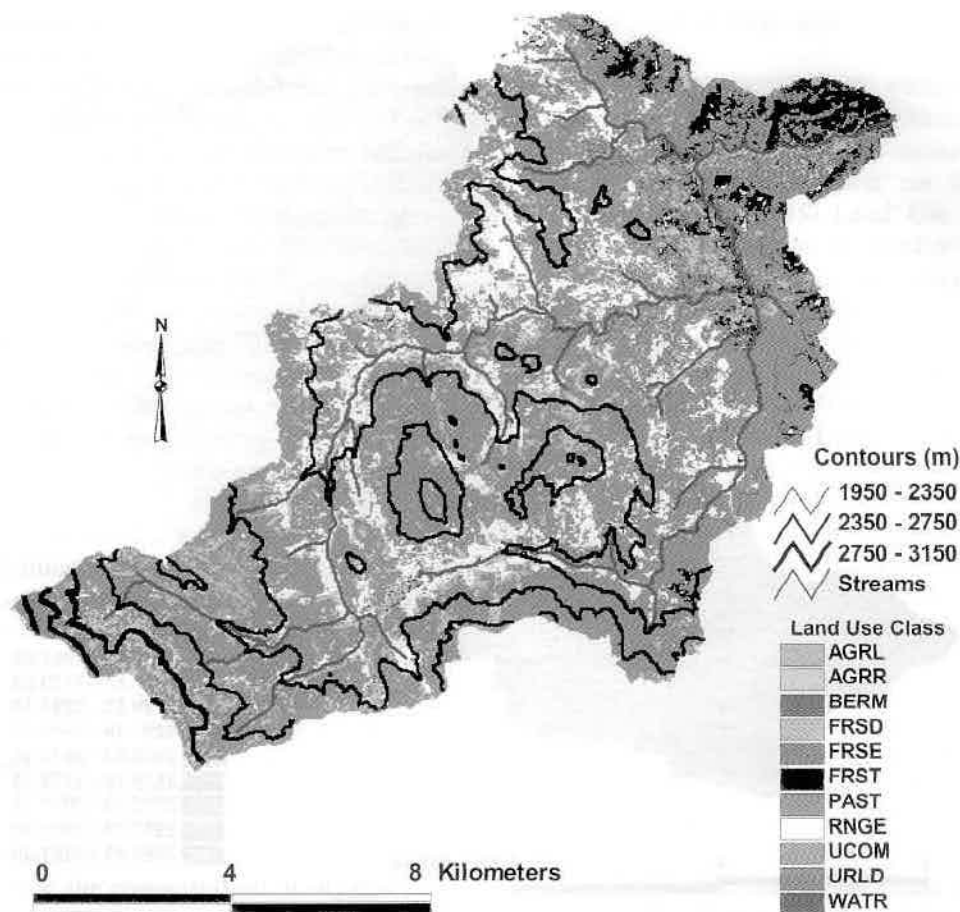


Fig. 2. Land cover classification for Turkey Creek Watershed, Colorado. *Note:* AGRL, Agricultural Land-Generic; AGRR, Agricultural Land-Row Crops; BERM, Bermudagrass; FRSD, Forest-Deciduous; FRSE, Forest-Evergreen; FRST, Forest-Mixed; PAST, Pasture; RNGE, Range-Grasses; UCOM, Commercial; URLD, Residential-Low-Density; WATR, Water.

sub-watershed size is required to adequately simulate fine sediment yield produced from upland sources and decreasing the size of sub-watersheds beyond this threshold does not substantially affect the computed fine sediment yield. The sensitivity of the distributed hydrological SWAT model to the pre-processing of soil and land use data was tested for modeling rainfall-runoff processes in the Thyle catchment in Belgium (Romanowicz et al., 2005). The results suggest the resolution and fragmentation of the original map objects are significantly affected by the internal aggregation procedures of the SWAT model. The catchment size threshold value (CSTV) is thereby a key parameter controlling the internal aggregation procedure in the model. It is shown that a parabolic function characterizes the relationship between the CSTV and the hydrological modeling performance of the uncalibrated model, suggesting that optimal uncalibrated modeling results are not obtained when the CSTV is minimized.

The number of sub-basins chosen depends on the size of the watershed and the amount of detail required to meet project goals. The sub-basin delineation should be detailed enough to capture significant variability. SWAT allows users to choose the number of sub-basins by deciding the initial threshold area, which in many studies is arbitrarily

chosen. The interface lists a minimum, maximum, and suggested sub-watershed area, which are shown in the stream definition section. This function plays a role in determining the detail of the stream network and the size and number of sub-basins. In this study, the minimum and maximum threshold values were 13 and 5329 ha. SWAT suggested minimum value was 200 ha, which gave only 32 sub-basins. To incorporate more complexity into a data set, it is recommended in the SWAT user manual (Neitsch et al., 2002b) that the user defines a greater number of sub-basins in the watershed rather than many HRUs within a few sub-basins. Thus, the number of sub-basins was increased to 61 by decreasing the threshold value to 100 ha.

3. Results

3.1. Comparison before model calibration

SWAT outputs were compared for the two data sets before calibration. Comparison using uncalibrated models is useful to evaluate the differences in model predictions because calibration masks the differences that may occur as a result of the soil data sets. In addition, the uncalibrated model results can show how good each data set predicts

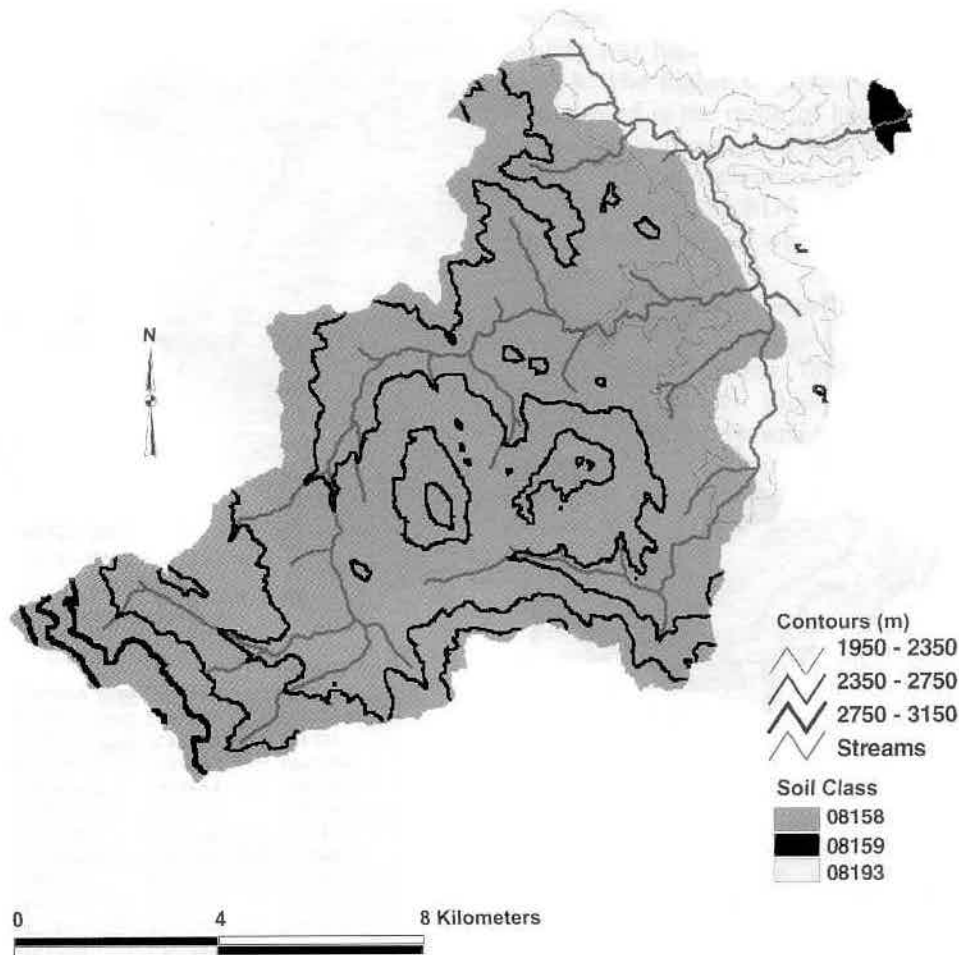


Fig. 3. STATSGO soil map for Turkey Creek Watershed, Colorado.

stream flow before calibration, which would indicate the effort required for calibration when using each data set. Our observation shows that the STATSGO data predicted stream flow well before calibration and would require less effort in calibration. Figs. 5–8 show stream flow and water-quality predictions from uncalibrated model. Fig. 5 shows the stream flow predictions from each soil type. Stream discharge using SSURGO soil data set was generally higher than STATSGO. Both soils resulted in over prediction of stream flow compared to measured data. One possible explanation for this result is that soil layers in the watershed are generally thin and underlain by bedrock. Thus, infiltration into the soil may be diverted by the bedrock to the stream. Evaluation of the soil data set shows there are more HRUs in soil hydrologic group D when SSURGO soil is used. SSURGO, with the highest spatial resolution, has 51 unique soil types in the watershed distributed in 1301 HRUs, while STATSGO has only three distributed in 261 HRUs. Each HRU has a unique soil and land use. The differences in runoff and soil properties can be attributed to the difference in the spatial resolution because all other inputs were the same for both setups. The finer resolution soil dataset (SSURGO) resulted in more areas with low infiltration potential. This resulted in a

higher discharge when SSURGO is used compared to STATSGO.

In contrast to stream flow, the sediment loading using STATSGO was higher than when SSURGO data were used. Based on the flow results alone one would expect sediment loading from SSURGO would be higher than STATSGO because runoff is higher when SSURGO data set is used. But it is also important to look at other parameters affecting sediment yield. An example is the soil erodibility factor, k_{usle} , which represents both susceptibility of soil to erosion and the rate of sediment runoff. Soils high in clay have low k_{usle} values, about 0.05–0.15, because they are resistant to detachment. Coarse-textured soils, such as sandy soils, also have low k_{usle} values, about 0.05–0.2, because of low runoff even though these soils are easily detached. Medium-textured soils, such as the silt loam soils, have moderate k_{usle} values, about 0.25–0.4. Soils having high silt content are most erodible of all soils. As the result of the fine resolution, more small size areas with lower k_{usle} values were identified when using SSURGO model. The area weighted average k_{usle} value was 0.19 when SSURGO was used and it was 0.32 when STATSGO was used. This has contributed to the higher sediment loading when STATSGO is used as shown in Fig. 6. Fig. 7

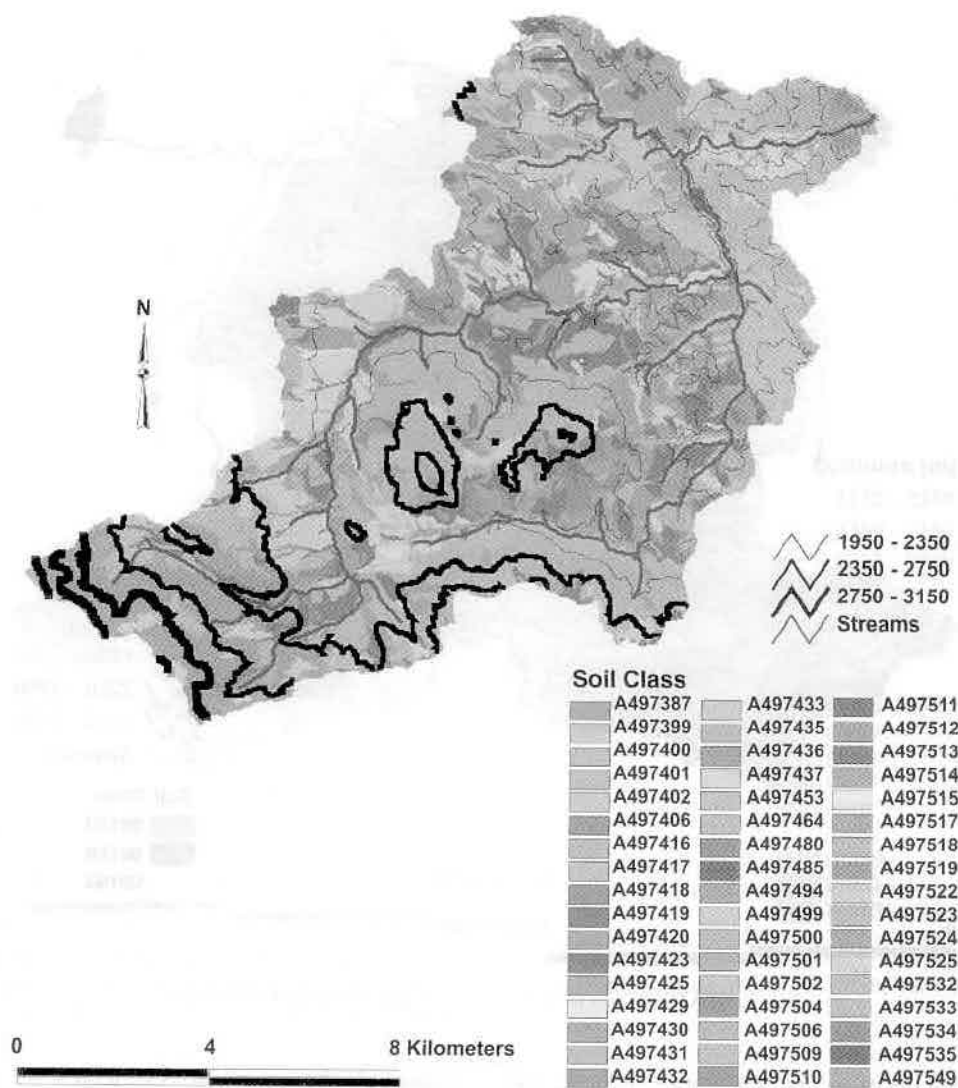


Fig. 4. SSURGO map for Turkey Creek Watershed, Colorado.

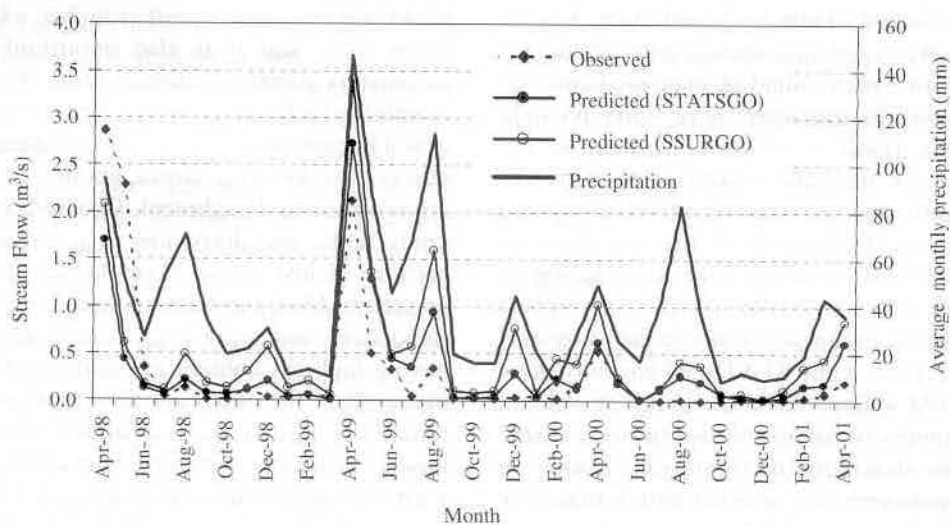


Fig. 5. Time series monthly average measured and SWAT predicted stream flow.

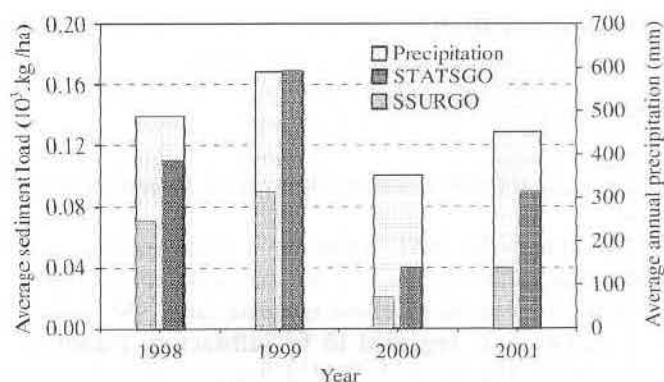


Fig. 6. Comparison of basin average annual sediment loading using SSURGO and STATSGO before calibration.

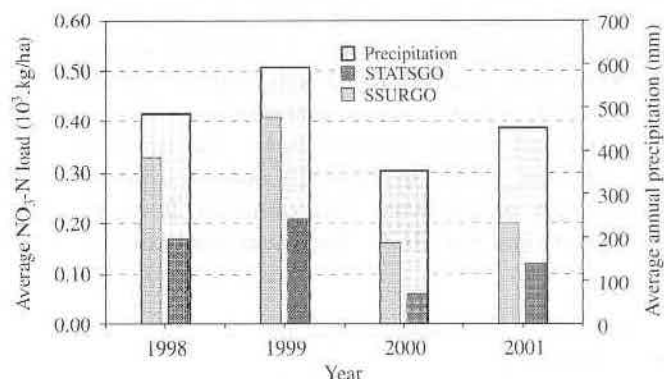


Fig. 7. Comparison of basin average annual NO₃ loading using SSURGO and STATSGO before calibration.

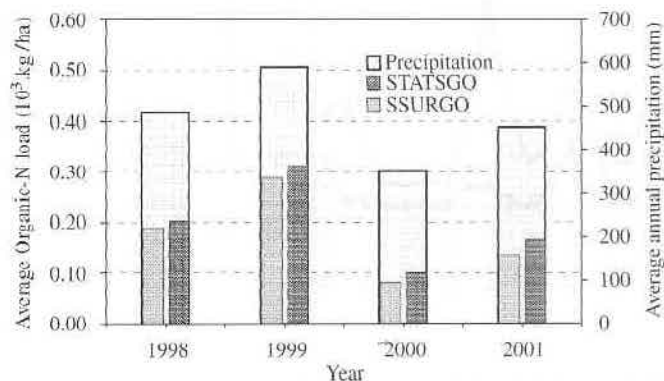


Fig. 8. Comparison of basin average annual Organic-N loading using SSURGO and STATSGO before calibration.

shows that the annual average basin NO₃ loading when SSURGO soil data set is used is higher than when STATSGO is used. This can be attributed to the fact that more flow was generated when SSURGO is used, because as mentioned earlier amounts of NO₃ contained in runoff and lateral flow in SWAT are estimated as products of the volume of water and the average concentration of nitrate in

the layer. However, annual average basin organic-N loading was higher when STATSGO is used as shown in Figs. 8. The higher Organic-N loading when STATSGO data is used is the result of higher sediment loading since Organic-N is transported with sediment. The results were similar for phosphorous, more dissolved P loading occurred when SSURGO was used and more organic-P loading was predicted when STATSGO is used.

3.2. Comparison after model calibration

As mentioned earlier, the stream flow predictions for uncalibrated model show that the STATSGO data predicted stream flow well before calibration and required less effort in calibration. However, calibration generally improves the reliability of the model predictions. The hydrologic calibration was performed using data from the gage station located at Turkey Creek at mouth of canyon near Morrison (Station 06710995). Gage data is available for this station from 1998 to 2001. The SWAT model in this study was calibrated for flow for the period of 1998–2001. Calibration was relatively faster and easier when STATSGO data was used although SSURGO provided more options for calibration because more soil types allow more adjustments.

The measured total flow was separated into base flow and surface flow using SWAT-base flow filter program (Arnold and Allen, 1999). About 57% of the flow was base flow. To compare the goodness of fit between simulated and measured data, relative error was used. Hydrologic calibration parameters for surface and base flow were adjusted to reduce relative error. Relative error is given by

$$\text{Relative error} = 100(\text{observed} - \text{simulated}) / \text{observed} \quad (1)$$

Calibration parameters were adjusted until relative error value is less than 5% for both soil types. Curve number, SOL-AWC and ESCO are basic water balance parameters recommended in the SWAT user manual (Neitsch et al., 2002b) to be adjusted to calibrate surface runoff. Model predictions before calibration were close to predicted values. Thus, slight adjustments were made to these parameters. Flow calibration with STATSGO and SSURGO was performed by adjusting the curve number on forestland covers, which account for about 70% of the watershed. The curve number values were uniformly reduced by 5% for both soil types to adjust SWAT predicted flow to mean daily measured flow. The remaining lands covers (agricultural, residential and pasture) account only for 30% and had reasonable CN values compared to literature values and were not altered. In addition to CN values, minor adjustment were made on other SWAT parameters such as soil evaporation compensation factor (ESCO) and available water capacity (SOIL_AWC) on all HRUs. For SSURGO soil data, ESCO was decreased by 2% and Soil_AWC was increased by 0.04 from default values. For STATSGO, ESCO was decreased by 1% and

Soil_AWC was increased by 0.01. SWAT predicted values were compared after calibration and time series monthly average measured and SWAT predicted flow values is shown in Fig. 9. The results of flow calibration are shown in Table 1. As shown in Table 1, model setups with both soil types did calibrate well although the results were slightly better when SSURGO soil is used. Calibration was done until the relative error given by Eq. (1) was less than 5% for total flow. The relative error for surface flow was less than 1% for both soil types. Errors during base flow periods are also less than 10%, although they were not as low as they are for surface and total flow.

Nash-Sutcliffe coefficient of efficiency (NSE) (Nash and Sutcliffe, 1970) and Root Mean Square Error (RMSE) were used to evaluate the goodness of fit of calibration. NSE determines the model efficiency as a fraction of the measured stream flow variance that is reproduced by the model. NSE given is calculated as

$$NSE = 1 - \frac{\sum(Q_o - Q_s)^2}{\sum(Q_o - \bar{Q}_o)^2} \quad (2)$$

where Q_o is measured stream flow, Q_s is simulated flow and $\bar{Q}_o$ is measured mean stream flow. RMSE is the root mean

squared errors given by

$$RMSE = \sqrt{\frac{\sum_i^n (Q_s - Q_o)^2}{n}} \quad (3)$$

where Q_s is simulated stream flow, Q_o is measured stream flow.

The closer the NSE value to 1.0 the better is the estimation of the stream flow by the model. $NSE \geq 0.75$ is considered to be an excellent estimate, and NSE between 0.75 and 0.36, is regarded to be satisfactory (Motovilov et al., 1999). The closer the RMSE value to zero, the better the estimation. The NSE value obtained was 0.61 and 0.70 for STATSGO and SSURGO data, respectively, and the RMSE values are 0.50 and 0.40 for STATSGO and SSURGO, respectively. Thus, both data provide a satisfactory calibration although the SSURGO data yield a slightly better fit. Scatter plots of average monthly measured and simulated flows are shown in Figs. 10 and 11 for STATSGO and SSURGO soils, respectively. The NSE and RMSE values give a quantitative comparison of observed and model predicted flows. The trend lines and the scattered points in Figs. 10 and 11 allow a visual comparison of predicted and observed flows for both STATSGO and SSURGO. The closer the trend line or the

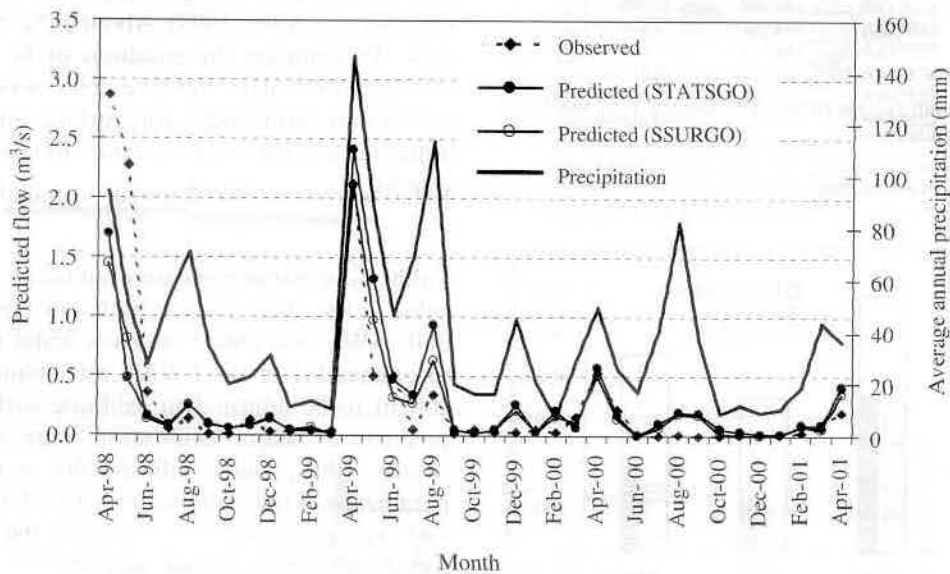


Fig. 9. Time series monthly average measured and SWAT predicted flow.

Table 1
Annual hydrologic calibration results on Turkey Creek Watershed

Soil type	Item	Surface flow (m ³ /s)	Base flow (m ³ /s)	Total flow (m ³ /s)
STATSGO	Measured	0.123	0.169	0.293
	Predicted	0.134	0.170	0.305
	Relative error (%)	0.33	9.30	4.10
SSURGO	Predicted	0.122	0.179	0.302
	Relative error (%)	−0.781	5.72	3.04

scattered points to 1–1 line ($y = x$ line), the closer the predicted value to the observed value. Thus, as shown in Figs. 10 and 11 for both STATSGO and SSURGO, the SWAT model predicted stream flow relatively better during high flow conditions compared to low flow conditions. This indicates that the model predicted surface runoff better

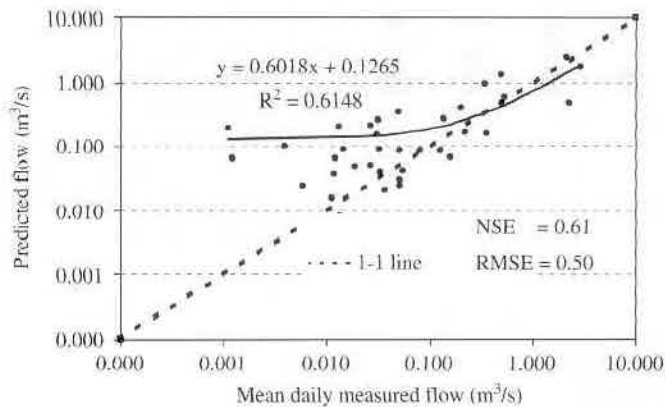


Fig. 10. Scatter plot of monthly average measured and SWAT predicted flow at Turkey Creek Watershed for STATSGO soil data set.

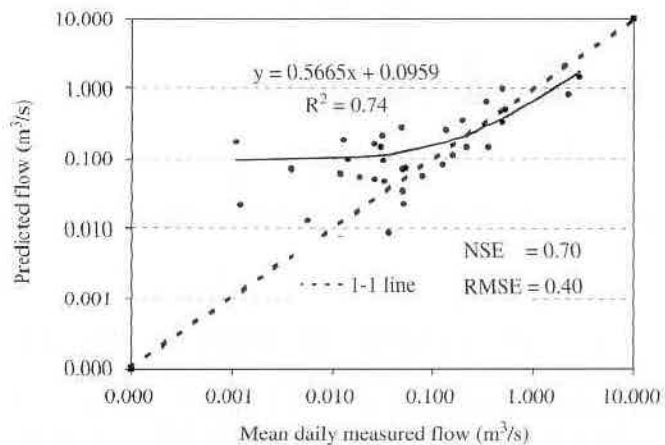


Fig. 11. Scatter plot of monthly average measured and SWAT predicted flow at Turkey Creek Watershed for SSURGO soil data set.

than base flow since surface flow is dominant during high flow conditions and base flow is dominant during low flow conditions. The lower relative error values for surface flow compared to base flow in Table 1 also shows relatively better model predictions for surface flow than base flow. In general, simulated total stream flow matched the measured total stream flow fairly well as given by the R^2 values of 0.62 and 0.74, respectively, for STATSGO and SSURGO.

Fig. 12 shows sediment yield after calibration both for calibration period (1998–2001) and validation period (2002–2005). As in the uncalibrated model, sediment yield was still higher for STATSGO data after stream flow calibration for both calibration and validation period although it was generally low for both soil types as a result of calibration, which reduced the total flow. Fig. 13 shows NO_3 loading after calibration for both soil types during calibration and validation periods. As in sediment loading NO_3 was generally low after calibration. In contrast to uncalibrated model, NO_3 loading was higher for STATSGO than SSURGO after calibration. This can be attributed to relatively higher stream flow values for STATSGO after calibration. Fig. 14 shows Organic-N loading after calibration for both soil types during calibration and validation periods. The trends for Organic-N are similar to sediment yield (higher sediment yield for STATSGO) because organic-N transport depends on sediment yield. The results in Figs. 12–14 are for the model calibrated for flow only. There were no sufficient data to perform water-quality calibration.

4. Discussion

As stated earlier, the results obtained vary depending on regional differences in climate, topography and soil types. Some indicated that there might be no significant increase in the accuracy of models, as a result of more precise topographic or soil information, which increases the input data collection and preparation. Others have determined that SSURGO soil provided an overall better prediction of the stream discharges than STATSGO; although both did a good and comparable job on predicting the high stream

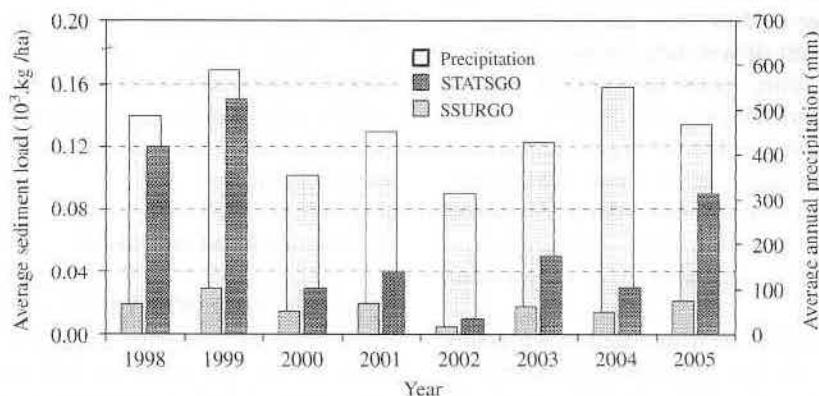


Fig. 12. Comparison of basin average annual sediment loading after calibration during calibration period (1998–2001) and validation period (2002–2005).

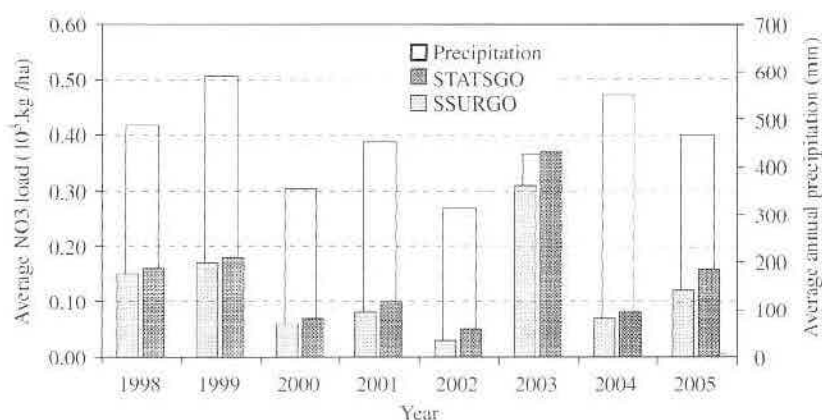


Fig. 13. Comparison of basin average annual NO₃ loading after calibration during calibration period (1998–2001) and validation period (2002–2005).

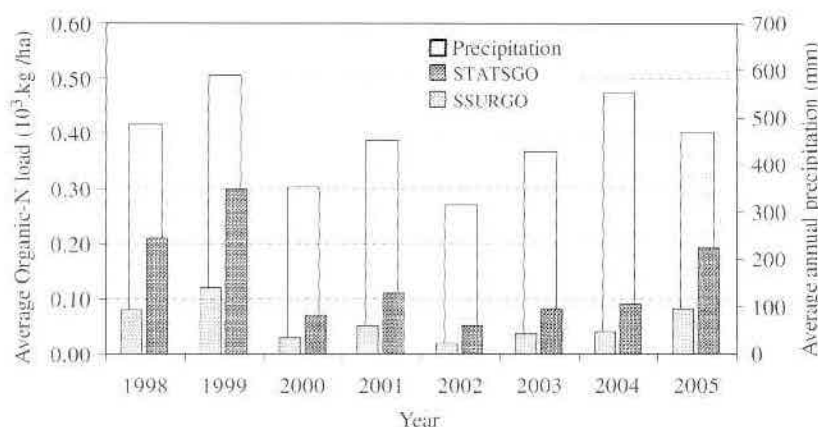


Fig. 14. Comparison of basin average annual Organic-N loading after calibration during calibration period (1998–2001) and validation period (2002–2005).

flows (Wang and Melesse, 2006). In our case study, SSURGO predicted less stream loading than STATSGO before and after calibration in terms of sediment and sediment-attached nutrients and vice versa for dissolved nutrients carried by the flow especially before calibration. When compared to mean daily measured flow, STATSGO performed better relative to SSURGO before calibration. SSURGO provided better results after calibration. Based on this research and other studies there are differences in model predictions as a result of soil data resolution, which can be insignificant in some cases as pointed out by Chaplot (2005). Much more effort is required to prepare and calibrate the model when a fine resolution soil data is used especially when applying it to large watersheds. Modelers need to weigh merits and demerits before selecting the type of data resolution they are going to use depending on the watershed size and level of accuracy required.

In addition, soil data resolution when used in watershed models affects other parameters physical parameters. As indicated by Romanowicz et al. (2005) in addition to the quality of the soil and land used data, the resolution and

fragmentation of original map objects are affected by catchment threshold value that controls the internal aggregation procedures in SWAT model. The fact that the high resolution SSURGO has more HRUs in a sub-basin will affect slope and slope length computations and sediment yield, the effect of which will be more prevalent in mountainous watersheds compared to flat topography because of relatively more variations in slope and slope length in mountainous watersheds like Turkey Creek Watershed. Climate is another important factor. Watershed wetness prior to the storm event of interest (antecedent moisture condition, AMC) is affected by the watershed climate including amount of precipitation and evaporation.

5. Summary and conclusion

In this study, STATSGO and SSURGO soil data sets were used as an input for SWAT model simulation and stream flow and water-quality predictions were compared for Turkey Creek Watershed, Colorado. The comparisons were made before and after calibration. It was determined

from comparison of the two soil data sets that some areas in the watershed were lumped in one soil type when low resolution soil (STATSGO) were grouped into different soil types as a result of discretization when SSURGO is used. The discretization resulted in more areas with soil types with low infiltration potential. Thus, before calibration SSURGO over predicted STATSGO in terms of total runoff. However, SSURGO predicted less stream loading than STATSGO in terms of sediment and sediment-attached nutrients components, and vice versa for dissolved nutrients carried by the flow. When compared to mean daily measured flow, before calibration, STATSGO performed better relative to SSURGO. SSURGO provided better results after calibration as evaluated by R^2 value (0.74 compared to 0.61 for STATSGO) and the NSE values (0.70 and 0.61 for SSURGO and STATSGO, respectively). These results suggest a relatively better estimation by SSURGO than STATSGO after calibration, although both are in the same satisfactory range. SSURGO provides more options for calibration, although the simulation runs slower when using SSURGO. Modelers need to weigh the benefits before selecting the type of data resolution they are going to use depending on the watershed size and level of accuracy required because more effort is required to prepare and calibrate the model when a fine resolution soil data is used.

Acknowledgments

This work was supported by the Environmental Science and Engineering Division of Colorado School of Mines with funding provided by the Water Environment Research Foundation (WERF). We also appreciate the technical support from people working on SWAT model development.

References

- Arnold, J.G., Allen, P.M., 1999. Automated methods for estimating base flow and ground water recharge from stream flow records. *Journal of the American Water Resources Association* 35 (2), 411–424.
- Arnold, J.G., Srinivasan, R., Mutiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association* 34 (1), 7389.
- Bingner, R.L., Garbrecht, J., Arnold, J.G., Srinivasan, R., 1997. Effect of watershed subdivision on simulation runoff and fine sediment yield. *Transactions of the ASAE* 40 (5), 1329–1335.
- Bosson, C.R., Caine, J.S., Stannard, D.L., Flynn, J.L., Stevens, M.R., Helny-Dash, J.S., 2003. Hydrologic Conditions and Assessment of Water Resources in the Turkey Creek Watershed, Jefferson County, Colorado, 1998–2001. US GEOLOGICAL SURVEY. Prepared in cooperation with the Jefferson County Planning and Zoning Department, Denver, Colorado. Water-Resources Investigations Report 03-4034.
- Brown, D.G., Bian, L., Walsh, S.J., 1993. Response of a distributed watershed erosion model to variations in input data aggregation levels. *Computers and Geosciences* 19 (4), 499–509.
- Chaplot, V., 2005. Impact of DEM mesh size and soil map precision for the prediction of water, sediment and NO_3 loads in a watershed. *Journal of Hydrology* 312, 207–222.
- Di Luzio, M., Srinivasan, R., Arnold, J.G., 2002. Integration of watershed tools and SWAT model into BASINS. *Journal of American Water Resources Association* 38 (4), 1127–1141.
- Haan, C.T., Barfield, B.J., Hayes, J.C., 1994. *Design Hydrology and Sedimentology for Small Catchments*. Academic Press, USA.
- Kroner, S.M., Cozzie, D.A., 1999. *Data Collection for The Hazardous Waste Identification Rule: Section 7.0 Soil Data*. US Environmental Protection Agency Office of Solid Waste, Washington, DC, 20460.
- Levick, L.R., Semmens, D.J., Guertin, D.P., Burns, I.S., Scott, S.N., Unkrich, C.L., Goodrich, D.C., 2004. Adding global soils data to the automated geospatial watershed assessment tool (AGWA). Second International Symposium on Transboundary Waters Management, Tucson, AZ, 2004 November 16–19. CD Proceedings, 8pp.
- Lin, H.S., Wheeler, D., Bell, J., Wilding, L., 2005. Assessment of soil spatial variability at multiple scales. *Ecological modelling* 182 (3–4), 271–290.
- McElroy, A.D., Chiu, S.Y., Nebgen, J.W., 1976. Load functions for assessment of water pollution from non-point sources. EPA document EPA 600/2-76-151. USEPA, Athens, GA.
- Murray, K.E., 2003. Spatial analysis and modeling of water resource issues in Colorado utilizing geographic information systems. A Ph.D. Dissertation submitted to the Faculty and Board of Trustees of the Colorado School of Mines, Golden, Colorado.
- Motovilov, Y.G., Gottschalk, L., Engeland, K., Rodhe, A., 1999. Validation of a distributed hydrological model against spatial observations. *Agricultural and Forest Meteorology* 98–99, 257–277.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, part 1—a discussion of principles. *Journal of Hydrology* 10 (3), 282–290.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., King, K.W., 2002a. *Soils and Water Assessment Tool. Theoretical Documentation, Version 2000*.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Srinivasan, R., Williams, J.R., 2002b. *Soil and Water Assessment Tool (SWAT). SWAT user manual, version 2000*. Water Resources Institute, College Station, Texas. TWRI Report TR-192.
- Peschel, J.M., Haan, P.K., Lacey, R.E., 2003. A SSURGO Pre-Processing Extension 798 for the Arc View Soil and Water Assessment Tool. ASAE Paper no. 032123, St.799 Joseph, Michigan.
- Prasanna, H.G., Mulla, D.J., 2005. Scale effects of STATSGO vs. SSURGO soil databases on water quality predictions. Watershed management to meet water quality standards and emerging (total maximum daily load) (TMDL). In: *Proceedings of the Third Conference* 5–9 March 2005, Atlanta, Georgia, USA.
- Romanowicz, A.A., Vancloster, M., Rounsevell, M., La Junesse, L., 2005. Sensitivity of the SWAT model to the soil and land use data parameterization: a case study in the Thyle catchment, Belgium. *Ecological modelling* 187 (1), 27–39.
- Soil Survey Division Staff, 1993. *Soil Survey Manual*. US Department of Agriculture Handbook no. 18. US Government Printing Office, Washington, DC.
- USDA, 1994. *State Soil Geographic (STATSGO) Data Base*. Natural Resources Conservation Service, Fort Worth, TX. Available online <http://www.ftw.nrcs.usda.gov/stat_data.html>.
- USDA, 1995. *Natural Resources Conservation Services (NRCS). Soil Survey Geographic (SSURGO) Data Base: Data Use information*. National Soil Survey Center, Lincoln, NE.
- USDA, 2006. *Natural Resources Conservation Service (NRCS). National Cartography and Geospatial Center. Soil Survey Geographic (SSURGO)*. Available online <<http://www.ncgc.nrcs.usda.gov/products/datasets/ssurgo/>>.

- Wang, X., Melesse, A.M., 2006. Effects of STATSGO and SSURGO as inputs on SWAT model's snowmelt simulation. *Journal of American Water Resources Association* 42 (5), 1217–1236.
- Wischmeier, W.H., Smith, D.D., 1978. Predicting rainfall losses: a guide to conservation planning. USDA Agricultural Handbook no. 537. US Gov. Print. Office, Washington, DC.
- Williams, J.R., 1975. Sediment-yield prediction with universal equation using runoff energy factor. In *Present and prospective technology for predicting sediment yield and sources*. ARS-S-40, U.S. Gov. Print. Office, Washington, DC, pp. 244–252.
- Williams, J.R., Hann, R.W., 1972. HYMO, a problem-oriented computer language for building hydrologic models. *Water Resources Research* 8 (1), 79–85.
- Zhang, W., Montgomery, D.R., 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research* 30, 1019–1028.

Are motorway wildlife passages worth building? Vertebrate use of road-crossing structures on a Spanish motorway

C. Mata*, I. Hervás, J. Herranz, F. Suárez, J.E. Malo

Departamento de Ecología, Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

Received 21 June 2006; received in revised form 31 January 2007; accepted 10 March 2007

Available online 30 April 2007

Abstract

Numerous road and railway construction projects include costly mitigation measures to offset the barrier effect produced on local fauna, despite the scarcity of data on the effectiveness of such mitigation measures. In this study, we evaluate the utility of different types of crossing structures. Vertebrate use of 43 transverse crossing structures along the A-52 motorway (north-western Spain) was studied during spring 2001. Research centered on wildlife passages (9), wildlife-adapted box culverts (7), functional passages (6 overpasses, 7 underpasses) and culverts (14), with marble dust being used to record animal tracks. A total of 424 track-days were recorded, with most of the larger vertebrate groups present in the area being detected. All crossing structure types were used by animals, although the intensity of use varied significantly among them (Kruskal–Wallis test, $p < 0.05$); culverts were used less frequently than other structures. Crossing structure type and width were identified as the most important factors in their selection for use. Wildlife passages and adapted culverts allowed crossing by certain species (wild boar, roe deer, Eurasian badger), which do not tend to cross elsewhere. These results highlight the importance of using both mixed-type structures and wildlife passages in reducing the barrier effect of roads.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Barrier effect; Habitat fragmentation; Road ecology; Vertebrate; Wildlife crossing structures

1. Introduction

Animal population isolation produced by habitat fragmentation constitutes one of the commonest causes of local extinctions (Hunt et al., 1987; Clarke et al., 1998; Lodé, 2000; Fahrig, 2003). Linear infrastructures not only physically destroy habitats but also become barriers, which considerably limit movement and dispersion of terrestrial vertebrates (Oxley et al., 1974; Mader, 1984; Camby and Maizeret, 1987; Walker et al., 2003). As a result, over the last few decades, the inclusion of faunal passages or the modification of existing culverts to serve as crossing structures has frequently been recommended in Environmental Impact Assessments and Statements (Beier and Loe, 1992; De Santo and Smith, 1993; van Bohemen, 1998). Wildlife crossing structures aim to re-establish the movement of animals between both sides of a road or railway line, helping to connect the areas affected by the

transport corridor (Saunders and Hobbs, 1991; Clergeau, 1993; Rodríguez et al., 1996; Putman, 1997).

Knowledge of the effectiveness of wildlife crossing structures is still scarce, although increasing quickly. The regular use of different wildlife passage types has been demonstrated for some vertebrate groups (Singer and Doherty, 1985; Foster and Humphrey, 1995; Bekker and Canters, 1997; Gloyne and Clevenger, 2001; Puky and Vogel, 2003; Taylor and Goldingay, 2003; Dodd et al., 2004) but others, like ungulates, are noteworthy for their reluctance to use most crossing structures (Reed, 1981; Vassant et al., 1993; Mata et al., 2005). The key elements for the design of effective mitigation measures are not yet well understood and different solutions may be needed for different faunal groups. The modification of construction projects to adapt culverts for vertebrates or to include wildlife passages imposes costs on new projects. Consequently, it is important to evaluate which characteristics of existing crossing structures affect their use by different species, in order to design cost-effective passages for target species in new projects. It is therefore important to

*Corresponding author. Tel./fax: +34 914978011.

E-mail address: cristina.mata@uam.es (C. Mata).

determine the actual use of over- and underpasses specifically designed for wildlife, due to the high construction costs they represent (Rosell et al., 2003).

The present study had two objectives. Firstly, to discover which vertebrate groups use road crossing structures. This was determined through an analysis of structures specifically designed for wildlife, as well as normal structural features of the highway, such as culverts, which are designed for other purposes but which could also be important for wildlife (Hunt et al., 1987; Yanes et al., 1995; Rodríguez et al., 1996). Secondly, we looked at the relative importance of crossing structure design since structural parameters of passages (e.g. their width and whether they are over- or underpasses) could determine their effective use by wildlife (Olbrich, 1984; Clevenger et al., 2001). Of the 43 crossing structures considered here only 16 were included in our earlier study (Mata et al. 2005) and the present data was obtained 1 year earlier.

2. Materials and methods

2.1. Study area

The study was undertaken along a stretch of the Spanish A-52 motorway between Camarzana de Tera (Km 34, Zamora province) and Orense (Km 217, Fig. 1). This four lane highway is fenced along its entire length. It was opened to traffic in 1998 and carries approximately 4500 vehicles daily, 23% of them heavy vehicles.

The first few kilometers pass through the flat countryside of the northern Spanish plain. The relief increases gradually up to the town of Puebla de Sanabria, after which the landscape is mountainous (720–960 m a.s.l.). The increase in elevation coincides with a decrease in temperature and an increase in rainfall (from 400 to 950 mm). The vegetation along the first 10 km is dominated by

continental Holm oak (*Quercus rotundifolia*), accompanied by gum cistus (*Cistus ladanifer*), broom spp. (*Cytisus scoparius* and *Cytisus multiflorus*), low shrubs (*Cistus laurifolius*, *Halimium umbellatum*, *Thymus mastichina*, *Thymus zygis*, *Lavandula stoechas* subsp. *pedunculata*) and pastures of *Agrostis castellana*. Cultivations of maize and non-irrigated cereals are also present but restricted to valley bottoms. There are also vineyards of limited extent. The next 45 km passes through patches of Pyrenean oak (*Quercus pyrenaica*), broom spp. and other low scrub (*Genista tridentata*, *Halimium ocymoides*, *Halimium lasianthum*) and moist meadows. The remaining section (in Orense province) is flanked by forests dominated by Pyrenean oak woodlands, some more heavily disturbed than others.

The progressive abandonment of fields and rural countryside, which is especially noticeable in Zamora province, is having an important impact on the fauna of the area. It has led to an intense regeneration of natural vegetation, leading in turn to an increase in the population of large herbivores, such as red deer (*Cervus elaphus*), roe deer (*Capreolus capreolus*) and wild boar (*Sus scrofa*). Additionally, the increase in herbivores has led to a dense wolf (*Canis lupus*) population (Blanco, 1998). These events led to the construction of specific crossing structures, and the modification of culverts along the highway, to reduce wildlife mortality from traffic.

2.2. Monitoring of crossing structures

A total of 39 crossing structures were studied along 57 km of road between Camarzana de Tera and Requejo (Zamora province). We also included four structures specifically designed or modified for wildlife (two overpasses, one underpass and one wildlife-adapted culvert), all of which were in Orense province (177.5–216.3 km). Our

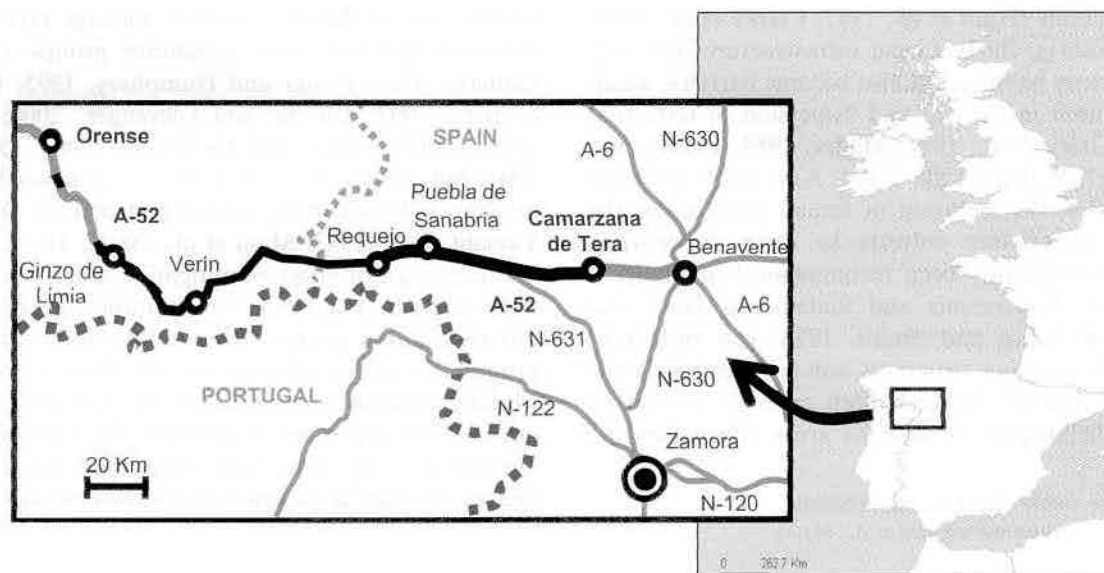


Fig. 1. Location of the study area and the monitored sections (in bold) of the A-52 motorway in the Iberian Peninsula.

selection of crossing structures aimed to include all those specifically designed for wildlife and to obtain a representative sample of the remainder, including all design types (see Table 1 and Fig. 2 for design details). The main difference between over- and underpasses designed for wildlife and other types is that the former have no vehicular access. Fauna-adapted culverts are characterised by their flat bases and enlarged cross-sectional areas.

Crossing structure use was monitored late March and early June 2001. Marble dust, a scentless substance, which is an excellent substrate both for recording high-quality impressions and for its persistence, was used to obtain tracks and signs (Yanes et al., 1995). A band of dust 1 m wide and 3–10 mm thick was spread across and perpendicular to the line of each passage in its midpoint. Animal tracks present in each passage were identified and recorded daily before being smoothed flat. Each passage was

monitored until ten days worth of reliable records were obtained, rejecting days when meteorological conditions prevented the reliable identification of tracks. Although some species may not be detected, a 10-day monitoring period has been found to be sufficient to detect a significant portion of the fauna using a passage (Malo et al., 2006).

Tracks and signs were identified following Strachan (1995), Bang and Dahlstrom (1997) and Blanco (1998). Data for those groups where identification to species was impossible were pooled. Consequently, apart from the individually identified species, the following groups were recorded: lizards (*Lacerta* spp., *Psammodromus* spp. and *Podarcis* spp.), ophidians (snakes), small mammals (mice, voles and shrews), water voles (*Arvicola sapidus* and *Arvicola terrestris*), rats (*Rattus rattus* and *Rattus norvegicus*), lagomorphs (*Oryctolagus cuniculus* and *Lepus granatensis*), small mustelids (*Mustela nivalis* and *Mustela*

Table 1

Characteristics of the monitored crossing structures: type, number, dimensions and function

	n	Dimensions (m)			Main function
		Width	Height	Length	
Circular culverts	14 (6)	Ø 1.80		35–62	Drainage
Adapted culverts	7 (0)	1.7–4	1.7–3	36–45	Drainage, adapted for wildlife
Open span underpasses	7 (2)	4–9	4–6	32–46 ^a	Rural tracks and livestock paths
Wildlife underpasses	5 (4)	14–20	5–8	30–32.5 ^b	Wildlife, closed to vehicles
Overpasses	6 (2)	7–8	—	58–65	Rural tracks
Wildlife overpasses	4 (2)	14.7–20	—	60–62	Wildlife, closed to vehicles

The number of them used in a previous study (Mata et al., 2005) is included in brackets.

^aOne case of 72 m.

^bOne case of 96 m.

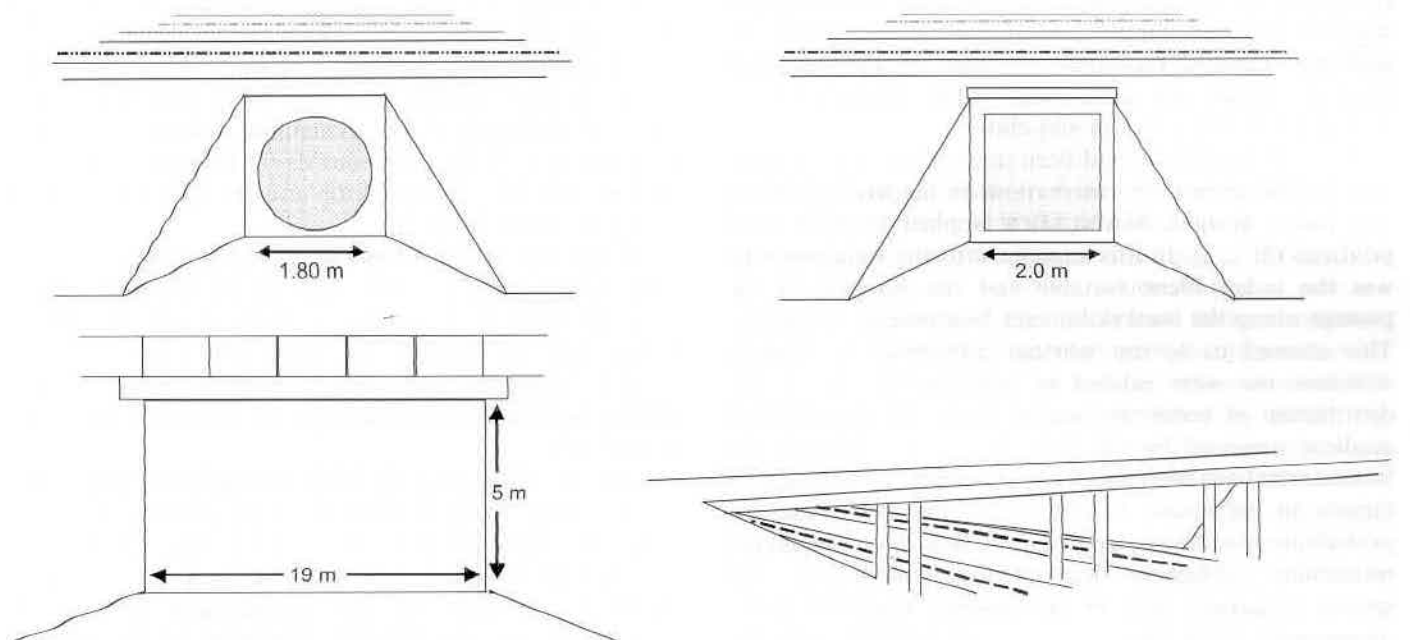


Fig. 2. Basic design of the different crossing structure types under study: circular culvert (top left), adapted culvert (top right), underpass (bottom left) and overpass (bottom right). Non-wildlife engineered passages of the two last types differ from wildlife passages in the absence of vehicular access to the latter.

erminea), large mustelids (*Martes foina* and *Martes martes*), cats (*Felis catus* and *Felis silvestris*), and canids (*Canis familiaris* and *C. lupus*).

2.3. Data analysis

The number of days on which tracks of a species (or faunal group) were registered has been considered as the basic unit for analysis, in order to minimize the problems of pseudoreplication deriving from counting several records of a particular species from the same passage on the same day.

The use of each crossing structure type by each species/group employed the following use index (UI):

$$UI = (n_{ij}/cs_j)/(N_i/CS),$$

where n_{ij} is the number of observations (animal crossing-days) of a species/group i in one type of crossing structure j , cs_j is the number of structures of type j , N_i is the number of observation days of a species (or group) i in the total structures, and CS is the total number of structures studied. This index allows the frequency of tracks observed in each passage type and that expected taking into account all passage types to be compared without bias due to sample size. Therefore, it is a selection index for animal use of a passage-type independent of use frequency, with a UI value of 1 when actual use is equal to that expected. The UI was compared between passage types using a Kruskal–Wallis test since the assumptions for an ANOVA were not met (Quinn and Keough, 2002).

General use patterns of the crossing structures were analyzed using multidimensional scaling (STATISTICA, 1998) to obtain an ordination of the crossing structures as a function of the frequencies of use by each species (animal crossing-days). Euclidean distance was used as the index of similarity between samples and ordination was carried out with the standard procedure provided by the statistical program. Given the adjustment values (stress < 0.1), a three-dimensional solution was chosen.

Once the ordination had been undertaken, the presence of separable groups of observations in the MDS diagram was tested using a MANCOVA applied to their axial positions (x , y , z). In this analysis, crossing structure type was the independent variable and the position of the passage along the road (kilometer location) the covariate. This allowed us to test whether differences in crossing structure use were related to passage type or to the distribution of vertebrate species along the geographical gradient traversed by the road. Differences between the factor levels have been tested a posteriori by testing pairs of factors in successive MANCOVAs and correcting the probability obtained by applying the Bonferroni sequential probability correction (Rice, 1989). Additionally, the species explaining most of the variance along the three-dimensional MDS axes (x , y , z) were detected using the Spearman rank correlations, selecting those species with values of $p < 0.05$ after applying the Bonferroni correction.

Finally, patterns of crossing structure use were analyzed by species. The crossing frequency of the different crossing structures by each species was analyzed using the Kruskal–Wallis test, given the non-normal distribution of the data. This analysis was carried out for species with at least ten observations and, since multiple comparisons were involved, Bonferroni corrections were applied to correct significance values.

The influence of structural variables on use patterns was also analyzed with the openness index (OI), which embraces cross-sectional area and passage length (Olbrich, 1984):

$$OI = \text{width} \times \text{height/length}.$$

As it is not possible to give a meaningful height value to overpasses (see, however, Clevenger and Waltho, 2005), the OI was adapted for these cases as $(OI = \text{width/length})$. Therefore, data from both wildlife- and other underpasses were analyzed separately from those from overpasses.

The relationship between the OI of the structures and their use (N of track-days) by different species was measured through the Spearman rank correlations due to data lacking normality. In this analysis, the Bonferroni correction was also applied.

3. Results

A total of 424 species track-days of animals was recorded, equivalent to 0.99 species passing per crossing structure per day. A total of 17 different faunal species and species-groups used the road crossing structures (Table 2), with the red fox (*Vulpes vulpes*) being the most common species recorded (0.27 records/day). Other canids (dogs and wolves) and Eurasian badgers (*Meles meles*) also had a high crossing frequency (0.19 detections/day and 0.15 detections/day, respectively). Other species using the crossing structures with medium frequency were cats (0.10 detections/day), lagomorphs and small mammals (each 0.07 detections/day). The remaining species had lower frequencies (< 0.04 detections/day). Records related to human activities showed little use by humans of most structure types (Table 2).

All crossing structure types were used by wildlife, though there were significant differences in the levels of use of different types (Fig. 3; $H_{5,43} = 12.07$; $p = 0.034$). Overpasses had the highest use index ($UI = 1.5$), whereas circular culverts and those culverts modified for wildlife were used less than expected ($UI = 0.6$ and 0.96 , respectively).

Positions obtained in the MDS allowed us to distinguish several groups in the distribution of the crossing structures (Fig. 3). This grouping was significantly determined by crossing structure type (MANCOVA test, $p < 0.01$; Table 3). In contrast, the geographical location of these (covariate km) was not significant to the suite of species using the crossing structures throughout the length of the section of motorway. A posteriori analysis

Table 2

Mean number of day-detections per crossing structure for species and species groups detected throughout the monitoring period (10 days for each crossing structure)

	Circular culverts <i>n</i> = 14	Adapted culverts <i>n</i> = 7	Open span underpasses <i>n</i> = 7	Wildlife underpasses <i>n</i> = 5	Overpasses <i>n</i> = 6	Wildlife overpasses <i>n</i> = 4	Mean
<i>Species</i>							
Western hedgehog (<i>Erinaceus europaeus</i>)	—	—	—	—	0.17	—	0.02
Polecat (<i>Mustela putorius</i>)	—	0.14	—	0.20	—	—	0.05
European badger (<i>Meles meles</i>)	1.00	3.14	2.71	1.60	0.17	—	1.49
Small-spotted genet (<i>Genetta genetta</i>)	—	—	0.29	—	—	—	0.05
Red fox (<i>Vulpes vulpes</i>)	1.07	2.43	4.71	3.60	3.00	3.50	2.67
Roe deer (<i>Capreolus capreolus</i>)	—	—	—	0.40	—	—	0.05
Red deer (<i>Cervus elaphus</i>)	—	—	—	0.40	0.17	—	0.07
Wild boar (<i>Sus scrofa</i>)	—	—	—	0.60	0.67	2.25	0.37
<i>Groups of species</i>							
Lizards	0.36	—	0.14	—	—	—	0.14
Ophidians	—	—	—	—	—	0.25	0.02
Small mammals	1.57	0.14	—	—	1.00	0.25	0.70
Rats	0.07	0.43	0.14	—	1.33	—	0.30
Lagomorphs	—	—	0.43	1.20	1.83	3.00	0.74
Small mustelids	0.29	—	0.14	0.40	0.17	—	0.19
Large mustelids	—	0.14	—	0.20	—	—	0.05
Cats	0.93	0.57	1.29	0.60	1.67	1.00	1.00
Canids (<i>Canis</i> spp.)	0.64	2.43	2.29	1.40	4.50	2.00	1.95
<i>Human activity</i>							
People	0.93	1.43	1.00	1.20	6.33	0.75	0.05
Livestock	0.14	0.86	0.57	0.20	2.50	—	1.49
Vehicles	—	0.86	1.00	—	3.50	0.50	0.05

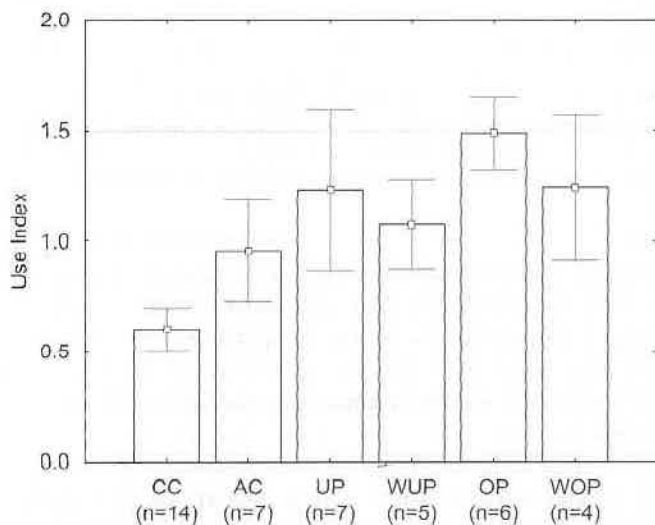


Fig. 3. Use index (mean + SE) recorded in the six types of crossing structures differentiated in the study. CC: circular culverts, AC: adapted culverts, U: underpasses, WU, wildlife underpasses, O: overpasses, and WOP: wildlife overpasses. *n*: number of monitored structures.

shows that the difference between the diverse crossing structure types was only significant after probability correction for wildlife overpasses vs. circular culverts (Table 4).

Table 3

Results of the MANCOVA on the effect of crossing structure type with the location (km post) as covariate on the position of the structures on axes *x*, *y*, *z* of the multidimensional scaling

Factors	Wilk's λ	<i>F</i>	d.f. 1	d.f. 2	<i>p</i>
Structure type	0.431	2.239	15	94	0.0098
Covariate (km post)	0.843	2.107	3	34	0.1175

This differential location of the crossing structure types in the Multidimensional Scaling can be interpreted through the correlations between species and axes of the analysis (Table 5). Thus, the negative values on the *x* axis are significantly associated (the Spearman rank correlation, $p < 0.05$ after Bonferroni correction) with the presence of canids, Eurasian badger and red fox where non-wildlife engineered underpasses occurred, whereas along the positive part of the axis where circular culverts were located there were no significant associations with any of the species. *y*-Axis is positively correlated with wild boar, lagomorphs and red fox, with overpasses and wildlife overpasses located along the positive part of the axis. In contrast, the two types of culverts are located along the negative values and give no significant relationship. Finally, the positive values along the *z*-axis are associated with the presence of Eurasian badger, and adapted culverts

Table 4

Results of a posteriori comparisons of differences between passage types and the fauna crossing them, simplified by their position on the x , y , z axes of the multidimensional scaling

		Wildlife overpasses	Overpasses	Wildlife underpasses	Open span underpasses	Adapted culverts
Circular culverts	Wilk's λ	0.254	0.526	0.636	0.618	0.503
	p	0.0004	0.0192	0.0878	0.0476	0.0192
Adapted culverts	Wilk's λ	0.401	0.517	0.683	0.894	
	p	0.1180	0.1344	0.4162	0.7863	
Open span underpasses	Wilk's λ	0.255	0.804	0.803		
	p	0.0327	0.5258	0.6508		
Wildlife underpasses	Wilk's λ	0.701	0.594			
	p	0.6649	0.3396			
Overpasses	Wilk's λ	0.577				
	p	0.3935				

The Wilk's λ value is given (upper line) and its probability value (lower line) corresponding to the MANCOVAs for pair comparisons. Significant results ($p < 0.05$) after Bonferroni sequential probability correction are shown in bold.

Table 5

Mean position along the x -, y - and z -axis of the multidimensional scaling of crossing structure types

	x	y	z
Circular culverts	0.380	-0.284	-0.044
Adapted culverts	-0.253	-0.252	0.172
Open span underpasses	-0.471	0.126	0.062
Wildlife underpasses	-0.083	0.155	0.164
Overpasses	-0.062	0.250	-0.376
Wildlife overpasses	0.135	0.644	0.104

and wildlife underpasses principally appear along this part of the axis. However, the negative part of the axis, along which overpasses are located, is not significantly related to any particular species.

The analysis by species of the use of the crossing structures has only demonstrated a differential use by wild boar and lagomorphs (Kruskal–Wallis test, $p < 0.05$ after Bonferroni correction; Table 6 and Fig. 4). Wild boar and lagomorphs mainly crossed on wildlife overpasses (0.22 and 0.30 detections/day, respectively), overpasses (0.07 and 0.18 detections/day) and wildlife underpasses (0.06 and 0.12 detections/day). None of them were recorded in the two culvert types, and only lagomorphs made some use of underpasses (0.04 detections/day).

Regarding the dimensions of underpasses, only small mammals showed a significant negative relationship with the OI ($r_s = -0.506$, $p < 0.05$ after Bonferroni correction). Accordingly, small mammals (shrews, mice and voles), preferentially used circular culverts to cross ($UI = 2.25$). Passage use by lagomorphs, fox, wild boar, roe deer and canids showed positive correlations with OI (Table 7) although they became non-significant after the application of the Bonferroni correction. Red foxes crossed principally using wider passages, showing a similar level of use in all of them (mean $\pm$ standard deviation: 1.39 ± 0.28). Roe deer, with two records, exclusively used wildlife underpasses.

Table 6

Results of the comparison of different crossing structure use by species along the A-52

Species	N	$H_{43, s}$	p
European badger	64	8.650	0.1246
Red fox	115	11.577	0.0638
Wild boar	16	19.350	0.0017
<i>Species groupings</i>			
Small mammals	30	14.700	0.0121
Rats	13	4.383	0.6030
Lagomorphs	32	17.009	0.0043
Cats	43	0.915	0.9465
Canids (<i>Canis</i> spp.)	84	13.422	0.0264

Heterogeneity (H) values and Kruskal–Wallis probability (p) values are shown for species with at least ten detections. N = number of records of each species for all crossing structures studied. Significant results ($p < 0.05$) after Bonferroni sequential probability correction are shown in bold.

Similarly, other canids also had a preference for using wide crossing structures, but they also had a higher frequency of using multipurpose passages, independently of their position relative to the road. No significant correlations with OI were found for the species using passages over the road (Table 7).

No significant relationships were found for the remaining species but it is worth noting the case of the Eurasian badger. This species almost exclusively used underpasses, and mainly wildlife adapted culverts ($UI = 2.24$).

4. Discussion

Our results highlight the broad spectrum of species which use the highway's crossing structures and the relevance of structure adaptation for wildlife. With the exception of amphibians, garden dormice (*Elyomys quercinus*), water voles (*A. sapidus* and *A. terrestris*), moles (*Talpa occidentalis*), terrapins (*Emys orbicularis*), red squirrels

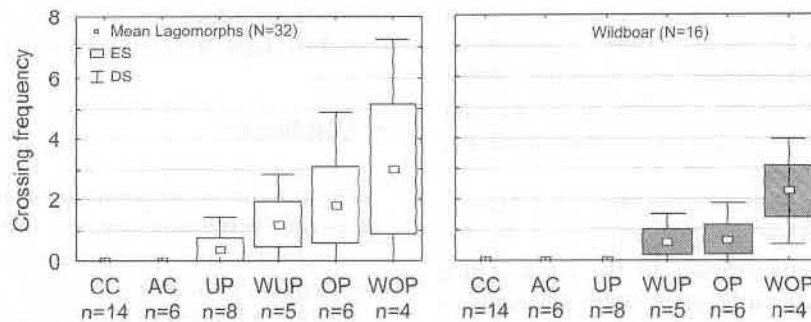


Fig. 4. Crossing frequency (N observations/10 days) of lagomorphs and wild boar in the different structure types (mean, SE, SD). CC: circular culverts, AC: adapted culverts, UP: underpasses, WUP: wildlife underpasses, OP: overpasses, WOP: wildlife-overpasses, n : number of crossing structures.

Table 7

Results of the Spearman correlations between *Openness index* of each passage and crossing frequency by species along the A-52

Species	Underpasses		Overpasses	
	r_s	p	r_s	p
Western hedgehog	—	—	−0.121	0.7402
Polecat	0.214	0.2327	—	—
European badger	0.208	0.2455	−0.121	0.7402
Small-spotted genet	0.186	0.3005	—	—
Red fox	0.467	0.0062	0.475	0.1651
Roe deer	0.387	0.0260	—	—
Red deer	0.279	0.1162	−0.121	0.7402
Wild boar	0.401	0.0207	0.482	0.1586
<i>Species groupings</i>				
Lizards	−0.056	0.7578	—	—
Ophidians	—	—	0.241	0.5023
Small mammals	−0.506	0.0026	0.000	1.0000
Rats	0.109	0.5472	−0.422	0.2247
Lagomorphs	0.472	0.0056	0.247	0.4920
Small mustelids	0.044	0.8094	−0.121	0.7402
Large mustelids	0.227	0.2040	—	—
Cats	0.073	0.6873	0.050	0.8920
Canids (<i>Canis</i> spp.)	0.380	0.0292	−0.465	0.1759

Openness indices for under- and over-passages are computed separately as explained in the text. The Spearman ρ value (r_s) is given together with its probability value (p) and the significant results ($p < 0.05$) after Bonferroni sequential probability correction are shown in bold.

(*Sciurus vulgaris*) and Eurasian otters (*Lutra lutra*), tracks showing passage of all vertebrate groups potentially present in the study area were recorded (Palomo and Gisbert, 2002). The absence of data for amphibians, water voles, terrapins and Eurasian otter is biased by the survey method used since it does not monitor movements through flooded passages. Most of these species have been recorded during other studies in underpasses of different types (Rosell and Velasco, 1999; Mata et al., 2005; C. Mata, unpublished data). Cold weather conditions during the present study could be responsible for the absence of records of amphibians and garden dormice in comparison with other data gathered in the area (Mata et al., 2005).

All crossing structure types were used by wildlife, both those specifically designed for that purpose and those

intended for other uses. The use indices demonstrated a relative use above that expected for all crossing structure types, except for culverts. Despite this, the potential importance of culverts, as the commonest crossing structures of roads and railways, should not be forgotten (e.g. a 91 km stretch of the A-52 under study has 106 culverts, C. Mata, unpublished data). Such structures do not represent an additional construction cost given their structural function and they allow the movement of small and medium-sized mammals from one side to the other (see also Huijser et al., 1999; Clevenger et al., 2001; Dodd et al., 2004). Moreover, culverts can be adapted at little extra cost to favour wildlife, as shown by the intensive use of the adapted culverts (mainly 2 m $\times$ 2 m boxes) by the Eurasian badger: this is noteworthy since badgers suffer high mortality from traffic (Clarke et al., 1998).

Over- and underpasses had high use indices and high crossing frequencies for a wide range of species, including the red fox and lagomorphs. In general terms, multipurpose crossing structures are clearly important for enhancing the permeability/crossability of roads and railways (Camby and Maizeret, 1987; Yanes et al., 1995; Ng et al., 2004). However, it is important to note that crossing structures of mixed human–faunal use have been demonstrated to be effective for fauna where there is little human activity (Rodríguez et al., 1996; Clevenger and Waltho, 2005). In our case, the maximum value for human use (that of overpasses) equals 1.2 pedestrians, vehicles or herds a day and the use of multipurpose crossing structures by wildlife could be radically different in areas with a higher intensity of human use.

The crossing structure design is shown to be a key factor for its use or selection by some species, thus determining its effectiveness for barrier effect mitigation (Ballon, 1985; Forman et al., 2003). This could justify the investment on wildlife crossing structures provided that target species rely mostly or exclusively on them. In our study, the difference in wildlife use of the various crossing structure types was greatest between culverts and wildlife overpasses, but differences were notable in the majority of comparisons. In simple terms, the width of the crossing structure was an important feature in their selection by wildlife, as revealed by lagomorphs, wild boar, deer, red foxes and other canids.

The relevance of passage width has been highlighted in other studies (Foster and Humphrey, 1995; Yanes et al., 1995; Rosell et al., 1997; Veenbaas and Brandjes, 1999), but our results also show that the position of the passage with reference to the road and its design for wildlife are key factors.

The use of crossing structures by rabbits and Iberian hares was related to passage design, width and by the position of the passage with respect to the road. Thus, overpasses were more frequently used than underpasses and no rabbit or hare crossed through a culvert, in agreement with the results presented by Rosell and Velasco (1999). In addition, the frequency of use of passages built for wildlife was between 1.5 and 3 times greater than that of non-wildlife engineered passages in the same position relative to the road. This tendency was already present in the previously published results (Mata et al., 2005), but differences were not as clear-cut.

The results obtained for wild boar and for ungulates also support the importance of building large wildlife passages. It is worth noting that the crossing frequencies of wild boar were higher in this study than in any other to date (Rodríguez et al., 1996; Rosell et al., 1997; Mata et al., 2005). The width and position of the crossing structures in relation to the road determined their selection by wild boar in a similar fashion to that discussed above for lagomorphs (see also Vassant et al., 1993). The use of wildlife overpasses by wild boar was three times greater than that for overpasses with mixed use, and the boars also used wildlife underpasses but not open span underpasses. These results contrast with those of Rosell and Velasco (1999) for roads in Catalonia (NE Spain), where the species has been observed using wide underpasses of mixed use. Although with fewer observations, data available for red and roe deer point in the same direction. Roe deer used only wildlife underpasses and had not previously been detected crossing this motorway. During this study, red deer crossed mainly through wildlife underpasses, and previous data showed the relevance of wildlife overpasses for them (Mata et al., 2005).

The results for the remaining species also showed a tendency to use wider crossing structures, although this finding was less clear-cut. The high frequency of use by red foxes of these passage types was notable, with an intensive use of both over- and underpasses (see also Trewhella and Harris, 1990; Rodríguez et al., 1997). Tendencies shown by other canids are difficult to interpret due to the presence of domestic and feral animals as well as wild individuals. Footprints do not allow the differentiation of wolves but photographic monitoring of passages carried out in the area after this study (Mata et al., 2005; C. Mata, unpublished data) showed them using mainly wildlife overpasses. In fact, the frequency of wildlife overpass use is 30–40 times higher than that of non-wildlife engineered over- and underpasses. However, since photographic systems show that domestic and feral dogs are at least six times more frequent on passages of the area than wolves,

the bulk of results from other canids presumably correspond to dogs crossing through most structure types.

5. Conclusions

Three main conclusions arise from our study: (i) most terrestrial vertebrate species use the different types of crossing structures present on motorways, with more or less frequency, and their correct design and maintenance could thus mitigate the barrier effect. However, the width and specific design of passages play a key role for determining their use by at least some species of concern, such as ungulates. Therefore, (ii) special attention should be paid to culvert adaptation and/or enlargement due to their significance for certain species (e.g. badgers) and relatively low cost, and (iii) the high construction costs of specific wildlife passages is justified in environmental terms. Moreover, it is important to note that wildlife passages can be kept free of human activity while the long-term effectiveness of mixed-use over- and underpasses can be threatened by an intensification of their use by humans.

Acknowledgments

This study was funded by a research agreement between the UAM, the Ministerio de Medio Ambiente and the Centro de Estudios y Experimentación (CEDEX), and by an FPI grant from the Comunidad de Madrid to Cristina Mata. Thanks are due to Juan Manuel Varela and Javier Cachón from CEDEX and to Antonio and Teresa from Hostal Las Ventas for their help at different stages of the project. Comments and suggestions by R. Goldingay and one anonymous referee improved the final text.

References

- Ballon, P., 1985. Bilan technique des aménagements réalisés en France pour réduire les impacts des grandes infrastructures linéaires sur les ongulés gibiers. In: 17^e Congrès de l'Union Internationale des Biologistes du Gibier. Brussels, pp. 679–689.
- Bang, P., Dahlstrom, P., 1997. Huellas y señales de los animales de Europa. Ediciones Omega, S.A., Barcelona.
- Beier, P., Loe, S., 1992. A checklist for evaluating impacts to wildlife movement corridors. *Wildlife Society Bulletin* 20, 434–440.
- Bekker, H.G.J., Canters, K., 1997. The continuing story of badgers and their tunnels. In: Canters, K. (Ed.), *Habitat Fragmentation and Infrastructure*. Maastricht, The Hague, Ministry of Transport, Public Works and Water Management, Delft, The Netherlands, pp. 344–353.
- Blanco, J.C., 1998. Mamíferos de España.
- Camby, A., Maizeret, C., 1987. Perméabilité des routes et autoroutes vis-à-vis des mammifères carnivores: exemple des études menées dans les landes de Gascogne. In: Bernard, J.M., Lansari, M., Tille, M. (Eds.), *Routes Et Faune Sauvage*. SETRA, Strasbourg, pp. 173–181.
- Clarke, G.P., White, P.C.L., Harris, S., 1998. Effects of roads on badger *Meles meles* populations in south-west England. *Biological Conservation* 86, 117–124.
- Clergeau, P., 1993. Utilisation des concepts de l'écologie du paysage pour l'élaboration d'un nouveau type de passage à faune. *Revue Gibier et Faune Sauvage* 10, 47–57.

- Clevenger, A.P., Waltho, N., 2005. Performance indices to identify attributes of highway crossing structures facilitating movement of large mammals. *Biological Conservation* 121, 453–464.
- Clevenger, A.P., Chruszcz, B., Gunson, K., 2001. Drainage culverts as habitat linkages and factors affecting passage by mammals. *Journal of Applied Ecology* 38, 1340–1349.
- De Santo, R.S., Smith, D.G., 1993. Environmental auditing: an introduction to issues of habitat fragmentation relative to transportation corridors with special reference to high-speed rail (HSR). *Environmental Management* 17, 111–114.
- Dodd, C.K., Barichivich, W.J., Smith, L.L., 2004. Effectiveness of a barrier wall and culverts in reducing wildlife mortality on a heavily traveled highway in Florida. *Biological Conservation* 118, 619–631.
- Fahrig, L., 2003. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology and Systematics* 34, 487–515.
- Forman, R., Sperling, D., Bissonette, J.A., Clevenger, A.P., Cutshall, C.D., Dale, V.H., Fahrig, L., France, R., Goldman, C.R., Heanue, K., 2003. *Road Ecology: Science and Solutions*. Island Press, Washington, DC, USA. 481pp.
- Foster, M.L., Humphrey, S.R., 1995. Use of highway underpasses by Florida panthers and other wildlife. *Wildlife Society Bulletin* 23, 95–100.
- Gloyne, C.C., Clevenger, A.P., 2001. Cougar *Puma concolor* use of wildlife crossing structures on the Trans-Canada highway in Banff national park, Alberta. *Wildlife Biology* 7, 117–124.
- Huijser, M.P., Van Der Grift, E.A., Bekker, G.J., 1999. Habitat fragmentation and infrastructure: a review of the proceedings of a European and North American conference. *Lutra* 41, 43–54.
- Hunt, A., Dickens, H.J., Whelan, R.J., 1987. Movements of mammals through tunnels under railway lines. *Australian Zoologist* 24, 89–93.
- Lodé, T., 2000. Effect of a motorway on mortality and isolation of wildlife populations. *AMBIO* 29, 163–166.
- Mader, H.J., 1984. Animal habitat isolation by roads and agricultural fields. *Biological Conservation* 29, 81–96.
- Malo, J.E., Hervás, I., Mata, C., Herranz, J., Suárez, F., 2006. How many days to monitor a wildlife passage? Species detection patterns and the estimation of vertebrate fauna using crossing structures in a motorway. In: Leroy, C., Garret, P., McDermott, K.P. (Eds.), *On the Road to Stewardship Proceedings of the International Conference on Ecology & Transportation (ICOET-05)*. Center for Transportation and the Environment, NC State University, Raleigh, NC, USA, pp. 406–413.
- Mata, C., Hervás, I., Herranz, J., Suárez, F., Malo, J.E., 2005. Complementary use by vertebrates of crossing structures along a fenced Spanish motorway. *Biological Conservation* 124, 397–405.
- Ng, S.J., Dole, J.W., Sauvajot, R.M., Riley, S.P.D., Valone, T.J., 2004. Use of highway undercrossings by wildlife in Southern California. *Biological Conservation* 115, 499–507.
- Olbrich, P., 1984. Study of the effectiveness of game warning reflectors and the suitability of game passages. *Zeitschrift für Jagdwissenschaft* 30, 101–116.
- Oxley, D.J., Fenton, M.B., Carmody, G.R., 1974. The effects of roads on populations of small mammals. *Journal of Applied Ecology* 11, 51–59.
- Palomo, L.J., Gisbert, J., 2002. *Atlas de los mamíferos terrestres de España*. Dirección General de Conservación de la Naturaleza-SECEM-SECEMU, Madrid.
- Puky, M., Vogel, Z., 2003. Amphibian Mitigation Measures on Hungarian Roads: Design, Efficiency, Problems and Possible Improvement, Need for a Co-ordinated European Environmental Education Strategy. *Habitat Fragmentation due to Transportation Infrastructure*. IENE, Brussels.
- Putman, R.J., 1997. Deer and road traffic accidents: options for management. *Journal of Environmental Management* 51, 43–57.
- Quinn, G.P., Keough, M.J., 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge, UK.
- Reed, D.F., 1981. Mule deer behavior at a highway underpass exit. *Journal of Wildlife Management* 45, 542–543.
- Rice, W.R., 1989. Analyzing tables of statistical tests. *Evolution* 43, 223–225.
- Rodriguez, A., Crema, G., Delibes, M., 1996. Use of non-wildlife passages across a high speed railway by terrestrial vertebrates. *Journal of Applied Ecology* 33, 1527–1540.
- Rodriguez, A., Crema, G., Delibes, M., 1997. Factors affecting crossing of red foxes and wildcats through non-wildlife passages across a high-speed railway. *Ecography* 20, 287–294.
- Rosell, C., Velasco, J.M., 1999. Manual de prevenció i correcció dels impactes de les infraestructures viàries sobre la fauna. Departament de Medi Ambient, Generalitat de Catalunya, Barcelona.
- Rosell, C., Parpal, J., Campeny, R., Jové, S., Pasquina, A., Velasco, J.M., 1997. Mitigation of barrier effect of linear infrastructures to wildlife. In: Canters, K. (Ed.), *Habitat Fragmentation and Infrastructure*. Maastricht, The Hague, Ministry of Transport, Public Works and Water Management, Delft, The Netherlands, pp. 367–372.
- Rosell, C., Alvarez, G., Cahil, S., Campeny, R., Rodriguez, A., Seiler, A., 2003. COST 341. La fragmentación del hábitat en relación con las infraestructuras de transporte en España. O.A. Parques Nacionales, Ministerio de Medio Ambiente, Madrid.
- Saunders, D.A., Hobbs, R.J. (Eds.), 1991. *Nature Conservation 2: The Role of the Corridors*. Surrey, Beatty, Chipping Norton, New South Wales, Australia.
- Singer, F.J., Doherty, J.L., 1985. Managing mountain goats at a highway crossing. *Wildlife Society Bulletin* 13, 469–477.
- StatSoft, Inc. 1998. STATISTICA, (data analysis software) version 6.0. Tulsa, OK, USA.
- Strachan, R., 1995. *Mammal Detective*. Whittet Books Limited, London.
- Taylor, B.D.T., Goldingay, R.L., 2003. Cutting the carnage: wildlife usage of road culverts in north-eastern new south Wales. *Wildlife Research* 30, 529–537.
- Trewhella, W.J., Harris, S., 1990. The effect of railway lines on urban fox (*Vulpes vulpes*) numbers and dispersal movements. *Journal of Zoology* 221, 321–326.
- Van Bohemen, H.D., 1998. Habitat fragmentation, infrastructure and ecological engineering. *Ecological Engineering* 11, 199–207.
- Vassant, J., Brandt, S., Jullien, J.M., 1993. Effects of the wildlife passage over motorway A5 on the populations of red deer and wild boar in the Massif d'Arc-en-Barrois. part 2: case of the population of wild boar. *Bulletin Mensuel de l'Office National de la Chasse* 184, 24–33.
- Veenbaas, G., Brandjes, G.J., 1999. The use of fauna passages along waterways under motorways. ICOWET III. The International Conference on Wildlife Ecology and Transportation, Florida, 7pp.
- Walker, R.S., Novaro, A.J., Branch, L.C., 2003. Effects of patch attributes, barriers, and distance between patches on the distribution of a rock-dwelling rodent (*Lagidium viscacia*). *Landscape Ecology* 18, 187–194.
- Yanes, M., Velasco, J.M., Suárez, F., 1995. Permeability of roads and railways to vertebrates: the importance of culverts. *Biological Conservation* 71, 217–222.

Feasibility of water purification technology in rural areas of developing countries

Dana M. Johnson^{a,*}, David R. Hokanson^b, Qiong Zhang^c,
Kevin D. Czupinski^a, Jinxian Tang^a

^a*School of Business and Economics, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA*

^b*Trussell Technologies Inc, Pasadena, CA, USA*

^c*Department of Civil & Environmental Engineering, College of Engineering, Sustainable Futures Institute, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA*

Received 19 May 2006; received in revised form 2 February 2007; accepted 10 March 2007

Available online 24 April 2007

Abstract

Water scarcity is threatening social and economic growth in rural areas of developing countries. There are potential markets for water purification technologies in these regions. The main focus of this article is to evaluate the social, economic and political feasibilities of providing water purification technologies to rural areas of developing countries. The findings of this research can serve as the basis for private investors interested in entering this market. Four representative regions were selected for the study. Economic, demographic, and environmental variables of each region were collected and analyzed along with domestic markets and political information. Rural areas of the developing world are populated with poor people unable to fulfill the basic needs for clean water and sanitation. These people represent an important group of potential users. Due to economic, social, and political risks in these areas, it is difficult to build a strong case for any business or organization focusing on immediate returns on capital investment. A plausible business strategy would be to approach the water purification market as a corporate responsibility and social investing in the short term. This would allow an organization to be well positioned once the economic ability of individuals, governments, and donor agencies are better aligned.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Water purification technology; Rural areas; Developing country; Economic feasibility; Social factors

1. Introduction

There are nearly 50 countries in the world experiencing water stress or water scarcity (FAO, 2005). An area under water stress is defined as having annual water supplies below 1700 m³ (450,000 gal) per person. A country facing water scarcity is defined as having annual water supplies below 1000 m³ (260,000 gal) per person (UNEP, 2005a, b). By the year 2025, it is anticipated another 70 countries will be added to the list. Appendix A contains a list of countries estimated to be water scarce in 2025. As part of its Millennium Development Goals, the United Nations has

set aggressive targets to reduce by half the proportion of people without access to safe drinking water by 2015.

Given the importance of clean water to human health and economic growth, and the serious scarcity of water in developing countries, exploring the water purification market could be a potential business opportunity for private investors or corporations. This paper evaluates the economic feasibility of providing water purification technologies in rural areas of developing countries, along with practical considerations of social and political factors that may affect market entrance strategy. The findings of this research provide interested organizations with key information about the feasibility of entering the market. Four cases were analyzed, including Sichuan, China; Madhya Pradesh, India; Honduras, Central America; and Mozambique, Africa. Two cases representing extremes are presented in more detail. The two extreme cases included

*Corresponding author. Tel.: +1 906 487 2803; fax: +1 906 487 2944.

E-mail addresses: dana@mtu.edu (D.M. Johnson), david.hokanson@trusselltech.com (D.R. Hokanson), qiong@mtu.edu (Q. Zhang).

Sichuan, China, as the most feasible choice and Mozambique, Africa, as the least favorable alternative.

This paper is organized as follows. Section 2 discusses data sources used for this study. Section 3 outlines the methodology used for selecting water purification technologies. Section 4 describes the methodology for selecting representative regions for this study. Section 5 discusses case studies for Sichuan Province, China, and Mozambique, Africa. Section 6 concludes the paper by summarizing the findings and giving suggestions to corporations that are interested in entering the water purification technology market in the researched regions. Section 7 offers suggestions for future research.

2. Data sources

Several sources for social, economic, and governmental data were utilized to support the overall findings of this paper, including diverse medium of professional and academic journals, textbooks, special topic books, and databases.¹

Several governmental and non-governmental sources were used to identify and assemble demographic and geographic data. For example, the United Nations (UN) Department of Economic and Social Affairs Population Division was used to obtain both urban and rural population levels, densities, and growth rates. The UN population reports contain both national and regional level data along with sources of local census data, if detail is required at that scale. Energy supply, distribution, and infrastructure data for the targeted regions was gathered from the International Energy Agency (IEA), the Organization for Economic Co-operation and Development (OECD), and the World Bank.

Water resource supply, withdrawal, and use data was gathered primarily from the UN Environment Programme's (UNEP) Global Environment Outlook (GEO) Data Portal. The GEO Data Portal is a collection of data sets used by the UNEP and its partners for various environmental reports and assessments. Databases available at the UNEP include water access statistics from the World Health Organization/UNICEF and water resources/usage data from the United Nations Food and Agriculture Organization. The World Health Organization (WHO) was the source for water treatment standards. Other data sources related to water rates and willingness-to-pay levels include the World Bank. The World's Water biennial reports, individual consulting reports, and technical literature reviews of valuation surveys and methods.

The UNEP GEO Data Portal offers country-specific information on: gross domestic product (GDP) per capita, illiteracy rate, and GDP annual growth rate. Literacy rates are available through United Nations Educational, Scientific, and Cultural Organization (UNESCO). The UN Statistics Division provides household size data for over 90

countries. For countries not included in the UN database, household size data is available through a variety of sources and is country-specific (e.g. Census of India). The World Bank Group (2005a) has a qualitative ranking of each country's governance in the world. The following six indicators aggregated illustrate a country's governance proficiency: Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption.

Numerous data sources were utilized for the business case studies to offer a broad perspective of the social, economic, and governmental viewpoints. The Central Intelligence Agency's (US CIA, 2005) World Factbook was accessed multiple times for the economic information. Region-specific resources used frequently included websites of the Madhya Pradesh, India government and the Bureau of Statistics of Sichuan, China. Country-specific resources include websites of the National Bureau of Statistics of China and Mozambique's National Institute of Statistics.

3. Methodology for selecting water purification technologies

For water purification technologies, the disinfection contribution, primary use, affordability, and appropriateness of each technology were analyzed as a basis for selection. Four criteria established for technologies were: (1) disinfection capability to remove or inactivate microbiological contaminants for safe drinking water; (2) primarily applied to domestic use; (3) affordable to the user; and (4) appropriate for cultural and social conditions.

Based on the above selection criteria and the technology information, the most appropriate technologies were:

- Hand pumps.
- Media filtration (sand filtration).
- Reverse osmosis (RO).
- Chemical disinfection (chlorine bleach).
- Ultraviolet radiation (UV).
- Flocculation/coagulation plus sedimentation (or filtration).

These represent the least costly alternatives, and in some cases, require no technology or energy/power.

3.1. Specification of community size for evaluation of water purification systems

For the selected technologies, a community with a population of 1000 people was selected as a representative size based on the fact that a reasonable estimate of the common population of the communities and neighborhoods in the selected countries or regions was 500–800 people.

¹Some databases were available at various websites.

Table 1
Key variable data for the four regions

Variables	Honduras	Mozambique	Madhya Pradesh, India	Sichuan, China
GDP per capita (\$ US)	1026 ⁽²⁰⁰⁴⁾	233 ⁽²⁰⁰⁴⁾	603 ^(2004a)	697 ⁽²⁰⁰²⁾
Gross domestic product—annual growth rate (%)	4.79	1.6	4.33	10.15
Political stability (% ranking)	38.4	63.8	22.2 ^a	51.4 ^a
Government effectiveness (% ranking)	27.3	42.8	54.1 ^a	63.4 ^a
Total population (thousands)	7257	19495	60385	86730
Percentage rural (%)	53.6	62	73.33	82
Literacy rate (% population over age 15)	76.2	53.5	64.11	86.45
Average household size (# persons)	5.4	4.1	5 ^a	4 ^a
Water use per capita (m ³ /year)	133	36	635 ^a	80.55 ^b
% of urban population with access to improved water ^c	99	76	> 85	92 ^a
% of rural population with access to improved water ^c	82	24	100	68 ^a

Sources of data: Ping (2005), IMF (2005), Office of the Registrar General-India, (2001), DDWS (2005), Government of Madhya Pradesh (2003a, b), NBSC (2003), UNESCO (2005), UNEP (2005a, b), UNESA (2004), The World Bank Group (2005a–d), US CIA (2005), BSS (2003), Dev and Mahendra (2004).

^aNational data, not province/state specific at this time.

^bRural water use; urban water use is 153.38.

^cThe definition of “improved” water is based on the assumption that one technology is better for health than none at all, implying that “improved” water and sanitation designate safe water and sanitation.

4. Methodology for selecting representative regions

Four representative regions were initially selected for the study. Economic, demographic, and environmental variables of each region were collected and analyzed along with domestic markets and political information. To ensure a broad representation of markets on a global scale, it was decided to select a province within China, a province within India, a country in Africa, and a country in Central or South America as the four representative regions. Criteria for selecting the four regions are provided below.

4.1. Selection of regions of study

The criteria for selecting representative regions were defined by ten key variables (i.e., GDP was counted as one variable). These included demographic, economic, political, and water resource related variables used to assess a region's suitability for this study. The ten variables and the data associated with each variable are summarized in Table 1 for the four selected regions. Demographic, economic, and water resource related variables are self-explanatory. Political variables include political stability² and government effectiveness.³ Demographic variables include literacy rates to

account for ease of information transfer pertaining to the business and technology of water purification systems.

The political stability and the government effectiveness are thought to be critical attributes for a firm to make a decision regarding the selection of a particular country for investment. The World Bank Group (2005a) conducted a study to rank these two attributes (among others) of nearly every country in the world. Madhya Pradesh and Sichuan's rankings for political stability and government effectiveness are for all of India and China, respectively. Therefore, for these two target regions, the values may not accurately represent the provincial political stability and government effectiveness. However, China and Mozambique represent the two countries most politically stable and with the highest government effectiveness of the four target regions. While India has an effective government, its political stability ranks lower than the fiftieth percentile. Honduras' rankings fall below the fiftieth percentile as well. Diversity with respect to government effectiveness and political stability exists among the four target regions through the spread of rankings.

When choosing the four target regions, literacy rates were considered to account for ease of information transfer pertaining to the business and technology of water purification systems. All four countries have literacy rates at 54% or greater, with Mozambique's literacy rate near 54% (The World Bank Group, 2005d). These literacy rates indicate the potential for educating technology operators and business entrepreneurs.

Western China is much less developed than eastern China. Twelve western provinces account for 75% of the total area and 28% of total population, but only 19.2% of total GDP of the entire country in 2002 (SATC, 2002). Northwest China is dominated by desert, and therefore, water resources are scarce. Northwestern provinces are sometimes afflicted with separatism and terrorism threats

²Political stability is defined as one of six dimensions of the Governance Indicators combining several indicators that measure perceptions of the likelihood that the government in power will be destabilized or overthrown by possibly unconstitutional and/or violent means, including domestic violence and terrorism (The World Bank Group, 2005d).

³Government effectiveness is defined as one of six dimensions of the Governance Indicators combining responses on the quality of public service provision, the quality of the bureaucracy, the competence of civil servants, the independence of the civil service from political pressures, and the credibility of the government's commitment to policies (The World Bank Group, 2005d).

Table 2
Key variables of the southwest provinces of China

Province	The amount of water supply (billion km ³)	Rural population (thousands)	GDP per capita (\$US)	Income per capita (rural) (\$US)
Guangxi	237.259	40,642	615	243
Guizhou	111.757	31,934	373	180
Sichuan	206.616	69,049	679	255
Tibet	424.349	2201	730	177
Yunnan	230.887	34,898	622	194

Source: NBSC (2003).

Table 3
Key variables of the candidate countries in Africa

Country	Renewable water resources (km ³)	Population (thousands)	GDP per capita (\$US)	Political stability (%)
Botswana	2.90	1801	3968	71.4
Cameron	273.00	16,564	702	37.3
Kenya	20.20	32,849	320	21.6
Madagascar	337.00	18,409	209	53.3
Malawi	16.14	12,572	142	54.6
Mozambique	99.00	19,495	222	63.8
Namibia	6.16	2032	2230	60
Nigeria	221.00	130,236	273	10.3
Tanzania	82.00	36,588	201	35.7
Zambia	80.20	11,043	404	44.3
Zimbabwe	14.10	12,963	528	8.6

Sources: UNEP (2005a), UNESA (2004), The World Bank Group (2005d).

from central Asia. Southwest China consists of 5 provinces, as listed in Table 2. Tibet was immediately eliminated as a province for the study due to the harsh environment and prohibitive transportation cost. Sichuan was the best selection due to its large rural population and relatively high income per capita.

North Africa was immediately discounted because of water resource scarcity. Then, any country afflicted with wars and turmoil was disregarded due to the extreme barriers that would place on emerging business and technology. These countries were Sierra Leone, Liberia, Cote d'Ivoire, Central African Republic, Sudan, Eritrea, Ethiopia, Somalia, Uganda, Burundi, Congo-Brazzaville, Congo, and Angola (Guest, 2004). The candidates for this project were eleven central and southern African countries with relatively peaceful environments, as listed in Table 3. Mozambique stood out among the remaining candidates with its high percentile ranking of political stability and relatively large population.

Madhya Pradesh, a province in India with a lower population percentage classified as rural, was selected to further diversify the targeted sites of the project. However, the overall population needed to be large enough to accommodate profitability of a water purification technol-

ogy. Madhya Pradesh has a total population of more than sixty million with nearly 25% rural.

Besides the ten key variables, geographic location is another criterion used in selecting target regions. The regions of study were selected based on geographic diversity. For example, Honduras and Mozambique are coastal nations, but Madhya Pradesh and Sichuan are land locked. Also, states in northeast India were avoided because of close proximity to China. Madhya Pradesh is located in the center of India and is representative of India with many of the states with large populations in the same region having similar water access issues.

From the preliminary data analysis, the following four regions were targeted for this study based on the ten selected variables:

- Honduras, Central America;
- Mozambique, Africa;
- Madhya Pradesh Province, India;
- Sichuan Province, China.

National data for China and India was used in some parts of this study when provincial data was not available. The two regions to be discussed in more detail in this paper include: a Chinese province, Sichuan, China, and a Sub-Saharan Africa country, Mozambique.

4.1.1. Water stress or scarcity

Aside from the water resource related variables listed in Table 1, water stress is also an important consideration in selecting representative regions. As shown in Appendix A, based on the earlier definition of water stress and scarcity, the four regions/countries chosen for analysis cover a wide range of water resource availability and consumption, with many facing water stress or scarcity conditions in the coming years. For example, many regions in China and India currently face water stress conditions. By 2025, Mozambique, India, and several provinces in China will be under water stress, with some Indian states facing water scarcity conditions (UNEP, 2005b). Honduras, however, has abundant water supplies and is not expected to suffer water supply shortages within the coming decades (UNEP, 2005b). Moreover, the four selected regions cover a wide range of annual domestic water consumption patterns, with Mozambique (4 m³/capita or 1060 gal/capita) and Madhya Pradesh (51 m³/capita or 13,500 gal/capita) being the extremes. All four regions also face challenges in providing clean water. Urban and rural areas with access to improved water supplies vary greatly between the selected regions.

In Table 1, the definition of "improved" water is based on the assumption that one technology is better for health than another, implying that "improved" water and sanitation designate safe water and sanitation. Although defining access as "improved" or "not improved" simplifies the meaning, there are contradictions to every situation. For example, in some locations an unprotected household

well may provide a better supply of water, both in terms of quantity and quality of water, than a household connection, which may be subject to intermittence and poor water quality. It is argued in the 2003 joint report by twenty-three UN agencies, *Water for People, Water for Life*, that methodology should be further developed so that progress can be assessed by measuring the number of people with access to “safe and sufficient” water and “safe and convenient” sanitation that meets basic welfare and hygiene needs (Water for People, 2004). The issue is not whether people have access to water and provision for sanitation, but whether the water supplies are safe, sufficient for their needs, and easily accessed at a price they can afford. It is also essential that there is a quality provision for sanitation that eliminates contact with human excreta and wastewater by making available toilets that are convenient, clean, easily accessed and affordable by all.

5. Case studies

The case studies section provides information on the two extreme cases: Sichuan Province, China and Mozambique, Africa. Sichuan Province, China is the most feasible possibility for introduction and implementation of water purification technologies. Mozambique, Africa represents the other end of the spectrum as the least feasible opportunity.

The level of detail and analysis varies based on the availability of information for the two regions examined in the case studies. General information includes the economic background, rural water supply and demand, water purification technology customers, water treatment technology market, and barriers and risks to entry. Additional information, based on data availability, includes private investment, international organizations and NGO investment, and rural water supply project financing practices.

5.1. Introduction to the case studies

5.1.1. Background in water privatization

In the early 1990s there was a shift in international development toward allowing transnational corporations to play a role in providing water and sanitation to the developing world. The belief was that individual countries could no longer meet their people's needs due to inefficiency and corruption. As a result, international development agencies, such as the World Bank, pushed for water management by the private sector. Although this seemed like an ideal solution, transnational corporations have struggled with privatization because they find difficulty in generating profits in the sale of services to the poor. This leads to communities accusing corporations of favoring projects for the wealthy (Robbins, 2003).

In recent years, private organizations have begun to consider both financial benefits and social responsibility in their attempts to meet the water needs of the poor. In

January of 2002, J.F. Talbot, chief of SAUR International, the fourth largest water company in the world, informed the World Bank that water could not be delivered to the poor. Talbot stated that “the scale of the need far out reaches the financial risk taking capacities of the private sector” (Hall, 2002).

At the same time it is still apparent in many developing countries that the majority of poor villages and urban slums are not yet served by piped water systems. Water sources available to the poor are polluted rivers, lakes, and shallow hand dug wells. To avoid these sources, some are dependent on street vendors that cost more over time than paying fees for piped water (an unaffordable investment). In addition, middle class consumers in many countries pay subsidized rates, becoming a burden to government and preventing the expansion of infrastructure to the poor (Siregar, 2003). It is conditions such as these that motivate organizations like the World Bank to push for privatization of the water market.

5.1.2. Role of international organizations and non-governmental organizations

Although international governmental and non-governmental organizations (NGOs) play a significant role in activities in rural areas of developing countries of the world, it is not certain that they will be customers for water purification technology. Based on research of current NGOs' practices, it can be concluded that some organizations are helping to install purification technology; the technology promoted is usually what is considered “appropriate technology”. Accordingly, technology must be “culturally and economically appropriate” so it can be operated and maintained by local people.

The NGOs are not in the business of promoting or marketing technology. Companies need to properly evaluate the environment and the market and not expect to implement technology that is not appropriate (i.e., no power or fuel source available). This is not suggesting that the NGOs do not work with private investors. It is stating that NGOs will in some cases work with private investors to evaluate what technology may be “culturally and economically appropriate” in a given setting.

Also, many organizations focus resources on providing water and sanitation education (and not technology) to local communities for families to understand the benefits of using clean water and the interaction between clean water, sanitation, and improved health. Significant investments in water technology focus on providing accessibility to water for those in need. Examples of this include hand digging or mechanical drilling a well, installing a hand pump, and constructing rainwater catchments and water storage tanks. The water is consumed in an untreated form, or subsequently boiled or treated with chlorine.

5.1.3. Potential customers and water market

Generally, it is difficult to collect customer information in rural areas of each region for several reasons. First, the

demand for sanitation and water supply has been weak because the perceived benefits of the purification technologies are outweighed by the investments needed to receive and sustain it (Deverill et al., 2002). Second, people in remote areas are hard to reach by organizations. Because it is difficult to identify end users, this research assumes that the customers would be communities, international organizations, or NGOs. These are the organizations willing to purchase the technology and deliver clean water to the end users.

A precise estimate of the size of the water purification market is not always possible. This analysis instead used information assembled from previous studies to develop broad estimates (Meng et al., 2004; WaterAid, 2003a; Water and Sanitation Program, 2002). The results of these earlier studies show there is a low willingness for rural populations to embrace new water supply technologies (WaterAid, 2003a). These represent immature markets for water purification technologies that need significant development.

5.2. Regional analysis Sichuan, China

5.2.1. Economic background

"Sichuan, located in southwestern China, is an agricultural province. Sichuan has the third largest population among all provinces in the country. Its rural population accounts for 80% of its total population. Boosted by the National Western Development Plan launched by the Chinese government a few years ago, Sichuan's economy is growing at a rapid pace with an average growth rate of 10.15% during 2000–2003 (BSS, 2003)," compared to a national average of 8.25% during the same period (The World Bank Group, 2005c). The growth rate is an average for both urban and rural areas. The rural growth rate is much lower than the average. This rate was not reported in statistics used for this study. Table 4 provides a summary of Sichuan's key economic indicators.

To stimulate domestic and foreign investment, China has been maintaining very low interest rates (lower than most countries) since 1998 (US Department of State, 2005a). The inflation rate has been successfully controlled between 0.7% and 4% in 2000–2004 (IMF, 2004). The exchange rate of Chinese Yuan to US dollar has been constant around 8.3 Yuan per US dollar for more than one decade, as a result of government manipulation. Recently, the government has expressed its intention to relax control over the exchange rate (The Economist, 2005), but it is difficult to forecast future fluctuations.

5.2.2. Rural area water supply and demand

Most rural residents obtain water directly from rain-water, surface water, and shallow wells. Water consumption per capita is estimated to be 80.55 l per day (NBSC, 2003). Province specific data is lacking to determine the percentage of residents who have access to improved water sources, but the nationwide percentage is 68% of the rural

Table 4
Sichuan economic indicators

Economic indicators	Sichuan	China
Population (2002)	86 Million ^a	1285 Million ^a
Urban (2002)	17 Million ^a	502 Million ^a
Rural (2002)	69 Million ^a	782 Million ^a
GDP (2002) (\$US)	59 Billion ^a	1266 Trillion ^a
Per capita net income (2002) (rural) (\$US)	255 ^a	299 ^a
GDP average annul growth rate (2000–2003)	10.15% ^b	8.25% ^c

^aNBSC (2003).

^bBSS (2003).

^cThe World Bank Group (2005c).

population (WHO, 2005), which is used throughout this study to estimate Sichuan's rural population with improved access to water supply. At the national level, the Chinese government has set a high priority for increasing rural access to improved, conveniently located water. The government encourages investment of this kind to help reduce poverty and improve health and living standards in rural China.

One factor impacting Sichuan's water supply is the construction of the world's largest water dam, the Three Gorges Project (which was started in 1994 and completed in 2006, ahead of schedule), in Hubei (a neighboring downstream province of Sichuan) (IRN, 2006). As a result of the many concerns regarding the possible pollution to the water supply of Sichuan, many treatment facilities are expected to be built in near future.

5.2.3. Private investment

Most water supply infrastructure is state-owned and state-operated. Since China joined the World Trade Organization (WTO) in 2002, the central government has been gradually introducing more market-driven practices. The government is also slowly lowering barriers to market entry to allow private investment and international investment. For example, the consortium of Vivendi (France) and Marubeni Company (Japan), a foreign owned water group, has recently been granted the right to operate the Sixth Water Plant in Chengdu (the capital city of Sichuan) (US DOC, 2005). The number of private investors and international investors is still small. Their presence has been regarded as positive to the development of the water supply sector. The local government and rural residents are continuing to welcome their investment.

5.2.4. International organizations and NGOs involvement

Since the 1980s, China's rural water supply projects have obtained financial support from various international organizations, including the World Bank, United Nations Development Program (UNDP), the United Nations Children's Fund (UNICEF), and the WHO, etc. (Meng et al., 2004). The World Bank is the major contributor

among these organizations, and has financed a total of \$US137 million in loans for four water supply projects since the 1980s. Table 5 presents a brief summary of the four projects. Sichuan was one of the beneficiary provinces in the first project (CRWS).

The Chinese government encourages NGOs to support rural water supply projects. For example, from 2000 to 2004 one major Chinese NGO, the All-China Women Federation, has raised approximately \$US18 million to support rural water supply projects in western provinces (Meng et al., 2004).

5.2.5. Rural water supply project financing practices

In China, partial user-financing for rural water supply projects has been in place for many years and has been mostly successful. Funding of a project comes from various sources, including contributions from different levels of government, bank loans, and end users. In a typical World Bank supported project, the Bank provides 50% of the capital through credit/loan (Water and Sanitation Program, 2002). The different levels of government finance 25%, with the end users contributing the remaining 25% in the form of cash or labor hours. The end users are also responsible for the Bank debt payback and the operating/maintenance cost through payment of water tariffs, which are regulated by local governments. Taking all this information into consideration, the users finance 75% of the investment cost (Water and Sanitation Program, 2002). For economically disadvantaged villages that cannot afford this share, it is expected that other players (governmental and/or non-governmental organizations) would provide stronger support.

5.2.6. Water purification technology customers and market

Past rural water supply projects were mainly funded by governments, international organizations, and NGOs. These could be the potential customers for the water purification technologies. With the opening of the water supply market, private investors could also be potential customers. Province specific data is lacking to determine the size of the market of water purification technologies in Sichuan. The total size of the market in China is enormous and provides a significant opportunity. It is reasonable to

estimate that Sichuan, the province with the third largest population in China, has a significant portion of this vast market. Most of the market demand is covered by local production, but the import market is promising. According to the US Commercial Service, China's market for imported water treatment equipment is about \$US2 billion per year (US DOC, 2005). Table 6 summarizes the market size data.

5.3. Regional analysis Mozambique

5.3.1. Economic background

Although it has experienced high GDP growth in recent years, Mozambique is one of the poorest countries in the world. The 7.25% growth rate is one of the highest in Africa, but poverty still remains pervasive in rural areas. Per capita income in rural Mozambique averages \$US65/year, while the median is only \$US30/year (Walter et al., 2004). The high GDP growth rate is observed because the scale of Mozambique's economy is small and is not an indicator of an economic boom. Table 7 provides a summary of the key economic indicators in Mozambique.

The inflation rate is high but has been steadily decreasing in recent years, from 16.8% to 12.9% in 2002–2004 (IMF, 2004). Mozambique has a small capital market of ten commercial banks. Taking out a loan is usually difficult and expensive because of high interest rates (US Department of State, 2005b). The foreign exchange rate of Mozambique meticals to US dollar is stable around

Table 6
Estimated size of the water and wastewater technologies market in China, 1999. (millions of \$US)

Technology	Local production	Imports	Total market size
Water treatment and supply	1506	646	2152
Wastewater treatment	1191	1273	3184
Total	3417	1919	5336

Source: US DOC (2005).

Table 5
Rural water supply and sanitation program in China supported by the world bank

Name of project	No. of province	No. of counties	No. of beneficiaries million people	Implementation period		WB credit/loan (M\$)
				Start	Complete	
CRWS	5	25	5.93	1985	1991	10.46 (C)
CRWSS	6	75	9.06	1992	1998	11.00 (C)
NRWS	5	40	4.60	1997	2003	70.00 (C)
FRWSS	4	27	3.08	1999	2005	30.00 (C)
						16.00 (L)
Total	20	167	22.67			137.46

Source: The World Bank Group (2002).

Table 7
Mozambique economic indicators

Economic indicators	Mozambique
Population (2005)	19.50 Million ^a
Urban (2005)	7.40 Million ^a
Rural (2005)	12.10 Million ^a
GDP (2003) (\$US)	4.3 Billion ^b
Per capita net income (2002) (rural) (\$US)	65 ^c
GDP average annual growth rate (2002–2003)	7.25% ^d

^aUNESA (2004).

^bIMF (2004).

^cWalter et al. (2004).

^dThe World Bank Group (2005b).

23,700 meticals per US dollar in recent years (US CIA, 2005).

Lack of clean water is a serious problem in Mozambique. Improving water quality and increasing access to safe water, in both urban and rural areas, are among the Government's poverty reduction strategies. Government efforts have yet to reach rural areas. Even in cities, the water price sometimes is so high that not all city residents can afford access. The Mozambican government initialized a "Poverty Reduction Strategy Paper (PARPA)" in 2001, which set the water sector as a priority (SDC, 2003). Under this strategy, there are two main goals in improving the country's water supply: (1) establish an integrated water resource management system; and (2) build sanitized water supplies for all areas. Given the limited resources and the current government efficiency, there is still a long way to go before the goals can be achieved.

5.3.2. Rural area water supply and demand

Twenty-four percent of the rural population has access to improved water sources (WHO, 2005). Shallow wells and hand pumps are the only means of obtaining water in rural areas.

5.3.3. Private investment

The Mozambique government has consistently failed to provide affordable and sustainable clean water for its rural population because of government's limited capacity and authority. The government is now encouraging the participation of the private sector (including international investors) in the construction of improved water sources. A core element of their strategy is the policy of public asset privatization, as indicated by the Enhanced Structural Adjustment Facility Paper (ESAF), another strategic document approved by the government recently (Zandamela, 2001). Because most water supply facilities are public assets, this policy offers many opportunities for foreign investors who can provide cheap and sustainable water supply technologies.

Two private sector participants are known for this study. Aguas de Mozambique (AdeM) (The World Bank Group, 2004), a consortium of SAUR International, IPE Aguas de

Portugal and a group of five local investors, has undertaken a 15-year lease contract to improve urban water systems in five cities in Mozambique since 1999 (WaterAid, 2003a). AdeM's operation has been turbulent and unsuccessful. AdeM's involvement is limited to urban areas only. A Mozambican private institution called EPAR implements water sanitation systems mainly in rural areas (WaterAid, 2003a). EPAR receives financial and technical support from both the Mozambican government and Oxfam Belgium, a major European NGO.

A recent sample study conducted by WaterAid (WaterAid, 2003a), a major water resource related NGO based in the UK, shows that private sector participation has barely tackled any water supply difficulties in Mozambique due to the long-lasting major problems in capacity building, community participation, finance and institution reform. The study cautions private investors' involvement in water supply projects in Mozambique. Extreme poverty, poor governance, inactive citizenry, and lack of responsible partners are among the concerns cited by the study.

5.3.4. International organizations and NGOs involvement

Water supply projects in Mozambique predominantly depend on the involvement of international organizations and NGOs around the world. Refer to Appendix B for the major participants and their projects in recent years.

5.3.5. Rural water supply project financing practices

Government financed projects require water consumers pay 2–10% of the capital cost of a water facility (Breslin, 2002). The remainder comes from the government's budget. For each dollar the government invests, 90 cents are expected to come from foreign aid (WaterAid, 2003b). While heavily relying on foreign aid itself, the government has faced sharp criticism for making its poor citizens pay their share of 2–10% (Breslin, 2002). In addition, NGO financed projects sometimes provide water consumers a range of options to contribute to the capital cost. For example, they can contribute maize (corn, a major crop in Mozambique) or in-kind labor instead of cash.

5.3.6. Water purification technology customers and market

The Mozambican government, international organizations, and NGOs are potential customers for the technologies identified in this study. EPAR and AdeM are potential private sector customers. Water supply/treatment technologies used in rural Mozambique are predominantly hand-dug wells and hand pumps. Hand pumps have proven to be extremely difficult to maintain. More than 35% of hand pumps across the whole country are dysfunctional (WaterAid, 2003a, b). Protected springs and rainwater harvesting are also used occasionally (Breslin, 2003). There is insufficient information available to determine market share for these technologies.

5.4. Profitability uncertainties due to high risks

Common risks and barriers associated with investment into the water purification technology industry in the four selected regions are listed below:

1. *Government corruption.* A major concern is the pervasive corruption in government. Bribery and “under the table” deals are common practice in doing business in these regions. This may add unforeseeable expenditures to initial investment.
2. *Affordability.* A risk specific to rural area water supply projects is that poor villagers usually cannot afford treated water. A vast amount of the rural population in these developing countries is still struggling to obtain basic living necessities. According to a WHO estimate, a household should spend 3.5% of its income for basic water supply (Walker et al., 1999). Based on this estimate, investors should not expect to finance any significant part of the project through water tariffs. Investors should also be aware that the revenue generated is likely to be inconsistent and unsustainable.
3. *Uncertainty of capital recovery.* The same problem leads to difficulties in capital recovery for any investors. Investors must capitalize on the non-cash contribution of the poor communities, as well as the social benefits, to justify their investment.

Additionally, each region/country has specific risks, as discussed below:

5.4.1. Sichuan, China

1. *Water tariff regulations.* The biggest risk comes from the water tariff regulation by the government. In poor areas government intervention in setting water prices is common. The variation of water prices is a major factor to be considered by any investor (US DOC, 2005).
2. *Import regulations.* If a water supply project requires imported technologies and equipment, then China’s constantly changing import regulations pose a threat to investors. Identical technologies and equipment may be subjected to different levies at different times and ports. This uncertainty complicates the investors’ cost control.

5.4.2. Mozambique, Africa

1. *Low literacy level and hygiene education.* The low literacy rate and hygiene education among rural residents undermine public awareness of the importance of clean water. As a consequence, rural residents have a low willingness to accept more advanced water treatment technologies (WaterAid, 2003a). This is a common phenomenon in Africa.
2. *State owned facilities.* Water supply infrastructure is traditionally owned and operated by the state. The worker union in the state-owned facilities strongly

opposes private sector participation in any utility services in the country (Zandamela, 2001).

3. *Insufficient statistical data.* Due to the constrained capacity and authority of the Mozambican government, statistical data needed by investors is insufficient. Even the available statistics are usually unreliable (WaterAid, 2003b).

5.5. Investment feasibility

Overall, based on the set of qualitative variables used in this study to evaluate and analyze the business case for water purification technology in rural areas of developing countries, it is difficult to build a strong case for any business or organization focusing on immediate returns on capital investment.

If a private investor is willing to take on business risk to pursue a market, Sichuan, China may represent a potential opportunity for limited water purification technologies. Sichuan has the largest population density and rural population percentage among the four regions studied. This could be a sign of possible market and demand growth for technologies. Sichuan’s economy is growing at a pace as fast as the national economy of China. Rural household income is likely to increase with the growth of the economy and lead to increased willingness to pay for improved water supply. Moreover, the Chinese government has implemented favorable policies to encourage foreign investment, focusing on poverty reduction and living condition improvement in Western China. For example, a 5-year tax break is available in Sichuan because of the National West Development Plan (SATC, 2001, 2002). Based on these facts, Sichuan, China could be the most favorable potential market for water purification technologies among the four regions.

On the other hand, it is not recommended to pursue business opportunities in Mozambique. Most water consumers in rural Mozambique are relying on unstable income sources. It is not unusual for people to fail to pay for water, no matter how low the water tariff would be. In many cases water consumers have to pay the water tariff with crops due to shortage of cash. This indicates a very unstable cash flow and unsustainable revenue source for the operation (Zandamela, 2001). The same problem leads to difficulties in capital recovery for any investors. Investors must capitalize on the non-cash contribution of the poor communities, as well as the social benefits to justify their investment (WaterAid, 2003a). The inability to pay for improved water associated with the low literacy rate and the lack of hygiene education among rural residents leads to low acceptance of advanced technologies. Another critical factor is the lack of reliable statistical data for investors to make a well-informed business decision. There are also other factors that would prevent a company from operating successfully in Mozambique, such as corruption, state-owned facilities preventing private business entry, etc. Considering these risks, it is unlikely pursuing business opportunities for water purification technology in Mozambique will be economically feasible.

6. Conclusions and recommendations

Rural areas of the world are populated with poor people unable to fulfill the basic needs for clean water and sanitation. Although this represents an important group of potential users, the population density in rural developing countries may be difficult to quantify, small in number, and difficult to reach. Because of these critical issues, markets for rural areas of developing countries may not represent economically feasible regions to implement water purification technologies. The development effort should focus on generating demand for clean water by changing the social and cultural norms that currently prevent inhabitants of rural areas of developing countries from understanding the necessity of clean water.

As part of its Millennium Development Goals the United Nations has set two aggressive goals which apply directly to the future of the water purification market.

- By 2015, reduce by half the proportion of people without access to safe drinking water.
- Target for 2015: Halve the proportion of people living on less than a dollar a day and those who suffer from hunger.

These goals were set at a September 2000 conference and have been approved by the world community (UNDP, 2005). It is for these reasons that a plausible business strategy may be to approach water purification as a corporate responsibility and social investing in the short term, which will allow one's organization to be positioned well once the economic ability of individuals, governments, and donor agencies are better aligned. As one individual from the NGO CARE stated during this research, taking the time now to learn about people and culture and develop appropriate business models without the pressure of making a profit will provide a competitive advantage when the market for water purification technology evolves over the next decade.

7. Future research

Population density is typically low in rural areas of developing countries and could be one of the larger impediments to growth in new or expanded markets for technology. Urban regions of the developing world may represent a larger potential market based on population trends and density. In fact, urbanization is one of the critical global trends shaping the future. For example, by 2025, it is estimated that two-thirds of the world's people will live in cities. In contrast, only one-third of the world's population was urban 35 years ago. In addition, more than 150,000 people are being added to urban populations in developing countries every day. Poorer people can be located in urban regions, and in some developing nations, may represent a large share of the potential consumers and end-users for growth in these regions.

Because country, province/state and region specific data was used in the current study, key indicators represent possible barriers to urban areas in these same countries, provinces/states, and regions. Future studies should focus on different countries and/or regions within these countries.

If accurate information regarding market size potential is desired, a longer time frame of 18–24 months may be required to gather information. A caveat is that even with more market research, the conclusions may remain the same.

Acknowledgments

We would like to thank Caterpillar Inc. for funding of this important research project. Without their financial support we would not have been able to provide this valuable contribution to research in the area of economic and social feasibility for implementation of water purification technologies in rural areas of lesser developing countries. The Sustainable Futures Institute (SFI) research team from Michigan Technological University appreciates Caterpillar's corporate social responsibility and contribution to further research in this developing area. The authors wish to recognize Ms. Alexis M. Troschinetz, Mr. Joshua R. Cowden, Dr. David W. Hand, and Dr. James R. Mihelcic for their collaboration on this project.

Appendix A

Countries estimated to be water scarce in 2025 Seckler et al., 1999 are given in Table A1

Table A1
Countries estimated to be water scarce in 2025 (Seckler et al., 1999)

Category 1 (absolute water scarcity)	Category 2 (economic water scarcity)	Category 3	Category 4
Afghanistan	Angola	Albania	Argentina
Egypt	Benin	Algeria	Austria
Iran	Botswana	Australia	Bangladesh
Iraq	Burkina Faso	Belize	Belgium
Israel	Burundi	Bolivia	Bulgaria
Jordan	Cameroon	Brazil	Canada
Kuwait	Chad	Cambodia	(China) ^a
Libya	Congo	Central African Republic	Costa Rica
		Chile	Cuba
Oman	Cote d'Ivoire	Columbia	Denmark
Pakistan	Ethiopia	El Salvador	Dominican Republic
Saudi Arabia	Gabon		Ecuador
		Gambia	Finland
Singapore	Ghana	Guatemala	France
South Africa	Guinea-Bissau	Guinea	Germany
Syria	Haiti	Honduras	Greece
Tunisia	Lesotho	Indonesia	
United Arab Emirates	Liberia		
Yemen	Mozambique	Kenya	Guyana

Table A1 (continued)

Category 1 (absolute water scarcity)	Category 2 (economic water scarcity)	Category 3	Category 4
China ^a	Niger	Lebanon	Hungary
India ^a	Nigeria	Madagascar	India ^a
	Paraguay	Malaysia	Italy
	Somalia	Mali	Jamaica
	Sudan	Mauritania	Japan
	Uganda	Morocco	Mexico
	Zaire	Myanmar	Netherlands
		Namibia	North Korea
		Nepal	Norway
		New Zealand	Panama
		Nicaragua	Philippines
		Peru	Poland
		Senegal	Portugal
		Tanzania	Romania
		Turkey	South Korea
		Venezuela	Spain
		Zambia	Sri Lanka
		Zimbabwe	Surinam
			Sweden
			Switzerland
			Thailand
			UK
			Uruguay
			USA
			Vietnam

Countries that face absolute water scarcity are recognized as category 1 and they will not be able to meet water needs in the year 2025. Countries that must more than double their efforts to meet 2025 water needs, but will not have the financial resources available to develop these water supplies, face economic water scarcity and are recognized as category 2. Countries that must increase water development between 25 and 100% to meet 2025 needs but have financial resources to do so are considered category 3. Countries that will have to increase water development modestly overall on average by only 5% to keep up with the 2025 needs are considered category 4.

^aThese countries have severe regional water scarcity. A portion of their populations (381 million people in China in 1990 and 280 million people in India in 1990) are in Category 1. The rest of their populations are in Category 4.

Appendix B

Major international and non-governmental organizations in Mozambique are given in Table B1

Table B1
Major international and non-governmental organizations in Mozambique

International organizations	
The Netherlands	Major donor since the late 1970s budget 3.5 million Euro per year
The World Bank	Supported two projects
Switzerland	Mainly focusing on rural water supply, training, institutional building and human resource development, budget \$US3.5 million per year
The African Development Bank	Planning rural water supply projects for Niassa and Nampula provinces

Table B1 (continued)

The European Union	Active with some technical assistance, watershed commissions and development and in a future sanitation project for Beira
UNICEF	Provide institutional support, participant in a rural water project in Zambezia province
Canada	Funds a rural water supply project in Inhambane province, implemented by Cowater, a Canadian firm
Japan	Funds a rural water supply project in Zambezia province implemented by a Japanese NGO
NGOs	
WaterAid	Active in rural Niassa and rural Zambezia provinces
CARE	Active in rural Inhambane province
World Vision	Active in rural Nampula province
Helvetas	Active in rural Maputo province and Cabo Delgado
MSF	Active in urban Maputo
AusAid	Funds World Vision project
Ireland Aid	Active in Niassa province

References

- Breslin, N., 2003. Demand Response Approach in Practice: Why Sustainability Remains Elusive. WaterAid.
- Breslin, N., 2002. Lessons from the Field. Update Series: Mana and Nipepe. WaterAid.
- BSS, 2003. Bureau of Statistic of Sichuan. Available at: http://www.sc.stats.gov.cn/stats_sc/zxtjxx/200309230049.htm.
- DDWS, 2005. Ministry of Rural Development, India. Available at: <http://ddws.nic.in/index.html>.
- Dev, S. M., 2004. Post Reform Regional Variations. Available at: <http://www.india-seminar.com/2004/537/537%20s.%20mahendra%20dev.htm>.
- Deverill, P., Bibby, S., Wedgwood, A., Smout, I., 2002. Designing Water Supply and Sanitation Projects to Meet Demand in Rural and Peri-urban Communities. Book 1: Concepts, Principles, and Practice. Water, Engineering and Development Centre, Loughborough University, Leicestershire, UK.
- The Economist, 2005. Putting things in order, China ought allow more flexibility in its exchange rate, sooner rather than later. The Economist 374 (8418), 82.
- FAO, 2005. UNEP Geo Data Portal. Available at: <http://geodata.grid.unep.ch/>.
- Government of Madhya Pradesh, 2003a. Annual Plan 2002–2003. Madhya Pradesh, India, Retrieved 8 March 2005. <http://www.mp.nic.in/annualplan/>.
- Government of Madhya Pradesh, 2003b. State Water Policy. State Resources Department, Madhya Pradesh, India, Retrieved 8 March 2005. <http://www.mpgovt.nic.in/index/departments.htm>.
- Guest, R., 2004. The good news in Africa. In: Franklin, D. (Ed.). The World in 2005. The Economist, London.
- Hall, D., 2002. The Water Multinationals 2002—Financial and other Problems. Public Services International Research Unit, London.
- IMF, 2004. World Economic Outlook Database. Available at: <http://www.imf.org/external/pubs/ft/weo/2004/02/data/index.htm>.
- IMF, 2005. Available at: <http://www.imf.org/>.
- IRN, 2006. International River Network: Linking Human Rights and Environmental Protection. <http://www.irn.org/programs/threec/>, accessed November 22, 2006.

- Meng, S., et al., 2004. Rural water supply and sanitation in China—scaling up services for the poor. Presented at Scaling Up Poverty Reduction: A Global Learning Process and Conference, Shanghai, China, May 25–27, 2004.
- NBSC, 2003. 2003 China Statistical Yearbook. National Bureau of Statistics, Retrieved 8 March 2005, <<http://www.stats.gov.cn/>>.
- Office of the Registrar General—India, 2001. Census of India 2001, Available at: <<http://www.censusindia.net/index.html>>.
- Ping, Z.J., 2005. The Current Situation of China's Electric Power Industry, Office of Chongqing Energy Conservation Technical Service Center, Austrian Energy Agency, Retrieved 13 March 2005, <http://www.eva.ac.at/projekte/china_power.htm>.
- Robbins, P.T., 2003. Transnational corporations and the disclosure of water privatization, *Journal of International Development*, 1073–1082.
- SATC, 2001. On preferential treatments for the national west development program. The 20th Directive (2001) from the Ministry of Finance of P.R. China, the State Administration of Taxation of China and the Customs General Administration P.R. China, Available at: <<http://www.chinatax.gov.cn/view.jsp?code=200309241007373159>>.
- SATC, 2002. On implementations of the state's taxation policies regarding the national west development program. The 47th Directive (2002) from the State Administration of Taxation of China, Available at: <<http://www.chinatax.gov.cn/view.jsp?code=200309241007432975>>.
- SDC, 2003. SDC Water and Sanitation Strategy for Mozambique 2003–2006, Available at: <http://www.sdc.admin.ch/ressources/product_55_cn_1155.pdf>.
- Seckler, D., Barker, R., Amarasinghe, U., 1999. Water scarcity in the twenty-first century. *International Journal of Water Resources Development* 15 (1–2), 29–42.
- Siregar, P.R., 2003. World Bank and ADB's Role in Privatizing Water in Asia, jubileesouth.org, Available at: <<http://www.jubileesouth.org/news/EpZyYyEyyIqGyKXRu.shtml>>.
- The World Bank Group, 2002. Rural Water Supply and Sanitation Program in China, Available at: <<http://www.worldbank.org.cn/English/content/ruralwater.pdf>>.
- The World Bank Group, 2004. PID for the second national water development project—supplemental credit. Report No. AB555, The World Bank Group.
- The World Bank Group, 2005a. UNEP Geo Data Portal, Retrieved 1 March 2005, <<http://geodata.grid.unep.ch/>>.
- The World Bank Group, 2005b. Data and Statistics, Available at: <<http://www.worldbank.org/data/>>.
- The World Bank Group, 2005c. World Development Indicators Database, Available at: <<http://devdata.worldbank.org/data-query/>>.
- The World Bank Group, 2005d. Governance Indicators: 1996–2004, Retrieved 25 February 2005, <<http://www.worldbank.org/wbi/governance/govdata/>>.
- UNDP, 2005. Millennium Development Goals: The global challenge, Available at: <<http://www.undp.org/mdg/abcs.html>>.
- UNEP, 2005a. UNEP Geo Data Portal, Retrieved 4 March 2005, <<http://geodata.grid.unep.ch/>>.
- UNEP, 2005b. Vital Water Graphics, Retrieved 15 February 2005, <<http://www.unep.org/vitalwater/index.htm>>.
- UNESA, 2004. World Urbanization Prospects: The 2003 Revision, Sales No. E.04.XIII.6, UNESA Population Division, New York.
- UNESCO, 2005. World Water Assessment Programme: Valuing Water, Retrieved 4 March 2005, <http://www.unesco.org/water/wwap/facts_figures/valuing_water.shtml>.
- US CIA, 2005. The World Fact Book 2005, Available at: <<http://www.odci.gov/cia/publications/factbook/>>.
- US DOC, 2005. Water Supply and Wastewater Treatment Market in China, PB2005-102181, US Department of Commerce, International Trade Administration, Washington, DC.
- US Department of State, 2005a. 2005 Investment Climate Statement—China, Available at: <<http://www.state.gov/e/eb/ifa/2005/42000.htm>>.
- US Department of State, 2005b. 2005 Investment Climate Statement—Mozambique, Available at: <<http://www.state.gov/e/eb/ifa/2005/42092.htm>>.
- Walker, I., Ordoñez, F., Serrano, P., Halpern, J., 1999. Pricing, Subsidies, and the Poor: Demand for Improved Water Services in Central America, World Bank Policy Research Working Paper No. 2468, Retrieved 18 March 2005, <http://ssrn.com/abstract=632538>.
- Walter, T., et al., 2004. Determinations of rural income, poverty, and perceived well-being in Mozambique in 2001–2002. Economics Directorate from the Ministry of Agriculture and Rural Development, Republic of Mozambique.
- WaterAid, 2003a. Small-scale Private Sector Participation in Niassa, Mozambique, Available at: <<http://www.wateraid.org.uk/>>.
- WaterAid, 2003b. Managing Diversity: WaterAid Mozambique Programme Evaluation Summary, Available at: <<http://www.wateraid.org.uk/>>.
- Water and Sanitation Program, 2002. Willingness to Charge and Willingness to Pay: The World Bank-assisted China Rural Water Supply and Sanitation Program, Available at: <http://www.wsp.org/publications/global_wtp_china.pdf>.
- Water For People, 2004. Honduras, Retrieved 5 May 2005, <<http://www.waterforpeople.org/latinHonduras.html>>.
- WHO. World Health Organization, 2005. UNEP Geo Data Portal, Available at: <<http://geodata.grid.unep.ch/>>.
- Zandamela, H., 2001. Lessons from Mozambique: The Maputo Water Concession, Available at: <<http://www.citizen.org>>.



ELSEVIER

Journal of Environmental Management 88 (2008) 428–436

Journal of
Environmental
Management

www.elsevier.com/locate/jenvman

Emission of CO₂, CH₄ and N₂O from freshwater marsh in northeast of China

Changchun Song^{a,*}, Jinbo Zhang^{a,b}, Yiyong Wang^a, Yaosi Wang^c, Zhichun Zhao^a^aChinese Academy Sciences, Northeast Institute of Geography and Agricultural Ecology, CAS, Changchun, Jilin 130012, China^bState Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China^cInstitute of Atmospheric Physics, CAS, Beijing 100029, China

Received 5 April 2006; received in revised form 13 November 2006; accepted 11 March 2007

Available online 22 May 2007

Abstract

The wetlands play an important role in carbon storage, especially at high latitudes, at which they store nearly one-third of global soil carbons. However, few studies have investigated the emissions of CO₂, CH₄ and N₂O in the long-term, especially effects of freeze–thaw cycles on these gases emissions in freshwater marsh ecosystems. In this paper, we collected greenhouse gas emission data from a freshwater marsh area in China for 4 years, evaluated their release variables and speculated on their potential atmospheric impact. For this paper, we report on the CO₂, CH₄ and N₂O emission rates recorded from June 2002 to November 2005 in the Sanjiang Plain of northeast China. We measured their interannual variations and fluctuations, as well as factors affecting their emissions, and estimated their regulation and freeze–thaw cycle impacts. Our results revealed obvious CO₂ and CH₄ emission fluctuations during the winter months, and during the freeze–thaw cycle, and a strong interannual variation during the growing season. Overall, we documented a close relationship between the CO₂ and CH₄ emissions, implicating some regulatory commonality. We determined that the marsh was a N₂O sink during the winter, but a significant source of N₂O during the freeze–thaw cycle as the temperature increased, especially in early summer. During the thaw–freeze period, the N₂O levels were positively correlated with the water depth. Additionally, water depth greatly governed the interannual variation of the N₂O emissions from the marshes during the thaw–freeze period.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: CO₂ emission; CH₄ emission; N₂O emission; Freeze–thaw period; Interannual variation; Northeast china; Winter

1. Introduction

The wetlands play an important role in carbon storage, especially at high latitudes, at which they store nearly one-third of global soil carbons (Gorham, 1991). The carbon balance of these zones is sensitive to climate changes (Alm et al., 1999a,b; Baron et al., 2000). According to Brix's (1990) research, only 15% of the wetland plant's net-assimilated carbon atoms are released to the atmosphere, which indicates that the wetland ecosystem can contribute to high CO₂ concentrations by acting as a carbon sink.

Wetlands, on the other hand, are among the primary sources of atmospheric CH₄, of which 110–1012 g are from the anoxic decomposition of organic matter in natural

wetlands each year. This accounts for 15–30% of the total annual CH₄ emissions (Cicerone and Oremland, 1988; Barklett and Harriss, 1993), 50–60% of which come from high-latitude peatland areas. Over the past several centuries, atmospheric CO₂ and CH₄ concentrations have increased rapidly (Houghton, 1995). Fluctuations in CO₂ levels directly influence plant production and the overall vegetation community structure. Moreover, these fluxes impact the evolution and carbon cycling in these and other ecosystems.

N₂O may reach the stratosphere by transport and diffusion, which in turn, may hasten the destruction of the stratospheric ozone layer. However, few direct measurements of winter N₂O fluxes have been obtained from the wetlands. Martikainen et al. (1993) reported N-rich northern peatlands drained for forestry can be sources of N oxides. Nykänen et al. (1995) reported that N₂O emissions peaked in the grassland areas during the onset of winter,

*Corresponding author. Tel.: +8643185542204; fax: +8643185542298.
E-mail address: songcc@neigae.ac.cn (C. Song).

and during the spring thaw, while Alm et al. (1999a,b) reported that winter N_2O fluxes were negligible at all virgin sites and drained bogs, but that the emissions on efficiently drained fens were 27–28% of their annual emission rates, respectively. The estimates of N_2O emission concentrations are ambiguous at best, indicating that more studies are required.

The seasonal freeze–thaw cycle is an important meteorological event in boreal areas. The emission of greenhouse gases during freeze–thaw periods probably results in altered atmospheric chemistry. The transfer and transformation of the soil nutrients greatly impact the carbon balance in these areas during this period, when the emission of CO_2 is about 3–50% that for the entire year (Lafleur et al., 2001; Wickland et al., 2001). In the boreal areas, the freeze–thaw fluctuations are the main emission control factors, as they may increase the intensity of microorganism respiration and concentrations of available nutrients (Muller et al., 2002). The N_2O emission rate increases concurrently (Schimel and Clein, 1996). The freeze–thaw cycle in these regions also has a significant impact on the biology–terrestrial global chemistry cycle patterning (Grogan et al., 2004). Dise (1993) studied CH_4 emissions in different wetland types, and discovered that during the winter months the emission levels were roughly 4–21% of the annual emissions. Jin et al. (1999) studied CH_4 release in the Tibetan Plateau wetland, and measured high CH_4 concentrations during the spring freeze–thaw period. Until now, Melloh and Crill (1996) had recorded the longest observation of CH_4 emission, which was in New Hampshire over a period of 5 years. They found that the CH_4 emission rate was roughly 2–9.2% that of the entire year. However, few studies have investigated the effect of freeze–thaw cycles on CO_2 , CH_4 and N_2O emissions in freshwater ecosystems.

Overall, these greenhouse gas emission studies were mainly conducted indoors under highly controlled conditions. Long-term reports regarding the measurements of CO_2 , CH_4 and N_2O emissions during this time were sparse. Thus, there is no reliable, sufficient or accurate field data for CO_2 , CH_4 and N_2O emission measurements. The Sanjiang Plain is the largest freshwater marsh in China, and belongs to the freeze–thaw area. Gas emission observations have been conducted at Sanjiang for more than 3 years. Objectives of this study were to understand the regulation of CO_2 , CH_4 and N_2O emissions over a long-term period, to investigate the main factors affecting interannual variations in CO_2 , CH_4 and N_2O emissions, and to comprehend how the freeze–thaw cycle impacts greenhouse gas release from freshwater marshes.

2. Materials and methods

2.1. Study site

We selected the *Carex lasiocarpa* freshwater marsh study site to measure greenhouse gas emissions. We set up

our study area at the Sanjiang Mire Wetland Experimental Station, Chinese Academy of Science, in the Sanjiang Plain of China, at approximately 47°35'N, 133°31'E. The average altitude at *Carex lasiocarpa* is between 55.4 and 57.9 m, the annual mean temperature is 1.9 °C and the non-frost period is 125 days (Song et al., 2003). The study site is in a seasonal frozen zone (Song et al., 2003), with annual precipitation between 550 and 600 mm, concentrated in July and August, a period that accounts for more than 65% of the annual precipitation. The soil characteristics at *Carex lasiocarpa* have been described in detail by Ding et al. (2002) and Song et al. (2003).

We placed the thermometers, with soil measurement sensitivities of ± 0.1 °C, at depths of 5, 10, 20, 30, 50, 70, 100, 120 and 150 cm. The DT500 data logger measures the thermometer readings every hour, and provides the output in a format that is the same as that of the meteorological factor data.

2.2. Data sampling and analysis

We measured gas emissions in triplicate. The static chamber method was adopted to measure the concentrations of CO_2 , CH_4 and N_2O levels. The chamber (50 cm × 50 cm × 50 cm) and its collar were made from stainless-steel (Song et al., 2003). During our observations, the chamber was placed into the collar with water to prevent leakage, and the vegetation was included within the chamber. In order to maintain a consistent temperature within the chamber, we wrapped the cylinder with a quilt. Inside the chamber, a small fan that was used to stir the air, and a thermometer sensor and a trinal-venthole were installed. The sampling time was 30 min, and samples were collected with an injector with a 60 ml capacity, once in every 10 min.

The CH_4 , CO_2 and N_2O emissions were measured with gas chromatography (Agilent 4890) equipped with FID and ECD. The CH_4 was separated from the other gases with a 2 m stainless-steel column, with an inner diameter of 2-mm 13XMS column (60/80 mesh). The CO_2 was separated with a 2 m stainless-steel column with an inner diameter of 2 mm Porapak Q (60/80 mesh). For measuring CH_4 and CO_2 , the FID operates at 200 °C using high-pure nitrogen as a carrier gas, at a flow rate of 30 ml/min. The N_2O portion was separated using a 1 m stainless-steel column with an inner diameter 2 mm Porapak Q (80/100 mesh), and was measured using the ECD, which was set at 330 °C. The ECD used high-pure nitrogen as a carrier gas, at a flow rate of 35 ml/min. The column temperatures were maintained at 55 °C for all separations.

The gas flux was calculated according to the following equation:

$$J = \frac{dc}{dt} \frac{M}{V_0} \frac{P}{P_0} \frac{T_0}{T} H,$$

where the dc/dt is the slope of the gas concentration curve variation, along with time. M is the mole mass of each gas.

P is the atmospheric pressure in the sampling site. T is the absolute temperature during sampling. V_0 , T_0 , P_0 are the gas mole volume, air absolute temperature and atmospheric pressure under standard conditions, respectively. H is the height of chamber above the water surface.

The data in this study were collected over 3 years, from June 2002 to September 2005. In the winter months (November–March, next year), we collected measurements in triplicate once every 2 weeks. However, during a large flux freeze–thaw fluctuation period (April–June), we collected emission measurements once daily. During the growing season (July–October), the CH_4 flux was measured twice weekly. When the determination coefficient of regression R^2 is larger than 0.95, the data is available. In order to maintain pristine site conditions, a trestle was set up near the study location. In all, we measured CO_2 , CH_4 , N_2O emissions, and air and soil temperatures.

2.3. Statistics

Statistical analysis was done with SPSS and Microsoft Excel for Windows 2000. The normality of variables was checked using $K-S$ Test, SPSS. The results showed that the distribution pattern of CO_2 , CH_4 , N_2O emissions were all normal distribution ($P < 0.05$, 0.01, and 0.01, respectively). Hence parametric tests were performed. One-way ANOVA was used to test for significance. The Spearman Correlation was performed to analyse correlations between gaseous fluxes and environmental parameters. Figures were drawn using Microsoft excel and Origin 7.5 software. In all analysis where $P < 0.05$, the compare and correlate tests were considered statistically significant.

3. Results and discussion

3.1. Seasonal variation of freeze–thaw characteristics in freshwater marshes

The deepest frozen soil appeared in January and February; then surface snow and ice began to thaw at 13:00–14:30 each day at the end of March. In April, surface of soil thawed with the increasing temperature, and it completely thaws up in July. The same pattern of freeze–thaw in this area occurred every year with little difference (Fig. 1). Seasonal freeze–thaw is an important event in northern China, and it impacts the evolution process of different ecosystem, controls the soil development, and affects the chemistry process of mass and energy cycle in the global ecosystem.

3.2. Temperature and precipitation

The soil temperatures at the study site were measured at a depth of 5 cm (Fig. 2). The graphic shown in Fig. 2 illustrates that the soil temperature increased dramatically, immediately following the snow–thaw period. The thawed

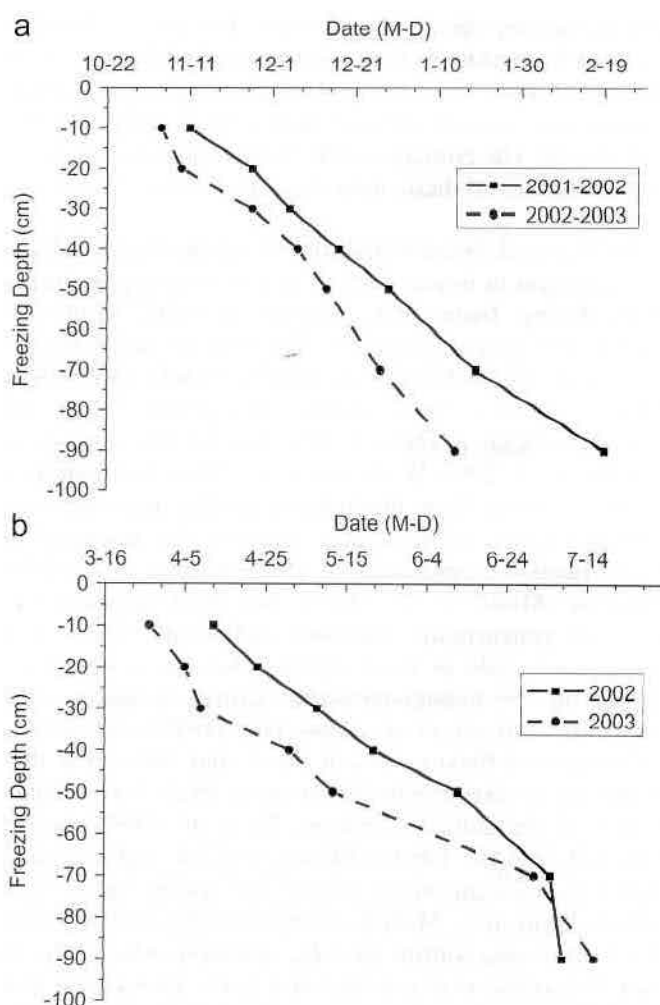


Fig. 1. (a) The seasonal freezing and thawing curve of the freshwater marsh soil is the freezing curve through soil profile from November to March of next year. (b) The thawing curve through soil profile from the end of March–July.

depth reached 5 cm by mid-April, at which time the soil temperature rapidly increased and reached its peak in mid-July. From April to mid-July, the temperature increased by roughly 30°C . After the temperature peaked, it decreased again at the end of September. By early November, the topsoil was frozen, and was up to its maximum depth by the following February. There were no significant differences in the soil temperatures among the 3 years ($P > 0.05$).

Likewise, there were no significant precipitation differences across the 4-year study period, including during the growing season ($P > 0.05$; Fig. 2). However, the precipitation in winter (November–March) of 2004 was higher than that measured in 2002, 2003 and 2005 ($P < 0.0001$).

There were significant differences in water depth during the freeze–thaw period (April–June) across 2003, 2004 and 2005 ($P < 0.05$; Fig. 3), but no obvious water depth differences from July to October across 2002, 2003, 2004 and 2005 ($P > 0.05$; Fig. 3).

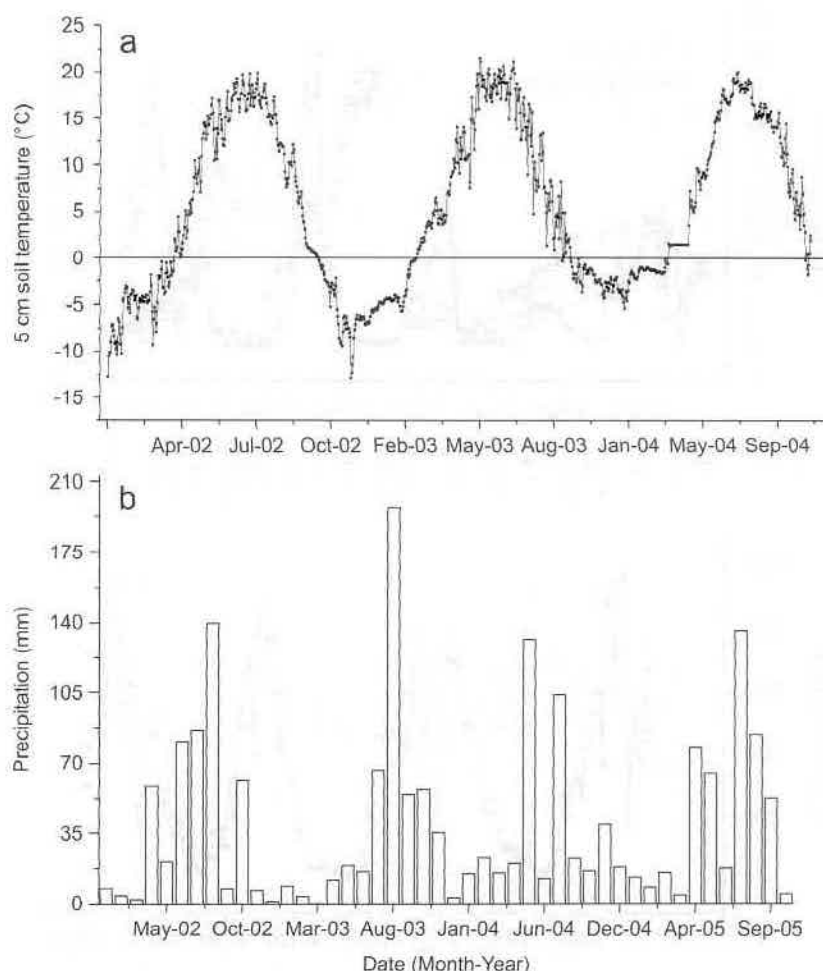


Fig. 2. Soil temperature at 5 cm below the ground surface (a) and monthly precipitation (b) from 2002 to 2005 in the study site.

3.3. The characteristics of CH_4 emission from the freshwater marsh

There was an obvious CH_4 emission event ($0.1\text{--}2.3\text{ mg m}^{-2}\text{ h}^{-1}$) from the freshwater marsh in the winter (Fig. 3), but no significant CH_4 emission rate differences among the winters of 2002, 2003 and 2004. The emission of CH_4 significantly increased during the freeze–thaw period (April–June), with the peak emissions occurring in June (Fig. 3). After the soil thawed completely, the emission of CH_4 reduced gradually at the end of June. The CH_4 flux during the freeze–thaw period in 2004 was much lower than that in 2003 and 2005. In general, the CH_4 flux, which was reduced during the growing season, accounted for roughly 3–5.5% of the total yearly emissions during winter (November–March), and for about 23.5–36.9% of the annual emissions during the freeze–thaw period (April–June). Dise (1993) reported that during winter, the CH_4 emission rate was about 4–21% of the total annual discharge. Melloh and Crill (1996) studied the freshwater wetlands in New Hampshire for 5 years, and found that in winter, the CH_4 discharge from the marsh was roughly 2–9.2% that of

the total yearly emission rate. In the winter, the emission of CH_4 is observable and important in the Sanjiang Plain. The lowest temperatures recorded for the surface soil and the rhizome were approximately -20 and -15 °C, respectively. Panikov and Dedysh (2000) determined that at -16 °C, the bacteria were still active, and CH_4 was emitted at these freezing conditions. The results in the Sanjiang Plain were consistent with their findings. The CH_4 emission rate increased obviously during the thaw–freeze period. This increase was likely caused by the activity of the bacteria, and the higher concentrations of carbon and nitrogen in the soil. The dead bacterial organisms comprise an important foundation for the mass production of greenhouse gases. Moreover, the effect of a freeze–thaw cycle could accelerate the disintegration and mineralization processes of carbon and nitrogen and increase the bacterial activities at the soil surface (Fig. 4).

During the thaw–freeze period that we investigated, the CH_4 flux was negatively correlated with water depth (Fig. 5). At the end of the April, plants budded and grew, but they were diminutive, and their aerenchyma was submerged in water, which made it difficult for CH_4 to

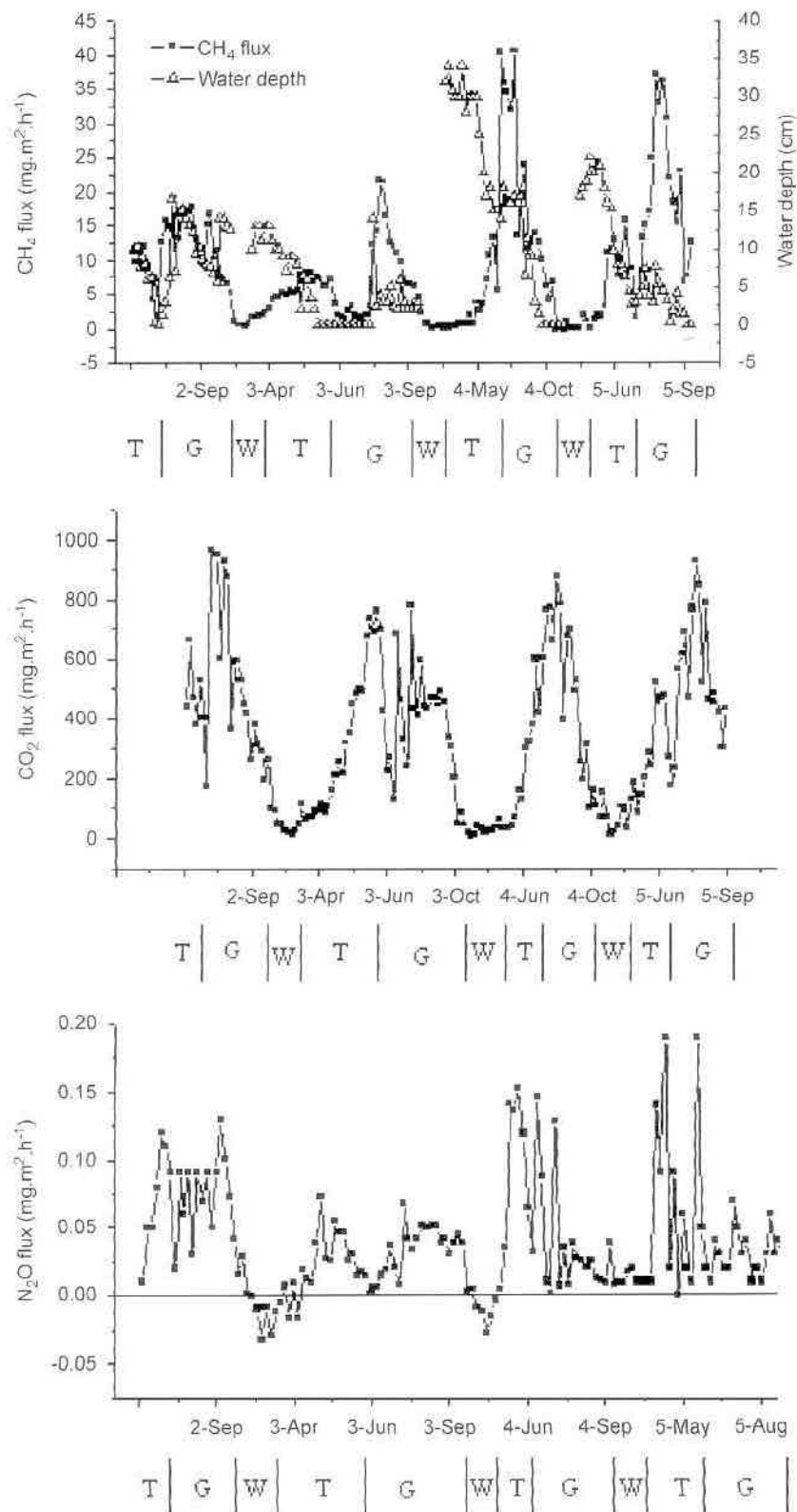


Fig. 3. Emission of CO_2 , CH_4 and N_2O related to water depth from June 2002 to December 2005 in the study site: (W) winter, (T) thaw-freeze period, (G) the growing season. (■) The gas flux and (△) water depth.

be transported from marsh into atmosphere through plant aerenchyma, but via the diffusion of air bubbles only. This situation compromised the transportation of CH_4 from the

marsh into the atmosphere. In this manner, the water depth greatly governed the interannual variation of CH_4 emissions from the marshes during the thaw-freeze period.

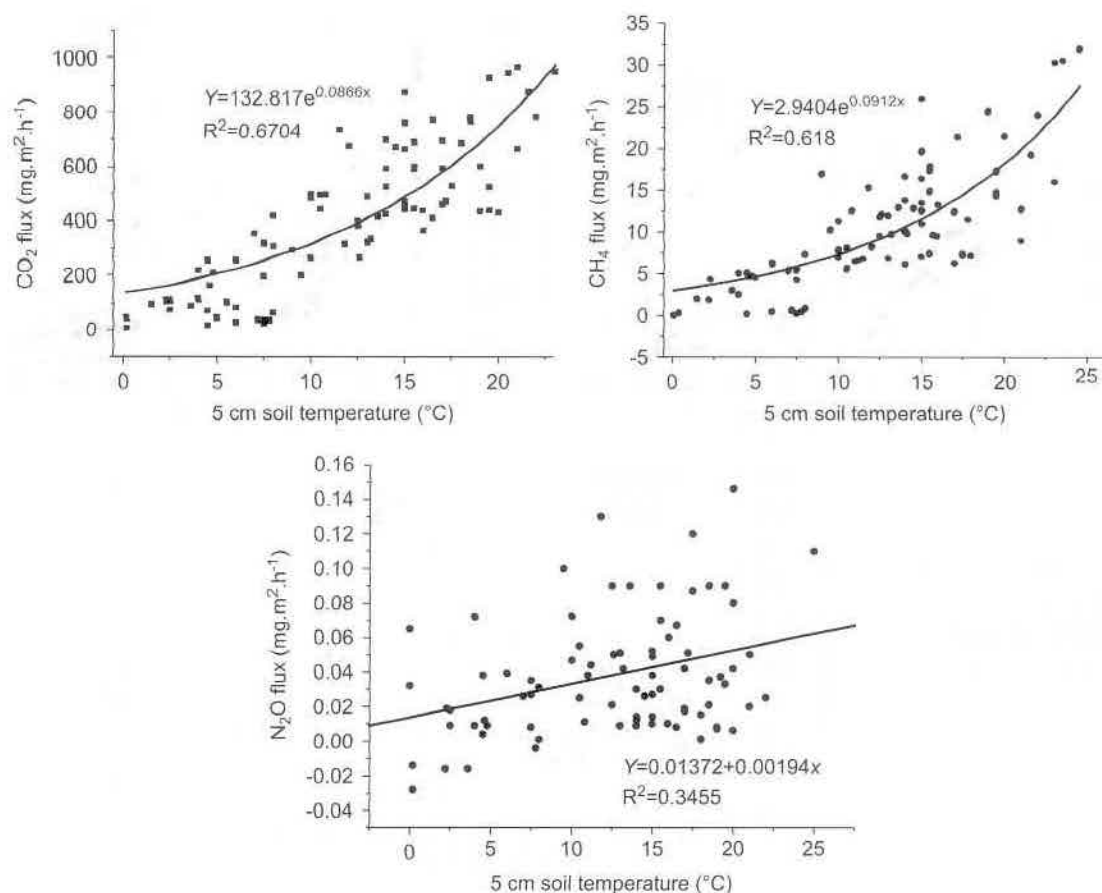


Fig. 4. Relationship between CO₂, CH₄ and N₂O flux and soil temperature.

The CH₄ emission flux varied widely with the seasons (Fig. 3), reaching high levels during July, August and September, which decreased rapidly after this growing season. The CH₄ emission flux peaked at the end of July–August. The maximum monthly mean CH₄ flux was 16.2–23.2 mg m⁻² h⁻¹ in August. The annual variation of CH₄ emissions was significant during the growing season ($P < 0.05$). The CH₄ emission dynamic was closely correlated with soil temperature (below ground 5 cm) during the growing season (Fig. 4). The temperature affected the annual CH₄ production and emission rates indirectly, via plant and microbial activities. However, there were no significant differences in soil temperature across 2002, 2003, 2004 and 2005 ($P > 0.05$). Therefore, soil temperature was not the primary factor inducing the interannual variation of CH₄ during the growing season. Rather, the marsh water depth was closely correlated with CH₄ flux during the growing season in 2003 and 2004 (Fig. 6). The standing water depth greatly governed the annual variation of the regional CH₄ marsh emissions by altering the amount of submerged plant litters. In this manner, the plants provided labile organic carbons for methanogens and controlled CH₄ concentrations in porewater (Ding et al., 2002). However, the relationship between the CH₄ flux and water depth during the growing season in 2003 and 2004 and across all 4 years was poor (Fig. 6).

Furthermore, the aboveground plant biomass and water depth during the growing season were not significantly different in 2002, 2003, 2004 and 2005. Therefore, the water depth was not the primary factor affecting the interannual variation of CH₄ discharge during the growing season. By examining the data graphed in Fig. 3, we found that higher water depths during the freeze–thaw period produced larger CH₄ fluctuations during the growing season. Ding (2003) determined that the concentrations of acetic acid and labile organic C (LOC) were the highest in June and the lowest in July in the same study site. Zhang et al. (2005) also found high DOC concentrations in May and the lowest DOC concentrations in July. These high DOC concentrations likely correlate with the higher water depth that engulfed the plants in May or June. Therefore, the interannual variation of CH₄ marsh emissions in the growing season mainly resulted from the difference in water depth during freeze–thaw period.

3.4. The characteristics of CO₂ emission in freshwater marshes

As with CH₄, the CO₂ emissions varied with the changing seasons (Fig. 3). The CO₂ secretions were associated mainly with the air temperature and water depth, but not with soil temperature. The emission flux

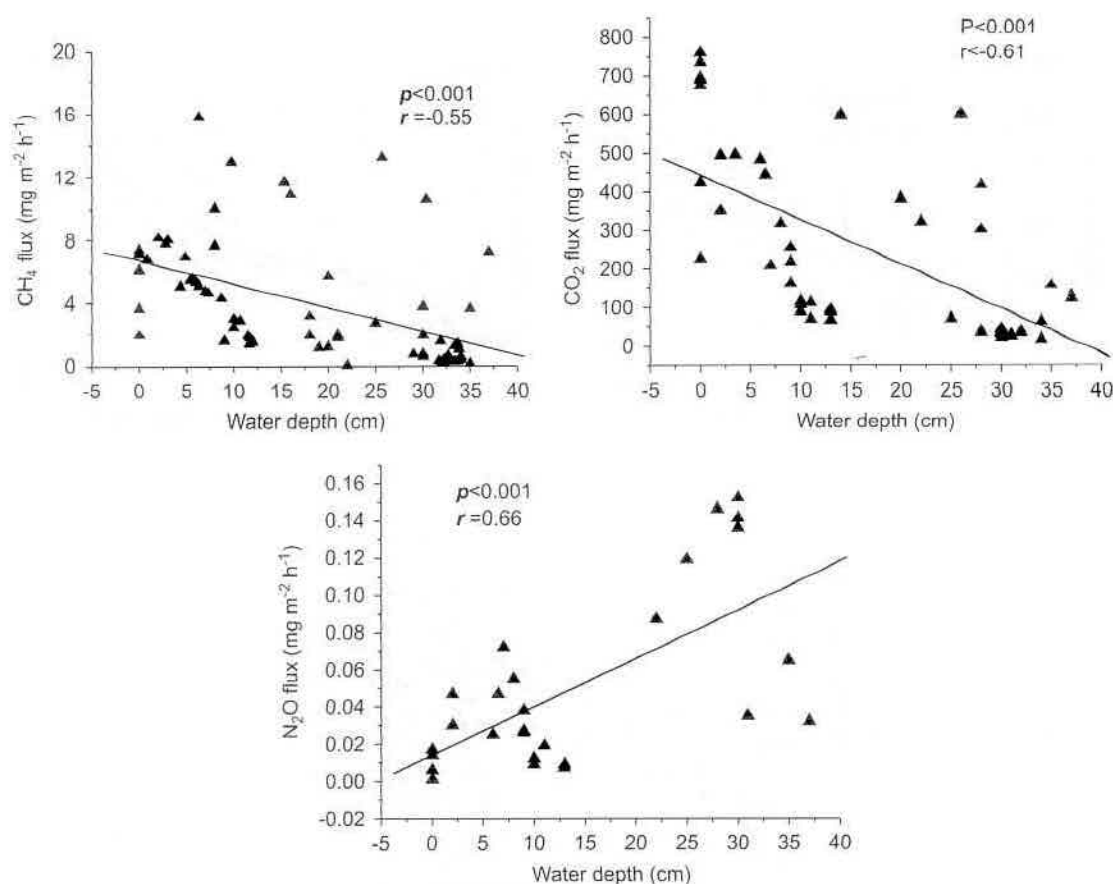


Fig. 5. Relationship between CO_2 , CH_4 and N_2O flux and water depth during the thaw–freeze period, from June 2002 to December 2005 in the study site.

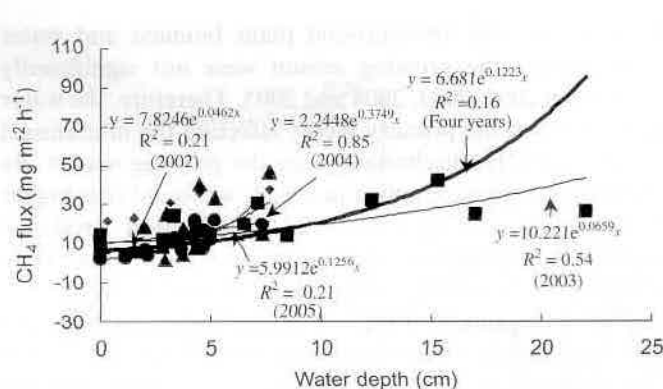


Fig. 6. Relationship between CH_4 flux and water depth during the growing season in (▲) 2002, (■) 2003, (●) 2004, (□) 2005, and among the 4 years.

peaks appeared in July and August, and then decreased rapidly after October, during the growth cessation period. The graph displayed in Fig. 7 revealed the close relationship between CO_2 and CH_4 emissions, implying that there are some common factors controlling the discharge of these two gases. As shown in Fig. 3, water depth during thaw–freeze period appears to be among the main factors affecting CO_2 and CH_4 emissions during the growing season. This parameter may thus influence the strong correlative relationship depicted in Fig. 7.

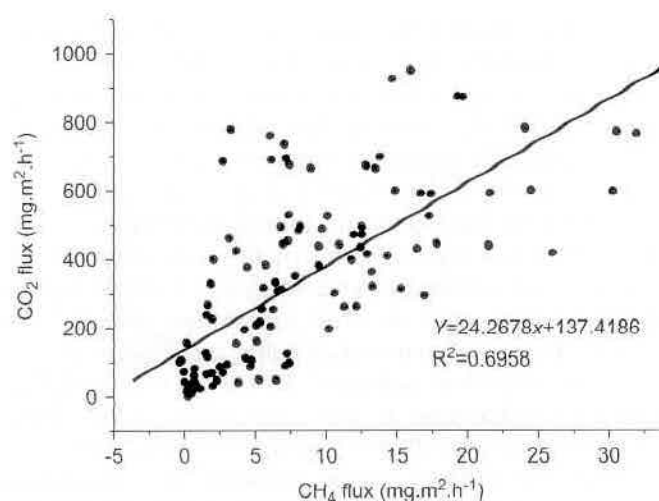


Fig. 7. Relationship between CH_4 and CO_2 emissions.

We measured no significant differences in CO_2 emissions during the winters of 2002, 2003 and 2004 ($P > 0.05$). Freshwater marshes usually emit CO_2 during winter either because CO_2 is stored by mesophilic microbes during the warm period, or CO_2 is produced by psychrophilic microorganisms in the colder months. Panikov and Dedysh (2000) support the first mechanism, whereby CO_2 is stored

in the water during the warm months, and emission increases rapidly during the freeze–thaw period, partly due to a sudden release of the ice-encapsulated CO_2 (Wickland et al., 2001). This release simultaneously stimulates the activity of bacteria, and increases the concentration of available nutrients (Brooks et al., 1998; Schimel and Klein, 1996; Elberling and Brandt, 2003). This in turn speeds the growth of perennial vegetation, enabling a rapid emission of CO_2 (Fig. 3). Similar to CH_4 , the CO_2 emissions during the freeze–thaw period in 2004 were much lower than those measured in 2003 and 2005. During the thaw–freeze period, the CO_2 flux was negatively correlated with water depth (Fig. 5). Hence, water depth greatly governed the interannual variation of CO_2 emissions from the marshes during this period.

3.5. N_2O emission in freshwater marsh

The secretion of N_2O from the soil occurs during nitrification and denitrification, and thus is controlled by the soil temperature and nitrogen availability. During the winter of 2002 and 2003, we measured a N_2O sink in the Sanjiang Plain. There was no significant difference between N_2O emissions in the winters of 2002 and 2003 ($P > 0.05$). However, there was an obvious and significant emission of N_2O during the winter of 2004. Liu et al. (2001) also determined that the wetland soil absorbed atmospheric N_2O at the same site. In contrast, Nykänen et al. (1995) reported that N_2O emissions peaked on the grassland during the onset of winter, and Alm et al. (1999a,b) reported that winter N_2O fluxes were negligible at all virgin sites and drained bogs, but that N_2O emissions on efficiently drained fens were 27–28% of their annual rates. The mechanisms of winter N_2O flux are unclear and deserve focused research attention.

During the freeze–thaw period, the marsh was a source of N_2O as the temperature increased (Fig. 3), especially in May. The N_2O discharge during the freeze–thaw period in 2004 was obviously higher than in 2003. During the thaw–freeze period, the N_2O level was positively correlated with the water depth (Fig. 5). Moreover, the water depth greatly governed the interannual variation of N_2O emissions from the marshes during this time. N_2O production, primarily through nitrification and denitrification processes, may comprise a substantial fraction of the total nitrogen lost from peatlands (Jacks et al., 1994). Soil moisture was an important factor controlling of N_2O emissions (Davidson et al. 2000; Neill et al., 2005). Neill et al. (2005) reported that forest soils did not reach water-filled pore space (WFPS) (<30%) sufficient to eliminate nitrification and that pasture soils did reach WFPS sufficient to promote denitrification. Because pasture soils were on average wetter than forest soils, denitrification was likely a relatively more important source of N_2O from pastures than in forests (Neill et al., 2005). According to the studies of recent years, nitrate was source of N_2O during the freeze–thaw period (Kammann

et al., 1998; Brumme et al., 1999), which resulted from the reduced oxygen supply from the soil interspace variations, the high soil water content and the high active concentrations of carbon and nitrogen. The WFPS >75% (not publication) and redox potential (Eh) variation between –100 and –150 mv in the 5 cm below ground (Ding et al., 2004), which was an anaerobic condition in the study site. High WFPS eliminated nitrification and promoted denitrification in the freshwater marsh, where N_2O production by denitrification predominated.

The field measurements and indoor experiments all showed that the activity of the bacteria was a main factor controlling N_2O emission dynamics. Inubushi et al. (2003) reported that N_2O levels were somewhat negatively correlated with the precipitation events in the tropical peatlands. The corresponding R^2 values suggested that the variation in the soil temperature and water depth could explain approximately only 35% and 44%, respectively. This suggested that other factors also simultaneously regulated the seasonal N_2O emission. Therefore, there was stochastic variability of the temporal emission pattern of N_2O .

4. Conclusions

We recorded obvious CH_4 and CO_2 emissions from the freshwater marsh during the winter months. There was close relationship between the emission rates of CO_2 and CH_4 , implying there were some common factors controlling their secretion dynamics. Water depth greatly governed the interannual variation of the CH_4 and CO_2 emissions during the thaw–freeze period. The interannual variation of CH_4 and CO_2 emissions from the marshes was significant during the growing season ($P < 0.05$). The interannual variation of the CH_4 and CO_2 emissions from marsh in the growing season mainly resulted from the difference in water depth during freeze–thaw period. The freshwater marsh in Sanjiang Plain of northeast China was a sink of N_2O in winter of 2002 and 2003. However, there was obvious emission of N_2O in winter of 2004. During freeze–thaw period, it was a source of N_2O with the temperature increasing, especially in May the emission of N_2O increased obviously. The water depth variable greatly governed the interannual variation of N_2O emissions from marshes during thaw–freeze period.

Acknowledgements

This work was founded by National Natural Science Foundation of China (40471124, 40431001), Chinese Academy of Sciences (KZCX2-YW-309).

References

- Alm, J., Saarnio, S., Nykänen, H., Silvola, J., Martikainen, P.J., 1999a. Winter CH_4 , CO_2 and N_2O fluxes on some natural and drained boreal peatlands. *Biogeochemistry* 44, 163–186.

- Alm, J., Schulman, L., Walden, J., Nykanen, H., et al., 1999b. Carbon balance of a boreal bog during a year with an exceptionally dry summer. *Ecology* 80 (1), 161–174.
- Barklett, K.B., Harriss, R.C., 1993. Review and assessment of methane flux from wetlands. *Chemosphere* 26 (4), 261–320.
- Baron, J.S., Hartman, M.D., Band, L.E., et al., 2000. Sensitivity of a high-elevation Rocky Mountain watershed to altered climate and CO₂. *Water Resources Research* 36 (1), 89–99.
- Brix, H., 1990. Gas exchange through the soil–atmosphere interphase and through dead culms of *Phragmites australis* in a constructed reed bed receiving domestic sewage. *Water Research* 24, 259–266.
- Brooks, P.D., Williams, M.W., Schmidt, S.K., 1998. Inorganic nitrogen and microbial biomass dynamics before and during spring snowmelt. *Biogeochemistry* 43, 1–15.
- Brumme, R., Borken, W., Finke, S., 1999. Hierarchical control on nitrous oxide emissions in forest ecosystems. *Global Biogeochemical Cycles* 13, 1137–1148.
- Cicerone, R.J., Oremland, R.S., 1988. Biogeochemical aspects of atmospheric methane. *Global Biogeochemical Cycles* 2 (4), 299–327.
- Ding, W.X., 2003. Mechanism of methane emission from the mire and cultivated mire soils. A Thesis Submitted to Graduate School, Chinese Academy of Sciences in Partial fulfillment of the Requirement for the Degree of Ph.D. p. 104.
- Ding, W.X., Cai, Z.C., Tsuruta, H., Li, X.P., 2002. Effect of standing water depth on methane emissions from freshwater marshes in northeast China. *Atmospheric Environment* 36 (33), 5149–5157.
- Ding, W.X., Cai, Z.C., Tsuruta, H., 2004. Diel variation in methane emissions from the stands of *Carex lasiocarpa* and *Deyeuxia angustifolia* in a cool temperate freshwater marsh. *Atmospheric Environment* 38, 181–188.
- Disc, N.B., 1993. Methane emission from Minnesota peatlands, spatial and seasonal variability. *Global Biogeochemical Cycles* 7 (1), 123–142.
- Elberling, B., Brandt, K.K., 2003. Uncoupling of microbial CO₂ production and CO₂ release in frozen soil and its implications for field studies of arctic C cycling. *Soil Biology & Biochemistry* 35, 263–272.
- Górham, E., 1991. Northern peatlands: role in the carbon cycle and probable responses to climatic warming. *Ecological Applications* 52, 182–195.
- Grogan, P., Michelsen, A., Ambus, P., et al., 2004. Freeze–thaw regime effects on carbon and nitrogen dynamics in sub-arctic heath tundra mesocosms. *Soil Biology & Biochemistry* 36, 641–654.
- Houghton, J.T., 1995. IPCC, climate change 1994: radiative forcing of climate change and an evaluation of the IPCC IS92 emission scenarios. Cambridge University Press, Cambridge UK, p. 337.
- Inubushi, K., Furukawa, Y., Hadi, A., Purnomo, E., Tsuruta, H., 2003. Seasonal changes of CO₂, CH₄ and N₂O fluxes in relation to land-use change in tropical peatlands located in coastal area of South Kalimantan. *Chemosphere* 52, 603–608.
- Jacks, G., Joëlsson, A., Fleische, r.S., 1994. Nitrogen retention in forest wetlands. *Ambio* 23, 358–362.
- Jin, H.J., Wu, J., Chen, G.D., Nakano, T., Sun, G.Y., 1999. Methane emissions from wetlands on Qinghai–Tibet Plateau. *Chinese Science Bulletin* 44, 2282–2287.
- Kammann, C., Grunhage, L., Muller, C., et al., 1998. Seasonal variability and mitigation options for N₂O emissions from different managed grasslands. *Environmental Pollution* 102, 179–186.
- Lafleur, P.M., Roulet, N.T., Admiral, S.W., 2001. Annual cycle of CO₂ exchange at a bog peatland. *Journal of Geophysical Research* 106 (3), 3071–3081.
- Liu, J.S., Wang, J.D., Li, Z.G., Yu, J.B., Zhang, X.L., Wang, C.M., Wang, Y., 2001. N₂O concentration and its emission characteristics in Sanjiang Plain wetland. *Environmental Science* 24, 33–39.
- Martikainen, P.J., Nykänen, H., Crill, P., Silvola, J., 1993. Effect of a lowered water table on nitrous oxide fluxes from northern peatlands. *Nature* 366, 51–53.
- Melloh, R.A., Crill, P.M., 1996. Winter methane dynamics in a temperate peatland. *Global Biogeochemical Cycles* 10 (2), 247–254.
- Muller, C., Martin, M., Stevens, R.J., et al., 2002. Processes leading to N₂O emissions in grassland soil during freezing and thawing. *Soil Biology & Biochemistry* 34, 1325–1331.
- Neill, C., Paul, A.S., Diana, C.G., Jerry, M.M., Brigitte, J.F., Marisa, C.P., Carlos, C.C., 2005. Rates and controls of nitrous oxide and nitric oxide emissions following conversion of forest to pasture in Rondônia, Nutrient Cycling in Agroecosystems 71, 1–15.
- Nykänen, H., Alm, J., Lång, K., Silvola, J., Martikainen, P.J., 1995. Emissions of CH₄, CO₂ and N₂O from a virgin fen and a fen drained for grassland in Finland. *Journal of Biogeography* 22, 351–357.
- Panikov, N.S., Dedysh, S.N., 2000. Cold season CH₄ and CO₂ emission from boreal peat bogs (West Siberia), winter fluxes and thaw activation dynamics. *Global Biogeochemical Cycles* 14, 1071–1080.
- Schimel, J.P., Clein, J.S., 1996. Microbial response to freeze–thaw cycles in tundra and taiga soils. *Soil Biology & Biochemistry* 28, 1061–1066.
- Song, C.C., Yan, B.X., Wang, Y.S., Wang, Y.Y., Lou, Y.J., Zhao, Z.C., 2003. Fluxes of carbon dioxide and methane from swamp and impact factors in Sanjiang Plain, China. *Chinese Science Bulletin* 48, 2749–2753.
- Wickland, K.P., Striegl, R.G., Mast, M.A., et al., 2001. Carbon gas exchange at a southern Rocky Mountain wetland, 1996–1998. *Global Biogeochemical Cycles* 15, 321–335.
- Zhang, J.B., Song, C.C., Yan, W.Y., 2005. Seasonal dynamics of dissolved organic carbon and its impact factors in the *Deyeuxia angustifolia* marsh soil. *Acta Scientiae Circumstantiae* 25 (10), 141–146 (in Chinese).

Sustainability of wastewater treatment technologies

Helen E. Muga, James R. Mihelcic*

Department of Civil and Environmental Engineering, Sustainable Futures Institute, Michigan Technological University, Houghton, MI 49931, USA

Received 8 September 2006; received in revised form 15 February 2007; accepted 11 March 2007

Available online 27 April 2007

Abstract

A set of indicators that incorporate environmental, societal, and economic sustainability were developed and used to investigate the sustainability of different wastewater treatment technologies, for plant capacities of <5 million gallons per day (MGD) or 18.9×10^3 cubic meters (m^3/day). The technologies evaluated were mechanical (i.e., activated sludge with secondary treatment), lagoon (facultative, anaerobic, and aerobic), and land treatment systems (e.g., slow rate irrigation, rapid infiltration, and overland flow). The economic indicators selected were capital, operation and management, and user costs because they determine the economic affordability of a particular technology to a community. Environmental indicators include energy use, because it indirectly measures resource utilization, and performance of the technology in removing conventional wastewater constituents such as biochemical oxygen demand, ammonia nitrogen, phosphorus, and pathogens. These indicators also determine the reuse potential of the treated wastewater. Societal indicators capture cultural acceptance of the technology through public participation and also measure whether there is improvement in the community from the specific technology through increased job opportunities, better education, or an improved local environment. While selection of a set of indicators is dependent on the geographic and demographic context of a particular community, the overall results of this study show that there are varying degrees of sustainability with each treatment technology.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Wastewater treatment; Activated sludge; Lagoon; Land treatment; Sustainable development; Indicators; Design

1. Introduction

Improvement in global health and sanitation and consequent reduction in the spread of disease depends largely on good hygiene practices, availability of health facilities, and reliable collection and treatment of wastewater. The World Health Organization (WHO) estimates that 2.4 billion people lack access to any type of sanitation equipment (WHO and UN Children's Fund, 2000).

Wastewater collection systems (i.e., sewer networks) and centralized and decentralized treatment systems are designed and managed primarily to protect human and environmental health. Though their benefits are widely recognized, there are other aspects of this infrastructure and associated technologies that are not so obvious and hence less acknowledged, yet they impact communities and the surrounding environment. For example a positive

aspect of the sewer network is the collection and transport of wastewater to appropriate treatment facilities, whereby pathogens and chemical constituents such as oxygen-depleting organic matter and phosphorus are removed before the treated water is returned to the environment. A negative aspect of such a network is that it can create an imbalance in water and nutrient fluxes and therefore distort natural hydrological and ecological regimes. For instance the discharge of large volumes of treated wastewater that contains low concentrations of chemical constituents may still lead to an excessive input of nutrients in a receiving water body, thus, leading to a water quality problem.

Transport of water and wastewater across watershed boundaries not only increases the embodied energy of a material and requires extensive infrastructure needs, but it may also result in adverse changes in an ecosystem's hydrology. In addition, treatment facilities, while they treat wastewater to a quality deemed safe for discharge, also consume considerable energy during their operational life, and consequently contribute to atmospheric carbon

*Corresponding author. Tel.: +1 906 487 2324; fax: +1 906 487 2943.
E-mail address: jm41@mtu.edu (J.R. Mihelcic).

dioxide emissions. These impacts, whether positive or negative, greatly affect local and global sustainability, be it in the construction, operation, or dismantling life stage and hence deserve discussion. In an era where there is growing concern of the local and global impact of our current environmental management strategies, and the need to reduce sanitation problems, disease, and poverty, there is a greater need to develop more environmentally responsible, appropriate wastewater treatment technologies whose performance is balanced by environmental, economic, and societal sustainability.

While there is no consensus on the definition of sustainability, what is clear is that it strives for the maintenance of economic well being, protection of the environment and prudent use of natural resources, and equitable social progress which recognizes the just needs of all individuals, communities, and the environment. Furthermore, it recognizes the need to design human and industrial systems that ensure humankind's use of natural resources and cycles do not lead to diminished quality of life due to either losses in future economic opportunities or adverse impacts on social conditions, human health and the environment (Mihelcic et al., 2003).

Although collection systems are important to overall wastewater management, this study is limited to wastewater treatment. Thus in light of the main aspects of sustainability, questions that deserve further analysis are how selection of a particular wastewater treatment technology affects overall sustainability and are there certain aspects of a particular treatment technology that makes it more balanced in terms of economic, environmental, and social sustainability?

Sustainability of wastewater treatment systems can be assessed through different assessment tools such as exergy analysis, economic analysis, and life cycle assessment (LCA). For this study, the use of a balanced set of indicators that provides a holistic assessment was chosen for evaluating the sustainability of the different wastewater treatment technologies. These wastewater treatment technologies include mechanical systems, lagoons systems, and land treatment systems. Mechanical systems such as activated sludge utilize physical, chemical and biological mechanisms to remove nutrients, pathogens, metals and other toxic compounds. Lagoon systems primarily use physical and biological processes to treat wastewater, while land treatment systems utilize soil and plants, without significant need for reactors and operational labor, energy, and chemicals (Metcalf and Eddy, 2003).

Selection of a particular set of indicators may vary from community to community depending on geography, culture, and population served. Several lists of sustainability indicators have been proposed to assess wastewater management and wastewater treatment technologies (Nilsson and Bergstrom, 1995; Balkema, 1998; Lundin et al., 1999; Balkema et al., 2002). These studies use a comprehensive, multi-disciplinary set of indicators; however, they have focused on evaluating one treatment technology and

have not compared different treatment technologies as proposed in this study. Furthermore, these studies were one dimensional in terms of evaluating sustainability because they only evaluated environmental stressors. Other studies measuring environmental and/or economic issues associated with wastewater treatment (Emmerson et al., 1995; Nilsson and Bergstrom, 1995; Hwang and Hanaki, 2000; Tsagarakis et al., 2002; Dixon et al., 2003) do not address societal issues; thus, they do not fully capture the overall sustainability that should be inclusive of a balance of economic, environmental and social considerations.

Selection of a particular wastewater treatment technology should not be based primarily on technical insight, but should also integrate the human and environmental activities that surround it. Accordingly, this paper initiates a discussion on this topic by first developing a set of indicators and then evaluating the environmental, economic, and societal impacts of three different wastewater treatment technologies that treat less than 5 million gallons per day (MGD) ($18.9 \times 10^3 \text{ m}^3$).

We acknowledge that the selection of indicators and evaluation of particular technologies in this paper is dependent on the geographic and demographic particulars of a community. Here, the results pertain to small communities ($< 5 \text{ MGD}$ or $18.9 \times 10^3 \text{ m}^3$) who wish to integrate the selection of an economically and environmentally viable wastewater treatment technology with social attributes such as workforce education level, employment, and open space; while also being concerned with localized warming from a built up infrastructure. The delicate balance associated with these attributes differs by community, region, and country. For example, higher labor costs may prohibit employment of additional workforce at a treatment plant and urban heat island stresses may not impact communities located in colder regions. Furthermore, this paper does not advocate analyzing a treatment technology only through use of sustainability indicators. Other tools such as LCA exist and can be utilized. This paper merely attempts to initiate a discussion on how to address a more integrated evaluation of the overall sustainability of wastewater treatment technologies.

2. Methodology

There currently exists a plethora of different frameworks to select sustainability indicators. Commonly used frameworks include those developed by the Organization for Economic Co-Operation and Development (OECD, 1998) and the United Nations Commission on Sustainable Development (UNCSD, 1996). However, some problems associated with selection and application of indicators exist, for example, indicators produced by one group may not be applicable to another, thus they may not be widely implemented. Geographical diversity of urban, rural and peri-urban areas may mean that an indicator appropriate to one locality may be inappropriate to another (Mitchell, 1996). Other reasons include difficulty of data availability (Bell and Morse, 2004; Mitchell, 1996), the numerous

methodologies available for deriving sets of sustainability indicators (Bell and Morse, 2004) and lack of consensus among peers regarding the definition of sustainability and indicators (Spangenberg, 2002).

In any case an indicator should refer to specific targets chosen, be able to indicate the success or lack of it in approaching them, and be sensitive and robust in their construction. Indicators should also be easy to fathom and be limited in number (UNDPSCD, 1995). Moreover, indicators should be linked to a target or goal that is ecologically sustainable (Meadows, 1998) and be able to “stand the test of time.”

The first step in this process was deciding on a suitable treatment plant capacity, over which all three different types of technologies exist. In determining the most appropriate plant capacity to perform this study, 16,255 US wastewater treatment plants were reviewed (U.S. EPA, 2004a). Over 80% of existing plants have a capacity less than 5 MGD ($18.9 \times 10^3 \text{ m}^3$) and this was the size range selected for this study.

The set of indicators selected were based on: (1) the relevancy of the indicators to different wastewater treatment technologies, and (2) their ability to indicate progress towards balanced sustainability or away from it, that is equal inclusion of economic, environmental and social aspects. The United Nations Department of Policy Coordination and Sustainable Development's criteria (UNDPSCD, 1995) were used to select the most appropriate indicators. Indicators in general should be (i) based on a sound scientific basis and widely acknowledged by scientific community; (ii) transparent, e.g., their selection calculation and meaning must be obvious even to non-experts; (iii) relevant, e.g., they must cover crucial aspects of sustainable development; (iv) quantifiable, e.g., they should be based on existing data and/or data that is easy to gather and to update; and, (v) limited in number according to their purposes they are being used for (UNDPSCD, 1995).

From the above criteria a set indicators useful for evaluating wastewater treatment technologies were selected. They are presented in Table 1 along with their unit of measure. These indicators measure the economic, environmental and social sustainability of the treatment system; however, readers should understand that the selection and interpretation of these indicators is context specific and based on opinions of the authors. Data related to each indicator were then obtained from sources that included government, professional societies, and academic textbooks.

A final consideration is the life stage over which the indicators are applied. The operational life stage of wastewater treatment was chosen for this study, primarily because of its length of time relative to the other life stages and its environmental impacts are believed to be greater than those during construction and end-of-life stages (Emmerson et al., 1995). For example, the operational stage has the highest energy consumption (95%) compared

Table 1

Set of indicators developed in this study for assessing sustainability of wastewater treatment technologies

Indicator	Unit of measure
Economic	
Capital costs	\$/GPD
Operation and management	\$/MGD
User cost	\$/month
Environmental	
Energy use	kWh/MGD (kWh/m ³)
Biochemical oxygen demand (BOD)	% Removal
Total suspended solids (TSS)	% Removal
Nitrogen (NH ₃)	% Removal
Phosphorus	% Removal
Pathogens	% Removal
Societal	
Public participation (in selecting treatment technology)	Qualitative measure
Community size served	Population/MGD
Aesthetics	Measured level of nuisance from odor
Staffing required to operate plant	Staff/MGD
Level of education	Operational requirements (operator license)
Open space availability	Acre/MGD

Selection and interpretation of indicators will be based on the geographic and demographic context of a particular setting.

to construction and refurbishment/demolition stages (Emmerson et al., 1995). This is significant because energy production and its use are associated with a large number of environmental problems including release of airborne pollutants associated with global warming and acidification. Moreover, environmental impacts from the construction stage account for only 5% of the total environmental impact (Emmerson et al., 1995).

3. Results and discussion

3.1. Economic sustainability

3.1.1. Capital costs

Fig. 1a shows comparable capital costs for the different wastewater treatment technologies. These costs are based on a gallon per day (GPD) or cubic meters per day (m³/day) of wastewater treated. Due to unavailability of data for higher capacities (i.e., 1–5 MGD) (3.8×10^3 – $18.9 \times 10^3 \text{ m}^3$), data were only available for plant capacities ranging from 0.1–1 MGD (3.8×10^2 – $3.8 \times 10^3 \text{ m}^3$). Fig. 1a shows there is significant difference (\$10.50/GPD versus \$4/GPD) (\$2770/m³ per day versus \$1060/m³ per day) in capital construction costs between a more costly mechanical and less costly lagoon system at the higher expected range. At the lower range of cost, there is a much smaller difference (\$3/GPD versus \$1/GPD) (\$790/m³ per day versus \$260/m³ per day) between both technologies (UNEP, 1997); however, mechanical systems are still more expensive.

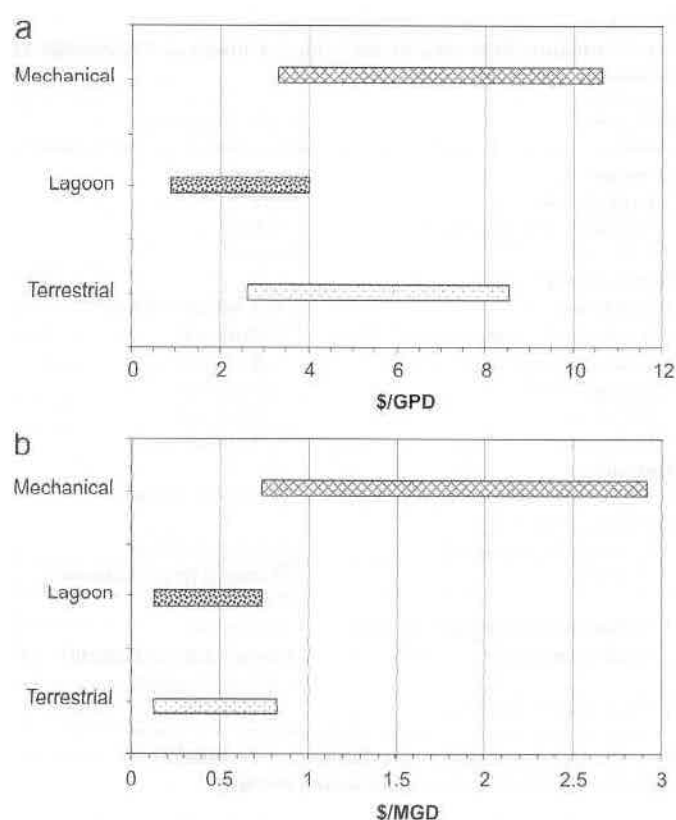


Fig. 1. (a) Comparative capital cost for the different wastewater treatment technologies. Plant size is 0.1–1.0 MGD (3.8×10^2 – 3.8×10^3 m³/day) (UNEP, 1997). (b) Comparative operations and management cost for the different wastewater treatment technologies. Plant size is 0.1–1.0 MGD (3.8×10^2 – 3.8×10^3 m³/day) (UNEP, 1997).

3.1.2. Operation and management

Operating and maintenance costs associated with wastewater treatment include labor, energy, and purchase of chemicals and replacement equipment. Fig. 1b shows this cost for mechanical treatment is approximately 4–5.5 times higher than a lagoon system and 4–6.5 times greater than a land treatment system. This can be attributed to more highly mechanized equipment and complex processes that require considerable energy inputs.

3.1.3. User costs

Wastewater treatment costs are usually dependent upon the type of treatment technology, its efficiency, and the discharge option used. Another factor is the population size served. About 70% of the US population is serviced by secondary or greater treatment level (U.S. EPA, 2004a). A majority of this population reside in urban areas.

The cost of wastewater treatment for a resident of a smaller population serviced by a treatment plant capacity of <9.8 MGD ($<37 \times 10^3$ m³), is much higher (\$0.10–\$1.24) than a resident of a larger population serviced by a larger plant capacity (>187 MGD) ($>708 \times 10^3$ m³) using similar treatment systems (Raftelis Financial Consulting, 2004), even when normalized on a per capita

basis. These costs can translate to affordability problems of residents in smaller communities, especially low-income residents (income of \$15,000 per annum), which may be evident through late or nonpayment of bills, disconnect notices, and service terminations.

Affordability problems are more pronounced in developing countries with expanding economies, where the average annual household income is well below US\$300, and the cost for conventional treatment (mechanical systems) may range from US\$300–1000. Cost-effective treatment technologies such as lagoon and land treatment systems have the potential to reduce these costs by at least one-half (reduced costs to the order of US\$100 per household) (Helmer and Hespanhol, 1997). These alternatives are beneficial in terms of money saved that could be spent elsewhere, for example on health care, education, or other environmental initiatives. As for developing countries, this monetary savings could be spent on improvements in the water supply with further benefits of reduced illnesses and hence cost savings in health care and enhanced economic activity (WHO, 2005).

3.2. Environmental sustainability

3.2.1. Energy use

A majority of operation and maintenance costs may be attributed to energy consumption during aeration and pumping of water and solids. Fig. 2 shows the energy requirements of a combination of various wastewater treatment technologies, excluding the use of methane gas from digesters as an energy source. The data clearly show that activated sludge requires more energy than either lagoon or land treatment systems. Aeration of lagoon systems may increase energy use considerably, by 3 times for a 0.1 MGD (3.8×10^2 m³) plant capacity, and 5 times for a 5 MGD (18.9×10^3 m³) plant capacity (Middlebrooks and Middlebrooks, 1979).

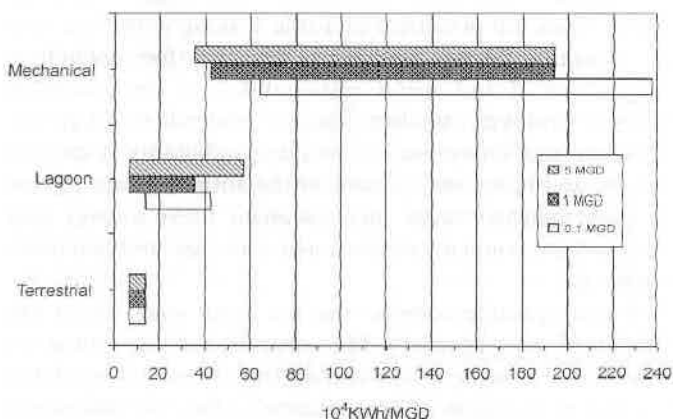


Fig. 2. Total energy requirements for various sizes and types of wastewater treatment plants located in the intermountain areas of the US. Total electricity requirements measured in kWh/MGD at various flow rates (Middlebrooks and Middlebrooks, 1979).

Some activated sludge systems may have lower energy consumption because of internal energy combustion of methane gas produced in-house, particularly from anaerobic digestion. In future studies, energy production related to waste management could be included as an attribute of a sustainable technology.

Energy use is often associated with global environmental problems such as carbon dioxide emissions. For example, an activated sludge system serving a population of 1000 people (<1 MGD or $<3.8 \times 10^3 \text{ m}^3$) has the potential to produce up to 1400 ton of carbon dioxide for operation and 50 ton of carbon dioxide for maintenance over a 15-year life (Emmerson et al., 1995).

Various opportunities exist in reducing energy use and associated impacts. These include the type of wastewater treatment technology selected; use of recycled materials for construction; correct sizing and rating of equipment for operation, especially for pumping requirements; and reuse of waste aggregate from demolition. Plant design can also more carefully incorporate issues of energy conservation and as mentioned previously, use of in-house methane production may reduce external energy needs.

There are also energy issues associated with materials that go into a treatment technology. For example, concrete is the primary component in many mechanical treatment systems and has embodied energy estimated between 70 and 6000 MJ/Mg (Horvath and Hendrickson, 1998). Building-related demolition debris is estimated at around 65 million tons annually in the United States (Franklin Associates, 1998). Two thirds of it by weight and about one half by volume is cement-based concrete (Wilson, 1993). Although mechanical treatment systems use more concrete than lagoon and land treatment systems they can reduce their overall environmental impact if designers consider recovery and reuse of any waste aggregate during the end-of-life stage.

3.2.2. Removal of water quality constituents

The conventional water quality constituents associated with wastewater treatment are: BOD, TSS, phosphorus, nitrogen and fecal coliforms. There are some major differences in removal efficiencies of each treatment technology. Mechanical treatment has high removal efficiency of BOD (90–95%), TSS (90–95%), and fecal coliforms (92–99.99%) but has low removal efficiency of phosphorus (10–20%) and total nitrogen (15–25%). Lagoon treatment has comparable high removal efficiency of TSS (90–95%), and fecal coliforms (90–99.90%) but has medium to high removal efficiency of BOD (75–95%), and low to medium removal efficiency of phosphorus (10–50%) and total nitrogen (10–60%). A land treatment system has a guaranteed high removal efficiency of fecal coliforms (90–99.90%) but has varying efficiencies for the other constituents. It has medium to high removal efficiency for BOD (67–100%), TSS (58–99%), and low to high removal efficiency of phosphorus (40–99%) and total nitrogen (38–95%) (U.S. EPA, 1975; Gilbert, 1976; Leach et al.,

1980; U.S. EPA, 1981, 1982; Pettygrove and Asano, 1984; Hannah et al., 1986; Reed, 1991; Reed et al., 1995; UNEP, 1997; Crites et al., 2000; U.S. EPA, 2002; Metcalf and Eddy, 2003). These effluent qualities ultimately determine if further treatment is required, what type of discharge options can be used and most importantly their potential for reuse.

Removal efficiencies for a lagoon are dependent on whether the lagoon is operating aerobically or facultatively. Aerobic lagoons have detention times of 3–10 days and facultative lagoons, 5–30 days (Reed et al., 1995). While both have high removal efficiencies, the longer detention time of facultative lagoons often leads to higher removal efficiencies of such constituents as nitrogen and pathogens (WEF, 1992). Furthermore, comparison of synthetic organic compound removal in facultative and aerobic lagoons show that a facultative lagoon is capable of removing between 77 and 96% (average of 86%), whereas as an aerobic lagoon can remove between 61 and 80% (average 68%) of the same organic compounds (WEF, 1992). Careful design of a facultative lagoon can reduce the detention time to 10–15 days, while still achieving high removal of pathogens (Oakley, 2005). Properly designed lagoon systems have the capability of removal efficiencies comparable to mechanical systems.

In terms of heavy metal removal, mechanical systems have a comparative removal performance (24–82%) to a lagoon system (32–79%) (Hannah et al., 1986). No data were obtained on metal removal for land treatment systems. In terms of volatile organic chemical removal, the land treatment system had a high removal efficiency of 82–99.99%, while the mechanical system had removal efficiency of 74–94%, and the lagoon system 60–80% (Crites et al., 2000).

Fate and removal of hazardous and toxic compounds is a generally accepted challenge in wastewater treatment and indicators that carry information on emissions or flows of these compounds are often recommended (e.g., Lundin et al., 1999; Malmqvist and Pamlquist, 2005; Palme et al., 2005). If wastewater treatment is to be sustainable in the long-term, better waste management strategies may have to be developed that integrate local economic activity, eliminate the use and improper disposal of household hazardous waste, and partner with health providers that prescribe pharmaceutical chemicals. The use of barriers at various points in wastewater treatment systems can have a tremendous impact in managing hazardous substances. For example, the decision to design a wastewater treatment system with either a combined or separated system (systems that combine gray black water versus systems that separate gray black water at the source, usually in the houses), has the potential to reduce cadmium inputs by 114 g per year (combined system) and 40 g per year (separated system) (Malmqvist and Pamlquist, 2005).

Furthermore, removal efficiencies of pathogens, heavy metals, and other toxic compounds have major implications on water reuse schemes. The specific type of wastewater

reuse initiatives ultimately defines the quality of wastewater required and the subsequent treatment processes needed to achieve this quality. In the US, these reuse schemes such as agricultural irrigation are guided by EPA guideline criteria for wastewater reuse (U.S. EPA, 2005a), which include the control of conventional parameters such as BOD, turbidity, coliforms, pH and residual chlorine, in order to protect public and the environmental health. The WHO has set other guidelines specifically for pathogen control (WHO, 1989). All three systems evaluated have the capability of producing effluent quality within the suggested EPA and WHO guidelines for reuse.

3.3. Societal sustainability

3.3.1. Public participation

Public participation is often a neglected aspect when selecting the most appropriate wastewater treatment technology for a particular community. While some regulations designate a specific technology through a “best technology” process, the perceptions and preferences of the public for the selection and implementation of a particular technology is important if technology is to be integrated with local and broader sustainability concerns. In the US community, participation is embedded in the regulatory process through requirements of public hearings and it does not appear that a particular treatment technology has an advantage to encourage more community participation in the decision making process. We acknowledge this is a difficult indicator to quantify. Palme et al. (2005) suggest a more measurable indicator as percentage (%) of users that are “aware and responsible”.

The elements considered important in selecting a sustainable treatment system will vary from community and region because of geographical and demographical realities specific to a locality. In any case, the elements currently used in wastewater treatment selection are usually performance and affordability. In developed countries a wastewater treatment plant's efficiency, reliability, sludge disposal and land requirements are considered critical over affordability (Sperling, 1996). Hence there is a tendency to select mechanical systems over alternative treatment systems. In developing countries, affordability and the appropriateness of the technology is considered critical. Thus these countries often select simple, cost-effective appropriate technology, over more mechanized technology (Sperling, 1996). Selecting a sophisticated treatment system for a community with low-income families may place undue financial hardship on them.

3.3.2. Community size served

The size of a community can dictate the type of treatment system selected, its capacity, and hence its sustainability. Increase in population often means a larger plant capacity is required. Mechanical and lagoon systems are more capable of servicing a larger population than land treatment systems. However, mechanical systems are often

chosen over lagoon systems to service these populations. A deciding factor in choosing mechanical systems that serve large populations, especially in urban areas is the land requirement or open space availability.

But how does community size relate to the sustainability of a technology that is selected for a particular community? Greater municipal pollution loadings are associated with urban areas because of their large material inputs and outputs compared to smaller communities. This may create a burden on the surrounding environment to which dissolved and solid residuals are returned because the surrounding has limits on how much pollutant loading it can accept. Accordingly, if wastewater systems are to be sustainable, then considerations of material balances, particularly water and chemical fluxes, are required to maintain a proper balance of nutrients in the environment; avoiding the accumulation of pollutants in one ecosystem or deficiency of nutrients in another.

3.3.3. Nuisance from odor

Wastewater treatment facilities, regardless of how well designed, at one time or another may generate odor as by-products of the wastewater collection and treatment process. The presence of odor in any wastewater treatment facility is typically an aesthetic problem that usually evokes public intervention and sometime regulatory agency involvement. All the treatment systems have the potential to produce odorous emissions.

Land treatment systems have the lowest odor potential than mechanical and lagoon systems, due to pre-treatment of the wastewater before land application. Odor problems may also arise, if large solids and algae have not been removed prior to land treatment. Principle odor problems typically occur at pumping stations, inlet and outlet piping, and manholes if any are present. Odors from lagoon systems may also be due to overloading or excessive surface scum that has been allowed to accumulate. Sludge accumulation from lagoon systems may also contribute to the problem, however this should not be the case, as sludge builds up slowly over a period of 10–15 years and removal is required at end of this period (U.S. EPA, 2002).

Table 2 indicates that mechanical systems have a greater potential for odor problems than lagoon systems. Common locations in these facilities from which odors are released include preliminary and primary treatment, sludge processing facilities, and sludge disposal. Of these locations, severe odor problems can occur at the headworks facilities (wet wells, screening facilities, and grit chambers), and more recently, sludge storage, thickening, stabilization, and dewatering facilities. Industrial waste discharged to these mechanical systems can also generate significant odorous emissions, some of which, if emitted at high concentrations, pose serious health and safety risk to plant personnel (WEF, 1992). Table 2 shows other unit processes that have odor potential. Odors from sludge conditioning, transport, and disposal are a common problem. Scum and foam buildup, filamentous growth and solids settleability also

Table 2
Odor potential for typical unit processes in a wastewater treatment plant

Unit processes	Odor potential
Treatment plant	
Primary clarifiers	High
Trickling filters	High
Aeration	Low
Lagoons	Moderate
Terrestrial	Low/Moderate
Secondary clarifiers	Low/Moderate
Sludge handling	
Thickening/Holding	High
Aerobic digestion	Moderate
Sludge storage basins	Moderate/High
Dewatering	High

Source: WEF, 1992.

contribute to the problem. Overall, mechanical systems appear to have the greatest odor potential, followed by lagoon systems, and land treatment systems.

3.3.4. Staffing required to maintain infrastructure

Table 3 indicates the number of staff required to operate and maintain a wastewater facility based on plant capacity. An average staff of 4 is required to maintain a plant capacity below 1 MGD ($3.8 \times 10^3 \text{ m}^3$), while an average staff of 8.4 and 18 is required to maintain a 1.1–5 MGD (4.2×10^3 – $18.9 \times 10^3 \text{ m}^3$) and 5–10 MGD (18.9×10^3 – $37.9 \times 10^3 \text{ m}^3$) plant capacity, respectively. Plant capacities >100 MGD ($>378.5 \times 10^3 \text{ m}^3$) require the most staff with an average staffing of 315 (WEF, 1992). This indicates that smaller plants serviced by lagoon, land treatment and some mechanical technologies require the least number of staff to maintain, and that large plants served predominately by mechanical systems require the most number of staff to maintain. However, when staffing is normalized to capacity, the results show a decrease in staffing as plant capacity increases. Considering the staffing per MGD, less mechanized systems such as lagoon and land treatment systems, which constitute a majority of smaller treatment plants (<5 MGD or $<18.9 \times 10^3 \text{ m}^3$), have the greatest potential to impact social and economic development through increased employment in the community than larger treatment plants of mechanical systems that may be located outside of the community where the wastewater is generated.

3.3.5. Level of education

Increased education is generally valued as an important indicator for sustainability. The level of mechanization of a treatment system often dictates the level of operator qualification required to operate the plant, and thus their education level. Licensing required by each state depends on the type of treatment plant, complexity and size (U.S. EPA, 2004b). A high school diploma or general equivalency diploma (GED) may be necessary for less compli-

Table 3
Staffing required to operate a wastewater treatment plant for a given plant capacity

Plant capacity MGD (m^3/day)	Average capacity MGD	Plant staffing (Average)	Staffing per Ave. capacity
<1.0 ($<3.8 \times 10^3$)	0.4	3.9	10.0
1.0–5.0 (3.8×10^3 – 19×10^3)	2.5	8.4	3.4
5.0–10.0 (19×10^3 – 38×10^3)	7.5	17.6	2.4
10.0–50.0 (38×10^3 – 190×10^3)	25.1	50.6	2.0
50–100 (190×10^3 – 380×10^3)	68.1	129.4	1.9
>100 (380×10^3)	181.7	314.7	1.7

Source: WEF, 1992.

cated systems, like a lagoon or land treatment system. Higher education is required to operate and maintain complex, sophisticated treatment systems such as activated sludge.

Each state has a tiered system for operator licensing. For example in Michigan the tiered system for operator licensing is from Class A to D. The level of education is important in determining the type of wastewater treatment plant an individual can operate and the type of license required. More complex and sophisticated processes such as mechanized treatment systems require a higher level of education for licensing and operation, than less mechanized processes like lagoon and land treatment systems. The more mechanized wastewater treatment system, (e.g., activated sludge system) usually requires a higher class license, (Class A or B), while less mechanized systems (e.g., lagoons, land treatment) require a lower class license (Class C or D).

Results of an EPA survey of 150 small treatment plants with debilitating problems, show that poor operator understanding and application of process control is the most frequently occurring problem that limits treatment plant performance (WEF, 1992). Smaller treatment facilities are more likely to gain from higher level of education and additional training as these facilities are not often reached by professional networks that can offer trouble shooting advice in the event of a process upset or failure. An operator with relevant knowledge of the processes within a facility can reduce the risk of a process failure, protect worker safety, and solve debilitating problems by responding promptly to any problems, then one who does not. Even the perfect plant design will not perform adequately without informed operation and responsible administration.

3.3.6. Open space availability

Mechanical treatment systems have lower hydraulic retention time of 3–8 h (Metcalf and Eddy, 2003), which

Table 4

Land required by the different treatment technologies to treat 1 million gallons per day (MGD) ($3.8 \times 10^3 \text{ m}^3/\text{day}$)

Treatment technology	Required land area Acre/MGD
Mechanical treatment (conventional activated sludge)	0.4
Lagoon treatment	
Facultative	49–161
Aerated ponds	5–16.3
Partial-mix aerated ponds	28–49
Land treatment	
Slow rate (loading rate, 2–18 ft/yr)	60–700
Rapid infiltration (loading rate, 18–360 ft/yr)	3–60
Overland flow (loading rate, 10–70 ft/yr)	16–112

Sources: Crites and Tchobanoglous (1998), Metcalf and Eddy (2003), National Research Council (1993).

translates to smaller land requirements compared to lagoon and land treatment systems that require a longer detention time and hence more land area (Table 4). Smaller land requirements can mean that land can be used for other economic or environmental purposes. We have stated previously that the value and use of open space requires context-based interpretation. Here it is assumed that open space is a benefit to the community. Although minimum land requirement may be seen as a benefit, especially in urban areas where land is at a premium, increased land area can also be considered a benefit as assumed in this study as this provides more open, green spaces to a community and hence increased ecological and social benefits. These open spaces have many climatic purposes, such as maintaining the ambient temperature at an optimum. Land can also serve as a buffer between a community and aesthetic concerns associated with wastewater treatment.

In Atlanta, Georgia, the estimated annual economic value of tree cover in improving Atlanta's air quality was valued at \$15 million. Additional annual economic benefits in improved air quality of \$7 million could be realized if Atlanta's tree cover were increased to 40% (Trust for Public Land, 2005). Lagoon and land treatment systems that require more land area, and hence contribute to green space, also may provide habitat for waterfowl and other wildlife, as well as areas for public education and recreation (U.S. EPA, 2005b). This is important in an economic sense because in 1990, it was estimated that outdoor recreation contributed \$40 billion to the US economy. Furthermore, recreation is ranked number 2 among all economic activities on US Forest Service lands (Trust for Public Land, 2005). Such green spaces attract visitors who for example, enjoy bird watching or open areas in general, and also encourage social encounters among different members of a community that is another valued benefit.

3.4. Overall sustainability

As seen in this study, the sustainability of the different wastewater treatment technologies evaluated varied with each sustainability indicator. These selected indicators were used to measure the economic, environmental and societal sustainability of a treatment system. Hence, in order to compare results and show the overall sustainability of each treatment system, the individual results from the three treatment technologies were displayed using target plots. Target plots have been used historically in LCA and environmental product design to evaluate overall sustainability. The plots make it easy to single out points that are far removed from the plot's center that require special attention. Furthermore, in this study they enable quick visual comparisons of environmental, economic and social attributes.

The target plot shown in Fig. 3a displays the three dimensions of wastewater sustainability, the scale of impact from these dimensions, and the sustainability indicators used in this study. The impact values for each sustainability indicator for the different treatment systems were rated on a scale of 1–3, with a scale of 1 being more preferable and thus situated closer to the center of the plot. Again, the rating is a context-based decision that can change with a community, region, and country. Due to space limitations the conversion of the data to the target plots is not shown. However, the authors can be contacted for supplemental information that explains the method. Figs. 3a–d indicate that using the indicators, mechanical treatment systems are overall less sustainable than lagoon and land treatment systems (i.e., greater area of the plot is taken up by the shaded area). This is especially true in the economic and societal indicators identified for this study.

The greatest economic impacts for mechanical systems arise from their higher capital, operation and management, and resulting user costs. The greatest environmental impact for such technology originates from its high energy use compared to other technologies. Mechanized processes may compromise aesthetics (i.e., odor forming potential) to a greater extent compared to the other technologies and contribute less to the economy of a community by employing the least number of staff per plant capacity, than other treatment systems. Despite these setbacks in attaining sustainability, mechanical systems are very efficient removing biochemical oxygen demand, total suspended solids, and pathogens.

Lagoon treatment systems can provide open space to a community along with having a low user cost and less odor forming potential while providing sufficient removal of nitrogen and phosphorus, and less compromising aesthetics. The user cost depends on whether the lagoon is aerated or not, usually reflected in operational and management costs. Aerating a lagoon (including power) costs \$350 to \$500 per year, in operational and management costs, whereas a facultative lagoon costs less than \$100 per year (U.S. EPA, 2002). In lagoons, the presence of

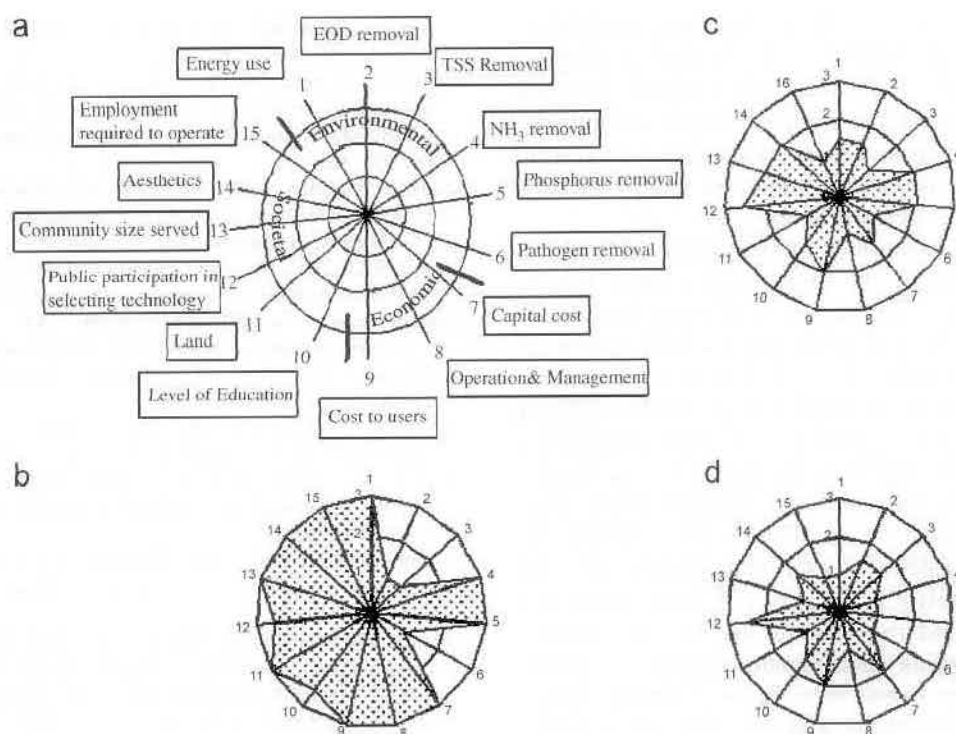


Fig. 3. (a) Target plot showing the three dimensions of wastewater sustainability, the scale of impacts from these three aspects and the sustainability indicators used in the assessment. Impact values closer to the center of the target plot are more preferable. (b) Summary of impacts from a mechanical treatment system. The plot indicates most of the impacts are further from the center. (c) Summary of impacts from a lagoon treatment system. The plot indicates impact values are close to the center. (d) Summary of impacts from a terrestrial treatment system. The plot indicates that most impact values are more closer to the center.

odor is seasonal; therefore, there is lower odor potential in the winter than summer months. These systems also have low energy use, and provide sufficient removal of biochemical oxygen demand, total suspended solids, and pathogens. Their lower capital and operation and management costs make them ideal systems for small communities. Although these systems are less mechanized, they can contribute more to the economic growth of a community in the long term by employing more staff per plant capacity than larger mechanical systems that employ fewer staff per plant capacity.

Of all three systems, land treatment systems were found to pose the least overall impact as shown by Fig. 3 results. Advantages of land treatment systems include lower capital and resulting user costs and low potential to produce odor. These systems also have low energy use during operation, and high removal of biochemical oxygen demand, total suspended solids, nitrogen, and pathogens. Like lagoon systems, land treatment systems were found to contribute to economic growth by employing more staff per plant capacity than the larger mechanical systems and also provide a community with open space.

The results suggest that using the sustainability indicators developed in this study, wastewater treatment technologies, such as lagoon and land treatment systems were shown to be a more sustainable choice, considering economic, societal, and environmental issues, when select-

ing a technology to serve communities with wastewater generation rates of less than 5 MGD ($18.9 \times 10^3 \text{ m}^3/\text{day}$).

4. Conclusions

The sustainability of mechanical, lagoon, and land treatment technologies for wastewater treatment was evaluated, using a set of sustainability indicators developed particularly for this study. The results showed the overall sustainability of a wastewater treatment technology is a function of economic, environmental and societal dimensions, and the selection and interpretation of indicators is influenced by an area's geographic and demographic situation. The results of this study are an attempt to look beyond the engineering cost and environmental performance associated with a particular treatment technology in order that selection of a technology associated with the management of wastewater treatment meets triple bottom line expectations for an equal balance of environmental, economic, and societal sustainability. One goal of this paper was to initiate a discussion on how to address a more integrated evaluation of the overall sustainability of wastewater treatment technologies.

While we acknowledge the many sources and range of data that went into this evaluation, and the difficulty in identifying a "best overall option", what is interesting is that this study demonstrated there are varying degrees of

sustainability in the way a particular treatment technology is selected and then operated. It is not as easy to design a wastewater treatment system that values education of the workforce, open space, and employment in the community, and minimizes aesthetic nuances associated with odorous air emissions, while also minimizing costs, energy use, and maximizing treatment performance.

If a particular wastewater management strategy is deemed non-sustainable, the impact will extend beyond its immediate operational vicinity and even into future generations. Therefore, traditional sustainability indicators for wastewater systems that have emphasized environmental stressors at the neglect of societal issues need to strive in the future to include current and intergenerational balanced impacts. In addition, the design of wastewater management systems that are better integrated into larger community needs could be considered. For example the reuse of treated wastewater and management of solid residuals could be better integrated with local agriculture activities which would re-distribute and return nutrients back to the surrounding environment, instead of concentrating nutrient fluxes in one receiving water body. Ideally the use of onsite-treatment systems like septic tanks, constructed wetlands, and even composting latrines has potential in contributing to sustainability as they rely on non-energy and chemical intensive processes that return nutrients to the surrounding environment. In either case, the realization of more sustainable wastewater management will require a more balanced approach in evaluating a particular management strategy's overall sustainability.

Acknowledgments

The authors gratefully acknowledge support for Ms. Muga from the Michigan Technological University Graduate School, the Sustainable Futures IGERT project sponsored by the U.S. National Science Foundation (under Grant no. DGE 0333401), and a Modialogo Worldwide Engineering Award provided by Daimler-Chrysler and UNESCO.

References

- Balkema, A., 1998. Sustainability criteria for the comparison of wastewater treatment technologies. Paper for the 11th European Junior Scientist Meeting, 12–15 February 1998, Germany.
- Balkema, A.J., Preisig, H.A., Otterpohl, R., Lambert, J.D., 2002. Indicators for the sustainability assessment of wastewater treatment systems. *Urban Water* 4, 153–161.
- Bell, S., Morse, S., 2004. Experiences with sustainability indicators and stakeholder participation: a case study relating to a 'Blue Plan' Project in Malta. *Sustainable Development* 12, 1–14.
- Crites, R., Tchobanoglous, G., 1998. *Small and Decentralized Wastewater Management Systems*. McGraw-Hill, Boston, 452pp.
- Crites, R.W., Reed, S.C., Bastian, R.K., 2000. *Land Treatment Systems for Municipal and Industrial Wastes*. McGraw-Hill, New York.
- Dixon, A., Simon, M., Burkitt, T., 2003. Assessing the environmental impact of two options for small-scale wastewater treatment: comparing a reedbed and an aerated biological filter using a life cycle approach. *Ecological Engineering* 20, 297–308.
- Emmerson, R.H.C., Morse, G.K., Lester, J.N., 1995. The life-cycle analysis of small-scale sewage treatment plants. *Journal of the Chartered Institution of Water and Environmental Management* 9 (3), 317–325.
- Franklin Associates, 1998. Characterization of building-related construction and demolition debris in the United States. Rep. EPA530-R-98-01: US-EPA, Washington, DC.
- Gilbert, R.G., 1976. Virus and bacteria removal from wastewater by land treatment. *Applied Environmental and Microbiology* 32, 333.
- Hannah, S.A., Austern, B.M., Erulp, A.E., Wise, R.H., 1986. Comparative removal of toxic pollutants by six wastewater treatment processes. *Journal of Water Pollution Control Federation* 58 (1), 27–34.
- Helmer, R., Hespanhol, L., 1997. Water pollution control: a guide to the use of water quality management principles: WHO/UNEP.
- Horvath, A., Hendrickson, C.T., 1998. Steel versus steel-reinforced bridges: environmental assessment. *Journal of Infrastructure Systems* 4 (3), 111–117.
- Hwang, Y., Hanaki, K., 2000. The generation of CO₂ in sewage sludge treatment systems: life cycle assessment. *Water Science and Technology* 41 (8), 107–113.
- Leach, L.E., Enfield, C.G., Harlin, C.C., 1980. Summary of long-term rapid infiltration studies. U.S. Environmental Protection Agency, EPA 600/2-80-165: R.S. Kerr Laboratory, Ada, Oklahoma.
- Lundin, M., Molander, S., Morrison, G.M., 1999. A set of indicators for the assessment of temporal variations in sustainability of sanitary systems. *Water Science and Technology* 39 (5), 235–242.
- Malmqvist, P.A., Pamliquist, H., 2005. Decision support tools for urban and wastewater systems—focusing on hazardous flows assessment. *Water Science and Technology* 51 (8), 41–49.
- Meadows D., 1998. Indicators and information systems for sustainable development. Hartland Four Corners VT. The Sustainability Institute (http://www.sustainer.org/dhm_archive/search.php), accessed February 5, 2007.
- Metcalf and Eddy, 2003. *Wastewater Engineering: Treatment and Reuse*. McGraw-Hill, Boston.
- Middlebrooks, E.J., Middlebrooks, C.H., 1979. Energy requirements for small flow wastewater treatment systems: Special Report 79-7 Corp of Engineers. Cold Regions Research and Engineering Laboratory: Hanover, New Hampshire.
- Mihelcic, J.R., Crittenden, J.C., Small, M.J., Shonnard, D.R., Hokanson, D.R., Zhang, Q., Chen, H., Sorby, S.A., James, V.U., Sutherland, J.W., Schnoor, J.L., 2003. Sustainability science and engineering: emergence of a new metadiscipline. *Environmental Science Technology* 37 (23), 5314–5324.
- Mitchell, G., 1996. Problems and fundamentals of sustainable development indicators. *Sustainable Development* 4, 1–11.
- National Research Council, 1993. *Managing Wastewater in Coastal Urban Areas*. National Academy of Sciences, Washington, DC.
- Nilsson, J., Bergstrom, S., 1995. Indicators for the assessment of ecological and economic consequences of municipal policies for resource use. *Ecological Economics* 14, 175–184.
- Oakley, S.M., 2005. A case study of the use of wastewater stabilization ponds in Honduras. *Small Flows Quarterly* 6 (2), 36–51.
- Organization for Economic Co-operation and Development (OECD), 1998. *Towards sustainable development: environmental indicators*. OECD, Paris.
- Palme, U., Lundin, M., Tillman, A.-M., Molander, S., 2005. Sustainable development indicators for wastewater systems—researchers and indicator users in a co-operative case study. *Resources, Conservation, and Recycling* 43, 293–311.
- Pettygrove, G.S., Asano, T., 1984. *Irrigation with reclaimed municipal wastewater—a guidance manual*. Rep. 84-1 wr, SWRCB; Sacramento, California.
- Raftelis Financial Consulting, 2004. *Water and Wastewater Rate Survey*. Pennsylvania; CRC Press/Lewis Publishers: Pennsylvania.

- Reed, S.C., 1991. Constructed wetlands for wastewater treatment. *BioCycle* 32, 44.
- Reed, S.C., Crites, R.W., Middlebrooks, E.J., 1995. *Natural Systems for Waste Management and Treatment*, second ed. McGraw-Hill Inc., New York.
- Spangenberg, J.H., 2002. Institutional sustainability indicators: an analysis of the institutions in Agenda 21 and a draft set of indicators for monitoring their effectivity. *Sustainable Development* 10, 103–115.
- Sperling, M.V., 1996. Comparison among the most frequently used systems for wastewater treatment in developing countries. In: *International Symposium on Technology Transfer*, vol. 1, pp. 107–120.
- Trust for Public Land, 2005. The Economic benefits of parks and open space: how land conservation helps communities grow smart and protect the bottom line (www.tpl.org), accessed October 2005.
- Tsagarakis, K.P., Mara, D.D., Angelakis, A.N., 2002. Application of cost criteria for selection of municipal wastewater treatment systems. *Water, Air & Soil Pollution* 142, 187–210.
- UNDP/UN Division of Sustainable Development, 1995. Department of Policy Co-ordination and Sustainable Development. Work Programme on Indicators for Sustainable Development. United Nations, New York.
- United Nations Commission on Sustainable Development, UNCSO, 1996. Indicators of sustainable development framework and methodologies. United Nations, New York.
- United Nations Environment Programme, UNEP—International Environmental Technology Centre, Unit of Sustainable Development and Environment, 1997. Source book of alternative technologies for freshwater augmentation in Latin America and the Caribbean. Washington, DC.
- U.S. Environmental Protection Agency, 1982. Process design manual land treatment of municipal wastewater, supplement on rapid infiltration and overland flow. EPA 625/1-81-013a; CERL, Cincinnati, Ohio.
- U.S. EPA, 1975. Wastewater treatment ponds. EPA-430/9-74-001, MCD-14; Washington, DC.
- U.S. EPA, 1981. Process design manual—land treatment of municipal wastewater. EPA-625/1-81-013, CERL, Cincinnati, Ohio.
- U.S. EPA, 2002. Onsite wastewater treatment systems manual, EPA/625/R-00/008, February 2002.
- U.S. EPA, 2004a. Office of wastewater management. Report to congress: clean water needs survey 2000 (<http://www.epa.gov/OW-OWM/html/mtb/swns/foi.htm>), accessed June 2004.
- U.S. EPA, 2004b. Federal register (www.epa.gov/fedrgstr/EPA-WATER/2001/April/Day-18/w9482.htm), accessed September 2004.
- U.S. EPA, 2005a. Manual: guidelines for water reuse (<http://www.epa.gov/ordntrnt/ORD/NRMRL/Pubs/625R92004/625R92004prel.pdf>), accessed December 2005.
- U.S. EPA, 2005b. Constructed wetlands for wastewater treatment and wildlife habitat: 17 case studies (<http://www.epa.gov/OWOW/wetlands/construc/index.html>), accessed July 2005.
- Water Environment Federation (WEF) and the American Society of Civil Engineers, 1992. *Design of Municipal Wastewater Treatment Plants*, vol. 11. Book Press Inc., Vermont.
- Wilson, A., 1993. Cement and concrete: environmental considerations. *Environmental Building*, News 2 (2), 7–12.
- World Health Organization (WHO), 1989. Health guidelines for the use of wastewater in agriculture and aquaculture. Technical Report Series no. 778, Report of a Scientific Group Meeting. World Health Organization, Geneva.
- World Health Organization (WHO), 2005. Evaluation of the costs and benefits and sanitation improvements at the global level (http://www.who.int/water_sanitation_health/wsh0404summary/), accessed July 2005.
- World Health Organization (WHO) and United Nations Children's Fund, 2000. *Global water. Supply and Sanitation Assessment 2000 Report*. World Health Organization and the United Nations Children's Fund Publishing, Geneva, Switzerland.

Probabilistic fugacity analysis of Lake Pontchartrain pollution after Hurricane Katrina

Sinem Gokgoz-Kilic*, Mustafa M. Aral

Multimedia Environmental Simulations Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received 16 June 2006; received in revised form 10 February 2007; accepted 11 March 2007
Available online 25 April 2007

Abstract

After Hurricane Katrina passed through the US Gulf Coast in August 2005, floodwaters covering New Orleans were pumped into Lake Pontchartrain as part of the rehabilitation process in order to make the city habitable again. The long-term consequences of this environmentally critical decision were difficult to assess at the time and were left to observation. In the aftermath of these natural disasters, and in cases of emergency, the proactive use of screening level models may prove to be an important factor in making appropriate decisions to identify cost effective and environmentally friendly mitigation solutions. In this paper, we propose such a model and demonstrate its use through the application of several hypothetical scenarios to examine the likely response of Lake Pontchartrain to the contaminant loading that were possibly in the New Orleans floodwaters. For this purpose, an unsteady-state fugacity model was developed in order to examine the environmental effects of contaminants with different physicochemical characteristics on Lake Pontchartrain. The three representative contaminants selected for this purpose are benzene, atrazine, and polychlorinated biphenyls (PCBs). The proposed approach yields continuous fugacity values for contaminants in the water, air, and sediment compartments of the lake system which are analogous to concentrations. Since contaminant data for the floodwaters are limited, an uncertainty analysis was also performed in this study. The effects of uncertainty in the model parameters were investigated through Monte Carlo analysis. Results indicate that the acceptable recovery of Lake Pontchartrain will require a long period of time. The computed time range for the levels of the three contaminants considered in this study to decrease to maximum contaminant levels (MCLs) is about 1 year to 68 years. The model can be implemented to assess the possible extent of damage inflicted by any storm event on the natural water resources of Southern Louisiana or similar environments elsewhere. Furthermore, the model developed can be used as a useful decision-making tool for planning and remediation in similar emergency situations by examining various potential contamination scenarios and their consequences.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Fugacity; Monte Carlo analysis; Management; Environment

1. Introduction

The city of New Orleans, Louisiana, was flooded as a result of Hurricane Katrina making large sections of the city uninhabitable. In an immediate attempt to make the city habitable again, a decision was made to drain the floodwaters directly into Lake Pontchartrain. Although draining the floodwaters into Lake Pontchartrain seemed to be an effective solution at the time, the long-term

pollution effects of this process were not examined thoroughly as the water being drained into the lake contained various potentially hazardous contaminants. Since industrial complexes and chemical warehouses along with other facilities were flooded, it is probable that hazardous chemicals were leaked into the floodwaters. It was assumed that natural processes such as dilution and decay in the lake would be effective in neutralizing contaminants over time. In this study, we examine the effectiveness of these natural processes for the recovery of Lake Pontchartrain. To provide preliminary insights into the long-term effects of the decision made to drain the

*Corresponding author. Tel.: +1 404 385 0565.

E-mail address: sinem@gatech.edu (S. Gokgoz-Kilic).

floodwaters into the lake without treatment, three representative chemicals with different physicochemical properties were considered—benzene, atrazine, and polychlorinated biphenyls (PCBs).

Large amounts of gasoline were observed in the New Orleans floodwaters. Thus, benzene, a constituent of gasoline and representative of volatile compounds, was considered to be one chemical compound of interest in this study. Warehouses that stored agricultural products were inundated by floodwaters. Thus, atrazine, the most commonly used agricultural herbicide, was considered to be another possible chemical compound in the floodwaters. Besides, atrazine is a good representative of more water-soluble chemical which may have possibly been in the floodwaters. The third chemical compound considered in this study is PCB, a highly toxic and highly hydrophobic chemical compound. The PCBs, which are still in existence even after the prohibition in 1970s, were selected to represent hydrophobic compounds. All three chemical compounds considered are potentially harmful to human health as well as biota in the lake. Selecting chemicals with different physicochemical characteristics also provides information about which chemicals are likely to cause the most severe and longest-lasting effects on the recovery of Lake Pontchartrain.

All chemicals behave differently in nature; they partition through different pathways when introduced into a surface water system such as a lake. Some of the contaminants volatilize, some of them adsorb to solid materials, and some of them stay in the aqueous phase. In any attempt to determine the fate of chemicals in surface water environments, the existence of sediment, suspended particles, and the air phase must also be considered. This paper addresses the existence of contaminants in three different phases by a detailed application of fugacity analysis to Lake Pontchartrain. Using fugacity analysis, a chemical equilibrium between phases can be established and partitioning among different phase compartments can be predicted. Since fugacity is continuous between phases (unlike concentration which is discontinuous between phases), the fugacity approach allows for monitoring of a chemical as it passes from one phase to another with relative ease (Mackay, 2001; Stumm and Morgan, 1996).

Extensive collection of data on the floodwaters was not done due to the need for immediate emergency response actions. Thus, no site-specific information is available on pollutant levels of the floodwaters resulting from Hurricane Katrina. This study, therefore, provides one alternative approach to conduct pollution analysis in the lake in the absence of such data. Nevertheless, the lack of data does introduce significant uncertainties in the study outcome. For this purpose, a Monte Carlo analysis was built into the developed model to evaluate the uncertainty effects. Therefore, probability distributions for mass transfer coefficients, decay rate coefficients, and most importantly source rate were used instead of single deterministic values. Consequently, the recovery time of the lake becomes a

probability density outcome instead of a deterministic outcome. In turn, a better understanding of the expected pollution levels in Lake Pontchartrain as well as the recovery time of the lake was achieved. Following a sensitivity analysis, most sensitive parameters were determined and used for the uncertainty analysis.

2. Fugacity model

A lake system consists of different phases that interact dynamically with each other. These interactions play an important role in the persistence of chemicals in each phase. A lake system consists of air, water, and sediment compartments. The water column is in contact with both air and sediment compartments, and therefore the basis of the model is established by three unsteady-state mass balance equations among these compartments.

All three compartments are assumed to be well mixed. The interactions between the air and water compartments involve wet and dry deposition from air to water and air–water diffusion. The interactions between the sediment and water phases involve diffusion, deposition of suspended particles onto sediments, and resuspension of particles from sediments. Additionally, water flows out of the lake system through tributaries, and there is wind-induced air flow just above the lake creating a mixing zone. Sediments are assumed to be immobile. Moreover, chemicals follow first-order decay reactions in each compartment. The compartments and the natural processes included in the model are shown in Fig. 1.

Fugacity is used instead of concentration in the development of mass balance equations since this approach provides an easier representation of mass flux among phases (Mackay, 2001; Stumm and Morgan, 1996). Fugacity, F , and concentration, C , are linearly related through a proportionality constant, Z , as

$$C = ZF. \quad (1)$$

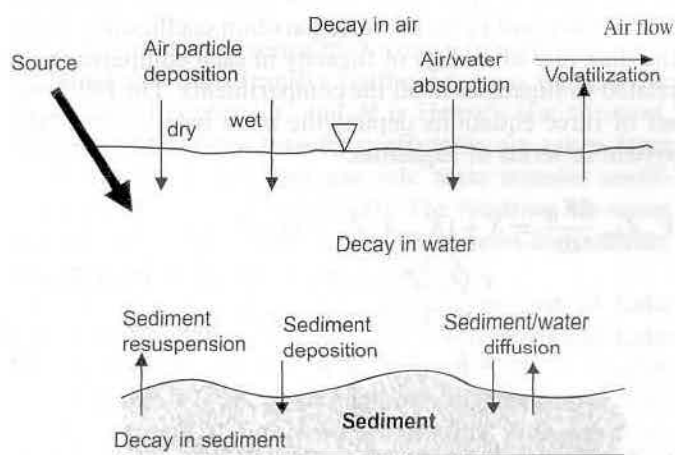


Fig. 1. Schematic view of the natural processes considered in the fugacity model.

Each phase contains particles. Air consists of gas and aerosols, the small particulate matter in air. Aerosols provide solid surfaces onto which a chemical can adsorb. Likewise, the water column contains suspended particles which form the solid phase available for adsorption. Only a portion of the bottom sediments is solid material as sediments form a porous media saturated with water. Thus, each phase has its bulk fugacity capacity representing respective physicochemical characteristics. Bulk fugacity capacities can be calculated using volume fractions. The fraction of aerosols in air is given in literature as $30 \mu\text{g}/\text{m}^3$, which corresponds to a volume fraction of 2×10^{-11} . The fraction of suspended solids in water is given in literature as 5 mg/L , which corresponds to volume fraction of 2.08×10^{-7} . Sediments are given to be 15% solid materials by volume (Mackay, 1989). Hence, the bulk fugacity capacity in water phase can be calculated using Eq. (2):

$$Z_{bw} = Z_p n_w + Z_w(1 - n_w), \quad (2)$$

where Z_{bw} is bulk water fugacity capacity, Z_p is water particle fugacity capacity, Z_w is pure water fugacity capacity, and n_w is the volume fraction of particles in water. Bulk fugacity capacity in air phase is calculated with Eq. (3):

$$Z_{ba} = Z_{aerosol} n_a + Z_a(1 - n_a), \quad (3)$$

where Z_{ba} is bulk air fugacity capacity, $Z_{aerosol}$ is aerosol fugacity capacity, Z_a is pure air fugacity capacity, and n_a is the volume fraction of aerosols in air. Bulk fugacity capacity in sediment phase is calculated with Eq. (4):

$$Z_{bs} = Z_s n_s + Z_w(1 - n_s), \quad (4)$$

where Z_{bs} is bulk sediment fugacity capacity, $Z_{sediment}$ is sediment fugacity capacity, Z_w is water fugacity capacity, and n_s is the volume fraction of solid particles in sediment.

The fugacity capacity of suspended sediments (water particles) is assumed to be the same as the fugacity capacity of sediments as they share similar physical characteristics (i.e. density and particle–water partition coefficient). Thus, the time rate of change of fugacity in each compartment is related to fugacities in all the compartments. The following set of three equations depicts the mass balance of a lake system in terms of fugacities:

$$\begin{aligned} V_w Z_{bw} \frac{dF_W}{dt} = & S + (K_{aw} A_w Z_w + Q_{dry} Z_{aerosol} \\ & + Q_{wet} Z_{aerosol}) F_A + (Q_{res} Z_s + K_{sw} A_s Z_w) F_S \\ & - (k_w V_w Z_w + Q_{dep} Z_p + Q_{out} Z_{bw} \\ & + K_{aw} A_w Z_w) F_W, \end{aligned} \quad (5)$$

$$\begin{aligned} V_a Z_{aw} \frac{dF_A}{dt} = & (K_{aw} A_w Z_w) F_W - (Q_{dry} Z_{aerosol} + k_a V_a Z_a \\ & + Q_{wet} Z_{aerosol} + Q_{air} Z_{ba} + K_{aw} A_w Z_w) F_A, \end{aligned} \quad (6)$$

$$\begin{aligned} V_s Z_{bs} \frac{dF_S}{dt} = & (Q_{dep} Z_p + K_{sw} A_s Z_w) F_W - (k_s V_s Z_s + Q_{res} Z_s \\ & + K_{sw} A_s Z_w + Q_{bury} Z_s) F_S \end{aligned} \quad (7)$$

in which F_W is the fugacity of the water compartment, F_A is the fugacity of the air compartment, and F_S is the fugacity of the sediment compartment. S is the source rate of contaminant in the water phase. V_w , V_a , and V_s represent the volumes of water, air, and sediment compartments, respectively. Similarly, A_w and A_s represent the surface areas of water and sediment. Q_{dry} is dry deposition rate from air, Q_{wet} is the rate of wet deposition from air, Q_{res} is the resuspension rate of sediment particles, Q_{dep} is the deposition rate of suspended particles in the water column, Q_{air} is the wind-induced air flow rate on the lake, Q_{out} is the rate of water flowing out of the lake, and Q_{bury} is the burial rate of bottom sediments. Z_w is the fugacity capacity of water, Z_a is the fugacity capacity of air, Z_s is the fugacity capacity of sediments, Z_p is the fugacity capacity of water particles (suspended solids), $Z_{aerosol}$ is the fugacity capacity of aerosols, Z_{bw} is the bulk water fugacity capacity, Z_{ba} is the bulk air fugacity capacity, and Z_{bs} is the bulk sediment fugacity capacity. K_{aw} is the mass transfer coefficient between air and water, whereas K_{sw} is the mass transfer coefficient between sediments and water. The terms k_w , k_a , and k_s are the first-order decay rate constants in water, air, and sediment compartments, respectively.

Eqs. (5) through (7) have three unknowns that can be solved simultaneously through numerical integration using finite-difference approximations. In this study, forward difference approximation is used for time derivatives, and time-averaged values are used for fugacity values of each component on the right-hand side of the equations as shown in Eq. (8):

$$\left. \begin{aligned} \frac{dF_i}{dt} &= \frac{F_i^{k+1} - F_i^k}{\Delta t} \\ F_i &= \frac{F_i^{k+1} + F_i^k}{2} \end{aligned} \right\}; \quad i = A, W, S. \quad (8)$$

Applying this process to the governing equations and lumping all known values together yield three equations with three unknowns as shown in Eq. (9):

$$\begin{aligned} A_1 F_W^{k+1} + A_2 F_A^{k+1} + A_3 F_S^{k+1} &= B_1, \\ A_4 F_W^{k+1} + A_5 F_A^{k+1} &= B_2, \\ A_6 F_W^{k+1} + A_7 F_S^{k+1} &= B_3. \end{aligned} \quad (9)$$

These equations can be written in matrix format as demonstrated by Eq. (10):

$$\begin{bmatrix} A_1 & A_2 & A_3 \\ A_4 & A_5 & 0 \\ A_6 & 0 & A_7 \end{bmatrix} \begin{bmatrix} F_W \\ F_A \\ F_S \end{bmatrix}^{k+1} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}^k. \quad (10)$$

Since all of the terms on the right-hand side of the above equations are known, the right-hand side terms are simplified and represented as a vector. The matrix equation

can be solved by a common matrix solver such as the LU decomposition scheme as used in this study. The source rate is also introduced into the equations above as a time-dependent initial loading. The source is calculated as a volume fraction of floodwaters during the pumping period and is converted to mass of contaminant per time using pumping rate, molecular weight, and density of the contaminant.

The constants in Eq. (10) are either physicochemical properties of the chemicals or represent the characteristics of the lake. In the following section, the information used to calculate these values is explained. The physicochemical properties of the selected chemicals are obtained from reference handbooks (Howard, 1989; Mackay et al., 1992) and summarized in Table 1.

The amount of floodwaters pumped into Lake Pontchartrain is about 10% of the lake volume, and the draining of floodwaters took about 1 week, which is used to calculate loading rate into the lake. Ten percent of the lake volume in 1 week corresponds to a pumping rate of $3.5 \times 10^6 \text{ m}^3/\text{h}$. Using this pumping rate, a source rate term that is the key input of the model is estimated. Since there are no measurements prior to the pumping of floodwaters into Lake Pontchartrain, various potential contaminant-loading scenarios can be investigated. To reflect the uncertainty in the source rate estimate, a uniform distribution is assigned to the source rate term. Source rates of 0.01% and 1% by volume of the inflow floodwaters are used as the lower and upper limits of the distribution, respectively. Any source rate between these values is given an equal probability of occurrence. In another scenario, a fixed contaminant-loading rate corresponding to the average of the uniform distribution (0.1%) is also used. By using this estimation method and fixed source rate values, the source rate terms for benzene, atrazine, and PCB are estimated as 3.84×10^7 , 1.88×10^7 , and $1.47 \times 10^7 \text{ mol/h}$, respectively. We should emphasize here that the source-loading scenarios selected above are for

demonstration purposes and may be varied depending on the field data that may be collected at a site for a specific and more rigorous application. Further, as stated earlier, a sensitivity analysis on this parameter was also conducted in an effort to estimate its impact on the outcome.

The time to recover from the contaminant load is calculated as the time to reach allowable concentrations given in Environmental Protection Agency (EPA) (2004) Edition of the Drinking Water Standards and Health Advisories (EPA, 2004). The maximum contaminant levels (MCL) for the aforementioned chemicals are 0.005 mg/L for benzene, 0.003 mg/L for atrazine, and 0.0005 mg/L for total PCBs.

The water outflow from the lake is chosen as $1.57 \times 10^6 \text{ m}^3/\text{h}$, which is the total water output via tributaries of the system (USGS, 2002). The deposition rate of $816 \text{ m}^3/\text{h}$ and the resuspension rate of $1000 \text{ m}^3/\text{h}$ are selected based on the guidelines in the work of Mackay (2001).

Air flow calculations are based on the annual average wind velocity. An average wind speed of $1.99 \times 10^4 \text{ m/h}$ is expected above Lake Pontchartrain predominantly in the SE–NW direction (USGS, 2002). The volumetric air flow rate is calculated using the cross-sectional area of the air compartment in SW–NE direction. This study mainly focuses on the response time of Lake Pontchartrain with respect to different contaminants which may exist in the floodwaters after Hurricane Katrina, while some of the volatilized chemical loss can re-enter the system through atmospheric deposition mechanisms. To incorporate this process into our model, only a small thickness of the air compartment is used as the mixing zone for tracking the contaminant immediately above the lake. The mixing zone is chosen to be the bottom 50 cm of the air compartment, which resulted in the air flow rate as $3.2 \times 10^7 \text{ m}^3/\text{h}$.

For determining air–water and sediment–water mass transfer rates required to calculate diffusive fluxes between compartments, the two-film theory is used (Valsaraj, 2000):

$$\frac{1}{K_{aw}} = \frac{1}{k_L} + \frac{1}{k_G H} \quad (11)$$

where K_{aw} is the air–water mass transfer coefficient, k_L is the liquid side mass transfer coefficient, k_G is the gas side mass transfer coefficient, and H is Henry's law constant. Sediment–water mass transfer coefficients are taken from Reible (1998). Liquid and gas side mass transfer coefficients are given in Valsaraj (2000). The resulting air–water and sediment–water mass transfer coefficients for benzene, atrazine, and PCBs are given in Table 2.

To advance further into the site-specific case of Lake Pontchartrain, data on the physical characteristics of Lake Pontchartrain are needed. Lake Pontchartrain is roughly oval in shape. It is about 40 miles (64.4 km) in length from east to west and about 24 miles (38.6 km) in length from north to south and forms the northern boundary of New Orleans. Cities adjacent to the lake are Mandeville and Madisonville to the north and Slidell to the northeast

Table 1
Physicochemical data on selected chemicals

Parameter	C ₆ H ₆ — benzene	C ₈ H ₁₄ ClN ₅ — atrazine	Total PCBs
Molecular weight (g/mol)	78.11	215.75	326
Log K_{ow}	2.13	2.75	6.66
Henry's Law constant (atm m ³ /mole)	5.55×10^{-3}	2.88×10^{-4}	2.9×10^{-5}
Z_{water} (mol/m ³ Pa)	1.78×10^{-3}	3480	0.0818
Z_{air} (mol/m ³ Pa)	4.04×10^{-4}	4.04×10^{-4}	4.04×10^{-4}
$Z_{sediment}$ (mol/m ³ Pa)	1.41×10^{-2}	77000	11,500
$Z_{bulk\ water}$ (mol/m ³ Pa)	1.8×10^{-3}	3480	0.0951
$Z_{bulk\ air}$ (mol/m ³ Pa)	4.04×10^{-4}	1.62×10^{-3}	4.58×10^{-4}
$Z_{bulk\ sediment}$ (mol/m ³ Pa)	2.96×10^{-3}	14,500	1730
Half life in water (h)	1700	17,000	139,000
Half life in air (h)	170	5	693
Half life in sediment (h)	17,000	1700	347,000

Table 2
Mass transfer coefficients

Compound	Overall air–water mass transfer Coefficient (m/h)	Sediment–water mass transfer coefficient (m/h)
Benzene	0.00958	0.0006
Atrazine	0.00001	0.0013
PCBs	0.0034	0.0004

Table 3
Physical data on Lake Pontchartrain

Surface area	1,632,000,000 m ² (630 mi ²)
Average depth	3.65 m (12 ft)
Total water outflow	1,565,000 m ³ /h (15350 cfs)
Wind speed on lake	10.75 knots (19910 m/h)
Total air outflow	3.2 × 10 ⁷ m ³ /h
Depth of air on lake	0.5 m
Volume of air	816,000,000 m ³
Depth of sediment	0.05 m
Volume of sediment	81,600,000 m ³
Density of sediment	2400 kg/m ³

Table 4
Removal percentages of contaminants from aqueous phase

Compound	Decay (%)	Deposition (%)	Outflow (%)	Volatilization (%)
PCBs	0.1	73	6.7	20.2
Benzene	20.5	0.03	7.3	72.2
Atrazine	13	0.45	85.8	0.9

(USGS, 2002). Other data on the physical characteristics of Lake Pontchartrain are given in Table 3.

The removal mechanisms from the water column of the lake, as can be seen in Eq. (5), are first-order decay, transfer into sediments, flow out via tributaries, and transfer into air compartment. The percentage of each removal mechanism for the three contaminants is calculated. The results of this calculation are given in Table 4.

As inferred by the data in Table 4, the three contaminants examined in this study behave very differently with respect to each other. Monitoring the removal mechanism of each contaminant from a compartment is made easier with the use of fugacity approach in the model. Upon examining the results of the analysis for these three contaminants, the expected behavior of other similar contaminants in the floodwaters may be easily understood. If the contaminant is as volatile as benzene, it will be removed from the lake rapidly. If the contaminant is more hydrophobic, the contaminant will remain in the system for a longer period of time.

Table 4 also infers that the main removal mechanism for benzene is volatilization. Higher values of the air–water mass transfer coefficient of benzene better enable the

transfer of benzene into the air compartment. Once passed into the air compartment, benzene is rapidly removed from the system by wind-induced air flow. It is observed that a negligible percentage of benzene is transferred into the sediments. The second most significant removal mechanism for benzene is decay.

The main removal mechanism for PCBs is deposition onto the sediments. This deposition is expected as PCBs are the most hydrophobic chemicals among the three selected contaminants. About three quarters of the PCBs introduced into the lake system are transferred to the sediment compartment. When a chemical passes into the sediments, it remains there for a long time, thus stays in the system for a long time. The loss mechanisms from the sediments are burial, decay, and release back into the water column. Therefore, PCB stays in the system longer than the other contaminants studied. In fact, that is the main reason for observation of PCBs in aquatic systems even three decades after being banned. Sediments become the secondary source of PCBs by slowly releasing the chemical back into the water column.

The main removal mechanism for atrazine is outflow through the tributaries. Removal occurs to a smaller degree through first-order decay in the water column. Atrazine does not attach onto solid surfaces or volatilize significantly. Atrazine stays mostly in the aqueous phase and is removed from the system in its dissolved form. In fact, atrazine has the highest water solubility among the three contaminants selected for this study.

From studying Table 4, it is seen that the three selected chemicals are indeed representative of very different behavior in nature. Examining the removal mechanisms of potential contaminants is a very effective decision tool in any emergency situation, whether it is a natural disaster or an accidental spill. In fact, this model can be a very useful tool to analyze the effects of many potential chemicals that could be introduced into natural waters after natural disasters or accidental spills.

3. Sensitivity analysis

After calculating all the constants in Eq. (10), a sensitivity analysis is performed on the proposed model. A sensitivity analysis of the values representing the transport and transformation processes in the lake allows for observation of parameters that have the most significant effects on the self cleaning capacity of Lake Pontchartrain. The most sensitive parameters were found to be the source rate, mass transfer coefficients, and decay rate constants. Decay rate constants directly affect the persistence of a chemical in the lake system. Given a higher decay rate in a specific phase, the removal rate of a chemical increases. Mass transfer coefficients are also very important in the recovery time of the system. If a chemical passes to sediments more rapidly, it stays in the system for a longer period of time. On the other hand, if a chemical passes to the air phase more rapidly, it is removed from the

system more rapidly due to wind-induced advection. We do not know the exact concentrations of chemicals in the floodwaters. Thus, using probability distributions instead of deterministic values for the contaminant loads is logical. The results of the sensitivity analysis with respect to the source rate are given in Table 5 demonstrating the effect of source rate on the recovery time of Lake Pontchartrain.

4. Monte Carlo analysis

Mass balance fugacity models, used to determine fates of chemicals, are increasingly recognized as useful tools to facilitate management of chemicals that are released into the environment. However, as models are only an approximation of the actual fates of chemicals, it is crucial to communicate clearly the uncertainties associated with the model. Monte Carlo analysis is currently the technique of choice to quantify uncertainty in environmental models. Monte Carlo analysis is very attractive for many applications in different fields for it is widely applicable and unconstrained by the type of model (MacLeod et al., 2002). In this approach, the most sensitive parameters recognized by the sensitivity analysis are passed to the fugacity model as random values sampled from a probability distribution.

In literature, mass transfer coefficients and decay rate constants have been assigned lognormal distributions (Citra, 2004; MacLeod et al., 2002). Choice of the lognormal distribution for many fate and transport model parameters is advantageous as lognormal distributions have a positive state space. The lognormal distribution has two descriptive parameters: a mean and a standard deviation that are related to a corresponding normal distribution.

Table 5
Sensitivity analysis results for source rate

Contaminant (%)	Contaminant ratio in floodwaters by volume		
	0.01	0.1	1
Benzene (years)	0.72	0.82	0.87
Atrazine (years)	1.05	1.27	1.50
PCB (years)	25.49	44.09	62.7

Table 6
Mean and standard deviation values used to generate random values for decay rate constants and mass transfer coefficients

Parameter/chemical	Benzene		Atrazine		PCBs	
	Mean	SD ^a	Mean	SD	Mean	SD
Water decay rate constant	0.00077 h ⁻¹	0.5	0.00126 h ⁻¹	0.5	0.000005 h ⁻¹	0.5
Air decay rate constant	0.003316 h ⁻¹	0.5	0.386 h ⁻¹	0.5	0.0001 h ⁻¹	0.5
Sediment decay rate constant	0.0000856 h ⁻¹	0.5	0.000199 h ⁻¹	0.5	0.000002 h ⁻¹	0.5
Overall air–water mass transfer coefficient	0.00958 m/h	0.5	0.00001 m/h	0.5	0.0034 m/h	0.5
Sediment–water mass transfer coefficient	0.0006 m/h	0.5	0.0013 m/h	0.5	0.0004 m/h	0.5

^aStandard deviation.

The pumping rate of floodwaters into Lake Pontchartrain is estimated to be $3.5 \times 10^6 \text{ m}^3/\text{h}$. The source rate term is estimated, assuming a fraction of the floodwaters (by volume) is the contaminant. In this study, a uniform distribution is assigned to the source rate term. Source rates calculated with 0.01% and 1% by volume of contaminant in the floodwaters are used as the lower and upper limits of the uniform distribution, respectively. Any source rate between these values is given an equal probability of occurrence.

A total of 10^4 trials are performed for the Monte Carlo runs. 10^4 random variables are created for the overall air–water mass transfer coefficient, the water–sediment mass transfer coefficient, the water decay rate constant, the air decay rate constant, the sediment decay rate constant, and the source rate. The Box–Muller algorithm is used (Gentle, 2003; Ronen, 1988) to generate random values from a lognormal distribution. The values found in chemical handbooks for these parameters are used as means for mass transfer coefficients and decay rate constants (Table 6). A standard deviation of 0.5 is used for all mass transfer coefficients and decay rate constants. This value gives a confidence factor of about 2.7 (MacLeod et al., 2002), which implies that about 95% of all random values lie between $1/2.7$ and 2.7 times the median. The inverse of the cumulative density function of a uniform distribution is used to generate random variables for the source rate. The model (Eq. (10)) is run 10^4 times through time, selecting a random value for the uncertain parameters at each time.

5. Results of Monte Carlo analysis

Prior to Monte Carlo Analysis, each contaminant is modeled using the mean values for each parameter including the mass transfer coefficients and decay rate constants. The source rate is calculated assuming that contaminant level in the floodwaters is 0.1% of the total floodwaters by volume. The response of the lake to the introduction of contaminants at these mean values can be seen in Fig. 2.

Floodwaters are pumped for the duration of a week into the lake. During the pumping period, the concentration of

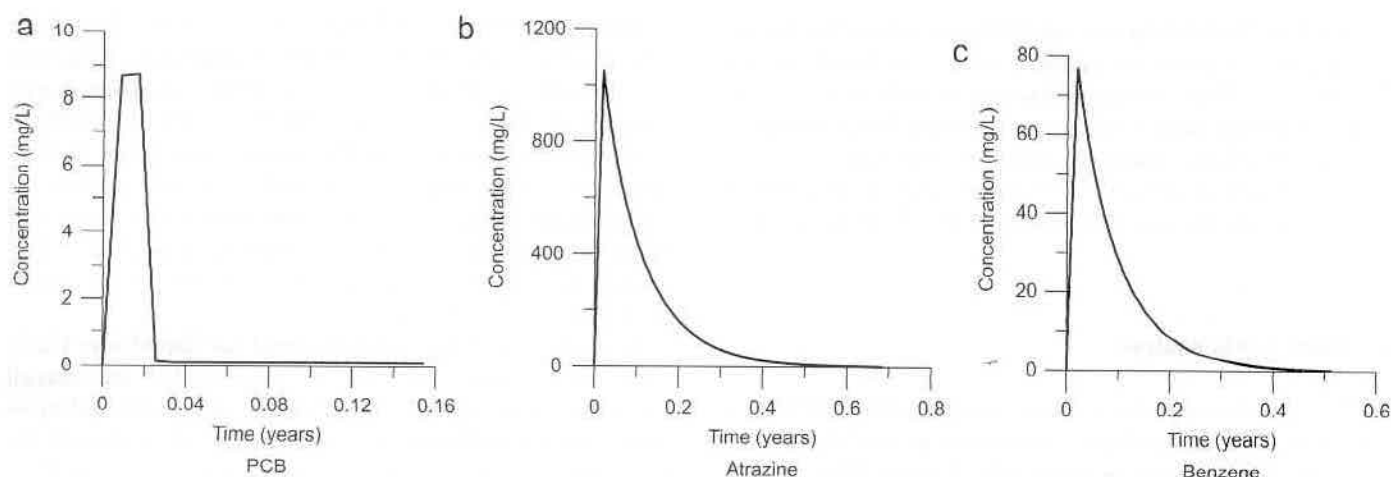


Fig. 2. Concentration profile of contaminants in the water column of Lake Pontchartrain.

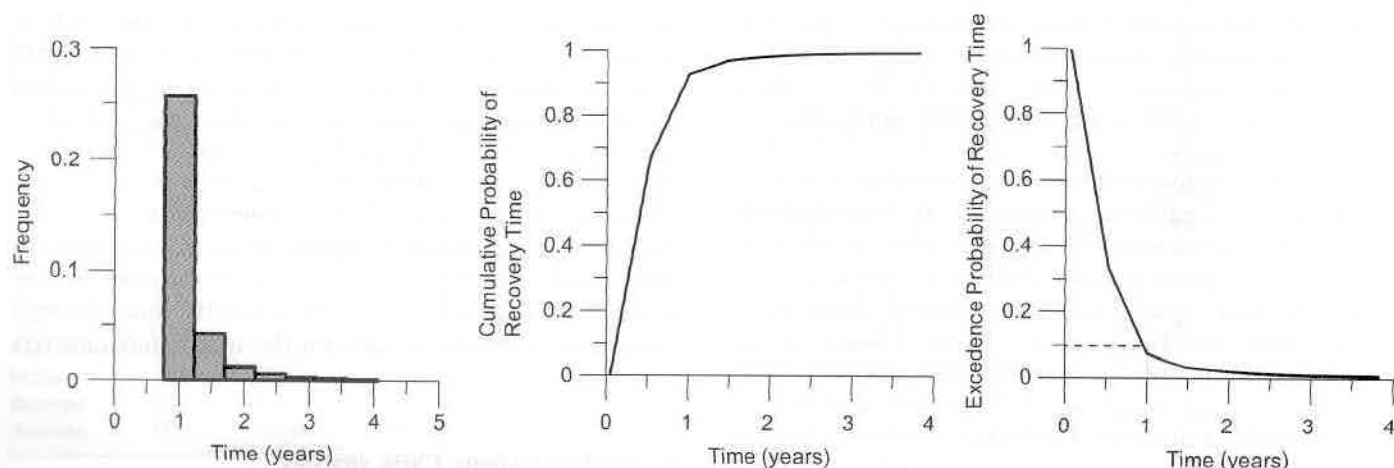


Fig. 3. Analysis results for recovery of Lake Pontchartrain from atrazine load with source rate fixed at 0.1%.

each contaminant reaches very high concentrations in the lake. After pumping stops, the concentration of contaminants decreases in the aqueous phase as a function of physico-chemical properties. The other compartments follow a similar behavior, so they are not shown in this figure.

PCB concentration reaches a maximum of about 9 mg/L during the week of pumping, which is a larger amount than the MCL values. Similarly, atrazine and benzene also reach very high concentrations as seen in Fig. 2.

For the Monte Carlo analysis, two cases are examined. In Case I, the source rate is kept constant at 0.1% (the mean value of the source rate distribution) while all decay rate constants and mass transfer coefficients are sampled from lognormal distributions and randomly used in the model. In Case II, random sampling of the source rate from a uniform distribution occurs in addition to the sampling of mass transfer coefficients and decay rate constants from lognormal distributions.

For each set of randomly selected groups of mass transfer coefficients, decay rate coefficients, and source rate, the time for the system to reach MCL values is

recorded. This value reflects the duration of time required for Lake Pontchartrain to recover from contaminant load by natural processes. In other words, this value is the self-cleansing time of the lake.

The resulting set of 10^4 time values is represented within three plots. The first plot is the frequency of occurrence of a time value. Given these distributions, cumulative and complementary cumulative probability distributions can also be calculated as shown in the same figure as discussed below. The complementary cumulative probability plot shows the probability of the lake exceeding a certain time in recovering from a contaminant load. This is a very useful way to interpret the uncertainty inherent in the results.

Fig. 3 shows the results of Monte Carlo simulation for atrazine for Case I, where the source rate is fixed at 0.1% by volume, while Fig. 4 shows the results for atrazine for Case II where source rate is randomly sampled along with mass transfer coefficients and decay rate coefficients. Figs. 5 and 6 show the results for benzene for Case I and Case II, respectively; similarly, Figs. 7 and 8 show the results for total PCBs for Case I and Case II, respectively.

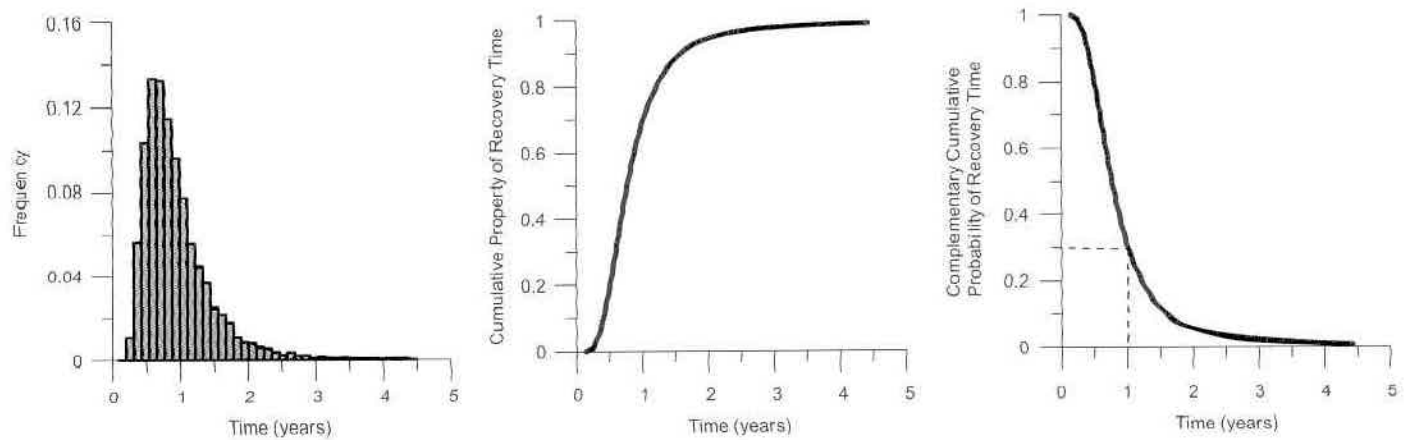


Fig. 4. Analysis results for recovery of Lake Pontchartrain from atrazine load with an uncertain source rate.

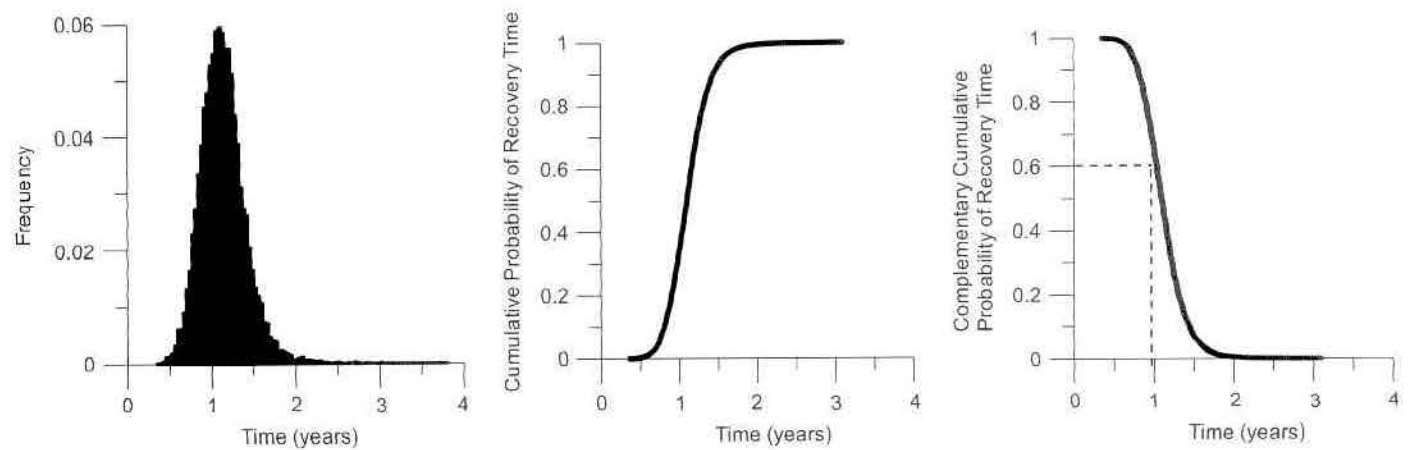


Fig. 5. Analysis results for recovery of Lake Pontchartrain from benzene load with source rate fixed at 0.1%.

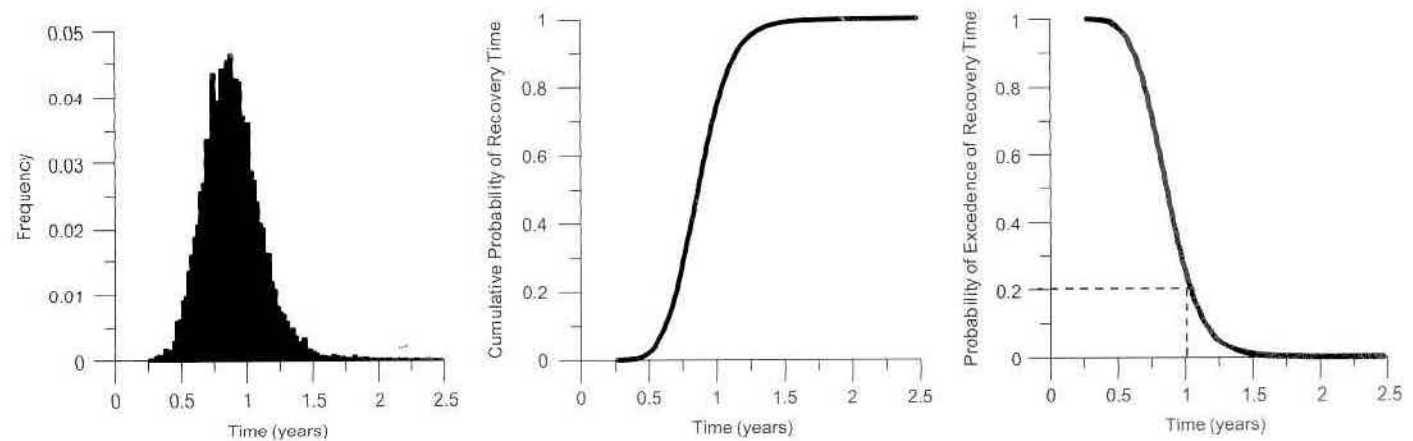


Fig. 6. Analysis results for recovery of Lake Pontchartrain from benzene load with an uncertain source rate.

6. Discussion of results

Atrazine has a mean recovery time of 1.32 years for Case I where the source rate is kept fixed at 0.1% (by floodwater volume). The maximum recovery time for the same conditions is 3 years. Given the boundary and initial

conditions considered in this model, the probability of exceedence for Lake Pontchartrain to recover from atrazine in 1 year is 0.1, as seen in the complementary cumulative probability plot in Fig. 3. Atrazine is the most soluble chemical in water among the three representative contaminants selected in this study. Decay within the

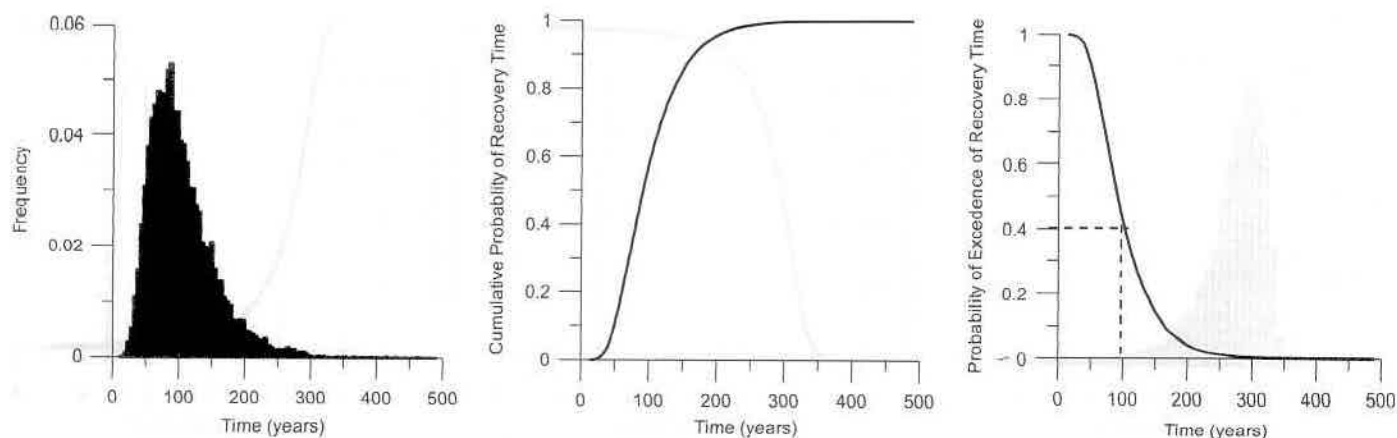


Fig. 7. Analysis results for recovery of Lake Pontchartrain from PCB load with source rate fixed at 0.1%.

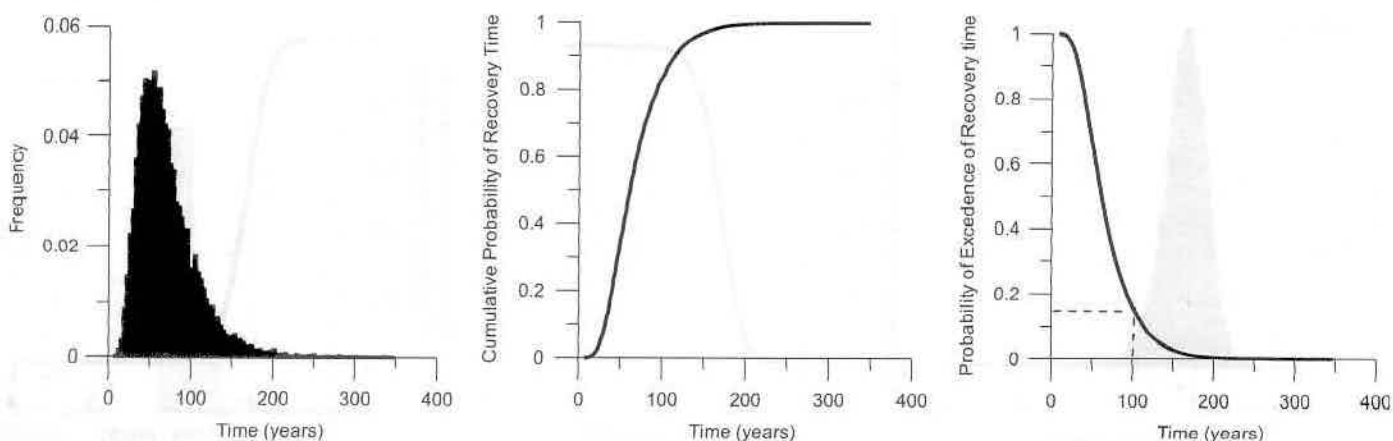


Fig. 8. Analysis results for recovery of Lake Pontchartrain from PCB load with an uncertain source rate.

relatively larger volume of the water compartment and outflow through tributaries allow for rapid removal of atrazine. For Case II, where the source rate is also randomly selected, atrazine has a mean recovery time of 0.9 years. For Case II, the probability of exceedence of MCL value within a year has a higher value of 0.3. The cumulative probability plot shows that atrazine will reach MCL value within 3 years (Fig. 4).

In Fig. 5, the mean time to recover from benzene loading is 1.12 years for Case I. The same figure also shows that although there is a 60% probability of exceedence of 1 year recovery time for the benzene concentration to reach MCL value, MCL value will be reached within 2 years. When source rate is also randomly sampled, as in Case II, the mean recovery time is 0.8 years for benzene (Fig. 6). In this case, the probability that benzene concentration will still be higher than MCL value after the first year is 0.2. In both cases, since benzene is the most volatile among the three contaminants, it is rapidly removed from Lake Pontchartrain system. Based on the input data used here, Lake Pontchartrain would recover from benzene contamination within 2 years.

PCB is the most hydrophobic of the contaminants and is expected to stay in the system for the longest period of time. As shown in Figs. 7 and 8, PCBs are indeed very persistent, and stay in the lake system the longest period of time. In Case I, with the fixed source rate, Lake Pontchartrain requires a mean recovery time period of 100 years and maximum recovery time period of 400 years to clean itself from PCB contamination. For the conditions considered in this study, the probability of exceedence from PCB loading in 100 years is 0.4. For Case II, the mean recovery time period drops to 65 years. However, the probability of recovery being more than 100 years is still high with a value of 0.15.

7. Conclusion

In this study, an unsteady-state fugacity model was developed in order to examine the self-cleaning capacity of Lake Pontchartrain from three potential contaminants in the floodwaters. Using this model, fugacity values and the analogous contaminant concentrations in water, air, and sediment compartments were calculated as a function of

time. Monte Carlo analysis was performed as a part of the developed model to reflect the uncertainties in the parameters used in the solution of the model.

It is observed that if the contaminant is hydrophobic, it will resist exiting the system. However, if the contaminant is volatile, it will stay in the system for a shorter period of time. The amount of contaminant load introduced into the system is a crucial parameter in determining the time of recovery. In evaluating the terms that are uncertain, use of Monte Carlo Simulation is an effective tool.

Although the model is designed to understand the behavior of contaminants in Lake Pontchartrain after Hurricane Katrina, it has various other applications. The model developed in this study offers a very simple yet effective procedure to assess environmental response in surface water systems. It can be used effectively in making emergency decisions and in designing emergency response scenarios as well as in analyzing the potential effect of each scenario on the environment. The model is general, thus it can be used to study the effect of any contaminant load for any shallow water system. The results would be very informative for decision makers who are forced to find solutions in very short time periods during emergencies. It can also be used proactively in advance of natural disasters to identify which chemicals may have the most severe and longest-lasting effects. This would be very useful in planning stages of various emergency response scenarios. It can be used as a screening level model to examine the consequences of short-term measures in emergency situations.

The analysis described above could be very effective for decision making and in preventing decisions that may lead to long-term pollution of surface water systems. If the results of this study were known at the time of Hurricane

Katrina, alternative solutions could have been considered. One possibility would be the use of primary treatment prior to discharging floodwaters into Lake Pontchartrain using large-scale portable treatment devices. That measure would have reduced the problem significantly rather than leaving the cleaning to natural attenuation alone.

References

- Citra, M.J., 2004. Incorporating Monte Carlo analysis into multimedia environmental fate models. *Environmental Toxicology and Chemistry* 23 (7), 1629–1633.
- EPA, 2004. Drinking Water Standards and Health Advisories (No. EPA 822-R-04-005). Environmental Protection Agency, Washington, DC.
- Gentle, J.E., 2003. Random Number Generation and Monte Carlo Methods, second ed. Springer, New York, NY.
- Howard, P.H., 1989. Handbook of Environmental Fate and Exposure Data for Organic Chemicals. Lewis Publishers, Boca Raton, FL.
- Mackay, D., 1989. Modeling long-term behavior of an organic contaminant in a large lake: application to PCBs in Lake Ontario. *Journal of Great Lakes Research* 15 (2), 283–297.
- Mackay, D., 2001. Multimedia Environmental Models: The Fugacity Approach, second ed. Lewis Publishers, Boca Raton, FL.
- Mackay, D., Shiu, W.Y., Ma, K.C., 1992. Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals. Lewis Publishers, Boca Raton, FL.
- MacLeod, M., Fraser, A.J., Mackay, D., 2002. Evaluating and expressing the propagation of uncertainty in chemical fate and bioaccumulation models. *Environmental Toxicology and Chemistry* 21 (4), 700–709.
- Reible, D.D., 1998. Fundamentals of Environmental Engineering. Lewis Publishers, Boca Raton, FL.
- Ronen, Y., 1988. Uncertainty Analysis. CRC Press, Inc., Boca Raton, FL.
- Stumm, W., Morgan, J.J., 1996. Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters. Wiley, New York, NY.
- USGS, 2002. Environmental Atlas of Lake Pontchartrain Basin (No. Open Report 02-206). United States Geological Survey, Reston, VA.
- Valsaraj, K.T., 2000. Elements of Environmental Engineering Thermodynamics and Kinetics. Lewis Publishers, Boca Raton, FL.

Fuzzy pricing for urban water resources: Model construction and application

Ranhang Zhao^{a,b,*}, Shouyu Chen^b

^a*School of Civil & Hydraulics Engineering, Shandong University, Jinan 250061, PR China*

^b*School of Civil & Hydraulics Engineering, Dalian University of Technology, Dalian 116024, PR China*

Received 8 June 2006; received in revised form 9 February 2007; accepted 11 March 2007

Available online 11 May 2007

Abstract

A rational water price system plays a crucial role in the optimal allocation of water resources. In this paper, a fuzzy pricing model for urban water resources is presented, which consists of a multi-criteria fuzzy evaluation model and a water resources price (WRP) computation model. Various factors affecting WRP are comprehensively evaluated with multiple levels and objectives in the multi-criteria fuzzy evaluation model, while the price vectors of water resources are constructed in the WRP computation model according to the definition of the bearing water price index, and then WRP is calculated. With the incorporation of an operator's knowledge, it considers iterative weights and subjective preference of operators for weight-assessment. The weights determined are more rational and the evaluation results are more realistic. Particularly, dual water supply is considered in the study. Different prices being fixed for water resources with different qualities conforms to the law of water resources value (WRV) itself. A high-quality groundwater price computation model is also proposed to provide optimal water allocation and to meet higher living standards. The developed model is applied in Jinan for evaluating its validity. The method presented in this paper offers some new directions in the research of WRP.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Fuzzy pricing model; Multi-criterion fuzzy evaluation; Urban water resources; Water price; Water resources value

1. Introduction

As one of most precious natural resources, water is indispensable for human beings and economic growth. It is also a matter of cost and value and not just quantity, which cannot be exempt from the principles of economics (Franklin et al., 2005). The crisis of water shortage has aroused extensive attention in the history of human kind, especially in recent decades (Gleick, 1993; Liu and He, 1999; Biswas, 1991). In China, with the rapid process of urbanization, economic growth, and increase in population, serious shortage of water resources, threats of flooding, and aquatic environmental deterioration have become urgent issues. In response to these concerns, the government has embarked on a large-scale investment

program to increase the capacity of groundwater and surface water sources on the one hand, for instance, south–north water transfer projects (Liu and He, 1999; Wang and Yang, 2005). On the other hand, much attention has been paid to other alleviating measures such as increasing water use efficiency, rainwater harvesting, and water recycling. Water price is considered as an instrument for satisfactory of water allocation. It has been proved that improving water price system for promoting water saving is an important approach for maintaining the continuous operation of hydraulic engineering and realizing sustainable development in urban water resources and economy (Rogers et al., 2002; Ioslovich and Gutman, 2001; Martinez-Espinoeira, 2002; Amigues et al., 1997; Wang and Xu, 2003).

In the past, most cities and utilities in China have provided water to their customers almost free of charge as a basic necessity because water was a relatively cheap and abundant resource. Current water pricing is inefficient in most utilities as prices are based on average embedded

*Corresponding author. School of Civil & Hydraulics Engineering, Shandong University, Jinan 250061, PR China. Tel.: +86 531 88395214; fax: +86 531 88392845.

E-mail address: ranhang-z@sdu.edu.cn (R. Zhao).

utility costs using historical accounting costs rather than the effective economic costs. Most water utilities do not consider factors such as the cyclicity of demands, the time-of-use, and real water resources value (WRV) (scarcity) (Elnaboulsi, 2001).

Besides, the traditional way of handling water supply used for a long period in most cities in China, was supplying water for all users by the standard for drinking water using one set of pipeline system. This old fashioned way was efficient for the situation of less water demand and single purpose. Now many changes occurred such as the lack of high-quality water resources, the diversified uses of water resources, and intensified water pollution, consequently most of the water supplied under the current approach cannot satisfy the demand for high-quality drinking water. So the quality-divided water supply is necessary for the security of drinking water, and high efficient utilization of water resources.

Dual water supply is a new way that supplies water with different qualities to different users with various use purposes (Fen et al., 2004). The water supply system with two kinds of pipelines is constructed, pipelines for high-quality water (e.g., groundwater) coming from advanced treatment of mains water or raw water with purge switches, which supplies for special industrial water (e.g., medicine, foodstuff, and other special industry) and potable domestic water (e.g., drinking water, bathing); and pipelines of the current water supply system for lower quality water supply for general industry, household water, and ecological environment (Qin, 2002). Accordingly, the current water price system must be reformed as different prices are set for water resources with different qualities.

Many studies have been conducted in water resources price (WRP). The main reported models include shadow price model (Franklin et al., 2005; Ioslovich and Gutman, 2001; van Soest et al., 2006; Wang et al., 1999), marginal opportunity cost model (MOC) (Mann et al., 1980; Mann and Schlenger, 1982; Hanke, 1981; Turvey, 1976; Robert and Richard, 1981a, b; Warford, 1997; Zhang, 2000), and fuzzy comprehensive evaluation model (Jiang, 1998), which offer alternatives for the evaluation of WRP. Recently, in order to meet the growing demand and diversified purposes for water, many investigators have increasingly researched on fixing different prices for water with different qualities (Hurlimann and McKa, 2006). These models have their own merits and demerits (Jiang et al., 1998).

Shadow price originated from mathematical programming and has a long history in economics. It is often used in optimization problems subject to constraints (Tinbergen, 1952, 1955; Schiffler, 1998). Recently, it has been developed significantly in theory (Baumgardner, 1991). In general, four computation methods are used for solving shadow price: linear programming, adjusting domestic market price, adjusting international market price, and opportunity cost method. In many applications these shadow prices directly correspond to prices of goods or services in particular future or current markets that, for one reason

or another, do not exist. While the applications in water resources (Jorgenson and Slesnick, 1983; Jorgenson, 1995; Franklin et al., 2005; Ioslovich and Gutman, 2001; van Soest et al., 2006; Wang et al., 1999) have some limitations. Shadow water price can only reflect the degree of some water resources shortage and the relation between water resources and whole economic benefit, it is remarkably different from production price and market price and cannot substitute WRV; as well as, it is very difficult to obtain shadow water price through adjusting domestic and international market prices, and in practice it is almost impossible to obtain shadow water price by solving a linear programming.

Since Turvey (1976), there has been an extensive use of MOC pricing methodology for water resources. MOC of water consists of marginal water production costs (MPC), marginal user costs (MUC), and marginal environment costs (MEC) (i.e., $MOC = MPC + MUC + MEC$) (Warford, 1997). Generally, MOC pricing of water services can assure the optimal allocation of scarce resources as it sends the right signal to both investors and consumers. However, there are some problems in the applications (Warford, 1997). It is often difficult to estimate MUC and MEC, which may involve prediction of costs of depletion or substitutes many years hence. Moreover, WRP is involved not only in quantity, but also in quality. For example, wastewater presents negative value while high-quality water indicates high value.

A comprehensive fuzzy model was proposed by Jiang (1998). Considering WRV system being fuzzy, Jiang (1998) constructed a comprehensive evaluation matrix R firstly, then calculated the comprehensive evaluation index V in accordance with resident's bearing capacity of water rate, and calculated WRP. The method is very important for quantitative evaluation of WRV with the comprehensiveness and maneuverability. But the evaluation model is linear and it is very difficult to embody the non-linear and complexity relation of water resources system.

Water resources system is complex, fuzzy and complicated, composed of three interactive parts of society, economy and environment. It has been found that the use of fuzzy set theory coupled with multi-criteria recognition methods is very promising (Wang and Xu, 2003; Li et al., 2005; Chen, 2001, 2002; Chen and Wang, 2002). In this paper, the fuzzy pattern recognition and iterative theories will be described for the evaluation of WRP, and the rational prices are proposed for urban water resources, which embody the self-characteristics and reflect the real and dynamic values of water resources. An integrated method, which amalgamates subjective preference of operators (i.e., physical weight) and iterative weights (mathematic weight), is presented for weight-assessment as the weights determined are more rational and the evaluation results are more realistic.

The paper is organized as follows. Following the introduction section, the evaluation system of affecting factors and indicators involved in urban WRP is described

in section two. In the third section, multi-criterion fuzzy pricing theory and model for urban WRR are presented. In the fourth section, the price vectors of water resources are constructed. A case study on the application of the models is included in Section 5, with conclusions drawn in Section 6.

2. Evaluating system of indicators involved in urban WRP

There are many influential factors involved in WRV. It mainly includes three aspects: (a) social aspects: population, policy, and historical culture; (b) economic aspects: Gross National Product, structure of industry, scale and level of development, and efficiency of water exploitation; (c) natural aspects: water resources quantity and quality, exploitation conditions, and degree of water environmental degeneration, as well as probability of foreign water resources utilization. The hierarchical structure of WRV evaluation is listed in Fig. 1.

It is difficult and maybe unnecessary to include all the factors in WRP model. The influential factors should be chosen according to actual conditions (Jiang et al., 1998). Meanwhile, the following four points should be followed up: (a) entirety, all aspects influencing WRR should be included; (b) delegate (characteristics), parameters chosen must be primary factors reflecting the local characteristics. The primary factors are different in different places resulting from that WRR is not separated and closely related with local conditions of society, economy and environment; (c) independence, intensions of factors chosen must not be overlapped or intercrossed; (d) maneuverability, for the convenience of model use and spread, the factors chosen can be reduced as far as possible on the base of considering for delegate, in order to reduce model parameters.

3. Multi-criteria fuzzy pricing theory and model for urban WRR

Assume $X = (x_1, x_2, \dots, x_n)$ be a finite set of n feasible alternatives in WRV operation with m criteria to be considered, then the decision-making matrix can be denoted as $X = (x_{ij})_{m \times n}$, where x_{ij} is the value of the i th criterion for the j th alternative for all $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$.

Generally, there exist three types of criteria, namely: excellent, inferior, intermediate, and can, respectively, be normalized by the following equations:

$$r_{ij} = \frac{x_{ij} - \min_j x_{ij}}{\max_j x_{ij} - \min_j x_{ij}} \quad \forall j, \quad (1)$$

$$r_{ij} = \frac{\max_j x_{ij} - x_{ij}}{\max_j x_{ij} - \min_j x_{ij}} \quad \forall j, \quad (2)$$

$$r_{ij} = 1 - \frac{|x_{ij} - x_i|}{\max_j |x_{ij} - x_i|} \quad \forall j, \quad (3)$$

in which r_{ij} is the relative membership of the j th alternative attributed to criterion i , $\max_j x_{ij}$ is the maximum value regarding criterion i , $\min_j x_{ij}$ is the minimum value regarding criterion i , x_i is the mid optimum regarding criterion i . Using Eqs. (1)–(3), the decision-making matrix can be normalized into the following relative membership matrix:

$$R = (r_{ij})_{m \times n}.$$

On the basis of m criteria, the alternative set can be assigned into c rankings with a membership function. Usually, for any criterion, as far as fuzzy superiority is concerned, the relative membership of the superior ranking (ranking 1) associated with superiority is prescribed to be 1, while the relative membership of the inferior ranking (ranking c) is prescribed to be 0. Due to the medium transitivity of fuzzy superiority, the relative memberships from ranking 1 to ranking c are linearly decreasing from 1 to 0, and the difference between the consecutive rankings is assumed to be $\Delta = 1/(c-1)$. For any criterion, the standard relative membership vector is expressed as

$$S = \left(1, \frac{c-2}{c-1}, \frac{c-3}{c-1}, \dots, 0\right) = (s_h) \quad \text{for all } h = 1, 2, \dots, c, \quad (4)$$

the relative memberships of the j th alternative $r_{1j}, r_{2j}, \dots, r_{mj}$ are compared with vector expression (4) one by one, the intervals between consecutive rankings of the relative memberships are $[a_{1j}, b_{1j}]$, $[a_{2j}, b_{2j}]$, $\dots$, and $[a_{mj}, b_{mj}]$, the ranking upper limit value b_j and lower limit value a_j of the j th alternative are as

$$\begin{cases} a_j = \min_i a_{ij}, \\ b_j = \max_i a_{ij}. \end{cases} \quad (5)$$

Let the relative membership grade matrix of the alternative set belonging to all rankings be

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{c1} & u_{c2} & \cdots & u_{cn} \end{bmatrix} = (u_{hj})_{c \times n} \quad (6)$$

in which u_{hj} is the relative membership grade of the j th alternative belonging to the h th ranking for all $h = 1, 2, \dots, c$ and $j = 1, 2, \dots, n$.

Due to j th alternative following the range $[a_j, b_j]$, the above matrix satisfies the following normalized condition:

$$\sum_{h=a_j}^{b_j} u_{hj} = 1. \quad (7)$$

For all n alternatives belonging to c rankings, the matrix U also satisfies the normalized condition.

$$\sum_{h=1}^c u_{hj} = 1. \quad (8)$$

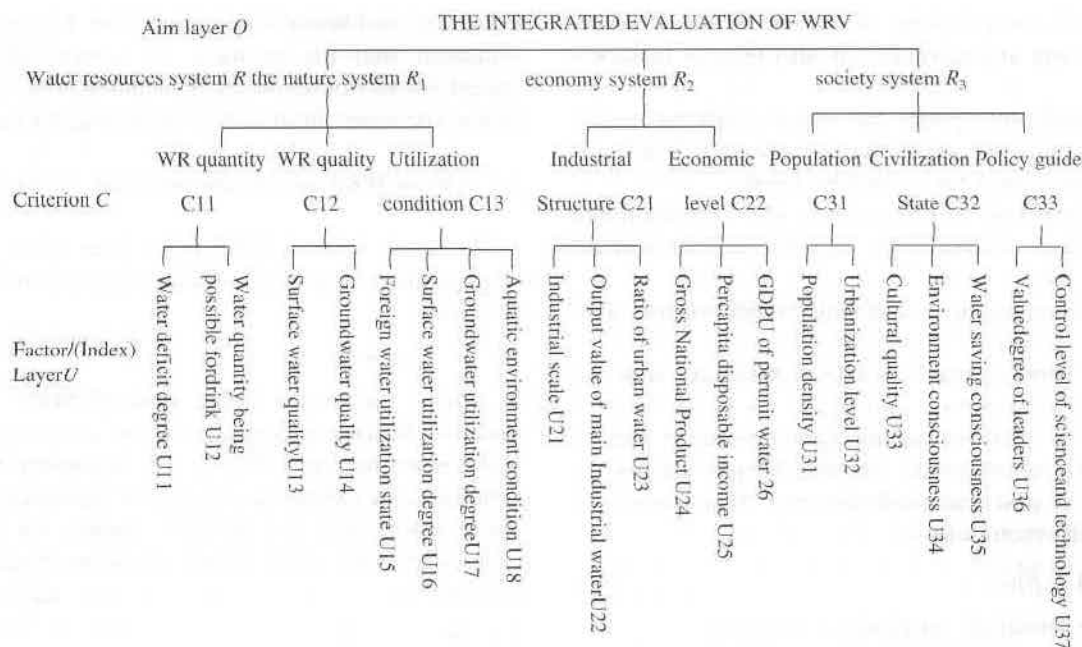


Fig. 1. The hierarchical structure chart of WRV evaluation.

Considering $a_j \geq 1$, $b_j \leq c$ and the constrained conditions (7) and (8), the Eq. (9) can be obtained

$$u_{hj} = 0 \quad \text{and} \quad h < a_j \quad \text{or} \quad h > b_j. \quad (9)$$

Let the criteria weighting vector be defined as follows:

$$W = (w_1, w_2, \dots, w_m), \quad \sum_{i=1}^m w_i = 1 \quad \forall j, \quad (10)$$

then the difference between the j th alternative and the h th ranking can be represented as the following weighted distance:

$$d_{hj} = \left\{ \sum_{i=1}^m [\omega_i (r_{ij} - s_h)]^2 \right\}^{1/2}. \quad (11)$$

Eq. (11) shows the difference between the j th alternative and h th ranking considering the objective weights. For the improving description, weighting generalized and weighted Euclidean distance is used to express the difference

$$D_{hj} = u_{hj} d_{hj} = u_{hj} \left\{ \sum_{i=1}^m [\omega_i (r_{ij} - s_h)]^2 \right\}^{1/2}, \quad (12)$$

where the objective weights and relative membership grade u_{hj} are both considered.

In order to obtain the relative membership grade of the j th alternative belonging to h th ranking, the following goal function is constructed:

$$\min \left\{ F(u_{hj}, w_i) = \sum_{h=1}^c D_{hj}^2 \right\}. \quad (13)$$

Based on Eq. (13) and constraint conditions in Eqs. (8) and (10), the following Langrange function is established:

$$L(u_{hj}, w_i, \lambda_1, \lambda_2) = \sum_{h=1}^c D_{hj}^2 - \lambda_1 \left(\sum_{h=1}^c u_{hj} - 1 \right) - \lambda_2 \left(\sum_{i=1}^m w_i - 1 \right), \quad (14)$$

where λ_1 and λ_2 are Langrange multipliers. Solving the equations

$$\frac{\partial L}{\partial u_{hj}} = 0, \quad \frac{\partial L}{\partial w_i} = 0, \quad \frac{\partial L}{\partial \lambda_1} = 0, \quad \frac{\partial L}{\partial \lambda_2} = 0, \quad (15)$$

the following multi-criteria fuzzy recognition models are obtained:

$$u_{hj} = \frac{1}{\sum_{k=1}^c \frac{\sum_{i=1}^m [w_i (r_{ij} - s_h)]^2}{\sum_{i=1}^m [w_i (r_{ij} - s_k)]^2}}, \quad (16)$$

$$w_i = \frac{1}{\sum_{i=1}^m \frac{\sum_{j=1}^n \sum_{h=1}^c [u_{hj} (r_{ij} - s_h)]^2}{\sum_{j=1}^n \sum_{h=1}^c [u_{hj} (r_{ij} - s_h)]^2}}. \quad (17)$$

The Eqs. (16) and (17) are the fuzzy pattern recognition and iteration formulas used to comprehensively evaluate water resources with multi-criteria. It is obvious that, if $c = 2$, the proposed models will be degenerated into the fuzzy optimization models used in Chen (2002).

The computational steps above seeking WRV with multi-criteria evaluation can be obtained as follows:

Step (1): According to the affecting factors involved in WRP and the characteristic values of evaluation indexes

establish the decision-making matrix X of fuzzy multi-criteria evaluation and normalize it into relative membership matrix R :

Step (2): Randomly specify the initial weighting vector $W^0 = (w_1^0, w_2^0, \dots, w_m^0)^T$, satisfying $w_i^0 > 0$ and $\sum_{i=1}^m w_i^0 = 1$, provide the precision δ for w_i and let $l = 0$;

Step (3): According to w^l , compute the corresponding membership grade matrix u_j^l by Eq. (16) and write it as: $u^l = (u_1^l, u_2^l, \dots, u_n^l)$;

Step (4): According to u^l , calculate weight vectors w^{l+1} by Eq. (17);

Step (5): If $\max_{1 \leq i \leq m} |w_i^{l+1} - w_i^l| \leq \delta$, turn into step (6); otherwise, turn into step (3) and let $l = l + 1$;

Step (6): w^{l+1} obtained are mathematic weight vectors w_s , the model can assign the physical weight vectors w_w according to the practical condition, and then, obtain the synthetic weight vectors as

$$W = \beta w_w + (1 - \beta) w_s, \quad (18)$$

where β is the sensitivity coefficient of weights, $0 \leq \beta \leq 1$.

Step (7): Integrated weight vectors can be obtained with sensitivity coefficients of weights. If the compositors of mathematic and physical weights are consistent, then the value of β can be given in $[0, 1]$ with equal difference, one can get a set of synthetic weight vectors, and get the final synthetic weights by averaging the synthetic weight vectors; if the compositors of mathematic and physical weights are inconsistent, according to the principia that the compositors of objective weights of synthetic and physical weights are consistent, a set of synthetic weight vectors can also be obtained from synthesizing physical and mathematic weight vectors with available sensitivity coefficients of weights, and the final synthetic weights can be obtained by averaging synthetic weight vectors;

Step (8): Substitute the final integrated weight vectors into Eq. (16) directly, and calculate the relative degrees of membership u_j^{l+1} for all objectives.

4. Calculation model of urban WRP vectors

The price vectors are decided by the social bearing capability. The bearing capability is the capability, including matter and psychology, that people still can keep the normal state under the stimulation from some signals (Gu et al., 2002; Jiang et al., 1998).

The bearing index k of water rate is used to express the social bearing capacity on WRP. The bearing index k is the proportion of the payment for water to the real income, expressed by the following formula:

$$k = \frac{O}{I}. \quad (19)$$

in which O is the payment for water rate, I is the real income, and k is the bearing index of water rate. The World Bank and some international lending institutions figured out that it is a practical and feasible criterion for industrial water to take 2–3% of industrial output as water rate as

Economic and Social Commission for Asia and the Pacific suggested that the payment for water rate should not exceed 3% of household income (Jiang et al., 1998). In this paper, the upper limit of k is prescribed to be 3%.

4.1. Urban WRP vectors computation

The upper limit of WRP is the price when the maximal bearing index is arrived, which is denoted as the following formula:

$$P_{W_1} = K_{\max} \bullet I / C, \quad (20)$$

in which, P_{W_1} is the upper limit of WRP, $K_{\max}$ is the maximal bearing index, C is water consumption. So the WRP is in the range of $[P_{W_1}, 0]$. According to the actual condition, the range $[P_{W_1}, 0]$, can be separated and taken as price vectors with the different ranges, for example, the equal difference, then, water resources price vectors are denoted as

$$P = (P_{W_1}, \dots, 0). \quad (21)$$

Combining the price vectors proposed with Eqs. (6), (16) and (17), the relative membership grades calculated in fuzzy pattern recognition model above, which are dimensionless and vector values, can be transferred into corresponding WRP which are scalar values. The expression can be obtained as

$$V = U \bullet P^T. \quad (22)$$

In Eq. (22), urban WRV is embodied by the prices.

4.2. Urban WRP vectors computation for dual water supply

In order to clearly, directly and synthetically reflect the social bearing capacity in groundwater resources price with high quality, the WRP vector of dual water supply is structured:

$$P_{F_i} = P_{F_{(i-1)}} - \Delta P_F, \quad i = 1, 2, \dots, m, \quad (23)$$

then

$$P_F = (P_{F_1}, P_{F_2}, \dots, P_{F_m}) \quad (24)$$

and

$$\Delta P_F = P_{F_1} / (m - 1), \quad (25)$$

in which P_{F_1} is social bearing capability to prices of high-quality water resources (i.e., the general bearing capability of all the departments using high-quality water, including resident, service business for food and drinking, as well as, special industries). P_{F_1} can be calculated by

$$P_{F_1} = W_1 K_1 I / (365 \times N_1) + 10000 W_2 K_2 / N_2 + 10000 W_3 K_3 / N_3, \quad (26)$$

where K_1 , K_2 and K_3 are the coefficients of bearing capability for resident, special industry and service business for food and drinking, respectively; N_1 is the necessary water consumption of high quality per capita per day, N_2 and N_3 are the water demands of output values of 10000

Yuan for special industry and service business, respectively. W_1 , W_2 and W_3 are ratios of high-quality water demands and the total water demands of resident, special industries and service business for food and drinking, respectively. I is average per capital disposable income.

The computation model of high-quality WRP for dual water supply can then be obtained:

$$V_F = U_F \bullet P_F^T. \quad (27)$$

5. Case study in Jinan city

As the capital of Shandong province of China, Jinan is a famous city for many springs of different types. However, with the development of society and economy, there exists serious shortage in water resources, and the contradictions between supply and demand are becoming more and more serious. So the research about the reasonable prices has practical meaning and far-reaching effects for settling water resources shortage, keeping the various springs gushing, renewing the connotation in spring-city, and improving the people's husbandly consciousness.

5.1. WRP computation for general water resources

5.1.1. Selection of affecting factors of WRR

According to above principles and Jinan's actual conditions, seven main factors are chosen in the model used in this paper. They are: water quality (to reflect the local quality of water resources), water quantity (to reflect the local quantity of water resources), per capita disposable income (to reflect the economic level of the native), gross national product (to reflect the requirement of water and development level of economy), population or population density (to reflect social factor and domestic demand), aquatic environmental degradation degree (to reflect the states of water resources development and utilization), and availability of replacement resource (to reflect the possibility of foreign water resources).

5.1.2. Multi-criteria fuzzy evaluation on WRP factors

Water quality data (in 2000) are taken as a case to explain the determination of membership grade vectors. The municipal water supply sources in Jinan mainly include Wohu mountain reservoir, Jinxiuchuan reservoir, Que mountain reservoir, and Yuqinghu reservoir. The water quality index of the four water sources are selected to reflect the local water quality, the monitoring data of 10 substances, such as, DO, BOD, COD etc., are taken as the basic data (Jinan Statistics Bureau, 2001), then characteristic value matrix of the objective set is as $X = (x_{ij})_{10 \times 4}$, and the relative membership matrix can be obtained as $R = (r_{ij})_{10 \times 4}$ by using normalizing Eqs. (1)–(3).

Referring to national standard of environmental quality for surface water (GB-3838-88) used currently in China, water quality is divided into five levels (i.e., $c=5$), with Eq. (4), the relative membership vectors of the

standard objectives can be obtained as $S = (s_h)$, for all $h = 1, 2, \dots, 5$. Taking initial weight vectors as $w^0 = (w_1^0, w_2^0, \dots, w_m^0)^T = (\frac{1}{10}, \frac{1}{10}, \dots, \frac{1}{10})^T$, substituting R , S and the weight vectors w^0 into Eq. (16), and making the loop iteration computation according to the steps above, mathematic weight vectors w_s are obtained until $\max_{1 \leq i \leq m} |w_i^{t+1} - w_i^t| \leq \delta$ (in this paper, $\delta = 0.0001$). The physical weight vectors w_w can be obtained on the basis of various opinions of relative experts and others. The integrated weight vectors W are obtained from sensitivity coefficients of weights with Eq. (18). Substituting W into Eq. (16), the relative degrees of membership $U = (u_{hj})$ can be obtained.

Then, according to the difference of service discharge in the four water heads, the situation of water quality in Jinan is calculated by weighted mean using the supplied proportion as weights, here, calculation process is omitted.

The relative membership degree vectors of water quality calculated are as follows:

$$U_1 = (0.548, 0.182, 0.135, 0.135, 0.000).$$

Similarly, the standard of per service discharge is determined by water-stress index (e.g., Malin Falkenmark), and per capita gross national product is classified and ranked with the bases of statistic data in statistical yearbook of China (Wang and Xu, 2003).

Then the relative membership degree vectors of the other six factors above are as follows:

$$U_2 = (0.102, 0.701, 0.197, 0.000, 0.000),$$

$$U_3 = (0.000, 0.000, 0.109, 0.891, 0.000),$$

$$U_4 = (0.000, 0.000, 0.818, 0.182, 0.000),$$

$$U_5 = (0.000, 0.000, 0.000, 0.793, 0.207),$$

$$U_6 = (0.598, 0.402, 0.000, 0.000, 0.000),$$

$$U_7 = (0.000, 0.299, 0.497, 0.204, 0.000).$$

The relative membership degree matrix of all factors for WRP is as follows:

$$R = \begin{bmatrix} 0.548 & 0.182 & 0.135 & 0.135 & 0.000 \\ 0.102 & 0.701 & 0.197 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.109 & 0.891 & 0.000 \\ 0.000 & 0.000 & 0.818 & 0.182 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.793 & 0.207 \\ 0.598 & 0.402 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.299 & 0.497 & 0.204 & 0.000 \end{bmatrix}.$$

Taking initial weight vectors as $w^0 = (w_1^0, w_2^0, \dots, w_m^0)^T = (\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7})^T$, substituting R , S and the weight vectors w^0 into Eqs. (16) and (17), making the loop iteration computation according to the steps above, fuzzy synthetic evaluation vectors of water resources in Jinan are obtained as

$$U = (0.234, 0.490, 0.205, 0.071, 0.000).$$

5.1.3. Calculating WRP V

It is crucial in the WRP vectors determination to determine the upper limit of the price. In 2000, the average amount of water consumption is 110 L per day per capita, and the actual disposable income per capita is 8471 Yuan in Jinan (Jinan Statistics Bureau, 2001); if the bearing index of water rate for domestic water is 0.03, then the upper limit of WRP according to Eq. (20) is as follows:

$$P_{W_1} = K_{\max} \bullet I/C = 0.03 \times 8471 / (110 \times 30 \times 12 \times 0.001) \\ = 6.42 \text{ Yuan/m}^3.$$

Then the number 6.42 is divided equally. The equal difference is 1.605, so the vectors of WRP are as

$$P = (6.420, 4.815, 3.210, 1.605, 0.000).$$

With Eq. (22), WRP in 2000 can be obtained as

$$V = (0.234, 0.490, 0.205, 0.071, 0.000) \\ \bullet (6.420, 4.815, 3.210, 1.605, 0.000)^T \\ = 4.63 \text{ Yuan/m}^3.$$

5.1.4. Result analysis

Comparing the WRP 4.63 Yuan/m³ obtained above with data in 2000 with the result of 4.88 Yuan/m³ (Hu et al., 2003), the two results are nearly consistent. While the actual water price in 2000 is 1.75 Yuan/m³, water rate only accounts for 0.9% of the resident real income, which is much less than 3% proposed by ESCAP. The comparison indicates that the current price is on the low side, resident bearing capacity on increasing water price can be further heightened.

During the last decade many studies have been devoted to identifying the response of water demand to price (Shin, 1985; Nieswiadomy and Molina, 1988, 1989; Opaluch, 1982, 1984; Chicoine and Ramamurthy, 1986; Billings, 1982; Hewitt and Hanemann, 1995; Corral et al., 1998; Pint, 1999; Martinez-Espinoira, 2002), indicating that water price changes have a measurable effect on aggregate water demand. For example, marginal price elasticity lies between −0.12 and −0.17 (Martinez-Espinoira, 2002), own-price elasticity ranges from −0.15 to −0.52 (Nieswiadomy, 1992; Nieswiadomy and Molina, 1989; Billings, 1987; Moncur, 1987; Agthe et al., 1986) for urban areas as in California own-price elasticity ranges from −0.22 to −0.37 (Renwick and Archibald, 1998; Berk et al., 1980).

Assuming own-price elasticity being −0.22, implying a 10% increase in price will reduce the aggregate quantity demanded by 2.2%. The increase in price from 1.75 Yuan/m³ in practice to 4.63 Yuan/m³ calculated above, would reduce the water demand by 36.3%. That is, the water demand would be reduced to 79.18 × 10⁶ m³ from 124.30 × 10⁶ m³ in 2000 in Jinan city. Therefore, the rational prices for water resources are very important for water saving and the relief of water resources shortage.

5.2. Fuzzy model of water pricing for dual water supply

The implementation of dual water supply and fixing a high price for high-quality water (groundwater) maybe an effective way for realizing the “semblance” (spring gushing) and the “connotation” (people in the city can drink high-quality groundwater) of the spring city being consistent, and restoring the visualization of spring city, as well as reducing pressure in water resources.

5.2.1. Selection of the influential factors

According to the practical conditions, five main factors are chosen in the model: water quality, water quantity, per disposable income, degree of aquatic environmental degradation, and availability of replacement resource.

5.2.2. Multi-criterion fuzzy evaluation for dual water supply

Using the same procedure as 5.1.2, the relative membership degree vectors of the five factors are as follows:

$$U_1 = (0.548, 0.182, 0.135, 0.135, 0.000), \\ U_2 = (0.892, 0.108, 0.000, 0.000, 0.000), \\ U_3 = (0.000, 0.000, 0.109, 0.891, 0.000), \\ U_4 = (0.486, 0.149, 0.000, 0.000, 0.000), \\ U_5 = (0.000, 0.000, 0.000, 0.793, 0.207).$$

The relative membership degree matrix of all factors for WRP is as follows:

$$R_F = \begin{bmatrix} 0.798 & 0.202 & 0.000 & 0.000 & 0.000 \\ 0.892 & 0.108 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.110 & 0.890 & 0.000 \\ 0.486 & 0.514 & 0.000 & 0.000 & 0.000 \\ 0.851 & 0.149 & 0.000 & 0.000 & 0.000 \end{bmatrix}.$$

Taking initial weight vectors as $w^0 = (w_1^0, w_2^0, \dots, w_m^0)^T = (\frac{1}{5}, \frac{1}{5}, \dots, \frac{1}{5})^T$, the loop iteration computation is made again according to the steps above and fuzzy synthetic evaluation vectors of water resources in Jinan are obtained as

$$U_F = (0.6814, 0.1468, 0.0430, 0.1288, 0.0000).$$

5.2.3. WRP V_F calculation

The per capita disposable income is 8,471 Yuan, groundwater exploited is 387.81 × 10⁶ m³, in which industrial water is 203.67 × 10⁶ m³ accounting for 52.5% of the total water, domestic water in city is 92.03 × 10⁶ m³ occupying 23% of the total, and farm irrigation and domestic water are 92.11 × 10⁶ m³ taking 23.8% in 2000 (Jinan Statistics Bureau, 2001). The highest social bearing capability to the high-quality groundwater can be calculated with Eq. (26) as 0.139 Yuan/L. So the price vectors are as

$$P_F = (0.139, 0.104, 0.070, 0.035, 0.000).$$

Substituting U_F and P_F into Eq. (27), high-quality groundwater resources price in 2000 can be obtained as

$$V_F = U_F \bullet P_F^T = 0.11750 \text{ (Yuan/L)} = 117.50 \text{ Yuan/m}^3.$$

5.2.4. Result analysis

From the above results computed, one can point out that high-quality groundwater resources price is quite higher than these of general water resources. Comparing it with the prices of “pure water” and “mineral water” for sale at present, it is obviously smaller than those. Therefore, from the water price’s point of view, dual water supply may be accepted at a certain degree.

Besides, with dual water supply operation, on the one hand, fixing a higher price for high-quality groundwater in spring area can be propitious for protecting groundwater and springs, at the same time, resident of the city is able to drink high-quality water. On the other hand, fixing the lower prices for the general water resources used in industry, environment and living for no drinking can reduce the influence of water prices on social bearing capabilities, and realize the sustainable developments of water resources and social economy.

6. Summary and conclusions

This paper presents a fuzzy pricing approach to fix prices for urban water resources, based on fuzzy pattern recognition and iterative theories. The methodology can be summarized in four steps: (a) select and evaluate the factors affecting WRP using the fuzzy theory of multi-criteria evaluation; (b) compute the relative membership degree matrix of all factors and determine the initial weight vectors; (c) make the loop iteration computations and obtain the fuzzy synthetic evaluation vectors of water resources; (d) construct WRP vectors and calculate WRP according to the bearing capacity of water rate. The case study at Jinan city in Shandong province of China illustrates the developed method.

The method presented in this paper offers some new directions in the research of WRP. It embodies the characteristics of WRV (including complexity, fuzziness and development), overcomes the limitations of traditional model of fuzzy comprehensive evaluation in which optimization criterion parameter α and distance parameter p were both taken as 1, and portrays the non-linear and complexity relation of water resources system. With the incorporation of operator’s knowledge, it considers iterative weights and subjective preference of operators for weight-assessment. The weights determined are more rational and the evaluation results are more realistic. Particularly, we propose that dual water supply should be adopted and fixing a higher price for high-quality water (groundwater).

For the case study, the prices are calculated using the data in 2000 at Jinan city as static values. Considering the dynamics of the factors affecting WRP such as water quality, water quantity, and others, the different prices

under different conditions can be calculated with the method proposed by modifying the criterion relative membership vectors through altering the criterion characteristic values. Therefore, this method is possible to consider the dynamics of WRP system with various factors.

The dual water supply is one of the important measures for optimal water resources allocation. The pricing model for dual water supply proposed in this paper is based on the full consideration for various users’ social bearing capability and demand of high-quality groundwater. Different prices being fixed for water resources with different qualities conforms to the law of WRV self. These prices computed in this paper can decrease the adverse impact of environmental pollution, reduce water treatment load, reduce the hidden danger of drinking water security, and avoid the waste of high-quality water.

The results indicate that the fuzzy model of fixing prices for urban water resources is practical and can be applied to other similar cities.

Acknowledgments

The authors would like to thank two anonymous referees for their careful reviews and constructive comments. We also thank Dr. Shuming Liu from the University of Exeter and Zhongbo Yu from the Department of Geoscience University of Nevada for their valuable suggestions and editorial revision. This work was funded partly by the Ministry of Water Resources (China) technological innovation grant (SCXC-2005-01), and specialized research fund for doctorate subjects of higher education (9014102).

References

- Agthe, D.E., Billings, R.B., Dobra, J.L., Raffice, K., 1986. A simultaneous equation demand model for block rates. *Water Resources Research* 22 (1), 1–4.
- Amigues, J.P., Ge'ard, G., Moreaux, M., 1997. Rational pricing of groundwater with a growing population. *Environmental Modeling and Assessment* 2, 323–331.
- Baumgardner, J., 1991. The interaction between forms of insurance contract and types of technical change in medical care. *Rand Journal of Economics* 22 (1), 36–53.
- Berk, R.A., Cooley, T.F., LaCivita, C.J., Parker, S., Sredl, K., Brewer, M., 1980. Reducing consumption in periods of acute scarcity: the case of water. *Social Science Research* 9 (2), 99–120.
- Billings, R.B., 1982. Specification of block rate price variables in demand models. *Land Economics* 58 (3), 386–394.
- Billings, R.B., 1987. Alternative demand model estimators for block rate pricing. *Water Resources Bulletin* 23 (2), 341–345.
- Biswas, A.K., 1991. Water for Sustainable Development in the 21st century. Address to 7th world Congress on Water Resources, Water International, Morocco.
- Chen, S.Y., 2001. The model and method for qualitative assessment of sustainable development of regional water resources. *Engineering Science* (2), 33–38.
- Chen, S.Y., 2002. Optimal Fuzzy Recognition Theory and Application of Complex Water Resources System. JiLin University Press, Jilin, China.
- Chen, S.Y., Wang, J.M., 2002. Fuzzy pattern recognition model for supervision of abundance of river basin water resources. *Water Resources Protection* 2, 1–3.

- Chicoine, D., Ramamurthy, G., 1986. Evidence on the specification of price in the study of domestic water demand. *Land Economics* 62 (1), 26–32.
- Corral, L., Fisher, A.C., Hatch, N., 1998. Price and non-price influences on water conservation: an econometric model of aggregate demand under non-linear budget constraint. Working paper No. 881, Department of Agricultural and Resource Economics and Policy, University of California at Berkeley.
- Elnaboulsi, J.C., 2001. Nonlinear pricing and capacity planning for water and wastewater services. *Water Resources Management* 15, 55–69.
- Fen, Z.T., Yang, X.W., Liu, Y.L., 2004. Discussion on the necessity and restraint factors of dual water supply in China cities. *China Management of Public Health Journal* 2, 167.
- Franklin, M.F., Annette, H.L., Ilan, A., Shaul, A., Zvi, E., Munther, J.H., Salem, G.H., Ammar, J., Anan, J., Uri, S., Hans, W., 2005. *Liquid Assets: An Economic Approach for Water Management and Conflict Resolution in the Middle East and Beyond*. RFF Press, Washington, USA.
- Gleick, P.H., 1993. *Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford University Press, UK.
- Gu, S.P., Lin, R.Y., Liu, H.L., 2002. The fuzzy pricing model of water resource. *Journal of Research on Developing Water Power* 2, 9–12.
- Hanke, S.H., 1981. On the marginal cost of water supply. *Water Engineering and Management* 120, 60–63.
- Hewitt, J.A., Hanemann, W.M., 1995. A discrete/continuous choice approach to residential water demand under block rate pricing. *Land Economics* 71 (2), 173–192.
- Hu, Y., Cao, S.L., Zhao, R.H., Wang, X.J., Li, W.Y., 2003. Research on water resource value model. *Journal of Shandong University* 5, 342–345.
- Hurlimann, A., McKa, J., 2006. Urban Australians using recycled water for domestic non-potable use—an evaluation of the attributes price, saltiness, colour and odour using conjoint analysis. *Journal of Environmental Management* 2, 2–12.
- Ioslovich, I., Gutman, P.O., 2001. A model for the global optimization of water prices and usage for the case of spatially distributed sources and consumers. *Mathematics and Computers in Simulation* 56, 347–356.
- Jiang, W.L., 1998. A study on model of water resources value. *Resources Science* 1 (1), 35–43.
- Jiang, W.L., Wu, X., Lin, T.F., 1998. Study on assessment for water resources value model. *Journal of Advance in Earth Sciences* 2, 179–183.
- Jinan Statistics Bureau, 2001. *Jinan Statistical Yearbook*. China Statistics Press, Beijing, China.
- Jorgenson, D.W., 1995. *Productivity*, vols. 1 and 2. The MIT Press, Cambridge, MA.
- Jorgenson, D.W., Slesnick, D.T., 1983. Individual and social cost of living indexes. In: Diewert, W.E., Montmarquette, C. (Eds.), *Price Level Measurement Statistics*. Canada, Ottawa, pp. 241–336.
- Li, Y.W., Chen, S.Y., Fu, T., 2005. Comprehensive evaluation of water resource carrying capacity based on fuzzy recognition. *Advances in Water Science* 9 (5), 726–729.
- Liu, C.M., He, X.W., 1999. *China Water Strategies for the 21st Century*. Press of Science and Technology, Beijing, China.
- Mann, P.C., Schlenger, D.L., 1982. Marginal cost and seasonal pricing of water service. *American Water Works Association Journal* 74, 6–11.
- Mann, P.C., Sanuders, R.J., Warford, J.J., 1980. A note on capital indivisibility and the definition of marginal cost. *Water Resources Research* 16, 602–604.
- Martinez-Espineira, R., 2002. Residential water demand in the northwest of Spain. *Environmental and Resource Economics* 21, 161–187.
- Moncur, J.E.T., 1987. Urban water pricing and drought management. *Water Resources Research* 23 (3), 393–398.
- Nieswiadomy, M.L., 1992. Estimating urban residential water demand: effects of price structure, conservation, and public education. *Water Resources Research* 28 (3), 609–615.
- Nieswiadomy, M.L., Molina, D.J., 1988. Urban water demand estimates under increasing block rates. *Growth and Change* 19 (1), 1–12.
- Nieswiadomy, M.L., Molina, D.J., 1989. Comparing residential water estimates under decreasing and increasing block rates using household data. *Land Economics* 65 (3), 280–289.
- Opaluch, J., 1982. Urban residential demand for water in the United States: further discussion. *Land Economics* 58 (2), 225–227.
- Opaluch, J., 1984. A test of consumer demand response to water prices: reply. *Land Economics* 60 (4), 417–421.
- Pint, E., 1999. Household responses to increased water rates during the California drought. *Land Economics* 75 (2), 246–266.
- Qin, Y.H., 2002. *Sanitation and Treatment Technology of Drinking Water*. Chemical Industry Press, Beijing, China.
- Renwick, M.E., Archibald, S.O., 1998. Demand side management policies for residential water use: who bears the conservation burden? *Land Economics* 74 (3), 343–359.
- Robert, M.C., Richard, G.S., 1981a. A water supply cost model incorporating cost variables. *Land Economics* 57, 18–32.
- Robert, M.C., Richard, G.S., 1981b. A regional water supply cost model. *Growth and Change* 12, 9–16.
- Rogers, P., Silva, R.D., Bhatia, Q.S., 2002. Water is an economic good: how to use prices to promote equity, efficiency, and sustainability. *Water Policy* 4, 1–17.
- Schiffler, M., 1998. *The Economics of Groundwater Management in Arid Countries: Theory, International Experience and a Case Study of Jordan*. Routledge, London.
- Shin, J., 1985. Perception of price when information is costly: evidence from residential electricity demand. *Review of Economics and Statistics* 67 (4), 591–598.
- Tinbergen, J., 1952. *On the Theory of Economic Policy*. North-Holland, Amsterdam.
- Tinbergen, J., 1955. *On the Theory of Economic Policy*. North-Holland, Amsterdam.
- Turvey, R., 1976. Analyzing the marginal cost of water supply. *Land Economics* 52, 158–168.
- van Soest, D., List, J.A., Jeppesen, T., 2006. Shadow prices, environmental stringency, and international competitiveness. *European Economic Review* 50 (5), 1151–1167.
- Wang, D.X., Wang, H., Yin, M.W., 1999. Implication of water resources and its value. *Advances in Water Science* 10 (2), 195–200.
- Wang, Y., Xu, T., 2003. Sources price of water resources in Tianjin city with fuzzy mathematic method. *Journal of Territory and Natural Resources Study* 3, 63–64.
- Wang, Y.S., Yang, Y.Y., 2005. South-north water transfer projects in China. *Journal of Yangtze River* 7, 2–5.
- Warford, J.J., 1997. Marginal opportunity cost pricing for municipal water supply. Special Papers (<http://203.116.43.77/publications/specialp2/ACF298.html>).
- Zhang, F., 2000. Marginal opportunity cost pricing for wastewater disposal: a case study of Wuxi, China. Research Reports (<http://203.116.43.77/publications/research1/ACF1C9.html>).

A systematic approach for the comparative assessment of stormwater pollutant removal potentials

Lian Scholes, D. Michael Revitt*, J. Bryan Ellis

Urban Pollution Research Centre, School of Health and Social Sciences, Middlesex University, Queensway, Enfield EN3 4SA, UK

Received 30 May 2006; received in revised form 4 January 2007; accepted 11 March 2007

Available online 25 April 2007

Abstract

This paper describes the development of a methodology to theoretically assess the stormwater pollutant removal performances of structural best management practices (BMPs). The method combines the categorisation of the relative importance of the primary removal processes within 15 different BMPs with an evaluation of the ability of each process to remove a pollutant in order to generate a value representing the pollutant removal potential for each BMP. The methodology is demonstrated by applying it separately to a set of general water quality indicators (total suspended solids, biochemical and chemical oxygen demand, nitrates, phosphates and faecal coliforms) to produce a ranked list of BMP pollutant removal efficiencies. Given the limited amount of available monitoring data relating to the differential pollutant removal capabilities of BMPs, the resulting prioritisation will support stakeholders in making urban drainage decisions from the perspective of pollutant removal. It can also provide inputs to existing urban hydrology models, which aim to predict the treatment performances of BMPs. The level of resilience of the proposed approach is tested using a sensitivity analysis and the limitations in terms of BMP design and application are discussed.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Best management practices (BMPs); Pollutant removal potentials; Unit operating processes; Stormwater

1. Introduction

The EU Water Framework Directive (WFD) (EU, 2000) identifies the control of diffuse pollution as a key factor in enabling good ecological status to be achieved in aquatic systems and therefore represents a clear driver for the review of current stormwater management practices. The conventional drainage approach to managing stormwater flows involves the direct removal of surface water through a series of pipes to the nearest watercourse to prevent local flooding, with little attention being paid to stormwater quality or its impact on receiving waters. However, the issue of urban water quality is increasingly taking centre stage and associated with this is a greater interest in the use of stormwater best management practices (BMPs). These represent a diverse range of source control procedures, which integrate stormwater quality and quantity control as well as enabling social and amenity perspectives to be

incorporated into stormwater management approaches. The term ‘BMPs’ covers a wide range of structural systems, all of which have the ability to improve stormwater quality and attenuate flow volumes through a combination of biological, physical and chemical processes.

1.1. Selection of BMPs

There are a variety of approaches and guidelines available for selecting the most appropriate type of BMP for a particular site with, for example, many states/counties in the USA having their own stormwater design manuals (e.g. Maryland Department of the Environment, 2000). Typically, these make recommendations in relation to catchment-specific factors such as soil type, available space, capacity to store a design storm event, operation and maintenance requirements and cost (CIRIA, 2000, 2001). In contrast, the potential for specific types of BMPs to remove a particular pollutant of concern, or even the treatment efficiency of BMPs in general, is rarely, if ever, used as a discriminatory criterion. However, as a result of

*Corresponding author. Tel.: +44 208 411 5308; fax: +44 208 411 5440.
E-mail address: m.revitt@mdx.ac.uk (D.M. Revitt).

the implementation of the EU WFD, and the increasing importance of pollution reduction accountability in the context of River Basin Management Plans, knowledge relating to the ability of different BMP systems to remove a particular pollutant is becoming a prime requirement.

A range of urban hydrology models are available which incorporate assessment of the performance of BMPs as part of the drainage network, although these simulations are primarily based on an evaluation of their hydraulic behaviour. An appropriate example is InfoWorks CS, which models the hydraulic behaviour of BMPs such as soakaways, infiltration trenches, swales and permeable pavements. Additionally, it contains a water quality module which can feedback into the hydraulic simulation, but this relates primarily to the modelling of physical processes, such as sediment build-up, to support pollution control through the identification of CSO problems rather than contributing to the prediction of pollutant removal treatment efficiencies in different BMPs (Wallingford Software, 2006). The Storm Water Management Model (SWMM) integrates BMP hydrological modelling with a consideration of the associated treatment performance (EPA, 2006). However, this requires operators to input their own BMP removal efficiencies, which in the absence of field data would need to be estimated.

A model which more specifically addresses the role of BMPs by incorporating ponds, bioretention systems, infiltration buffer strips, sedimentation basins, pollutant traps, wetlands and swales into stormwater management strategies is the decision-support model MUSIC, Model for Urban Stormwater Improvement Conceptualisation (CRC, 2006). MUSIC models BMP performances using algorithms originally developed for Continuously Stirred Tank Reactors (CSTRs) with different numbers of CSTRs being used to mimic different types of BMPs. Pollutant removal is evaluated using the k – C^* model modification of the first-order kinetic uptake model developed by Kadlec and Knight (1996) to predict the removal of BOD in wetlands. The selection of appropriate values for the first-order rate constant (k) and the background pollutant concentration (C^*) are therefore of critical importance in robustly predicting pollutant removal. However, because the removal processes which occur in BMPs are highly heterogeneous in terms of both space and time (for example, physical processes may dominate during storm events with biological and chemical processes being of greater importance in the longer term), choosing a single ' k ' value which covers all these variables is extremely complex. The selection of C^* is also problematic as it may vary greatly in relation to factors such as variations in the time-period between storm events and rainfall intensity. A range of default ' k ' and C^* values, based on considerations of the relevant physical processes, are provided within the model for total suspended solids (TSS), total phosphorous and total nitrogen for seven different BMPs. Users are recommended to select from a range of values (e.g. ' k ' values from 500 to 5000 m year^{-1} for TSS removal within a

wetland) based on factors relating to the specific BMP characteristics and a sensitivity analysis to determine the impact of varying ' k ' and ' C^* ' values on overall treatment performance. Such a pragmatic approach is necessitated by the current lack of comparable data on pollutant removals across the many different types of BMPs and indicates the need for a fuller investigation into the relative contributions of the pertinent biological, chemical and physical processes.

As a contribution to meeting these identified needs, this paper describes the development of a systematic approach, based on fundamental scientific principles, for the prediction of the comparative pollutant removal potentials operating within a range of BMPs. The results can be utilised to inform the selection of k values which integrate biological, chemical and physical processes within modelling routines, such as MUSIC, as well as contributing to the wider decision-making framework which typically involves the consideration of a diversity of factors ranging from catchment size to rainfall-runoff coefficients and costings. Hence, the approach described in this paper is not meant to be used as a stand-alone procedure, but one which can contribute to stormwater management decision-making processes with respect to the control of the discharge of specific substances to receiving water bodies until further field data becomes available. Full details on the development of this novel approach are presented, together with its application to TSS, BOD, COD, nitrates, phosphates and faecal coliforms. Results of this procedure are critically discussed and compared, where possible, to observed field data and the validity tested through a sensitivity analysis.

2. Unit removal processes in BMPs

2.1. Identification and controlling factors

The performances of individual stormwater BMPs may vary from site to site in relation to variables such as design specifications, local hydrologic and climatic conditions, and system age (e.g. Ellis et al., 2003). In developing the described methodology, the BMP-type descriptions are as outlined in Table 1. In addition, the individual BMP devices are assumed to be operating at their design efficiency and to be functioning on a 'stand-alone' basis, i.e. not part of a hybrid treatment train. The vulnerability of the treatment potential to variable hydraulic conditions has been taken into account by incorporating the impact of extreme events into the predicted efficiencies of the identified removal processes within different BMPs. The possible impacts of variables, such as system ageing, are partially incorporated through the inclusion of how the different removal processes are likely to be maintained throughout the lifetime of a BMP.

The primary biological, chemical and physical processes associated with pollutant removal in structural BMPs are identified in Fig. 1 and reflect the fundamental unit

Table 1
Descriptions of different structural BMPs

System type	Description
Filter drains	Gravelled trench systems where stormwater can drain through the gravel to be collected in a pipe; unplanted but host to algal growth
Porous asphalt	Open graded powdered/crushed stone with binder; high void ratio; no geotextile liner present
Porous paving	Continuous surface with high void content, porous blocks or solid blocks with adjoining infiltration spaces; an associated reservoir structure provides storage; no geotextile liner present; host to algal growth
Sedimentation tank	Symmetrical concrete structure containing appropriate depth of water to assist the settling of suspended solids under quiescent conditions
Filter strip	Grassed or vegetated strip of ground that stormwater flows across
Swales	Vegetated broad shallow channels for transporting stormwater
Soakaways	Underground chamber or rock-filled volume; stormwater soaks into the ground via the base and sides; unplanted but host to algal growth
Infiltration trench	A long thin soakaway; unplanted but host to algal growth
Infiltration basin	Detains stormwater above ground which then soaks away into the ground through a vegetated or rock base
Retention ponds	Contain some water at all times and retains incoming stormwater; frequently with vegetated margins
Detention basins	Dry most of the time and able to store rainwater during wet conditions; often possess a grassed surface
Extended detention basin	Dry most of the time and able to store rainwater during wet conditions for up to 24 h; grassed surface and may have a low basal marsh
Lagoons	Pond designed for the settlement of suspended solids; fringing vegetation can sometimes occur
Constructed wetlands	Vegetated system with extended retention time
• Sub-surface flow	Typically contain a gravel substrate, planted with reeds, through which the water flows
• Surface flow	Typically contain a soil substrate, planted with reeds, over which the water flows

operating processes (UOPs) familiar to traditional water and wastewater engineers. This UOP approach provides an alternative, and more radical, methodological basis for the selection of BMPs, but one which is being more widely considered within stormwater engineering (Quigley et al., 2005; Scholes et al., 2005). Fig. 1 shows how these fundamental UOP behavioural properties can be integrated with pollutant-specific characteristics to develop a combined value, which represents the potential for a specific pollutant to be removed within a particular BMP system. Repetition of this procedure for each BMP enables a combined value to be developed which represents the

relative potential for each BMP to remove the pollutant under consideration. These combined values can then be ranked in a descending order to generate a hierarchy of BMPs with regard to removal of the specific pollutant of concern, which can also be used to support the relative selection of 'k' values for use in stormwater management modelling.

The primary pollutant removal mechanisms found in BMPs can be divided into two categories depending on whether they result (i) in the direct removal of a pollutant from the water column (e.g. settling; adsorption to substrate; microbial degradation; filtration; plant uptake; volatilisation and photolysis) or (ii) whether they contribute indirectly to the removal of a pollutant (e.g. precipitation; adsorption to suspended solids). The indirect processes contribute to the overall removal of pollutants by occurring before processes such as settling or filtration. Adsorption to suspended solids will be controlled by factors such as the particulate surface area and surface composition. The latter influences the extent of physical adhesion and/or chemi-sorption with the presence of coatings of organic matter particularly affecting the adsorption of organic pollutants. In contrast, precipitation is mainly controlled by variations in the temperature and/or the chemical composition of the water although both processes are assisted by the existence of still as opposed to turbulent conditions.

2.2. Relative importance of the different BMP removal processes

Currently, there is only limited field data available to describe the differential pollutant removal capabilities of BMPs and in the absence of experimental results, it is appropriate to consider a theoretical approach as a pragmatic contribution to managing the uncertainties associated with this acknowledged data gap. This can be achieved by utilising existing knowledge at the UOP level to predict the primary removal mechanisms which occur in each type of treatment system and making informed judgements about the comparative importance of each removal process on both intra- and inter-BMP system bases. The adoption of such a prioritisation approach is commonly applied as part of risk management procedures and has been initiated in this study by designating the relative importance of each removal mechanism within specific BMPs as being of either:

- high importance, i.e. considered to be a dominant removal process within the BMP;
- medium importance, i.e. a process which contributes significantly to the overall BMP pollutant removal capability;
- low importance, i.e. a process which makes only a small contribution to pollutant removal or
- not applicable (NA) where it is not relevant to a particular BMP option.

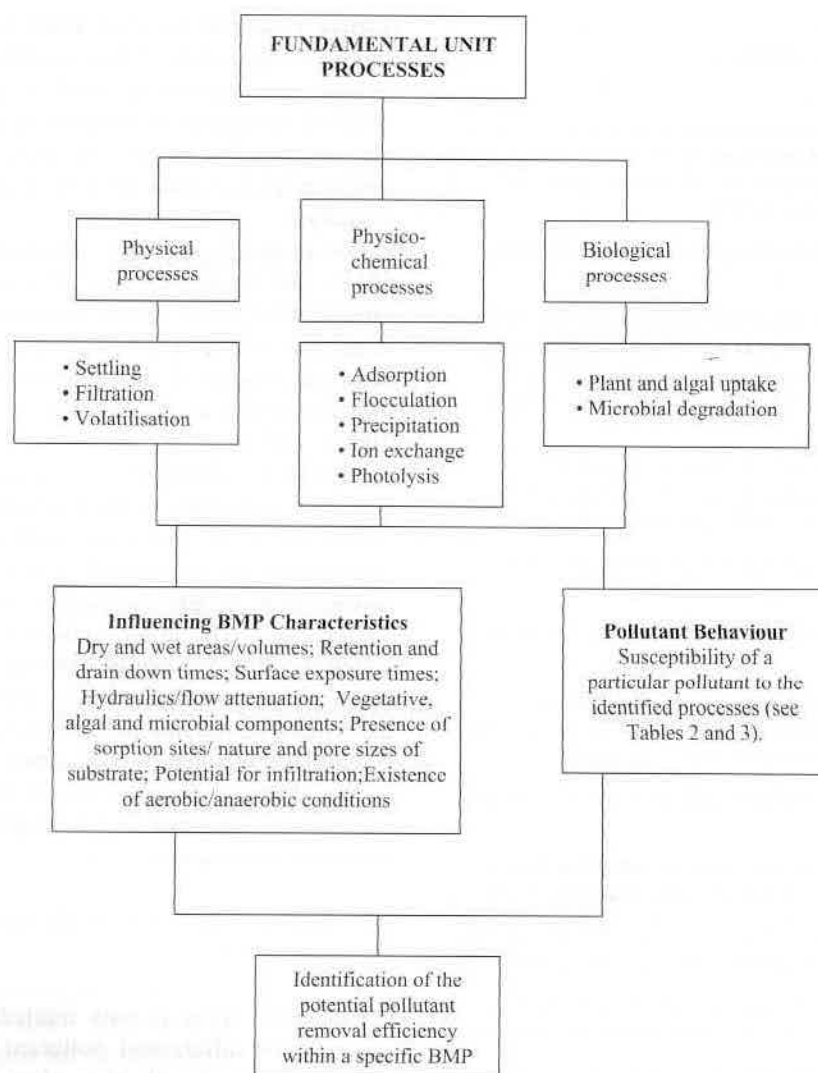


Fig. 1. Fundamental unit processes in relation to BMP characteristics and pollutant behaviour.

Where appropriate, these functional categories are further sub-divided into medium/high and low/medium but it is not considered justifiable to introduce additional sub-classifications. Developing such a systematic approach to support decision-making in the absence of existing data requires the use of expert judgement as a basis for assigning rankings and is therefore a subjective process although experience gained in the wastewater-engineering field provides a sound basis for UOP characterisation. There are also many areas of professional practice such as occupational risk management, where a combined data/professional judgement approach to decision-making is necessary due to operational demands. Embedded within the use of this approach is the understanding that, where expert judgement forms a significant part of the ranking procedure, it is recommended that wider participation should be considered (CERM, Risk Ranking, 1997). The rankings identified and assigned to the various removal processes presented in this paper reflect both the opinions of the authors and collaborative partners and end-users

associated with the completed DayWater project (DayWater, 2002). The premises on which the rankings have been assigned are clearly stated to facilitate further discussion and consultation and to assist the development of a more refined and explicit framework, where this is required.

2.3. Descriptions of the identified unit operating processes

2.3.1. Adsorption to substrate

Adsorption to substrate refers to the physico-chemical adherence of pollutants to an artificial substrate (e.g. the gravel matrix of a filter drain), a natural substrate (e.g. vegetation within a swale) or an introduced substrate (e.g. the deposited benthic sediment within a detention pond) and is influenced by those factors previously described for adsorption to suspended solids. It is an important potential removal process in filter drains, porous paving (with underground reservoir) (Legret and Colandini, 1999), sub-surface flow (SSF) constructed wetlands, infiltration

basins, soakaways and infiltration trenches due to the close contact achieved between stormwater and substrate surface during the infiltration of an effluent through the relevant permeable material (Table 2). The hydraulic pathways taken by stormwater within swales, filter strips, surface flow (SF) constructed wetlands, detention basins and extended detention basins will result in lower direct contact times with the available substrate and therefore less potential for adsorption. The regular draining down of detention basins following a storm event will encourage adsorption in comparison to retention ponds and lagoons where a permanent water body exists.

2.3.2. Settling

Settling is the vertical movement of discrete or agglomerating suspended sediment particles to the base of a water column (Ellis et al., 2004) and is highly dependent on the retention of a quiescent water volume within the BMP system. It basically represents the Type I and Type II settling processes which occur in conventional wastewater primary settlement tanks (Tchobanoglous et al., 2003). Hence, it will be a predominant mechanism in retention ponds (Pettersson et al., 1999), infiltration basins (Barbosa and Hvitved-Jacobsen, 2001) and extended detention basins (Revitt, 2004), but slightly less important in detention ponds (reduced retention time) and lagoons and sedimentation tanks (comparatively lower volumes and surface areas) (Table 2). Although, the presence of macrophytes in both types of constructed wetland contributes to the formation of quiescent conditions, the

presence of dense stands of vegetation also effectively lowers the stationary water column volume, through which settling can occur, leading to the allocation of a comparatively lower potential for settling in these systems. The other types of BMP are assigned 'low/medium' or 'low' removal potentials for settling processes due to the absence of a persistently still water body, which would facilitate particle deposition.

2.3.3. Microbial degradation

Microbial degradation is facilitated by the availability of attachment sites and nutrients within a BMP and both aerobic and anaerobic processes are enhanced by the occurrence of high contact ratios between stormwater and substrate material. Microbial degradation is therefore strongly encouraged within SSF constructed wetlands (Ellis et al., 2003) and infiltration basins (Table 2). Filter drains, porous paving, soakaways, infiltration trenches, retention ponds, extended detention basins and SF constructed wetlands do not typically provide the same diversity of microbial attachment sites and are consequently assigned a medium significance for this process. The prolonged contact of stormwater with an established microbial population is less feasible for detention basins (non-permanent water body), filter strips and swales (low retention times) resulting in the allocation of 'low/medium' removal potentials for these BMPs. The remaining treatment systems (porous asphalt, sedimentation tanks and lagoons) are assigned the lowest importance for this removal process due to the comparatively lower potentials for stormwater to interact with substrates acting as hosts to diverse microbial communities.

2.3.4. Filtration

This UOP in BMPs occurs by the same mechanisms as those present in conventional water treatment plant sand filtration units where physical sieving removes particulate pollutants as they pass through a porous substrate or hydraulic barrier (Ellis et al., 2004). Hence, the potential for filtration to occur is considered to be most effective in porous paving and porous asphalt due to surface filtration (Revitt, 2004), particularly in porous asphalt due to the low pore size of the crushed construction material (Table 3). Infiltration trenches, infiltration basins, soakaways and SSF constructed wetlands involve the passage of stormwater through a sub-surface substrate but filtration is less efficient due to the greater void sizes within gravel, which is typically used as the substrate. Filter drains possess similar substrate structures but do not provide the opportunity for further infiltration to ground. The allocated medium removal potential also applies to SF constructed wetlands and swales, where the filtering function of surface vegetation combines with possible soil infiltration to remove pollutants (Bäckström, 2003). Similar processes are possible in filter strips but shorter contact times between stormwater and the grassed surface result in a lower filtration potential. The remaining BMPs are considered to

Table 2
Relative importance of substrate adsorption, settling and microbial degradation in BMPs

BMPs	Adsorption to substrate	Settling	Microbial degradation
Filter drain	Medium/high	Low/medium	Medium
Porous asphalt	Low/medium	Low	Low
Porous paving	High	Low/medium	Medium
Filter strip	Medium	Low	Low/medium
Swales	Medium	Low/medium	Low/medium
Soakaways	Medium/high	Low/medium	Medium
Infiltration trench	Medium/high	Low/medium	Medium
Infiltration basin	High	High	High
Sedimentation tank	Low	Medium/high	Low
Retention ponds	Low/medium	High	Medium
Detention basins	Medium	Medium/high	Low/medium
Extended detention basin	Medium	High	Medium
Lagoons	Low/medium	Medium/high	Low
Constructed wetlands (SSF)	Medium/high	Medium	High
Constructed wetlands (SF)	Medium	Medium	Medium

Table 3

Relative importance of filtration, volatilisation, photolysis and plant uptake in BMPs

BMPs	Filtration	Plant uptake	Volatilisation	Photolysis
Filter drain	Medium	Low	Low	NA
Porous asphalt	High	NA	Low	Low
Porous paving	High	Low	Low	NA
Filter strip	Low/ medium	Medium	Low/medium	Low/ medium
Swales	Medium	Medium	Medium	Low/ medium
Soakaways	Medium/ high	Low	Low	NA
Infiltration trench	Medium/ high	Low	Low	NA
Infiltration basin	Medium/ high	Low/ medium	Medium	Low/ medium
Sedimentation tank	NA	NA	Low	Low
Retention ponds	Low	Low	Medium	Low/ medium
Detention basins	Low	Low	Medium	Low/ medium
Extended detention basin	Low	Low	Medium	Low/ medium
Lagoons	Low	Low	Low/medium	Low
Constructed wetlands (SSF)	Medium/ high	Medium/ high	Low/medium	Low
Constructed wetlands (SF)	Medium	Medium	Medium	Low

have low filtration potentials because of limited contact between stormwater and the basal sediments/substrates.

2.3.5. Plant uptake

The presence of terrestrial or aquatic vegetation provides the potential for plant uptake to occur and therefore this is not an applicable process in non-vegetated BMPs such as porous asphalt and sedimentation tanks. Retention ponds with fringing macrophytes present are allocated a 'low' classification for plant uptake because of the limited contact between aquatic vegetation and the bulk of the pollutants (Table 3). Similarly in detention basins and extended detention basins, surface grass coverage can participate in pollutant removal when stormwater is present. Pollutant bioaccumulation by cell tissue at a low level is also possible in porous paving, filter drains, soakaways and infiltration trenches due to algal growth on the sub-surface gravel or other filler material. A slightly increased plant uptake is envisaged for infiltration basins due to a combination of a naturally grassed surface with an algal-coated gravel substrate. Swales and filter strips are permanently grassed structures, which are allocated medium removal potentials because of efficient contact ratios between stormwater and vegetation. Similar contact efficiencies between stormwater and vegetation are expected in SF constructed wetlands. However, the potential for plant uptake will be highest in SSF constructed

wetlands due to the increased contact between stormwater and the elaborate root systems of aquatic macrophytes.

2.3.6. Volatilisation and photolysis

Both processes are strongly dependent on surface exposure but whereas photolysis requires direct exposure to sunlight, volatilisation can occur from the open spaces within a BMP structure. Photolytic degradation will be negligible in filter drains, porous paving, soakaways and infiltration trenches due to the rapid incorporation of stormwater into the BMP structure. A low effectiveness is predicted in sedimentation tanks and lagoons (normally low surface areas and associated retention times), in both types of constructed wetland (restricted exposure of stormwater to sunlight due to the presence of dense stands of vegetation) and in porous asphalt (rapid infiltration into the surfacing material where interaction with UV/visible radiation is still possible). The level of photolysis will be highest (low/medium potential) in filter strips, swales, infiltration basins, retention ponds, detention basins and extended detention basins due to a combination of enhanced surface areas and exposure times (Table 3).

The degree of volatilisation is highest (medium removal potential) in extended detention basins, detention basins, retention ponds, infiltration basins, SF constructed wetlands and swales, where stormwater exposure times and surface area exposure to wind/ambient pressure differentials are optimised, relative to SSF constructed wetlands, lagoons and filter strips (low/medium potentials). Sedimentation tanks, infiltration trenches, soakaways, porous paving, porous asphalt and filter drains are all allocated 'low' removal potentials due to the lower surface exposures associated with these systems (Table 3).

2.4. Susceptibilities of TSS, BOD, COD, nitrates, phosphates and faecal coliforms to the identified primary BMP removal processes

The abilities of the general water quality indicators (TSS, BOD, COD, nitrates, phosphates and faecal coliforms) to be removed by each of the identified mechanisms are described in Table 4. These classifications for the pollutant removal potentials have been derived by consideration of the relevant physico-chemical data, when available, as well as the existing scientific knowledge of the environmental behaviours of the pollutants. Thus, the nutrients (nitrates and phosphates) are both assigned 'high' removal potentials for plant uptake but they differ markedly for adsorption, settling and filtration because of their vastly different water solubilities. Both nutrients are resistant to aerobic biodegradation but can be assigned 'low' and 'low/medium' potentials for removal under anaerobic conditions. The classifications quoted in Table 4 represent the combined impacts of both biodegradation processes. BOD and COD demonstrate similar susceptibilities to the range of removal processes with the slightly higher composition of intractable organics in COD resulting in marginally

Table 4

Potential for direct BMP processes to remove TSS, BOD, COD, nitrates, phosphates and faecal coliforms

	TSS	BOD	COD	Nitrates	Phosphates	Faecal coliforms
Adsorption	Medium	Medium	Low/medium	Low	High	Medium
Settling	High	Medium	Medium	Low	High	High
Microbial degradation ^a	Low	Medium	Low/medium	Low	Low	Low/medium
Filtration	High	Medium	Medium	Low	High	High
Volatilisation	NA	Low	Low	NA	NA	NA
Photolysis	NA	Low	Low	NA	NA	Low/medium
Plant uptake	NA	Medium	Low/medium	High	High	NA

NA = not applicable.

^aBased on a consideration of aerobic and anaerobic processes.

reduced tendencies to undergo adsorption, plant uptake and microbial degradation. In the latter process, anaerobic degradation is considered to represent a low potential for removal for both oxygen demanding parameters but the corresponding aerobic process can produce 'high' and 'medium' removals respectively for BOD and COD resulting in the combined classifications shown in Table 4. Faecal coliforms and TSS are expected to show similar removal capabilities with the only known exception being the susceptibility of the former to photolysis (van der Steen et al., 2000). The potentials of the selected pollutants to undergo both settling and filtration have been based on their abilities to adsorb to suspended solids and to precipitate, which have been identified as relevant indirect processes. Although they are dependent on similar precursor mechanisms, filtration and settling are clearly separate removal processes within a structural BMP.

3. Development and testing of an approach for the prediction of removal potentials

3.1. Aggregation of data on the relative importance of removal processes within BMPs with the potential for the selected parameters to be removed by the same processes

The potential for a pollutant to be removed within a BMP can be considered to be a function of the type and magnitude of the removal processes which occur within a BMP in combination with the susceptibility of the pollutant to be removed by these removal processes. The developed methodology enables these two data sets to be combined by adopting the risk-rating approach of converting the classifications of high, medium, low and 'not applicable' to the numeric values 3, 2, 1 and 0, respectively (Boyle, 2000). Intermediate values of 1.5 and 2.5 were allocated to low/medium and medium/high classifications as appropriate. The potential of a particular pollutant (e.g. TSS) to removal by settling can be represented by the combination of the derived numeric value representing this process with the corresponding value representing the importance of settling within a particular BMP (e.g. infiltration trench) (Tables 2, 4 and 5). A multiplicative approach has been used to produce the combined values to

Table 5

Potential for removal of TSS by an infiltration trench

	Significance of process to the pollutant	Significance of process to BMP	Combined value
Adsorption	2	2.5	5
Settling	3	1.5	4.5
Microbial degradation	1	2	2
Filtration	3	2.5	7.5
Volatilisation ^a	0	1	0
Photolysis ^a	0	0	0
Plant uptake	0	1	0
		Overall value	19

^aAssigned a weighting of 0.5.

highlight the 'extremes' (i.e. the best and worst values) and provide greater discriminatory power than would be achieved by addition. An additional factor in the calculation is that photolysis and volatilisation are both assigned weightings of 0.5 relative to the other removal processes to signify their typically lower contributions to the overall pollutant removal capability of BMPs. The separate values calculated for each removal process can then be summed to give a single overall value representing the removal potential of TSS in an infiltration trench, as displayed in Table 5.

Repeating this procedure for each BMP and then ranking the overall values in descending order of magnitude effectively enables an order of preference for the relative potential of BMPs to remove TSS to be generated. In addition to TSS, this procedure was also applied to BOD, COD, nitrates, phosphates and faecal coliforms, and the ranked results are presented in Figs. 2 and 3 where a ranking of 1 identifies the BMP possessing the highest removal potential for the identified pollutant. The values upon which the rankings are based are ordinal and not numeric and therefore represent the order of predicted BMP performances relative to each other but do not have any quantitative meaning in terms of actual removal performance.

Figs. 2 and 3 show that settlement tanks are consistently predicted to be the worst performing BMP system followed

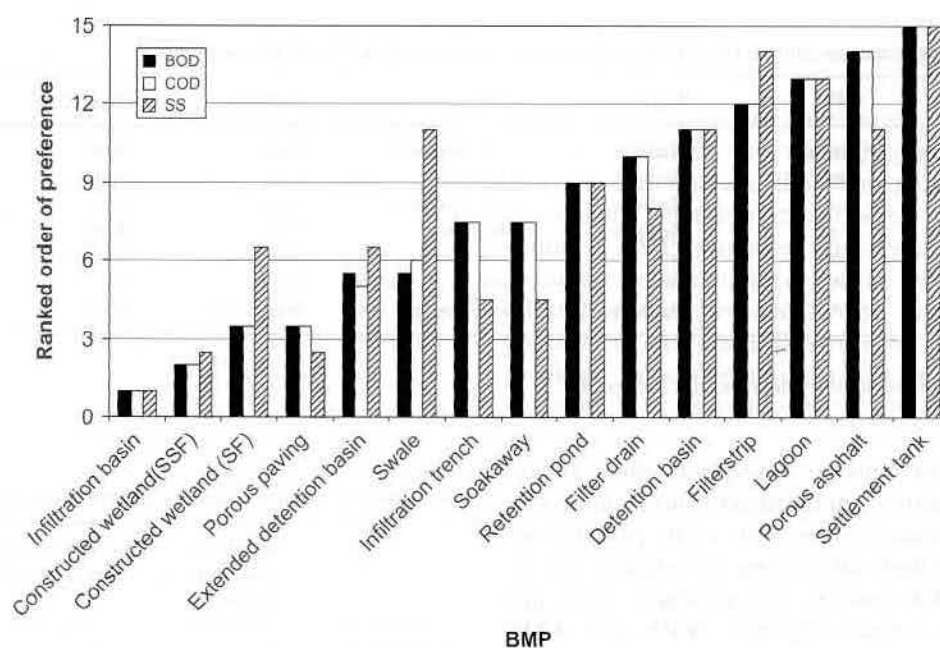


Fig. 2. Predicted order of preference for the use of BMPs to remove BOD, COD and TSS.

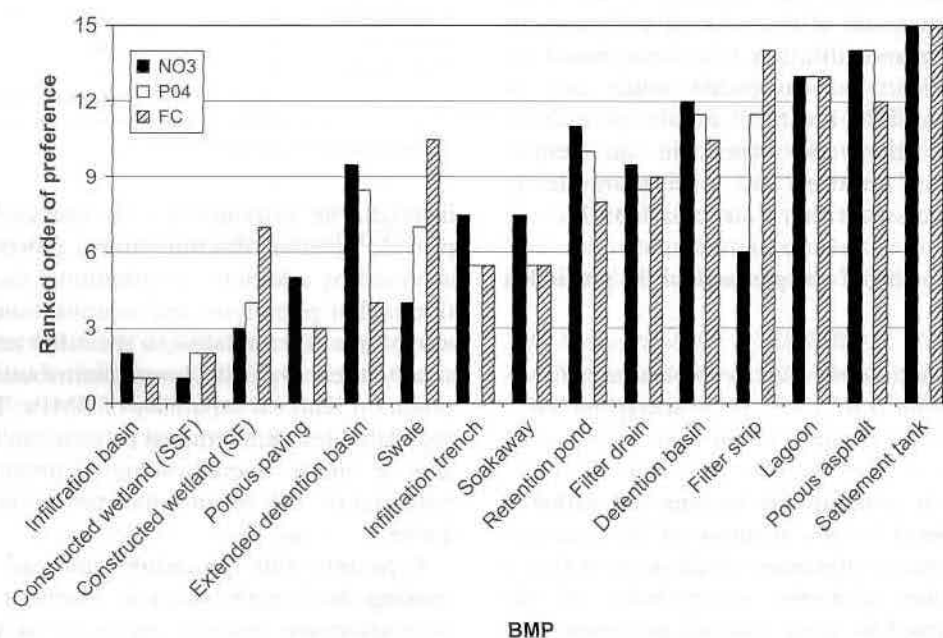


Fig. 3. Predicted order of preference for the use of BMPs to remove nitrates, phosphates and faecal coliforms.

by porous asphalt, although TSS and faecal coliforms have slightly enhanced removals for the latter. In contrast, infiltration basins offer the highest predicted removal efficiencies for five of the six pollutants evaluated. The exception is nitrates where SSF constructed wetlands achieve the highest ranking. This type of constructed wetland is more efficient than the SF type for the range of considered pollutants. Porous paving is predicted to behave equally as well, if not better than SF constructed wetlands, except for nitrates, as the removal of this pollutant is

strongly influenced by the importance of plant uptake processes (Table 3). Extended detention basins generally demonstrate the best performances for BMPs relying predominantly on 'stormwater volume capture' as the major pollutant removal process, followed by retention ponds, detention basins and lagoons. Swales exhibit a large variability in the predicted hierarchy for their removal of the different pollutants with TSS and faecal coliforms being the least efficiently retained. Filter strips show a similar variability but with overall lower efficiencies. For

BMPs utilising infiltration as the major pollutant removal process, infiltration trenches and soakaways show identical mid-range removal characteristics and are predicted to consistently out-perform filter drains.

3.2. Comparison of predicted pollutant removal efficiencies with monitored values

The predicted ranked order of BMPs with respect to their pollutant removal potentials can only be confidently compared with existing monitoring data for TSS as this is the only pollutant for which 5 independent data sets exist for a realistic number of BMPs throughout Europe and N. America (ASCE/USEPA, 2006; Ellis et al., 2003; Scholes et al., 2005; Schueler, 1997; USEPA, 2006; Winer, 2000; Yu et al., 1993). The mean values are plotted in Fig. 4, together with the standard deviations which demonstrate the variabilities in this field data resulting from the use of different monitoring strategies and associated evaluation methods and the dependence, in some cases, of the BMP performance on the magnitude of influent concentrations and loadings. There is a tendency for monitored pollutant capture rates to decrease and be more variable as influent concentrations fall. In this initial stage of the model development, the theoretical considerations are based on all BMPs operating at their design potential. However, the monitored systems will be at different stages of their operational lives and hence subject to factors such as ageing, which are not yet fully incorporated in the developed approach. However, in spite of these differences, there is a good agreement between the systematically predicted hierarchy for suspended solids removal efficiencies (Fig. 2) and those determined for the monitored data (Fig. 4) with, for example, 4 of the top 5 BMPs predicted

using the theoretical approach also appearing in the top 5 BMPs identified using field data. The major discrepancy appears to be for swales, which perform considerably better in practice than is predicted theoretically. It may be that the theoretical interpretations of swale performance have under-estimated the potential for particulate removal by filtering (by the grass sward) and settling (due to vegetative interactions reducing the rate of flow). This has been tested by increasing the relative importances of settling and filtration in swales from low/medium to medium and from medium to medium/high classifications, respectively. The impact is an expected improved TSS removal behaviour in swales, which is now similar to that predicted for infiltration trenches and SSF constructed wetlands.

In contrast, constructed wetlands (particularly SSF systems) and extended detention basins appear to under-perform relative to their predicted removal performances for TSS. For extended detention basins, only the minimum data requirement (5 independent values) was available and therefore it is possible that the calculated mean values are not truly representative of the optimum performance. The monitored value for SSF constructed wetlands was $81.0 \pm 11.1\%$, which is consistent with the predicted high hierarchical position. The most reliable monitored data (based on 43 independent measurements) is for retention ponds ($73.8 \pm 15.9\%$) and this suggests a TSS removal potential which is entirely consistent with the predicted performance which is indicated to be less efficient than a SSF wetland. Monitoring results, above the minimum 5 data sets, are also available for retention ponds for the other pollutants (except phosphates), and hence it is possible to observe if there are any comparable trends between the different pollutants for this particular BMP.

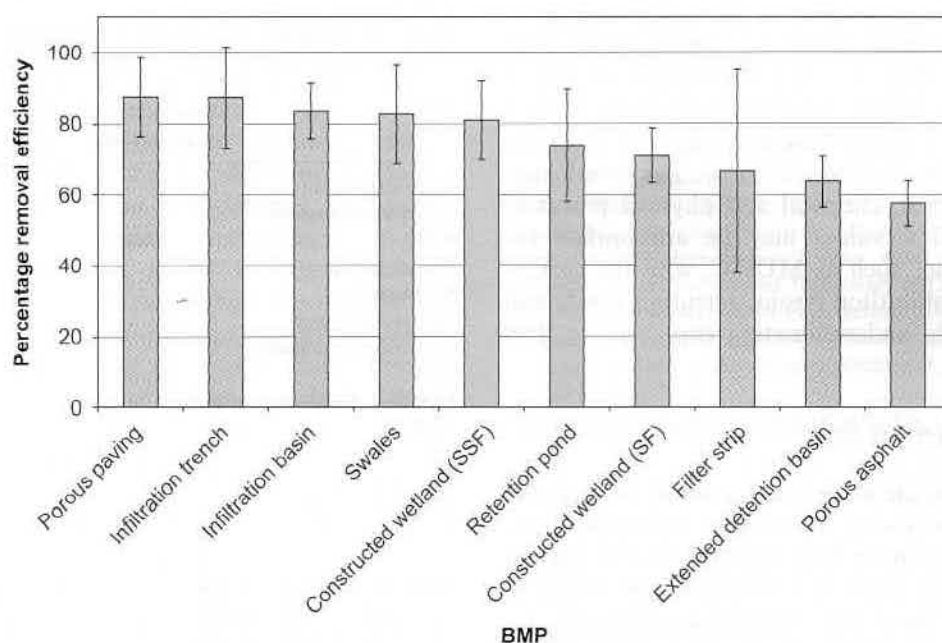


Fig. 4. Average monitored TSS percentage removal efficiencies ($\pm$ SD) for BMPs.

The measured values for BOD ($51.9 \pm 25.7\%$) and COD ($55.9 \pm 25.6\%$) are similar and consistent with the relative hierarchical positions predicted theoretically (Fig. 2). The reduced ability of this BMP to remove nitrates ($33.6 \pm 43.1\%$) compared to faecal coliforms ($71.3 \pm 14.0\%$) is also in agreement with the predicted order of preference indicated in Fig. 3 for these pollutants.

3.3. Support in the selection of k values within urban hydrological modelling

The development of a BMP order of preference in relation to a particular pollutant of concern not only provides an important contribution to the wider decision-making process but also provides relevant input to the selection of appropriate k values within stormwater management models. Currently, the MUSIC model proposes a range of k values for TSS removal within ponds, infiltration systems, wetlands, swales and sedimentation basins. A comparison of the descriptions provided within the MUSIC user manual with those in Table 1 identifies these BMPs as corresponding to retention ponds, infiltration basins, constructed wetlands (SSF for the purposes of this example), swales and settlement tanks, respectively as defined in this study. The extensive ranges of k values recommended for these BMPs are:

- infiltration basins ($200\text{--}1000 \text{ m year}^{-1}$),
- constructed wetlands ($500\text{--}5000 \text{ m year}^{-1}$),
- retention ponds ($200\text{--}1000 \text{ m year}^{-1}$),
- swales ($4000\text{--}15\,000 \text{ m year}^{-1}$) and
- settlement tanks ($4000\text{--}15\,000 \text{ m year}^{-1}$).

The BMP order of preference generated for TSS removal (Fig. 2) would suggest that the k values assigned to the above systems decrease in value in the order infiltration basin > constructed wetland (SSF) > retention pond > swale > settlement tank. There is a clear discrepancy for swales and settlement tanks, which have been assigned the highest ranges of proposed k values based on physical processes only. The differences resulting from an integrated consideration of biological, chemical and physical processes suggest that revised k values may be appropriate for modelling approaches, such as MUSIC, with the current values assigned to infiltration basins, retention ponds and constructed wetlands underestimating their role in TSS removal.

3.4. Sensitivity analysis on the predicted removal priorities

Sensitivity analysis identifies the variability in response of the results generated using a particular approach to any assumptions that may have been made about the data or methods used to derive them. It is therefore appropriate to test the sensitivity of the orders of preference generated (i.e. the approach output) to changes in the approach input (i.e. variations in the weightings placed on input parameters).

To achieve this, a simple factorial approach was applied to testing the influence of each removal process in turn on the overall pollutant removal capability of each BMP. This involved setting the weighting of a specific removal process to 0% (i.e. effectively assuming a process does not occur) and then recalculating the order of BMP preference to assess the influence of the chosen process on the overall potential of a system to remove a particular pollutant. This approach also provides an opportunity to determine how well the developed approach simulates the systems it sets out to represent by supporting the identification of the most influential removal processes within each BMP.

The results of applying this process to the removal of TSS by BMPs are compared to the default ranked order of preference (i.e. before any alteration of weightings took place) in Table 6. As can be seen from the ranges of values, demonstrated by the highest and lowest ranked positions, the responses of individual BMPs varied considerably. The relative increases and decreases in the ranked positions, reflect the differential responses of BMPs to comparable variations in weightings applied sequentially to each of the removal processes.

Seven of the BMPs are identified in Table 6 as producing overall changes in position of ≥ 5 places. These represent the BMPs, which are most vulnerable to changes in the weightings applied to the removal processes. It is interesting to note that for these seven BMPs, the greatest decreases within the order of preference for retention ponds, detention ponds and extended detention basins, in relation to their potential to remove TSS, was instigated by applying a 0% weighting to sedimentation. In contrast, the same allocation to filtration had the greatest impact on the relative TSS removal potential of porous paving,

Table 6
Impact of separately reducing the contributions of individual removal processes on the ranked position of BMPs with respect to TSS removal

BMP	Default ranked order of preference	Highest and lowest ranked positions
Infiltration basin	1	1–2
Constructed wetland (SSF)	2.5	2–3.5
Porous paving	2.5	1–6 ^a
Soakaways	4.5	4.5–9.5 ^a
Infiltration trench	4.5	4.5–9.5 ^a
Constructed wetland (SF)	6.5	6–8
Extended detention basin	6.5	2–11 ^a
Filter drain	8	6.5–10
Retention ponds	9	3.5–13 ^a
Detention basins	11	5–12 ^a
Porous asphalt	11	6.5–15 ^a
Swales	11	9–13
Lagoons	13	9.5–14
Filter strip	14	10–14
Sedimentation tank	15	12–15

^aBMPs which show an overall change in ranked position of ≥ 5 places.

soakaways, infiltration trenches and porous asphalt. The differential identification of sedimentation and filtration as being the most influential removal processes within BMPs, which function by stormwater storage and infiltration, respectively, enables confidence to be expressed in the developed approach.

4. Conclusions

This report describes the development of a methodology, based on the input of theoretical data and knowledge, to identify the relative performances of BMPs for the removal of pollutants commonly associated with stormwater. Comparison of the predicted order of preference for the use of BMPs to remove TSS with existing field data shows good agreement for most systems although for swales, there is evidence of an under-prediction of the ability to remove TSS. However, it is demonstrated that the model can easily be refined to more closely simulate the prevalent field conditions. The use of a simple factorial approach to sensitivity analysis identified sedimentation as the most influential TSS removal process in retention ponds, detention ponds and extended detention basins compared to filtration in porous paving, porous asphalt, infiltration trenches and soakaways. This provides initial confidence that the assumptions made in the development of this systematic framework approach are reasonable and credible. The predicted orders of preference for BMPs for the different pollutants can provide important inputs into existing modelling procedures where available data is deficient but the selection of appropriate k values is essential.

A major problem for stormwater managers in decision-making involving water quality is the limitation in the current amount of available and reliable monitoring data relating to the behaviours of different pollutants within BMPs. However, irrespective of this data gap, stormwater managers are required to make decisions and adopt urban drainage schemes to achieve compliance with the EU WFD. As further field data becomes available it will be possible to calibrate and refine the described systematic approach using a more robust field data set, and also to classify removal processes using quantifiable (or at least end-point) values. However, in the interim period the described methodology provides relevant information, which can support and inform discussions related to diffuse pollution control (as prioritised under the EU WFD) as well as feed into the more comprehensive considerations required within an integrated approach to urban stormwater management.

Acknowledgements

The results presented in this publication have been obtained within the framework of the EC-funded research project DayWater “Adaptive Decision Support System for Stormwater Pollution Control”, organised within the

“Energy, Environment and Sustainable Development” Programme in the Fifth Framework Programme for “Science Research and Technological Development” of the European Commission.

References

- ASCE/USEPA. 2006. International Stormwater Best Management Practices (BMP) Database, <<http://www.bmpdatabase.org>> (verified May, 2006).
- Bäckström, M., 2003. Grassed swales for stormwater pollution control during rain and snowmelt. *Water Science and Technology* 48 (9), 123–134.
- Barbosa, A.E., Hvitved-Jacobsen, T., 2001. Infiltration pond design for highway runoff treatment in semiarid climates. *Journal of Environmental Engineering* 127 (11), 1014–1022.
- Boyle, T., 2000. Health and Safety: Risk Management 2000. The Lavenham Press Ltd., UK, p. 492 (ISBN 0901357273).
- CERM, Risk Ranking, 1997. Centre for Environmental and Risk Management (CERM) for the Health and Safety Executive. HMSO, Norwich (ISBN 0 7176 1344 5).
- CIRIA, 2000. Sustainable Urban Drainage Systems: A Design Manual for England & Wales, Report C522. Construction Industry Research & Information Association, London, UK.
- CIRIA, 2001. Sustainable Urban Drainage Systems: Best Practice Manual, Report C523. Construction Industry Research & Information Association, London, UK.
- CRC, 2006. Catchment Modelling Toolkit: Overview of MUSIC, Version 3. <<http://www.toolkit.net.au/cgi-bin/WebObjects/toolkit.woa/1/wa/productDetails?productID=1000000&wsid=IUbxbpc637hpta0Z29Zwig>> (verified 25 May 2006).
- DayWater, 2002. An Adaptive Decision Support System (ADSS) for the Integration of Stormwater Source Control into Sustainable Urban Water Management Strategies, <<http://www.daywater.org>> (verified 18 May 2006).
- Ellis, J.B., Chocat, B., Fujita, S., Rauch, W., Marsalek, J., 2004. Urban Drainage: A Multilingual Glossary. IWA Publishing, London, UK, 512pp (ISBN: 190022206X).
- Ellis, J.B., Shutes R.B.E., Revitt, D.M., 2003. Constructed wetlands and links with sustainable drainage systems. Technical Report P2-159/TR1, Environment Agency, 178pp (ISBN 1857059182).
- EPA, 2006. Storm Water Management Model (SWMM), <www.epa.gov/ednrmr/models/swmm/index.htm> (verified 25 May 2006).
- EU, 2000. Directive of the European Parliament and of the Council of 23 October 2000 establishing a Framework for Community Action in the Field of Water Policy, 2000/60/EC.
- Kadlec, R.H., Knight, R.L., 1996. Treatment Wetlands. Lewis Publishers, Boca Raton, FL, USA.
- Legret, M., Colandini, V., 1999. Effects of a porous pavement with reservoir structure on runoff water: water quality and fate of heavy metals. *Water Science and Technology* 39 (2), 111–117.
- Maryland Department of the Environment, 2000. Maryland Stormwater Design Manual. <http://www.mde.state.md.us/Programs/WaterPrograms/SedimentandStormwater/stormwater_design/index.asp> (verified 25 May 2006).
- Pettersson, T., German J., Svensson, G., 1999. Pollutant removal efficiency in two stormwater ponds in Sweden. In: Joliffe, I.B., Ball, J.E. (Eds.), Proceedings of the Eighth International Conference on Urban Storm Drainage, vol. 2, 30 August–3 September, Sydney, Australia, pp. 866–873.
- Quigley, M.M., Strecker, E.W., Leisenring, M., Huber, W.C., Heaney, J., Weinstein, N., Sansalone, J., Bodine, D., 2005. The integrated unit process design approach for source control and best management practice design. In: Proceedings of the 10th International Conference on Urban Drainage, 21–26 August 2005, Copenhagen, Denmark, CD-Rom.

- Revitt, D.M., 2004. Water pollution impacts of transport. In: Hester, R.E., Harrison, R.M. (Eds.), *Transport and the Environment. Issues in Environmental Science and Technology*, vol. 20, Royal Society of Chemistry, Cambridge, UK, pp. 81–109 (ISBN: 0854042954).
- Scholes, L., Revitt, D.M., Ellis, J.B., 2005. Predicting the pollutant removal potentials of sustainable drainage systems. In: Newman, A.P., Pratt, C.J., Davies, J.W., Blakeman, J.M., *Proceedings of the Third National Conference on Sustainable Drainage*, Coventry, UK, pp. 199–210 (ISBN: 1846000076).
- Schueler, T., 1997. Influence of groundwater on performance of storm water ponds in Florida. *Watershed Protection Techniques* 2 (4), 525–528.
- Tehobanoglous, G., Burton, F.L., Stensel, H.D., 2003. *Wastewater Engineering: Treatment and Reuse*, fourth ed. McGraw-Hill, Boston, USA.
- USEPA, 2006. National Menu of Stormwater Best Management Practices. <<http://cfpub.epa.gov/npdes/stormwater/menuofbmps/menu.cfm>> (verified 25 May 2006).
- Van der Steen, P., Brenner, A., Shabtai, Y., Oron, G., 2000. Improved fecal coliform decay in integrated duckweed and algal ponds. *Water Science and Technology* 42 (10–11), 363–370.
- Wallingford Software, 2006. InfoWorks CS. <http://www.wallingford-software.com/products/infoworks_cs/> (verified 25 May 2006).
- Winer, R., 2000. National Pollutant Removal Performance Database for Stormwater Treatment Practices, second ed. Center for Watershed Protection, Ellicott City, MD, USA.
- Yu, S., Barnes, S., Gerde, V., 1993. Testing of Best Management Practices for Controlling Highway Runoff, Report No. FHWA/VA93-R16. Virginia Transportation Research Council, 60pp.

A fuzzy analytic network process for multi-criteria evaluation of contaminated site remedial countermeasures

Michael Angelo B. Promentilla*, T. Furuichi, K. Ishii, N. Tanikawa

Laboratory of Sound Material-Cycle Systems Planning, Graduate School of Engineering, Hokkaido University, Kita 13, Nishi 8, Kita-ku, Sapporo 060-8628, Hokkaido, Japan

Received 9 May 2006; received in revised form 5 December 2006; accepted 11 March 2007
Available online 30 April 2007

Abstract

The Analytic Network Process (ANP) has been proposed to incorporate interdependence and feedback effect in the prioritization of remedial countermeasures using a hierarchical network decision model, but this approach seems to be incapable of capturing the vagueness and fuzziness during value judgment elicitation. The aim of this paper is to present an evaluation method using a fuzzy ANP (FANP) approach to address this shortcoming. Triangular fuzzy numbers (TFN) and their degree of fuzziness are used in the semantic scale as human judgment expressed in natural language is most often vague and fuzzy. The method employs the alpha-cuts, interval arithmetic and optimism index to transform the fuzzy comparative judgment matrix into set of crisp matrices, and then calculates the desired priorities using the eigenvector method. A numerical example, which was drawn from a real-life case study of an uncontrolled landfill in Japan, is presented to demonstrate the process. Results from the sensitivity analysis describe how the fuzziness in judgment could affect the solution robustness of the prioritization method. The proposed FANP approach therefore could effectively deal with the uncertain judgment inherent in the decision making process and derive the meaningful priorities explicitly from a complex decision structure in the evaluation of contaminated site remedial countermeasures.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Remedial countermeasures; Hierarchical network; Fuzzy Analytic Network Process (FANP); Triangular fuzzy number; Uncontrolled landfill

1. Introduction

Contaminated sites management (CSM) is inevitably a balancing act of many diverse factors such as social, economic, political, legal, technical, and scientific issues. As often found in environmental management issues, the decision making involved in CSM is a process characterized by complexity and uncertainty, as well as, multiple and conflicting criteria. Thus, the importance of the application of multiple criteria decision analysis/aid (MCDA) methods in CSM is now widely recognized (Bardos et al., 2001; Pollard et al., 2004; Kiker et al., 2005). For example, the MCDA can make explicit the trade-offs among multiple conflicting criteria and can quantify the uncertainties

necessary for comparison of available remedial and abatement alternatives. Such process can help decision makers (DM), as well as, technical experts and stakeholders to systematically consider and apply value judgments to come up with a strategic choice of remedial options. MCDA may also provide methods for participatory decision-making where stakeholder values are elicited and explicitly incorporated into the decision process. However, even though MCDA could provide demonstrable advantages, formal applications of MCDA to CSM are still rare (Linkov et al., 2004).

The Analytic Network Process (ANP), a generalized form of the widely used Analytic Hierarchy Process (AHP), has been proposed as a suitable MCDA tool to evaluate the alternatives during the conceptual planning and design of remedial countermeasures (Promentilla et al., 2005a, 2006a). These papers introduced the hierarchical network decision structure to incorporate interdependence and

*Corresponding author. Tel.: +81 11 706 7284; fax: +81 11 706 7287.

E-mail addresses: mpromentilla@yahoo.com, mabp@kanri-er.eng.hokudai.ac.jp (M.A.B. Promentilla).

feedback effect in the prioritization of the remedial countermeasures. However, the ANP-based decision model seems to be ineffective in dealing with the inherent fuzziness or uncertainty in judgment during the pairwise comparison process. Although the use of the discrete scale of 1–9 to represent the verbal judgment in pairwise comparisons has the advantage of simplicity, it does not take into account the uncertainty associated with the mapping of one's perception or judgment to a number. In real-life decision-making situation, the DMs or stakeholders could be uncertain about their own level of preference, due to incomplete information or knowledge, complexity and uncertainty within the decision environment. They also tend to specify preferences in the form of natural language expressions which are most often vague and uncertain. Such conditions may occur when managing contaminated sites particularly during the planning stage of site remediation. Therefore, this paper proposes a fuzzy ANP approach, i.e., a fuzzy extension of ANP, to evaluate and prioritize the contaminated site remedial countermeasures.

The organization of this paper is as follows. In Section 2, the background of the methodology is discussed followed by the description of the proposed fuzzy ANP and its algorithm in Section 3. In Section 4, an illustrative numerical example is presented, which was drawn from the case study of a contaminated site caused by uncontrolled landfill in Japan. In Section 5, results from the prioritization are explained and some discussions on the sensitivity analysis follow. Finally, conclusions are presented in Section 6.

2. Background of methodology

2.1. The ANP and a prioritization method based on hierarchical network (PRMEHN)

The ANP is a generalization of the Analytic Hierarchy Process, popularly known as AHP. AHP is a theory of prioritization that derives relative scales of absolute numbers known as 'priorities' from judgments expressed numerically on an absolute fundamental scale (Saaty, 2005). The AHP/ANP framework has three basic features which are useful in multi-criteria decision-making problems:

(1) modeling the system's complexity, (2) measuring on a ratio scale and (3) synthesizing (Table 1). The local priorities in ANP are established in the same manner as they are in AHP using pairwise comparisons and judgments. However, the supermatrix approach which became popularly known as the ANP approach is becoming an attractive tool to understand more of the complex decision problem as it overcomes the limitation of the AHP's linear hierarchy structure (Saaty, 1996, 2001).

The supermatrix was introduced to serve as a unifying framework for the study of priorities in hierarchy and in 'systems with feedback' (Saaty, 1980). Consider a system that can be decomposed into m clusters $C_1, C_2, \dots, C_m$ and let the elements in C_k having n_k elements be denoted as $e_{k1}, e_{k2}, \dots, e_{kn_k}$. Then, a supermatrix representation of this system (Fig. 1a) can be viewed as a partitioned matrix that describes the interaction between the elements and clusters

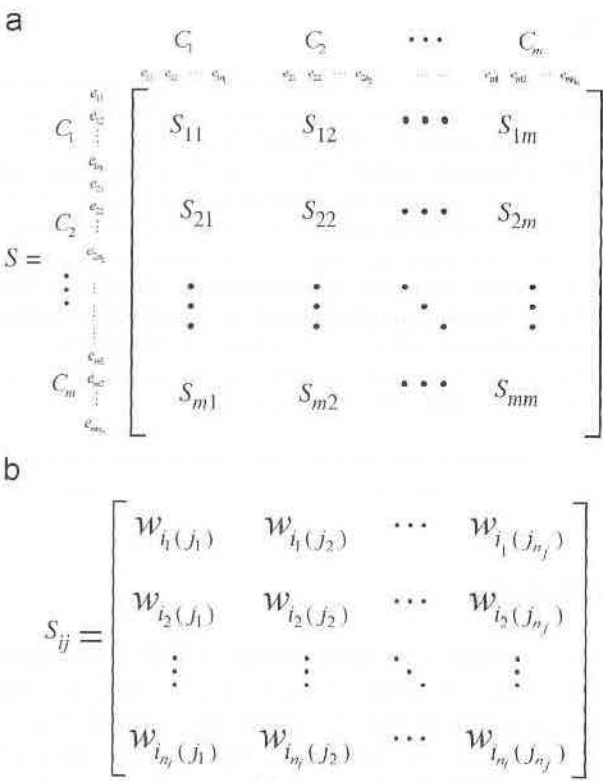


Fig. 1. A generalized supermatrix and an example of block matrix of the supermatrix.

Table 1
The main functions of AHP and ANP

Functions/features	AHP	ANP
(1) Structuring system's complexity	One-directional hierarchical structure with the assumption of mutual independence of elements.	A non-linear network structure of interdependence and feedback. Hierarchy, a linear structure, is a special case.
(2) Measuring on a ratio scale	Local priorities derived from pairwise comparison matrices	Local (or initial) priorities derived from pairwise comparison matrices
(3) Synthesizing	Global or overall priorities from hierarchic composition or additive synthesis	Overall (or final) priorities from supermatrix calculations

of a system. In this so-called supermatrix, the block matrix contains a column of priority weights which represent the impact of all elements in the i th cluster on each of the elements in the j th cluster (Fig. 1b). Some of its entries may be zero corresponding to those elements which have no path of direct interaction (or influence) to other elements.

The prioritization method based on hierarchical network (PRMEHN) model, which was used to evaluate remedial countermeasures of a contaminated site, is built from the ANP concepts (Promentilla et al., 2006b,c). The main feature of this method is its underlying problem structure represented as a strongly connected hierarchical network to understand the complexity of the contaminated site remediation problem. Fig. 2 is an example of a digraph and supermatrix representation of the four interacting clusters in a three-level hierarchical network.

The directions of the arc (or arrow) and loop signify dependence such that the arc and loop pointing into a cluster indicate that its elements influence the elements in the cluster from which the arc is emanating (Fig. 2a). This structural model incorporates the following types of dependence in a multilevel hierarchy: (1) hierarchic functional dependence as described by the downward arc from the upper level to the lower level clusters, (2) feedback

dependence as described by the upward arc from the lower-level to the upper-level cluster, (3) inner-dependence or self-feedback as indicated by loop at each cluster, (4) interdependence among clusters in a level as indicated by the (horizontal) arc between clusters belonging to the same level and (5) feedback control loop as indicated by the arc pointing to the goal cluster. The feedback control loop can be viewed as a structural dependence indicating that all the elements defined in the decision structure are relevant and influenced by the goal element, making the systems strongly connected.

The input to the supermatrix of a hierarchical network would depend on the presence and type of dependence described in the digraph as shown in Fig. 2b. The entries to the block matrices (S_{ij}) in the initial supermatrix are the estimated priorities that provide the relative strength of dominance of an element over another element in the cluster with respect to a common element from which the arc emanates. There are several algorithms to measure such priorities as described in Srdjevic (2005). The AHP eigenvector method is one of the popular methods used to quantify the relative dominance of the elements from pairwise comparison matrices. Saaty (1980) proposed the following eigenvalue formulation to obtain the desired ratio-scale priority vector (or weights) w of n elements:

$$Aw = \lambda_{\max} w, \quad e^T w = 1, \quad (1)$$

where $A = [a_{ij}]$ is the positive reciprocal pairwise comparison matrix, $\lambda_{\max}$ is the principal eigenvalue of the matrix A and $e^T = [1, 1, \dots, 1]$ is a unit row (or summation) vector. If a DM provides 'consistent' judgments (i.e., the resulting comparison matrix A satisfies the cardinal consistency property $a_{ij} \cdot a_{jk} = a_{ik}$), then $\lambda = n$; otherwise, $\lambda > n$. For details of this method and its features to measure the inconsistency of judgment, refer to Saaty (1980).

Overall (ratio-scale) priorities can be derived based on the synthesizing concept of the supermatrix. In raising the supermatrix into a large power, the transmission of influence along all possible paths defined in the decision structure is captured in the process (Saaty, 2001). The convergence of the initial priorities to a steady state or equilibrium value in the so-called limit supermatrix provides a set of meaningful synthesized priorities from the underlying decision structure. Since the PRMEHN model defines a strongly connected digraph, its supermatrix representation is a primitive irreducible matrix. Thus, the limit supermatrix L exists when the initial supermatrix is standardized by its principal eigenvalue as shown by the following equation:

$$\lim_{p \rightarrow \infty} (S/\lambda_{\max})^p = \lim_{p \rightarrow \infty} (\bar{S})^p = L. \quad (2)$$

Every column of this limit matrix is a unique positive column eigenvector associated with the principal eigenvalue $\lambda_{\max}$ (see Nikaido (1968) for the mathematical proof). This principal column eigenvector corresponds to the stable priorities from the limit supermatrix and can be

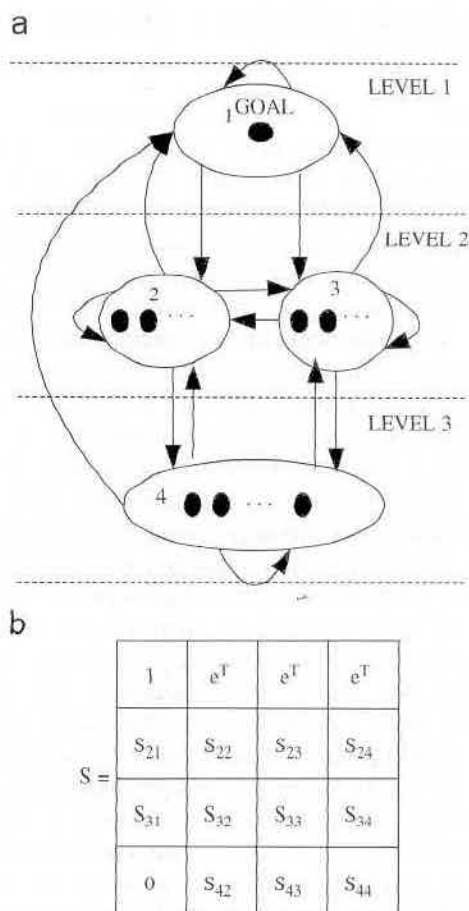


Fig. 2. An example of the digraph structure and supermatrix of a hierarchical network.

used to measure the overall relative dominance of one element over another in a hierarchical network structure (Promentilla et al., 2006b, c).

2.2. Fuzzy set theory (FST) and fuzzy numbers

In most cases, many decision problems are too complex to be understood with certainty and all the necessary information are rarely available, but people can manage to come up with a solution or decision by using knowledge that is imprecise rather than precise. Fuzzy set theory (FST) resembles human reasoning in its use of approximate information and uncertainty to generate decisions. FST is a mathematical theory introduced by Zadeh (1965) to model the uncertainty attributed to the vagueness and imprecision in real systems, particularly that of the human cognitive processes. This theory can be viewed as a generalization of the classical set theory to deal with classes of non-sharp boundaries. Thus, any methodology or theory implementing ‘crisp’ definitions such as the classical set theory, arithmetic and programming may be ‘fuzzified’ by generalizing the concept of a crisp set to the fuzzy set with blurred boundaries. The benefit of extending crisp theory or analytical methods to fuzzy techniques is in its strength in solving real-life problems, which inevitably entails some degree of uncertainty due to imprecision and fuzziness (Klir and Folger, 1988; Zimmermann, 2001).

A fuzzy number is a special fuzzy set denoted as $F = \{(x, \mu_F(x), x \in R)\}$, where x takes values on the real line, $R: -\infty < x < +\infty$ and $\mu_F(x)$ is a continuous mapping from R to the closed interval $[0, 1]$. In the literature, there are several forms of fuzzy numbers and one of the most commonly used is the triangular fuzzy number (TFN). The TFN is a special form of L–R fuzzy number (see Fig. 3) denoted as $\tilde{M} = (l, m, u)$, where $l \leq m \leq u$. It has the following triangular-type membership function:

$$\mu_{\tilde{M}}(x) = \begin{cases} 0 & x < l, \\ \frac{x-l}{m-l} & l \leq x \leq m, \\ \frac{u-x}{u-m} & m \leq x \leq u, \\ 0 & x > u, \end{cases} \quad (3)$$

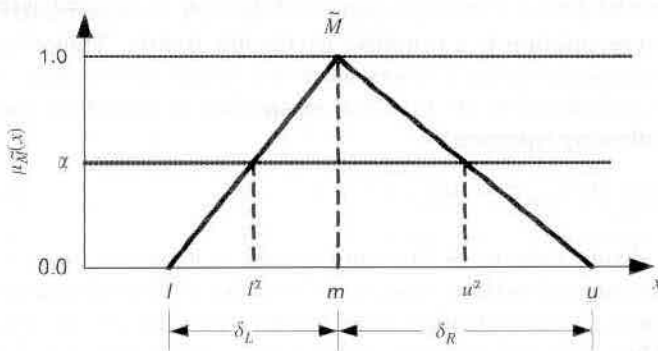


Fig. 3. A triangular fuzzy number $\tilde{M}$.

Alternatively, by defining the interval of confidence level denoted as α (also known as the alpha-cut), the TFN can be characterized as

$$\forall \alpha \in [0, 1] \tilde{M}_\alpha = [l^\alpha, u^\alpha] = [l + (m-l)\alpha, u - (u-m)\alpha]. \quad (4)$$

The main arithmetic operations for positive fuzzy numbers, as described by using the interval of confidence, have been shown by Kaufmann and Gupta (1985) to be

$$\forall m_L, m_R, n_L, n_R \in R^+, \tilde{M}_\alpha = [m_L^\alpha, m_R^\alpha], \tilde{N}_\alpha = [n_L^\alpha, n_R^\alpha], \alpha \in [0, 1].$$

Addition:

$$\tilde{M} + \tilde{N} = [(m_L^\alpha + n_L^\alpha), (m_R^\alpha + n_R^\alpha)]. \quad (5)$$

Subtraction:

$$\tilde{M} \ominus \tilde{N} = [(m_L^\alpha - n_R^\alpha), (m_R^\alpha - n_L^\alpha)]. \quad (6)$$

Multiplication:

$$\tilde{M} \otimes \tilde{N} = [(m_L^\alpha \times n_L^\alpha), (m_R^\alpha \times n_R^\alpha)]. \quad (7)$$

Inverse:

$$\tilde{M}^{-1} = [(1/m_R^\alpha), (1/m_L^\alpha)]. \quad (8)$$

2.3. Fuzzy pairwise comparative judgment matrix and fuzzy AHP (FAHP)/ANP

Suppose $E = \{e_1, e_2, \dots, e_n\}$ be a set of elements and E_c is a controlling element or a point of view that is a common aspect, characteristic or property of those elements with which a human DM is concerned. In conventional AHP/ANP, the pairwise comparison process requires that a crisp number denoted as a_{ij} be assigned to each pair of elements (e_i, e_j) representing in the perception of DM his value judgment. This judgment can be expressed verbally or by a corresponding numerical intensity to derive the ratio of the priority of the dominant element to that of the other element from pairwise comparisons. Accordingly, the DM is asked first which one of the pair of elements dominates the other with respect to a property or controlling element. If it is *just equal*, then this corresponds to the scale value of 1. But for example, if e_i is perceived to dominate e_j with respect to a property (E_c), then a numerical scale ranging from 2 to 9 is being used to represent the intensity of dominance (e.g., moderate, strong, very strong, absolute) that reflects the DM's value judgment. The first two columns of Table 2 describe this numerical intensity and the corresponding verbal definition of the scale proposed by Saaty (1980). However, the use of these crisp numbers has been criticized because of their inability to adequately handle the uncertainty and imprecision associated with the mapping of the DM's perception or subjective judgment (Deng, 1999). Thus, several studies have proposed the use of fuzzy sets or fuzzy numbers $\tilde{a}_{ij}$ to represent such uncertain judgments in the pairwise comparison matrix.

Table 2
The dominance scale for pairwise comparative judgment $\tilde{a}_{ij} \geq 1$

Numerical scale	Linguistic scale [D(element <i>i</i>) $\geq$ D(element <i>j</i>)] ^a	Fuzzy scale: TFN (<i>l_{ij}</i> , <i>m_{ij}</i> , <i>u_{ij}</i>) ^b
1	Just equal	(1,1,1)
2	Equal to moderate	(max(<i>L_B</i> , 2– δ), 2, min(<i>U_B</i> , 2+ δ))
3	Moderate dominance	(max(<i>L_B</i> , 3– δ), 3, min(<i>U_B</i> , 3+ δ))
4	Moderate to strong	(max(<i>L_B</i> , 4– δ), 4, min(<i>U_B</i> , 4+ δ))
5	Strong dominance	(max(<i>L_B</i> , 5– δ), 5, min(<i>U_B</i> , 5+ δ))
6	Strong to very strong	(max(<i>L_B</i> , 6– δ), 6, min(<i>U_B</i> , 6+ δ))
7	Very strong dominance	(max(<i>L_B</i> , 7– δ), 7, min(<i>U_B</i> , 7+ δ))
8	Very strong to absolute	(max(<i>L_B</i> , 8– δ), 8, min(<i>U_B</i> , 8+ δ))
9	Absolute dominance	(max(<i>L_B</i> , 9– δ), 9, min(<i>U_B</i> , 9+ δ))

^aFor pairwise verbal comparisons, dominance of element *i* over element *j* may be interpreted as importance, preference, likelihood or influence.

^bLB and UB refers to the lower bound and upper bound of the scale ($\tilde{a}_{ij} \geq 1$), respectively.

The fuzzy extension of such pairwise comparison matrix is referred to as the fuzzy pairwise comparative judgment matrix or simply as fuzzy pairwise comparison matrix $\tilde{A} = [\tilde{a}_{ij}]$.

The fuzzy pairwise comparison matrix is therefore one of the main characteristics of various fuzzy AHP (FAHP) methods proposed in the literature. During the past two decades, these so-called FAHP methods have applied systematic approaches that employ the concepts of FST and hierarchical structure analysis. For instance, the earliest work in FAHP appeared in Van Laarhoven and Pedrycz (1983), which extend the AHP to a fuzzy hierarchical analysis using comparison matrices with TFNs. They obtain fuzzy priorities by applying a fuzzy version of the logarithmic least-squares method. A modification of their method was later proposed in Boender et al. (1989) to make a more robust approach in the normalization of local priorities. On the other hand, Buckley (1985) used the fuzzification of the geometric mean method to determine the fuzzy priorities from judgments whose membership functions were trapezoidal, and later Buckley et al. (2001) proposed an evolutionary algorithm of computing fuzzy priorities by direct computation of fuzzy eigenvalues and fuzzy eigenvectors. There are also other papers proposing different procedures related to fuzzy hierarchical analysis or FAHP (see Csutora and Buckley (2001), Mikhailov (2003a), and Kahraman et al. (2004) for related discussions).

While there are many proposed FAHP methods with notable differences in their theoretical foundation, only a handful of these methods are being applied in real-life problems because of computational complexity or counter-

intuitive results. Many attempted to ‘fuzzify’ AHP but limited studies had been done to investigate the fuzzy extension of ANP including its practical application to real-life decision problems. Since the supermatrix priority derivation process in the ANP requires complex matrix operations on real numbers, the most practical approach to extend FAHP to the ANP framework is to derive first the crisp priorities or weights from the fuzzy comparison matrices (Mikhailov and Singh, 2003b).

Some researchers (Kahraman et al., 2006; Buyukozkan et al., 2004) applied a fuzzy ANP approach to quality function deployment problems. Their method is an extension of FAHP approach proposed by Chang (1996), which derives crisp local priorities from fuzzy comparison matrix using the extent analysis method and possibility theory. However, they used a rather simplified supermatrix calculation which appears, in our opinion, far removed from that of the original ANP. In addition, their proposed algorithm may yield a zero value of initial weights or local priorities to some elements of the decision structure. A computed zero local priority (e.g., the importance weight of a criterion with respect to the goal) from the extent analysis could be problematic because some paths of interactions will not be considered in the supermatrix calculations. It should be noted that an unimportant element (i.e., having the least local priority value) in a cluster may still be important overall (i.e., having the highest overall priority) because of dependence and feedback in the decision structure (Saaty, 2001). Thus, the extension of Chang’s extent analysis method to the ANP framework may have some drawbacks in the sense that the outcome could not capture all the possible interactions in the decision structure.

In this paper, an algorithm introduced by Cheng and Mon (1994) was used for the proposed fuzzy extension of ANP. Their fuzzy-scale-based algorithm used alpha-cuts, interval arithmetic, and linear convex combinations defined by optimism index values to transform the fuzzy comparison matrix into a set of crisp matrices; and then computed the set of crisp local weights from eigenvector of the said matrix. Although several authors have recently applied this FAHP method (Kwong and Bai, 2002; Buyukozkan et al., 2004; Ayag, 2005), to our knowledge, it has never been tested within the supermatrix framework.

3. The proposed fuzzy ANP

Before we turn to the technical discussion of the proposed fuzzy ANP, it should be pointed out that the concept of ‘fuzziness’ in human judgment within the framework of ANP is independent from the traditional concept of ‘inconsistency’ in judgment. Note that it is conceivable for a DM to express highly inconsistent preferences with a very low level of ‘fuzziness’. In AHP/ANP, inconsistency of judgments in pairwise comparisons is measured by the consistency ratio (CR) proposed by Saaty (1980) as a test of the reliability of the decision

outcomes. If the computed CR is more than 0.10, the DM is asked to reconsider and modify his judgment to improve the consistency according to his understanding. In this sense, allowing a modicum of inconsistency in pairwise comparison matrix can be seen as one of the attempts to incorporate 'fuzziness' indirectly in the decision model of AHP/ANP. Higher value of CR implies that inconsistency occurs because of some 'errors in judgment' on the part of the DM. Aside from recognizing that real-world inconsistencies do exist, Forman and Selly (2001) identified some of the causes of inconsistency such as lack of information, lack of concentration during judgment process, and inadequate model structure. On the other hand, Bryson et al. (1995) asserted that the intransitivity and inconsistency may also be 'errors in judgment' forced on a DM who sees the preference situation and the preference score as occurring over a range, but is required to give a point estimate. He argued that this pair of phenomena appears, not primarily because of 'errors in judgment', but rather because of the vagueness that is inherent in subjective and qualitative evaluations.

The aim of fuzzy ANP is to capture the 'fuzziness' or the vagueness-type uncertainties in the evaluation of remedial countermeasures particularly at the initial phase of remediation planning. Due to the complexity and uncertainty involved, as well as, the inherent subjective nature of human judgments, it is sometimes unrealistic and infeasible to acquire exact judgments in pairwise comparisons. It is more natural or easier to provide verbal judgments when giving subjective assessment. For instance, the DM may say with confidence that alternative *A* is very strongly more preferred than alternative *B* with respect to a criterion, but when asked to give an exact ratio of how strongly *A* dominates *B*, he may have difficulties in giving a precise numerical value. In addition, other researchers (e.g., Goddard, 1983; Kirkwood and Sarin, 1985; Ngwenyama and Bryson, 1999) have also noted that it is difficult to map qualitative preferences to point estimates. This introduces uncertainties in the decision-making process as natural language is most often vague and fuzzy. Such source of imprecision in real complex decision making is inevitable and this type of uncertainties can be handled using FST. Moreover, fuzzy-based techniques can be viewed as a generalized form of interval analysis that addresses uncertain or vague information. Therefore, we modified the conventional ANP to fuzzy ANP by employing fuzzy numbers in the semantic scale, instead of Saaty's crisp 9-point scale. In the next section, the proposed fuzzy scale and prioritization approach are further discussed.

3.1. Fuzzy scale

In the proposed method, the TFN were utilized to extend the standard AHP 9-point scaling scheme in order to capture the vagueness and uncertainty of judgment in the pairwise verbal comparison. Representing the scale as TFN

(l_{ij}, m_{ij}, u_{ij}) explicitly describes the fuzziness of the judgment denoted as $\tilde{a}_{ij}$. This will be characterized by the degree of fuzziness (δ) or the spread of fuzzy number ($\delta_L + \delta_R$). The δ_L is a parameter defined as the absolute distance from the lower bound (l_{ij}) to the modal value (m_{ij}) while δ_R is the absolute distance from the modal value (m_{ij}) to the upper-bound value (u_{ij}) (see Fig. 3). However, instead of using a fixed predefined membership function commonly found in FAHP literature, the following membership function for the 'fuzzification' of Saaty's fundamental scale will be defined according to the degree of fuzziness, lower bound and upper bound of the scale (see Table 2). Here the effect of these parameters on the proposed fuzzy scale was elucidated.

Note that the value judgments elicited during pairwise comparisons are associated with numbers from the scale to represent how many times the larger of the two elements dominates the smaller one with respect to a property or criterion which they have in common. According to AHP's reciprocal axiom, the smaller element would have the reciprocal or inverse value when compared with the larger element (Saaty, 1980). Consequently, the proposed membership function will be based on a premise that the support of any fuzzy scale should be mapped in either the value of 0–1, or 1– ∞ subdomains. Thus, the lower and upper bound of the dominance scale (i.e., $\tilde{a}_{ij} \geq 1$) are set equal to 1 and ∞ , respectively.

For example, consider the representation of a verbal judgment expressed as $\tilde{a}_{ij} = \tilde{3}$. If the following membership function $(m_{ij} - \delta, m_{ij}, m_{ij} + \delta)$ will be used in which $\delta = 3$, then the associated fuzzy number is (0, 3, 6) and its reciprocal is $\tilde{3}^{-1}$ or (1/6, 1/3, ∞). Note that at $l_{ij} = 0$, the subdomain of the support (0–1) becomes meaningless for a judgment ($\tilde{a}_{ij} = \tilde{3}$ in which the element *i* has moderate dominance over element *j*). Likewise, at $u_{ij} = \infty$, the subdomain of the support (1– ∞) becomes meaningless for the reciprocal of this judgment. Accordingly, the defined fuzziness in the scale ($m_{ij} \neq 1$) does not consider the possibility of reversal in the judgment initially given by a human DM. This means that the DM is assumed to be certain about the direction of preference in spite of his inability of assigning precise numerical representation to his verbal judgment. It is then more appropriate to use the following TFN ($\tilde{M} \rightarrow \tilde{a}_{ij}$) to describe the fuzzy judgment with scale value of $m_{ij} > 1$:

$$\begin{aligned}\tilde{M} &= (\max(1, m_{ij} - \delta), m_{ij}, \min(\infty, m_{ij} + \delta)) \text{ or} \\ \tilde{M} &= (\max(1, m_{ij} - \delta), m_{ij}, m_{ij} + \delta).\end{aligned}\quad (9)$$

Using Eq. (8), the reciprocal of the TFN $\tilde{M}$ ($m_{ij} > 1$) can be denoted as

$$\tilde{M}^{-1} = \left(\frac{1}{m_{ij} + \delta}, \frac{1}{m_{ij}}, \frac{1}{\max(1, m_{ij} - \delta)} \right). \quad (10)$$

Figs. 4a and 4b describe an example of the fuzzy scale for δ values of 1 and 3, respectively.

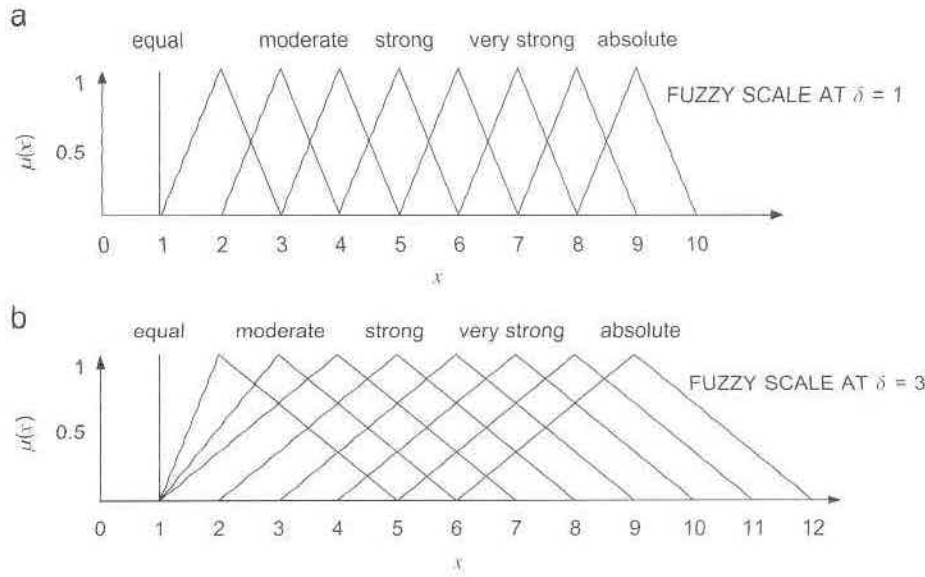


Fig. 4. Fuzzy scale and the corresponding membership function at (a) $\delta = 1$ and (b) $\delta = 3$.

3.2. Outline of the procedure

The Cheng and Mon's fuzzy-scale-based algorithm was incorporated in the PRMEHN model. The algorithm of the proposed fuzzy ANP can be described in the following steps:

- (1) Decompose the multiple criteria evaluation problem into a strongly connected hierarchical network.
- (2) Elicit value judgments from DM to compare each pair of elements with respect to an element. The TFNs in the fuzzy scale are used to indicate the relative dominance of one element over another. Then, construct the fuzzy pairwise comparison matrix such that:

$$\tilde{A} = \begin{bmatrix} 1 & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ \tilde{a}_{21} & 1 & \cdots & \tilde{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{n1} & \tilde{a}_{n2} & \cdots & 1 \end{bmatrix}, \quad (11)$$

where $\tilde{a}_{ij} \geq 1$ and $\tilde{a}_{ji} = 1/\tilde{a}_{ij}$.

- (3) Solve the fuzzy eigenvalue problem:

$$\tilde{A}\tilde{w} = \tilde{\lambda}\tilde{w}, \quad (12)$$

where $\tilde{A}$ is an $n \times n$ fuzzy matrix containing fuzzy numbers $\tilde{a}_{ij}$ and $\tilde{w}$ is a positive $n \times 1$ fuzzy vector containing fuzzy numbers. To perform fuzzy multiplications ($\otimes$) and additions ($\oplus$), the equation $\tilde{A}\tilde{w} = \tilde{\lambda}\tilde{w}$ is equivalent to

$$(\tilde{a}_{i1} \otimes \tilde{w}_1) \oplus \cdots \oplus (\tilde{a}_{in} \otimes \tilde{w}_n) = \tilde{\lambda} \otimes \tilde{w}_i \quad \forall 1 \leq i \leq n, \quad (13)$$

where $\tilde{A} = [\tilde{a}_{ij}]$, $\tilde{w} = (w_1, w_2, \dots, w_n)$ and $\tilde{\lambda}$ is the fuzzy eigenvalue.

Using alpha-cut (α) which incorporates the DM's confidence over his judgment, the interval values for $0 \leq \alpha \leq 1$ are

$$\tilde{a}_{ij}^\alpha = [a_{ijl}^\alpha, a_{iju}^\alpha], \quad \tilde{w}_i^\alpha = [w_{il}^\alpha, w_{iu}^\alpha], \quad \tilde{\lambda}_i^\alpha = [\lambda_l^\alpha, \lambda_u^\alpha].$$

Using the interval arithmetic operations, the following equation will be

$$[a_{i1l}^\alpha w_{1l}^\alpha, a_{i1u}^\alpha w_{1u}^\alpha] \oplus \cdots \oplus [a_{inl}^\alpha w_{nl}^\alpha, a_{inu}^\alpha w_{nu}^\alpha] = [\lambda_l^\alpha w_{il}^\alpha, \lambda_u^\alpha w_{iu}^\alpha], \quad (14a)$$

or $\forall 1 \leq i \leq n$

$$a_{i1l}^\alpha w_{1l}^\alpha + \cdots + a_{inl}^\alpha w_{nl}^\alpha = \lambda_l^\alpha w_{il}^\alpha, \quad a_{i1u}^\alpha w_{1u}^\alpha + \cdots + a_{inu}^\alpha w_{nu}^\alpha = \lambda_u^\alpha w_{iu}^\alpha. \quad (14b)$$

- (4) Assign an optimism index ω value. The optimism index is a linear convex combination defined as

$$\hat{a}_{ij}^\alpha = \omega a_{iju}^\alpha + (1 - \omega) a_{ijl}^\alpha \quad \forall \omega \in [0, 1], \quad (15)$$

The use of optimism index allows the algorithm to reflect the DM's attitude towards the fuzziness of judgment. For example, an 'optimistic' DM ($\omega = 1$) is apt to assign higher values of the crisp interval at a given level of confidence (α) while a 'pessimistic' DM ($\omega = 0$) tends to favor the lower values of the said interval.

- (5) With α fixed, construct the following matrix after setting the ω parameter:

$$\hat{A} = \begin{bmatrix} 1 & \hat{a}_{12}^\alpha & \cdots & \hat{a}_{1n}^\alpha \\ \hat{a}_{21}^\alpha & 1 & \cdots & \hat{a}_{2n}^\alpha \\ \vdots & \vdots & \ddots & \vdots \\ \hat{a}_{n1}^\alpha & \hat{a}_{n2}^\alpha & \cdots & 1 \end{bmatrix}. \quad (16)$$

- (6) From Eq. (16), calculate the local priorities or the normalized principal right eigenvector for all α -cut values [0,1] at a given fixed ω value. Computational-wise, the principal right eigenvector can be approximated from the following equation (Saaty, 1980):

$$\lim_{k \rightarrow \infty} \frac{\hat{A}^k e}{e^T \hat{A}^k e} = c \hat{w}, \tag{17}$$

where $e^T = [1, 1, \dots, 1]$ is a unit row vector, c is a constant and $\hat{w}$ is the eigenvector corresponding to the principal eigenvalue of the primitive matrix $\hat{A}$. Thus, the computation of principal eigenvector can easily be implemented in a spreadsheet (e.g., Microsoft Excel 2003); for example, by consecutively squaring the matrix, normalizing the row sums each time, and stopping the procedure when the normalized row sums converge to a limiting value.

- (7) Compute the overall priorities from principal column eigenvector of the initial supermatrix of the hierarchical network. The supermatrix is initially populated by the pertinent eigenvectors computed from step 6. Computational-wise, this column eigenvector can be approximated from an algorithm which can also be easily implemented in a spreadsheet using the following equation:

$$\lim_{p \rightarrow \infty} \frac{\hat{S}^p e}{e^T \hat{S}^p e} = q \hat{v}, \tag{18}$$

where $e^T = [1,1,\dots,1]$ is a unit row vector, q is a constant and $\hat{v}$ is the eigenvector corresponding to the principal eigenvalue of the primitive irreducible supermatrix $\hat{S}$. Then, the meaningful overall priorities can be derived from the column eigenvector normalized according to pertinent clusters or levels.

- (8) Repeat calculations from Steps 4 to 7 to derive the set of desired priorities for all α -cuts at different ω values.

4. A fuzzy ANP application: prioritization of remedial countermeasures

4.1. Problem description

Consider a decision problem described in Promentilla et al. (2005b, 2006a), i.e., to select the most appropriate

remedial alternatives for a contaminated site caused by illegal dumping of hazardous industrial wastes to a privately owned uncontrolled landfill located in one of the prefectures in Japan. In this decision context, the final DM is the prefectural government authority who will prioritize the feasible remedial alternatives and select the appropriate countermeasure for the site. The three remedial countermeasure strategies are briefly described relative to (1) site remediation containing the waste layer, (2) prevention of contaminant spreading to the surrounding area (plume control) and (3) remediation of the surrounding area (see Table 3).

In this situation, there is no single dominating alternative for all the criteria. For example, A1 is perceived to dominate in terms of environmental effectiveness and social acceptability. However, its desirability may be relatively low in terms of financial affordability and implementability because of its potentially high clean-up cost, as well as, difficulties in finding a nearby incineration or treatment facility for the excavated waste. On the other hand, A2 may have the least clean-up cost but it may also have some difficulties in effectively controlling the plume. In the case of A3, it is perceived to dominate in terms of implementability but its desirability is relatively lower with respect to the other criteria in which A1 or A2 dominates. In order to prioritize these three alternatives for further remediation design and analysis, it is necessary to have a systematic and transparent way to evaluate these alternatives and measure their overall relative desirability.

4.2. Decision structure under consideration

Fig. 5 shows the digraph structure and the generalized supermatrix representation of a three-level hierarchical network decision model to be considered in this case study. The decision problem is decomposed into three clusters namely the goal (G), the criteria (C), and the alternatives (A) cluster. The goal or focus of the problem is to prioritize the identified set of remedial alternatives described in Table 3. The evaluation criteria to achieve this goal are defined as follows: (1) environmental effectiveness (EE) to protect public health and environmental resources, (2) financial affordability (FA) with regards to the overall cost

Table 3
The remedial countermeasure strategies (alternatives) in this illustrative example

Remedial alternatives (A)	Site remediation containing the waste layer	Prevention of contaminant spreading	Remediation of surrounding area
Alternative 1 (A1)	Removal of all waste materials and off-site treatment (e.g., incineration, vitrification) and disposal of the treated waste		In situ remediation (e.g., enhanced bioremediation, natural attenuation)
Alternative 2 (A2)	In situ remediation (e.g., enhanced bioremediation, soil washing, etc.)	Capping and hydraulic control (e.g., groundwater extraction)	In situ remediation (e.g., enhanced bioremediation, natural attenuation)
Alternative 3 (A3)	In situ remediation (e.g., enhanced bioremediation, soil washing, etc.)	Capping, vertical cut-off wall (e.g., sheet piling, chemical grout, etc.)	In situ remediation (e.g., enhanced bioremediation, natural attenuation)

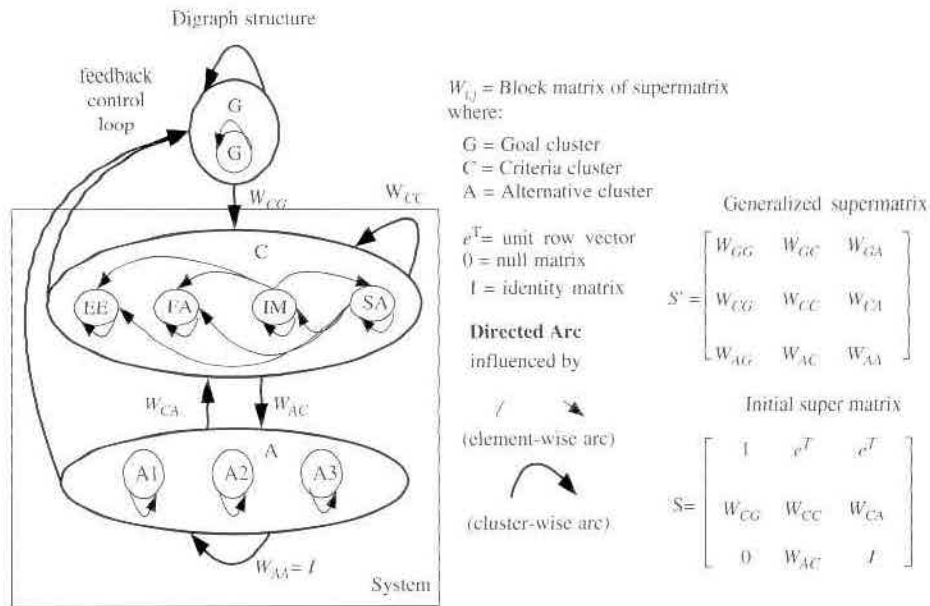


Fig. 5. The decision structure under consideration and its supermatrix representation.

of clean-up, (3) implementability (IM) in terms of administrative and technological feasibility and (4) the social acceptability (SA) of the countermeasure according to the perception of stakeholders. It should be noted that the 3-level hiernet model used in this case study is primarily for the purpose of illustration. A more complex multi-level hiernet model could also be applied depending on how the problem will be structured by the analyst or DMs.

4.3. A numerical example

For the purpose of brevity, the details of the pairwise comparison matrices in this case study were not discussed in this paper. Only a sample of the numerical calculations was shown to demonstrate the proposed method. For instance, suppose a DM was asked to express his judgment about the relative dominance of the remedial alternatives with respect to environmental effectiveness criterion. The questionnaire can be formulated like this: "Comparing Alternative 1 (A1) and Alternative 2 (A2), which one is more preferable with respect to environmental effectiveness (EE), and by how much is it preferred over the other?" For example, if the DM considers that A1 is preferred *very strongly* over A2 with respect to EE, then this judgment corresponds to $\tilde{a}_{12} = \tilde{7}$ and $\tilde{a}_{21} = 1/\tilde{7}$. After the value judgment elicitation process, the corresponding fuzzy pairwise comparative judgment matrix is established:

$$\tilde{A}_{AC_1} = \begin{bmatrix} 1 & \tilde{7} & \tilde{3} \\ \tilde{7}^{-1} & 1 & \tilde{5}^{-1} \\ \tilde{3}^{-1} & \tilde{5} & 1 \end{bmatrix}, \quad (19)$$

where subscript A refers to alternatives cluster (A) and subscript C_1 refers to the first element (EE) of the criteria cluster (C).

Given the membership function of TFN described in Table 1, the interval value $\tilde{a}_{ij}^z = [a_{ij}^z, a_{ji}^z]$ of the fuzzy numbers at a given δ and α were computed using Eq. (4). For example, if $\delta = 1$, the interval value of TFN ($m_{ij} > 1$) and its reciprocal at a given α -cut are as follows:

$$\tilde{a}_{ij}^z = [l_{ij} + \alpha, u_{ij} - \alpha], \quad \tilde{a}_{ji}^z = \frac{1}{\tilde{a}_{ij}^z} = \left[\frac{1}{u_{ij} - \alpha}, \frac{1}{l_{ij} + \alpha} \right], \quad (20)$$

where l_{ij} and u_{ij} are the lower and upper bounds of the support of the fuzzy number (Fig. 4a), respectively.

Let $\alpha = 0.5$ and $\omega = 0.5$, given the fuzzy comparative judgment matrix in Eq. (19), the following matrix and its principal right eigenvector were derived:

$$\hat{A}_{AC_1} = \begin{bmatrix} 1 & 7 & 3 \\ 0.143 & 1 & 0.202 \\ 0.343 & 5 & 1 \end{bmatrix}, \quad w_{AC_1} = \begin{bmatrix} 0.648 \\ 0.072 \\ 0.280 \end{bmatrix}. \quad (21)$$

Following the same calculations, all the pertinent normalized eigenvectors or local priorities which are needed to populate the initial supermatrix of the hierarchical network were computed. Fig. 6 is an example of the initial supermatrix and the derived normalized overall priorities at $\alpha = 0.50$ and $\omega = 0.50$. For instance, the column eigenvector derived in Eq. (21) is one of the entries (as shown in bold font) in the initial supermatrix (Fig. 6). Similar calculations were also done at different α -cuts for ω values of 0.0, 0.5 and 1.0. Figs. 7 and 8 illustrate the derived overall priorities of criteria and alternatives at

a									b		
S	G	EE	FA	IM	SA	A1	A2	A3	e-vector priorities		
G	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	G	0.382	1.000
EE	0.250	1.000	0.000	0.099	0.185	0.410	0.067	0.071	EE	0.100	0.262
FA	0.250	0.000	1.000	0.086	0.054	0.084	0.689	0.562	FA	0.129	0.337
IM	0.250	0.000	0.000	0.695	0.081	0.055	0.172	0.303	IM	0.075	0.196
SA	0.250	0.000	0.000	0.120	0.679	0.451	0.072	0.064	SA	0.078	0.205
A1	0.000	0.648	0.054	0.119	0.672	1.000	0.000	0.000	A1	0.082	0.349
A2	0.000	0.072	0.586	0.272	0.102	0.000	1.000	0.000	A2	0.069	0.290
A3	0.000	0.280	0.361	0.609	0.226	0.000	0.000	1.000	A3	0.085	0.361

Fig. 6. The initial supermatrix and the derived normalized overall priorities at $\alpha = 0.5$, $\omega = 0.5$ in this numerical example.

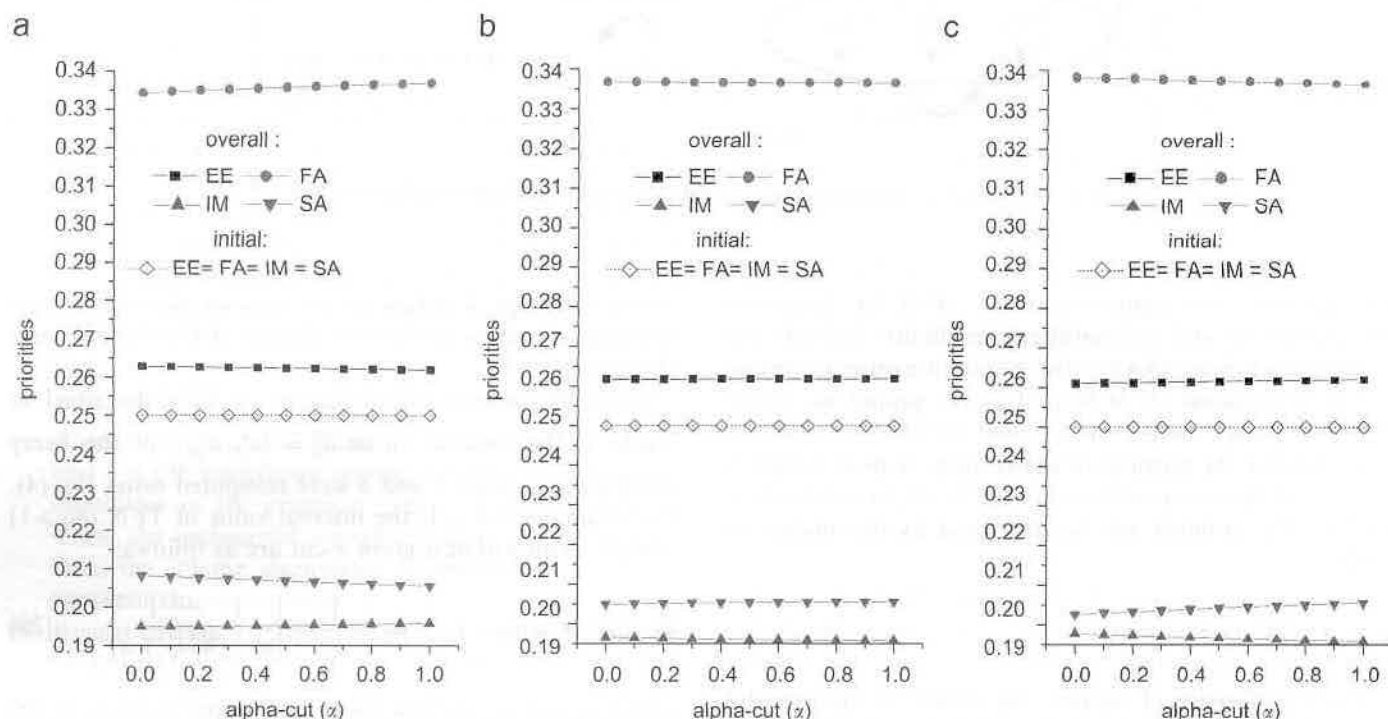


Fig. 7. Overall priorities of criteria at different α and ω values in this numerical example ($\delta = 1$).

different alpha-cuts and ω values in this numerical example.

5. Discussion

Outputs from the fuzzy ANP model are set of possible values of overall priorities according to DM's level of confidence and his attitude toward fuzziness over the evaluation. As mentioned before, the alpha-cut concept represents the DM's degree of confidence in the fuzzy assessment. As alpha-cut (α) approaches to zero, the wider the interval of numerical intensity from the scale becomes. This means that $\alpha = 0$ entails lack of confidence over the subjective assessment and thus consider the full range of the fuzziness; whereas the higher value of α represents a more confident DM, and reaches maximum at $\alpha = 1$ when the value approaches to the most likely value, i.e., the

modal value of the fuzzy scale. Further, given the desired confidence over the assessment ($0 \leq \alpha < 1$), the attitude toward fuzziness is represented by the ω value. Accordingly, when ω value approaches to 0, it reflects DM's attitude of representing his verbal judgment in terms of the lower values of the alpha-cut interval. On the other hand, when ω value approaches to 1, it reflects DM's attitude of representing his verbal judgment in terms of the higher values of the alpha-cut interval.

From Figs. 7 and 8, the results indicate the extent of robustness of the prioritization method when fuzziness of judgment is incorporated in the evaluation process. In this numerical example, the DM gave initially an equal importance to each criterion with respect to the goal. This equal weighting of criteria results to local weights of 0.25 for each criterion in the initial supermatrix (see Fig. 6). However, as interdependence among criteria and

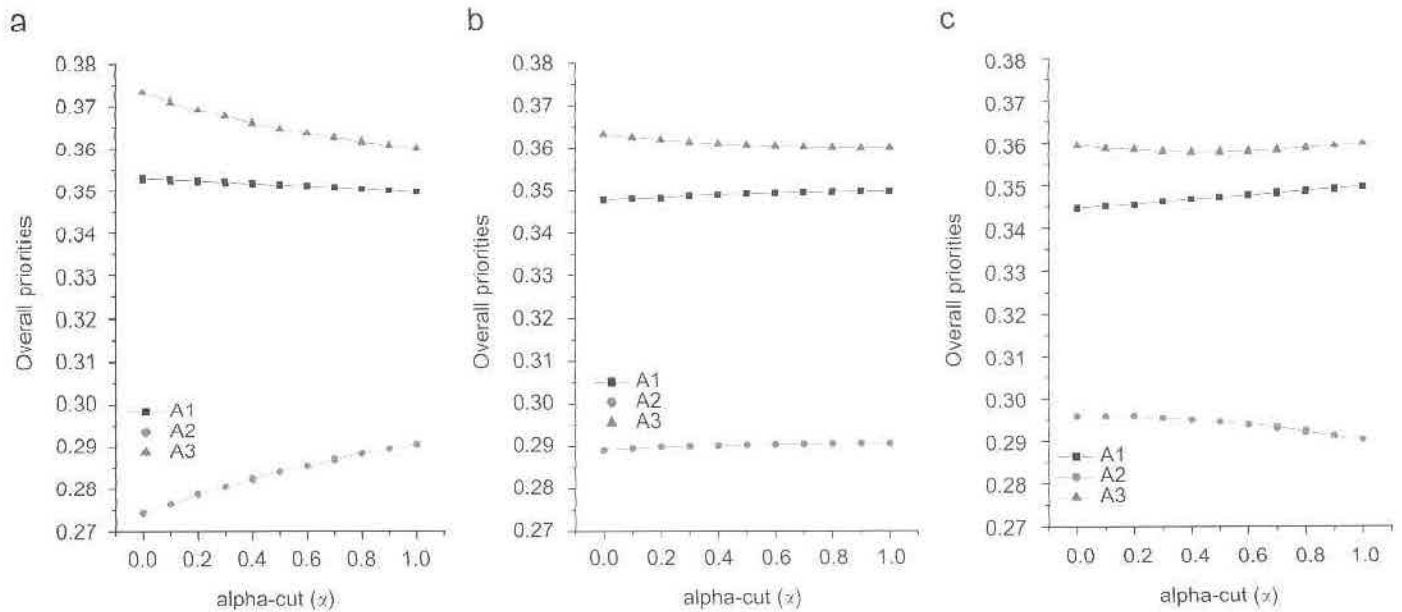


Fig. 8. Overall priorities of alternatives at different α and ω values in this numerical example ($\delta = 1$).

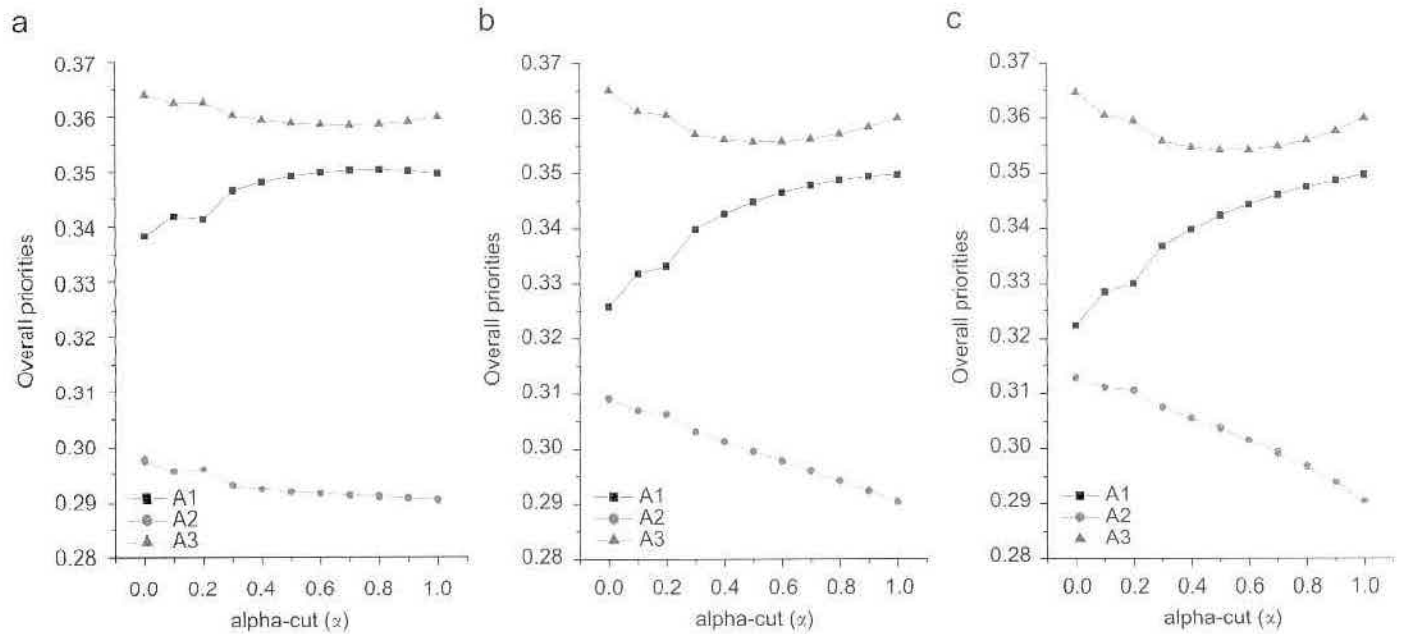


Fig. 9. Overall priorities of alternatives at different α and ω values in this numerical example ($\delta = 3$).

feedback dependence from the alternatives were captured in the evaluation process, FA became the most important criterion followed by EE. The incorporation of fuzziness ($\delta = 1$) in the evaluation process did not significantly affect the relative importance of the criteria as indicated by the computed overall priorities at different alpha-cuts and ω values (Fig. 7). Moreover, Fig. 8 describes the robust rank order of the countermeasure strategies ($A3 > A1 > A2$) at varying degree of DM's level of confidence (α) and attitude (ω) toward the fuzziness of his judgment.

In order to understand more the effect of fuzziness of judgment on the prioritization results, the degree of fuzziness was increased from $\delta = 1$ to 3 for the same decision structure and verbal judgment. These results indicate how the degree of fuzziness could influence the degree of dominance of an alternative over another at different alpha-cuts. In Fig. 9, notice that the rank order of the alternatives ($A3 > A1 > A2$) did not change at different alpha-cuts and ω . However, the results at $\delta = 3$ differs from that of $\delta = 1$ in terms of the significant changes on the degree of preference between the alternatives. At more

Table 4
Fuzzy pairwise comparative judgment matrix of the relative importance of criteria with respect to goal in Scenario 1

Goal	EE	FA	IM	SA
EE	1	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$
FA	4	1	2	2
IM	3	$\frac{1}{2}$	1	2
SA	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1

Table 5
Fuzzy pairwise comparative judgment matrix of the relative importance of criteria with respect to goal in Scenario 2

Goal	EE	FA	IM	SA
EE	1	3	2	2
FA	$\frac{1}{3}$	1	$\frac{1}{2}$	$\frac{1}{3}$
IM	$\frac{1}{2}$	2	1	$\frac{1}{3}$
SA	$\frac{1}{2}$	3	3	1

uncertain conditions, the degree of preference of *A3* over *A1* increases while the degree of preference of *A1* over *A2* decreases as shown in Fig. 9.

Another way to understand the behavior of the decision model is to conduct a ‘what-if analysis’ by varying the initial judgment on the relative importance of criteria. Since the base scenario assumes equal weighting of criteria from initial judgment, another two possible scenarios were considered in this study. Scenario 1 could represent a situation wherein a stakeholder/DM such as the local government authority considers financial affordability and implementability as more important than the environmental effectiveness and social acceptability criterion in the prioritization of remedial countermeasure. On the other hand, Scenario 2 could represent a situation wherein a stakeholder/DM such as the local community considers environmental effectiveness and social acceptability as the more important criteria than financial affordability and implementability. Tables 4 and 5 describe an example of the pairwise comparison matrix in Scenarios 1 and 2, respectively. To facilitate the discussion, this paper presented some of the results obtained from applying fuzzy ANP to these two scenarios.

An example of the derived priorities of the criteria in Scenario 1 was shown in Fig. 10. This figure describes the local (or initial) priorities and the overall priorities of criteria in Scenario 1 at different alpha-cuts (α) for $\omega = 0.5$ and $\delta = 1$. Comparing the overall priorities of criteria with their local priorities, the results showed that the relative dominance of FA and EE increased while that of IM decreased after the supermatrix calculation. Notice that the preference order of the criteria (FA > IM > EE > SA) based on the overall priorities was still consistent with that of the

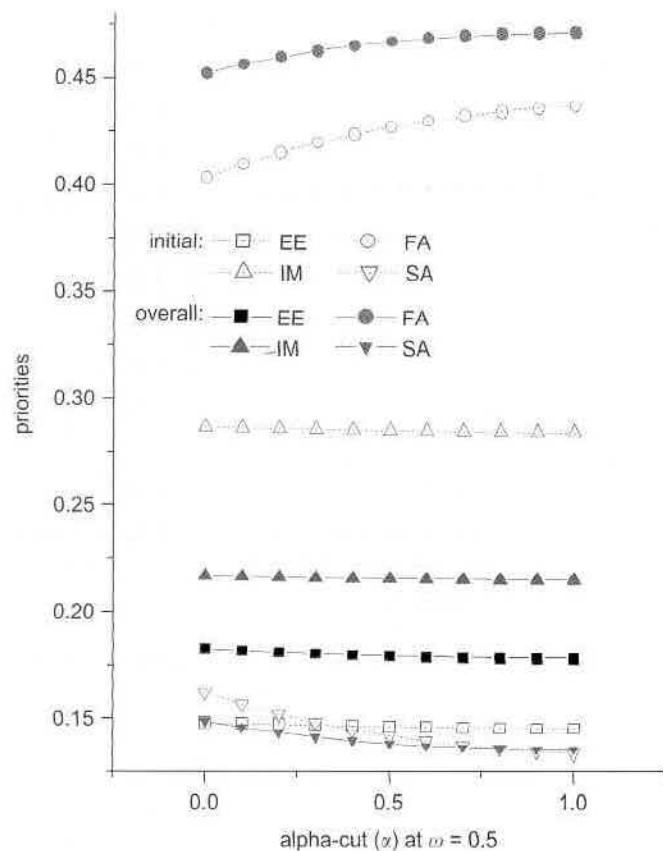


Fig. 10. Initial and overall priorities of criteria at different α given $\omega = 0.5$ for Scenario 1 ($\delta = 1$).

initial priorities, except at more uncertain conditions where preference reversal between EE and SA occurred. At $\alpha < 0.2$, the local priority of SA was a little bit higher than that of EE. However, as all the possible interactions in the decision structure were captured, the resulting overall priorities of EE became higher than that of SA at all alpha-cut levels.

In Figs. 11 and 12, it was shown how the degree of fuzziness could affect the rank order and overall priorities of alternatives in Scenario 1. At $\delta = 1$, the decision model provides a robust rank order ($A3 > A2 > A1$). It follows that the financial affordability and implementability have a significant impact on increasing the relative desirability of *A3* and *A2*. Thus, *A3* is most likely to be selected among the three alternatives given the condition of Scenario 1. However, a rank reversal phenomenon was also observed between *A3* and *A2* at $\delta = 3$ and $\omega = 1$ as shown in Fig. 12. At uncertain conditions wherein the DM tends to overestimate his judgment ($\omega = 1$), the relative desirability of *A2* increased and became greater than that of *A3* in Scenario 1. In this case, rank reversal may occur because of some fuzziness inherent in DM’s judgment when asked to provide a point estimate from subjective assessment.

In the case of Scenario 2, an example of the local and overall priorities of the criteria at different alpha-cuts was shown in Fig. 13. These set of priorities were derived from

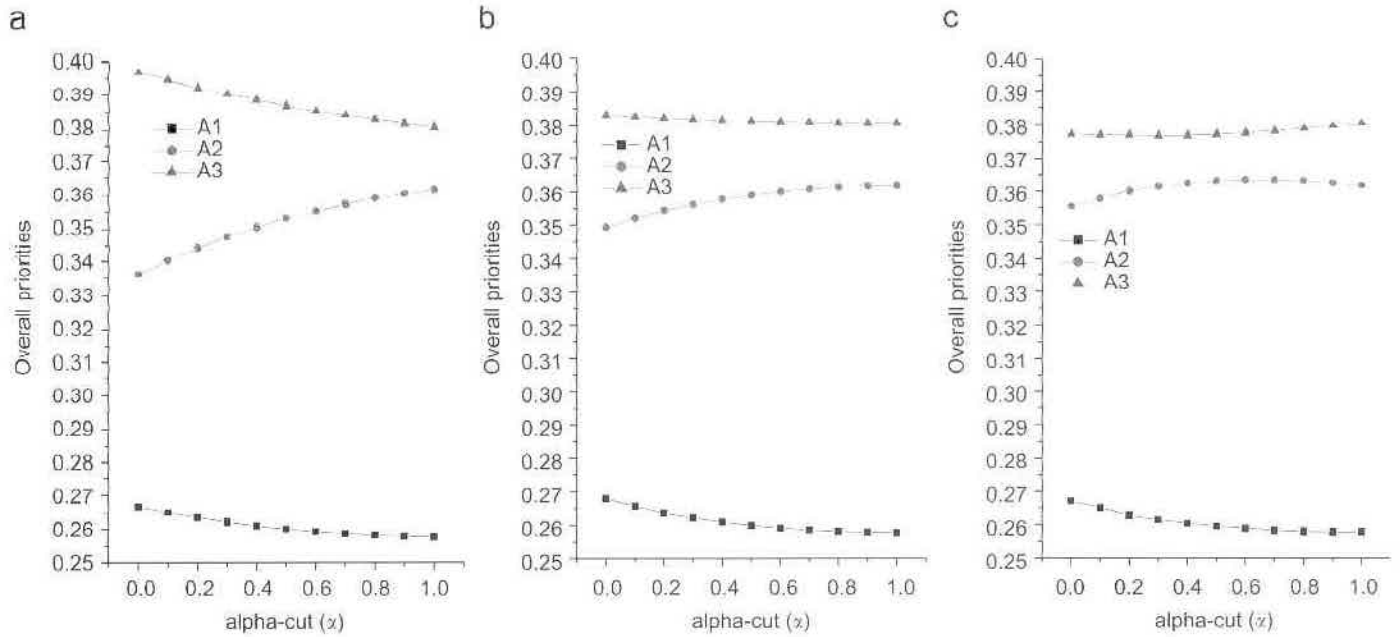


Fig. 11. Overall priorities of alternatives at different z and ω values for Scenario 1 ($\delta = 1$).

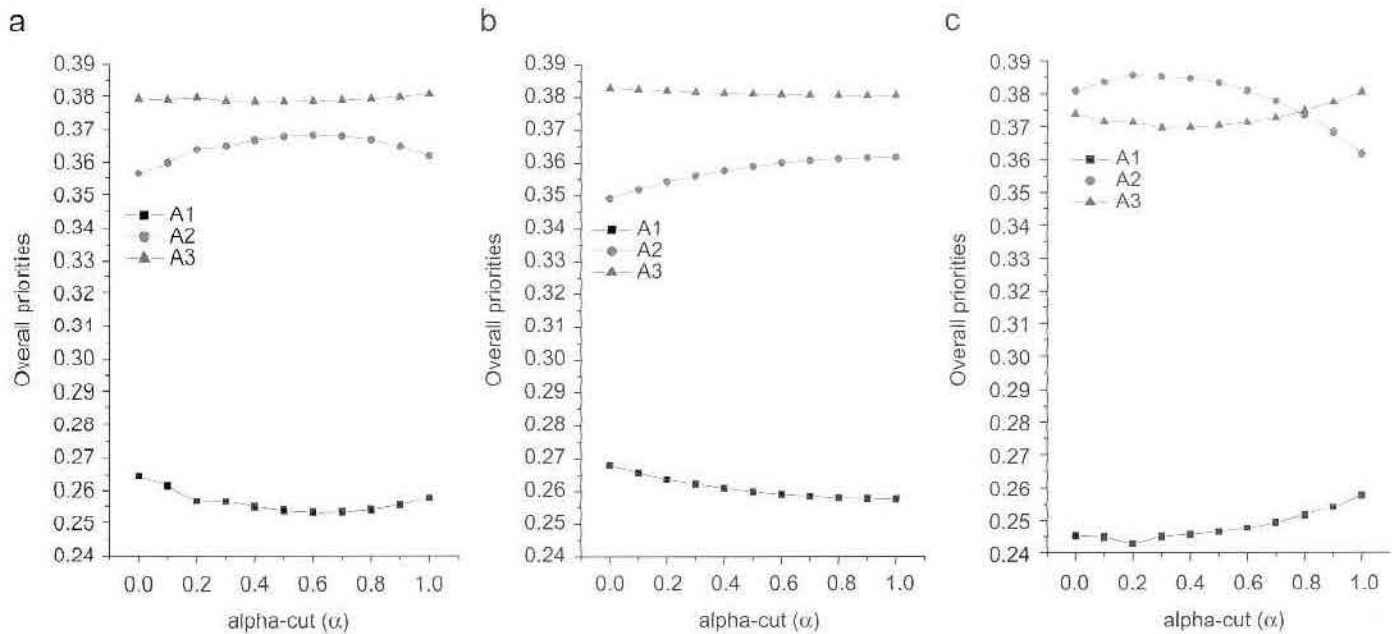


Fig. 12. Overall priorities of alternatives at different α and ω values for Scenario 1 ($\delta = 3$).

fuzzy ANP for $\omega = 0.5$ and $\delta = 3$. The results indicated how the initial (local) priorities were adjusted to the final values (or overall priorities) after supermatrix calculation. For instance, although FA had the least initial priority, the relative priority of FA increased to a final value while that of the other criteria decreased because of the interdependence and feedback effect in the decision structure. Accordingly, a different preference order was observed from the overall priorities of criteria ($EE > SA > FA > IM$) in contrast to that of their initial priorities ($EE >$

$SA > IM > FA$). At more uncertain conditions ($z < 0.1$), a preference reversal was observed between FA and SA but the difference between their derived overall priorities was small (see Fig. 13).

Figs. 14 and 15 show the set of overall priorities of alternatives derived from fuzzy ANP for $\delta = 1$ and 3, respectively. These results describe the robustness of the rank order of the alternatives ($A1 > A3 > A2$) in spite of the fuzziness of judgment. As expected, A1 becomes the most preferred alternative since A1 dominates the other two

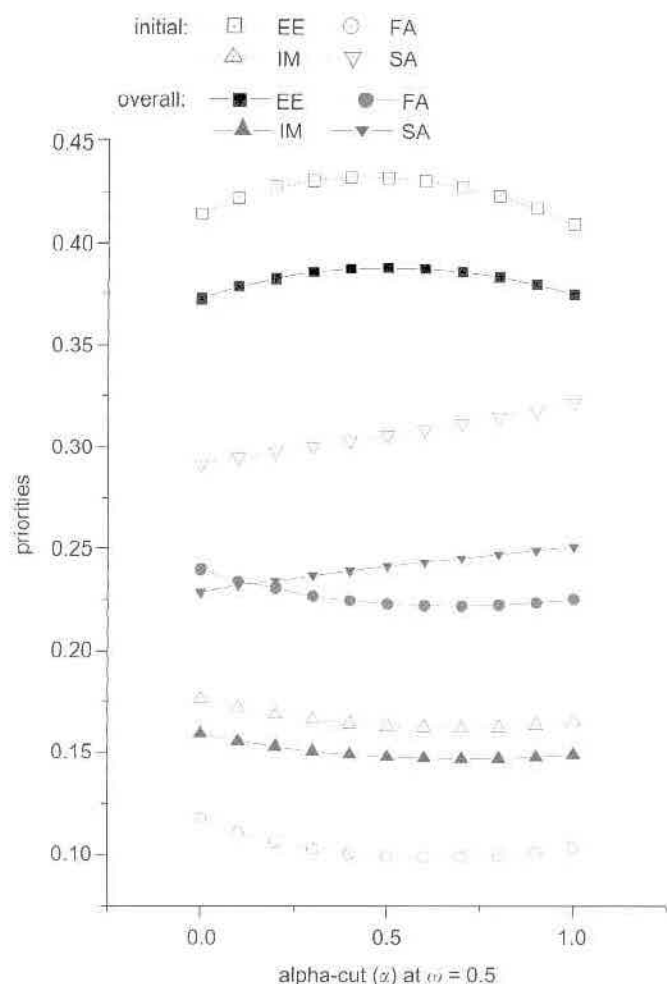


Fig. 13. Initial and overall priorities of criteria at different α given $\omega = 0.5$ for Scenario 2 ($\delta = 3$).

alternatives with respect to EE and SA. In this case, A1 is most likely to be selected among the three alternatives given the condition of Scenario 2.

Fig. 16 summarizes the results of the fuzzy ANP model in this section. This figure describes the high-low-crisp plot of each alternative which is analogous to the high-low-close plot of a financial stock chart. The highest and lowest priority value derived from fuzzy ANP (i.e., $\forall \alpha \in [0, 1]$, $\omega \in [0, 1]$) at a given δ are represented by the ends of vertical bar. On the other hand, the crisp value which is derived from fuzzy ANP for $\alpha = 1$ is shown as horizontal bar. The priorities derived from conventional ANP and fuzzy ANP were also put side by side in this figure for the purpose of comparison.

It is interesting to note that ANP can be viewed as a special case of fuzzy ANP at $\delta = 0$. As indicated by the horizontal bar, the set of crisp priorities from ANP belongs to the set of possible values of priorities derived from fuzzy ANP. This should be expected since the proposed fuzzy scale takes a modal value which is equivalent to the Saaty's 9-point scale. In other words, if we either reduce the degree of fuzziness (δ) to 0 or increase the confidence level (α) to 1.0, the outcome from fuzzy ANP is equivalent to that of conventional ANP. This entails the possibility that DM can map his judgment to a precise numerical value because he has sufficient information or confidence to do so.

One may argue that the results from fuzzy ANP can be obtained with the regular ANP and sensitivity analysis. In practice, the traditional sensitivity analysis used in AHP/ANP requires that a DM may perturb in deterministic fashion the weight of one or more criteria and/or alternatives in order to note changes in the rank of alternatives. Generally, this approach is rather limited to

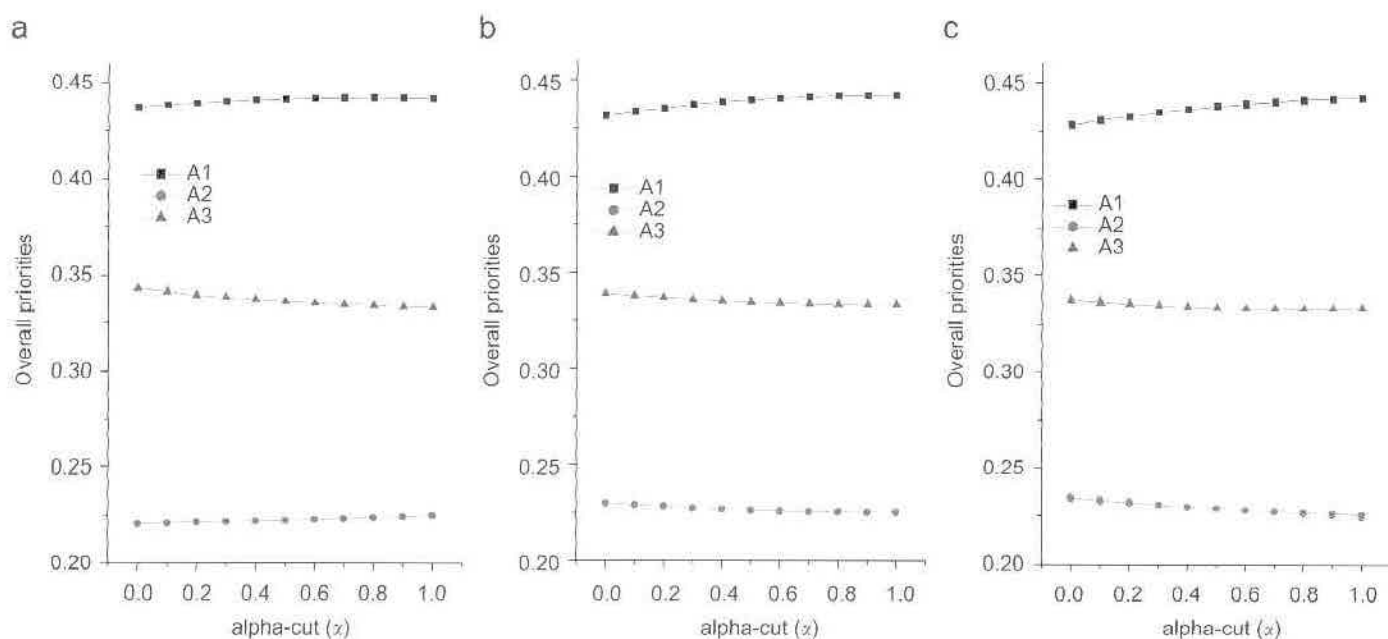


Fig. 14. Overall priorities of alternatives at different α and ω values for Scenario 2 ($\delta = 1$).

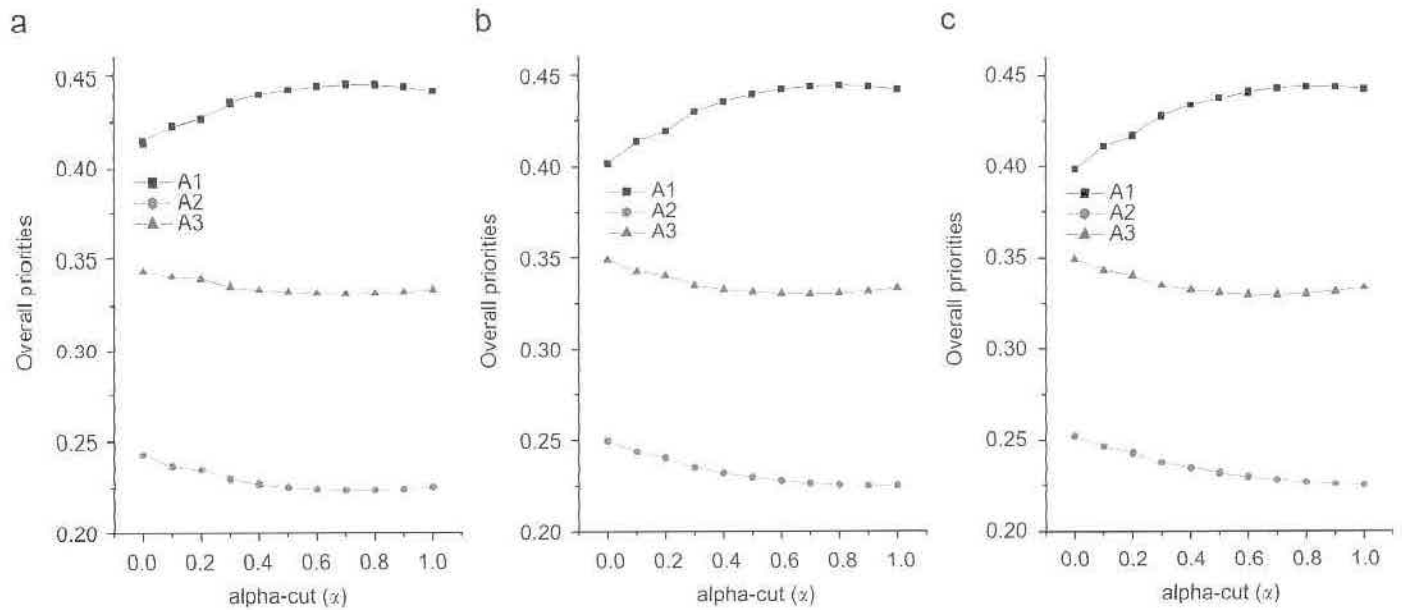


Fig. 15. Overall priorities of alternatives at different α and ω values for Scenario 2 ($\delta = 3$).

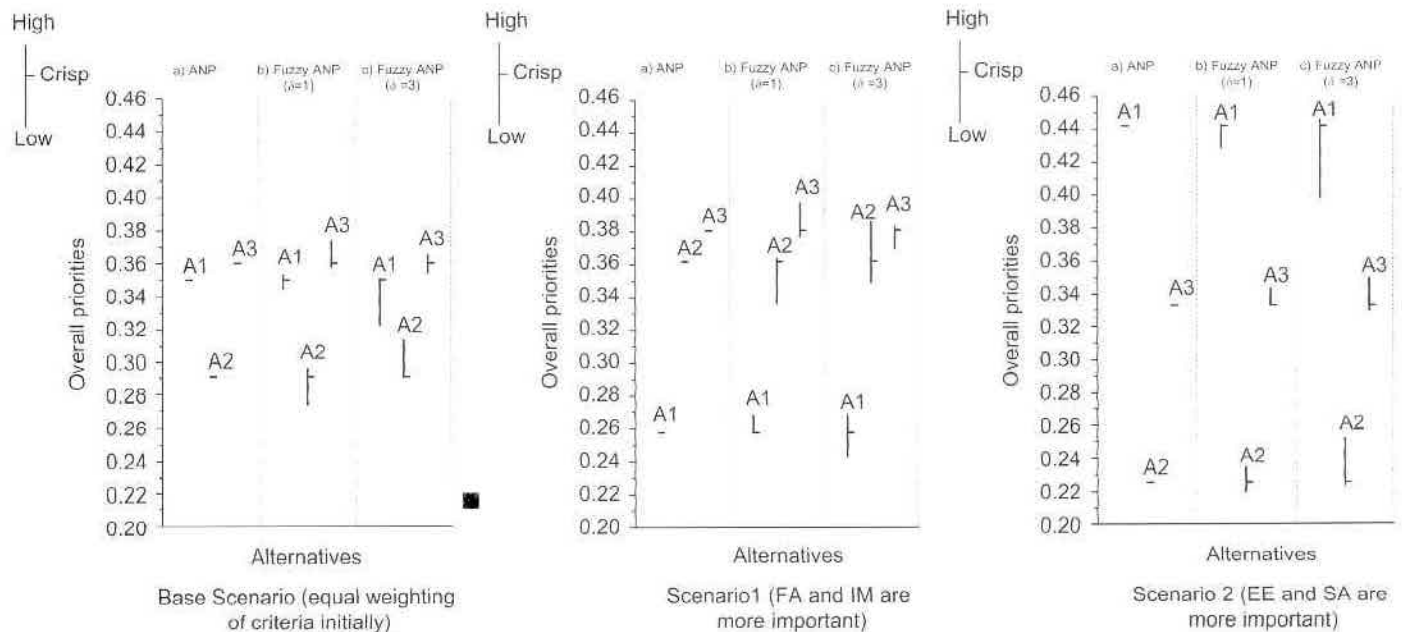


Fig. 16. The high-low-crisp chart of fuzzy ANP output in this numerical example.

single parameter test, i.e., all other parameters are held fixed except for a single one that is allowed to vary. However, the magnitude or the extent to which one should perturb the weights is most often arbitrary. The relationship between realistic variations in judgments may not be known relative to perturbation in the weights themselves, thus the conclusions drawn from such an analysis may not be meaningful and applicable. On the other hand, we believe that the concept of fuzzy ANP in handling uncertainty is quite different from that of regular ANP with sensitivity analysis. Fuzzy ANP accommodates the

subjectivity of human judgment as being expressed in natural language which entails ‘fuzziness’ in real-life decision-making problems. Modeling explicitly the effect of δ on the prioritization results could serve as a learning tool to understand the robustness of the decision model in a fuzzy environment. For instance, results from Fig. 16 could provide information that adds a new dimension to the analysis. They may increase the confidence in the outcome of the model or may expose a high sensitivity of the decision model to the fuzziness of the subjective assessment. In either case, they provide new and useful

information which could complement with the traditional sensitivity analysis. Thus, varying the δ parameter of the fuzzy scale for each scenario may also be viewed as another approach to robust analysis in the evaluation process.

As for the reliability of results that stems from consistency of judgment, Saaty's CR can be used for the modal value ($\alpha = 1$) of the fuzzy assessment. This is a necessary condition to satisfy the consistency in the fuzzy pairwise comparison matrix (Gogus and Boucher, 1998). We believe that the applicability of evaluating cardinal consistency is rather limited to point estimates of priorities that were derived from judgments represented (at least in ratio scale) as exact numbers. In this study, the computed CR values from the modal value of the fuzzy comparison matrices were within the tolerable range of 0.00–0.15. However, we do wish to remind readers that this approach may have some limitations and alternative ways to measure that inconsistency should be pursued in future fuzzy ANP research (see related discussions in Gogus and Boucher (1998) and Leung and Cao, 2000).

6. Conclusions

In this paper, a fuzzy ANP approach was applied in the prioritization of remedial countermeasures. Fuzzy ANP could adequately handle the judgments derived from perception-based information which are intrinsically imprecise, reflecting the bounded ability of human mind to resolve detail and store information. Fuzzy numbers were used to 'fuzzify' the crisp scale as human judgment expressed in natural language is most often vague and fuzzy. This is the first attempt to extend the algorithm introduced by Cheng and Mon (1994) within the super-matrix framework. The main feature of the proposed method is the redefinition of fuzzy scale according to the degree of fuzziness, confidence level, and attitude toward fuzziness. This could provide another approach to sensitivity analysis in addressing the uncertainty attributed to the fuzziness of human judgment. The set of possible values of overall priorities derived at different δ , α and ω values may also provide additional information on the solution robustness of the prioritization method.

Using a numerical example which was drawn from a real-life case study of an uncontrolled landfill in Japan, the applicability of the proposed method as a decision aiding tool in CSM was demonstrated. This study does not pretend to present a comprehensive sensitivity analysis; but for the purpose of illustration, this paper examined the influence of changes in the value of the degree of fuzziness, as well as, the changes in the relative importance of the criteria on the outcome of the decision model. In this case, *A3* (capping, vertical cut-off wall, and in situ remediation of the uncontrolled landfill site and the surrounding area) is the most preferred alternative when implementability and financial affordability are the dominating criteria. However, a rank reversal may occur wherein *A2* (capping, pump and treatment method, and in situ remediation of the

uncontrolled landfill site and the surrounding area) becomes the most preferred countermeasure when the DM tends to overestimate the preference value of *A2* over the other alternatives with respect to financial affordability. In a scenario in which environmental effectiveness and social acceptability are the dominating criteria, *A1* (excavation of all wastes from the landfill site, off-site treatment and disposal) is the most preferred alternative followed by *A3*.

Fuzzy ANP can be viewed as an extension of ANP (i.e., a special case of fuzzy ANP) to accommodate explicitly the 'fuzziness' in the evaluation process. When compared with the conventional ANP, some additional advantages of fuzzy ANP can be identified. First, it could better model the vagueness and imprecision associated with the pairwise comparison process. Thus, it may be cognitively less demanding and more intuitive for the DMs. Second, it may allow an adequate reflection of the DMs' attitude toward 'fuzziness' and their degree of confidence in the subjective assessments. Fuzzy ANP, as being applied in the evaluation of remedial countermeasures, could improve the decision process through better learning, clarification and transparency. Hence, this study demonstrates the effectiveness of the fuzzy ANP to handle not only the complexity of the problem structure at hand, but also the inherent uncertainty associated with the subjectivity of human judgment. In addition, the proposed algorithm can be easily implemented in a spreadsheet and integrated in a decision support system. Future research would involve further investigation on the practicability of the proposed method as a decision support tool.

Acknowledgments

The authors would like to acknowledge the two anonymous reviewers for their valuable comments that improved the content of this paper.

References

- Ayag, Z., 2005. A fuzzy AHP-based simulation approach to concept evaluation in a NPD environment. *IEEE Transactions* 37, 827–842.
- Bardos, R.P., Mariotti, C., Marot, F., Sullivan, T., 2001. Framework for decision support used in contaminated land management in Europe and North America. In: NATO/CCMS. Evaluation of Demonstrated and Emerging Technologies for the Treatment of and Clean-up of Contaminated Land and Groundwater (Phase III) (special session: decision support tools No. 245). EPA 542-R-01-002. United States Environmental Protection Agency, Washington, DC, pp. 9–30.
- Boender, C., deGraan, J.G., Lootsma, F.A., 1989. Multi-criteria decision analysis with fuzzy pairwise comparisons. *Fuzzy Sets and Systems* 29, 133–143.
- Bryson, N., Mobolurin, A., Ngwenyama, O., 1995. *European Journal of Operational Research* 83, 639–654.
- Buckley, J.J., 1985. Fuzzy hierarchical analysis. *Fuzzy Sets and Systems* 17, 233–247.
- Buckley, J.J., Feuring, T., Hayashi, Y., 2001. Fuzzy hierarchical analysis revisited. *European Journal of Operational Research* 129, 48–64.
- Buyukozkan, G., Ertay, T., Kahraman, S., Ruan, D., 2004. Determining the importance weights for the design requirements in the House of

- Quality using the fuzzy analytic network approach. *International Journal of Intelligent Systems* 19, 443–461.
- Chang, D.-Y., 1996. Application of the extent analysis method on fuzzy AHP. *European Journal of Operational Research* 95 (3), 649–655.
- Cheng, C.-H., Mon, D.-L., 1994. Evaluating weapon system by analytical hierarchy process based on fuzzy scales. *Fuzzy Sets and Systems* 63, 1–10.
- Csutora, R., Buckley, J.J., 2001. Fuzzy hierarchical analysis: the Lambda-Max method. *Fuzzy Sets and Systems* 120 (2), 181–195.
- Deng, H., 1999. Multicriteria analysis with fuzzy pairwise comparison. *International Journal of Approximate Reasoning* 21, 215–231.
- Forman, E., Selly, M.A., 2001. *Decision by Objectives: How to Convince Others That You Are Right*. World Scientific, Singapore.
- Goddard, S.T., 1983. Ranking in tournaments and group decision making. *Management Sciences* 29 (12), 1384–1392.
- Gogus, O., Boucher, T., 1998. Strong transitivity, rationality and weak monotonicity in fuzzy pairwise comparisons. *Fuzzy Sets and Systems* 94, 133–144.
- Kahraman, C., Cebecia, U., Ruan, D., 2004. Multi-attribute comparison of catering service companies using fuzzy AHP: the case of Turkey. *International Journal of Production Economics* 87, 171–184.
- Kahraman, C., Ertay, T., Buyukozkan, G., 2006. A fuzzy optimization model for QFD planning process using analytic network approach. *European Journal of Operational Research* 171 (2), 390–411.
- Kaufmann, A., Gupta, M.M., 1985. *Introduction to Fuzzy Arithmetic: Theory and Applications*. Van Nostrand Reinhold, New York, USA.
- Kiker, G.A., Bridges, T.S., Varghese, A., Seager, T.P., Linkov, I., 2005. Application of multicriteria decision analysis in environmental decision making. *Integr Environ Assess Manag* 1 (2), 95–108.
- Kirkwood, C.W., Sarin, R.K., 1985. Ranking with partial information: a method and an application. *Operations Research* 33 (1), 34–48.
- Klir, G.J., Folger, T., 1988. *Fuzzy Sets, Uncertainty and Information*. Prentice-Hall, Englewood Cliffs, NJ.
- Kwong, C., Bai, H., 2002. A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function deployment. *Journal of Intelligent Manufacturing* 13, 367–377.
- Leung, L.C., Cao, D., 2000. On consistency and ranking of alternatives in fuzzy AHP. *European Journal of Operational Research* 124, 102–113.
- Linkov, I., Vargee, S., Jamil, S., Seager, T.P., Kiker, G., Bridges, T., 2004. Multi-criteria decision analysis: a framework for structuring remedial decisions at contaminated sites. In: Linkov, I., Ramadan, A. (Eds.), *Comparative Risk Assessment and Environmental Decision Making*. Kluwer Academic Publishers, Amsterdam, pp. 15–54.
- Mikhailov, L., 2003a. Deriving priorities from fuzzy pairwise comparison judgments. *Fuzzy Sets and Systems* 134 (3), 365–385.
- Mikhailov, L., Singh, M., 2003b. Fuzzy Analytic Network Process and its application to the development of decision support systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews* 33 (1), 33–41.
- Nikaido, H., 1968. *Convex Structures and Economic Theory*. Academic Press, New York, USA.
- Ngwenyama, O.K., Bryson, N., 1999. Eliciting and mapping qualitative preferences to numeric rankings in group decision making. *European Journal of Operational Research* 116, 487–497.
- Promentilla, M.A.B., Furuichi T., Ishii, K., Tanikawa, N., 2005a. An ANP-based decision analytic model for evaluation of remedial countermeasure strategies of contaminated sites. In: *Proceedings of 20th International Conference on Solid Waste Technology and Management*, Philadelphia, PA, USA.
- Promentilla, M.A.B., Furuichi T., Ishii, K., Tanikawa, N., 2005b. A multi-criteria decision analytic process for evaluation of remedial countermeasures of contaminated site: a case study of uncontrolled landfill. In: *Proceedings (Sardinia 2005) of Tenth International Waste Management and Landfill Symposium*, Cagliari, Italy.
- Promentilla, M.A.B., Furuichi, T., Ishii, K., Tanikawa, N., 2006a. Evaluation of remedial countermeasures using the analytic network process. *Waste Management* 26 (12), 1410–1421.
- Promentilla, M.A.B., Furuichi, T., Ishii, K., Tanikawa, N., 2006b. A prioritization method using hierarchical network model for multi-criteria evaluation of remedial countermeasures. *Journal of Environmental Systems and Engineering, JSCE* 62 (3), 308–324.
- Promentilla, M.A.B., Furuichi, T., Ishii, K., Tanikawa, N., 2006c. Evaluation of remedial alternatives using a prioritization method based on a hierarchical network model. In: *Proceedings of Second International Engineering Research Conference*, Cebu City, Philippines.
- Pollard, S.J.T., Brookes, A., Earl, N., Lowe, J., Kearney, T., Nathanail, C.P., 2004. Integrating decision tools for the sustainable management of land contamination. *Science of the Total Environment* 325, 15–28.
- Saaty, T.L., 1980. *The Analytic Hierarchy Process*. McGraw-Hill, Inc., New York, USA.
- Saaty, T.L., 1996. *Decision Making with Dependence and Feedback: The Analytic Network Process*, first ed. RWS Publications, Pittsburgh, USA.
- Saaty, T.L., 2001. *Decision Making with Dependence and Feedback: The Analytic Network Process*, second ed. RWS Publications, Pittsburgh, USA.
- Saaty, T.L., 2005. The analytic hierarchy and analytic network processes for the measurement of intangible criteria and for decision-making. In: Figueira, J., Greco, S., Ehrgott, M. (Eds.), *Multi-criteria Decision Analysis: State of the Art Surveys*. Springer International Series in Operations Research and Management Science, pp. 345–406.
- Srdjevic, B., 2005. Combining different prioritization methods in the analytic hierarchy process synthesis. *Computers & Operations Research* 32, 1897–1919.
- Van Laarhoven, P.J.M., Pedrycz, W., 1983. A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems* 11, 229–241.
- Zadeh, L., 1965. Fuzzy sets. *Information Control* 8, 338–353.
- Zimmermann, H.J., 2001. *Fuzzy Set Theory and its Applications*, fourth ed. Kluwer Academic Publishers, London.

Validation of a hybrid life-cycle inventory analysis method

Robert H. Crawford*

Faculty of Architecture, Building and Planning, The University of Melbourne, Parkville, Victoria 3010, Australia

Received 24 October 2006; received in revised form 2 March 2007; accepted 16 March 2007

Available online 7 May 2007

Abstract

The life-cycle inventory analysis step of a life-cycle assessment (LCA) may currently suffer from several limitations, mainly concerned with the use of incomplete and unreliable data sources and methods of assessment. Many past LCA studies have used traditional inventory analysis methods, namely process analysis and input–output analysis. More recently, hybrid inventory analysis methods have been developed, combining these two traditional methods in an attempt to minimise their limitations. In light of recent improvements, these hybrid methods need to be compared and validated, as these too have been considered to have several limitations. This paper evaluates a recently developed hybrid inventory analysis method which aims to improve the limitations of previous methods. It was found that the truncation associated with process analysis can be up to 87%, reflecting the considerable shortcomings in the quantity of process data currently available. Capital inputs were found to account for up to 22% of the total inputs to a particular product. These findings suggest that current best-practice methods are sufficiently accurate for most typical applications, but this is heavily dependant upon data quality and availability. The use of input–output data assists in improving the system boundary completeness of life-cycle inventories. However, the use of input–output analysis alone does not always provide an accurate model for replacing process data. Further improvements in the quantity of process data currently available are needed to increase the reliability of life-cycle inventories. © 2007 Elsevier Ltd. All rights reserved.

Keywords: Life-cycle assessment; Life-cycle inventory; Input–output analysis; Hybrid analysis; Embodied energy

1. Introduction

There are a range of methods available for assessing the environmental impacts of products. Whilst many of these methods have proved adequate to an extent for a particular purpose, they are characterised by many disadvantages. Life-cycle assessment (LCA) is one of the most popular tools for evaluating the environmental impact that a particular product or process has throughout its entire life. Although LCA is currently considered by many to be the most appropriate and reliable tool for assessing these environmental impacts, it can suffer from a number of limitations. The most significant of these limitations is the incompleteness and lack of reliability associated with current inventory data sources, both process and input–output (I–O) data. This stage of an LCA is referred to as life-cycle inventory (LCI) analysis, where input and output

flows are quantified. The limitations associated with LCI have a subsequent impact on the reliability of overall LCA findings.

A greater level of completeness and reliability in LCI is needed to permit a more accurate and precise assessment of the life-cycle environmental loadings from the manufacture of a particular product. A recently developed approach of assessing the life-cycle inputs to a product that attempts to fill in the ‘gaps’ of previous methods of LCI has been proposed (Suh, 2002; Treloar, 1997). This newly developed method of LCI, namely hybrid LCI, combines both process and I–O data. Hybrid LCI methods based on process data have shown to be incomplete. Hybrid LCI methods based on I–O data involve substituting available process data into an I–O model minimising the errors associated with process-based hybrid LCI methods. Recent improvements to this hybrid LCI method include the use of current I–O data and the inclusion of capital input data. However, until now, this LCI method has not yet been broadly tested for its ability to produce more reliable and more complete

*Tel.: +61 3 8627 6000; fax: +61 3 8627 6060.

E-mail address: rhcr@unimelb.edu.au.

results than traditional LCI methods based on process or I–O data alone.

The aim of this study was to assess the reliability and completeness of an I–O-based hybrid LCI analysis method.

2. Background

It is widely accepted that LCA is one of the best tools for environmental assessment of a wide range of products and processes. However, it is arguably best suited to primary or secondary industrial processes. LCA can be used to assess and evaluate the impacts that products or processes have on the environment over their entire life, from 'cradle to grave'. The LCI step is considered one of the most crucial steps involved in any LCA. Although LCI methods may be appropriate for assessing the environmental impacts of various products, there is a need to re-evaluate the adequacy of existing LCI methods in light of an increasing focus towards greenhouse gas emission reduction and the current improvements to hybrid LCI methods.

Inventory analysis involves the collection of data and calculations in order to quantify the inputs and outputs to the product system over its entire life-cycle. The inputs may include energy and natural resources, whilst outputs may include emissions to air, water and land (ISO 14041, 1999). The extent of this analysis, and the two subsequent steps of an LCA (impact assessment and interpretation), depends on the scope and goals defined in the first step of an LCA. The inventory analysis step is one of the most time and cost consuming processes in an LCA (Suh and Huppes, 2002). If the scope of the study is not adequately defined, excessive time may be wasted obtaining and analysing data that is beyond the scope required for the intended purpose of the study.

The accuracy and extent of an LCI is dependent on which of the main analysis methods is chosen: process analysis, I–O analysis or hybrid analysis (Treloar, 1997). The two base methods of LCI are susceptible to different types of errors and have different benefits. The most widely used of these methods, process analysis, can be significantly incomplete. This is primarily due to the complexity of the upstream requirements for goods and services (Lave et al., 1995). The magnitude of this incompleteness varies with the type of product or process and depth of study but can be 50% or more (Lenzen, 2001a; Treloar, 1998). While the accuracy of the process analysis method can be high, it is only relevant to the particular system considered and can be subject to considerable variability. These errors can be exacerbated as more and more process data are collected, due to the flawed paradigm.

Results of studies by Bullard et al. (1978), Lenzen and Treloar (2003), Miller and Blair (1985) and Peet and Baines (1986) have suggested that even extensive process-based inventories for complex products do not achieve sufficient system completeness. This is not as important in some comparative studies as the system completeness may be identical for both products or processes. However, when

the level of system completeness is not even able to be estimated, validation can be very difficult. This issue can have a much greater impact when considering the environmental impacts of a single product or process. In this case, I–O data may be useful for assessing the system boundary completeness.

The second base method of LCI is I–O analysis. This method uses national average data for each sector of the economy and is considered by many researchers to be more comprehensive than process analysis (e.g. *inter alia* Lave et al., 1995; Lenzen, 2000, 2001a; Leontief, 1966; Suh and Huppes, 2002; Treloar, 1997, 1998). This method has a systemically complete system boundary, which can therefore potentially solve the major drawback of the process analysis method. However, I–O analysis is generally used as a 'black box', with little understanding of the values being assumed in the model for each process.

Furthermore, because they are based on many inherent assumptions appropriate for national modelling, a perfect I–O model may not result in valid results for a particular product (Carnegie Mellon University, 2002; Suh, 2000). Conversely, imperfect I–O data may provide reasonable estimates for particular products. For this reason, the I–O model should be seen as a scoping tool or an estimation method for missing data from the process analysis method. While I–O analysis is systemically complete, some I–O models are inappropriately constructed, and may leave out significant aspects of the economy (for example, capital investment, Lenzen 2001a; Lenzen and Treloar, 2004). Some of the other main limitations of I–O analysis include: the age of the I–O data; homogeneity assumption; proportionality assumption; the conversion of economic data to energy data; the use of national averages; and sector classification and aggregation (Lenzen, 2001a; Miller and Blair, 1985).

Hybrid techniques attempt to combine the benefits of both base methods, while minimising their respective limitations. Process-based hybrid analysis (Bullard et al., 1978) is almost exclusively based on incomplete process data, suffering similar limitations to those outlined above for the two base LCI methods (Treloar, 1998). The I–O systemic completeness is only applied to the components of the model upstream from the process data. Downstream and horizontal truncation can still occur, to significant levels. To some extent, these errors can be compounded, despite the practitioner's best efforts to minimise them.

I–O-based hybrid analysis combines process data and I–O data in a different way to process-based hybrid analysis, in order to exclude downstream and horizontal truncation. The direct inputs to a specific product or process being studied are calculated using process analysis. While process data are not usually easy to obtain, its use maximises the reliability of the analysis at this stage. Further upstream indirect processes are accounted for by either further applications of process analysis or I–O analysis when the process data are unavailable or is

considered too time consuming to collect relative to the significance of the process in question (Treloar, 1997). The I–O model is disaggregated to allow the activities for which process data are available to be subtracted, leaving a remainder that can be applied to the study in a holistic manner to fill all the remaining gaps (as demonstrated in Treloar et al., 2001). The use of this method can help to minimise time by determining the inputs to a product that may not be significant, in order to focus on those that are. The I–O-based hybrid analysis method proposed by Treloar (1997) has been used in several life-cycle and embodied energy studies (Crawford, 2004; Crawford et al., 2003, 2006), which have demonstrated the possible significance of the choice of LCI methods. This I–O-based hybrid analysis method is currently preferred, but has yet to be evaluated against the base LCI methods and the process-based hybrid analysis method, particularly considering recent improvements to the method.

The inputs required through the manufacture of machinery and other capital equipment have traditionally been ignored in LCI. This is due to the difficulty in determining the time that the equipment was used in production for amortising these inputs (Alcorn, 1997) and in calculating the inputs embodied in products through the purchase and use of capital equipment. These inputs are as much a part of a products life-cycle as is any other direct or indirect input. Casler (1983) has demonstrated that the quantification of capital inputs can be easily achieved through the use of I–O data. Gorree et al. (2002) and Lenzen (2001b) have estimated capital inputs to account for between 10% and 17% of total inputs of an LCI. Unlike previous LCI methods, the I–O-based hybrid analysis method evaluated in this study considers these capital inputs, and hence provides a more comprehensive inventory assessment.

Whilst there are many problems and limitations associated with current LCI methods, particularly through the application of the methodological processes, the need to improve LCI methodology has been the focus of research worldwide. These improvements involve minimising the current problems with LCI in order to maximise the level of reliability and completeness of the results obtained, to enable a more accurate reflection of the environmental impacts of a particular product.

In the past, researchers have used a number of techniques to evaluate the various methods of LCI. These have included error analysis (various types, including truncation error analysis), gap analysis and comparative analysis (Lenzen, 2001a; Treloar, 1998). Error analysis is used to assess the accuracy associated with the use of I–O data for modelling system inputs (Lenzen, 2000). Truncation error analysis is used to assess the extent of incompleteness associated with the use of process data in a process-based hybrid analysis context (Bullard et al., 1978). Error analysis evaluates initial data inputs. However, often these are so complex that the sum of the effects of the errors in a national I–O model can be quite different

when applied to an individual product. This process can then introduce further errors, which are often ignored (e.g. Bullard et al., 1978).

Other methods have since been developed to overcome this limitation in error analysis by focusing on the outputs of the LCI methods. Gap analysis is used to assess the difference between process analysis results and hybrid analysis results, as an evaluation of the completeness of each method (Treloar, 1998). Comparative analysis is used to compare the I–O values for the process analysis components that are used in an I–O-based hybrid analysis as a measure of the reliability associated with the I–O data (Treloar, 1998). These last two methods are best used together to evaluate and compare LCI methods as applied to individual products, thus overcoming the limitations of error analysis (Bullard et al., 1978). These methods have been developed and demonstrated by Treloar (1998), however they have only been applied to a single residential building. This study by Treloar (1998) showed that there was both a large gap between process and hybrid analysis results (49%) and little correlation between I–O and process analysis values.

3. Method

In order to assess the I–O-based hybrid analysis method, to provide a more accurate representation of its completeness and reliability, the results from this LCI method need to be compared to the results from traditional and other hybrid LCI methods. Each method needs to be applied to a range of case studies, to ensure a range of LCI values are produced at various levels of detail. This level of analysis has not previously been performed. The application to a range of different case studies will improve the relevance of the study across the types of products likely to be evaluated in the future.

The significance of the production and consumption of energy towards global environmental impacts has led to more streamlined methods of LCA, such as life-cycle energy analysis (LCEA), being developed. Due to the significance of energy consumption from within the manufacturing and construction industries in particular, this assessment of the I–O-based hybrid analysis method is based on energy as the sole parameter, in particular, the embodied energy component of an LCEA.

The steps involved in the I–O-based hybrid analysis method are detailed below. For an I–O-based hybrid analysis, the steps for process analysis, I–O analysis and process-based hybrid analysis are also required (Fig. 1).

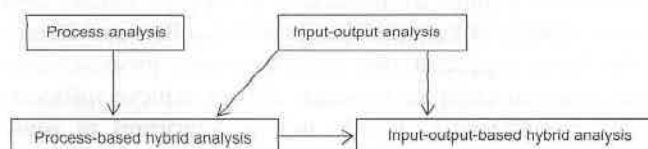


Fig. 1. Outline of I–O-based hybrid analysis framework.

3.1. Input–output-based hybrid analysis

An I–O-based hybrid analysis uses the data gathered from a traditional process analysis, and the result of a process-based hybrid analysis and increases the completeness of the LCI even further, with the use of I–O data. For this study, process data were derived based on the quantities of material inputs into the main products and a material energy intensity database (Grant, 2002).

National I–O tables, produced by the Australian Bureau of Statistics (ABS), using I–O data for Australia for the financial year 1996–1997 were combined with national energy data from the Australian Bureau of Agricultural and Resource Economics (ABARE) to develop an energy-based I–O model of the economy. A number of these models have been developed for Australia (e.g. Lenzen, 2000; Lenzen and Lundie, 2002; Treloar, 1997). The I–O tables were divided into 106 sectors of the Australian economy (for example: ‘household appliances’, ‘road transport’, ‘residential construction’). Each one of these economic sectors has a respective direct energy intensity and total energy intensity, both quantified in GJ/\$1000 of product, representing the amount of energy used directly and in total to produce \$1000 worth of products from that specific sector. It is therefore necessary to determine to which sector the product being studied belongs in order to determine the energy intensities to be applied to that product. The input–output theory underlying these calculations is extensively documented elsewhere (Lenzen, 2001b; Leontief, 1966; Miller and Blair, 1985; Proops, 1977) and shall therefore not be repeated here. The retail price of the product was obtained from the supplier of the product, or if this was unavailable, for example for buildings, an estimate was made based on literature and/or necessary assumptions.

The delivered quantities of basic materials obtained through the process analysis were then used as the basis for a process-based hybrid analysis. For this, a number of hybrid material energy intensity figures, combining both process and I–O data, were derived. A hybrid energy intensity figure was calculated for all of the most common basic materials, such as steel, timber and glass. These figures are expressed in GJ/unit (usually t, kg, m² or m³) of material and represent a simplified method of incorporating process data into the analysis, giving the amount of energy embodied in, for example, a kilogram of that material:

$$EI_M = PEI_M + (TEI_n - TEI_M) \times \frac{S_M}{1000}, \quad (1)$$

where EI_M is the hybrid energy intensity of the basic material; PEI_M the material process energy intensity; TEI_n the total energy intensity of the I–O sector (n) to which the basic material belongs; TEI_M the total energy intensity of the I–O path representing the basic material; and S_M the total price of the basic material.

Once the hybrid material energy intensities had been calculated, they were multiplied by the delivered quantities of basic materials of the product. These individual material embodied energy figures were then summed to obtain the process-based hybrid embodied energy value for the product.

The next step was to extract the inputs from the relevant sector of the economy from which the product belongs, by using an algorithm developed by Treloar (1997) and the energy-based model based on 1996–1997 I–O data. Then, from the inputs extracted, the inputs that have been counted in the process analysis inventory are identified. The total energy intensity of each of the inputs represented in the process analysis inventory is subtracted from the total energy intensity of the sector. If a process analysis value is available then the relevant input from the input extraction must be subtracted from the total energy intensity of the sector to avoid double counting. The remainder of the unmodified inputs (the total energy intensity of the sector minus those inputs subtracted, in GJ/\$1000) are then multiplied by the price of the product (\$) and divided by 1000 to give the additional energy inputs for the product, in GJ. The process-based hybrid analysis value is then added to this figure, minus the direct energy component (as this is included in the remainder of unmodified inputs), to give the I–O-based hybrid embodied energy total:

$$EE_t = Q_M \times W \times EI_M + (TEI_n - TEI_M) \times \frac{S_{BP}}{1000}, \quad (2)$$

where EE_t is the total embodied energy through I–O-based hybrid analysis; Q_M the quantity of materials in the basic product; W the wastage multiplier of the respective material; EI_M the material hybrid energy intensity; TEI_n the total energy intensity of the I–O sector (n) to which the product being assessed belongs; TEI_M the total energy intensity of the I–O path representing the basic material; and S_{BP} the total price of the basic product.

Table 1 provides an example of the I–O-based hybrid LCI assessment method.

Table 1
Example of an I–O-based hybrid analysis

Process	Embodied energy (GJ)
Process data for quantified materials	15.99
Input–output data used to fill upstream gaps	14.31
Process-based hybrid analysis total	30.30 (a)
<i>Input–output data for specific product sector</i>	
Direct energy intensity	0.51 (b)
Total energy intensity	64.10 (c)
Inputs covered by process data	24.27 (d)
Remainder (c–b–d)	39.33 (e)
Total (a + e)	69.63
Proportion of process data	23%

N.B. Figures may not sum due to rounding.

Traditionally, the consideration of capital inputs, such as the machinery used for the manufacture of materials, has been ignored. Even some of the most recent and comprehensive LCI methods and studies have neglected to include these capital inputs, mainly due to difficulties involved in its collection. Lenzen (2001b) and Lenzen and Treloar (2004) have shown that these capital inputs may be quite significant. Recent improvements to the hybrid LCI method evaluated in this study include the consideration of these capital inputs and the use of a model based on 1996–1997 I–O data, providing a more comprehensive and complex system boundary than ever before.

3.2. Evaluation method

Many previous evaluation methods have been appropriate for evaluating the results from traditional LCI methods, such as I–O analysis or process analysis. However, a method is needed to evaluate an LCI method that combines both process and I–O data; in this case, I–O-based hybrid analysis. Gap analysis and comparative analysis are considered most appropriate for this evaluation of the I–O-based hybrid analysis method as they focus on the evaluation of final LCI results, rather than initial data inputs, minimising the errors and limitations associated with similar previous evaluations. The use of two methods for evaluating the I–O-based hybrid analysis method is necessary since currently no single method exists for evaluating both completeness and reliability of hybrid LCI methods. Whilst gap analysis is capable of assessing the gap between the process analysis and I–O-based hybrid analysis results, to measure the completeness of the process analysis results, comparative analysis is required to evaluate the reliability of the I–O data used in the I–O-based hybrid analysis. The combination of these two evaluation methods provides a more thorough evaluation of the appropriateness of using the I–O-based hybrid analysis method.

The following provides a description of the detailed evaluation methods applied to the LCI results of a range of products, components and materials to provide a more detailed assessment of the use of the I–O-based hybrid analysis method. The evaluation methods that have been used for the evaluation of the LCI methods in this study include:

- gap analysis and
- comparative analysis.

Gap analysis has been used to assess the difference between the process analysis values and the I–O-based hybrid analysis values for each of the case studies, as a measure of completeness with respect to the I–O system boundary. This is done by subtracting the equivalent process analysis value from the I–O-based hybrid analysis value. The purpose of this is to show the embodied energy added through the use of I–O data, through the application of the

more comprehensive I–O-based hybrid analysis method. This gap can also be expressed as the percentage of completeness of the process analysis value when compared to the respective I–O-based hybrid analysis value:

$$\text{GAP} = \frac{\text{IOHA} - \text{PA}}{\text{IOHA}} \times 100,$$

where PA is the embodied energy of the main product through process analysis and IOHA the embodied energy of the main product through I–O-based hybrid analysis.

Comparative analysis is used to evaluate the process analysis and I–O analysis values for the whole product, individual components and materials, as a measure of reliability. The I–O analysis values are those extracted from the I–O model which are, through the process of the I–O-based hybrid analysis, substituted with process data. The process analysis values are therefore those substituted in place of the I–O figures. These two comparable figures are then evaluated to compare the I–O values against the process analysis values.

3.3. Application of evaluation methods to case studies

In order to evaluate the reliability and completeness of the I–O-based hybrid analysis method, the evaluation techniques are applied to a range of different components, materials, systems and products and at different levels of detail. Due to the significant environmental concerns and complexity associated with the construction and operation of buildings (Treloar, 1998), the items selected were those related directly to buildings. A number of case studies were identified, all of which required a reasonable amount of process data quantities to be available to enable a comparison to the equivalent I–O values.

Once the I–O-based hybrid analysis was applied to each case study and the results of the LCI determined, the evaluation methods were applied. The results obtained from these evaluations were then used to assess the completeness and reliability associated with the I–O-based hybrid analysis method.

3.4. Description of case studies

The case studies selected to test the reliability and completeness of the I–O-based hybrid analysis method include two commercial office buildings, a residential building, a Velodrome (all located in Melbourne, Australia) and a number of building related products including: a solar hot water system, building integrated photovoltaic (BiPV) system and a washing machine.

The first two case studies are two commercial office buildings. The gross floor area (GFA) of the first of these buildings is 47 000 m² with 8 and 15 storeys in the front and rear halves, respectively. The building construction type is of a reinforced concrete structure, clad in precast concrete, glazed panels, aluminium framed curtain walling and granite veneer panels. The roof consists of insulated steel

decking, and internal partitions consist of plasterboard clad steel framed and concrete walls and glazed screens. The building services were also included in the analysis.

The second commercial office building has a GFA of 11 588 m² over three storeys and consists of a 200 mm thick reinforced concrete slab substructure; steel columns; steel roof frame and decking; aluminium framed toughened glass glazed roof; precast concrete and aluminium cladding panels for external walls; frameless toughened glass solar double glazed windows; aluminium framed toughened glass curtain walls, louvers, screens and windows; and internal walls of single skin brickwork, 10 mm plasterboard clad steel and 200 mm thick autoclaved aerated concrete blocks.

The third case study is a single storey detached brick veneer house with a GFA of 109 m². The substructure of the building consists of a timber sub-floor on concrete stumps and the external walls are of brick veneer construction. The structural framing is a traditional timber frame clad with plasterboard internally and the roof is clad with concrete tiles. The ceiling is insulated with R2.5 fibreglass batts, and the walls and roof have reflective foil. The windows are single glazed and timber framed. The appliances and external structures, including a timber deck and pergola and concrete porch, were also included in the analysis.

The fourth case study is the Velodrome used for the 2006 Melbourne Commonwealth Games and consists of a two-storey building, with a GFA of 8947 m², which also includes change rooms, function rooms and spectator seating. The construction of the building incorporates concrete slabs and walls; steel trusses and framing; metal clad roof and external walls; and aluminium windows.

Case study five is an electric-booster solar hot water system. The system used for the study is a closed circuit system, comprising of a storage capacity of 300 L and a collector area of 3.96 m². The storage tank consists of a steel container with steel end domes; vitreous enamel lining; magnesium anode; polyurethane foam insulation; steel jacket; aluminium outer casing; and polypropylene casing ends. The collectors consist of a steel absorber plate with a black polyester powder coat finish, surrounded by polyester insulation, cased in aluminium, with a tempered glass top surface and aluminium glazing angles.

The sixth case study is a BiPV system. A BiPV system is commonly used in buildings to generate electrical energy through the conversion of solar energy, offsetting the energy consumed in the everyday operation of a building. This system consists of two mono-crystalline silicon (c:Si) 75 W photovoltaic modules in aluminium frames with a total area of 1.26 m². Each module contains silicon cells laminated between sheets of ethylene vinyl acetate (EVA) and high-transmissivity low-iron 3 mm tempered glass, fixed to a number of timber rafters as part of a typical residential roofing system. An advantage of this type of integrated installation is the embodied energy credit from

the displacement of conventional roofing materials, as accounted for in this study.

Case study seven is a typical household clothes washing machine. The washing machine used in the study consists of a large number of parts, including metal panels and frame; switches; control units; hoses; fixings; pump; motors; pulleys; and packaging. These parts are made from a number of main materials including steel, plastic, aluminium, copper and cardboard.

These seven case studies have been summarised in Tables 2 and 3.

The quantities of materials used in each of the case studies were determined. Information regarding components, materials, masses, areas and volumes was obtained from various public domain sources. These usually included the architectural plans, specifications, bill of quantities, manufacturers of the various products or through assumptions where information was unavailable or unknown.

4. Results and discussion

The LCI results of each of the case studies, based on each of the LCI methods, are shown below (Tables 4 and 5). These results show both the process analysis and I–O analysis proportions for the I–O-based hybrid analysis values.

A large range of values is evident for the buildings studied (Table 4). These results range from 6.4 GJ/m² for the residential building, using I–O analysis, to 29.4 GJ/m² for the first commercial building using I–O-based hybrid analysis.

Table 2
Building case study details

Case study	No. of storeys	GFA (m ²)	Construction type
Commercial building (1)	8/15	47 000	Concrete frame and precast concrete
Commercial building (2)	3	11 588	Steel frame and precast concrete
Residential building (3)	1	109	Brick veneer
Velodrome (4)	2	8947	Steel frame and precast concrete

Table 3
Building product case study details

Case study	Price (\$)	Size	Description
Solar hot water system (5)	2986	300 L tank, 3.96 m ² collector	Electric boosted, two collectors
Building integrated PVs (6)	1560	1.26 m ² (2 × 0.63 m ²)	75 W c:Si modules
Washing machine (7)	1050	5 kg	Typical model

Table 4
Embodied energy of building case studies for each LCI method

Case study	PA (a) (GJ/m ²)	I-OA (GJ/m ²)	PHA (GJ/m ²)	I-O-based hybrid analysis		
				Total (b) (GJ/m ²)	% PA value	% I-O value (gap)
1	10.1	16.0	19.4	29.4	34	66
2	8.0	12.5	17.8	25.8	31	69
3	6.9	6.4	14.4	17.0	41	59
4	8.1	11.0	16.0	21.8	37	63

N.B. 'gap' = $(b-a)/b \times 100$, PA = process analysis; I-OA = input-output analysis; PHA = process-based hybrid analysis.

Table 5
Embodied energy of building product case studies for each LCI method

Case study	PA (a) (GJ)	I-OA (GJ)	PHA (GJ)	I-O-based hybrid analysis		
				Total (b) (GJ)	% PA value	% I-O value (gap)
5	16	64	30	69	23	77
6	7.9	29	8.0	29	28	72
7	2.4	23	6.0	19	13	87

N.B. 'gap' = $(b-a)/b \times 100$.

For all of the case studies, other than the residential building, the process analysis results were lower than the respective I-O analysis results. This reflects the views of Lenzen and Dey (2000) who found that process analysis results are consistently lower than results calculated using I-O analysis. The exception to this (residential building) may be due to any or all of the following:

- an underestimation of the price of the building as used in the I-O analysis;
- a considerably large proportion of process data being available;
- other limitations associated with I-O analysis.

The process-based hybrid analysis figures show the effect of one method of combining process data and I-O data. All of the building case study process-based hybrid analysis values are moderately higher than the respective I-O analysis values. The reason for this may be due to the possible underestimation of the price of the buildings used in the calculation of the I-O analysis results. This may mean that I-O analysis provides a significant underestimation of the embodied energy of the buildings studied. Therefore, the results of the gap evaluation are also likely to be conservative. As expected, all of the process-based hybrid analysis values are also higher than the process analysis values due to the use of I-O data to fill some of the gaps in the process analysis method.

The I-O-based hybrid analysis results show a significant increase in the total values for the building case studies over the process-based hybrid analysis results (Table 4). The percentage of the total I-O-based hybrid values made up of I-O data for each building case study ranges from 59% to 69% with an average across these case studies of

~64%. These figures represent the gap in traditional process analysis.

A large range of values is evident for the building products studied. These values range from 2.4 GJ for the washing machine using the process analysis method to 69 GJ for the solar hot water system using the I-O-based hybrid analysis method (Table 5).

Table 5 shows that the process analysis values of each case study are considerably lower than the values calculated through the I-O analysis. For example, the process analysis value of the washing machine is 2.4 GJ, compared to the I-O analysis value of 23 GJ. This may be due to only a small amount of process data being available for these case studies or the price for the individual products being higher than the I-O sector average. This result also reflects the incompleteness associated with process analysis.

The process-based hybrid analysis values for the building product case studies are significantly lower than the respective I-O analysis values (Table 5). This may be due to these products being complex and/or a small amount of process data being available. However, as was expected, they are higher than the respective process analysis results. This is due to the inclusion of I-O data to fill the gaps in the process analysis results. This contrasts with the building case study process-based hybrid analysis values, which are significantly higher than the respective I-O analysis values (Table 4).

The I-O-based hybrid analysis results showed a significant increase in the total values over the process analysis and process-based hybrid analysis results for the building case studies (Table 4). This was to be expected due to the use of I-O data to fill the gaps in these process-based LCI methods. However, for the building product case

studies, these increases are even more significant, with the I–O-based hybrid analysis values more than double the equivalent process-based hybrid analysis values (Table 5). The main reason for this may be due to the building products being more complex, or only a small amount of process data being available. These I–O-based hybrid analysis values are almost identical to the I–O analysis values, showing that the process data collected may not have had a significant impact on the final results. This may also be partly due to the combined errors associated with the data used. This indicates that there may be a lack of available process data for these products. This is also reflected in the proportion of I–O data in the I–O-based hybrid analysis values for each building product case study, ranging from 72% to 87% with an average across these case studies of 79%. These values reflect the considerable gap in traditional process analysis.

4.1. Gap analysis

The gap between the process analysis value and I–O-based hybrid analysis value of the building and building product case studies ranges from 59% (residential building), and can be up to 87% (washing machine) (up from 80% in previous assessments, showing the increased complexity of the improved model) (Tables 4 and 5), with an average gap of 70%. This evaluation has shown that traditional process analysis suffers from a large truncation error, and that I–O-based hybrid analysis, through its increased complexity, provides a more comprehensive evaluation of energy inputs to buildings and building products. The truncation errors associated with process analysis have been identified by Lenzen (2000, 2001a) to be in the range of 50%. The results obtained from this study support Lenzen, and show that this gap is even greater for certain buildings and building products, much greater than those previously reported (Lave et al., 1995; Lenzen, 2000, 2001a; Lenzen and Treloar, 2002, 2003).

What was found to be significant in the evaluation of the gap between process analysis and I–O-based hybrid analysis results was that the gap for all of the building case studies was less than the gap for any of the building product case studies. The reasons for this may include any or all of the following:

- material inputs differing between buildings and building products;
- variations in the quality and quantity of process data available;
- difference in I–O total energy intensities and complexity for building sectors ('residential building' and 'other construction') versus building product sectors.

The gap between the process-based hybrid analysis and I–O-based hybrid analysis results ranged from 15% to 34% for the building case studies and 53% to 72% for the building product case studies. The reason for the consider-

able gap in process analysis and process-based hybrid analysis results is due to the truncation errors associated with process data (Lenzen, 2000, 2001a). The significance of this gap shows that even process-based hybrid analysis does not always provide a comprehensive assessment of the LCI of particular products. A breakdown of the gap between process analysis and I–O-based hybrid analysis results of the material inputs to each product was performed to provide a more detailed and less aggregated evaluation than on a whole product basis. However, on this more detailed level, the gap between traditional process analysis and I–O-based hybrid analysis results remains significant.

4.2. Comparative analysis

The results of the comparative analysis for all of the case studies are shown in Fig. 2. A logarithmic scale was used to avoid smaller values being lost for the sake of visual comparison. The initial I–O values of material inputs and whole products for each case study are compared to the modified process analysis values to determine the validity of the I–O values, against the more reliable process analysis values, in particular for use in accounting for those inputs which a typical process analysis does not consider. The I–O values for which equivalent process data were not available could not be compared. However, if the equivalent I–O values of the inputs for which process data have actually been collected are compared, then the results of this comparison could apply to those I–O values used to complete the system boundary in an I–O-based hybrid analysis. The intent here is to indicate that the national average I–O data are not always a perfect model for the process data. A strong correlation is not required to enable the use of I–O data to fill the gaps in traditional process

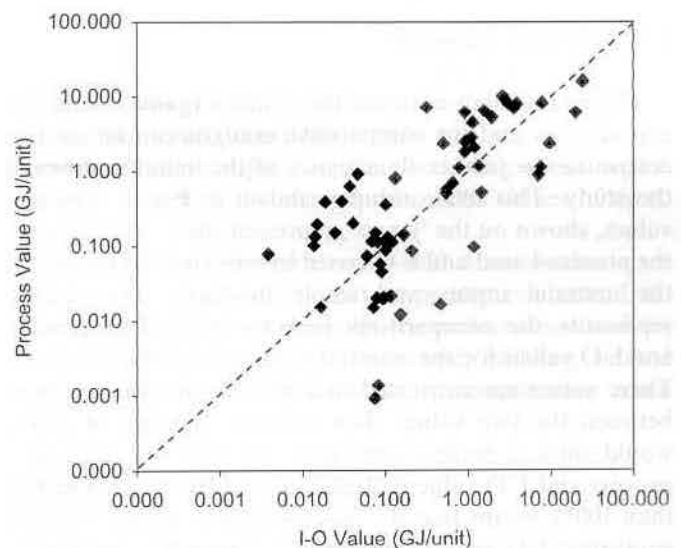


Fig. 2. Comparison of I–O values and process values for all case studies, by product and material.

analysis, although it would be beneficial in terms of validating the reliability of the I–O data.

With the relatively even distribution of values both above and below the bottom-left to top-right diagonal (59% and 41% of products and materials, respectively), the initial I–O analysis values are shown to be fairly conservative. The I–O values used to fill gaps in the hybrid model, when no process data are available, are therefore also likely to be conservative. While this comparison shows a strong relationship for a very small number of materials and products (~17%), there is very little correlation for the overwhelming majority of the comparisons (average correlation of 0.67), well below what is considered to be a strong correlation (>0.9).

The comparative analysis between I–O and process data showed that the values provided by I–O data do not give an accurate indication or representation of the equivalent process values, on a material input or whole product level. Therefore, the use of I–O data to account for that process data which is typically not collected in an LCI may not be appropriate. However, the gap in traditional process analysis has been shown to be too large to ignore and thus, even with the current limitations associated with I–O data, its use may be more beneficial than not using it at all. This current study has shown the lack of close correlation between process and I–O values across a broad range of building types and building products. However, due to the truncation errors associated with process data, resulting in large gaps between these values and I–O-based hybrid analysis values, this is more likely to be due to the limitations of current process data in Australia, particularly the insufficient quantity and possibly quality of such data. With this in mind, the aim must be to collect more process data before a more sophisticated evaluation of the I–O-based hybrid analysis method can be made. The quality of the available Australian process data can then be evaluated, allowing for further evaluation of LCI methods to occur.

4.3. Combined evaluation

The relationship between the values obtained from the gap analysis and the comparative analysis can be used to determine the further significance of the initial findings of the study. This relationship is shown in Fig. 3. The gap values, shown on the 'y-axis', represent the % gap between the process-based and I–O-based hybrid analysis results for the material inputs and whole products. The 'x-axis' represents the comparisons between equivalent process and I–O values for the material inputs and whole products. These values are converted to a percentage of correlation between the two values. For example, a value of 100% would show a perfect correlation between the equivalent process and I–O value (correlation of 1.0). A value higher than 100% means that the process value is higher than the equivalent I–O value. Therefore, a value of 0% correlation would mean that no process data were available and that the gap is 100%.

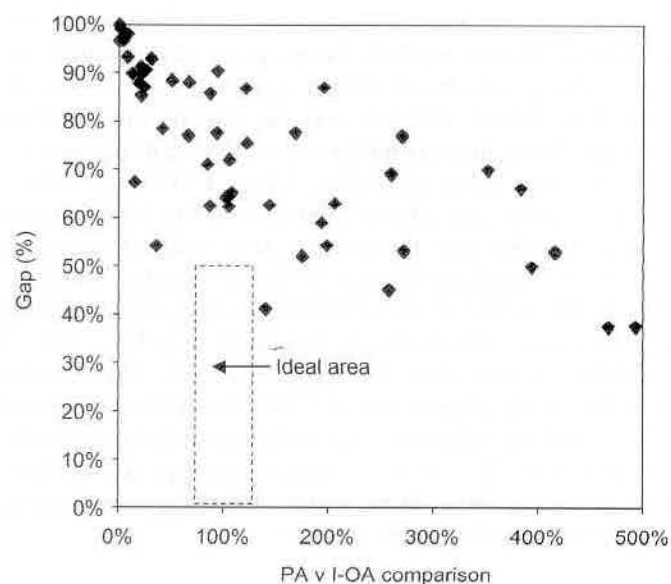


Fig. 3. Relationship of gap and process versus I–O value comparisons, by material inputs and whole products.

Ideally, for I–O data to be shown to provide an accurate representation of equivalent material input and whole product process data values and for achieving a reliable and comprehensive LCI, there would be a relatively low gap (closer to 0% on the y-axis, and no more than 50%) and a close correlation between process and I–O values (close to 100%). This ideal situation is shown by the area surrounded by the dotted line in Fig. 3. All of the points in the graph lie outside this ideal area. Thus, it is again shown that there is a weak correlation between equivalent process and I–O values and that there are large truncation errors associated with currently available process data (as shown by the large gap). Due to this, it appears that I–O data do not provide an accurate representation of the equivalent process data for material inputs and whole products. The significant gap for all of the inputs shows the need for improvements in the completeness of process-based LCI methods.

Due to the state of currently available Australian data, this level of evaluation of LCI results is considered to be sufficient as the errors and limitations of these current LCI data sources and methods restrict any more sophisticated evaluation from being relevant or necessary. If it were shown that a majority of these points had a low gap and strong correlation, then it may be possible that the current evaluation methods are not sophisticated enough in view of current levels of available process data.

The relationship between the gap and the comparison of I–O and process values showed that although there were a number of points with a fairly low gap and a higher process value than I–O value, the majority of points showed a high gap and a lower process value than I–O value. Typically, the gap between the process analysis and I–O-based hybrid analysis results ranged from 35% upwards, with the higher proportion of these above 50% and even up to 100% in

some cases. In all cases this gap is significant, while the exact degree of this gap may be dependant on a number of factors including:

- quantity and quality of process data available;
- variation in material inputs;
- energy intensity of material inputs;
- energy intensity of I–O sectors used.

For example, a particular material will have a certain I–O total energy intensity, while the process-based energy intensity of this material may be greater than the national average due to a particular manufacturing process or location.

As the evaluation between the gap and comparison between process and I–O values has shown that none of the points have a low gap and strong correlation, it can be assumed that the quantity of currently available process data are not sufficient for the LCI of buildings and building products. However, as for I–O data, having some process data is better than having none at all, and may be adequate until more process data are collected.

4.4. Capital inputs

Whilst the I–O-based hybrid analysis method used in this study has been in existence for a number of years, the inclusion of capital inputs (within any method for that matter) is a new development. Previous studies based on the case studies used in the current study (Crawford, 2004), using an I–O model based on data exclusive of capital equipment inputs showed lower results than those calculated in this current study. Table 6 shows the I–O-based hybrid analysis results of these previous studies and based on these results, the percentage that these capital inputs contribute to the current results.

It has been suggested in the past that capital inputs may account for up to 10–17% of the total embodied energy of any particular product (Gorree et al., 2002; Lenzen, 2001b). This study has shown that these capital inputs may account for up to 22% of a product's embodied energy, with an average across the buildings and building products looked at in this study, of 15%. This value is

reflected by the views of Gorree et al. (2002) and Lenzen (2001b). As capital inputs are rarely accounted for in LCI, the exclusion of these inputs is one reason for the incompleteness associated with past LCI studies. The inclusion of capital equipment inputs in the I–O model has shown to increase the gap between I–O-based hybrid analysis and process analysis results even further.

5. Conclusions

This paper has assessed the reliability and completeness of an I–O-based hybrid LCI method. Traditional LCI methods, I–O analysis, process analysis and process-based hybrid analysis, were used in this evaluation and applied across a range of building types and products.

It was found that current levels of available process data are insufficient. This is demonstrated by gaps of up to 87% shown for process analysis methods. The use of I–O data to fill the gaps associated with traditional process analysis methods was evaluated. The comparison between equivalent process and I–O values showed that I–O data do not provide a reliable representation of the equivalent process values for material inputs and whole products. Therefore, the use of I–O data to account for inadequate or missing process data is not reliable. However, as there is currently no other method for filling the gaps in traditional process analysis, and as I–O data are considered to be more complete than process data and the errors possibly somewhat lower, using I–O data to fill the gaps in traditional process analysis appears to be better than not using any data at all. This appears to be of even greater importance for more complex products, such as building related products. The main advantage of the I–O-based hybrid LCI method is that it improves the system boundary completeness. However, with the errors still evident in process and I–O data, when conducting an LCI using either form of data, it is essential that the potential inaccuracies associated with the LCI, as a result of these errors, are stated.

Whilst previous LCI studies have neglected to account for capital energy inputs, even those based on comprehensive hybrid approaches, this study has shown the significance of these capital inputs. Excluding these capital inputs can result in underestimates in LCI values of up to 22%.

This study showed that current evaluation methods are sufficiently accurate for the data typically available. I–O-based hybrid analysis is currently the preferred method for the LCI of Australian buildings and building related products due to its superior level of completeness when compared to traditional methods.

6. Further research

This study has shown that there is a significant lack of process data currently available in Australia and further improvements in the quantity of this data are needed to

Table 6
Previous case study results and percentage of capital component in current results

Case study	Previous IOHA results (GJ/unit)	% Capital in current results
1	27.2	7.5
2	23.6	8.5
3	13.4	21
4	19.7	10
5	54	22
6	25	14
7	15	21

increase the reliability of life-cycle inventories. Subsequent re-validation of LCI methods may be necessary in the light of expected progressive improvements in data availability.

If future comparisons between equivalent process and I–O values were to show a close correlation, then it may be necessary to re-evaluate whether I–O data have become a reliable indicator of missing or inadequate process data, and therefore whether any further collection of process data is required.

Acknowledgements

The author would like to acknowledge the assistance of Assoc. Prof. Graham Treloar (The University of Melbourne) in reviewing and providing invaluable comments on this paper. This work has been supported by an Australian Research Council (ARC) Linkage grant, LP0667653 with industry partner Williams Boag Architects, Melbourne.

References

- Alcorn, A., 1997. Recent Work in Embodied Energy Analysis of New Zealand Building Materials. In: *Proceedings: ANZAScA 1997*. Brisbane, Australia, pp. 91–98.
- Bullard, C.W., Penner, P.S., Pilati, D.A., 1978. Net energy analysis: handbook for combining process and input–output analysis. *Resources and Energy* 1, 267–313.
- Carnegie Mellon University 2002. Economic input–output life-cycle assessment <<http://www.eiolca.net/>> (last accessed 5th November).
- Casler, S., 1983. Correcting input–output coefficients for capital depreciation. *Energy Systems and Policy* 7 (3), 171–193.
- Crawford, R.H., 2004. Using input–output data in life-cycle inventory analysis. Ph.D. Thesis, Deakin University, Geelong.
- Crawford, R.H., Treloar, G.J., Ilozor, B.D., Love, P.E.D., 2003. Comparative greenhouse emissions analysis of domestic solar hot water systems. *Building Research and Information* 31 (1), 34–47.
- Crawford, R.H., Treloar, G.J., Fuller, R.J., Bazilian, M., 2006. Life-cycle energy analysis of building integrated photovoltaic systems (BiPVs) with heat recovery unit. *Renewable and Sustainable Energy Reviews* 10 (6), 559–575.
- Gorree, M., Guinee, J.B., Huppes, G., van Oers, L., 2002. Environmental life-cycle assessment of linoleum. *International Journal of Life-Cycle Assessment* 7 (3), 158–166.
- Grant, T., 2002. Australian Material Inventory Database of Life-Cycle Assessment Value for Materials. RMIT, Melbourne.
- ISO 14041, 1999. Environmental Management—Life-Cycle Assessment—Goal and Scope Definition and Inventory Analysis, Standards Australia, Australia.
- Lave, L.B., Cobas-Flores, E., Hendrickson, C.T., McMichael, F., 1995. Life-cycle assessment: using input–output analysis to estimate economy-wide discharges. *Environmental Science and Technology* 29 (9), 420A–426A.
- Lenzen, M., 2000. Life-cycle assessment of Australian transport. In: *Proceedings of the Second National Conference on Life-Cycle Assessment*. Melbourne.
- Lenzen, M., 2001a. Errors in conventional and input–output-based life-cycle inventories. *Journal of Industrial Ecology* 4 (4), 127–148.
- Lenzen, M., 2001b. A generalized input–output multiplier calculus for Australia. *Economic Systems Research* 13 (1), 65–92.
- Lenzen, M., Dey, C.J., 2000. Truncation error in embodied energy analyses of basic iron and steel products. *Energy* 25, 577–585.
- Lenzen, M., Lundie, S., 2002. Input–output model of the Australian economy based on published 1996–97 Australian input–output data.
- Lenzen, M., Treloar, G.J., 2002. Energy embodied in buildings: wood versus concrete. *Energy Policy* 30 (3), 249–255.
- Lenzen, M., Treloar, G.J., 2003. Differential convergence of life-cycle inventories toward upstream production layers: implications for life-cycle assessment. *Journal of Industrial Ecology* 6 (3–4), 137–160.
- Lenzen, M., Treloar, G.J., 2004. Endogenising capital—a comparison of two methods. *Journal of Applied Input–Output Analysis* 10, 1–11.
- Leontief, W., 1966. *Input–Output Economics*. Oxford University Press, New York, USA.
- Miller, R.E., Blair, P.D., 1985. *Input–Output Analysis—Foundations and Extensions*. Prentice-Hall, Englewood Cliffs, NJ.
- Peet, N.J., Baines, J.T., 1986. Energy analysis: a review of theory and applications. Report No.126, New Zealand Energy Research and Development Committee, Auckland, 56pp.
- Proops, J.L.R., 1977. Input–output analysis and energy intensities: a comparison of methodologies. *Applied Mathematical Modelling* 1, 181–186.
- Suh, S., 2000. Missing inventory estimation tool (MIET). <http://www.leidenuniv.nl/cml/_ssp/software/> (last accessed 5th November).
- Suh, S., 2002. The hybrid approach merging IO and process LCA. In: *Proceedings of the 16th Discussion Forum on Life-Cycle Assessment*. Lausanne, 9th April, 21pp.
- Suh, S., Huppes, G., 2002. Missing inventory estimation tool using input–output analysis. *International Journal of Life-Cycle Assessment* 7 (3), 134–140.
- Treloar, G.J., 1997. Extracting embodied energy paths from input–output tables: towards an input–output-based hybrid energy analysis method. *Economic Systems Research* 9 (4), 375–391.
- Treloar, G.J., 1998. A comprehensive embodied energy analysis framework. Ph.D. Thesis, Deakin University, Geelong.
- Treloar, G.J., Love, P.E.D., Holt, G.D., 2001. Using national input–output data for embodied energy analysis of individual residential buildings. *Construction Management and Economics* 19, 49–61.



ELSEVIER

Journal of Environmental Management 88 (2008) 507–516

Journal of
Environmental
Management

www.elsevier.com/locate/jenvman

A heuristic optimization approach for Air Quality Monitoring Network design with the simultaneous consideration of multiple pollutants

A. Elkamel^{a,*}, E. Fatehifar^b, M. Taheri^c, M.S. Al-Rashidi^d, A. Lohi^{a,e}

^aChemical Engineering Department, School of Engineering, University of Waterloo, Waterloo, Ont., Canada

^bEnvironmental Engineering Research Center, Faculty of Chemical Engineering, Sahand University of Technology, Sahand New Town, Tabriz, Iran

^cDepartment of Chemical Engineering, School of Engineering, Shiraz University, Shiraz, Iran

^dChemical Engineering Department, Loughborough University, Leicestershire LE11 3TU, UK

^eChemical Engineering Department, Faculty of Engineering and Applied Science, University of Ryerson, Toronto, Ont., Canada

Received 3 September 2006; received in revised form 25 February 2007; accepted 16 March 2007

Available online 9 May 2007

Abstract

An interactive optimization methodology for allocating the number and configuration of an Air Quality Monitoring Network (AQMN) in a vast area to identify the impact of multiple pollutants is described. A mathematical model based on the multiple cell approach (MCA) was used to create monthly spatial distributions for the concentrations of the pollutants emitted from different emission sources. These spatial temporal patterns were subject to a heuristic optimization algorithm to identify the optimal configuration of a monitoring network. The objective of the optimization is to provide maximum information about multi-pollutants (i.e., CO, NO_x and SO₂) emitted from each source within a given area. The model was applied to a network of existing refinery stacks and the results indicate that three stations can provide a total coverage of more than 70%. In addition, the effect of the spatial correlation coefficient (R_C) on total area coverage was analyzed. The modeling results show that as the cutoff correlation coefficient R_C is increased from 0.75 to 0.95, the number of monitoring stations required for total coverage is increased. A high R_C -based network may not necessarily cover the entire region, but the covered region will be well represented. A low R_C -based network, on the other hand, would offer more coverage of the region, but the covered region may not be satisfactorily represented.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Air Quality Monitoring Network; Multi-pollutant; Optimization; Industrial area

1. Introduction

Air Quality Monitoring Network (AQMN) design to measure pollutant levels in the atmosphere is of a fundamental significance when environmental impact studies are developed. The United States Environmental Protection Agency (US-EPA, 1999), for instance, indicates that such networks enable us to identify the degree by which pollution is actually taking place nearby pollutants sources; to establish a basis for comparison of air quality standards with the actual air quality levels; and to monitor source compliance with regulations as well as to provide a

database for research and evaluation of effects of urban, land-use and transportation planning.

The design of environmental site sampling selection has been well studied in meteorological sciences (Gandin, 1970), hydrology (Andricevic, 1990; Kassim and Kottegoda, 1991), and geological sciences (De Marsily et al., 1984; Russo, 1984). Several attempts were also made in order to design a representative AQMN (Munn, 1981; Trujillo-Ventura and Ellis, 1991; Chang and Tseng, 1999; Tseng and Chang, 2001). This was not only due to the important nature of the problem involved in atmospheric processes but also to the complex distribution of the emission sources. Air pollutants are emitted from a variety of sources and are transported and dispersed by the atmosphere; the turbulent processes in which can only be described statistically. As a result, pollutant concentrations

*Corresponding author. Tel.: +1 519 888 4567x7157;
fax: +1 519 746 4979.
E-mail address: aelkamel@uwaterloo.ca (A. Elkamel).

not only fluctuate in time but also differ from place to place. Such a condition makes the selection of sites to give a reliable and realistic picture of air quality a major issue and at the same time a challenging task. Among the few studies on AQMN design, Koda and Seinfeld (1978) presented a methodology for distributing a number of monitoring stations in a study area in order to obtain the maximum sensitivity of the collected data to achieve the variations in the emissions of the sources of interest. The developed methodology needs the availability of a mathematical model to estimate the ground level concentrations of pollutants under different meteorological scenarios. The design of an AQMN in the greater London area was conducted by Handscombe and Elson (1982) based on the concept of a spatial correlation analysis. As another example, Graves et al. (1981) performed the design of an AQMN in Fulton County, Georgia in which only non-reactive pollutants were considered. A statistical measure of information content was used as a mean for assessing the effectiveness of a particular monitoring network configuration in Canada (Pickett and Whiting, 1981); and a study on interpolation techniques was undertaken in the Netherlands resulting in an evaluation of the errors involved in the spatial analysis (Egmond and Onderdelinden, 1981).

The consideration of multi-pollutant AQMN design with respect to different objectives was presented in a series of papers by Modak and Lohani (1985a,b), in which the design principles of a Minimum Spanning Tree (MST) algorithm for single or multiple pollutants with respect to one or two objectives was illustrated. McElroy et al. (1986) applied air quality simulation models and population exposure information to produce representative combined patterns and then employed the concepts of 'sphere of influence' (SOI) and 'figure of merit' (FOM) developed by Liu (1986) to determine the minimum number of sites required. Trujillo-Ventura and Ellis (1991) further discussed the issues of multi-objective air pollution monitoring network design. Arbeloa et al. (1993) presented a methodology based on the concepts of potential violations to take into account the number of times that the maximum concentration values tolerated by law are exceeded.

In this paper, an interactive optimization methodology to represent multi-objective multi-pollutants AQMN design is presented. The optimization algorithm is an extension of the ideas presented by Modak and Lohani (1985a,b) and those by Arbeloa et al. (1993). A MATLAB program was written based on the methodology discussed in this paper. The developed algorithm was used to treat the output results of the mathematical modeling based on the multiple cell approach (MCA) delineated in Fatehifar et al. (2006) to represent the optimal multi-pollutant AQMN around Tabriz Refinery area. The objective of this AQMN design is that station location must be optimized with respect to sensitivity of monitoring stations to

evaluate the impact of multi-pollutant gases such as CO, NO_x and SO₂ released from Refinery stacks.

2. Methodology description

The main theme of the methodology delineated below is to build a visible solution space for the selection of the optimum monitoring sites. Such visible solution can be obtained from concentration values from experimental campaigns of measures of atmospheric pollution or estimations of such concentrations from a mathematical model. In general, the use of a mathematical model is recommended for creating a more extensive database. In this way, an initial hypothetical network as dense as desired can be generated within the vicinity of the emission sources.

For the designed network to be effective, air quality data should be collected during a sufficient period for the final data set to be representative of the area in which the network is to be located. One year is usually considered a minimum for reflecting the meteorological variability of the area (Munn, 1981). In this study, a mathematical model based on a multiple cell approach (MCA) is used to generate discrete grid points each of which represents a candidate location. Starting from this set of potential sites in which a monitoring station can be located, the basic idea is to formulate an optimization strategy based on different siting objectives and to take into account the coverage area and the detection ability of legal violations of each allocated station.

Monitoring objectives considered in the optimal allocation of AQMN through multi-objective analysis may include: (1) monitoring stations should be located in areas of highly populated regions; (2) monitoring stations should be located in areas where pollution concentrations are expected to be the highest; (3) monitoring stations should be located in areas where the highest frequency of violations can be detected; (4) monitoring stations should be located in areas where significant economic growth is expected to occur; and (5) monitoring stations should be located near major industrial sources (Tseng and Chang, 2001).

2.1. Dispersion model

The multiple cell approach (MCA) used in this work is based on the solution of the three-dimensional diffusion equation to predict the dispersion of emitted pollutants from refinery stacks. The prediction of the pollutants emitted from the stacks can be broken down into the following steps:

1. Describe the geometry of the domain.
2. Describe the position of stacks in the selected domain.
3. Introduce appropriate boundary conditions.
4. Introduce sources, sinks, and the dispersion characteristics for the entire domain.

5. Select values or required equations for parameters in the model.
6. Divide the domain into cells and solve the finite difference equations.
7. Find the concentrations at all grids for all pollutants simultaneously.
8. Output results in a visualized form.

Step 2 in the above procedure differs slightly from the single stack case as we are considering multiple stacks at different locations. The modeling process starts with first writing a conservation of mass equation for each pollutant and then dividing the air space into an array of cells (Fig. 1).

The conservation of mass for a fluid element and for a given pollutant S may be written as

$$\frac{\partial C^s}{\partial t} = -\frac{\partial(U_x C^s)}{\partial x} - \frac{\partial(U_y C^s)}{\partial y} - \frac{\partial(U_z C^s)}{\partial z} + \frac{\partial}{\partial x} \left(K_x \frac{\partial C^s}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_y \frac{\partial C^s}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_z \frac{\partial C^s}{\partial z} \right) + E^s - (k_1^s + k_2^s) C^s + Q(C^s), \quad s = 1, 2, \dots, q, \quad (1)$$

where C^s is the concentration of the chemical species involved in the model; U_x , U_y and U_z are wind velocities in the x -, y - and z -directions; K_x , K_y and K_z are diffusion coefficients; E^s are the emission sources; K_1^s and K_2^s are deposition coefficients (for dry deposition and wet deposition, respectively); and $Q(C^s)$ represents chemical reactions. The assumptions made are:

1. Steady state conditions ($\partial C / \partial t = 0$).
2. $U_y = U_z = 0$ (wind velocity in x -direction only).
3. Transport by bulk motion in the x -direction exceeds diffusion in the x -direction ($K_x = 0$).
4. There is no deposition in the system ($K_1^s = K_2^s = 0$).
5. There is no reaction in the system ($Q = 0$).

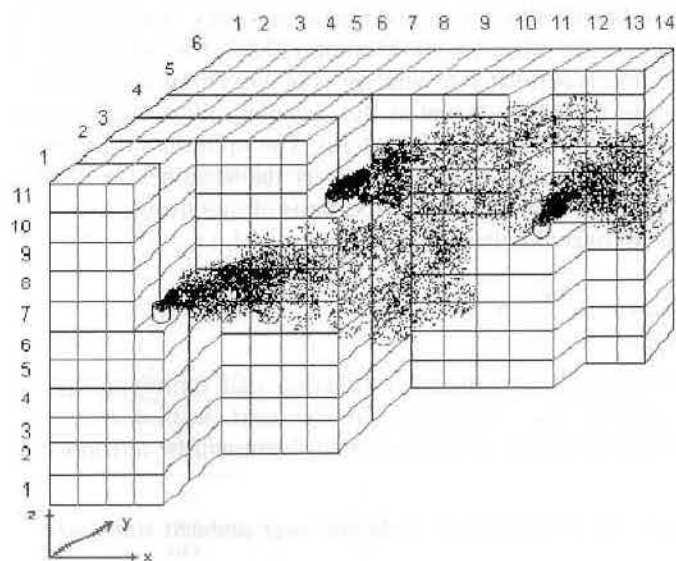


Fig. 1. Plume rise, pollutant dispersion and grid generation for network of refinery stacks.

Furthermore, the boundary and initial conditions required to solve (1) are

$$\text{at } x = 0, \quad C(0, j, k) = 0,$$

$$\text{at } y = 0, \quad \frac{\partial C}{\partial y} = 0,$$

$$\text{at } y = W, \quad \frac{\partial C}{\partial y} = 0,$$

$$\text{at } z = 0, \quad \frac{\partial C}{\partial z} = 0,$$

$$\text{at } z = \text{mixing length}, \quad \frac{\partial C}{\partial z} = 0, \quad (2)$$

where W and mixing height are illustrated in Fig. 2. Discretizing Eq. (1) according to the cells of Fig. 1 and applying the boundary conditions shown in Eq. (2) along with the stated assumptions leads to the finite difference representation given by

$$\begin{aligned} U_{xk} C_{i+1,j,k}^s \Delta y \Delta z + (K_y)_k \Delta x \Delta z (C_{i+1,j,k}^s - C_{i+1,j-1,k}^s) / \Delta y \\ + (K_y)_k \Delta x \Delta z (C_{i+1,j,k}^s - C_{i+1,j+1,k}^s) / \Delta y \\ + (K_z)_{k+1/2} \Delta x \Delta y (C_{i+1,j,k}^s - C_{i+1,j,k+1}^s) / \Delta z \\ + (K_z)_{k-1/2} \Delta x \Delta y (C_{i+1,j,k}^s - C_{i+1,j,k-1}^s) / \Delta z \\ = U_{xk} C_{i,j,k}^s \Delta y \Delta z + E_{i,j,k}^s \Delta y \Delta z, \end{aligned} \quad (3)$$

where the values of wind speed and eddy diffusivity are presumed known for every cell. Table 1 shows equations used to estimate the values of wind speed and eddy diffusivities. This is an explicit algebraic formulation and may be unstable in some conditions. The stability condition for this system is described as (Constantinides and Mostoufi, 1999)

$$\Delta x \leq \frac{U_x}{2K_z \left(\frac{5}{\Delta y^2} + \frac{1}{\Delta z^2} \right)}. \quad (4)$$

For illustration purposes, Fig. 1 indicates three stacks in positions (4, 1, 7), (8, 3, 9), and (12, 5, 6). In these cells, the term E^s due to stack emission must be considered. This term is zero for all other cells. More details about the modeling of the above phenomena can be found in our previous paper (Fatehifar et al., 2006).

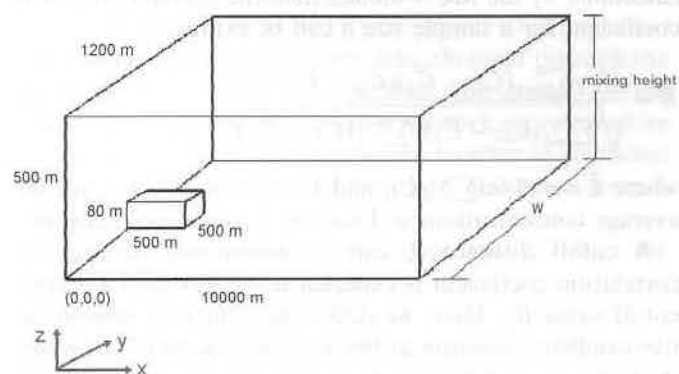


Fig. 2. Selected domain for simulation.

Table 1

Wind velocity and Eddy diffusivity for various stability categories (Ragland, 1973; Ragland and Dennis, 1975; Zannetti, 1990)

Stability	Wind velocity	Eddy diffusivity
<i>In surface layer, $0 < z < z_0$</i>		
Neutral	$\frac{u^*}{0.4} \ln \left(\frac{z+z_0}{z_0} \right)$	$K_z = \frac{u^*}{f} \left(7.396 \times 10^{-4} + 6.082 \times 10^{-2} \left(\frac{zf}{u^*} \right) + 2.532 \left(\frac{zf}{u^*} \right)^2 - 12.72 \left(\frac{zf}{u^*} \right)^3 + 15.17 \left(\frac{zf}{u^*} \right)^4 \right), \quad 0 \leq \frac{zf}{u^*} \leq 0.45$ $K_z = 0, \quad \frac{zf}{u^*} > 0.45$ $K_y = 5K_z$
Stable	$\frac{u^*}{0.4} \left[\ln \left(\frac{z+z_0}{z_0} \right) + \frac{5.2z}{L} \right]$	$K_z = 0.4u^*z / \left(1 + \frac{5.2z}{L} \right)$ $K_y = 6K_z$
Unstable	$\frac{u^*}{0.4} \left(\frac{2(\tan^{-1} x - \tan^{-1} x_0) + \ln \left(\frac{x-1}{x_0-1} \right)}{-\ln \left(\frac{x+1}{x_0+1} \right)} \right)$ $x = [1 - 15(z-z_0)/L]^{1/4}, \quad x_0 = [1 - 15z_0/L]^{1/4}$	$K_z = 0.4u^*z \left(1 - \frac{15z}{6L} \right)^{1/4}$ $K_y = 2K_z$
<i>Upper surface layer, $z_{sl} < z < z_m$</i>		
Neutral	$(u_g - u_{sl}) \left(\left(\frac{z-z_{sl}}{z_m-z_{sl}} \right)^{0.2} + u_{sl} \right)$	$K_z = K_z \quad \text{at } z = z_{sl}$ $K_y = 5K_z$
Stable	$(u_g - u_{sl}) \left(\left(\frac{z-z_{sl}}{z_m-z_{sl}} \right)^{0.5} + u_{sl} \right)$	$K_z = 0.4u_*L$ $K_y = 6K_z$
Unstable	$(u_g - u_{sl}) \left(\left(\frac{z-z_{sl}}{z_m-z_{sl}} \right)^{0.2} + u_{sl} \right)$	$K_z = 160u_*^2 \left(1 - \frac{6000u_*^*}{L} \right)^{(1/4)}$ $K_y = 2K_z$

2.2. First objective: spatial coverage representation

The spatial area coverage of a monitoring location also called SOI by Liu (1986) is defined as the zone over which the air quality data for a given monitoring location can be considered representative. Identification of representation for the SOI is based on the use of the spatial correlation coefficient (R). This coefficient gives an indication of the relationship among the monitoring locations to be selected in the designed optimal monitoring network. Suppose, for instance, that two adjacent monitoring locations have the following time series pollutant concentrations $C_1 = (C_{11}, C_{12}, C_{13}, \dots, C_{1n})$ and $C_2 = (C_{21}, C_{22}, C_{23}, \dots, C_{2n})$ calculated by the MCA model, then the spatial correlation coefficient for a sample size n can be expressed as

$$R = \frac{\sum_{i=1}^n (C_{1i} - \bar{C}_1)(C_{2i} - \bar{C}_2)}{\sqrt{\sum_{i=1}^n (C_{1i} - \bar{C}_1)^2 \sum_{i=1}^n (C_{2i} - \bar{C}_2)^2}}, \quad (5)$$

where $\bar{C}_1 = (1/n) \sum_{i=1}^n C_{1i}$ and $\bar{C}_2 = (1/n) \sum_{i=1}^n C_{2i}$ are the average concentrations at locations 1 and 2, respectively.

A cutoff distance d_c can be established so that the correlation coefficient is expected to be less than a certain cutoff value R_c . Here, we define the zone of a representative candidate location as the area surrounding it in which the spatial correlation coefficient of this location with the nearby locations is higher than a cutoff value R_c . This

means that the air quality data measured at this location can be considered representative and extrapolative with a certain degree of confidence to any location in the designed network within the area.

When an SOI for a monitoring location is calculated, the coverage area of this sphere is defined as the number of potential monitoring sites placed inside it. The coverage area can be quantified in terms of pattern scores. A pattern score (N_p^i) for an i th candidate location is defined as the number of locations correlated to time series of concentrations predicted using the MCA model at this location, above the specified cutoff correlation coefficient R_c . The optimization problem could therefore be deliberated as the identification of the optimum number (m) and locations of monitoring sites such as an overall pattern score is maximized, i.e.,

$$\text{Maximize } \sum_{i=1}^m N_p^i, \quad (6)$$

So far only one objective function that represents spatial coverage was discussed. In the next section a second objective of the optimization problem will be introduced.

2.3. Second objective: violations over ambient standards

A number of objective functions have been used in the past to measure the capability of a designed network to

detect violations of air quality standards. The adapted approach taken in this work considers the potential of a monitoring site for the detection of violations in terms of violation scores. This choice is due to the need to incorporate in the optimization algorithm the relative significance of the various alternative solutions. A candidate location with a high violation score is considered to be of a high potential for detection of violations. The computation of violation scores is essentially a weighted scoring that corresponds to concentrations of SO_2 , NO_x and CO above prescribed thresholds. These scores are therefore dependent on the threshold levels and the weighing factors between each threshold range and the weighing function.

A decision on threshold levels and weighing factors is indeed pollutant-specific and dependent upon the averaging time considered. Several weighing functions can be employed such as linear, segmented linear, non-linear, and segmented non-linear (Ott, 1977). In this study, a segmented non-linear weighing function is used (Table 2). The weights in the table represent an air pollution index that ranges from 0.5 to 5 according to the severity of threshold exceedance. The violation score for each candidate location is calculated from the equation

$$N_v^i = \sum_{j=1}^T \sum_{k=1}^{N_k} \frac{(w_{k+1} - w_k)(x_i - x_k)X}{(x_{k+1} - x_k)}, \quad (7)$$

where N_v^i is the violation score for the i th candidate location, w_k the weighing factor corresponding to threshold x_k , x_k the k th threshold, $X = 0$ if $(x_i - x_k) \leq 0$, $X = 1$ otherwise, N^T the total number of thresholds, T the total number of simulated observations.

The weighing function associated with the violation scores is important because the severity of threshold violations reported by two monitoring stations may differ. In this work, the values for thresholds of SO_2 , NO_x and CO were chosen based on the US-EPA limit values of common pollutants and are also given in Table 2.

Since the interest of the optimization problem is to achieve a maximum detection of violations over ambient air standards, the second objective of the optimization problem can be formulated as

$$\text{Maximize } \sum_{i=1}^m N_v^i, \quad (8)$$

Table 2
Air pollution index assigned to weighing factors values for the violation score (Sarigiannis, 1999)

SO_2 ($\mu\text{g}/\text{m}^3$)	NO_x ($\mu\text{g}/\text{m}^3$)	CO ($\mu\text{g}/\text{m}^3$)	Weighing factors
80	30	4000	0.5
120	80	6000	1
140	100	8000	2.5
160	130	13,000	3
190	160	20,000	5

where N_v^i is the potential of violation of the i th position, and m is the number of monitors.

2.4. Formulation of a composite objective function

In order to meet the two objectives described above a composite objective function is formulated as

$$F = N_p \times N_v^b. \quad (9)$$

The parameter b is used to weigh the relative importance given to the two objectives and its value depends on the purpose of the network to be developed. This structure of the objective function ensures that if any of the two sub-objectives becomes zero the overall objective function (F) will also have a zero value (Modak and Lohani, 1985a,b). This condition would imply that if a location has a high violation score, but zero effective pattern score (i.e., it is already covered by other locations), then it is discarded in the optimal design. The optimization problem can now be state as

$$\text{Maximize } \sum_{i=1}^m F_i \quad (10)$$

such that $(\text{SOI}^1 \cup \text{SOI}^2 \cup \dots \cup \text{SOI}^m) = (\text{SOI}^1 \cup \text{SOI}^2 \cup \dots \cup \text{SOI}^N)$; total coverage or $m = m^0$; partial coverage, where F_i is the objective function of the i th candidate location, m the number of monitors, m^0 the maximum number of monitors based on the budgetary constraint, N the number of candidate locations, and SOI^i the set of monitoring locations correlated with location i .

3. A heuristic algorithm for single or multi-pollutant network design

Due to the computational complexity of the above optimization problem, we present here a simple heuristic algorithm for generating a near-optimal pollution monitoring network. Work is in progress to develop an exact solution methodology. The basic idea of the heuristic is to build up a single feasible solution space in a deterministic sequential fashion. A description of the heuristic algorithm for the solution of the above problem based on one pollutant network design is given below:

1. Start from the concentration data obtained through the mathematical model (i.e., multiple cell model) of the pollution phenomena to form a network represented by a matrix of $M \times N$ where M is the number of predicted concentrations at potential locations numbered from 1 to N .
2. Calculate the spatial correlation matrix R_{ij} for $i = 1, 2, \dots, N$; $j = 1, 2, \dots, N$.
3. Calculate SOI_i for $i = 1, 2, \dots, N$.
4. Calculate the violation scores N_v^i , $i = 1, 2, \dots, N$.
5. Define the coverage area in term of pattern score N_p^i for each SOI obtained, for $i = 1, 2, \dots, N$.

6. Calculate the overall objective function F_i , for $i = 1, 2, \dots, N$ using Eq. (11) with a suitable value for b .
7. Choose the i th candidate location (monitoring station) of maximum F .
8. In order to avoid overlaps, the i th location(s) belonging to the SOI of the station being selected in step 7 are deleted from all sets SOI_i to which they belong.
9. Back to step 5. The loop ends when the number of stations is adequate.

In this heuristic, it is necessary to specify the adequate number of stations at the start of the algorithm. This number usually depends on a budgetary constraint such that if an organization has \$10,000,000 to construct a monitoring network and each station costs \$1,000,000 then at most 10 monitoring stations can be built.

To extend the above heuristic to multiple pollutants, we formulate a weighted sum objective function as

$$U^i(u) = \sum_{j=1}^P w_j F_{ij}^i \quad (11)$$

$$\sum_{j=1}^P w_j = 1, \quad (12)$$

where w_j is the importance associated with pollutant j , F_{ij}^i the objective function for pollutant j at location i , $U^i(u)$ the additive/overall objective for P pollutants at location i .

The solution algorithm for the multi-pollutants network design becomes a straightforward extension of the single pollutant case. The numerous steps of this algorithm are similar to those discussed above for the single pollutant, except that the objective function is calculated as described by Eq. (11). Fig. 3 shows a flowchart depicting the above algorithm. This algorithm has been implemented in MATLAB.

4. Illustrative case study, results, and discussion

The heuristic algorithm discussed above was used to identify the optimal design of a monitoring network around Tabriz Refinery. Tabriz Refinery with 2 km² area is located at 15 km of Tabriz-Azarshahr road, Azarbyjan Shargi province, Iran. Fig. 4 shows the location of refinery stacks, and wind direction (needed in modeling pollutants dispersion). Fig. 5 shows ground level concentrations of NO_x and SO₂ in the study area. Finally, it is noted that Tabriz Refinery is located at a smooth plate.

For potential sites, the MCA model uses six years (from 1990 to 1995) real meteorological data collected from the vicinity of the refinery stacks. A matrix of 72 × 868 for SO₂, NO_x, and CO concentrations that has been generated at the specified 868 candidate locations was used as an input to the above algorithm. These candidate locations serve as potential monitoring stations with 72 predicted concentrations (12 months for 1990–1995) ready to be selected by the algorithm. The wind velocities and temperatures data used in the modeling of pollutants dispersion for the period of study are given in Tables 3 and 4.

4.1. Single pollutant monitoring network optimization modeling results

The heuristic algorithm was used to find the optimal locations and the configuration of the monitoring networks for sulfur dioxide (SO₂), nitrogen oxide (NO_x) and carbon monoxide (CO), separately within the area under study. The six year (monthly) predicted concentrations of SO₂, NO_x, and CO by the multiple cell model within Tabriz Refinery area were fairly low, and the violations of the EPA ambient air quality standard (AAQS) were never exceeded. For the purpose of illustration therefore, the concentrations of SO₂, CO and NO_x were scaled up by factors of 10,000, 1000 and 100, respectively. It is to be noted that a constant multiplier such as 100 or 1000 does not change the correlation coefficient matrix, so the analysis is not affected as far as the patterns are concerned.

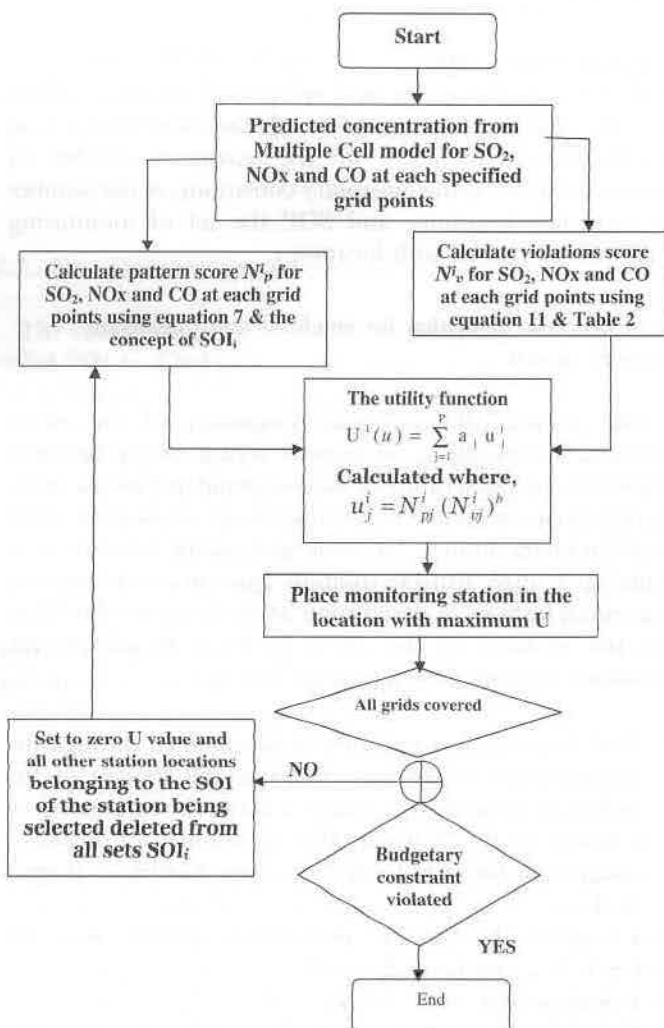


Fig. 3. Flowchart determination procedure for allocation of monitoring stations.

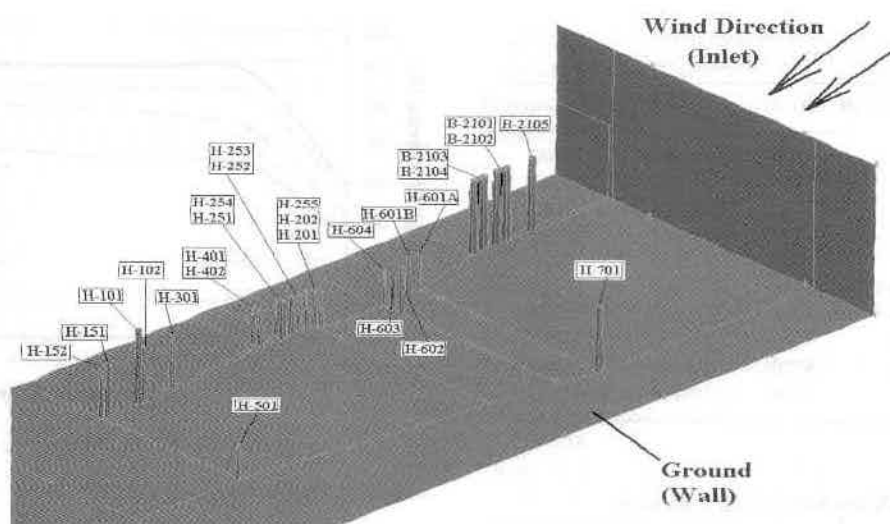
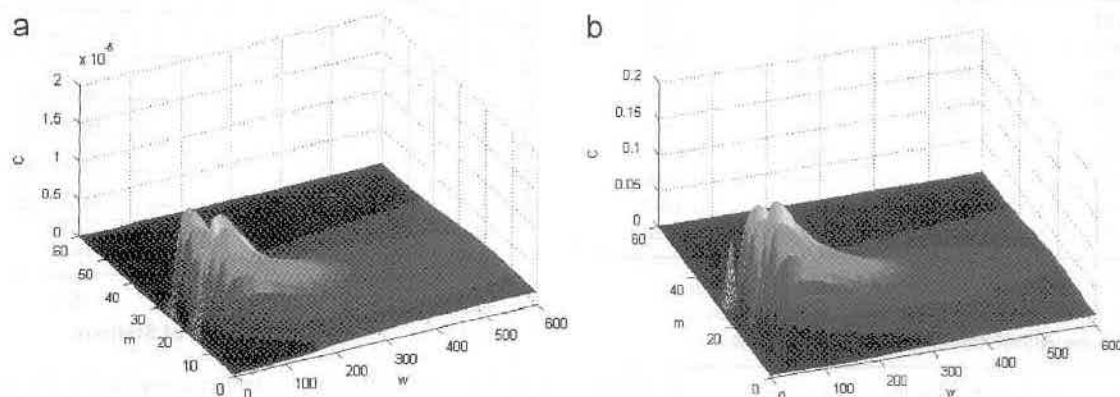


Fig. 4. Location of Tabriz Refinery Stacks.

Fig. 5. (a) NO_x and (b) SO_2 concentration distributions at ground level. C is in g/m^3 , w is number of grids in x -direction and m is number of grids in y -direction.Table 3
Monthly mean temperature distribution for different years ($^{\circ}\text{C}$)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1995	1.2	2.8	6.8	11	17.7	21.2	25.3	27	21	12.8	6.8	8.1
1994	-0.8	0.6	6.3	13.5	16.4	21.5	24.7	23.2	19.9	14.7	7.1	-2.4
1993	-4.6	-1.3	3	10.4	15.3	21.3	25.6	24.8	21.7	13.6	3.2	0.3
1992	-4	-1.2	2	9.8	13.1	20.4	23.4	22.9	20	13.9	6.3	0.8
1991	-1.8	-0.8	5.4	12.8	13.5	22.4	26	26.5	21.2	13.1	7.3	-0.5
1990	-2.6	-0.3	4.3	10	16.9	23.4	26.5	24.8	22.2	14	8.7	1.4

Table 4
Monthly mean wind velocity distribution for different years (m/s)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1995	2.1	2.15	2.15	4.05	3.95	3.6	4.45	4.15	3.3	3.15	2.45	2.75
1994	1.65	3.05	1.55	2.6	3.65	4.3	4.55	4.75	4.2	3.3	2.3	2.55
1993	2.25	2.75	2.25	5.65	4.1	3.75	4.6	4.35	4.3	4.15	3.5	2.3
1992	1.7	2.3	3.2	4.75	3.2	3.8	3.75	4.1	3.7	4.55	3.25	3.15
1991	1.6	1.9	4.05	4.9	4.55	4.85	4.75	4.4	3.75	3.55	2.95	2.2
1990	2	2.8	3.1	4.4	3.2	4.95	4.9	4.9	3.65	2.6	1.6	2.75

The heuristic algorithm was implemented for different cutoff R_C values in the correlation coefficient matrix. The value of R_C was varied from 0.75 to 0.95 in order to study the effect of R_C on the coverage effectiveness of the monitoring networks. The optimal locations for different R_C values are shown in Table 5. For a stipulated budget, an air quality monitoring organization can maintain either a high or a low value R_C based network. A high R_C based network may not necessarily cover the entire region, but the covered region will be well represented. A low R_C based network on the other hand, would offer more coverage of the region, but the covered region may not be satisfactorily represented. The ultimate decision in such a case is of course left to the air quality monitoring organization.

Tables 6 and 7 show optimal locations of AQMN for NO_x and SO_2 for a maximum number of five stations. The number of stations is of course subject to a budgetary constraint. It should be noted that a network designed for a higher value of R_C is not necessarily a simple extension, in terms of a mere addition of some more monitors, as compared to that of a network obtained at a lower R_C value. Rather, the expansion of a network to a higher

Table 5
Optimal locations for AQMN for CO (x and y are in m)

Station	$R_C = 0.75$	$R_C = 0.80$	$R_C = 0.85$	$R_C = 0.90$	$R_C = 0.95$
1	$X = 800$ $Y = 320$	800 320	800 320	800 320	800 320
2	7440 320	8480 320	7760 320	8240 320	7280 320
3	2240 320	2320 320	2800 160	2800 160	9920 160
4	9920 640	2960 160	2400 320	2480 320	2480 320
5	3120 160	8320 640	8720 160	9920 160	2480 160

Table 6
Optimal locations for AQMN for NO_x (x and y are in m)

Station	$R_C = 0.80$	$R_C = 0.85$	$R_C = 0.90$	$R_C = 0.95$
1	4000 320	4240 320	4400 320	4320 320
2	6400 480	4240 480	5360 480	6140 480
3	6320 160	5680 160	6140 160	3120 320
4	4240 480	4400 480	7680 480	4160 480
5	5040 320	5280 320	5520 320	5600 320

Table 7
Optimal locations for AQMN for SO₂ (x and y are in m)

Station	$R_C = 0.90$	$R_C = 0.95$
1	3200 160	3280 160
2	8640 320	7600 320
3	2480 320	9920 160
4	4800 160	2480 320
5	4880 800	4800 800

reliability of representation will require an additional number of monitors but also the relocation of existing ones.

Figs. 6–8 show the coverage efficiencies of AQMN for different R_C values. As shown in the figures, three stations for AQM are enough.

4.2. Multi-pollutant monitoring network optimization modeling results

The heuristic algorithm was used to find the “optimal” locations and the configuration of the monitoring networks for monitoring simultaneously SO₂, NO_x and CO within the area under study. For this purpose, the six year

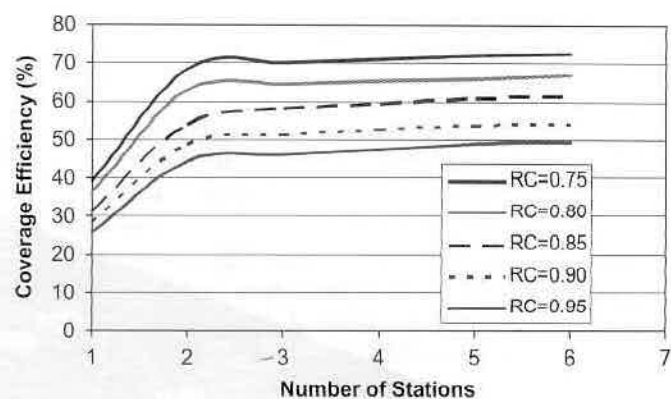


Fig. 6. Coverage efficiency vs. number of stations for CO as a function of R_C .

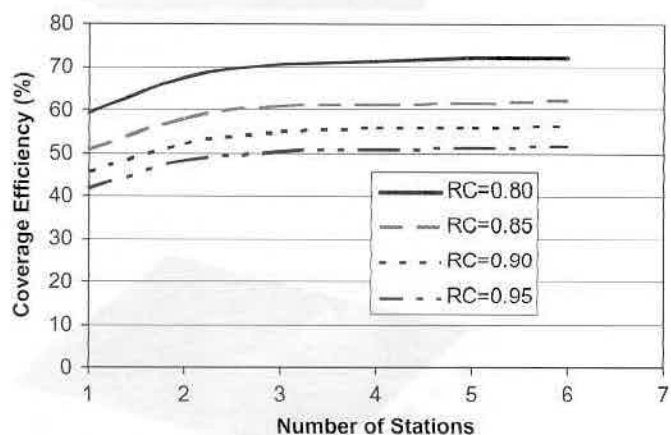


Fig. 7. Coverage efficiency vs. number of stations for NO_x as a function of R_C .

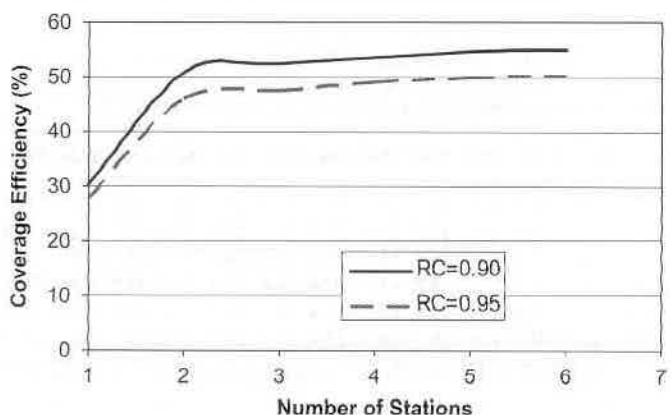


Fig. 8. Coverage efficiency vs. number of stations for SO₂ as a function of R_C .

(monthly) predicted concentrations of SO₂, NO_x, and CO by the multiple cell model within Tabriz Refinery area were used. Table 8 shows the optimal locations of the monitors for a maximum number of five stations. Fig. 9 shows the coverage efficiency of the AQMN for different R_C values. As shown in the figure, three stations for AQM are enough; similar to the single pollutant case.

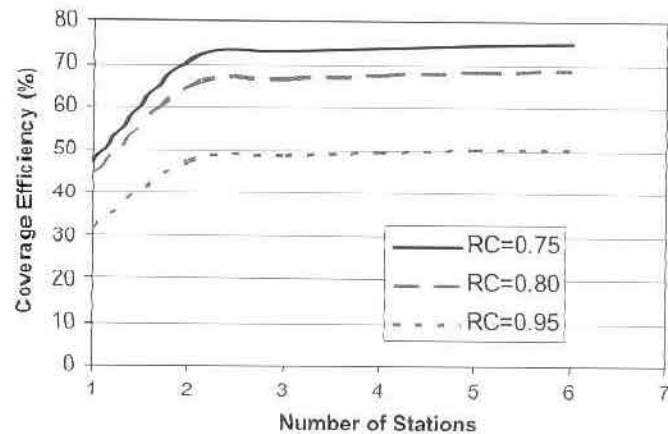


Fig. 9. Coverage efficiency vs. number of stations for multi-pollutants as a function of R_C .

Table 8

Optimal locations for AQMN for multi-pollutant (x and y are in m)

Station	$R_C = 0.75$	$R_C = 0.80$	$R_C = 0.95$
1	2880 160	3120 160	3040 160
2	8320 320	8560 320	7600 320
3	2720 480	2640 480	7920 160
4	3840 640	3920 640	2480 480
5	9760 640	10,000 1200	5440 800

5. Conclusion

AQMN represents an essential tool to monitor and control atmospheric pollution. The use of some specific criteria in conjunction with mathematical models for the spatial and temporal dispersion of pollutants within the atmosphere provides a general approach to determine the optimal number and location of monitoring stations.

In this study, an improved methodology for the optimization of an AQMN with respect to multi-pollutants is presented. For this purpose, the authors proposed model (Fatchifar et al., 2006) was used to generate a monthly database from 1990 to 1995 around Tabriz refinery. This database was then employed in an optimization methodology for generating an AQMN. The generated monitoring network provides maximum information about the multi-pollutants (CO , NO_x and SO_2) emitted from each stack within the refinery area. The effect of the correlation coefficient on total coverage and effectiveness of the network was also studied. The results show that for the area under study three stations are enough and can give a total coverage area of about 70%.

Acknowledgments

The authors would like to thank Tabriz Refinery Research and Development Department for their kind

cooperation and for providing the case study. The financial support of the Natural Sciences and Engineering Research Council (NSERC) of Canada is also gratefully acknowledged.

References

- Andricevic, R., 1990. A real-time approach to management and monitoring of groundwater hydraulic. *Water Resources Research* 26, 875–885.
- Arbeloa, S.F.J., Caseiras, C.P., Andres, P.M.L., 1993. Air quality monitoring: optimization of a network around a hypothetical potash plant in open countryside. *Atmospheric Environment* 27A (5), 729–738.
- Chang, N.B., Tseng, C.C., 1999. Optimal design of a multi-pollutant Air Quality Monitoring Network in a metropolitan region using Kaohsiung, Taiwan as an example. *Environmental Monitoring and Assessment* 57, 121–148.
- Constantinides, A., Mostoufi, N., 1999. *Numerical Methods for Chemical Engineers with MATLAB Applications*. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- De Marsily, G., Lavedan, G., Boucher, M., Fasanino, G., 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In: Verly, G., et al. (Eds.), *Geostatistics for Natural Resources Characterization, Part 2*. D. Reidel Publishing, Dordrecht, pp. 831–849.
- Egmond, N.D.V., Onderdelinden, D., 1981. Objective analysis of air pollution monitoring network data: spatial interpolation and network density. *Atmospheric Environment* 15 (6), 1035–1046.
- Fatchifar, E., Elkamel, A., Taheri, M., 2006. A MATLAB-Based modeling and simulation program for dispersion of multi-pollutants from an industrial stack for the educational use in a course on air pollution control. *Computer Applications in Engineering Education* 14 (4), 300–312.
- Gandin, L.S., 1970. The planning of meteorological station networks. WMO No. 265.
- Graves, R.J., Lee, T.D., McGinnis, L.F.J., 1981. Air Monitoring Network Design: case study. *Journal of Environmental Engineering ASCE* 107 (5), 941.
- Handcombe, C.M., Elson, D.M., 1982. Rationalisation of the National Survey of Air Pollution Monitoring Network of the United Kingdom using spatial correlation analysis: a case study of the Greater London area. *Atmospheric Environment* 4B (3), 395.
- Kassim, A.H.M., Kottegoda, N.T., 1991. Rainfall network design through comparative kriging methods. *Hydrological Sciences Journal HSIJODN* 36 (3), 223–240.
- Koda, M., Seinfeld, J.H., 1978. *Air monitoring siting by objective*. EPA-600/4-7-036, U.S. Environmental Protection Agency, Las Vegas, Nevada.
- Liu, M.K., 1986. Methodology for designing Air Quality Monitoring Networks: I. Theoretical aspects. *Environmental Monitoring and Assessment* 6, 1–11.
- McElroy, J.L., Behar, J.V., Meyers, T.C., Liu, M.K., 1986. Methodology for designing Air Quality Monitoring Networks: II. Application to Las Vegas, Nevada, for carbon monoxide. *Environmental Monitoring and Assessment* 6, 13.
- Modak, P.M., Lohani, B.N., 1985a. Optimization of ambient Air Quality Monitoring Networks: Part I. *Environmental Monitoring and Assessment* 5, 1–19.
- Modak, P.M., Lohani, B.N., 1985b. Optimization of ambient Air Quality Monitoring Networks: Part II. *Environmental Monitoring and Assessment* 5, 21–38.
- Munn, R.E., 1981. *The Design of Air Quality Monitoring Networks*. MacMillan, London.
- Ott, W.J., 1977. Development of criteria for siting air monitoring stations. *Journal of the Air Pollution Control Association* 27, 543–547.

- Pickett, E.E., Whiting, R.G., 1981. The design of cost-effective Air Quality Monitoring Networks. *Environmental Monitoring and Assessment* 1, 59.
- Ragland, K.W., 1973. Multiple box model for dispersion of air pollutants from area sources. *Atmospheric Environment* 7, 1017–1032.
- Ragland, K.W., Dennis, R.L., 1975. Point source atmospheric diffusion model with variable wind and diffusivity profiles. *Atmospheric Environment* 9, 175–189.
- Russo, D., 1984. Design of an optimal sampling network for estimating the variogram. *Social Science Society of American Journal* 48, 708–716.
- Sarigiannis, D.A., 1999. Multi objective optimization of ambient air quality networks. European Commission, Institute for Systems, Informatics and Safety Systems Analysis and Information Assessment Unit.
- Trujillo-Ventura, A., Ellis, J.H., 1991. Multi-objective air pollution monitoring network design. *Atmospheric Environment* 25A, 469–479.
- Tseng, C.C., Chang, N.B., 2001. Assessing relocation strategies of urban air quality monitoring stations by GA-based compromise programming. *Environment International* 26, 523–541.
- US Environmental Protection Agency (US-EPA) (1999). Guidance for network design and optimum site exposure for $PM_{2.5}$ and PM_{10} . EPA-451R-99-022.
- Zanetti, P., 1990. Air Pollution Modeling Theories, Computational Methods and Available Software. Computational Mechanics Publications.

Thermophilic adaptation of a mesophilic anaerobic sludge for food waste treatment

Luis Ortega^{a,b}, Suzelle Barrington^{a,*}, Serge R. Guiot^b

^aDepartment of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Ste-Anne-de-Bellevue, Que., Canada H9X 3V9

^bBiotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montréal, Que., Canada H4P 2R2

Received 20 August 2006; received in revised form 4 March 2007; accepted 16 March 2007

Available online 27 September 2007

Abstract

As opposed to mesophilic, thermophilic anaerobic digestion of food waste can increase the biogas output of reactors. To facilitate the transition of anaerobic digesters, this paper investigated the impact of adapting mesophilic sludge to thermophilic conditions. A 5 L bench scale reactor was seeded with mesophilic granular sludge obtained from an up-flow anaerobic sludge blanket digester. After 13 days of operation at 35 °C, the reactor temperature was instantaneously increased to 55 °C and operated at this temperature until day 21. The biomass was then fed food waste on days 21, 42 and 63, each time with an F/M (Food/Microorganism) ratio increasing from 0.12 to 4.43 gVS/gVSS. Sludge samples were collected on days 0, 21, 42 and 63 to conduct substrate activity tests, and reactor biogas production was monitored during the full experimental period. The sludge collected on day 21 demonstrated that the abrupt temperature change had no pasteurization effect, but rather lead to a biomass with a fermentative activity of 3.58 g Glucose/gVSS/d and a methanogenic activity of 0.47 and 0.26 g Substrate/gVSS/d, related respectively, to acetoclastic and hydrogenophilic microorganisms. At 55 °C, an ultimate gas production (G_0) and a biodegradation potential (B_0) of 0.2–1.4 L_{STP}/gVS_{fed} and of 0.1–0.84 L_{STP} CH₄/gVS_{fed} were obtained, respectively. For the treatment of food waste, a fully adapted inoculum was developed by eliminating the initial time-consuming acclimatization stage from mesophilic to thermophilic conditions. The feeding stage was initiated within 20 days, but to increase the population of thermophilic methanogenic microorganisms, a substrate supply program must be carefully observed.

© 2007 Published by Elsevier Ltd.

Keywords: Anaerobic digestion; Biogas production; Food waste; Mesophilic conditions; Thermophilic adaptation.

1. Introduction

In North America and Europe, up to 35% of the total mass of generated municipal solid waste can be made up of organic components (Rhynder et al., 1995). To divert the organic fraction of municipal solid wastes (OFMSW) from landfills while beneficially reusing such waste, more energy-efficient practices are required. For OFMSW, the conventional mesophilic-anaerobic/composting coupled system offers an energy potential of 150 kWh/t_{MSW} which is 21 times higher than that reported for landfill municipal solid waste of 6.9 kW-h/t_{MSW} recovered from collected biogas (Baldasano and Soriano, 2000). Despite this higher energy yield, there are only a few conventional anaerobic digesters

(AD) in North America for the transformation of OFMSW into green-energy carriers (USEPA, 2003). In comparison, Europeans have long accepted to treat the OFMSW using AD, but also know that the choice of technologies depends on the substrate characteristics.

Thermophilic conditions to operate one or two-stage AD systems can further improve the energy efficiency of the technology. According to De Baere (2000) and as opposed to mesophilic plants, the start-up of the first thermophilic plant in 1992 has increased the cumulative capacity (t/year) of treating OFMSW from 56 000 to 280 000 t/year, and this from 1997 through 2000. Under thermophilic conditions, a high level of pathogen kill-off is obtained and more heat and electricity are recovered. Nevertheless, the optimal transition of AD systems from mesophilic to thermophilic conditions is not clearly defined. The transition depends on the microbial adaptation to new operational conditions or

*Corresponding author. Tel.: +1 514 398 7776; fax: +1 514 398 8387.
E-mail address: barrington@macdonald.mcgill.ca (S. Barrington).

Nomenclature

Bo	ultimate biodegradation (methane) potential
C/N	carbon to nitrogen ratio
COD	chemical oxygen demand
C0, C1, C2, C3, C4	batch stages
F/M	food to microorganism ratio
Go	ultimate gas production potential
L _{STP}	litre of gas corrected to standard pressure (103.25 kPa) and temperature (273.15 K)
MSW	municipal solid waste

OFMSW	the organic fraction of municipal solid waste
OLR	organic loading rate
SA	substrate consumption rate
SC-OFMSW	selectively collected organic fraction of municipal solid waste
TKN	total Kjeldahl nitrogen
TS	total solids
VFA	volatile fatty acids
VS	volatile solids
VSS	volatile suspended solids

the presence of different groups of microorganisms capable of performing their biological functions at specific temperatures.

When the groups present are not capable of performing under mesophilic and thermophilic conditions, a longer temperature shift period is required. Also, a rapid temperature change from mesophilic to thermophilic may bring about a population shift if the groups are not compatible (Mata-Alvarez, 2002). Nozhevnikova et al. (1999) have assumed that thermophilic and mesophilic methanogenic communities differ in terms of microbial species. This assumption may not be completely true as some mechanisms may permit the microorganisms to evolve from one specific state to another depending on the transition conditions. Nielsen and Petersen (2000) mentioned that the quick and satisfactory start up of full-scale thermophilic anaerobic digesters in Denmark has been achieved with and without a transitional period. The first two days of starting the plant at thermophilic temperatures were associated with a high content of volatile fatty acids and a high biogas CO₂ concentration, requiring the careful monitoring of the loading rate. This period was avoided when the digester was immediately started using thermophilic temperatures. For food waste mechanically sorted from MSW and treated under either a dry or wet anaerobic process, Mata-Alvarez (2002) observed a transition from mesophilic to thermophilic conditions (35–55 °C in 10 d) requiring significant variations in organic loading (from 15% to 40% over 2 d) without permanent effect on the process performance.

The objective of this experiment was therefore to test the development of a thermophilic anaerobic inoculum for the treatment of food waste, by subjecting anaerobic mesophilic sludge to a fast temperature change. The adaptation was conducted using a bench scale reactor seeded with a mesophilic granular sludge obtained from two Upflow Anaerobic Sludge Blanket digester (UASB), one treating apple wastewater while the other treating milk wastewater. Substrate activity tests were conducted to characterize the changes in biomass, and the new thermophilic inoculum was fed selectively collected OFMSW (SC-OFMSW) to test its treatment capacity.

2. Methodology

The project consisted in achieving the adaptation of:

- (1) A granular UASB biomass from mesophilic to thermophilic conditions, and;
- (2) The resulting thermophilic biomass to a new type of substrate (food waste) at different loading rates, under completely mixed and thermophilically controlled anaerobic conditions using a bench scale reactor.

2.1. Substrate

The SC-OFMSW used in this study was the food waste collected from the solid waste produced by a cafeteria. Immediately after collecting the cafeteria solid waste, the organic fraction was manually sorted and the plastic, glass, metal and wood portions were removed. Despite manual sorting, the substrate was mostly made up of food scraps (meat, egg, fruit and vegetables, bread, pastries), but also contained some paper napkins and packaging, plastic packaging, wood sticks and plastic table-settings.

The resulting substrate was shredded using a 50 mm stainless-steel blade and then chopped into a paste using an 87/Polytron/PTA 36/2 blade connected to a CH-6010 Kiennz-Lu electric rotor (*Kinematica*, Switzerland). All particles larger than 4.75 mm were separated with a No. 4 USA standard testing sieve (*W.S. Tyler*, USA). During the first shredding operation, tap water was added to drop the solid content to 21.8%. The meat and egg fractions were visibly high explaining the high Total Kjeldahl Nitrogen (TKN) of 17% and carbon/nitrogen (C/N) ratio of 3:1 associated with the substrate (Table 1).

2.2. Seed

A large number of sludge types were possible for this study, but two types of anaerobic granular sludge were readily available and were therefore selected. The seed selected consisted of a 50/50% mixture of granular and non-granular sludge taken from two separate-UASB reactors treating dairy process wastewater and apple-juice

process wastewater, respectively. The total and volatile solids are typical for anaerobic granular sludge (Table 1).

2.3. Reactor set-up

The experimental 5 L glass reactor is illustrated in Fig. 1. The reactor's working volume was 4 L. The system included a 6–600 peristaltic pump (Cole-Parmer, Chicago, IL) and a Type-RZR50 stirrer (Cafra, Warton, ON). The temperature was automatically controlled using a heating system surrounding the digester and encompassing a 305 mm-LT-type probe (Cole-Parmer, Vernon Hill, IL), a Digi-Sense standard model temperature controller (Cole-Parmer, Vernon Hill, IL) and a Brisk Heat silicone extruded, flexible electric heating tape (Thermolyne, Dubuque, IO). The pH monitoring system consisted of

an epoxy-pH probe (Cole-Parmer, Niles, IL) and an Accumet pH-meter model 825 MP (Fisher, USA). After recovering the condensate, the biogas generated was sampled through the biogas sampling port and measured by means of a Wet Tip-like water displacement system.

2.4. Experimental stages

The first stage of the experiment took 20 days (stage C0) and consisted in adapting the seed to thermophilic conditions. At the beginning of stage C0, the reactor was seeded with 4 L of the seed mixture (Table 1) and operated at 35 °C to generate inoculum, without exogenous substrate supply. On day 13, the process temperature was instantaneously increased from 35 to 55 °C, and maintained at 55 °C until day 20.

This C0 stage was followed from day 21 to 84, by the thermophilic batch stages C1, C2, and C3 each lasting 21 d, to assess the treatment capability of the new biomass at 55 °C and at various SC-OFMSW loading rates. At the start of each stage C1, C2 and C3, the SC-OFMSW was fed with a food to microorganism ratio (F/M) increasing sequentially along with the stages, from 0.12 to 4.43 gVS/gVSS. Because of medium acidification at the beginning of C3, a fourth 21-days stage (C4) was carried out to restore methanogenic conditions. At the beginning of stage C4 (day 84), the reactor biomass received a mixture of 1.6 L of sludge produced at the end of stage C2, 0.3 L of the original seed blend and 0.25 L of a 0.2 M NaHCO₃ solution at pH of 8, to restore the reactor's pH to 7.0. Biomass activity was monitored during stage C4 by measuring gas production and subjecting the final (21 d) sludge to substrate activity (SA) tests.

Table 1
Experimental substrate (SC-OFMSW) and seed characteristics

Parameter	SC-OFMSW	Seed mixture
Density (kg/L)	1.22	1.042
Solid (%)	21.8	4.2
TS (g/L)	218.2	42.1
VS (% TS)	95.1	67.0
SS (g/L)	161.8	38.1
VSS (% SS)	96.5	70.0
Total COD (g/gVSS)	1.16	0.97
Soluble COD (% ICOD)	n.e.	12.12
TKN (% TS)	17	n.e.
C/N	3.1	n.e.

Notes: n.e., not evaluated.

The seed mixture consisted of 50% UASB sludge treating apple wastewater and 50% UASB sludge treating milk wastewater.

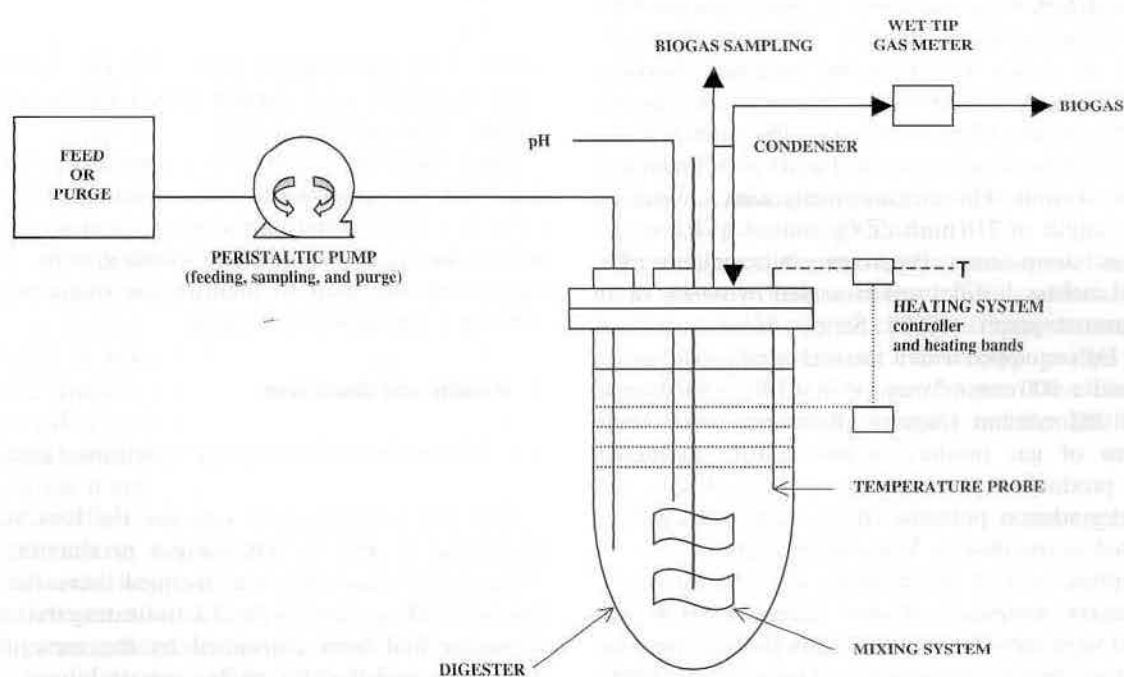


Fig. 1. Schematic diagram of the batch reactor system.

During the complete 104 d experimental period, gas production was continuously measured and samples were taken at regular 24 h intervals. A sludge sample was obtained from the reactor on days 0 and 20 of stage C0, and used for substrate activity tests (SA). Both samples were incubated at 35 and 55 °C during such SA tests to check for the presence and character of the remaining mesophilic and thermophilic microorganisms. Sludge samples were also collected from the reactor at the end of stages C1, C2, C3 and C4, for additional SA conducted under an incubation temperature of 55 °C.

2.5. Analytical methods

To calculate the F/M ratio supplied during stages C1–C3, the substrate, the seed sludge and the reactor's biomass were analyzed for chemical oxygen demand (COD), total and suspended solids (TS and SS), and total and suspended volatile solids (VS and VSS), according to standard methods (APHA et al., 1995).

Acetic, propionic and butyric acids were measured using an Agilent 6890 gas chromatograph (Wilmington, DE) equipped with a flame ionization detector of 0.2 µL. The samples were fortified at a ratio of 1:1 (V/V) using an internal standard of *iso*-butyric acid dissolved in 6% formic acid and directly injected into a glass column of 1 m × 2 mm *Carbopack C* (60–80 mesh) coated with 0.3% *Carbowax* 20 M and 0.1% H₃PO₄. The column was held at 130 °C for 4 min and helium as carrier gas was injected at a rate of 20 mL/min. The injector and the detector were both maintained at 200 °C.

For the SA tests, glucose was measured using an HPLC (Waters Chromatography Division, Milford, MA, USA) equipped with an injector (model 717+), a photodiode array (PDA) detector (model 2996), a pump (model 600) and a 2414 refractive index detector. The separation unit consisted of an ICsep ICE-ION-300 column (*Transgenomic*, San Jose, CA) of 300 mm × 7.8 mm id and an ion guard GC-801 column (*Transgenomics*). The mobile phase consisted of 0.035 N of sulfuric acid at a pH of 4, flowing at a rate of 0.4 mL/min. The measurements were conducted using a wavelength of 210 nm.

The biogas composition (hydrogen, nitrogen, oxygen, methane, and carbon dioxide) was measured by means of an HP gas chromatograph (68900 Series, *Hewlett-Packard*, Wilmington, DE) equipped with a thermal conductivity detector (TCD) and a 900 mm × 3 mm (36 in × 1/8 in) 60/80 mesh Chromosorb 102 column (*Supelco*, Bellefonte, PA). Daily determinations of gas production rate (GPR: L_{STP}/L/d), ultimate gas production potential (Go in L_{STP}/gVS_{fed}), and ultimate biodegradation potential (Bo in L_{STP}–CH₄/gVS_{fed}) were calculated as specified by Mata-Alvarez (2002).

The mesophilic and/or thermophilic microbial diversity of the biomass produced during stages C0–C4 was evaluated through substrate activity tests (SA) carried out in 60 mL serum bottles according to Guiot et al. (1995). Triplicate bottles were inoculated with biomass diluted

with a phosphate buffer to obtain 2000 mg VSS/L for hydrogenophilic activity and 5000 mg VSS/L for the bottles used for measuring activity of glucose, acetic and propionic acid. The bottles were flushed with N₂/CO₂ (80%/20%) to remove all oxygen and incubated in a rotary shaker (*New Brunswick Scientific Co.*, Edison, NJ). While being incubated at either 35 or 55 °C, the serum bottles and their content were shaken at 100 rpm for the soluble substrate tests and at 400 rpm for the hydrogen substrate test.

2.6. Statistical analysis

For the C0 stage, the statistical significance for the biomass activity (SA) before and after the temperature shift and incubated at temperatures of 35 and 55 °C, was determined using a *t*-test with paired data (Wardlaw, 1985). The SA results obtained were tested for significant differences using:

$$t = \frac{h(N)^{1/2}}{Sh} \quad (1)$$

where *t* is the value of student *t*-found, *N* the test number of pairs, *h* the mean of differences of SA-values of each bottle-pair (mesophilic and thermophilic) and *Sh* the standard deviation of *h*.

The *t*-found value was compared to that of a Student *t*-Statistic table for a degree of freedom of 2 (2 d.f.).

Likewise, the statistical significance of the SA measured under incubation at 55 °C for samples of biomass taken at the end of C1, C2 and C3 was determined by a one-way analysis of variance or *F*-test (Wardlaw, 1985). The values obtained from the SA of these sludge samples were tested for significant difference using the following equation:

$$F = \frac{Vb}{Vw} \quad (2)$$

where *F* is the variance ratio, *Vb* the between groups variance (stages C1, C2 and C3) and *Vw* the within groups variance (between triplicates).

The *F*-value was compared at a critical *F*-value table for 2 d.f. between groups and 6 d.f. within groups. A multiple range test for comparing means in an analysis of variance was evaluated for significance within groups. The multiple range test was used to identify the treatment(s) demonstrating a significant difference.

3. Results and discussion

3.1. Mesophilic to thermophilic transition (stage C0)

The biogas production rate for the first stage, C0, is illustrated in Fig. 2. The biogas production peaked at 1.9 L_{STP}/day after 24 h, but dropped thereafter to reach a low of 0.32 L_{STP}/day on day 13, indicating that all available substrate had been consumed by the mesophilic microorganisms and that the sludge was stabilized. On day 13, the system's temperature was instantaneously increased

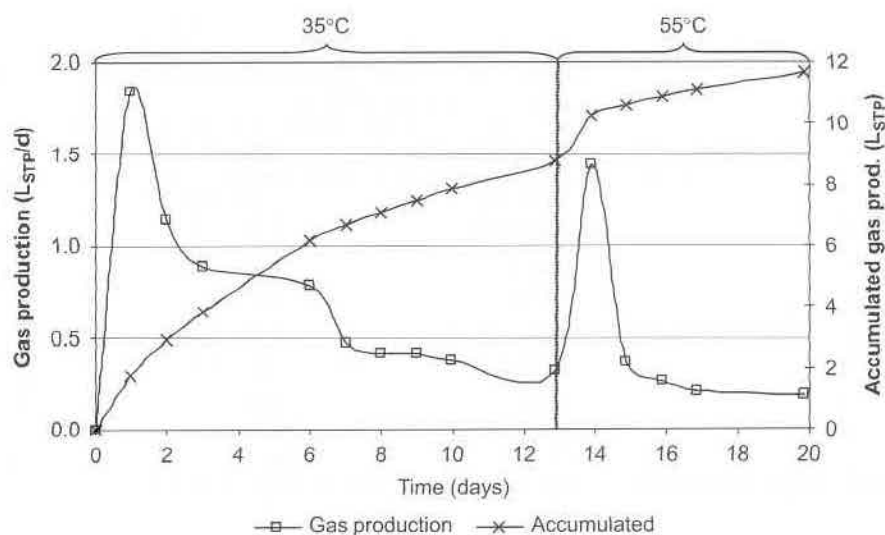


Fig. 2. Gas production profile during the C0 stage.

from 35 to 55 °C and the reactor's content reached 55 °C within 10 minutes. Thermophilic conditions were maintained thereafter.

After 22 h of thermophilic operation at 55 °C, 1.4 L_{STP} of gas was produced, only 50% of which was produced by the microbial activity. The other 50% biogas volume resulted from the increase in temperature of the reactor and the lower CO₂ water solubility. Accordingly, this temperature change also affected the water solubility of carbon dioxide and the methane fraction of the biogas composition which dropped after day 1 from 65% to 38% and came up slightly to reach 41% on day 20. Because methane was still being produced, the methanogenic microorganisms had survived the temperature transition.

During the mesophilic phase (from day 0 to 12), the volumetric production of methane varied between 0.27 and 0.05 L_{STP}/L/d, while during the thermophilic phase (day 13–20), the volumetric production of methane varied between 0.14 and 0.02 L_{STP}/L/d. Therefore, the temperature transition was successful in extracting a slight but additional volume of gas from the food waste.

The shift of microbial population within the sludge collected at the beginning and at the end of the seeding stage (C0) was evaluated by means of SA tests. The SA values obtained at both 35 and 55 °C demonstrated that increasing the temperature did not produce a pasteurization effect on either bacterial or methanogenic populations. Except for hydrogenophilic methanogens, the thermophilic population of the 0 and 20 d sludge showed no significant difference ($P > 0.10$) when comparing the SA values, but the abrupt change in reactor temperature, from 35 to 55 °C on day 13, resulted in a positive selection of thermophilic fermentative bacteria (Table 2). Thermophilic glucose consuming bacteria were observed to almost double their activity from 1.85 to 3.58 g Glucose/gVSS/d, from day 0 to 20.

For the mesophilic glucose consuming bacteria, their activity was significantly reduced ($0.05 > P > 0.02$) by 58% on day 20 after being exposed to 55 °C, and as compared to day 0. Although not significantly different ($P > 0.10$), mesophilic or thermophilic propionic acid consuming bacteria reduced their activity from 0.09 to 0.06 and from 0.03 to 0.02 g propionate/gVSS/d, respectively, after increasing the reactor temperature to 55 °C.

Similarly, the activity of the hydrogen consuming mesophilic methanogens was significantly reduced ($0.05 > P > 0.02$) during stage C0, because by the end of day 20, their SA dropped 100 fold to a final value of 0.01 g Hydrogen/gVSS/d. Although not statistically different ($0.10 > P > 0.05$), thermophilic hydrogen-consuming methanogens also reduced their SA from 2.52 to 0.26 g Hydrogen/gVSS/d. Interestingly, mesophilic acetoclastic methanogens were not affected by the abrupt change of temperature, because they doubled their SA from 0.62 to 1.18 g Acetate/gVSS/d, likely as a result of starvation. Without being significantly different, thermophilic acetoclastic methanogens slightly reduced their activity from 0.6 to 0.47 g Acetate/gVSS/d.

In summary, the mesophilic seed used for this experiment already had a thermophilic potential which was expressed by the sudden change of temperature carried out on day 13. Nevertheless, the trophic groups which were significantly affected by the temperature change were the mesophilic fermentative bacteria and mesophilic hydrogen-consuming methanogens.

3.2. Effect of F/M on thermophilic fermentation (stages C1–C3)

Following the C0 stage, the SC-OFMSW-feeding stage was initiated through three batch cycles (C1–C3) during which the substrate was supplied every 21 days using a F/M ratios of 0.12, 1.15 and 4.43 gVS/gVSS (Table 3a).

Table 2

Substrate activity for the C0 initial and adapted sludge measured at 35 and 55 °C

Substrate	Incubation T (°C)	Activity (g substrate/gVSS /d)		Probability
		Initial sludge (day 0)	Adapted sludge (day 20)	
Glucose	35	5.09±1.82	2.12±1.09	>0.02
	55	1.85±0.49	3.58±2.06	>0.10
Propionate	35	0.09±0.02	0.06±0.006	>0.10
	55	0.03±0.02	0.02±0.010	>0.10
Acetate	35	0.62±0.4	1.18±0.39	>0.10
	55	0.60±0.11	0.47±0.51	>0.10
Hydrogen	35	1.04±0.39	0.01±0.001	>0.02
	55	2.52±1.12	0.26±0.020	>0.05

Notes: Initial sludge = mesophilic sludge; adapted sludge = sludge adapted after 7 days of operation at 55 °C.

Table 3a

Operation conditions and reactor performance from C1, C2 and C3

Stage	F/M (gVS/gVSS)	Average pH	Go (L _{STP} /gVS _{fed})	Bo (L _{STP} CH ₄ /gVS _{fed})
C1	0.12	—	1.40	0.84
C2	1.15	7.30	0.89	0.61
C3	4.43	5.17	0.20	0.01

Notes: Go: Ultimate gas production.

Bo: Ultimate methane production.

Table 3b

Calculation of FM and equivalent OLR during C1, C2 and C3

Stage	Vol. fed ^a (L)	Vol. sludge ^b (L)	Vol. r ^(a+b) L	VSS sludge (gVSS/L)	F/M (gVS/gVSS)	OLR (gVS/L ^b /day)
C1	0.05	3.95	4.0	21.88	0.12	0.12
C2	0.4	3.60	4.0	20.05	1.15	1.09
C3	1.4	2.60	4.0	25.22	4.43	5.32

The concentration of VS in the SC-OFMSW was constant (207.508 gVS/L).

For the OLR calculation, the amount of SC-MSW^a fed was considered equivalent to that applied continuously over the 21 d cycle.

The average process temperatures were controlled at 55.2±0.1, 55.3±0.1 and 56.5±2.5 °C, respectively.

The reactor biomass accommodated the low and intermediate F/M ratios (C1 and C2), which were calculated as equivalent to a constant loading rate for basis of comparison with other research projects (Table 3b). Although applied at the start of the stage using a F/M of 0.12 gVS/gVSS, the equivalent constant VS loading rate (OLR) applied uniformly over the 21 d period, was 0.12 gVS/L/d for C1. This OLR is low compared to the typical organic load of 3.6 gVS/L/d suggested by ten Brummeler (2000) for continuous systems. The intermediate (C2) F/M ratio was increased ten fold to 1.15 gVS/gVSS (OLR of 1.09 gVS/L/d) and still did not negatively affect the reactor biomass, since the gas production increased substantially from 11 L_{STP} (C1) to 56 L_{STP} during stage C2. The F/M ratio of 4.43 gVS/gVSS applied during stage C3

produced two short burst of gas production: the first of 106.3 L_{STP}/d occurred 6 h after feeding, and; the second of 15.95 L_{STP}/d occurred two days after feeding (Fig. 3).

Within two days, the F/M ratio of stage C3 acidified the reactor medium and its pH fell from 7.75 to 5.45 (Fig. 4), as a result of accumulating volatile fatty acids (VFA). Within the first day of C3, the concentrations of acetate, propionate and butyrate reached levels of 4580, 65 and 7000 mg/L, respectively. By the end of C3, only 45 and 40% of the acetate and butyrate, respectively, were consumed, while the propionate concentration remained unchanged.

Stages C1 and C2 resulted in ultimate gas (Go) and methane production (Bo) values exceeding those of 0.85 L/gVS_{fed} and 0.46 L-CH₄/gVS_{fed}, respectively, for SC-OFMSW degradation, presented by Mata-Alvarez (2002). Nevertheless, the acidification of the reactor during

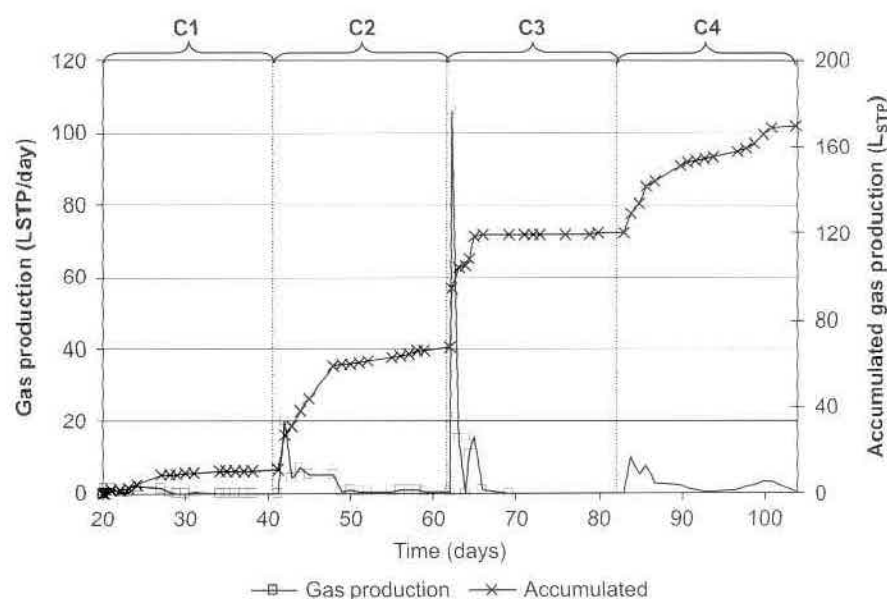


Fig. 3. Daily and accumulated gas production profile from stages C1 to C4.

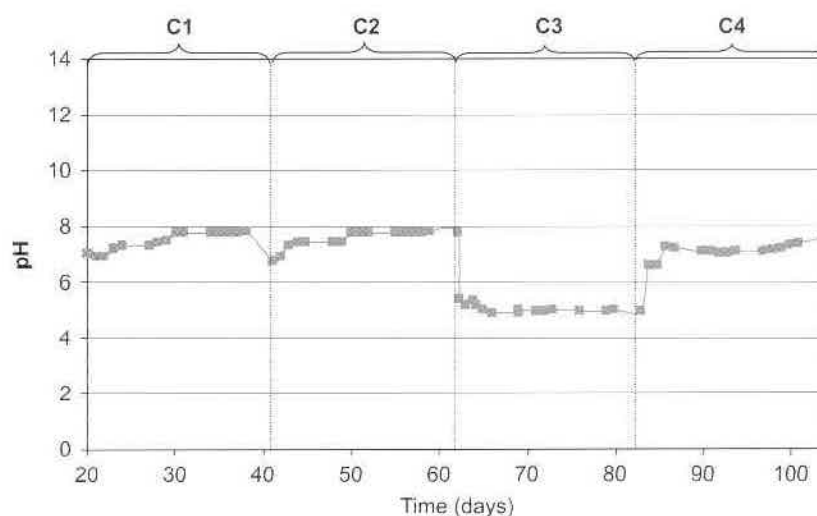


Fig. 4. Daily pH profile from stages C1 to C4.

stage C3, following a high F/M application, reduced the biogas-production capacity and both values for G_0 and B_0 dropped from 0.89 to 0.2 L_{STP}/gVS_{fed} and from 0.61 to 0.01 $L_{STP}CH_4/gVS_{fed}$, respectively.

Fig. 5 shows the profile of the gas composition throughout C1–C4. The concentration of methane increased from 60% during stage C1 to 64% during stage C2. During stage C3, methanogenesis was displaced towards hydrogen production because on day 64, the biogas contained 30% hydrogen suggesting strong fermentative activity. Notably, a decrease in total gas production rate from 106.3 to 15.95 L_{STP}/d was observed during this period (Fig. 3). To check for hydrogen production on day 65, the reactor's headspace was replaced with a mixture of inert gas (80% N_2 /20% CO_2) and biogas H_2

content was once more monitored. Hydrogen was immediately observed again reaching 55% by the end of day 69. In general, methanogenic activity and anaerobic digestion are tightly regulated by the production of H_2 . For all organisms involved and to conserve energy, the anaerobic media must sustain a H_2 partial pressure within the window of 6×10^{-5} – 10^{-2} atm (Ferry, 1993). Accordingly, the H_2 produced during C3 resulted either from the consumption of rapidly available carbohydrates by fermentative bacteria or the degradation of butyric acid by syntrophic bacteria; more likely the first burst of H_2 , accompanied by a small volume of methane, was related to the action of hydrogen-producing, carbohydrate-fermenting bacteria. After stripping H_2 from the media and the reactor headspace on day 65, the second

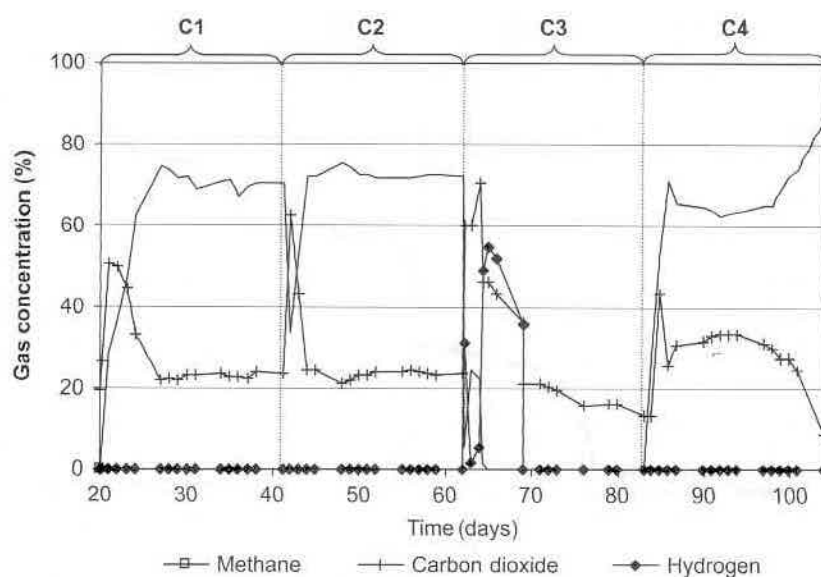


Fig. 5. Daily gas composition from stages C1 to C4.

Table 4

Substrate activity/evolution of thermophilic sludge during the adaptation stage

Sludge	Substrate activity (g/gVSS/d)			
	Glucose	Propionate	Acetate	Hydrogen
C1	0.21 ± 0.27^b	0.011 ± 0.006^c	0.08 ± 0.01^d	0.23 ± 0.06^f
C2	2.27 ± 0.98^a	0.005 ± 0.017^c	0.88 ± 0.12^d	0.36 ± 0.04^e
C3	1.93 ± 0.86^a	0.007 ± 0.019^c	-0.005 ± 0.09^e	-0.17 ± 0.07^f

Notes: Values with the same letter are not significantly different within groups at the level of $P \leq 0.05$.

stage of H_2 production was largely related to syntrophic activity, since fermentative bacteria could not be inhibited. While degrading butyric acid, *Syntrophomonas wolfei* produces propionate, H_2 and CO_2 (McCarty and Mosey, 1991). Since during the second burst of H_2 after day 65, there was no uptake of H_2 and CO_2 for the production of methane, the build up of H_2 and its media partial pressure thus inhibited syntrophic activity. Notably, after day 71 and until day 84, gas production ceased completely (Fig. 3) and no more H_2 was observed in the headspace of the reactor (Fig. 5).

3.3. Substrate activity test for thermophilic fermentation (stages C1–C3)

The SA tests conducted on the biomass collected at the end of stages C1, C2 and C3 indicated that the F/M ratio affected the capability of the biomass to adapt to different substrates (Table 4). Among all three stages, C2 improved the fermentative biomass activity while C3 reduced it by more than 50%. Hydrogen production rather than uptake was observed with the C3 biomass indicating that H_2

producing fermentative bacteria outpaced the hydrogenophilic capacity of the biomass to convert H_2 into methane. Similarly, the uptake of acetate for the production of methane was affected by the C3 biomass because the activity of acetoclastic methanogens dropped from 0.88 during C2 to -0.005 g Acetate/gVSS/d.

The analysis of variance conducted on the SA values for stages C1, C2, and C3 suggested that the selected F/M program significantly affected the fermentative (glucose degrading) bacteria (Table 4), where stage C1 produced significantly lower values than stages C2 and C3 ($P < 0.05$). As well, both acetoclastic and hydrogenophilic methanogenesis were specially affected by the F/M supplied ($P < 0.01$), where stage C2 produced significantly higher values than stages C1 and C3. Finally, there was no significant difference in syntrophic-propionic degrading activity ($P > 0.05$), as all stages had activity results under 0.015 g of propionic acid/g VSS/d.

3.4. Restoration of thermophilic sludge (stage C4)

After stage C3 and on day 84, a new non-substrate cycle (C4) was started to restore the pH of the medium and the methanogenic activity. For this purpose, 2.15 L of the reactor's acidified medium were replaced with a mixture of 1.6 L of sludge produced at the end of stage C2, 0.3 L of the original seed blend, and 0.25 L of a 0.2 M $NaHCO_3$ solution at pH 8. At the end of this 21-day stage (C4), the biomass was producing gas (Fig. 3) and the pH was restored to 7.0 (Fig. 5).

With C4, both fermentative (glucose consumption) and acetoclastic methanogenesis activities were recovered, giving SA values of 0.6 and 0.26 g substrate/gVSS/d. The propionate activity was recovered and measured 0.09 g Propionic/gVSS/d while the hydrogenophilic methanogenesis activity was at a low value of 0.02 g H_2 /gVSS/d.

4. Conclusions

The present experiment demonstrated that a mesophilic wastewater-treating anaerobic sludge can be easily and quickly adapted for the treatment of the organic fraction of municipal solid waste (OFMSW) under thermophilic conditions. Nevertheless, the organic load program must be carefully selected and monitored, since the methanogenic activity can easily be displaced towards a fermentative activity. The usual, time-consuming acclimatization stage can be avoided when starting-up the thermophilic anaerobic treatment of OFMSW. Favorable conditions were quickly achieved (20 d) and the substrate feeding stage could be initiated immediately afterwards.

The instantaneous upgrading procedure from mesophilic to thermophilic conditions did not have a pasteurization effect on anaerobic microorganisms, but positively selected the mesophilic acetoclastic methanogens and mesophilic fermentative bacteria. However, the specific activity of both mesophilic and thermophilic propionate-oxidizing syntrophs and hydrogenophilic methanogens was reduced.

Under batch conditions, the reactor could sustain a moderate F/M ratio of 1.5 gVS/gVSS while a high F/M ratio of 4.43 gVS/gVSS displaced methanogenesis towards H_2 production. Although H_2 production was not the goal of this process development, its presence is of interest as an alternative energy source. More research is needed with that respect, as well as repeating the test under continuous conditions.

Acknowledgments

The technical assistance of A. Corriveau and S. Deschamps is gratefully acknowledged. This research was supported by the National Research Council of Canada and the Natural Science and Engineering Research Council of Canada. Additional support was

granted by the Mexican Council of Science and Technology (CONACYT in Spanish), towards a graduate scholarship for Mr. L. Ortega.

References

- Baldasano, J.M., Soriano, C., 2000. Emission of greenhouse gases from anaerobic digestion processes: comparison with other municipal solid waste treatments. *Water Science and Technology* 41 (3), 275.
- De Baere, L., 2000. Anaerobic digestion of solid waste: state-of-the-art. *Water Science and Technology* 41 (3), 283.
- Ferry, J.G., 1993. *Methanogenesis: Ecology, Physiology, Biochemistry & Genetics*. Chapman & Hall, USA.
- Guiot, S.R., Safi, B., Frigon, J.-C., Mercier, P., Mulligan, C., Tremblay, R., Samson, R., 1995. Performances of a full-scale novel multiplate anaerobic reactor treating cheese whey effluent. *Biotechnology and Bioengineering* 45, 398–405.
- Mata-Alvarez, J., 2002. *Biomethanization of the Organic Fraction Municipal Solid Wastes*. IWA Publishing, UK.
- McCarty, P.L., Mosey, F.E., 1991. Modeling of anaerobic digestion processes (a discussion of concepts). *Water Science and Technology* 24 (8), 17–33.
- Nielsen, B., Petersen, G., 2000. Thermophilic anaerobic digestion and pasteurization. Practical experience from Danish wastewater treatment. *Water Science and Technology* 42 (9), 65.
- Nozhevnikova, A.N., Kostsyurbenko, O.R., Parshina, S.N., 1999. Anaerobic manure treatment under extreme temperature conditions. *Water Science and Technology* 40 (1), 215.
- Rhynder, C.R., Schwartz, L.J., Wenger, R.B., Kohrell, M.G., 1995. *Waste Management and Resource Recovery*, first ed. Lewis Publishers (CRC Press), USA.
- Standard methods for the examination of water and wastewater, 1995, 19th ed. American Public Health Association/American Water Works Association/Water Environmental Federation, Washington, DC, USA.
- ten Brummeler, E., 2000. Full scale experience with the Biocel process. *Water Science and Technology* 41 (3), 299.
- US Environmental Protection Agency (USEPA), 2003. Landfill Methane Outreach Program. Available from the web via the Internet <www.epa.gov/lmop/about.htm#ifgte>.
- Wardlaw, A.C., 1985. *Practical Statistics for Experimental Biologists*. Wiley, Great Britain.

Groundwater risk assessment at a heavily industrialised catchment and the associated impacts on a peri-urban wetland

Elias Dimitriou^{a,*}, Ioannis Karaouzas^a, Konstantinos Sarantakos^b, Ierotheos Zacharias^c,
Konstantinos Bogdanos^a, Aristidis Diapoulis^a

^aHellenic Centre for Marine Research, Institute of Inland Waters, 46.7 km of Athens – Sounio Av., P.O. Box 712, 19013 Anavissos, Attika, Greece

^bHellenic Centre for Marine Research, Institute of Oceanography, 46.7 km of Athens – Sounio Av., P.O. Box 712, 19013 Anavissos, Attika, Greece

^cDepartment of Environmental and Natural Resources Management, University of Ioannina, 2 Seferi Str., 30100 Agrinio, Greece

Received 27 June 2006; received in revised form 5 December 2006; accepted 16 March 2007

Available online 17 May 2007

Abstract

Industrial and agricultural activities often impose significant pressures to the groundwater quality and consequently degrade wetland ecosystems that depend mostly on subsurface water flow. Groundwater vulnerability and risk mapping is a widely used approach to assess the natural protection of aquifers and the associated pollution potential from human activities. In the particular study, the relatively new Pan-European methodology (COP method) has been applied in a highly industrialized peri-urban wetland catchment, located close to Athens city, to map the intrinsic vulnerability of the aquifer and evaluate the risk potential originating from local land uses. Groundwater analysis results for various parameters, including Phenols, PCBs and nutrients, have been used to validate the vulnerability and risk estimations while a biological assessment occurred to associate the mapping results with the wetland's ecological status. The results indicated that even though the natural protection of the aquifer is relatively high due to the dominant hydrogeologic and geomorphologic conditions, the groundwater pollution risk is considerable, mainly because of the existing hazardous land uses. The water quality of the groundwater accredited these findings and the ecological status of this peri-urban wetland also indicated significant impacts from industrial effluents.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Risk assessment; Groundwater vulnerability; Wetland; Pollution; Hydrogeology

1. Introduction

Groundwater pollution is one of the most common environmental problems nowadays (Fatta et al., 2002). Leachates from landfills and uncontrolled industrial effluents disposal cause significant deterioration in the water quality of many aquifers throughout the world (Gailey and Gorelick, 1993; Kjeldsen et al., 1993). Vulnerability and risk mapping for groundwater resources is a widely used approach for determining areas with high pollution potential and delineating aquifer protection

zones (Babiker et al., 2005). Several methodologies have been developed aiming to estimate groundwater vulnerability and pollution risk under different hydrogeological and socioeconomic conditions (Tesoriero et al., 1998; Lake et al., 2003; Al-Adamat et al., 2003). However, a universal approach has not been developed yet. Intrinsic vulnerability is a term first used in the early 1960s to describe the protective function provided by the overlying layers, the geomorphology and the hydrological regime to the main aquifer of an area (Vrba and Zaporozec, 1994). Hydrogeologic models, statistical methods and tracer surveys are often implemented to evaluate vulnerability assessment results while the development of state-of-the-art cartographic tools, such as GIS facilitate and accelerate the entire process (Babiker et al., 2005; Dixon, 2005). Groundwater risk mapping combines hazard originating from land uses and vulnerability as a physical property to assess the

*Corresponding author. Tel.: +30 2291076389; fax: +30 2291076419.

E-mail addresses: elias@ath.hcmr.gr (E. Dimitriou),
ikarz@ath.hcmr.gr (I. Karaouzas), ksarant@ath.hcmr.gr
(K. Sarantakos), izachari@cc.uoi.gr (I. Zacharias), ebogda@ath.hcmr.gr
(K. Bogdanos), adiapo@ath.hcmr.gr (A. Diapoulis).

aquifer's pollution potential. COP methodology is a relatively new approach for vulnerability and risk mapping developed under the COST Action 620 EU project (COST 620, 1998; Goldscheider et al., 2003) targeting in the protection of carbonate aquifers (Andreo et al., 2006). In this study, the particular methodology has been applied to assess the pollution risk for surface and underground water bodies in a peri-urban wetland area that receives pressures from industrial and agricultural activities as well as from urban waste disposal. Recent water quality and ecological assessment results have been used to validate the risk mapping outcome and evaluate its applicability on the local conditions while the impacts imposed on this peri-urban wetland have been illustrated.

2. Material and methods

2.1. Description of the study area

Koumoundourou Lake catchment is a suburban area located in the western part of Attica region. It has an extent of 39.8 km², the elevation ranges from 0 to 540 m and the mean slope of the catchment is 8%. The study catchment is an industrial area that hosts oil refinery companies, asbestos factories and the main landfill site of Attica region (population served: approx. 3 million people) at about 12 km distance from Athens city centre.

Koumoundourou Lake is located inshore of Elefsis Bay (Fig. 1), has a surface elevation of 1.41 m above sea level, covers an area of 147 000 m² (Zacharias et al., 2003) and its shoreline perimeter is approximately 1700 m. The mean depth of the lake is 1.5 m and the main water recharge occurs through important submerged springs on the northwest part of the lakeshore. Seawater intrusion, freshwater from the springs and rainfall are mixed, producing a meromictic brackish environment (10.50‰ average salinity levels during December 2002) (Zacharias et al., 2003). The geological formations of the lake's catchment comprise Quaternary deposits covering a surface extend of 20 km², which overly moderately karstified limestone and dolomite formations that have a coverage of 19.5 km² (Zacharias et al., 2003). Below the calcareous formations Neopalaiozoic impermeable schists and graywackes exist that appear only in limited surface outcrops of 0.3 km².

In the southern side of the lake a controlled flow weir is installed that outflows excess water towards the sea to avoid flooding events in the adjacent highway (Athens–Corinth) that separates the sea from the lake shoreline. The annual rainfall height is approximately 360 mm while 70% of this value is recorded during the period November to February. The mean annual evaporation is approximately 1630 mm (Conides and Parpoura, 1997). The climate of the study region is characterised as Mediterranean subtropical due to its short annual temperature range. The combination of increased air temperatures due to the intensified

industrialization of the area and the location of the lake (being close the sea), create an even milder microclimate.

The lake is located in a highly industrialized area surrounded by many significant pollution sources. On the north side of the lake an army camp is located, which serves as an oil supply depot station. On the north-west side the national oil refinery plant is based while on the south-west side a marble-cutting small industry is found. Oil spills and leaks from both the refinery and the military depot form thin layers on the water surface and contaminate underground water and aquifers. Leachates derived from the major landfill site of Athens, situated at the north side of the lake around 7 km away, contribute further to the degradation of the lake and groundwater quality of its catchment (Fatta et al., 1999) (Fig. 1).

2.2. Groundwater vulnerability and risk assessment at a catchment level

For the groundwater intrinsic vulnerability and risk assessment, the COP method was used (Andreo et al., 2006), which has been developed under the European COST Action 620 project aiming to the protection of karstic aquifers. The specific methodology can also be implemented in other types of aquifers apart from karstic and is based on the concept of classifying the study area in 5 vulnerability zones (very low to very high) according to the local geomorphology, the properties of the overlying soils and lithological formations as well as the existence or absence of vegetation (Andreo et al., 2006). Koumoundourou lake catchment has been digitized with the contribution of ArcView program (GIS package) and topographic maps of scale 1:25 000 acquired from the Hellenic Military Service. The grid size used for map calculations was 25 m. At the first stage of this study the concentration of flow map (C map) has been developed by assigning weighting values in the grids of the base map describing the slope, vegetation cover and the karstic features of the area as suggested in the aforementioned method (Andreo et al., 2006). This map expresses how fast the surface water and thus the potential pollutants as well, reach the aquifer. In this index the karstic features of the aquifer are described as developed/scarcely developed karst, fissured carbonate and non karstic terrain while the existence of permeable/impermeable surface layers plays also a role in the weighting assignment. Then, the protective function provided by the overlying layers of the aquifer has been depicted in the relevant map (O map) by examining the soil type and thickness as well as the lithological composition and depth of the unsaturated zone. The precipitation component of the vulnerability map which may reduce natural protection is expressed by the P map incorporating both the quantity and intensity of precipitation and their spatial distribution in the study area. Four meteorological stations have been used, homogeneously distributed around and inside the study

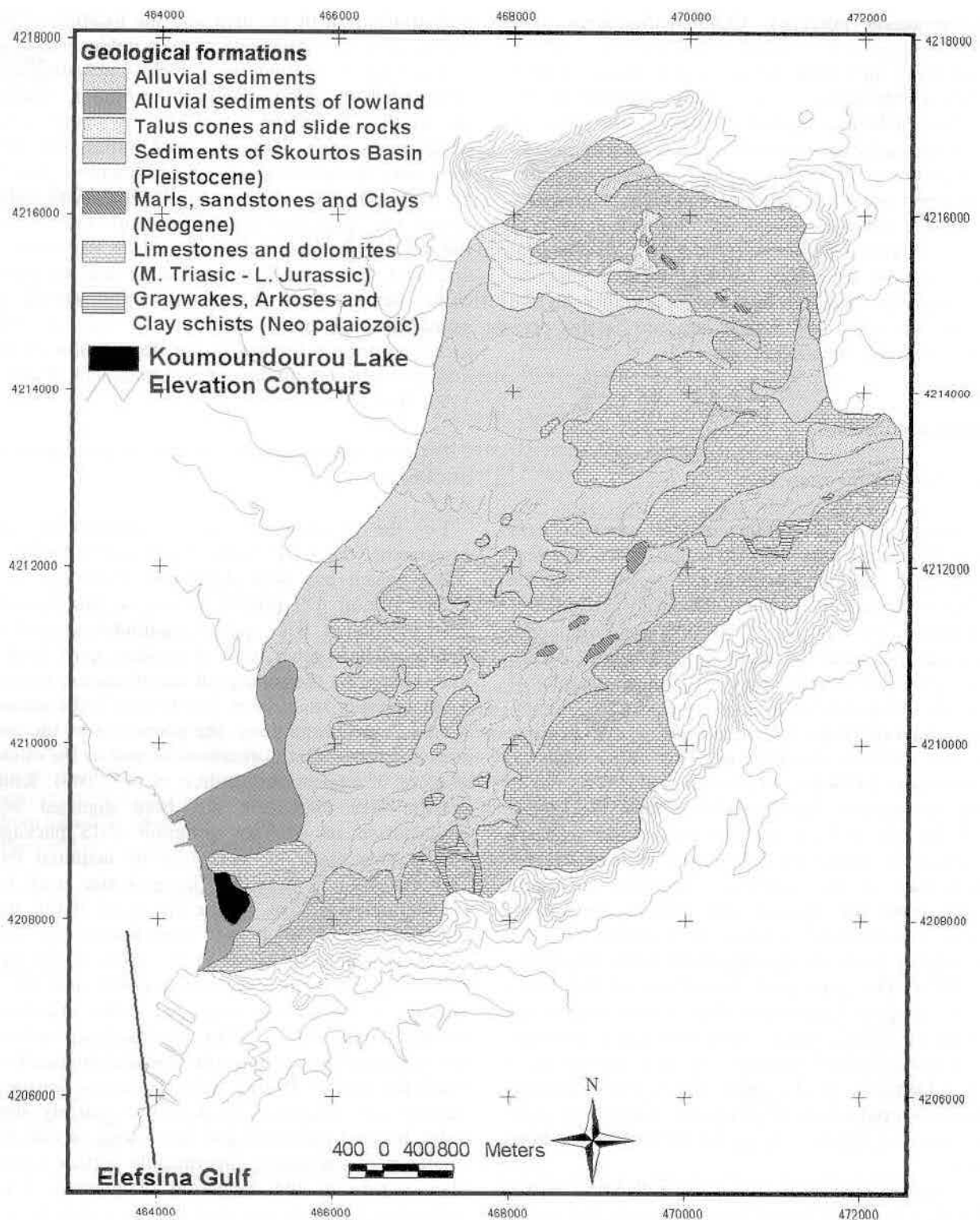


Fig. 1. Geological map of Koumoundourou Lake catchment.

catchment and average annual rainfall values of the period 1990–2000 have been interpolated by applying the krigging method (Kitanidis, 1997). The vulnerability map (COP map) has been developed by multiplying the above maps through the Spatial Analyst component of ArcView program.

For the hazard map, aerial photos of the year 2000 have been acquired, digitized and imported in Image Analysis GIS module. Supervised interpretation of the aerial photos was performed by using past land use maps of the area and ground truth visits. A detailed inventory of the industrial installations of the study catchment has been developed

and the land use map of the area has been transformed to hazard map by assigning hazard weighting values in each land category polygon according to the Pan-European methodology (Andreo et al., 2006). In this method, a hazard inventory list has been proposed by the COST Action project team with weighting factors for each potential hazard depending on the toxicity, solubility and mobility of the associated pollutants (De Ketelaere et al., 2004). The final hazard index is acquired after multiplying the weighting factor of each land use polygon with a ranking and a reduction factor that express the relative quantity of pollutants that can be released from each activity and the possibility for contamination to occur based on the infrastructure condition, respectively.

The risk map of the area has been developed by estimating the risk intensity index for each grid cell of the study map as stated by the formula:

$$R_i = 1/H_i V_i$$

where R_i is the risk intensity index, H_i is the hazard index from the respective land use map and V_i is the intrinsic vulnerability index from the COP map.

2.3. Physicochemical, heavy metal and biological analyses

Physical and chemical variables (pH, conductivity, nutrients, trace elements, etc.), heavy metals (Zn, Pb, Cd, Ni, Cr, etc.) and toxic compounds (PCBs, phenols) were compiled from various publications, including technical reports, to provide an overview of the environmental status of the lake and its catchment.

Regarding biological data, benthic invertebrate samples were collected from six sampling sites at depths ranging between 0.5 and 2 m. Benthic samples were collected with a bottom sampler (surface covered—0.045 m²), preserved with ethanol and transferred to the laboratory. Benthic species were removed from the samples with soft tweezers and identified at the lowest taxonomic level where possible.

The results from the physicochemical analyses in the lake and aquifer water as well as from the biological survey have been compared to the aforementioned risk zones to validate the mapping outcome and to depict the linkage between the potential impacts on the local environment and specific human activities or land uses.

3. Results and discussion

3.1. Groundwater vulnerability and risk assessment at a catchment level

In the particularly study area no obvious shallow holes and sinks are observed while the degree of karstification is moderate according to recent hydrogeologic investigations in the area (Zacharias et al., 2003). Therefore the second scenario of the COP method is used (Andreo et al., 2006) which is applied to areas outside shallow hole recharge zones and the karstic features of the study catchment have

been classified as scarcely developed karst with permeable and in some cases absent surface layers. This process led in assigning the value of 0.5 to the calcareous formations of the catchment (min: 0.25—developed karst with absent surface layers, max: 1—non karstic terrains). The rest of the study area covered by other than calcareous formations has been classified as non karstic terrains regarding the C index, according to the relevant methodology.

The hydrographical network of the area can be characterized as sparse and not well developed partly due to the human interventions in the catchment's geomorphology, as well as due to the low rainfall values. Particularly, there are few episodic streams in the surrounding mountains that discharge, after rainfall events only, limited amount of water in the sedimentary plain and some of them directly to the sea. There is no indication of influent streams and thus no significant connection between the limestone outcrops and surface water bodies exist. The only seasonal stream that discharges water in the sea for few winter months is located close to the western part of the catchment, flows over the relatively low permeability formations and in its downstream section is canalised. Therefore, the hydrographical network in this particular case does not seem to interact significantly with the main aquifer and cannot create fast flow components to the groundwater.

The vulnerability map of the study area indicates three different classes (high, low and very low vulnerability, Fig. 2). The high vulnerability zone coincides to the limestone outcrops in the central and northern part of the catchment while the low vulnerability areas represent the thick and medium to low permeability sediment formations in the lowland that offer significant pollution protection due to high natural attenuation potential. Further, very low vulnerability values have been assigned to the spatially limited flysch outcrops and to the clayey alluvial sediments of the south-western part of the catchment. Approximately 49% of the total extent of the study area is of high vulnerability while the remaining 51% illustrates relatively high natural protection as far as the groundwater is concerned (Table 1).

In the hazard map of the area, 85% of the catchment is classified as very low and low hazard, which coincides to the abandoned, forested and some limited urban zones (site no. 3 in Fig. 3, Table 1b). It should be stated that the urban areas in the particular catchment cover only a small proportion of the study area since this is officially characterised as an industrial zone and therefore no significant pressure can be assigned to this land category. Moreover, the agricultural sector even though used to be a substantial activity during the previous decades, nowadays cannot be considered developed in the specific catchment due to the limited extent available for commercial farming and because of the economic pressure from industry for its spatial expansion. Approximately 12% of the study area is of moderate hazard mainly due to the existing official landfill of Athens city (site no. 4 in Fig. 3), active mining

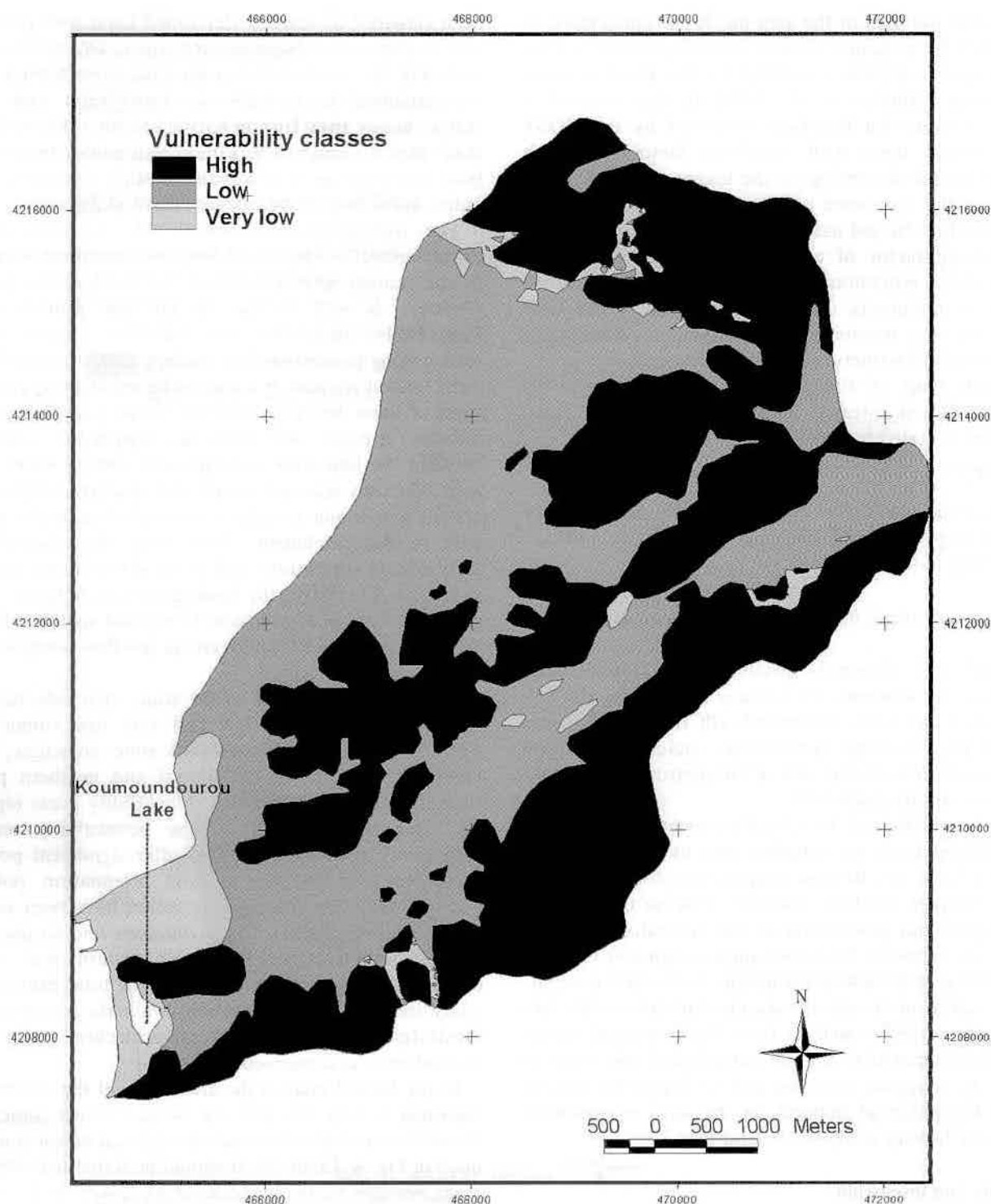


Fig. 2. Vulnerability map of Koumoundourou Lake catchment.

installations in the western part of the catchment and large scale industries (site no. 2 in Fig. 3, plastic equipment manufacturers and asbestos factories). It should be mentioned that about 60% of the particular landfill, which receives urban solid waste from approximately 3 million people has been constructed on the high vulnerability zone (limestone formations) at the upstream part of the catchment. Only 3% of the study area is of high hazard

potential and this is due to the existence of major oil refineries in the southwestern part of the catchment (site no. 1 in Fig. 3).

The combination of vulnerability and hazard map provided the groundwater pollution risk map of Koumoundourou lake catchment (Fig. 4, Table 2). The low and very low risk zones cover 83% of the entire area which comprise urban zones on medium to low permeability

Table 1
Vulnerability (a) and hazard (b) classes and extent in Koumoundourou Lake catchment

(a)		
Vulnerability classes	Area (km ²)	Percent of total
High	19.49	48.96
Low	17.78	44.68
Very low	2.54	6.37
(b)		
Hazard classes	Area (km ²)	Percent of total
Very low	28.84	72.45
Low	5.20	13.06
Moderate	4.86	12.20
High	1.10	2.76

sediments as well as natural areas over limestone and sediment formations (Fig. 1). The moderate zone of the risk map (14% of the catchment) represents mostly the large scale industrial installations in the lowland sedimentary section of the study area. Moderate pollution risk for the groundwater depicts also the oil refinery downstream, the sedimentary part of the landfill and oil warehouses close to Koumoundourou lake (Fig. 4). High and very high risk zones cover approximately 4% of the catchment and comprise the part of landfill lying on limestone formations, active mining installations and oil refineries/warehouses on permeable surface layers.

The area west of Koumoundourou Lake illustrates low pollution risk since there is only a thin boundary with clayey sediments that separates the lake from the adjacent sea shore while east and north of the lake there are moderate risk zones even though the surface layers offer significant protection due to the location of the oil refinery and warehouses in close proximity.

The main aquifer in the area is located in the Triassic limestone formation in a level slightly above the mean sea elevation (Fatta et al., 2002). Pumping tests occurred under a recent hydrogeologic investigation (Zacharias et al., 2003) in several boreholes, homogeneously distributed in the study area, which indicated groundwater velocities that range between 3 m/day and 20 m/day. These tests were conducted in alluvial sediments (falling head test—MAAG) and in limestone formations (Lugeon) and were repeated 2–3 times at each borehole to validate the results. The piezometric map of the catchment has been developed mainly for qualitative assessment purposes by using groundwater level measurements from the 10 boreholes and kriging interpolation technique. The resulted map indicates a dominant flow direction of NE–SW (Fig. 5) which means that groundwater in this area travels with a relatively high velocity from upstream towards the broader area of Koumoundourou Lake and the adjacent coastal zone and discharges through submerged springs. Therefore, natural attenuation in the specific aquifer system is relatively low and consequently any conservative pollutant reaching the main groundwater body would eventually end

up in Elefsina gulf and in Koumoundourou Lake. This is in accordance with chemical analysis results occurred in well and lake samples for PCBs and PAHs (Figs. 5 and 6). PAHs results (Fig. 5) illustrate that the lake water sample has the second highest concentration of all the analyzed samples while the highest value has been recorded in a well located downstream of an oil warehouse installation. As far as PCBs results are concerned the lake water sample presents the highest concentration while the second highest value has been observed in the borehole stated above.

3.2. Groundwater and lake water quality

3.2.1. Lake water quality

Dissolved oxygen concentrations fluctuated considerably in the lake where the lowest concentration recorded was 6.8 mg/l and the highest 12.9 mg/l (Table 3). Due to the shallow character of the lake that allows atmospheric oxygen exchange and to the photosynthetic activity of several aquatic plants (see Section 3.3) oxygen levels can be considered low but acceptable. These fluctuations in the dissolved oxygen levels are also a result of periodic oil film formation in the surface of the lake that enhances bacterial activity and restricts atmospheric oxygen exchange. Salinity levels are very high and vary significantly due to temporally irregular freshwater recharge from submerged springs resulting to the variation of salinity layers. As a result of different salinity layers, conductivity levels fluctuated also considerably (0.19–2.31 S/m). Light transmission was relatively acceptable thus allowing aquatic plants to easily photosynthesize. Concentrations of nitrogen compounds were extremely high, especially those of N–NO₃ indicating significant pressures from human activities (Table 3).

Regarding heavy metal concentrations, these fluctuated significantly between sampling locations in the lake. According to the studies of Koussouris (1994), Conides et al. (1996) and Conides and Parpoura (1997), highest levels of Mn, Zn, Pb, Cu and Fe were detected at the northwestern and southwestern part of the lake, near its shores. This is probably because of the submerged springs located in the northern part of the lakeshore that transfer heavy metals introduced in the groundwater upstream in Koumoundourou catchment by the industrial activities. The concentrations of the rest of the heavy metals remained the same with no significant fluctuations. Overall, heavy metal concentrations were generally in similar levels with other Greek lakes (Skoulikidis et al., 1998; Sotiriou et al., 1991; Voutsinou, 1989; Fytianos et al., 1985).

3.2.2. Groundwater quality

The results of groundwater analysis from various studies (Fatta et al., 2002; Paraschoudis, 2002; Hellenic Ministry of Agriculture, 1990–1997) showed that groundwater quality is significantly deteriorated. Groundwater is almost brackish and levels of nitrogen compounds and chloride ions are exceptionally high (Table 4). The increased levels

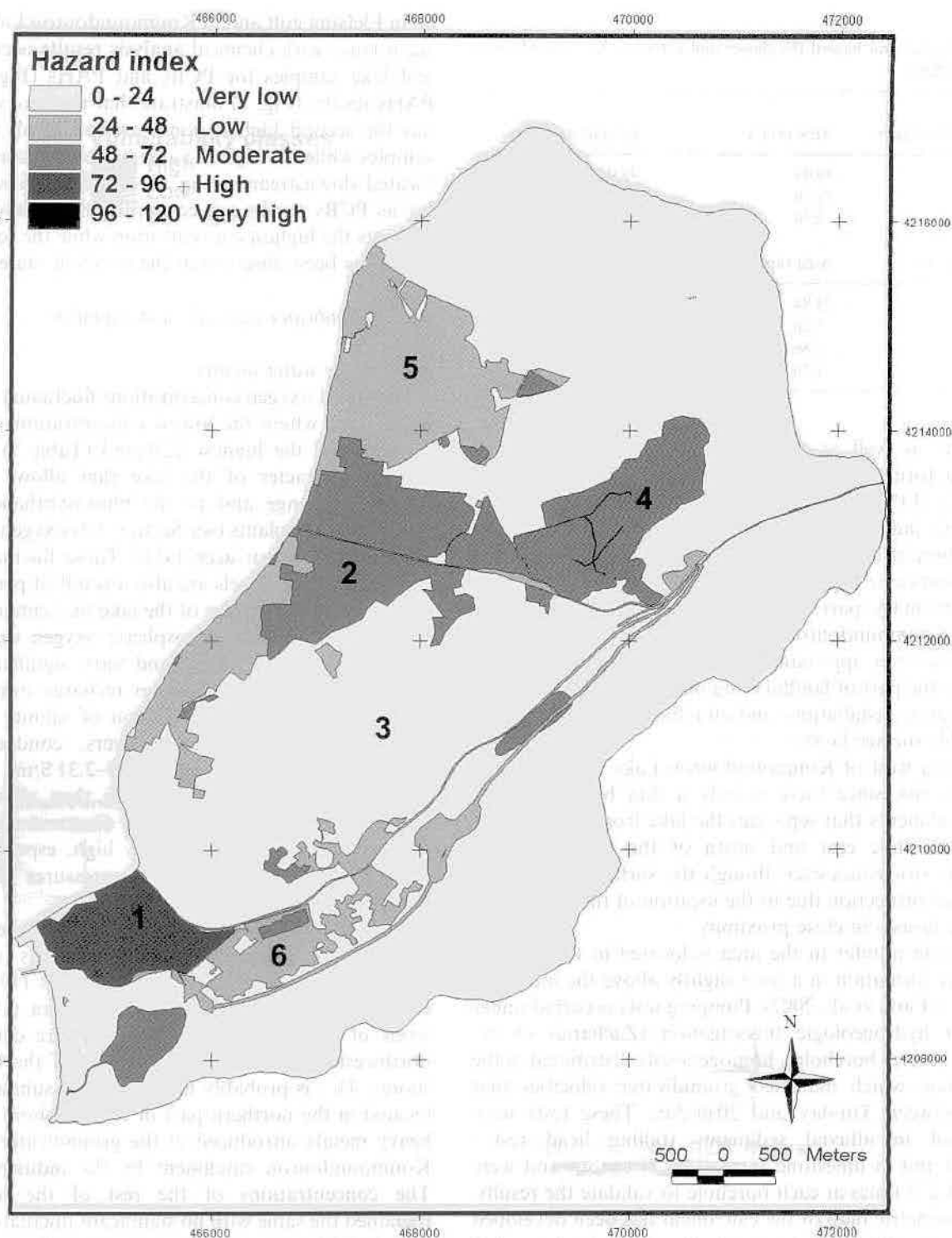


Fig. 3. Hazard map of Koumoundourou Lake catchment.

of chloride ions are mainly attributed to water abstractions that limit freshwater and allows mixing with seawater. Seawater mixing with groundwater is merely dependent with the distance from the sea. Near Elefsis Bay, chloride levels are about 1900 ppm, whereas further inshore at 2 km distance chloride levels reach 2500–3000 ppm (Paraschoudis, 2002). Moreover, the nitrate concentrations in the groundwater is very high since it fluctuates from 211 to 885 mg/l

(Table 4), which is 4 to 17 times higher than the threshold presented in the Nitrates Directive and in the Directive on Water for Human Consumption (50 mg/l). The respective concentrations in the lake are even higher (Table 3a) since the residence time of the water and consequently of the pollutants is greater in the lake than in the groundwater. Moreover, the lake is also receiving pressure from the adjacent military camp in which the relevant sewer is

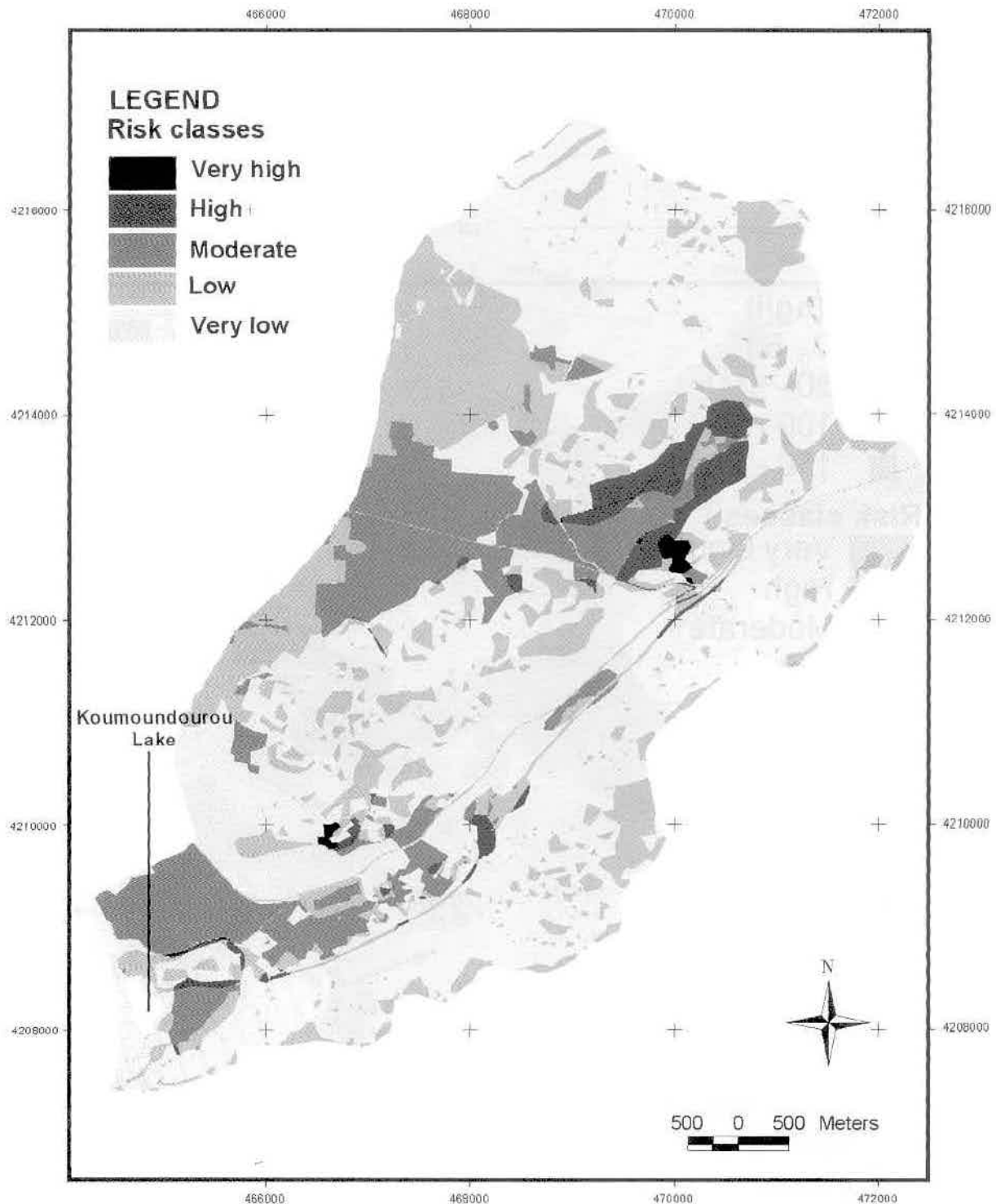


Fig. 4. Groundwater risk map of Koumoundourou lake catchment.

discharged and this further increases the nitrate values of the lake water. Nitrites and phosphates are also extremely elevated in the study area's groundwater and lake water which indicates substantial pressure from uncontrolled disposal of urban waste. Phosphates values in the lake water are relatively low in relation to the groundwater and this could be explained by the biological uptake that dominates in a lake ecosystem as well as by the phosphate

storage in the sediment which is more pronounced in a wetland area than in the groundwater.

3.3. Biological quality of the lake

The aquatic vegetation of the lake comprised of Halophytes (*Arthrocnemum fruticosum* and *Cakile maritima*), *Juncus marinus*, *Juncus effuses* and *Carex* species as

Table 2
Groundwater risk classes and extent in the study area

Risk classes	Area (km ²)	Percent of total
Very high	0.14	0.35
High	1.29	3.24
Moderate	5.47	13.75
Low	8.97	22.54
Very low	23.92	60.11

well as other common aquatic plants such as *Limonium gmelinii* (Koussouris, 1994). A large area of the lake's riparian zone was covered by *Phragmites australis*, which was also present inside the lake and especially abundant in areas that were secluded by oil. *Phragmites australis* is a very common species in nutrient rich environments and therefore the elevated nitrate concentrations measured in the lake are also validated by the dominance of this particular species. Additionally, *Chara hispida* and *Ruppia*

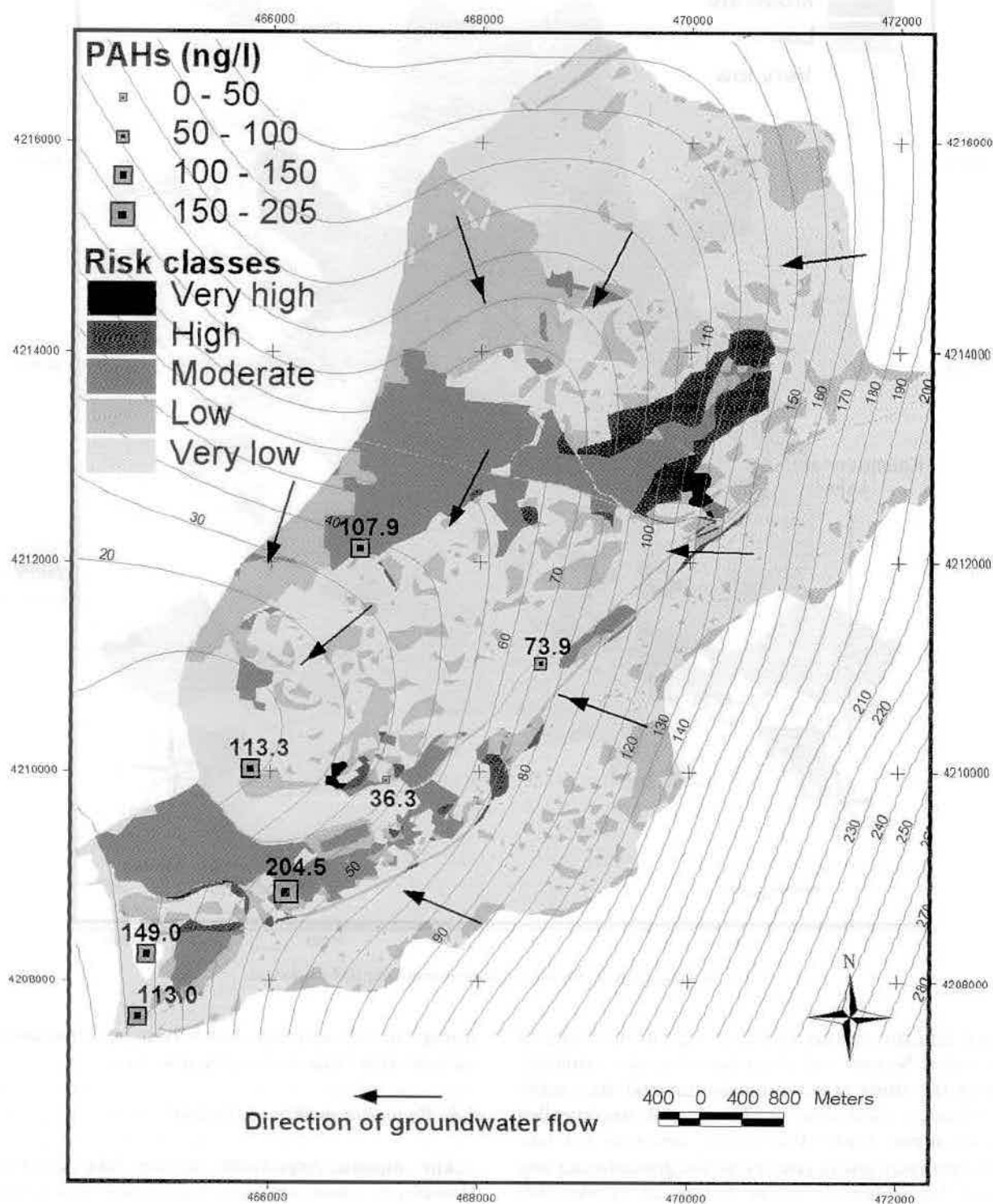


Fig. 5. Groundwater flow map and PAHs concentration in monitoring boreholes.

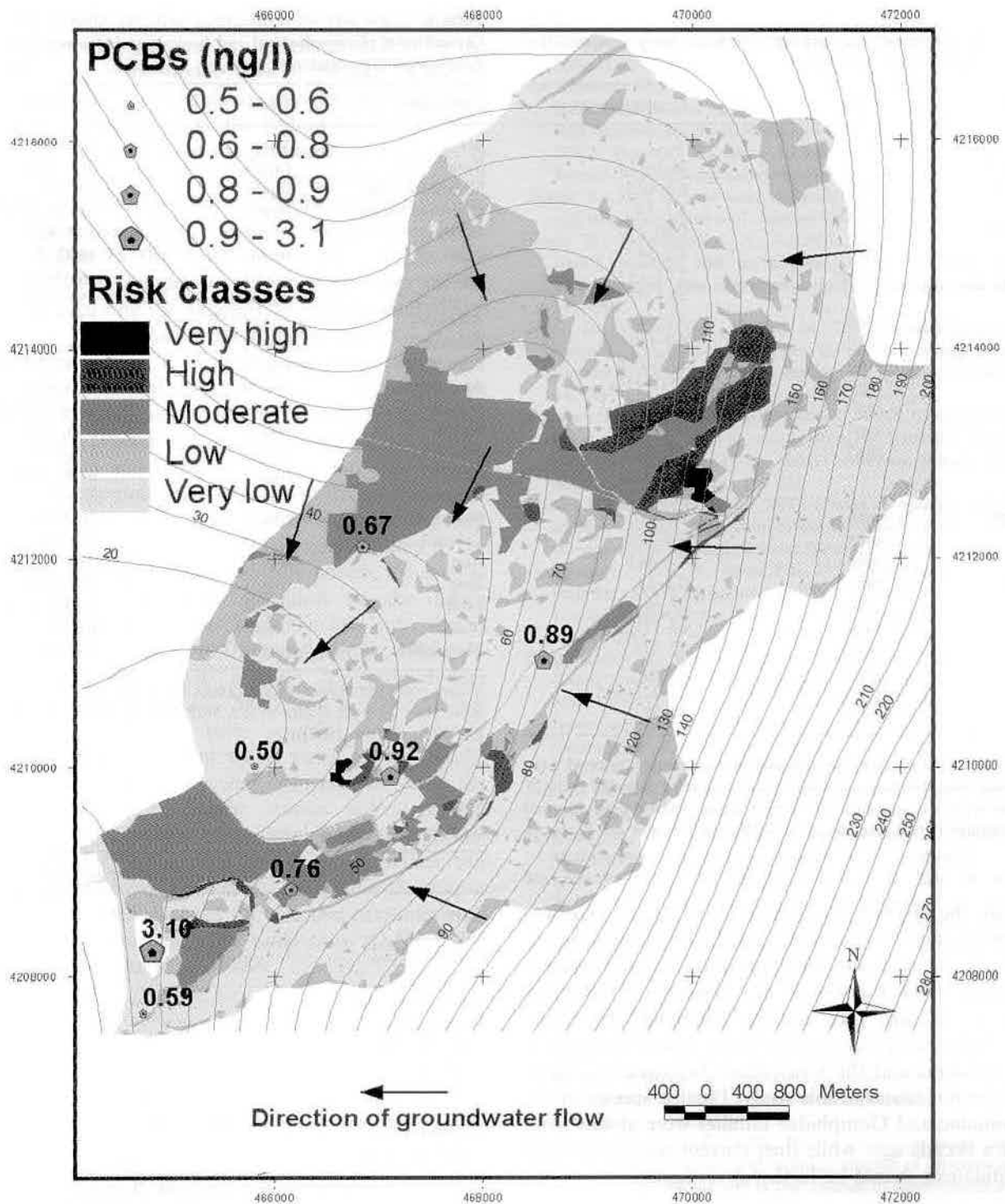


Fig. 6. Groundwater flow map and PCBs concentration in monitoring boreholes and the respective risk map.

cirrhus were found in relatively large numbers in the lake and around its shore while *Potamogeton* species were also present in high abundance at the deepest part of the lake. Benthic invertebrates collected during several periods were those of highly pollution tolerant species (Table 5). The invertebrate community of the lake was dominated by the Polychaeta *Nereis diversicolor*, which was found in exceptionally high abundances. The rest of the community

included the Oligochaeta *Tubifex tubifex*, the Crustacean *Tanais coquilini* and the Gastropod *Hydrobia ulva*. At very low abundances, species of the Chironomidae (Diptera) and Gomphidae (Odonata) families were collected (one taxon from each family). A large number of dead Bivalve shells were found in the benthic samples thus suggesting that a different benthic community used to inhabit the lake in the recent past.

Table 3
Lake water physicochemical (a) and sediment heavy metal concentration ranges (b) in Koumoundourou Lake

Parameters	Units	Concentrations
(a)		
Temperature	°C	21.1–22.5
Dissolved O ₂	mg/l	6.8–12.9
pH	Value	6.1–7.6
Conductivity	S/m	0.19–2.31
Salinity	‰	1.2–14.7
Density	Kg/m ³	2.75–13.3
Light	‰	49.1–97.2
BOD	mg/l	3.0–4.0
COD	mg/l	9.3–130
N–NO ₂	mg/l	17.7–22.6
N–NO ₃	mg/l	1237–1394
N–NH ₃	mg/l	0–62.5
P–PO ₄	mg/l	0–6.9
Si–SiO ₄	mg/l	3.8–4.6
(b)		
Manganese	ppm	46–800
Cobalt	ppm	8.0–12.0
Nickel	ppm	8.0–44.0
Cadmium	ppm	0.1–2.8
Chromium	ppm	21–56
Zinc	ppm	62–715
Lead	ppm	12–220
Copper	ppm	19–90
Iron	ppm	2332–11 835
Barium	ppm	14–37

Physicochemical parameters represent lake water samples collected from six sampling sites and heavy metals sediment samples collected from 11 sites. Data represent concentration ranges from 1994 to 1996 acquired from Koussouris (1994), Conides et al. (1996) and Conides and Parpoura (1997).

Overall, the diversity of benthic fauna of the lake was very limited and was dominated by the Polychaeta *Nereis diversicolor*. Compared with previous related studies (Koussouris, 1994; Conides et al., 1996) the invertebrate community of the lake almost remained the same. A second species, however of the genus *Nereis* was present (*Nereis caudata*) and the Crustacean *Leptocellia savignii*, which were not found in this study. Finally, species of the Chironomidae and Gomphidae families were absent from the lake a decade ago, while their current occurrence with low abundances should be further investigated.

4. Conclusions

The groundwater vulnerability map of Koumoundourou lake catchment indicated relatively high natural protection in the lowland due to the existence of thick and medium to low permeability sedimentary layers covering the aquifer. The high vulnerability zone encountered in this area coincides with the hilly limestone outcrops in the central and eastern part of the catchment extending over 48% of the total area. Nevertheless, even though most of the industrial installations are located in the relatively well

Table 4
Groundwater physicochemical and heavy metal characteristics of Lake Koumoundourou catchment area

Parameters	Units	Concentration
pH	Value	7.16–8.2
Total Solids	(mg/l)	369–2294.5
Suspended solids	(mg/l)	50–130
Dissolved solids	(mg/l)	245.83–2170.33
BOD ₅	(mg/l)	1.17–4.17
COD	(mg/l)	10.42–78
Hardness	(mg/l)	560–807.17
Conductivity	(mS/cm)	650–5700
Alkalinity	(asCaCO ₃) (mg/l)	52.17–230
Sulphate	(mg/l)	4.17–728
Phosphate	(mg/l)	11.33–171.33
Ammonium	(mg/l)	0.93–67.75
Nitrate	(mg/l)	211.17–884.83
Nitrite	(mg/l)	0.78–298.67
Bicarbonate	(mg/l)	150–550
Chloride	(mg/l)	47.17–3000
Potassium	(mg/l)	3.2–87.35
Sodium	(mg/l)	20–674.4
Calcium	(mg/l)	57–380
Iron	(mg/l)	0.51–1.92
Zinc	(mg/l)	0.05–0.19
Lead	(mg/l)	0.26–0.3
Nickel	(mg/l)	0.02–0.12
Magnesium	(mg/l)	34–200

Values represent concentration ranges from 1997 to 2001 sampled from more than 20 wells (Fatta et al., 2002; Paraschoudis, 2002; Ministry of Agriculture (1990–1997)).

Table 5
Benthic community of Lake Koumoundourou

Taxa	Lake Koumoundourou sampling locations					
	1	2	3	4	5	6
Oligochaeta						
<i>Tubifex tubifex</i>		+				
Polychaeta						
<i>Nereis diversicolor</i>	+	+	+	+	+	+
Gastropods						
<i>Hydrobia ulvae</i>	+	+	+	+		+
Crustacea						
<i>Tanais covoimii</i>					+	
Insecta						
<i>Chironomus</i> sp.	+	+	+	+	+	+
<i>Gomphidae</i> un.sp.		+				

protected lowland zone the groundwater risk map indicates moderate to high pollution risk due to the significant hazard that the particular human activities incorporate. The high risk zone is mainly observed in the landfill area (site no. 4 in Fig. 3) and especially in the part of it overlying limestone formations while very high risk areas are only encountered in sites where specific industries are located on high vulnerability substrate. Furthermore, a very high risk zone exists in the southern part of the coastal oil refinery (site no. 1 in Fig. 3) in close proximity to Koumoundourou

lake. These results are also accredited by the water quality analyses in groundwater and in the lake as well as by the biological survey in the broader area of the lake. PCBs and PAHs values increase in the direction of groundwater flow and in Koumoundourou lake the highest levels have been detected. This is because of the groundwater recharge in the lake that occurs through submerged springs and enriches it with leakages from industry and agricultural activities. Moreover, heavy metal concentrations were relatively higher in the Northeastern part of the lake (Koussouris, 1994; Conides et al., 1996) supporting the risk map results regarding the adjacent oil refinery interaction. Nevertheless, overall heavy metal concentrations in the lake are not extremely high during the study period and this is because of several restoration measures that allowed the removal of significant amount of oil from the lake over the past 5 years (Poulovassilis et al., 1994).

The water flow and turnover is such that, even though the lake is heavily polluted by oil, phytoplankton blooms are not promoted during any time of the year and the light penetration is relatively high. This can be attributed to the fact that the oil film that periodically covers the lake is forced, either by the wind direction or the lake's meromictic status, to outflow from the lake into Elefsis Bay (Conides & Parpoura, 1997; Conides et al., 1996).

In this study, the COP method for groundwater vulnerability (Andreo et al., 2006) and a relevant risk mapping approach have been applied in a heavily industrialised lake catchment while water quality results from the aquifer and the lake has been used to verify the risk assessment effort. Similar studies indicated that implementing groundwater vulnerability and risk mapping models to identify impacts from human activities can provide accurate results (Lake et al., 2003) while the output from such models can be used for effective environmental management by the responsible authorities (Osborn and Cook, 1997; Giupponi et al., 1999).

References

- Al-Adamat, R.A.N., Foster, I.D.L., Baban, S.M.J., 2003. Groundwater vulnerability and risk mapping for the Basaltic aquifer of the Azraq basin of Jordan using GIS, Remote sensing and DRASTIC. *Applied Geography* 23, 303–324.
- Andreo, B., Goldscheider, N., Vadillo, J., Vias, J.M., Neukum, C., Sinreich, M., Jimenez, P., Brechenmacher, J., Carrasco, F., Hotzl, H., Perles, M.J., Zwahlen, F., 2006. Karst groundwater protection: first application of a Pan-European approach to vulnerability, hazard and risk mapping in the Sierra de L bar (Southern Spain). *Science of the Total Environment* 357 (1–3), 54–73.
- Babiker, I.S., Mohamed, A.A.M., Hiyama, T., Kato, K., 2005. A GIS-based DRASTIC model for assessing aquifer vulnerability in Kakumigahara Heights, Gifu Prefecture, central Japan. *Science of the Total Environment* 345, 127–140.
- Conides, A., Diapoulis, A., Koussouris, T., 1996. Ecological study of an oil polluted coastal lake ecosystem in Greece. *Fresenius Environmental Bulletin* 5, 324–332.
- Conides, A., Parpoura, A.R., 1997. A study of oil pollution effects on the ecology of a coastal lake ecosystem. *The Environmentalist* 17, 297–306.
- COST 620, 1998. Vulnerability and risk mapping for the protection of carbonate (karst) aquifers—Minutes from the 3rd MC-Meeting, 8.S., Li ge (Belgium), 12–13 March.
- De Ketelaere, D., Hotzl, H., Neukum, C., Civita, M., Sappa, G., 2004. Hazard analysis and mapping. In: Zwahlen, F. (Ed.), *Vulnerability and Risk Mapping for the Protection of Carbonate (karst) Aquifers*. EUR 20912, Brussels: European Commission, Directorate-General XII Science, Research and Development, pp. 86–105.
- Dixon, B., 2005. Groundwater vulnerability mapping: a GIS and fuzzy rule based integrated tool. *Applied Geography* 25, 327–347.
- Fatta, D., Papadopoulos, A., Loizidou, M., 1999. A study on the landfill leachate and its impacts on the groundwater quality of the greater area. *Environmental Geochemistry and Health* 21, 175–190.
- Fatta, D., Naoum, D., Loizidou, M., 2002. Integrated environmental monitoring and simulation system for use as a management decision support tool in urban areas. *Journal of Environmental Management* 64, 333–343.
- Fytianos, K., Samanidou, V., Angelidis, T., 1985. Comparative study of heavy metals pollution in various rivers and lakes of northern Greece. In: *Proceedings of the Heavy Metal Conference*, Athens, Pan-Hellenic Society of Chemists, pp. 119–121.
- Gailey, R.M., Gorelick, S.M., 1993. Design of optimal, reliable plume capture schemes: application to the Gloucester landfill ground-water contamination problem. *Ground Water* 31 (1), 107–114.
- Giupponi, C., Eiselt, B., Ghetti, P.F., 1999. A multicriteria approach for mapping risks of agricultural pollution for water resources: the Venice Lagoon watershed case study. *Journal of Environmental Management* 56, 259–269.
- Goldscheider, N., Brechenmacher, J., Hotzl, H., Neukum, C., 2003. Engen, Swabian Alb, Germany. In: Zwahlen, F. (Ed.), *COST Action 620—Vulnerability and Risk Mapping for the Protection of Carbonate (karst) aquifers*. Final Report, European Commission Directorate-General XII Science, Research and Development, Report EUR, COST-ACTION 620, Luxembourg, pp. 200–216.
- Hellenic Ministry of Agriculture, 1990. Water quality characteristics of lakes and rivers, 1990–1997.
- Kitanidis, P., 1997. *Introduction to Geostatistics. Applications in Hydrogeology*. Cambridge University Press, Cambridge.
- Kjeldsen, P., Bjerg, P., R gge, K., Pedersen, J.K., Skov, B., Foverskov, A., Wurtz, S., Christensen, T.H., 1993. Assessing the variability in leachate migration from an old municipal landfill. In: Cossu, R., Christensen, H.T., Stegmann, R. (Eds.), *Proceedings of the Fourth International Landfill Symposium*, Sardinia, Italy, pp. 1519–1531.
- Koussouris, Th., 1994. Environmental upgrading of Lake Koumoundourou. Technical Report, HCMR, Athens.
- Lake, I.R., Lovett, A.A., Hiscock, K.M., Betson, M., Foley, A., Sonnenberg, C., Evers, S., Fletcher, S., 2003. Evaluating factors influencing groundwater vulnerability to nitrate pollution: developing the potential of GIS. *Journal of Environmental Management* 68, 315–328.
- Osborn, S., Cook, H.F., 1997. Nitrate vulnerable zones and nitrate sensitive areas: a policy and technical analysis of groundwater source protection in England and Wales. *Journal of Environmental Planning and Management* 40, 217–233.
- Paraschoudis, V., 2002. Hydrogeological study of Western Attica (Megara–Thrasio Plains). Hellenic Ministry of Agriculture, Department of Geology-Hydrology, Athens, 165 pp.
- Poulovassilis, A., Mimides, T., Faltseta, S., Gianopoulos, P., Sgoumpoulou, A., 1994. Hydrogeology of Lake Koumoundourou. Pollution sources and its protection. Agronomic University of Athens.
- Skoulikidis, N., Bertahas, I., Koussouris, T., 1998. The environmental state of freshwater resources in Greece. *Environmental Geology* 36, 1–17.
- Sotiriou, N., Ouzounis, K., Lahamatis, S., 1991. Chlorinated hydrocarbons and heavy metals in Lake Vistonis in northern Greece. *Environmental Pollution and its impact on Life in the Mediterranean Region*.

- Tesoriero, A.J., Inkpen, E.L., Voss, F.D., 1998. Assessing ground-water vulnerability using logistic regression. In: Proceedings for the Source Water Assessment and Protection 98 Conference, Dallas, pp. 157–165.
- Voutsinou, F., 1989. Heavy metal concentration on surface sediments of two Ionian Sea lagoons. In: Lekkas, Th.D. (Ed.), Second Congress of Environmental Science and Technology. pp. 261–270 (in Greek).
- Vrba, J., Zaporozec, A., 1994. Guidebook on Mapping Groundwater Vulnerability. International Contributions to Hydrogeology 16. Verlag Heinz Heise, Hannover.
- Zacharias, I., Sarantakos, K., Dimitriou, E., Koussouris, T., 2003. Hydrogeologic report for site selection of sampling wells in the hydrogeologic catchment of Lake Koumoundourou, 1st Technical Report. HCMR/IIW, Athens.

Invasion of *Pinus halepensis* Mill. following a wildfire in an Argentine grassland nature reserve

Sergio M. Zalba^{a,*}, Yannina A. Cuevas^a, Roberto M. Boó^b

^aGEKKO, Grupo de Estudios en Conservación y Manejo, Departamento de Biología, Bioquímica y Farmacia, Universidad Nacional del Sur (UNS), San Juan 670 (8000), Bahía Blanca, Argentina

^bDepartamento de Agronomía UNS, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS), Argentina

Received 28 April 2006; received in revised form 23 February 2007; accepted 16 March 2007

Available online 27 April 2007

Abstract

The Ernesto Tornquist nature reserve is a relict of native Pampas vegetation in Argentina. Alien trees were introduced to the reserve in the 1950s, mainly to “improve” the natural landscape, resulting in the arrival of a totally new life form. In 1987, a fire affected an area planted with *Pinus halepensis* resulting in its massive expansion. In 1999, we removed trees from 17 circular plots of 10 m diameter placed systematically within the area that was colonized after the fire. Trunks were cut 20 cm from the ground and growth rings were counted. We studied the age structure of the population in order to reconstruct the colonizing events after the fire. We found that recruitment occurred throughout this period, except in the three years after the disturbance. We suggest that this delay in recruitment might be caused by low seedling survival under water stress conditions due to low rainfall, combined with scarce vegetation cover after fire. This could have been associated with an initial reduction in propagule pressure due to the scarcity of surviving trees in the vicinity and with the fact that fire occurred after the peak of seed release, during an extremely dry summer, probably killing a great number of seeds that were already in the soil. In the following years, recruitment was probably aided by pioneer trees and later by seeds shed from established pines. Alien trees had been allowed to reach maturity due to wildfire prevention and control in the years preceding the fire and the accumulated dry matter resulted in increased fire intensity that reduced the ability of grasses to re-sprout. As a consequence, the invasion window that allowed the expansion of pines remained open for at least 12 years.

© 2007 Published by Elsevier Ltd.

Keywords: Age structure; Conservation; Disturbance; Invasion window; Invasive alien species

1. Introduction

Biological invasions are among the most important human-driven agents of change in natural environments around the world, and are usually associated with significant impacts on economy, culture and biodiversity (Baskin, 2002; Williamson, 1996). Historically human activities have caused the accidental or planned dispersal of many species beyond their original distribution ranges and this process is presently at its highest expression. According to Williamson (1996) only a small percentage of species introduced into a

new area are able to establish self-sustaining populations, and only a small number of them become invasive. Nevertheless, the probability of invasion success increases significantly according to three main factors: previous selection of species ecologically suited to the new location, propagule pressure and the degree of “human assistance” the species receives during the immediate post-introduction phase, when it is especially susceptible to environmental and demographic stochasticity. Forestry species have a particularly high frequency of introduction and huge efforts are directed towards getting them acclimatized to new localities. They have been introduced for a variety of purposes including firewood, food production, timber, land stabilization, wind-breaks or for “improving” the landscape (Mack et al., 2000; Richardson, 1998; Richardson and Bond, 1991).

*Corresponding author. Tel.: +54 291 4595100x2420; fax: +54 291 4595130.

E-mail address: szalba@criba.edu.ar (S.M. Zalba).

Tree species, such as pines and eucalypts, have been introduced all over the world, covering millions of hectares. Eucalypts have shown poor invasive potential, but the genus *Pinus* includes some of the species with the greatest invasive potential. Species such as *P. radiata*, *P. elliotti*, *P. halepensis*, *P. patula* and *P. pinaster* have invaded extensive regions in Argentina, Australia, Brazil, Chile, Colombia, New Zealand, South Africa, Uruguay and Venezuela (GISP, 2005; Richardson and Rejmánek, 2004). They have modified huge areas of grassland, shrubland and woodland, replacing dominant life forms, changing structural diversity and altering the frequency and intensity of fires (Chilvers and Burdon, 1983; Macdonald and Jarman, 1985; Richardson and Bond, 1991). Three species of pines, *P. radiata*, *P. halepensis*, and *P. pinaster* have been cited as aggressive invaders in Argentina (Inbiar, 2005; Simberloff et al., 2002; Zalba and Villamil, 2002) and the first two are considered as major threats to biodiversity conservation in the last relicts of native grassland in the Pampas phytogeographic province (Zalba, 2000).

The ecological features that increase the ability of serotinous pines to colonize and invade new areas are related to their life-trait attributes (Rejmánek and Richardson, 1996). They are self-compatible, rapidly maturing plants that produce a large quantity of small, light, winged, wind-dispersed seeds (Cronk and Fuller, 1995; Richardson et al., 1994). Their cones remain closed for long periods, thus maintaining a persistent canopy seed bank (Daskalakou and Thanos, 1996; Izhaki et al., 2000; Thanos and Daskalakou, 2000). Fire promotes the opening of serotinous cones and favours seedling establishment due to the addition of nutrients to the soil and the elimination of vegetation cover (Rejmánek, 1999; Rejmánek and Richardson, 1996; Thanos, 1999). Therefore, fire is a key factor for the establishment of spontaneous populations of serotinous pines. Knowledge of the relationship between the alien invasive species and this disturbance constitutes a powerful tool for managing the invasion problem. *P. halepensis* is well known as a naturalized or invasive species in Australia, Israel, New Zealand, South Africa and USA (Richardson and Higgins, 1998; Richardson and Rejmánek, 2004), but its invasion dynamics have not been studied in any part of South America. In this study, we analyse the age structure of an Aleppo pine population that is invading natural grassland in the Argentine Pampas in order to determine the effect of fire on its recruitment. We also evaluate whether fire is responsible for opening an invasion window that favours the advance of the alien species.

1.1. Study area

The work was carried out in a *P. halepensis* stand located on the south-western slopes of the Ventania mountains, in the Ernesto Tornquist Provincial Park, Buenos Aires province, Argentina (Fig. 1). The Ventania mountains are found from 38°10'S to 62°8'W in a general NW–SE direction, and they include the highest peaks of the Pampas

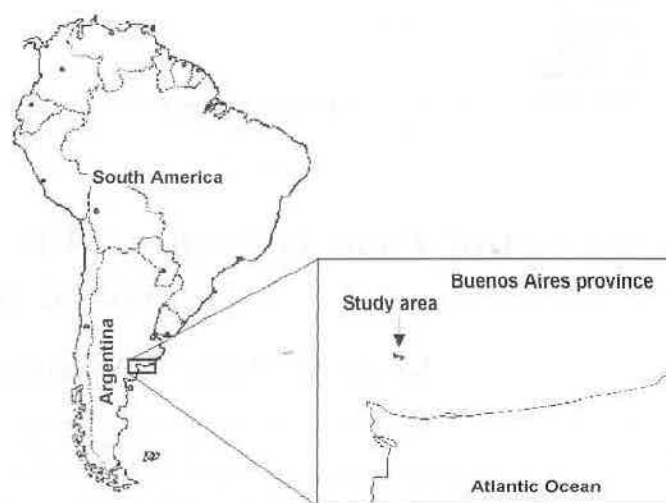


Fig. 1. Location of the study area. Ernesto Tornquist Provincial Park includes part of a mid-height mountain range in the south of Buenos Aires province, Argentina (38°04'S, 61°58'W).

region, which barely exceed 1200 m above sea level (Kristensen and Frangi, 1995).

The climate is temperate, with a mean annual temperature of 14.6 °C. Average annual rainfall ranges from 600 to 800 mm, concentrated in the spring and part of summer (Frangi and Bottino, 1995). The region belongs to the Southern District of the Pampas Phytogeographic Region, where grass steppe is the dominant vegetation (Cabrera, 1976) and *Stipa*, *Piptochaetium*, *Festuca* and *Briza* are among the most common grass genera. There are also some shrub communities mainly dominated by *Eupatorium buniifolium*, *Discaria longispina* and *Geoffroea decorticans* (Kristensen and Frangi, 1995).

Ernesto Tornquist Provincial Park is one of the last relicts of native vegetation in the Pampas region (Bertonatti and Corcuera, 2000; Bilenca and Miñarro, 2004) and has an area of ca. 6700 ha. More than 50% of the vascular flora belonging to this ecosystem is found in the reserve, including 37 endemic taxa (Long and Grassini, 1997), which highlights the role of this park as the most valuable unit for biodiversity conservation in the region as a whole (Kristensen and Frangi, 1995).

The reserve is undergoing severe invasion by Aleppo pine. The species occurs in a high priority area for conservation, causing significant reductions in native plant species diversity and richness, and allowing the introduction of other alien plants and the establishment of opportunistic bird species (Zalba, 1994, 2001). This species was originally planted in the park in the 1950s, with trees clustered in small plantations and also scattered along trails and roadsides. Data from aerial photographs 1:10 000 revealed that pines covered an area of ca. 10 ha in 1981 (Zalba 1994). In January 1987, a fire affected 80% of the area covered by pines. Unfortunately, there are no specific records about the extent of the fire or its intensity, or the percentage of trees that were killed by fire. We can infer,

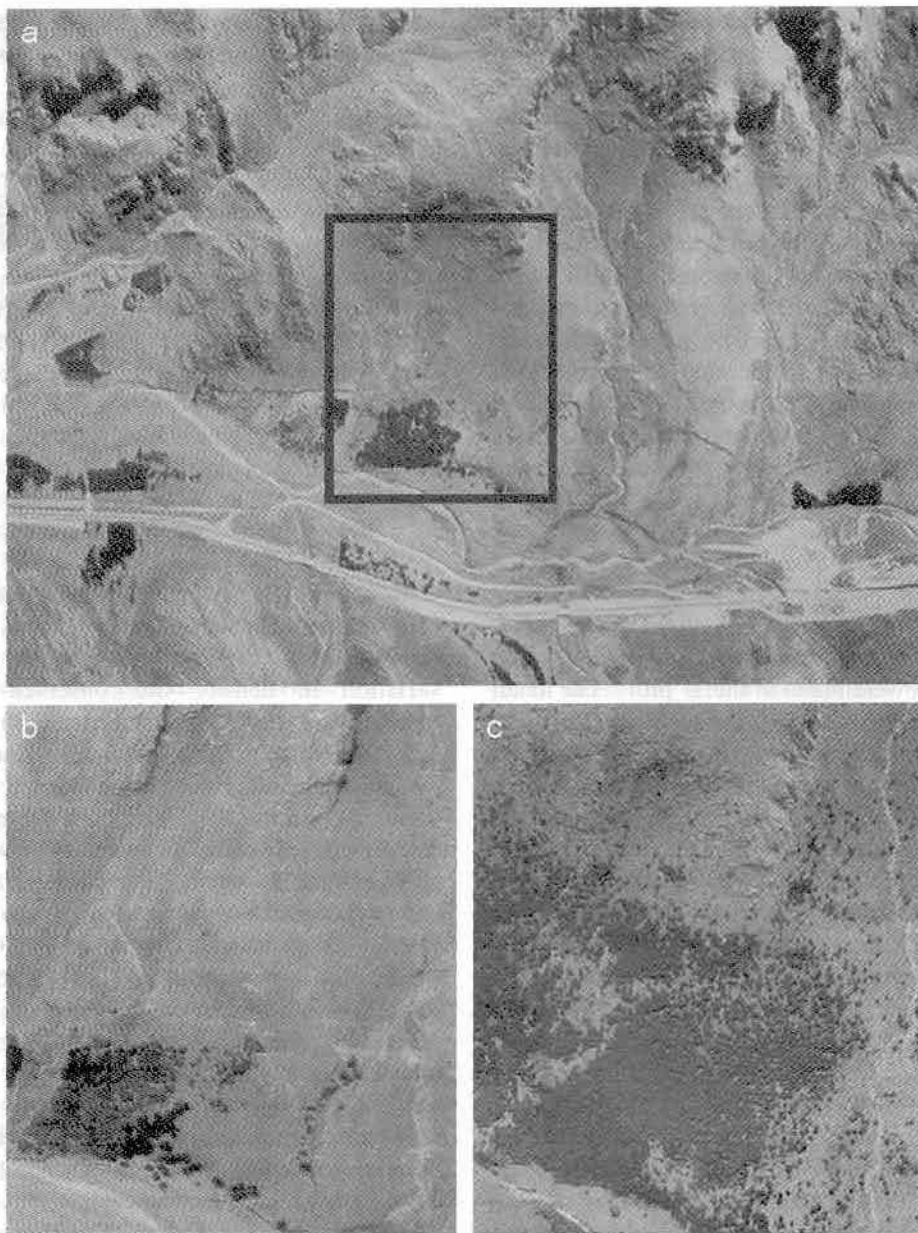


Fig. 2. Expansion of Aleppo pines in Ernesto Tornquist Provincial Park: (a) 1967 (scale 1:10 000); (b) 1981; and (c) 2003 (scale 1:6000).

from interviews with people that lived in the area at the time of the fire and also by considering the age of the oldest trees in this sector, that fire killed most of the pines in the area directly affected, leaving unburnt clusters at distances ranging from 250 to 400 m from the burnt area. A twenty-fold increase in the area of occupancy of *P. halepensis* was recorded following this fire (Zalba, 1994) (Fig. 2).

2. Materials and methods

In June 1999, 17 circular plots of 10 m diameter were established at two different altitudes (500 and 700 m above sea level) in a dense stand of Aleppo pine, as part of a control strategy implemented in the reserve to restore natural environments invaded by pines. Aerial photo-

graphs showed that the area under study was devoid of trees before the 1987 fire, and this was confirmed by the lack of trees older than 12 years in this site. All the trees inside these plots were cut down. Pines with trunk diameters of more than 1 cm were cut with hand and chain saws 20 cm above ground level, while the smaller ones were pulled by hand. The number of pines taken from each plot was recorded, the diameters of cut trunks measured, and the number of growth rings counted. For the smaller trees, the number of growth rings recorded in the field was confirmed by looking at cross-sections dyed with safranin under the microscope. In the case of larger trees, growth rings were counted in the field, and trunks were treated with bleach (sodium hypochlorite) when rings were difficult to define.

The age structure of the population was inferred from data on the density of each age class and changes in recruitment were analysed taking into account the time elapsed since the fire. Age structure was evaluated in the light of information on rainfall during the main recruitment season (March–May) (Cuevas and Zalba, 2004).

Density and basal area were calculated when the trees were felled. Differences in density, basal area and mean age between pines in the sites at different altitudes were assessed by means of “*t*” tests. Linear regression analyses were used to evaluate the relationship between the diameter of the trunks at a height of 20 cm and the number of growth rings at the different altitudes. Both regression curves were compared in order to determine if there were differences in growth rates at different altitudes.

3. Results

3.1. Age structure

A total of 199 pines were felled in the 17 plots. The mean age for all measured trees was 5.16 years (SE = 0.17) and the mode was between 5 and 6 years, corresponding to trees that recruited in 1994 and 1993, respectively. As the sampling was undertaken at the time when most of the germination takes place in the region (Cuevas and Zalba, 2004) we were not able to record seedlings corresponding to the last peak of seed release (summer 1998–1999) and so the youngest plants recorded belonged to the recruitment peak of the year preceding the cut (1998). The oldest pines were 10 years old in the higher (700 m) site and 12 years old in the lower (500 m) one, coinciding in the latter case with the fire that affected the area in 1987.

The age structure of the population shows a multi-modal profile, with several peaks of recruitment during the period under consideration (1987 to 1999) (Fig. 3). One of the peaks corresponds to young trees one year old, representing 8.5% of the population, the majority of which showed

evident signs of weakness, with dry needles and branches. The other two peaks coincide with trees aged five and six years, and eight years, respectively. Only 3.7% of the studied pines (trees that were 10–12 years old at the time of felling) were recruited in the first three years after the fire. A good correlation can be seen between the density of trees of different ages and the amount of rain during the recruitment period (March–May). Rainfall during the recruitment periods corresponding to the first four years after the fire were slightly below the historical average of the area (212 mm, SE = 9.8). In addition, precipitation during the 1986–1987 summer was only one third of the average since 1915, probably determining extremely low water runoff and soil humidity at the moment of the fire.

3.2. Density, basal area and trunk diameter–age relationship

Mean pine density in 1999 was 0.55 trees/m² (SE = 0.20), with no significant differences in this variable between the 500 and 700 m sites ($t = 1.7$, $p = 0.12$). However, a great variation in density was observed within each site, especially at the lower altitude, with densities two and four times higher than the mean value for the site in the two plots located close to the border of the stand. When these two plots were excluded from the analysis, the consequent reduction in variability allowed the detection of a significantly higher density at the lower altitude (0.32 ± 0.04 plants/m²) compared to the higher one (0.21 ± 0.04 plants/m²) ($t = 2.074$, $p = 0.05$).

Mean basal area was 35.27 m²/ha (SE = 4.01), with no significant differences between altitudes ($t = 0.82$, $p = 0.425$) and there was still no difference even when the outer plots of the lower site were excluded ($t = 0.58$, $p = 0.57$).

Highly significant linear relationships were detected between trunk diameter and age (number of growth rings) for all the trees considered together and at the different altitudes separately (Fig. 4). The curves belonging to these

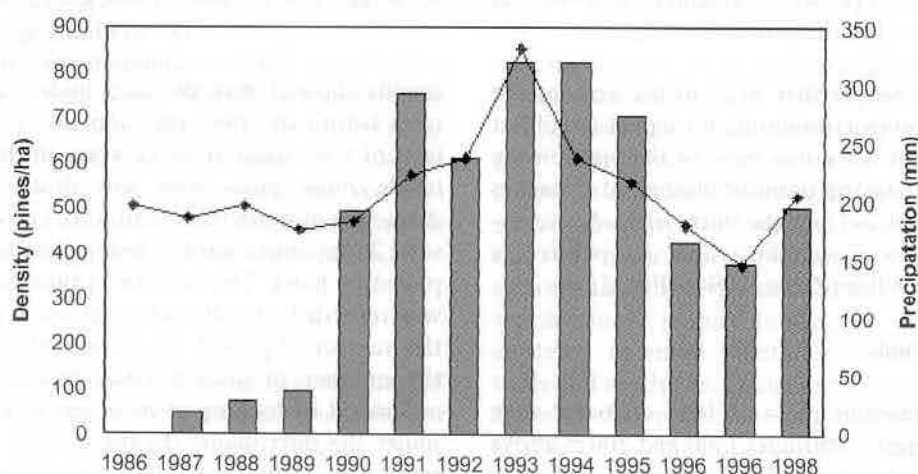


Fig. 3. Age structure in a stand of *Pinus halepensis* Mill. determined by counting growth rings on 199 trees (columns) and rainfall measured during the recruitment period (March–May) (line).

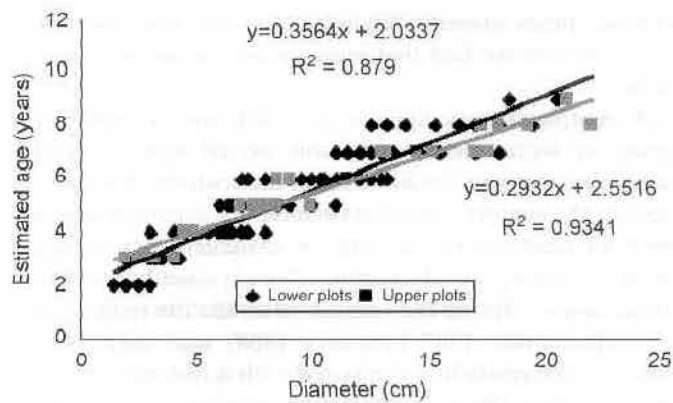


Fig. 4. Linear regression curves corresponding to pines growing at 500 (black line) and 700 (grey line) m above sea level.

two zones showed significant differences ($p = 0.0013$). The one corresponding to the higher altitude ($Y = 0.293X + 2.551$) has a smaller slope than the one at the lower altitude ($Y = 0.356X + 2.033$). This reveals significant differences in the growth rate, favouring trees at the higher altitude.

4. Discussion

The effect of disturbance on the success of invasive species has been extensively discussed (Hobbs, 1989, 1991; Lozon and MacIsaac, 1997). In particular, the relationship between disturbance and success of alien invasive plants in the Pampas grasslands of Argentina was assessed experimentally by Facelli and León (1986) and Mazia et al. (2001) working with elm (*Ulmus pumila*), privet (*Ligustrum lucidum*), and honey locust (*Gleditsia triacanthos*) and they concluded that native grasses interfere with establishment processes and that invasion only becomes possible when grasslands are strongly disturbed by vegetation clipping or topsoil disturbance. In a similar way, our results stress the importance of fire as a key factor that triggered the invasion of *P. halepensis* in the natural grassland that we studied. This situation has already been reported by many authors working in the area where this species is native (Izhaki et al., 2000; Ne'eman and Izhaki, 1998, 1999; Ne'eman et al., 1992; Thanos and Daskalidou, 2000; Thanos et al., 1996; Tsimoni and Karagiannakidou, 2000) and also in regions where it acts as an invasive alien (Higgins and Richardson, 1998; Richardson, 1988; Richardson and Higgins, 1998; Rouget et al., 2001). Nevertheless, some interesting features arose when the age structure of pines that invaded our study area was compared with those reported elsewhere, and this could help to fill a gap in our understanding of the invasive nature of this species.

One of these features relates to the length of the recruitment period. The 1987 fire in our study site seemed to promote massive seed release and recruitment, as happens in all the other described locations, the difference

being that in our case recruitment continued for at least the next 12 years. In South Africa Aleppo pine recruitment completely stopped seven years after fire (Richardson, 1988), whereas in Greece it started after fire with a gradual decrease during the first two post-fire years (Thanos et al., 1996). Our results indicate continuous establishment during the study period which coincides with results reported by Trabaud et al. (1985) for Montpellier, France, where Aleppo pine density increased during 15 years after fire. The soil seed bank in *P. halepensis* is ephemeral (Daskalidou and Thanos, 1996; Izhaki et al., 2000; Skordilis and Thanos, 1995; Thanos 2000; Thanos et al., 1996), so the seed source for this continuous recruitment must have come from trees that remained unaffected by fire, as well as from cones in trees that were partially burnt but not killed by fire, which in turn were able to cast seeds during the first years after fire (Trabaud et al., 1985). A few years later, newly established trees must have constituted new seed sources, as suggested below.

Another interesting feature of the case reported in this study is the relative low success of cohorts recruiting in the years after fire, which determines the shape of the age structure curve. Most studies on the response to fire by *P. halepensis* report population structures that are almost even-aged, with a main peak in the first year after the disturbance (Richardson, 1988; Trabaud et al., 1985). The time of establishment of the older trees in our study (12 years) coincides with the year after the fire. Fire was probably responsible for increasing pine recruitment, either by stimulating seed release or reducing competition with grasses in the regeneration niche, or both. Nevertheless, seedling establishment in our study area was not very successful during the first three years after the fire; consequently, the main peak in the age structure curve is displaced forward. We suggest that this delay in recruitment might be caused by low germination and/or seedling survival under water stress conditions resulting from the combination of low rainfall and lack of protection due to scarce vegetation cover after fire. Thanos et al. (1996), working with a native Aleppo pine forest in Greece, pointed out high mortality, exceeding 80%, during the first three years after fire, mainly in response to competition with other plants, attacks by insects and other animals, fungal infections and water and cold stress. Also in Greece, Daskalidou and Thanos (2004) found seedling mortalities ranging from 23% to 50% of the total cohort in relation to water availability during the early establishment phase after fire and during the following drought tolerant phase when germination and early seedling development and establishment take place. Trabaud et al. (1985) also reported a high mortality of young pine classes in Montpellier, France, that was probably due to dry climatic conditions that affected seedling survival; moreover, they found two areas where there was no re-colonization during the first three and four years after fire, probably due to high mortality. Pausas et al. (2003) reported a significant decrease in seedling recruitment during the 29 months

following a fire. Survival challenges seem to be especially critical far away from the large burnt canopies, where inter- and intra-specific competition are higher, and mineral nutrition is poorer (Ne'eman et al., 1992). Traubad et al. (1985) found that a light vegetation cover increases the survival of seedlings, especially during summer droughts when they are young. In a similar way, Pausas et al. (2004) observed that the main source of variation in post-fire regeneration was the number of branches and trunks covering the soil surface. Branches from burnt trees provide micro-climatic conditions that favour the establishment of pines, emphasising the importance of physical stress for tree establishment. As has been said, precipitation during the recruitment periods corresponding to the first four years following fire was slightly below the historical average for the area, and this, combined with the reduction in vegetation cover that followed the fire in our study site, probably increased insolation and water stress. This hypothesis coincides with results obtained in another study in the same area that showed that Aleppo pine seedlings growing in sites subjected to high defoliation levels (and so more exposed to high solar radiation and evapotranspiration) suffer from higher mortality (de Villalobos com. pers.).

Seedlings that were successfully established in the years after fire in our study site might have changed the micro-climatic conditions, increasing the probability of survival of those that germinated later. This could explain the peak in recruitment that started in the fourth year after fire. These more appropriate conditions for germination and seedling survival may have continued until the eighth year, when pine density increased to a level that generated density-dependent mortality resulting in a poorer representation of the following age classes. The observed drop in recruitment in 1992 might be the effect of unusually low precipitation (there was no rainfall in the period February–March 1992).

The following recruitment peak in 1993–1994 probably originated from the seeds produced by pines established soon after the fire. Traubad and de Chanterac (1985) report that the juvenile period in *P. halepensis* in France lasts for more than 10 years, whereas Thanos and Daskalakou (2000) state that this period is no longer than six years in Greece where it is apparently common to see reproductive trees of 3–4 years in populations recruited after a fire (Thanos et al., 1996). This last record coincides with the reproductive behaviour of the species in our study area where seeds are released from five year old trees and 25% of the 11 year old trees have open cones (Zalba et al., 2003).

Recruitment was high again in the last two years of our study, but the majority of these new plants growing in a fully stocked population will probably die young. This negative density-dependent effect is also reflected in the lower growth rate recorded for pines growing in the lower altitude site, where tree density is significantly higher than at the higher altitude. In the study area, tree density is higher in the foot-hills, with no difference in basal area in respect to the higher altitude site, indicating that on

average pines growing higher up in the mountains are larger despite the fact that recruitment started later in this area.

A complementary hypothesis which would explain the delay in recruitment is an unexpected low propagule pressure following the fire due to the scarcity of surviving trees in the vicinity, and also because fire occurred after the peak of seed release, during an extremely dry summer, probably killing a great number of seeds already in the soil. Total rainfall during the summer after the fire in our study area (December 1987–February 1988) was only 74 mm, which is extremely low compared with a historical average of 226.05 mm ($DS = 87.74$) for the same months during the period 1915–2005, being the second lowest rainfall in 90 years. Recent discussions about the reproductive biology of *P. halepensis* include a shift from a traditional strong emphasis on fire as the major factor affecting regeneration (Ne'eman et al., 2004). The species has usually been considered as a post-fire obligate seeder, which regenerates only from seeds in the post-fire environment, nevertheless it is now proposed that most of the seeds of *P. halepensis* are produced in regular cones that release their seeds under dry weather conditions (Lavi et al., 2005; Nathan et al., 1999; Ne'eman et al., 2004). The unusually dry conditions prior to the fire probably stimulated seed release and those seeds were then killed by fire. Moreover, another factor is that the studied population was established in an area which was previously devoid of trees, at distances ranging from 250 to 400 m from the parent plants, and therefore the area was not directly under the canopy and seed rain was probably not as intense as in other studies of Aleppo pine regeneration. It is interesting to note that this delay in recruitment did not occur in later fires that affected different sectors of the study site in December 2001, December 2002 and February 2003 (Zucchini and Zalba, 2003) when an intense recruitment was observed in the first year after fire in all cases. The main differences between these more recent fires and the one in 1987 were that in the former recruitment occurred over previously invaded areas, and not before especially dry summers.

Both reduced initial recruitment and high mortality could explain the observed density that is among the lowest values reported for Aleppo pines recovering from fire (Pausas et al., 2004).

The Pampas grasslands were almost devoid of native trees before European settlement, with the exception of a few scattered individuals restricted to habitats protected from fire, and so alien trees are a totally new life form in this ecosystem. Wildfires, which were the main cause for the absence of trees in this ecosystem, are now a significant factor for serotinous pine expansion. The invasion window triggered by fire remained opened for at least 12 years and this might be a consequence of a fire suppression policy undergoing in this region. If fires occur at intervals that largely exceed the natural disturbance regimes in the area, it is possible that the accumulation of biomass results in fires of such intensity that the restoration ability of natural

grasslands is limited, increasing the period during which they remain vulnerable to invasion. In a similar way, the absence of fire could result in a reduction of the abundance of plant species capable of rapid recovery and growth after fire, which might be the clue to the resistance of grassland to alien woody plants. Prescribed fire has been used as a management tool for controlling invasive pines in South Africa, where management actions are directed to mimic the natural frequency of this disturbance, which in turn favours the restoration and conservation of natural plant communities (van Wilgen et al., 1992). Depending on management decisions, fire can be the initiator of an invasion, or a key factor in serotinous pines control if the frequency between successive fires is greater than the period that elapses between pine recruitment and maturation and if it helps to maintain healthier natural grasslands. Fires in the study area are known to increase green biomass and net primary productivity and to produce an initial increment in the frequency of annual exotic herbs (Frangi et al., 1980), with the recovery of natural grassland structure and composition in about two years (Kristensen, 1993). Aleppo pines in our area start to release seeds at the age of five years (Zalba et al., 2003), and so initial burning frequencies from two to five years could be appropriate for effectively controlling the invasion. Further research is needed to assess burning frequencies that are compatible with long-term conservation of grassland biodiversity and prevention of reinvasion by pines. A strategy of adaptive management might be the best option for assessing the optimal burning frequency, considering the urgent need to stop pines from spreading.

5. Conclusions

Fire appears to be a key factor for triggering the invasion of *P. halepensis* in natural Pampas grassland in Argentina.

Propagule pressure and/or rainfall during the germination and seedling establishment period (March–May) seem to be critical for Aleppo pine recruitment in the area.

Seedling established in the years after fire might increase the probability of survival of those plants that germinate later, by changing micro-climatic conditions.

Despite high seed production, density-dependent mortality reduces trees survival at the end of the study.

Burning can be an effective control method if applied with a frequency that permits grassland recovery but inhibits pine maturation and seed release. In the study area this might initially be between two and five years, but more research is needed to carefully evaluate the relationship between natural grassland, fire and invasive species in order to assess an optimal fire frequency to maintain the ecosystem's natural resistance.

Acknowledgements

This work was funded by CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), CIC (Comisión

de Investigaciones Científicas de la Provincia de Buenos Aires) and Universidad Nacional del Sur, Argentina. We also wish to thank two anonymous reviewers for their suggestions which resulted in significant improvements to this manuscript.

References

- Baskin, Y., 2002. A plague of rats and rubber-vines. The growing threat of species invasions. The Scientific Committee of Problems on the Environment (SCOPE). Island Press, Washington.
- Bertonatti, C., Corcuera, J., 2000. Situación Ambiental Argentina 2000. Fundación Vida Silvestre Argentina.
- Bileca, D., Miñarro, F., 2004. Identificación de áreas valiosas de pastizal en las pampas y campos de Argentina, Uruguay y sur de Brasil. Fundación Vida Silvestre Argentina, JM Kaplan Foundation.
- Cabrera, A.L., 1976. Regiones Fitogeográficas Argentinas. In: Parodi, L.R. (Ed.), Encyclopedia Argentina de Agricultura y Jardinería. Tomo II. Edición ACME, Argentina, pp. 1–85.
- Chilvers, G.A., Burdon, J.J., 1983. Further studies on a native Australian eucalypt forest invaded by exotic pines. *Oecologia* 59, 239–245.
- Cronk, Q.C.B., Fuller, J.L., 1995. Plants invaders: A People and Plants Conservation Manual. Chapman & Hall, Cambridge University Press, UK.
- Cuevas, Y.A., Zalba, S.M., 2004. Banco de semillas de *Pinus halepensis*, una especie invasora del pastizal pampeano. In: Abstracts II Reunión Binacional de Ecología. Mendoza, Argentina, p. 251.
- Daskalakou, E.N., Thanos, C.A., 1996. Aleppo pine (*Pinus halepensis*) post-fire regeneration: the role of canopy and soil seed banks. *International Journal of Wildland Fire* 6 (2), 59–66.
- Daskalakou, E.N., Thanos, C.A., 2004. Post-fire regeneration of Aleppo pine—the temporal pattern of seedling recruitment. *Plant Ecology* 171, 81–89.
- Facelli, J.M., León, R.J.C., 1986. El establecimiento espontáneo de árboles en la Pampa. Un enfoque experimental. *Phytocoenologia* 14 (2), 263–274.
- Frangi, J.L., Bottino, O.J., 1995. Comunidades vegetales de la Sierra de la Ventana, provincia de Buenos Aires, Argentina. *Revista de la Facultad de Agronomía de La Plata* 71 (1), 93–133.
- Frangi, J.M., Ranco, G., Sánchez, N., Vicari, L.R., Rovetta, G.S., 1980. Efecto del fuego sobre la composición y dinámica de la biomasa de un pastizal de Sierra de la Ventana (Buenos Aires, Argentina). *Darwiniana* 22 (4), 565–585.
- GISP, 2005. Sudamérica Invasida. El Creciente peligro de las Especies Exóticas Invasoras. Programa Mundial sobre Especies Invasoras. Kirstenbosch, South Africa.
- Higgins, S.L., Richardson, D.M., 1998. Pine invasions in the southern hemisphere: modelling interactions between organism, environment and disturbance. *Plant Ecology* 135, 79–93.
- Hobbs, R.J., 1989. The nature and effect of disturbance relative to invasions. In: Drake, J.A., Mooney, H.A., di Castri, F., Groves, R.H., Kruger, F.J., Rejmánek, M., Williamson, M. (Eds.), *Biological Invasions: A Global Perspective*. Wiley, Chichester, England, pp. 389–405.
- Hobbs, R.J., 1991. Disturbance, a precursor to weed invasion in native vegetation. *Plant Protection Quarterly* 6 (3), 99–104.
- Inbiar, 2005. Base de Datos sobre Invasiones Biológicas Argentinas. IABIN, Inter-American Biodiversity Information Network. Proyecto I3N. <www.uns.edu.ar/inbiar>. Visited December 2005.
- Izhaki, I., Henig-Sever, N., Ne'eman, N., 2000. Soil seed banks in Mediterranean Aleppo pine forests: the effect of heat, cover and ash on seedling emergence. *Journal of Ecology* 88, 667–675.
- Kristensen, M.J., 1993. Efectos del fuego sobre la estructura de pastizales naturales (Sierra de la Ventana, provincia de Buenos Aires). Resúmenes de Medio Ambiente I Jornadas Nacionales y VI Regionales, Universidad Nacional de La Plata, La Plata, Argentina.

- Kristensen, M.J., Frangi, J.L., 1995. La Sierra de la Ventana: una isla de biodiversidad. *Ciencia Hoy* 5 (30), 25–34.
- Lavi, A., Perevolotsky, A., Kigel, J., Noy-Meir, I., 2005. Invasion of *Pinus halepensis* from plantations into adjacent natural habitats. *Applied Vegetation Science* 8, 85–92.
- Long, M.A., Grassini, C.M., 1997. Actualización del conocimiento florístico del Parque Provincial Ernesto Tornquist. Ministerio de Asuntos Agrarios Provincia de Buenos Aires y Universidad Nacional del Sur, Argentina.
- Lozon, J.D., MacIsaac, H.J., 1997. Biological invasions. Are they dependent on disturbance? *Environmental Review* 5, 131–144.
- Macdonald, I.A.W., Jarman, M.L., 1985. Invasive alien plants in the terrestrial ecosystems of Natal, South Africa. South African National Scientific Programmes Report 118, South Africa.
- Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M., Bazzaz, F.A., 2000. Biotic invasions: causes, epidemiology, global consequences and control. *Ecological Applications* 10 (3), 689–710.
- Mazia, C.N., Chaneton, E.J., Ghersa, C.M., León, R.J.C., 2001. Limits to tree species invasion in pampean grassland and forest plant communities. *Oecologia* 128, 594–602.
- Nathan, R., Safriel, U.N., Noy-Meir, I., Schiller, G., 1999. Seed release without fire in *Pinus halepensis*, a Mediterranean serotinous wind-dispersed tree. *Journal of Ecology* 87, 659–669.
- Ne'eman, G., Izhaki, I., 1998. Stability of pre- and post-fire spatial structure of pine trees in Aleppo pine forest. *Ecography* 21, 535–542.
- Ne'eman, G., Izhaki, I., 1999. The effect of stand age and microhabitat on soil seed banks in Mediterranean Aleppo pine forests after fire. *Plant Ecology* 144, 115–125.
- Ne'eman, G., Lahav, H., Izhaki, I., 1992. The effect of burned pine trees on post fire germination and establishment. In: Thanos, C.A. (Ed.), *Proceedings of the Sixth International Conference on Mediterranean Climate Ecosystems, MEDECOS VI, Plant Animal Interactions in Mediterranean-type Ecosystems*, September 23–27, 1991, Maleme, Crete Greece, pp. 361–366.
- Ne'eman, G., Goubitz, S., Nathan, R., 2004. Reproductive traits of *Pinus halepensis* in the light of fire—a critical review. *Plant Ecology* 171, 69–79.
- Pausas, J.G., Ouadah, N., Ferran, A., Gimeno, T., Vallejo, R., 2003. Fire severity and seedling establishment in *Pinus halepensis* woodland, eastern Iberian Peninsula. *Plant Ecology* 169, 205–213.
- Pausas, J.G., Ribeiro, E., Vallejo, R., 2004. Post-fire regeneration variability of *Pinus halepensis* in the eastern Iberian Peninsula. *Forest Ecology and Management* 203, 251–259.
- Rejmánek, M., 1999. Invasive plant species and invasive ecosystems. In: Sandlund, O.T., Schei, P.J., Viken, A. (Eds.), *Invasive Species and Biodiversity Management, Population and Community Biology Series*. The Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 79–102.
- Rejmánek, M., Richardson, D.M., 1996. What attributes make some plant species more invasive? *Ecology* 77 (6), 1655–1661.
- Richardson, D.M., 1988. Age structure and regeneration after fire in a self-sown *Pinus halepensis* forest on the Cape Peninsula, South Africa. *South African Journal of Botany* 54 (2), 140–144.
- Richardson, D.M., 1998. Forestry trees as invasive aliens. *Conservation Biology* 12 (1), 18–26.
- Richardson, D.M., Bond, W.J., 1991. Determinants of plant distribution: evidence from pine invasions. *The American Naturalist* 137, 639–668.
- Richardson, D.M., Higgins, S.I., 1998. Pines as invaders in the Southern Hemisphere. In: Richardson, D.M. (Ed.), *Ecology and Biogeography of Pinus*. Cambridge University Press, Cambridge, pp. 450–473.
- Richardson, D.M., Rejmánek, M., 2004. Invasive conifers: a global survey and predictive framework. *Diversity and Distributions* 10, 321–331.
- Richardson, D.M., Williams, P.A., Hobbs, R.J., 1994. Pine invasions in the Southern Hemisphere: determinants of spread and invadability. *Journal of Biogeography* 21, 511–527.
- Rouget, M., Richardson, D.M., Milton, S.J., Polakow, D., 2001. Predicting invasion dynamics of four alien *Pinus* species in a highly fragmented semi-arid shrubland in South Africa. *Plant Ecology* 152, 79–92.
- Simberloff, D., Reiva, M.A., Núñez, M., 2002. Gringos en el bosque: introduced tree invasion in a native *Nothofagus/Austrocedrus* forest. *Biological Invasions* 4, 35–53.
- Skordilis, A., Thanos, C.A., 1995. Seed stratification and germination strategy in the Mediterranean pines *Pinus brutia* and *P. halepensis*. *Seed Science Research* 5, 151–160.
- Thanos, C.A., 1999. Fire effects on forest vegetation, the case of Mediterranean pine forests in Greece. In: Eftichidis, G., Balabnis, P., Ghazi, A. (Eds.), *Wildfire Management. Algosystems SA and European Commission DGXII*, Athens, pp. 323–336.
- Thanos, C.A., 2000. Ecophysiology of seed germination in *Pinus halepensis* and *P. brutia*. In: Ne'eman, G., Trabaud, L. (Eds.), *Ecology, Biogeography and Management of Pinus halepensis and P. brutia Forest Ecosystems in the Mediterranean Basin*. Backhuys Publishers, Leiden, The Netherlands, pp. 37–50.
- Thanos, C.A., Daskalakou, E.N., 2000. Reproduction in *Pinus halepensis* and *P. brutia*. In: Ne'eman, G., Trabaud, L. (Eds.), *Ecology, Biogeography and Management of Pinus halepensis and P. brutia Forest Ecosystems in the Mediterranean Basin*. Backhuys Publishers, Leiden, The Netherlands, pp. 79–90.
- Thanos, C.A., Daskalakou, E.N., Nikolaidou, S., 1996. Early post-fire regeneration of a *Pinus halepensis* forest on Mount Parnis, Greece. *Journal of Vegetation Science* 7, 273–280.
- Trabaud, L., de Chanterac, B., 1985. The influence of fire on the phenological behaviour of Mediterranean plant species in Bas-Languedoc (southern France). *Vegetatio* 60, 119–130.
- Trabaud, L., Michels, C., Grosman, J., 1985. Recovery of burnt *Pinus halepensis* Mill. forests II. Pine reconstruction after wildfire. *Forest Ecology and Management* 13 (3–4), 167–179.
- Tsitsoni, T., Karagiannakidou, V., 2000. Site quality and stand structure in *Pinus halepensis* forests of north Greece. *Forestry* 73 (1), 51–64.
- van Wilgen, B.W., Bond, W.J., Richardson, D.M., 1992. Ecosystem management. In: Cowling, R.M. (Ed.), *The Ecology of Fynbos: Nutrients, Fire and Diversity*. Oxford University Press, Cape Town, South Africa, pp. 345–371.
- Williamson, M., 1996. *Biological Invasions*. Chapman & Hall, London.
- Zalba, S.M., 1994. Plantas leñosas exóticas en el Parque Provincial Ernesto Tornquist (Buenos Aires): Evaluación de Impacto y Propuesta de Control. Tesis de Maestría en Manejo de Vida Silvestre. Universidad Nacional de Córdoba, Argentina.
- Zalba, S.M., 2000. El pastizal pampeano, los árboles exóticos y la fauna silvestre: un problema con múltiples dimensiones. In: Bertonatti, C., Corcuera, J. (Eds.), *Diagnóstico Ambiental de la República Argentina*. Fundación Vida Silvestre, Argentina, pp. 332–337.
- Zalba, S.M., 2001. Efectos de la forestación con especies exóticas sobre comunidades de aves del Pastizal Pampeano. Tesis Doctoral en Biología. Universidad Nacional del Sur, Bahía Blanca, Argentina.
- Zalba, S.M., Villamil, C.B., 2002. Woody plant invasion in relictual grasslands. *Biological Invasions* 4, 55–72.
- Zalba, S.M., Barrionuevo, L.B., Cuevas, Y.A., de Villalobos, A.E., 2003. Restauración de ambientes naturales afectados por especies exóticas en el Parque Provincial Ernesto Tornquist. In: Abstracts II Jornadas Interdisciplinarias del Sudoeste Bonaerense, Editorial de la Universidad Nacional del Sur, Bahía Blanca, Argentina.
- Zucchini, E.E., Zalba, S.M., 2003. Efecto de un incendio sobre el establecimiento y la supervivencia de plántulas de *Pinus halepensis* en el Parque Provincial E. Tornquist. In: Abstracts XXIX Jornadas Argentinas de Botánica, San Luis, Argentina, p. 74.

Developing a sustainable agro-system for central Nepal using reduced tillage and straw mulching

Kishor Atreya^{a,*}, Subodh Sharma^b, Roshan M. Bajracharya^b, Neeranjan P. Rajbhandari^c

^aAlternative Development and Research Center (ADRC – Nepal), GPO Box 20078, Kathmandu, Nepal

^bDepartment of Environmental Science and Engineering, Kathmandu University, Nepal

^cSustainable Soil Management Program, Helvetas, Kathmandu, Nepal

Received 19 March 2006; received in revised form 24 February 2007; accepted 16 March 2007

Available online 3 May 2007

Abstract

In Nepal, soil erosion under maize (*Zea mays*) agro-ecosystems is most critical during the pre-monsoon season. Very few field experiments have been conducted on reduced tillage and rice straw (*Oryza sativa*) mulching, although these conservation approaches have been recommended. Thus, a five replicate field experiment was established in 2001 at Kathmandu University (1500 m above sea level) on land with 18% slope to evaluate the efficiency of reduced tillage and mulching on soil and nutrient losses and maize yield. The results showed non-significant differences among conservation approaches on runoff and maize yield. Mulching and reduced tillage significantly lowered annual and pre-monsoon soil and nutrient losses compared to conventional tillage. Soil organic matter (SOM) and nitrogen losses associated with eroded sediment were significantly higher in conventional tillage. However, due to limited availability and high opportunity cost of rice straw, reduced tillage would be a better option for soil and nutrient conservation without sacrificing economic yield in upland maize agro-ecosystems.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Himalayan region; Mulching; Reduced tillage; Residue; Runoff; Soil erosion

1. Introduction

Accelerated soil erosion has been a chronic problem ever since the dawn of settled agriculture and has adverse economic and environmental impacts (Lal, 1998). For rainfed agro-ecosystems in Nepal, soil erosion is most severe during the pre-monsoon season (Ries, 1995; Schreier and Shah, 1995; Sherr and Yadav, 1996; Tripathi et al., 2000a; Schreier et al., 2001; Atreya, 2002; Gardner and Gerrard, 2003; Atreya et al., 2006). During the pre-monsoon season, rainfall events tend to be of high intensity accompanied by strong winds and occasionally hail. Further, before the onset of pre-monsoon, the agricultural lands are tilled several times, which makes soil dry, bare and devoid of vegetative cover. Thus, soil and nutrient losses from the upland terraces are usually high even before the onset of actual monsoon. Gardner and Gerrard (2003)

clearly state that “in the case of middle hills, [soil and nutrient] losses will be seriously underestimated if the small number of high magnitude storms in late April and May after plowing but before development of weed cover, are not measured”. The rate of soil loss gradually diminishes as the soil cover from maize (*Zea mays*) crop and other vegetation increases following the onset of the monsoon season.

There is a growing literature that indicates conventional tillage to be the dominant practice leading to accelerating soil erosion especially in slopping farms of the Himalaya. In the Indian Himalaya, Sharma et al. (2001) and Jain et al. (2000) found very high rates of overland flow, soil and nutrient losses from bare agricultural fields following traditional practices compared to other land uses. In Nepal, the tillage intensity is a major factor that enhances soil erosion during pre-monsoon in upland agro-ecosystems. Thapa (1996) argues that farmlands have been undergoing unsustainable rates of soil erosion and soil nutrients depletion due to frequent hoeing and plowing.

*Corresponding author. Tel.: +977 1 6211252.

E-mail address: k.atreya@gmail.com (K. Atreya).

Farmers believe that the exposure of land to the sun, rain and air for a long period helps to improve soil fertility (Thapa and Poudel, 2002). Therefore, they plow lands immediately after the harvest of crops and leave them exposed without any vegetative cover. Atreya et al. (2006) pointed out that Nepalese farmers in central hills till land three times before maize sowing viz. (i) before the onset of pre-monsoon season (nearly 2 months earlier, February/March), (ii) after 20–25 days (March/April) of first tillage, and (iii) at the time of maize sowing, which is after 2 to 3 pre-monsoon rainfall events (April/May). Due to this tillage operation, soils become extremely dry and dusty, without any vegetative cover allowing the high intensity pre-monsoon rain wash away a significant portion of top soil.

As alternatives to the conventional tillage system, various studies (Acharya, 1999; Rajbhandari, 2000; Bajracharya, 2001; Ransom and Rajbhandari, 2001; Gardner and Gerrard, 2003) reported minimum tillage and mulching to be potential practices for mitigating soil erosion during pre-monsoon. Tripathi et al. (2000b) also recommended developing options such as mulching and reduced tillage in order to enhance sustainability of agriculture in the Himalaya. Use of crop residue mulch has profound beneficial effects on soil properties, microclimates, and agronomic productivity. Mulching conserves soil and water, improves soil structure, improves soil organic matter content, adds cations to the soil, regulates soil temperature, and restore the productivity of degraded lands (Srivastava et al., 1993). Mulching dissipates raindrop energy and thereby decreases soil detachment. It also encourages infiltration rates by minimizing surface sealing, and thus reduces runoff and soil loss (Adekolu et al., 2007). Edwards et al. (2000) reported that mulching of barley (*Hordeum vulgare*) straw at 4 t/ha reduced soil loss by almost 50% compared to control. Tolk et al. (1999) also noted that mulching at 6.7 t/ha increased grain and stover yields of maize by 17% and 19%, respectively, compared to bare treatment. Similarly, Smolikowski et al. (2001) and Barton et al. (2004) also reported reduced runoff and soil loss by mulching, but no effects on crop yields. Reduced tillage also decreases runoff through reducing surface crusting and increasing infiltration (Cassel et al., 1995; Acharya, 1999; FAO, 2000). The practice of crop residue retention and minimum tillage, in association with basal fertilizer application, increases the supply of carbon, nitrogen and enhances the degree of synchronization between crop demand and nitrogen supply. Kushwaha and Singh (2005) pointed that retention of a small fraction of above-ground biomass and its incorporation in the soil through minimum tillage enhances crop productivity and soil fertility in a cereal based tropical dry land agro-ecosystem.

The practice of covering the plowed land by crop residues for soil and nutrient conservation has been documented in hills of Nepal (Paudel and Thapa, 2001; Khanal and Watanabe, 2006). However, only farmers with

a relatively large household size and small land holdings used this method as availability of residue was a constraint for others (Paudel and Thapa, 2001). Applications of mulching and reduced tillage for soil and nutrient conservation during the pre-monsoon season have been recommended as potential researchable options in Nepal. However, to date very few field experiments have been conducted on reduced tillage and rice-straw (*Oryza sativa*) mulching. Further, crop yield in sloping upland has been in decline in most areas, at the same time farmers continue to cultivate despite negative benefit–cost ratios as there are few other alternatives for livelihood. There is an urgent need to increase/stabilize crop yields and reverse this trend. Therefore, this study evaluated the effect of the two conservation practices, namely reduced tillage and mulching on soil and nutrient losses and maize yield in maize – mustard (*Brassica juncea*) and maize + soybean (*Glycine max*) – mustard agro-ecosystems.

2. Materials and methods

2.1. Runoff plots and experimental design

Twelve runoff plots (two replicates), each of 6 m × 3 m with an average slope of 18% and northeast aspect, were established within the Kathmandu University premises (1500 m above sea level) to measure soil and nutrient losses and crop yield during 2001. Another 18 productivity plots (three replications) were also setup in nearby farmers' field, each of dimensions 3 m × 3 m, for crop yield only. This was done basically to increase the replication and thus, to minimize the experimental errors for crop yield. It should be noted that the university is situated in a rural area, in Kavreplanchowk district of Nepal (Fig. 1) and the experimental site was farmed by locals in a traditional manner. The runoff plots were bordered by galvanized iron sheets, 20 cm into the soil to prevent horizontal movement of soil water and 25 cm above the ground to check soil particle splash between plots. The collection system of runoff plots consisted of two drums at the low end of each plot connected with a five-slot splitter that drains one-fifth of the runoff into the second drum (see Fig. 2). The sediment mixture of the collection drums were stirred vigorously before sampling after each storm generating soil loss. Each sample was filtered using Whatman No. 42 filter



Fig. 1. Location of the study area.

paper, oven dried at 105 °C for 24 h, for calculation of soil loss. These oven dried samples were used for soil nutrient analysis.

The experiment was conducted as a factorial randomized complete block design. The three main factors were conventional tillage (CT), CT with rice-straw mulching (5 t/ha), and reduced tillage (RT). Main plots were split on the basis of cropping pattern, i.e., maize mono-cropped and maize and soybean mix-cropped, i.e., grown simultaneously. Thus, there were altogether six treatments viz. (1) conventional till, sole maize, (2) conventional till, maize + soybean, (3) conventional till and mulching, sole maize, (4) conventional till and mulching, maize + soybean, (5) reduced till, sole maize, and (6) reduced till, maize + soybean.

2.2. Benchmark soil properties of the runoff and productivity plots

A baseline of soil physical properties of the runoff plots were established at the inception of the experiment for future comparison during monitoring of soil and nutrient losses. Slope and aspect, soil pH, and soil chemical parameters like organic matter (OM), total nitrogen (N), and exchangeable potassium (K_2O) were similar at both sites, with the exception of available phosphorus (P_2O_5), which was significantly higher in the productivity plots (Table 1). Based on USDA Soil Taxonomy (SMSS, 1985), soils of the runoff plots were classified as mixed mesic lithic

Dystocrept. Soils of the runoff plots were sandy loam having dark yellowish brown (10 YR 4/4) surface layer. Soil depth of top layer was 15 cm and that of entire profile was only 50 cm. The initial gravimetric soil water content, and infiltration rate, determined before the onset of pre-monsoon in the year 2001 were 5.3% and 4.31 cm/h, respectively. Double ring infiltrometer (Gupta, 2000) was used to determine infiltration due to simplicity and wide adoption even in small areas (Rao and Sharma, 1997). Slopes of erosion plots ranged from 16% to 21%. Bulk density was 1.38 g/cm³.

2.3. Soil chemical analyses

The study was concerned with the loss of primary nutrients namely: N, P_2O_5 , K_2O and OM. N is the most important limiting nutrient to plant growth and development in Nepal (NARC, 2000). Due to high solubility and transport, some researchers (Shah, 1996; Gardner et al., 2000) observed a decreasing trend of K in soils. Brown and Schreier (2000) also found significant deficits in N and P, especially in maize agro-ecosystems. For this study, soil pH (1:1), soil OM (Walkley and Black), total N (Modified Kjeldahl digestion), available P (Modified Olsen) and exchangeable K (Flame Photometer) of the eroded sediments were analyzed (Farm Resources and Systems Research Division, 1980).

2.4. Agronomic practices

Plots under conventional tillage, as generally done by local farmers, were tilled thrice (1) before the onset of pre-monsoon season (February/March); (2) after 20–25 days of first tillage (March/April), and (3) at the time of maize sowing, which is after 2 to 3 pre-monsoon rainfall events (April/May). Plots under mulching were also tilled thrice, and after sowing maize, rice straw at 5 t/ha was surface applied along the slope and straw was incorporated into soil after first weeding (30 days after planting, DAP). Plots under reduced tillage were tilled only once at the beginning by opening narrow furrows across slope using a hand-hoe i.e., during maize sowing. First and second weeding (60 DAP) was done in all plots, even in reduced till plots, because here the objective of the reduced tillage was to check soil erosion during pre-monsoon rain, which is

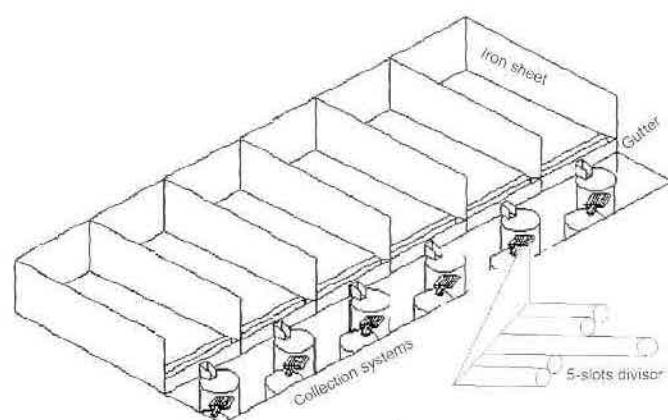


Fig. 2. Layout of an erosion block and runoff collection systems.

Table 1
Comparison of soil properties of the runoff and productivity plots

	No. of plots	Aspect	Slope (%)	pH	SOM (%)	N (%)	P_2O_5 (kg/ha)	K_2O (kg/ha)
Runoff plots	12	Northeast	16–21	3.23	0.6142	0.0461	66	330
Productivity plots	18	Northeast	12–17	3.08	0.7889	0.0471	127	421
t-test				ns	ns	ns	**	ns

Note: ns = not significant.

** = significant at probability level of 0.01.

believed to be most serious in terms of soil and nutrient loss. Farmers in hills of Nepal cultivate land by hand-hoe or ox-drawn wooden plough. Due to the small size of plots used in this study, a small pointed hand-hoe was employed to till the land. Hand-hoe technique is commonly used on terraces narrower than about 1.5 m, or used by those farmers who possess only small parcels of land and do not own oxen (Bajracharya, 2001). Most importantly, this technique is used intensively in and around the Kathmandu valley area including the study site due to religious belief, which does not permit use of oxen for land plow.

Test crops were maize cultivar 'Arun 2' and soybean cultivar 'Ransom'. Maize was planted with 75 cm × 25 cm spacing across slope for all treatments. For the maize + soybean mixed-cropped, soybean was planted in between rows of maize, spacing 25 cm, under reduced till and mulch plots. However, in conventional plots, the farmer practice of planting at random spacing in plots across slope was followed. Test crops were planted on 31 April and harvested on 25 August in 2001, whereas, because of sufficient soil moisture due to early rain in 2002, crops were planted on 10 April and harvested on 9 August. For additional details refer to Atreya et al. (2006).

2.5. Data analysis

Statistical analyses were performed using Statistix for Windows, version 2.2 (Analytical Software, 2000). Soil and nutrient losses were monitored only from two replications; therefore, two sample *t*-test was performed to assess the differences between conservation approaches. However, analysis of variance (ANOVA) was also used to describe

the differences of treatment means in terms of annual runoff, soil and nutrient loss and crop yield. Seasonal dynamics of soil and nutrient losses were also described with the help of *t*-test and ANOVA at 0.05 probabilities.

3. Results and discussion

3.1. Precipitation at experimental site

The amount of rainfall received in different months is presented in Fig. 3. Total rainfall received were 1445 and 2143 mm in 2001 and 2002, respectively, indicating a dry year in first and wet year in the latter. During pre-monsoon season (March–May), 15% of annual rainfall occurred in 2001 and 16% of annual rainfall in 2002. Total numbers of rain events that eroded soil from the runoff plots were 30 in 2001 and 24 in 2002. Of these, only seven (in 2001) and 11 (in 2002) events occurred during pre-monsoon season.

3.2. Runoff

The unusual conditions created by runoff plot construction during the establishment year 2001 should be noted. Soils along plot boundaries were substantially disturbed and there was very little cover (upland weeds), even in reduced till plots due to trampling during plot establishment. This likely affected the system outputs; especially increased soil and nutrient losses from reduced till plots in 2001.

With this caution, statistical analysis showed that the conservation approaches were non-significant for annual

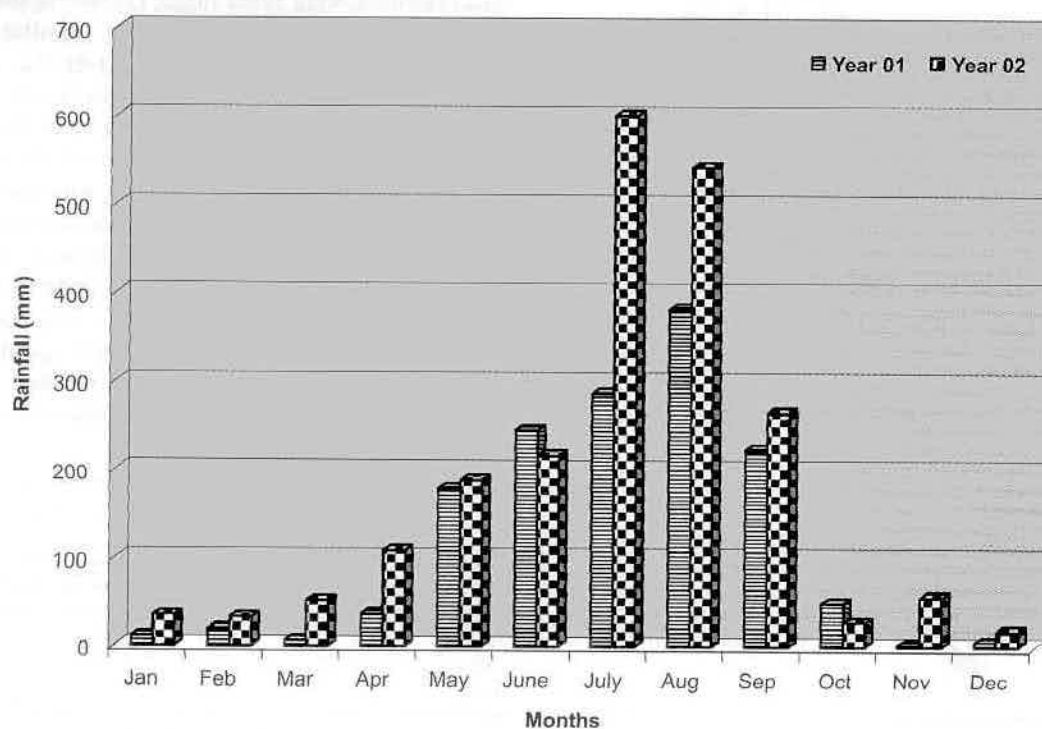


Fig. 3. Monthly rainfalls at experimental site (2001–2002).

Table 2

Treatment and seasonal dynamics on runoff, soil and nutrient losses in hills of central Nepal (2001–2002)

Treatment	Annual runoff (mm)	Soil loss (t/ha)			SOM loss (kg/ha)			Annual nutrient losses (kg/ha)		
		Pre-monsoon	Monsoon	Annual	Pre-monsoon	Monsoon	Annual	Total N	P ₂ O ₅	K ₂ O
Conventional till, sole maize	129	12.4	3.4	15.8	262.7	74.1	336.9	18.8	1.9	4.0
Conventional till, maize + soybean	117	14.9	2.3	17.3	269.6	41.7	311.4	18.8	1.9	4.4
Mulching, sole maize	123	5.7	3.5	9.3	108.7	61.9	170.6	9.6	1.0	2.3
Mulching, maize + soybean	110	5.6	2.7	8.3	102.6	49.5	152.1	9.1	0.8	2.2
Reduced till, sole maize	128	7.2	3.5	10.8	167.7	59.2	226.8	11.9	0.9	2.6
Reduced till, maize + soybean	134	7.9	3.4	11.4	146.8	60.4	207.2	11.7	1.1	2.7
F-test										
Replication	ns	ns	*	ns	ns	ns	ns	ns	ns	*
Treatment (T)	ns	**	ns	**	**	ns	**	**	**	**
Year (Y)	ns	ns	**	**	ns	**	ns	ns	**	**
T × Y	ns	ns	ns	ns	*	ns	ns	ns	*	*
LSD for T (0.05)	—	2.1	—	3.1	55.0	—	64.7	3.9	0.3	0.7

*Significant at probability level of 0.05.

**Significant at probability level of 0.01.

ns = not significant.

runoff; however, reduced tillage produced slightly higher runoff than conventional tillage and mulching (Table 2). Highest runoff was observed in reduced till maize + soybean (134 mm) followed by conventional till sole maize (129 mm). Previous studies claimed that reduced tillage significantly decreases runoff through reducing surface crusting and increasing infiltration (Cassel et al., 1995; Acharya, 1999; FAO, 2000). However, in the present study, slightly higher runoff from reduced tillage plots could be due to compaction from trampling and less porous soil conditions due to lack of complete tillage. A study in Nigeria (Adekalu et al., 2006) under rainfall simulator showed that as the compaction level increases runoff increased and soil loss decreased significantly. In conventional tillage, the plots were freshly plowed having higher surface roughness and porosity than reduced till plots; therefore, retained and absorbed more water. Atreya et al. (2006) and Basic et al. (2001) have also made similar arguments. Moreover, farmer practice of maize sowing automatically tends to create small ridges on the soil surface across slope (Atreya, 2002; Atreya et al., 2006), which may have increased time for infiltration, thus, decreased total runoff.

On the other hand, the reductions in runoff with mulch was small (5–18%), but consistent. For example, conventional till sole maize was nearly 5% lower with mulch than without (123 mm vs. 129 mm) and the conventional till maize + soybean with mulch was 18% lower than reduced till maize + soybean without mulch (110 mm vs. 134 mm). However, as shown in previous studies, the anticipated

reduction in runoff due to mulching was not observed. Lal (1976a,b) showed negligible runoff by mulching compared to bare-fallow. Kukal et al. (1993) also documented 58% reduction in runoff by the application of 4 t/ha of mulch. Bhatt and Khera (2006) also showed that compared with un-mulched, mulch spread over whole plot reduced runoff by 33%. In this study straw mulching produced comparable runoff with other treatments which was presumably due to the length and placement of straw mulch. About 50 cm long rice straw was surface applied along the slope, which may have facilitated rainfall to drain out of the plots into the collection systems.

3.3. Soil loss

Annual soil loss was significantly affected by treatments and by years. Soil loss during pre-monsoon season was only affected by treatments, whereas, monsoon loss was affected by replication and by years (Table 2). Conventional till maize + soybean produced the highest annual (17.3 t/ha) as well as pre-monsoon soil loss (14.9 t/ha), followed by conventional till sole maize (15.8 and 12.4 t/ha, respectively). A significant difference by years was observed for annual and monsoon soil loss, but not for pre-monsoon loss because of yearly variation in rainfall amount. As compared to conventional tillage practice, annual and pre-monsoon soil losses were significantly lowered by mulching and reduced tillage (Fig. 4 and Table 3). Reduced tillage minimized 33% of annual and nearly 50% of pre-monsoon soil loss. Similarly,

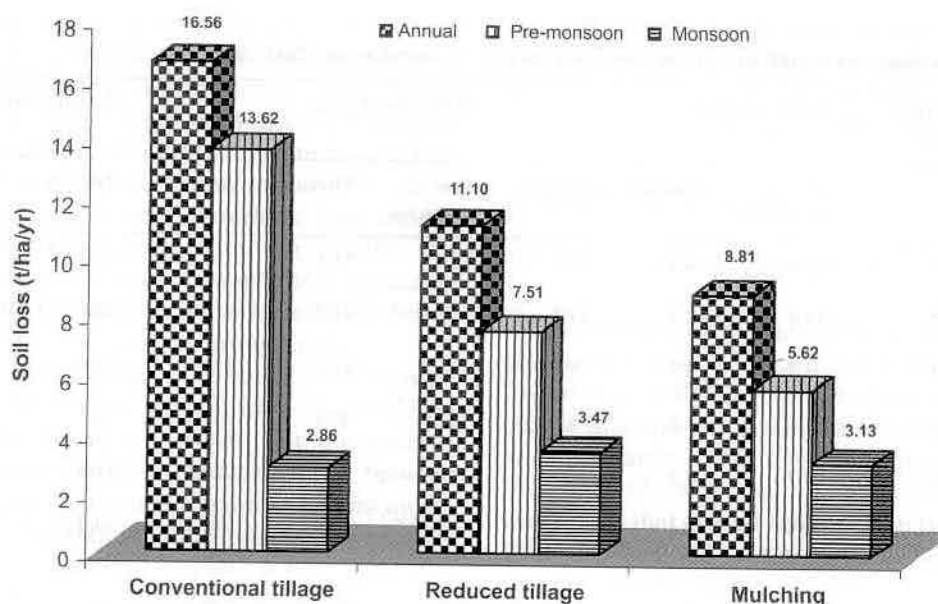


Fig. 4. Effects of conservation approaches on soil loss in central Nepal.

Table 3
Two-sample *t*-test results^a

Soil and nutrient loss	<i>t</i> -test values	
	CT versus RT	CT versus mulching
Annual runoff	−0.74 ^b	0.652 ^b
Annual soil	3.27**	9.02**
Pre-monsoon ^c	5.90**	10.81**
Monsoon	−0.67 ^b	−0.36 ^b
Soil organic matter	3.91**	8.01**
Total N	5.53**	9.04**
Available P ₂ O ₅	2.42*	2.61*
Exchangeable K ₂ O	3.45**	7.53**

*Significant at probability level of 0.05.

**Significant at probability level of 0.01.

^aDegree of freedom = 14.

^bNot significant.

^cSoil loss during March–May was considered as pre-monsoon and that of June–October as Monsoon.

mulching also reduced annual soil loss by nearly 53% and pre-monsoon soil loss by 59%. Annual runoff was similar for all treatments, but soil loss was significantly different. This was reflected in differences in sediment concentration among the treatments. Sediment concentrations per thousand ml of runoff were observed to be the highest for conventional till followed by reduced till and mulching.

In Central Nepal, Maskey et al. (1992) also found a 77% reduction in soil loss by minimum tillage as compared to the farmers' practice. Soil loss during pre-monsoon season was significantly higher in terms of amount and intensity. During this season, only seven (out of 30) and 11 (out of 24) rainfall events occurred in 2001 and 2002, respectively. Significant soil loss during pre-monsoon period from

conventional tillage was due mainly to freshly plowed conditions with minimal vegetation cover. Freshly plowed field is susceptible to erosion and 40–50 days is required before maize crop cover can provide enough cover protection to reduce soil loss. In hills of central Nepal, farmers plant the land only after the first few pre-monsoon rain events. The mid-hills also experience local ascending winds (March–May) that lead to violent thunderstorms accompanied by hail before the arrival of actual monsoon front (Shah and Friend, 1992). This leads to significantly higher soil losses during pre-monsoon season compared to monsoon. It is not possible to establish early live vegetative cover in upland areas due to insufficient moisture and lack of irrigation; therefore, it is necessary to modify the current tillage practices for upland agriculture in the hills of Nepal in order to conserve soil resources.

3.4. Soil organic matter (SOM) and nutrient losses

Treatments were significantly different in annual and pre-monsoon SOM loss (Table 2). SOM loss was the highest in conventional till sole maize (337 kg/ha) followed by conventional tillage maize + soybean (311 kg/ha) and reduced till sole maize (227 kg/ha). The interaction between treatments by year was also significant, especially in pre-monsoon, indicating inconsistent treatment responses to SOM losses in the two years. This was likely due to the differences in soil surface conditions and vegetative (weed) cover during the two years. In 2002 significant reduction in soil and SOM loss was observed in reduced till plots because of presence of weeds during pre-monsoon, which were destroyed in the first year due to plot construction (Atreya et al., 2006). As compared to conventional tillage, conservation approaches reduced

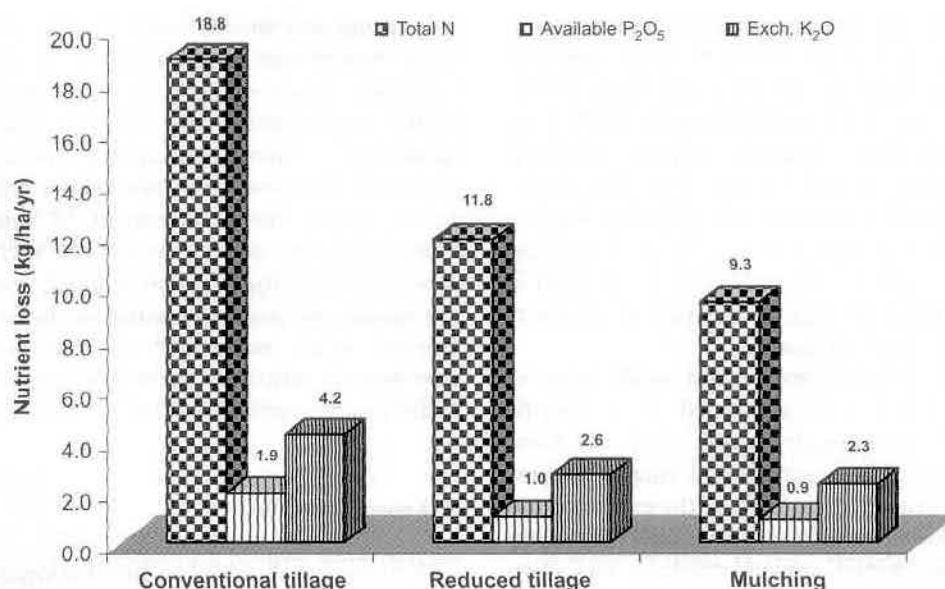


Fig. 5. Effects of conservation approaches on soil nutrient losses in central Nepal.

SOM from maize agro-ecosystems, which was also statistically significant (Table 3).

Although there were treatments effects on nutrient losses by water erosion (Table 2), nutrient losses in eroded sediment, for this study, appeared to be minor with the exception of nitrogen (Fig. 5). Sediment bound nitrogen loss was highest in conventional till plots, and lowest in mulched plots. Maskey et al. (1992), Pilbeam et al. (2000), and Brown et al. (1999) also argue that SOM and nitrogen losses through water erosion during the rainy season are serious problems in Nepal.

This study showed that mulching and reduced tillage was effective in controlling soil and nutrient losses from upland hills of Nepal. However, it should be noted that natural deposition under traditional cultivation practices was not monitored in the plot experiment (Gardner and Gerrard, 2003) and re-deposition of sediments may occur in lower terraces (Schreier et al., 2001), which was also not accounted in plots. Thus, it is important to understand the dynamics of soil and nutrient transports within as well as among the terraces. Therefore, quantification of soil and nutrient losses from multiple terraces is recommended. However, the seasonal interactions on the intensity and magnitude of soil and nutrient losses at plot scale is not only important for the timing of conservation planning, but also important for understanding and selecting types of conservation measures, especially in the fragile hillsides of Nepal.

Although rice straw was observed to be effective in controlling soil and nutrient loss, farmers tend not to accept this practice. Most farmers, especially in the mid-hills, use rice-straw as fodder for livestock; some farmers use it as thatch material for roof, and if available in surplus, they sell it to neighbors and industries, such as paper factories. The value of soil protection and improved

Table 4

Mean treatment effects on maize height, grain and stover yield

Treatments	Cob height (cm)	Plant height (cm)	Grain yield (t/ha)	Stover yield (t/ha)
Conventional till, sole maize	107	215	3.82	4.10
Conventional till, maize + soybean	122	224	3.98	3.89
Mulching, sole maize	101	208	4.10	3.16
Mulching, maize + soybean	102	200	4.00	3.20
Reduced till, sole maize	80	169	3.49	2.74
Reduced till, maize + soybean	86	187	3.94	2.72
F-Test	*	*	ns	ns
LSD 0.05	24	44	—	—

*Significant at probability level of 0.05.

ns = not significant.

crop yields would need to exceed the cost of residue to be feasible. Due to limited availability and high cost of rice straw, further research is needed to quantify the economic benefits of soil and nutrient conservation provided by rice straw mulch and/or evaluate alternative mulch materials.

3.5. Maize parameters

All treatments produced statistically similar grain and stover yields (Table 4). Grain yield was highest in mulching plots. Even under reduced-till condition, there was no significant sacrifice of grain yield as compared to conventional till. However, in previous studies conducted in Nepal, minimum/no tillage was found unsuitable for crop

yield and thus, farmers did not accept minimum tillage due to low crop yield (Maskey et al., 1992). In some instances the crop even failed completely (Khadka and Shah, 1987). But a recent review paper in China (Wang et al., 2007) have showed increased crop yields, however, farmers' adoption of such tillage practices is still limited. For this study, reduced tillage significantly lowered cob and plant heights and as a result, dry stover yield was also lower in reduced till plots. The shorter plant height and lower stover yield in reduced till plots could be due to restriction in roots development at early stages of growth.

The present study revealed good crop yields even in reduced till plots, which was attributed to a specific combination of soil type (sandy loam), crop rotations and agronomic practices. Conservation tillage systems are specific to soil characteristics. Soils at the experimental sites were sandy loam, which is favorable for crop establishment. During winter season, mustard was sown in each plot, and residue was taken out completely. In preparing land for mustard, maize stubbles were removed (generally these are used as fuel materials by farmers) and roots were incorporated into the soil. The cultivation of mustard in winter season further facilitates tillage operations for maize crop, even under reduced till plots. Furthermore, the reduced tillage here is different from that in other studies. Firstly, tillage intensity was significantly minimized (only during the pre-monsoon period), soil was less disturbed by opening small furrows at 75 cm spacing during maize planting (only once against thrice in conventional tillage), which could restrict root development at early stages. Secondly, reduced till plots were also tilled during first and second weeding. Weeding operations after 30 and 60 days of planting and application of chemical fertilizers (top dressing) at the first weeding in reduced till plots helped plant roots development and maintained yield. Although such weeding operations and fertilizer application were done in all treatments, better response was found in reduced tilled plots. Thus, a slight modification in tillage intensity and fine tuning agronomic practices may be required to conserve plant available resources in cultivating upland maize agro-environment in hills of Nepal.

4. Conclusion and recommendations

Mulching and reduced tillage significantly lowered soil and nutrient losses compared to conventional tillage. Soil nitrogen and organic matter losses, as particulate organic matter associated with eroded sediment, was quite high in conventional tillage. Thus, there is an urgent need to modify conventional tillage practices to protect plant available nutrients in agricultural land, especially in hills of central Nepal, which has 42% of the total farmland and 44% of the population of the nation. Most of the soil loss occurred during high intensity pre-monsoon rain under conventional tillage because of freshly tilled fields with minimal vegetation cover. Future research thrusts on

conserving soil should focus on this critical period. Rice straw mulch was found effective in conserving soil and nutrients, however, due to its limited availability and higher opportunity cost, farmers may not adopt this technology. Alternative mulch materials need to be explored. This research showed that reduced tillage could be an option for management of soil resources in the Himalayan agro-ecosystem without sacrificing crop yields. It was further observed that a mulch residue was beneficial for runoff and soil loss control in the hill region of Nepal. Further work is needed to determine the long-term cost-benefit ratios for these practices in order to develop a sustainable agro-ecosystem for this region.

Acknowledgments

Authors are thankful to International Maize and Wheat Improvement Centre (CIMMYT), South Asia Regional Office, Kathmandu, Nepal for the financial support. We are also grateful to J.K. Ransom, B.P. Tripathi, and P.B. Shah for their support during the study period.

References

- Acharya, G.P., 1999. Review on soil and soil fertility losses from the cultivated hill lands of Nepal and their conservation. Lumle Review Paper No 99/1. Agricultural Research Station, Lumle, Kaski, Nepal.
- Adekalu, K.O., Okunade, D.A., Osunbitan, J.A., 2006. Compaction and mulching effects on soil loss and runoff from two southwestern Nigeria agricultural soils. *Geoderma* 137, 226–230.
- Adekalu, K.O., Olorunfemi, I.A., Osunbitan, J.A., 2007. Grass mulching effect on infiltration, surface runoff and soil loss of three agricultural soils in Nigeria. *Bioresource Technology* 98, 912–917.
- Analytical Software, 2000. Statistix for Windows, Version 2.2. USA.
- Atreya, K., 2002. Management effects on soil and nutrient losses and crop productivity in maize-based cropping systems of the central Nepal mid-hills. M.Sc. Thesis, Kathmandu University, Nepal.
- Atreya, K., Sharma, S., Bajracharya, R.M., Rajbhandari, N., 2006. Applications of reduced tillage in hills of central Nepal. *Soil & Tillage Research* 88, 16–29.
- Bajracharya, R.M., 2001. Land preparation: an integral part of farming systems in the mid-hills of Nepal. *Nepal Journal of Science and Technology* 3, 15–24.
- Barton, A.P., Fullen, M.A., Mitchell, D.J., Hocking, T.J., Liu, L., Bo, Z.W., Zheng, Y., Xia, Z.Y., 2004. Effects of soil conservation measures on erosion rates and crop productivity on subtropical Ultisols in Yunnan Province, China. *Agriculture, Ecosystem and Environment* 104, 343–357.
- Basic, F., Kisić, I., Butorac, A., Nestroy, O., Mesic, M., 2001. Runoff and soil loss under different tillage methods in Stagnic Luvisols in central Croatia. *Soil & Tillage Research* 62, 145–151.
- Bhatt, R., Khera, K.L., 2006. Effect of tillage and mode of straw mulch application on soil erosion in the submontaneous tract of Punjab, India. *Soil & Tillage Research* 88, 107–115.
- Brown, S., Schreier, H., 2000. Nutrient budget: a sustainability index. In: Allen, R., Schreier, H., Brown, S., Shah, P.B. (Eds.), *The People and Resource Dynamics Project: The First Three Years (1996–1999)*, ICIMOD, Kathmandu, Nepal.
- Brown, S., Schreier, H., Shah, P.B., Lavkulich, L.M., 1999. Modeling of soil nutrient budgets: an assessment of agricultural sustainability in Nepal. *Soil Use and Management* 15, 101–108.

- Cassel, D.K., Raczowski, C.W., Denton, H.P., 1995. Tillage effects on corn production and soil physical conditions. *Soil Science Society of America Journal* 59, 1436–1443.
- Edwards, L., Burney, J.R., Richter, G., MacRae, A.H., 2000. Evaluation of compost and straw mulching on soil-loss characteristics in erosion plots of potatoes in Prince Edward Island, Canada. *Agriculture, Ecosystem and Environment* 81, 217–222.
- Farm Resources and Systems Research Division, 1980. Standard methods of analysis for soil, plant tissue, water and fertilizer. Philippines Council for Agriculture and Resources Research, Los Banos, Laguna.
- Food and Agriculture Organization (FAO), 2000. Manual on integrated soil management and conservation practices. FAO Land and Water Bulletin No 8. International Institute of Tropical Agriculture and FAO.
- Gardner, R.A.M., Gerrard, A.J., 2003. Runoff and soil erosion on cultivated rainfed terraces in the middle hills of Nepal. *Applied Geography* 22, 23–45.
- Gardner, R., Mawdesley, K., Tripathi, B.P., Gaskin, S., Adams, S., 2000. Soil erosion and nutrient loss in the middle hills of Nepal (1996–1998). ARS, Lumle and SSD (NARC) and Queen Mary and Westfield College, University of London, UK.
- Gupta, P.K., 2000. Methods in Environmental Analysis: Water, Soil and Air. Agro bios, Jodhpur, India.
- Jain, A., Rai, S.C., Sharma, E., 2000. Hydro-ecological analysis of a sacred lake watershed system in relation to land-user/cover change from Sikkim Himalaya. *Catena* 40, 263–278.
- Khadka, S.R., Shah, S.C., 1987. Effect of different tillage practices on spring maize production. *Journal of the Institute of Agriculture and Animal Science* 9, 91–95.
- Khanal, N.R., Watanabe, T., 2006. Abandonment of agricultural land and its consequences: a case study in the Sikles area, Gandaki basin, Nepal Himalaya. *Mountain Research & Development* 26, 32–40.
- Kukul, S.S., Khera, K.L., Hadda, M.S., 1993. Soil erosion management on arable lands of submontane Punjab, India: a review. *Arid Soil Research Management* 7, 369–375.
- Kushwaha, C.P., Singh, K.P., 2005. Crop productivity and soil fertility in a tropical dryland agro-ecosystem: impact of residue and tillage management. *Experimental Agriculture* 41, 39–50.
- Lal, R., 1976a. Soil erosion on alfisols in western Nigeria, i. effects of slope, crop rotation and residue management. *Geoderma* 16, 363–375.
- Lal, R., 1976b. Soil erosion on alfisols in western Nigeria, ii. effects of mulch rates. *Geoderma* 16, 377–387.
- Lal, R., 1998. Soil erosion impact on agronomic productivity and environmental quality. *Critical Review, Plant Science* 17, 319–464.
- Maskey, R.B., Joshi, D., Maharjan, P.L., 1992. Management of slopping lands for sustainable agriculture in Nepal, in: IBSRAM (International Board for Soil Research and Management) (Ed.), Technical Report on the management of slopping lands for sustainable agriculture in Asia. Phase I, 1988–1991, Network Document No. 2, Thailand.
- Nepal Agricultural Research Council (NARC), 2000. Annual Report 1998/1999. NARC, Khumaltar, Nepal.
- Paudel, G.S., Thapa, G.B., 2001. Changing farmers' land management practices in the hills of Nepal. *Environmental Management* 28, 789–803.
- Pilbeam, C.J., Tripathi, B.P., Sherchan, D.P., Gregory, P.J., Gaunt, J., 2000. Nitrogen balances for households in the mid-hills of Nepal. *Agriculture, Ecosystem and Environment* 79, 61–72.
- Rajbhandari, N.P., 2000. Declining soil fertility constraints maize production in the hills: a review of recent surveys of farmers practices, perceptions and conceptualizing a basis for proper targeting of maize research in the hills. In: Tripathi, B.P., Rajbhandari, N.P., Ransom, J.K. (Eds.), Improved Soil Fertility Management for Sustainable Maize Production: Proceedings of a workshop group meeting of the Hill Maize research project. Nepal Agricultural Research Council/HMRP/CIMMYT, Nepal.
- Ransom, J.K., Rajbhandari, N.P., 2001. Increasing the productivity and sustainability of maize-based cropping systems in the hills of Nepal. In: Maize Program (Ed.), Maize Research Highlights 1999–2000. International Maize and Wheat Improvement Center (CIMMYT), Nepal.
- Ries, J.B., 1995. Does soil erosion in the high mountain region of the eastern Nepalese Himalayas affect the plains? *Physics and Chemistry of the Earth* 20, 251–269.
- Roa, K.P.C., Sharma, S., 1997. Water infiltration into soil. In: Laryea, K.B., Pathak, P., Katyal, J.C. (Eds.), Measuring Soil Processes in Agricultural Research. Technical Manual Number 3, ICRISAT and ICAR, India.
- Schreier, H., Shah, P.B., 1995. Understanding degradation processes in the middle mountain in Nepal. In: Shengji, P. (Ed.), Rehabilitation of Degraded Lands in Mountain Ecosystems of the Hindu-Kush-Himalayan Region. International Centre for Integrated Mountain Development (ICIMOD), Nepal.
- Schreier, H., Brown, S., Shah, P.B., Shrestha, B., Nakarmi, G., Allen, R., 2001. Human interactions in soil and geomorphic processes in Nepal: the role of soil fertility in degradation and rehabilitation processes. *International Journal of Applied Earth Observation and Geoinformation* 3, 93–98.
- Shah, P.B., 1996. Soil fertility and erosion based unsustainability concerns in Nepal. In: Joshi, D. (Ed.), Workshop on Soil Fertility and Plant Nutrition Management. NARC—Soil Science Division, Nepal.
- Shah, S.G., Friend, J.A., 1992. Mixed crop systems in the Himalaya. In: Pearson, C.J. (Ed.), Ecosystems of the World: Field Crop Ecosystems. Elsevier, Amsterdam, pp. 291–310.
- Sharma, E., Rai, S.C., Sharma, R., 2001. Soil, water and nutrient conservation in mountain farming systems: case study from the Sikkim Himalaya. *Journal of Environmental Management* 61, 123–135.
- Sherr, S.J., Yadav, S., 1996. Land degradation in the developing world: implications for foods, agriculture, and the environment to 2020. Food, Agriculture, and the Environment Discussion Paper 14. International Food Policy Research Institute, USA.
- Smolikowski, B., Puig, H., Roose, E., 2001. Influence of soil protection techniques on runoff, erosion and plant production on semi-arid hillsides of Cabo Verde. *Agriculture, Ecosystem and Environment* 87, 67–80.
- Soil Management Support Services (SMSS), 1985. Keys to soil taxonomy. Soil management support services technical monograph no. 6. United States Department of Agriculture and Agency for International Development, USA.
- Srivastava, J.P., Tamboli, P.M., English, J.C., Lal, R., Stewart, B.A., 1993. Conserving soil moisture and fertility in the warm seasonally dry tropics. World Bank Technical Paper Number 221. The World Bank, Washington DC, USA.
- Thapa, G.B., 1996. Land use, land management and environment in a subsistence mountain economy in Nepal. *Agriculture, Ecosystem and Environment* 57, 57–71.
- Thapa, G.B., Paudel, G.S., 2002. Farmland degradation in the mountains of Nepal: a study of watersheds 'with' and 'without' external intervention. *Land Degradation and Development* 13, 479–493.
- Tolk, J.A., Howell, T.A., Evett, S.R., 1999. Effect of mulch, irrigation, and soil type on water use and yield of maize. *Soil Tillage Research* 50, 137–147.
- Tripathi, B.P., Rajbhandari, N.P., Ransom, J.K., 2000a. Improved soil fertility management for sustainable maize production: proceedings of a working group meeting of the Hill Maize Research Project. Nepal Agricultural Research Council/HMRP/CIMMYT, Nepal.
- Tripathi, B.P., Shrestha, S.P., Acharya, G.P., 2000b. Summary and updating with 1999 season soil and nutrient losses from bari land terraces in the western hills of Nepal. Lumle Technical Paper No 2000/3. Agricultural Research Station, Lumle, Nepal.
- Wang, X.B., Cia, D.X., Hoogmoed, W.B., Oenema, O., Perdok, U.D., 2007. Developments in conservation tillage in rainfed regions of North China. *Soil & Tillage Research* 93, 239–250.

Influence of pyrene combination state in soils on its treatment efficiency by Fenton oxidation

Hong-wen Sun*, Qi-she Yan

College of Environmental Science and Engineering, Nankai University, Tianjin 300071, China

Received 18 March 2006; received in revised form 1 March 2007; accepted 16 March 2007

Available online 22 May 2007

Abstract

Interactions of hydrophobic organic compounds (HOCs) with soil organic matter (SOM) determine their combination state in soils, and therefore strongly influence their mobility, bioavailability, and chemical reactivity. Contact time (aging) of an HOC in soil also strongly influences its combination state and environmental fate. We studied Fenton oxidation of pyrene in three different soils to reveal the influences of SOM, contact time, and combination state on the efficiency of vigorous chemical reactions. Pyrene degradation efficiency depended strongly on the dose of oxidant (H_2O_2) and catalyst (Fe^{2+}); the greatest degradation was achieved at an oxidant to catalyst molar ratio of 10:1. Pyrene degradation differed among the three soils, ranging from 65.4% to 88.9%. Pyrene degradation efficiency decreased with increasing SOM content, and the aromatic carbon content in SOM was the key parameter. We hypothesize that pyrene molecules that combine with the compact net structure of aromatic SOM are less accessible to Fenton oxidation. Furthermore, pyrene degradation efficiency decreased considerably after aged for 30 days, but further aging to 60 and 180 days did not significantly change degradation efficiency. The Fenton oxidation efficiency of pyrene in both unaged and aged soils was greater than the corresponding desorption rate during the same period, perhaps because Fenton reaction can make pyrene more accessible to the oxidant through the enhancement of HOCs' desorption by generating reductant species or by destroying SOM through oxidation.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Fenton oxidation; Aging; Combination state; Pyrene; Desorption; SOM

1. Introduction

The combination state of hydrophobic organic compounds (HOCs) in soils is an important factor controlling their fate, reactivity, and bioavailability in the environment. When an HOC enters soil, its combination state can vary greatly as a result of the high heterogeneity of the soil. Soil organic matter (SOM) is believed to be the key structure that binds HOCs (Chiou et al., 2000). Based on its solubility in acids and bases, SOM is divided into three fractions: fulvic acid (FA) is the fraction soluble in both acids and bases, humic acid (HA) is the fraction soluble in bases but not acids, and humin is the residue soluble in neither bases nor acids. HOCs combined with these different fractions of the SOM behave differently because the fine structures of the three fractions differ (Huang et al.,

2003). FA and HA have relatively flexible and amorphous structures; in contrast, humin has a relatively condensed, rigid, three-dimensional structures. Therefore, it has been proposed that HOC molecules exist in SOM as different combination states: some are combined with those "easily desorbing and available sites", some combined with "difficultly desorbing and less available sites", and the others with "irreversible and non-available sites" (Sun and Li, 2005).

Moreover, it has been reported that HOCs become increasingly unavailable with increasing contact time (aging) in soils (Alexander, 2000), which suggests that the combination state of the HOC changes further during aging, with some molecules moving from the "easily desorbing and available sites" to the "difficultly desorbing and less available points", and even to the "irreversible and non-bioavailable sites". It has been well documented (Walker et al., 2005; Conrad et al., 2002; Macleod and Semple, 2000; Kan et al., 1994) that desorption of HOCs in

*Corresponding author. Tel.: +86 22 2350 9241.

E-mail address: sunhongwen@nankai.edu.cn (H.-w. Sun).

soils, their availability to bio-uptake and biodegradation, and their extraction by mild solvents all decrease considerably during aging. However, whether or not the combination state of an HOC, which is controlled by soil properties and contact time, influences its reactivity in chemical reactions such as Fenton oxidation, has seldom been studied (Bogan and Trbovic, 2003).

Polycyclic aromatic hydrocarbons (PAHs) have attracted much attention for a long time due to their potential carcinogenic or mutagenic effects on human beings. Therefore, it is important to find an effective technique for purifying PAH-contaminated soils. However, biodegradation of PAHs is slow and incomplete, especially at high concentrations (Mueller et al., 1991). Recently, in-situ chemical oxidation (ISCO) has been shown to be an efficient and economical remediation technique for PAH-contaminated soils (Goi and Trapido, 2004; Watts et al., 2002; Kang and Hua, 2005). Fenton reaction, which uses H_2O_2 as oxidant and Fe^{2+} as catalyst to generate hydroxyl radicals ($\cdot\text{OH}$), is a promising ISCO technique due to its high efficiency, low cost, and environmental benignancy. It has been successfully applied in the remediation engineering of PAH-contaminated sites (Quan et al., 2003; Martinez et al., 2003). Moreover, several studies have reported that Fenton oxidation can enhance the biodegradability of PAHs (Nam et al., 2001; Lee and Hosomi, 2001; Kulik et al., 2005; Palmroth et al., 2006). Several studies have reported that the treatment efficiency of HOCs by ISCO is affected by soil type (Gates-Anderson et al., 2001; Goi and Trapido, 2004). Bogan and Trbovic (2003) have found that total organic carbon, humin content, and soil porosity all influenced the degradability of sorbed PAH by Fenton oxidation. The susceptibility of PAH to chemical oxidation was a function of SOM in four of the soils (those with SOM greater than approximately 10%), but was strongly dependent on soil porosity for low-SOM soils. There have also been many studies of aging effects on the bioavailability of HOCs in soils (Laor et al., 1996; Harkey et al., 1995; Conrad et al., 2002). However, the effect of aging on Fenton oxidation efficiency has seldom been discussed in literature (Bogan and Trbovic, 2003). Hence, the objectives of the present study were: (1) to determine the influence of soil type and aging on the treatment efficiency of a representative PAH (pyrene) in soil slurries using Fenton reaction; (2) to examine the effect of pyrene combination state on Fenton oxidation efficiency by comparison with its desorption. To determine the optimum conditions for Fenton reaction, the influences of the doses of the oxidant and the catalyst on reaction efficiency were observed first, then the optimal combination of the two reactants was used in subsequent experiments.

2. Materials and methods

2.1. Soil samples

Uncontaminated Chinese soils were sampled as the surface (0–20 cm) horizontal soils in Henan Province, and

in Tianjin. These soils were named Soil 1, 2, and 3, respectively, and their characteristics are shown in Table 1. The soil samples were air-dried, passed through a 2-mm sieve, and stored in closed containers in a dry place, in the dark at room temperature, prior to use. The content of SOM was determined using the potassium dichromate outside-heating method (Liu et al., 1996) after removing visible plant residues. HA is soluble in bases, but not acids; FA is soluble in both acids and bases; and humin is the insoluble residue that remains after HA and FA extractions. The content of each of the three fractions can thus be measured using a series of extractions: 0.1 mol/L diposphoric acid and 0.1 mol/L sodium hydroxide to remove HA and FA, followed by 0.5 mol/L sulfuric acid (pH = 2–3) to remove FA. pH was measured in a 1:1 (w/w) soil– CaCl_2 water suspension. The grain-size distribution was determined using a densimeter. Iron content was determined using the *o*-phenanthroline photometric method.

2.2. Chemicals

Pyrene (98% purity) was purchased from the Aldrich Chemical Company, US. Pyrene stock solution (1 g/L) was prepared in HPLC-grade methanol. Aqueous solution containing 0.01 mol/L CaCl_2 and 200 mg/L NaN_3 was adopted as the test solution in the desorption experiments in order to create a specific ionic strength and to prevent microbial activity, respectively. Ethyl acetate used as organic solvent to extract pyrene from soils; and H_2O_2 and FeSO_4 were of analytical grade. Acetonitrile used for the mobile phase in the high-performance liquid chromatography (HPLC) analysis was of HPLC grade.

2.3. Preparation of unaged and aged soil samples

To obtain pyrene-contaminated soil samples, specific volume of pyrene–methanol stock solution (1 g/L) was added to a given mass of soil in tubes to make up a final pyrene concentration of 40 mg/kg (dry soil). The tubes were then shaken using a Vortex Mixer for 1 min at 15-min intervals for 1 h in a hood to evaporate the methanol and disperse pyrene molecules in the soils. Specific amount of 200 mg/L NaN_3 aqueous solution was added to inhibit

Table 1
Selected properties of the tested soils

	Soil 1	Soil 2	Soil 3
pH	7.46	7.72	7.66
SOM (%)	1.55	1.86	5.20
Fulvic acid (%)	0.24	0.22	1.22
Humic acid (%)	0.75	0.80	1.24
Humin (%)	0.56	0.84	2.74
Sand (%)	70.10	49.95	67.79
Silt (%)	23.25	34.55	25.27
Clay (%)	6.62	15.50	6.90
Iron (%)	2.79	2.98	2.60
Texture type	Loamy sand	Sandy loam	Loamy sand

biological activity and keep soil moisture at 30% (wet weight basis). The tubes were then tightly capped with Teflon septum and kept in the dark at $25 \pm 1^\circ\text{C}$ overnight or for 30, 60, and 180 days to obtain unaged and 30-, 60-, and 180-day-aged soils, respectively.

2.4. Characterization of HA and humin

Solid-state cross-polarization magic-angle-spinning and total-side band-suppression (CPMAS-TOSS) ^{13}C NMR spectra of the extracted HA and humin were obtained using a Varian Infinity Plus 400 spectrometer operated at the ^{13}C frequency of 100.5 MHz. The instrument was run under the following conditions: contact time, 1 ms; spinning speed, 5 kHz; 90° ^1H pulse, 2.2 μs ; acquisition delay, 4 s; and from 5000 to 10,000 scans (Kang and Xing, 2005).

2.5. Treatment of the soil samples by Fenton oxidation

In the Fenton oxidation, 2 g of pyrene-contaminated soil samples were slurried in 10 mL of de-ionized water containing certain concentration (0–80 mM) of FeSO_4 . Henceforth, we will refer to this mixture as the 1:5 slurry. pH of the tested solution was adjusted to 3.0 (± 0.2) with appropriate amounts of 1 M HCl and 1 M NaOH. Specific concentrations (0–500 mM) of H_2O_2 were then added according to the design of the experiments. The slurry was stirred using a magnetic bar to ensure complete mixing of the reactants for a certain period (15–180 min). The soil samples were then recovered and analyzed to determine their pyrene content. All Fenton oxidation experiments were conducted in duplicate.

2.6. Tenax beads—assisting desorption

To compare pyrene desorption and oxidation, desorption of pyrene from unaged and 30-day-aged soils was measured using Tenax beads (35–60 mesh, Chromatographic Specialties, Canada) to assist desorption. The procedure is similar to that previously described in the literature (Sun and Li, 2005). In this approach, 1 g of the contaminated soils was suspended in round-bottom glass tubes containing 10 mL of the test solution and 0.5 g of Tenax beads. The tubes were capped with Teflon septum and shaken at 150 rpm for a designated period of time (0–10 days). At the end of each desorption period, the tubes were centrifuged for 20 min at 3500 rpm. The Tenax beads were then carefully removed using a clean stainless-steel spatula and transferred into a clear 25-mL glass flask. An additional 0.5 g of clean Tenax beads was added into the tube to continue the desorption. The contaminated Tenax was then extracted with a 1:1 (v/v) mixture of acetone and *n*-hexane in three successive extractions, with each extraction carried out in a horizontal shaker overnight. The combined acetone and *n*-hexane was evaporated to near-dryness under a gentle stream of nitrogen in a water bath held at 40°C , and the residue was re-dissolved

to certain volume using acetonitrile. The contaminant concentration in the acetonitrile was analyzed by means of HPLC. Pyrene recovery from the Tenax beads using this method was more than 98%, which was determined by adding a known quantity of pyrene onto Tenax beads. The percentage of pyrene desorbed from the soils was then calculated. Desorption experiments were conducted in duplicate.

2.7. Soxhlet extraction

Soil samples were obtained after centrifugation of the slurry. The separated solid phase was mixed with anhydrous Na_2SO_4 (1:1 w/w), allowed to stand overnight, and ground into a fine powder prior to Soxhlet extraction. For the Soxhlet extraction, the soil sample was transferred into a cellulose extraction thimble, which was then placed in a Soxhlet extraction apparatus. A 200-mL round-bottom flask containing 85 mL of ethyl acetate was connected to the apparatus, and the extraction was performed for 6 h with a reflux cycle of 5 cycles per hour (Ravelet et al., 2001). After extraction, the solvent was evaporated to near-dryness under a gentle stream of nitrogen in a water bath held at 40°C . The residue was dissolved in 10 mL of acetonitrile for HPLC analysis. Recovery of pyrene in the Soxhlet extraction of the contaminated soil samples ranged from 85% to 90%, and the oxidation percentage was calculated using the sample concentration after an average extraction recovery of 88% was accounted for.

2.8. Analytical procedure

The pyrene concentration in the solvent was quantified by means of HPLC. The HPLC system was a model SCL-10AVP HPLC (Shimadzu, Japan) equipped with a programmable fluorescence detector. A reverse-phase column (VP-ODS Kromasil C_{18} , 150 mm $\times$ 4.6 mm $\times$ 5 μm , Shimadzu, Japan) was employed at ambient temperatures. The mobile phase, containing 85% acetonitrile and 15% water (MilliQ quality), flowed at 1.0 mL/min, and the injection volume was 20 μL . Fluorescence detection was performed at an excitation wavelength of 260 nm and an emission wavelength of 373 nm. The amount of pyrene was determined using an external standard for calibration.

3. Results and discussion

3.1. Effect of H_2O_2 and FeSO_4 dosage on pyrene degradation

Fenton oxidation efficiency for HOCs varies greatly as a function of the reaction conditions, and this variation has been well-documented in aqueous phase (Martinez et al., 2003). The optimum conditions for Fenton oxidation in soil slurries appear to be similar to those in aqueous solutions (Watts et al., 2002). In the present study, the optimum dosages of the oxidant (H_2O_2) and the catalyst

(Fe^{2+}) for pyrene oxidation in soil slurries were determined at a fixed pH of 3. This pH was chosen because Fenton reaction efficiencies are the highest at a pH around 3 in both aqueous and soil systems (Quan et al., 2003). To investigate the effect of H_2O_2 dosage, batch experiments were carried out on 40 mg/kg (dry soil) pyrene in a soil slurry at different H_2O_2 concentrations (0–500 mM) with 20 mM Fe^{2+} as the catalyst and the solution temperature held at $25 \pm 1^\circ\text{C}$. The resulting levels of pyrene degradation are shown in Fig. 1. Not surprisingly, the degradation percentage increased with increasing H_2O_2 . For example, the percentages were $40.3 \pm 2.4\%$, $31.3 \pm 1.2\%$, and $15.5 \pm 1.3\%$ at 25 mM H_2O_2 in Soils 1, 2, and 3, respectively. When the H_2O_2 concentration increased to 200 mM, the degradation percentage increased to $88.9 \pm 2.7\%$, $79.4 \pm 2.1\%$, and $65.4 \pm 2.0\%$, respectively. However, when the H_2O_2 concentration continued increasing beyond this level (to 400 and 500 mM), the oxidation efficiency did not increase significantly.

To observe the effect of the Fe^{2+} concentration, pyrene degradation was analyzed at FeSO_4 concentrations ranging from 5 to 80 mM at a fixed H_2O_2 concentration of 200 mM (Fig. 2). Fig. 2 shows that pyrene degradation increased with increasing Fe^{2+} concentration for Fe^{2+} concentrations less than 20 mM, then stabilized or decreased slightly; the greatest degradation was achieved at 20 mM FeSO_4 . Hence, the optimum condition for pyrene oxidation in soil slurry appears to be 200 mM H_2O_2 plus 20 mM FeSO_4 , with a resulting oxidant to catalyst molar ratio of 10:1.

The greatest oxidation efficiency was achieved at the above-mentioned optimum oxidant to catalyst ratio, which agrees with previous results (Lunar et al., 2000). This is because Fenton oxidation is a complicated reaction system. The Fenton initiation reactions are (Walling and Goosen, 1973):

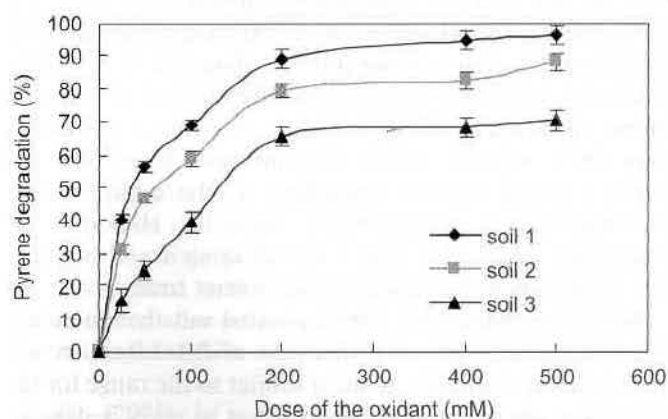
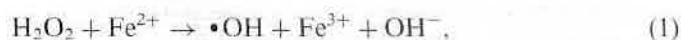


Fig. 1. Effects of H_2O_2 dose on pyrene degradation in three different soils using the Fenton reaction (30-min reaction time in a 1:5 slurry with 20 mM Fe^{2+} at pH 3).

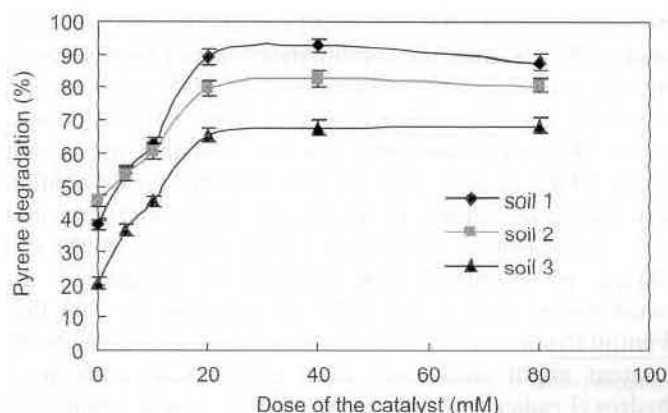


Fig. 2. Effects of Fe^{2+} dose on pyrene degradation in three different soils using Fenton reaction (30-min reaction time in a 1:5 slurry with 200 mM H_2O_2 at pH 3).

As demonstrated in Eq. (1), H_2O_2 acts as the oxidant, and soluble Fe^{2+} serves as the catalyst in Fenton oxidation. Attack on HOCs by the $\bullet\text{OH}$ radicals generated through the reaction in Eq. (1) is believed to be responsible for destruction of the HOCs (Watts et al., 2002). In order to continually generate $\bullet\text{OH}$, the Fe^{3+} produced in Eq. (1) must be reduced to Fe^{2+} via additional steps (Eq. (2)). Additionally, it has been pointed out that the generation of $\bullet\text{OH}$ and the subsequent free-radical attack on the target HOCs will proceed when soluble Fe^{2+} species are present. Hence, if the total soluble Fe^{2+} concentration decreases, the catalytic cycles of soluble Fe^{2+} cease and the rate of Fenton oxidation decreases (Bogan and Trbovic, 2003). This is why, at certain concentrations, pyrene degradation increased with increasing dosages of both H_2O_2 and Fe^{2+} . The removal rate of pyrene was limited by the availability of soluble Fe^{2+} when H_2O_2 was present in excess (Fig. 1), thereby consuming too much Fe^{2+} and consequently creating a low level of Fe^{2+} in the reaction system. On the other hand, when Fe^{2+} concentration was too high (>20 mM in this study; Fig. 2), a considerable quantity of H_2O_2 may have been consumed in vain (i.e., not used to generate $\bullet\text{OH}$ but instead used to generate water via other reactions). Hence, an optimum oxidant to catalyst ratio exists for Fenton oxidation of HOCs in soil systems, and the mechanisms are the same as in aqueous systems.

3.2. Effect of SOM on pyrene oxidation efficiency

The time-dependent oxidation efficiencies of pyrene in the three soils with different SOM content are shown in Fig. 3. Oxidation took place quickly, and the majority of the reaction finished within 30 min. The percentage of pyrene degradation differed noticeably among the three soils, with values of $88.9 \pm 2.7\%$, $79.4 \pm 2.1\%$, and $65.4 \pm 2.0\%$ for unaged Soils 1, 2, and 3, respectively. Lower degradation efficiency was found for soils with a higher SOM content. It has been well-documented that SOM plays an important role in determining the mobility and

bioavailability of HOCs in soils (Sun and Li, 2005; Walker et al., 2005) because the combination state of HOCs bound in soils varied greatly with respect to SOM content and structure. Usually, soils with higher SOM content possess more “difficultly desorbing and less available sites”, and hence HOCs in these soils are less available for desorption and biological uptake. In the present study, we found that even in the face of a vigorous Fenton reaction, SOM still played an important role through its impact on the combination state of the HOC. In addition, the fact that Fenton oxidation efficiency decreased with increasing SOM content might also result from greater consumption of hydroxyl radicals at high SOM content, which would lead to a lower rate of pyrene degradation.

SOM is a highly heterogeneous system of macromolecules. From the view of solubility in acids and bases, SOM can be divided into three fractions: FA (soluble in acids), HA (soluble in bases), and humin (insoluble in either acids or bases). The three SOM fractions have different effects on the mobility and bioavailability of the HOC that combines with them due to differences in their fine structure. In order to identify the key factor that affects the efficiency of Fenton oxidation, the percentages of the three SOM fractions were measured (Table 1), and the fine structures of HA and humin in the three soils were observed by CPMAS-TOSS ^{13}C NMR (Table 2). The resulting NMR spectrum revealed that humin contains more aromatic carbons (108–160 ppm) than HA does, and that their aliphaticity is lower than that of HA; in contrast, HA contains a higher content of amide or carboxylic acid groups (160–190 ppm) than humin does. These groups are polar, thus sorption of HOCs on these groups is reported to be weak (Kang and Xing, 2005). We analyzed the correlation between pyrene degradation in the three soils and their SOM fractions (Table 3). The pyrene degradation percentage was negatively correlated with all the parameters analyzed, but the strongest correlation ($R^2 = -0.955$) was for the percentage of aromatic carbon in both HA and humin, which equals the sum of the

percentage of aromatic carbon in each fraction (Table 2 multiplied by the corresponding HA and humin percentages (Table 1). From these results, the aromatic carbon component of SOM appears to be the key factor affecting pyrene degradation, and HOCs that combine with aromatic carbons become less available to Fenton oxidation. This conclusion should be verified using additional soil samples. Bogan and Trbovic (2003) also reported that humin in SOM could reduce the susceptibility of PAHs to Fenton oxidation.

3.3. Effect of contact time on pyrene oxidation efficiency

The effect of contact time (aging) on Fenton oxidation is shown in Fig. 4. The results reveal that the rate of pyrene degradation had decreased considerably after 30 days of aging, with the pyrene degradation percentage decreasing to $74.7 \pm 3.1\%$, $63.0 \pm 2.0\%$, and $42.3 \pm 2.5\%$, for Soils 1, 2, and 3, respectively. However, the oxidation efficiencies in samples aged for 60 and 180 days did not differ significantly from those of samples aged for 30 days. The aging period may have allowed some of the pyrene that had originally combined with the “easily accessible sites” to transfer to the “difficultly accessible sites”, where the availability of the molecule to the oxidation reaction is lower. This process is called sequestration, and it is believed that sequestration is a natural process in soils that reduces the availability of HOCs to biodegradation and other reactions (Alexander, 2000).

The mechanism of sequestration has been proposed to result from slow diffusion of HOCs within the nanopores of organic–mineral complexes or the compact structure of SOM (Pignatello and Xing, 1996). Many authors have concluded that interactions between SOM and HOCs occurs as a two-step process, in which sequestration begins with partitioning of the HOC molecules into the HA and FA fractions (which have amorphous structures), followed by diffusion into the humin–mineral fraction, which has a rigid net structure (Huang et al., 2003; Pignatello and Xing, 1996). The release of HOC molecules that have combined with the HA and FA fractions is believed to be relatively easy, whereas the release of those that have combined with the humin fraction is more difficult (Sun and Yan, 2007). Nam and Kim (2002) studied the distribution of phenanthrene in different SOM fractions, and found that more than 90% of the initial amount was found in the humin–mineral fraction regardless of the aging period. The recovery of phenanthrene from the HA and FA fractions is about 11% after 1 day of aging in soil samples, and decreased with the increasing contact time.

Moreover, Alexander (2000) pointed out that there are many pores or voids with diameter of 0.3–1.0 nm in soil aggregates, and this size range is similar to the range for the molecular size of HOCs. Malekani et al. (1997) demonstrated that the high porosity in humin–mineral complexes mainly comes from humin, not from the mineral surfaces. The molecular volume of pyrene is 0.186 nm^3 (Brion and

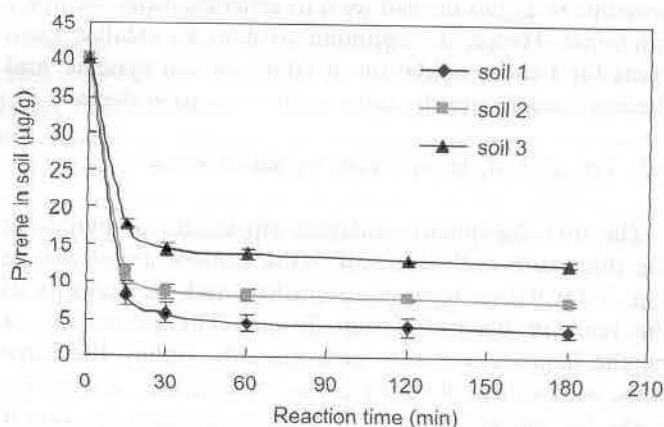


Fig. 3. Pyrene degradation dynamics in different soils under optimal conditions for Fenton oxidation (in a 1:5 slurry with 200 mM H_2O_2 plus 20 mM FeSO_4 at pH 3).

Table 2
Integration of solid-state ^{13}C NMR spectra and aliphaticity (the ratio of the sum of aliphatic to aromatic carbons) for three HA and three humin samples^a

Sample	Distribution (%) of chemical shift (ppm)				Aliphatic C (%)	Aromatic C (%)	Aliphaticity
	0–60	60–108	108–160	160–190			
HA 1	56.8	11.8	10.1	21.3	68.6	10.1	6.79
Humin 1	16.6	47.3	36.1	—	63.9	36.1	1.77
HA 2	44.4	16.8	19.9	18.9	61.2	19.9	3.08
Humin 2	7.0	55.9	37.0	—	62.9	37.0	1.7
HA 3	59.3	8.1	17.8	14.8	67.4	17.8	3.79
Humin 3	4.6	60.1	35.3	—	64.7	35.3	1.83

^aAliphatic C represents the total aliphatic carbon region (0–108 ppm); aromatic C represents the total aromatic carbon region (108–160 ppm); aliphaticity represents the ratio aliphatic C (0–108 ppm)/aromatic C (108–160 ppm).

Table 3
Correlation coefficients (R^2) for the relationship between pyrene degradation in the three soils and their SOM parameters

Total SOM	HA	Humin	HA + humin	Aromatic carbon in (HA + humin) (%)
–0.890	–0.900	–0.915	–0.912	–0.955

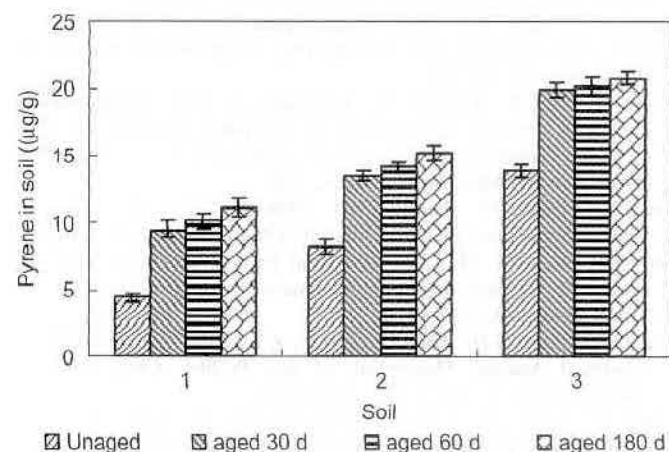


Fig. 4. Effect of contact time (aging) on the efficiency of pyrene degradation by Fenton oxidation in different soils (30-min reaction time in a 1:5 slurry with 200 mM H_2O_2 and 20 mM Fe^{2+}).

Pelletier, 2005), which is similar to the size of the soil pores. Hence, it is plausible to hypothesize that molecules trapped in micropores or nanopores are less available than those found in relatively larger macropores (i.e., pores that are larger compared with the size of the pollutant molecule). Hence, the smaller pores in SOM and in soil minerals can serve as sequestration sites for HOCs.

3.4. Comparison of pyrene desorption and degradation

Desorption dynamics were assessed for six pyrene-contaminated soils (three soils, with unaged and 30-days-

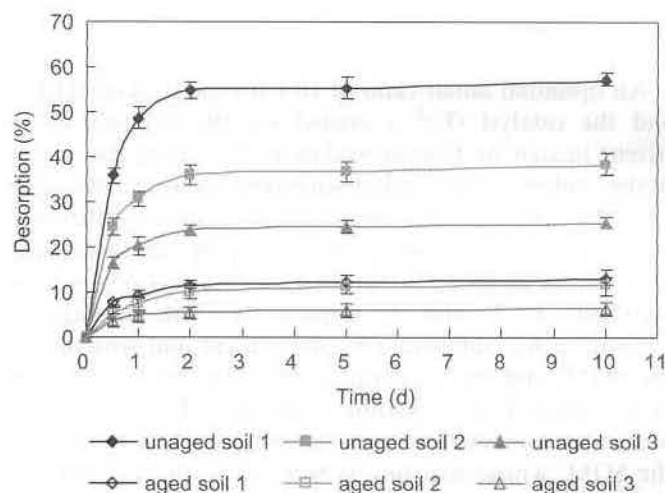


Fig. 5. Desorption percentages of sorbed pyrene in 30 days-aged and unaged samples of the three soils using desorption assisted by Tenax beads.

aged samples for each; Fig. 5). Pyrene desorbed quickly within the first day, but after 2 days, desorption seemed to cease, and a steady-state was achieved in all the six samples. These results demonstrate that pyrene quickly became adsorbed to the soils and subsequently resisted desorption, with desorption percentages of only 54.7%, 35.9%, and 23.7% for unaged Soils 1, 2, and 3, respectively. The desorption percentages of the aged soils were much lower, with the corresponding values of 13.12%, 11.34%, and 6.25%. On the other hand, pyrene degradation by Fenton oxidation was completed in 0.5 h, resulting in about 65–89% degradation of the parent compound. This seems to indicate that Fenton reaction can oxidize pyrene sorbed to soil particles. However, there may be other explanations. First, SOM content was reduced during the Fenton reaction due to the oxidation, and this would release many of the sorbed pyrene molecules. In our previous study, we found that Fenton oxidation reduced the SOM content of the same three soils from 1.55% to 1.28%, from 1.82% to 1.49%, and from 5.20% to 4.10% for Soils 1, 2, and 3, respectively (Sun and Yan, 2007). Moreover, the Fenton reaction can generate both oxidant

and reductant species, and it has been proposed that the reductant species have the potential to enhance desorption of sorbed HOCs (Watts et al., 1999, 2002). Therefore, a vigorous Fenton reaction may provide a complex reaction matrix that can change the content and structure of the SOM, and as a result, the combination states of the contaminant. Although the Fenton oxidation percentage for pyrene is higher than its desorption percentage, oxidation still depends strongly on the pyrene's combination state in the soil. As was discussed above, oxidation efficiency differed greatly among the three soils due to their different SOM contents and structures. Furthermore, oxidation decreased further after aging, presumably as a result of sequestration at less accessible sites.

4. Conclusions

An optimum molar ratio of 10:1 for the oxidant (H_2O_2) and the catalyst (Fe^{2+}) existed for the degradation of pyrene in soils by Fenton oxidation. Increasing the dosage of the oxidant or the catalyst separately, causing a deviation from this optimum ratio, decreased the oxidation efficiency. Pyrene degradation varied greatly among the three soils, with values of $88.9 \pm 2.7\%$, $79.4 \pm 2.1\%$, and $65.4 \pm 2.0\%$ for Soils 1, 2, and 3, respectively. The degradation percentage was influenced by the content and structure of the SOM, and the aromatic carbon content was the key factor controlling oxidation efficiency. That is, pyrene molecules combined with the aromatic carbon fraction of the SOM, whose structure is believed to be compact and rigid, were much less accessible to Fenton oxidation. Pyrene degradation had decreased after 30 days of aging, reaching values of $74.7 \pm 3.1\%$, $63.0 \pm 2.0\%$, and $42.3 \pm 2.5\%$ for Soils 1, 2, and 3, respectively. This is because during aging, more pyrene molecules became trapped in less accessible sites. Fenton oxidation was largely complete by 30 min, whereas desorption only occurred at rates of 54.7%, 35.9%, and 23.7% for unaged Soils 1, 2, and 3, and these values are considerably lower than the corresponding degradation percentage. Though Fenton oxidation is strongly influenced by the combination state of the pollutant molecules in soils, the reaction can change the combination state of the HOCs by reducing SOM content and enhancing desorption of the HOCs by reductant species.

Acknowledgments

The authors thank the Ministry of Science and Technology of China (2004cb-418504) and the National Foundation of Natural Science (40273036) of China for their financial support.

References

- Alexander, M., 2000. Aging, bioavailability, and overestimation of risk from environmental pollutants. *Environmental Science and Technology* 34, 4259–4264.
- Bogan, B.W., Trbovic, V., 2003. Effect of sequestration on PAH degradability with Fenton's reagent: roles of total organic carbon, humin, and soil porosity. *Journal of Hazardous Material B* 100, 285–300.
- Brion, D., Pelletier, E., 2005. Modelling PAHs adsorption and sequestration in freshwater and marine sediments. *Chemosphere* 61, 867–876.
- Chiou, C.T., Kile, D.E., Ruthereord, D.W., 2000. Sorption of selected organic compounds from water to a peat soil and its humic-acid and humin fractions: potential sources of the sorption nonlinearity. *Environmental Science and Technology* 34, 1254–1258.
- Conrad, A.U., Comber, S.D., Simkiss, K., 2002. Pyrene bioavailability: effect of sediment–chemical contact time on routes of uptake in an oligochaete worm. *Chemosphere* 49, 447–454.
- Gates-Anderson, D.D., Siegrist, R.L., Cline, S.R., 2001. Comparison of potassium permanganate and hydrogen peroxide as chemical oxidants for organically contaminated soils. *Journal Environmental Engineering* 127, 337–347.
- Goi, A., Trapido, M., 2004. Degradation of polycyclic aromatic hydrocarbons in soil: the Fenton reagent versus ozonation. *Environmental Technology* 25, 155–164.
- Harkey, G.A., Van Hoof, P.L., Landrum, P.F., 1995. Bioavailability of polycyclic aromatic hydrocarbons from a historically contaminated sediment core. *Environmental Science and Technology* 14, 1551–1560.
- Huang, W.L., Peng, P.A., Yua, Z.Q., Fu, J.M., 2003. Effects of organic matter heterogeneity on sorption and desorption of organic contaminants by soils and sediments. *Review Applied Geochemistry* 18, 955–972.
- Kan, A.T., Fu, G., Tomson, M.B., 1994. Adsorption/desorption hysteresis in organic pollutants and soil/sediment interaction. *Environmental Science and Technology* 28, 859–867.
- Kang, N., Hua, L., 2005. Enhanced chemical oxidation of aromatic hydrocarbons in soil systems. *Chemosphere* 61, 909–922.
- Kang, S.H., Xing, B.S., 2005. Phenanthrene sorption to sequentially extracted soil humic acids and humins. *Environmental Science and Technology* 39, 134–140.
- Kulik, N., Goi, A., Trapido, M., Tuhkanen, T., 2005. Degradation of polycyclic aromatic hydrocarbons by combined chemical pre-oxidation and bioremediation in creosote contaminated soil. *Journal of Environmental Management* 78, 382–391.
- Laor, Y., Strom, P.F., Farmer, W.J., 1996. The effect of sorption on phenanthrene bioavailability. *Journal of Biotechnology* 51, 227–234.
- Lee, B.D., Hosomi, M., 2001. A hybrid Fenton oxidation—microbial treatment for soil highly contaminated with benz(a)anthracene. *Chemosphere* 43, 1127–1132.
- Liu, G.S., Jiang, N.H., Zhan, L.D., Liu, Z.L., 1996. *Soil Physical and Chemical Analysis Description of Soil Profiles*. China Standard Methods Press, Beijing, China (in Chinese).
- Lunar, L., Sicilia, D., Rubio, S., Pérez-Bendito, D., Nickel, U., 2000. Degradation of photographic developers by Fenton's reagent: condition optimization and kinetics for metal oxidation. *Water Research* 34, 1791–1802.
- Macleod, C.J.A., Semple, K.T., 2000. Influence of contact time on extractability and degradation of pyrene in soils. *Environmental Science and Technology* 34, 4952–4957.
- Malekani, K., Rice, J.A., Lin, J.S., 1997. The effect of sequential removal of organic matter on the surface morphology of humin. *Soil Science* 162, 333–342.
- Martínez, N.S.S., Fernández, J.F., Segura, X.F., Ferrer, A.S., 2003. Pre-oxidation of an extremely polluted industrial wastewater by the Fenton's reagent. *Journal of Hazardous Material B* 101, 315–322.
- Mueller, J.G., Middaugh, D.P., Lantz, S.E., Chapman, P.J., 1991. Biodegradation of creosote and pentachlorophenol in contaminated groundwater: chemical and biological assessment. *Applied Environmental Microbiology* 57, 1277–1285.
- Nam, K., Kim, J.Y., 2002. Role of loosely bound humic substances and humin in the bioavailability of phenanthrene aged in soil. *Environmental Pollution* 118, 427–433.
- Nam, K., Rodríguez, W., Kukor, J.J., 2001. Enhanced degradation of polycyclic aromatic hydrocarbons by biodegradation combined with a modified Fenton reaction. *Chemosphere* 45, 11–20.

- Palmroth, M.R.T., Langwaldt, J.H., Aunola, T.A., Goi, A., Puhakka, J.A., Tuhkanen, T.A., 2006. Treatment of PAH contaminated soil by combination of Fenton's reaction and biodegradation. *Journal of Chemical Technology and Biotechnology* 81, 598–607.
- Pignatello, J.J., Xing, B., 1996. Mechanisms of slow sorption of organic chemicals to natural particles. *Environmental Science and Technology* 30, 1–11.
- Quan, H.N., Teel, A.L., Watts, R.J., 2003. Effect of contaminant hydrophobicity on hydrogen peroxide dosage requirements in the Fenton-like treatment of soils. *Journal of Hazardous Material B* 102, 277–289.
- Ravelet, C., Grosset, C., Montuelle, B., Benoit-Guyod, J.L., Alary, J., 2001. Liquid chromatography study of pyrene degradation by two micromycetes in freshwater sediment. *Chemosphere* 44, 1541–1546.
- Sun, H., Li, J., 2005. Availability of pyrene in unaged and aged soils to earthworm uptake, butanol extraction and SFE. *Water, Air and Soil Pollution* 166, 353–365.
- Sun, H., Yan, Q., 2007. Influence of Fenton oxidation on soil organic matter and its sorption and desorption of pyrene. *Journal of Hazardous Material*, in press (available on line, DOI:10.1016/j.jhazmat.2006.10.005).
- Walker, A., Rodriguez-Cruz, M.S., Mitchell, M.J., 2005. Influence of ageing of residues on the availability of herbicides for leaching. *Environmental Pollution* 133, 43–51.
- Walling, C., Goosen, A., 1973. Mechanism of the ferric ion catalyzed decomposition of hydrogen peroxide. Effect of organic substrates. *Journal of American Chemical Society* 95, 2987–2991.
- Watts, R.J., Bottenberg, B.C., Hess, T.F., Jensen, M.D., Teel, A.L., 1999. Role of reductants in the enhanced desorption and transformation of chloraliphatic compounds by modified Fenton's reactions. *Environmental Science and Technology* 33, 3432–3437.
- Watts, R.J., Stanton, P.C., Howsawheng, J., Teel, A.L., 2002. Mineralization of a sorbed polycyclic aromatic hydrocarbon in two soils using catalyzed hydrogen peroxide. *Water Research* 36, 4283–4292.

Environmental and economic feasibility study of a total wastewater treatment network system

Seong-Rin Lim, Donghee Park, Jong Moon Park*

Advanced Environmental Biotechnology Research Center, Department of Chemical Engineering, School of Environmental Science and Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Pohang 790-784, South Korea

Received 20 October 2006; received in revised form 17 January 2007; accepted 17 March 2007

Available online 4 May 2007

Abstract

The synthesis of distributed wastewater treatment plants (WTPs) has been studied to reduce capital and operating costs associated with wastewater treatment. In this study, the environmental and economic feasibility of a total wastewater treatment network system (TWTNS) including distributed and terminal WTPs was estimated using life cycle assessment (LCA) and life cycle costing (LCC) methods. Wastewater sources and existing distributed and terminal WTPs in an iron and steel plant were networked. The TWTNS was generated from the optimal solution to a mathematical optimization model and compared to a conventional wastewater treatment system (CWTS). The environmental effect scores of the TWTNS were from 29.6% to 68.3% higher than those of the CWTS because of higher electricity consumption required to pump wastewater to the networked WTPs. However, the life cycle cost of the TWTNS was lower than that of the CWTS by 10.1% because of the decrease of the labor cost resulting from the closing of three distributed WTPs. Overall, the TWTNS was no more eco-efficient than the CWTS because the increase of environmental burdens outweighed the decrease of economic costs.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Environmental and economic feasibility; Life cycle assessment; Life cycle costing; Synthesis; Wastewater treatment.

1. Introduction

Wastewater treatment strategies required to effectively meet the discharge permits of environmental regulations have been important concerns in many industries. The synthesis of distributed wastewater treatment plants (WTPs) has been regarded as a way to reduce initial capital investment and operating costs in wastewater treatment (Wang and Smith, 1994). This strategy focused on minimizing the flowrate of wastewater treated in distributed WTPs and maximizing the flowrate of wastewater bypassing distributed WTPs, with the quality of the final effluent mixed with the bypassed and treated wastewater streams not exceeding the discharge limits of environmental regulations (Wang and Smith, 1994; Kuo and Smith, 1997; Galan and Grossmann, 1998; Huang et al., 1999; Mann and Liu, 1999; Hernandez-Suarez et al.,

2004). However, the environmental and economic feasibility of the distributed wastewater treatment network systems has not been estimated, even though there have been increasing concerns for sustainable development and cleaner production with the aim of conserving natural resources and achieving economic growth.

Heuristically graphical approaches to the optimal design of distributed wastewater treatment network systems have been proposed as an extension of water pinch technology (Wang and Smith, 1994; Kuo and Smith, 1997). A minimum wastewater treatment flowrate is obtained with a concentration–composite curve, and the heuristic design methodology is used to generate and merge the independent wastewater treatment subnetworks used to remove each contaminant in wastewater streams.

Mathematical optimization related to superstructure models has also been applied to synthesize distributed WTPs. Galan and Grossmann (1998) studied a search procedure used to yield a global or near global optimum solution on the basis of the successive solution of a relaxed

*Corresponding author. Tel.: +82 54 279 2275; fax: +82 54 279 8299.

E-mail address: jmpark@postech.ac.kr (J.M. Park).

Nomenclature

Sets

- C = { c | c is a contaminant in wastewater}, $c = 1, 2, \dots, Nc$
- WW = { ww | ww is a wastewater source}, $ww = 1, 2, \dots, Nm$
- TP = { tp | tp is a distributed WTP}, $tp = 1, 2, \dots, Nn$
 = { tp_{in} | tp_{in} is a distributed WTP}, $tp_{in} = 1, 2, \dots, Nn$
 = { tp_{out} | tp_{out} is a distributed WTP}, $tp_{out} = 1, 2, \dots, Nn$
- TTP = { ttp | ttp is a terminal WTP}, $ttp = 1, 2, \dots, Nk$
 = { ttp_{in} | ttp_{in} is a terminal WTP}, $ttp_{in} = 1, 2, \dots, Nk$
 = { ttp_{out} | ttp_{out} is a terminal WTP}, $ttp_{out} = 1, 2, \dots, Nk$

Variables

- $C_{c,tp_{in}}$ concentration at the inlet of a distributed WTP
- $C_{c,tp_{out}}$ concentration at the outlet of a distributed WTP
- $C_{c,ttp_{in}}$ concentration at the inlet of a terminal WTP
- $C_{c,ttp_{out}}$ concentration at the outlet of a terminal WTP
- $F_{tp_{in}}$ flowrate at the inlet of a distributed WTP
- $F_{ww,tp_{in}}$ flowrate from a wastewater source to the inlet of a terminal WTP
- $F_{tp_{out}}$ flowrate at the outlet of a distributed WTP

- $F_{tp_{out},tp_{in}}$ flowrate from the outlet of a distributed WTP to the inlet of a distributed WTP
- $F_{tp_{out},ttp_{in}}$ flowrate from the outlet of a distributed WTP to the inlet of a terminal WTP
- $F_{ttp_{in}}$ flowrate at the inlet of a terminal WTP
- $F_{ttp_{out},dis}$ flowrate at the outlet of a terminal WTP
- $F_{ww,tp_{in}}$ flowrate from a wastewater source to the inlet of a distributed WTP
- $F'_{tp_{in}}$ total flowrate of wastewater treated in the distributed WTPs
- PV present value
- P payment

Parameters

- $C_{c,ttp_{out}}^{max}$ maximum concentration at the outlet of a terminal WTP
- $C_{c,ww}$ concentration of a wastewater source
- e escalation rate
- $F_{tp_{in}}^{max}$ maximum flowrate at the inlet of a distributed WTP
- $F_{ttp_{in}}^{max}$ maximum flowrate at the inlet of a terminal WTP
- F_{ww} flowrate of a wastewater source
- i interest rate
- $L_{tp_{in}}^{max}$ maximum contaminant load of a distributed WTP
- $L_{ttp_{in}}^{max}$ maximum contaminant load of a terminal WTP
- $R_{c,tp}$ removal efficiency of a distributed WTP
- $R_{c,ttp}$ removal efficiency of a terminal WTP
- t time

linear model and an original nonconvex nonlinear problem. Huang et al. (1999) proposed a mathematical model to determine the optimal water use and treatment network with the lowest freshwater consumption rate and minimum wastewater treatment capacity. Hernandez-Suarez et al. (2004) developed a superstructure decomposition and parametric optimization approach to obtain a globally optimal network system. However, terminal WTPs have been excluded in the synthesis of the distributed WTPs.

In this study, a terminal WTP was included in a superstructure model used for the synthesis of existing WTPs to reflect real wastewater treatment situations. This was because the direct discharge of untreated and diluted wastewater into sea or surface water, which has been employed for the distributed wastewater treatment network systems in the previous studies, is prohibited by environmental regulations in South Korea, even though the quality of a final effluent meets discharge permits. Therefore, a total wastewater treatment network system (TWTNS) consisting of distributed and terminal WTPs was employed in this study.

The concept of an eco-design has become more important in many industries for sustainable development. Eco-design is performed to enhance the environmental and economic performances of systems, processes, products and services. Life cycle assessment (LCA) and life cycle costing (LCC) methods have been used to evaluate the environmental and economic performance during the life cycle (Friedrich, 2002; Gluch and Baumann, 2004; Carpentieri et al., 2005; Spitzley et al., 2005).

The environmental and economic feasibility of the TWTNS was estimated to determine whether the TWTNS was in accordance with sustainable development and cleaner production. Existing distributed and terminal WTPs in an iron and steel industry were used to synthesize the TWTNS in this study. The TWTNS was generated from the optimal solution to a mathematical optimization model. A conventional wastewater treatment system (CWTS) was used as a control for the estimation of the environmental and economic feasibility of the TWTNS. The CWTS did not include streams required to bypass distributed WTPs and to dilute the contaminant

concentration of wastewater. The two wastewater treatment systems were specifically designed. LCA and LCC were performed to evaluate the environmental and economic performances of the CWTS and TWTNS on the basis of their design results.

2. Methods

Wastewater sources and the existing distributed and terminal WTPs in the iron and steel plant were used to generate the TWTNS. Operational data and field information, such as the flowrates and concentrations of the wastewater sources, and the removal efficiency and maximum capacity of the WTPs were examined for the synthesis. Characteristics of the seven wastewater sources are described in Table 1, and those of the seven distributed and one terminal WTPs are shown in Table 2. The wastewater sources were from steel making, iron making, continuous casting, cold forming, and electroplating processes. The seven distributed WTPs were used to treat the seven wastewater sources one-to-one. The terminal WTP was used to polish the effluents of the seven distributed WTPs, in order to consistently meet discharge limits. All the WTPs consisted of physical and chemical treatment processes, such as coagulation, flocculation and sedimentation. A distance matrix for interconnections between the wastewater sources and sinks is shown in Table 3. The distance included the length of pipes derived from the elevations of distributed and terminal WTPs. The discharge quality of the terminal WTP is set at 20 mg/L for

COD_{cr}, 5 mg/L for SS, and 8 mg/L for F⁻ to consistently meet the discharge limits of environmental regulations.

2.1. Wastewater treatment network synthesis

Fig. 1 shows the superstructure model used to generate the TWTNS. This model included all possible interconnections: from the wastewater sources to the inlets of the distributed and terminal WTPs and from the outlets of the distributed WTPs to their inlets and to the inlets of the terminal WTPs. Effluents from the terminal WTPs were discharged into sea or river. It was assumed that mixers combined all possible streams into one stream and that splitters divided a given stream into all possible streams. It was also assumed that there was no loss of wastewater in the distributed and terminal WTPs. This was because the difference between the losses in the TWTNS and CWTS is negligible, even though the flowrates of their effluents generally decrease.

The objective function used to optimize the TWTNS was to minimize the total flowrate of wastewater treated in the distributed WTPs, to reduce the number of the distributed WTPs used and fully utilize the superfluous capacity of the distributed and terminal WTPs. The mathematical formulation of the superstructure model is as follows:

For the objective function used for the TWTNS,

$$\text{Minimize } F_{\text{tp in}}^t = \sum_{\text{tp in} \in \text{TP}} F_{\text{tp in}} \quad (1)$$

Subject to the followings:

For the overall mass balance of the TWTNS,

$$\sum_{\text{ww} \in \text{WW}} F_{\text{ww}} - \sum_{\text{tp out} \in \text{TTP}} F_{\text{tp out,dis}} = 0. \quad (2)$$

For the mass balances of the wastewater source splitters,

$$F_{\text{ww}} - \sum_{\text{tp in} \in \text{TP}} F_{\text{ww,tp in}} - \sum_{\text{tp in} \in \text{TTP}} F_{\text{ww,tp in}} = 0 \quad (3)$$

For the mass balances of the distributed WTP inlet mixers,

$$\sum_{\text{ww} \in \text{WW}} F_{\text{ww,tp in}} + \sum_{\text{tp out} \in \text{TP}} F_{\text{tp out,tp in}} - F_{\text{tp in}} = 0, \quad (4)$$

$$\sum_{\text{ww} \in \text{WW}} F_{\text{ww,tp in}} C_{\text{c,ww}} + \sum_{\text{tp out} \in \text{TP}} F_{\text{tp out,tp in}} C_{\text{c,tp out}} - F_{\text{tp in}} C_{\text{c,tp in}} = 0. \quad (5)$$

For the mass balances of the distributed WTPs,

$$F_{\text{tp in}} - F_{\text{tp out}} = 0, \quad (6)$$

$$F_{\text{tp in}} C_{\text{c,tp in}} (1 - R_{\text{c,tp}}) - F_{\text{tp out}} C_{\text{c,tp out}} = 0. \quad (7)$$

For the mass balances of the distributed WTP outlet splitters,

$$F_{\text{tp out}} - \sum_{\text{tp in} \in \text{TP}} F_{\text{tp out,tp in}} - \sum_{\text{tp in} \in \text{TTP}} F_{\text{tp out,tp in}} = 0. \quad (8)$$

Table 1

Wastewater stream data used to synthesize the distributed and terminal WTPs

	Flowrate (m ³ /h)	Contaminant	Concentration (mg/L)
WW1	56.8	COD _{cr}	75
		SS	67
		F ⁻	109
WW2	7.0	COD _{cr}	48
		SS	10
		F ⁻	115
WW3	20.9	COD _{cr}	54
		SS	24
		F ⁻	10
WW4	2.6	COD _{cr}	63
		SS	2
		F ⁻	5
WW5	10.5	COD _{cr}	62
		SS	10
		F ⁻	89
WW6	187.5	COD _{cr}	74
		SS	7
		F ⁻	10
WW7	179.2	COD _{cr}	32
		SS	15
		F ⁻	2

WW: wastewater, COD_{cr}: chemical oxygen demand by dichromate, SS: suspended solid, F⁻: fluoride ion.

Table 2
Operational and design data of the distributed and terminal WTPs

	$F_{\text{tp in}}^{\text{max}}, F_{\text{tp in}}^{\text{max}}$ (m^3/h)	Contaminant	Removal ratio (%)	$L_{\text{tp in}}^{\text{max}}, L_{\text{tp in}}^{\text{max}}$ (kg/h)
TP1	65	COD _{cr}	72	5.5
		SS	90	4.8
		F ⁻	95	7.8
TP2	10	COD _{cr}	72	0.5
		SS	90	0.3
		F ⁻	95	1.0
TP3	30	COD _{cr}	63	1.4
		SS	88	1.6
		F ⁻	10	0.3
TP4	5	COD _{cr}	73	0.2
		SS	85	0.1
		F ⁻	16	0.1
TP5	12	COD _{cr}	75	0.8
		SS	88	0.5
		F ⁻	92	1.3
TP6	210	COD _{cr}	85	17.5
		SS	86	10.7
		F ⁻	12	2.2
TP7	205	COD _{cr}	80	7.3
		SS	82	6.2
		F ⁻	17	0.5
TTP	540	COD _{cr}	50	8.3
		SS	93	18.2
		F ⁻	20	3.1

TP: distributed WTP, TTP: terminal WTP.

Table 3
Distance matrix

	TP1	TP2	TP3	TP4	TP5	TP6	TP7
TP2	410						
TP3	1710	1840					
TP4	680	820	1090				
TP5	550	680	2300	1470			
TP6	4310	4440	5150	4440	4170		
TP7	2140	2280	3390	2550	1870	2300	
TTP	3470	3600	1900	2980	4060	7310	5010

TP: distributed WTP, TTP: terminal WTP, unit: meter.

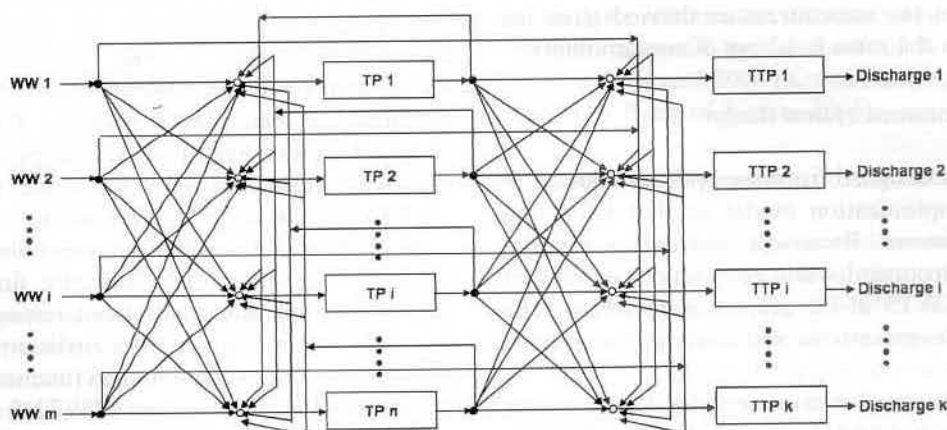


Fig. 1. Generalized superstructure model used to generate a total wastewater treatment network system (TWTNS), (WW: wastewater, TP: distributed WTP, TTP: terminal WTP).

For the mass balances of the terminal WTP inlet mixers,

$$\sum_{tp \in TP} F_{tp, out, ttp, in} + \sum_{ww \in WW} F_{ww, ttp, in} - F_{ttp, in} = 0, \quad (9)$$

$$\sum_{tp \in TP} F_{tp, out, ttp, in} C_{c, tp, out} + \sum_{ww \in WW} F_{ww, ttp, in} C_{c, ww} - F_{ttp, in} C_{c, ttp, in} = 0. \quad (10)$$

For the mass balances of the terminal WTPs,

$$F_{ttp, in} - F_{ttp, out, dis} = 0, \quad (11)$$

$$F_{ttp, in} C_{c, ttp, in} (1 - R_{c, ttp}) - F_{ttp, out, dis} C_{c, ttp, out} = 0. \quad (12)$$

For the constraints of flowrates and loads on the distributed and terminal WTPs,

$$F_{tp, in} - F_{tp, in}^{max} \leq 0, \quad (13)$$

$$F_{tp, in} C_{c, tp, in} - L_{tp, in}^{max} \leq 0, \quad (14)$$

$$F_{ttp, in} - F_{ttp, in}^{max} \leq 0, \quad (15)$$

$$F_{ttp, in} C_{c, ttp, in} - L_{ttp, in}^{max} \leq 0. \quad (16)$$

For the constraint of concentrations on the discharge quality of the terminal WTP required to consistently meet the discharge limits of environmental regulations,

$$C_{c, ttp, out} - C_{c, ttp, out}^{max} \leq 0. \quad (17)$$

The TWTNS was generated from the optimal solution to the above nonlinear programming (NLP) model. GAMS/MINOS (GAMS Development Corporation, 2005) was used as an NLP solver to obtain the optimum solution to the mathematical optimization model. Linear programming (LP) models generated by fixing the flowrates or concentrations in the NLP model were used to determine the initial points of the NLP model. However, it should be noted that even local solutions are useful in industrial applications if they can achieve a significant reduction of the total flowrate of wastewater treated in distributed WTPs. This is because global optima cannot easily be obtained because of the nonconvexities derived from the bilinear variables in the mass balances of contaminants.

2.2. Wastewater treatment system design

The TWTNS was designed from the optimal solution to the mathematical optimization model so that LCA and LCC could be performed. Because a control was required to estimate the environmental and economic feasibility of the TWTNS, the CWTS in the plant was revised on the basis of the same design criteria and assumptions applied to the TWTNS.

Pipe diameter and head loss were calculated simultaneously with a flowrate and a pipe length. The flowrate data were obtained from the optimal solution to the mathematical optimization model and operational data.

The pipe length was obtained from the distance matrix among the wastewater sources, and the distributed and terminal WTPs. The Darcy–Weisbach equation was used to calculate the head loss (McGhee, 1991). The maximum head loss criteria were used to select nominal pipe diameter and were set at 200 and 20 kPa for pumping and gravity flows, respectively. Note that wastewater in the streams from the distributed WTPs to the terminal WTP was transferred by gravity, and wastewater in the others by pumps. Carbon steel was used as the pipe material. The Korean Standard, KS D3507 (Korean Standards Information Center, 2006), was used to obtain specific data, such as nominal diameter, wall thickness and weight. The minimum nominal pipe diameter was 0.0254 m for the simplicity of this design.

Pumps and electric motors were specified in detail in relation to the flowrate and pressure requirement. The discharge pressure of the pump was determined by adding the head losses in the pipes to the pressure required for the WTP, which was assumed to be 100 kPa.

The rates of chemicals consumption and sludge generation in the TWTNS were assumed to be equal to those in the CWTS. This was because the dosages of coagulants and flocculants used and the quantity of sludge generated are generally proportional to contaminant loads, which were the same in the CWTS and TWTNS. In addition, it is difficult to accurately estimate the amounts of chemicals used and sludge generated because the properties of wastewater such as pH, alkalinity, and the types of coagulants and flocculants have significant effects on the dosage of coagulants and flocculants and the quantity of sludge generated. The concentrations of the final effluent in the TWTNS were also assumed to be the same as those in the CWTS, because of the same chemicals dosages and contaminant loads.

2.3. Life cycle assessment

An LCA methodology was used to calculate and compare the environmental burdens associated with the CWTS and TWTNS. All environmental burdens generated from inputs and outputs throughout their life cycle were considered by assessing the design results of the two wastewater treatment systems. The LCA procedure was performed in accordance with the ISO 14040 series of standards (ISO, 1997): goal and scope definition, life cycle inventory analysis (LCI), life cycle impact assessment (LCIA), and life cycle interpretation.

The goal and scope definition included the goal, system, function, functional unit, reference flow, system boundaries, allocation, data quality and assumptions. The goal of this LCA was to estimate the environmental feasibility of the TWTNS. The system and its function were defined as a wastewater treatment system (TWTNS or CWTS) used to consistently meet the discharge permits of the environmental regulations. The functional unit was defined as the wastewater treatment system used to treat the seven

wastewater sources shown in Table 1 during its life cycle (15 years), and the reference flow was set to one unit of the wastewater treatment system. The system boundaries included all items in the wastewater treatment system, with the exception of the existing distributed and terminal WTPs and existing pump pits which could be neglected as a baseline in the comparison of the TWTNS and CWTS. However, the operations and maintenance (O&M) of mechanical equipment (agitators, flocculators, and sludge scrapers) in the distributed and terminal WTPs was included in the system boundaries of this LCA. This was because some of the distributed WTPs and their mechanical equipment were not used in the TWTNS as a result of the synthesis. The allocation was not required in this LCA. The same data qualities were used for a comparative assessment because most of the data were calculated from the same design criteria and assumptions. The service life of the wastewater treatment systems was assumed to be 15 years with respect to the lifetime of the pipes and mechanical equipment.

A life cycle inventory analysis (LCI) was performed to quantify all the inputs and outputs associated with each wastewater treatment system throughout its construction, operations and maintenance (O&M), and disposal stages. The GaBi 4.0 (IKP and PE Europe GMBH, 2004) and Ecoinvent v1.2 (Swiss Center for Life Cycle Inventories, 2005) databases were used for the LCI. The inventory in the construction stage included the manufacture of the pipes, equipment (pumps and motors), as well as associated transportation and piping works. The inventory in the O&M stage included the consumption of electricity for pumping and wastewater treatment, and maintenance and repair (M&R). Inputs and outputs in the M&R were assumed to be 3% of those in the construction stage taking into account that an M&R cost can be assumed to be 3% of a construction cost (Peters and Timmerhaus, 1991). The disposal stage included the recycling of steel, iron, and copper.

A life cycle impact assessment (LCIA) was performed to evaluate the significance of potential environmental impacts on the basis of the results of the LCI. The CML 2001 methodology (IKP and PE Europe GMBH, 2004) was used for the classification and characterization to evaluate environmental effect scores from all contributors in the TWTNS and CWTS. Environmental impact categories consist of abiotic depletion, acidification, eutrophication, freshwater aquatic ecotoxicity, global warming, human toxicity, marine aquatic ecotoxicity, ozone layer depletion, photochemical ozone creation, radioactive radiation, and terrestrial ecotoxicity.

A life cycle interpretation was carried out to comprehensively estimate the results of the preceding steps. In this stage, the environmental feasibility of the synthesis of the distributed and terminal WTPs was finally estimated by comparing the environmental effect scores of the TWTNS with those of the CWTS. The overall environmental scores of the CWTS were used as a control.

2.4. Life cycle costing

An LCC methodology was used to estimate the economic feasibility of the TWTNS. A life cycle for the LCC was divided into four categories: design and supervision, construction, O&M, and disposal. The design and supervision stage consisted of basic and detailed designs and supervision. The cost in the construction stage was divided into costs for piping, equipment (pump and motor) and construction expenses, as well as a contractor's overhead and profits. The O&M cost included the costs of electricity, labor, and M&R. The electricity cost was required for pumping, as well as wastewater treatment, such as the operation of agitators, flocculators and sludge scrapers. The labor cost consisted of operators' wages, as well as their expenses and overheads. The M&R cost was assumed to be 3% of a construction cost (Peters and Timmerhaus, 1991). The disposal stage consisted of pipe decommissioning, scrap (iron, steel and copper) recycling and construction expenses, as well as a contractor's overhead and profits. The costs were estimated in detail using databases containing price and cost information (Korea Price Information Corp., 2006a, b). The service life of the LCC was set at 15 years, as in the LCA.

All the costs of the contributors classified above and the life cycle cost of the TWTNS were compared to those of the CWTS to estimate the economic feasibility of the TWTNS. Future costs were discounted to present values taking into account the time value of money, and summed to obtain the life cycle cost of each wastewater treatment system. O&M costs recur annually, and disposal costs are incurred at the end of the service life. Accordingly, these costs should be converted into present values in order for them to be equally estimated and compared to an initial capital investment cost (Peters and Timmerhaus, 1991; Gluch and Baumann, 2004). The present value was estimated using the following equation:

$$PV = P(1 + e)^t / (1 + i)^t \quad (18)$$

To discount future costs, the interest rate was set at 5.7%, the yield of treasury bonds (5 years) over the last 10 years in South Korea, and the escalation rate was assumed to be the 3.0% targeted by the Bank of Korea for the period between 2004 and 2006 (Korea National Statistical Office, 2006; The Bank of Korea, 2006).

3. Results and discussion

The CWTS used as a control to evaluate the environmental and economic feasibility of the TWTNS is illustrated in Fig. 2a. The TWTNS shown in Fig. 2b was generated from the optimal solution to the mathematical optimization model reflecting the superstructure model shown in Fig. 1. The three distributed WTPs (TP2, TP3, TP4) were not used in the TWTNS, because the capacities of the other distributed WTPs were sufficient to treat wastewater from other sources, and a portion of

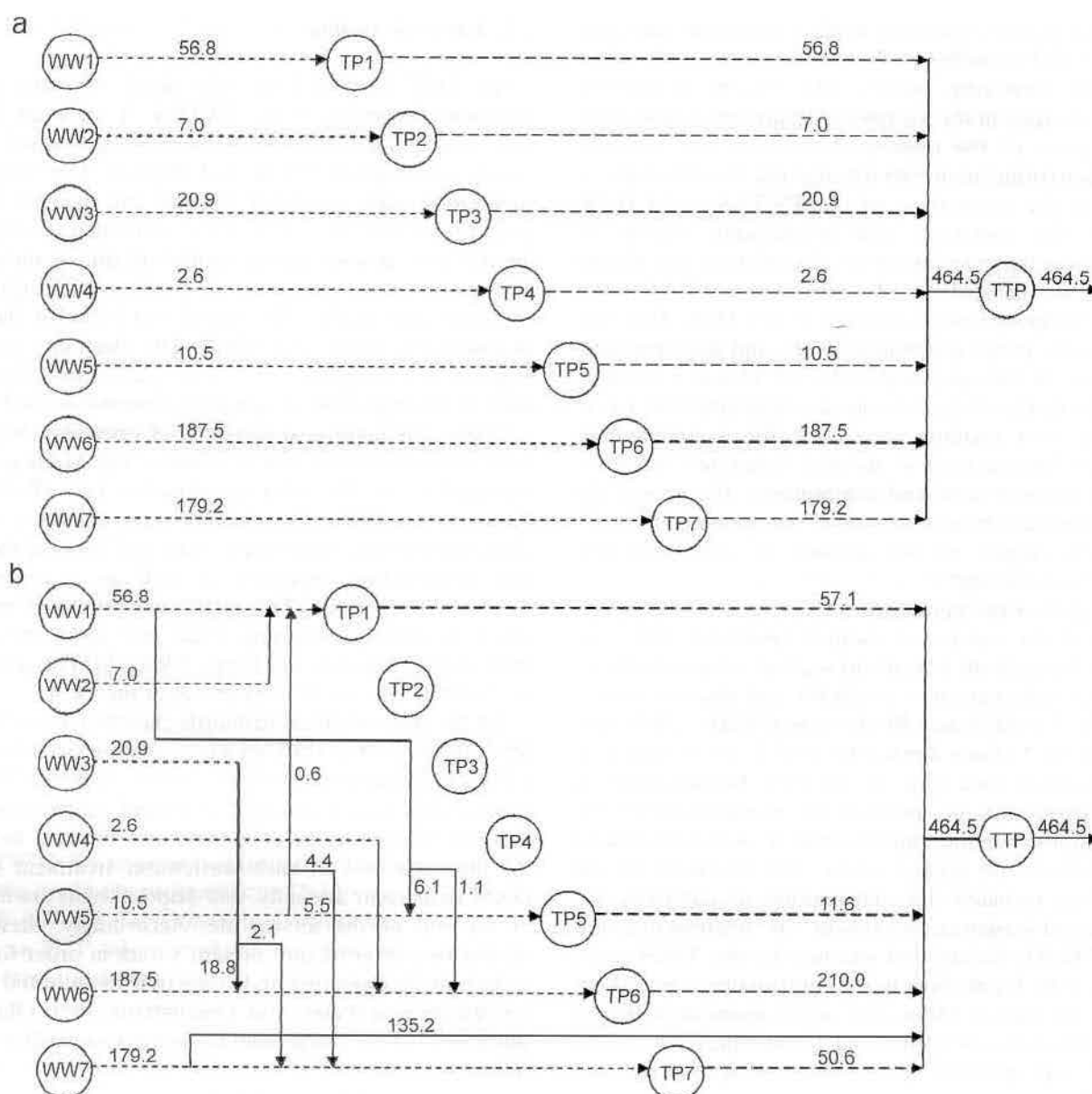


Fig. 2. Comparison of the wastewater treatment systems: (a) conventional wastewater treatment system (CWTS); (b) total wastewater treatment network system (TWTNS). (WW: wastewater, TP: distributed WTP, TTP: terminal WTP).

wastewater (WW7) not treated in the distributed WTPs was bypassed to the terminal WTP. This was attributed to the decrease of wastewater generation rates in the wastewater sources as a result of the considerable effort which had already been made to reduce water consumption rates in the plant: water reuse, recycling and recirculation. In other words, the present wastewater generation rates in the wastewater sources were less than the design capacities of the existing WTPs. The operators in the iron and steel plant were in charge of the integrated operations of production processes and the distributed WTP, allowing them to directly cope with the variations of the flowrate and concentrations of wastewater depending on the operations of production processes. Therefore, the closing of the three distributed WTPs would allow the operators to concentrate

on the production processes and enhance their productivity, besides reducing the labor cost. The interconnections diluting the influents of the distributed WTPs with their effluents were not generated through the synthesis, because of the objective function used to minimize the total flowrate of wastewater treated in the distributed WTPs.

Tradeoffs among the items were incurred as a result of the synthesis of the distributed and terminal WTPs, as summarized in Table 4. The total wastewater flowrate treated in the distributed WTPs of the TWTNS was 29.1% lower than that of the CWTS because the objective function drove a portion of the wastewater to bypass the distributed WTP. The total length of pipes in the TWTNS was 60.0% greater than that in the CWTS because of the interconnections for the networking in the TWTNS.

Table 4
Summary of the design results of the conventional wastewater treatment system (CWTS) and total wastewater treatment network system (TWTNS)

Item		Unit	CWTS	TWTNS
Pipe	Length	m	28,510	45,630
	Weight	kg	1,504,182	1,572,751
Pump	Weight	kg	603	616
	Weight	kg	659	684
Motor	Electricity (pumping)	kW	48.0	96.2
	Electricity (wastewater treatment)	kW	17.1	15.7
Utility consumption	Distributed WTPs	m ³ /h	464.5	329.3
	Terminal WTP	m ³ /h	464.5	464.5

However, it should be noted that the total weight of pipes in the TWTNS was just 4.6% greater than that in the CWTS, because the three streams by gravity requiring the larger diameter of pipes were removed in the TWTNS. The total weight of pumps and motors in the TWTNS were 2.2% and 3.8% greater than those in the CWTS, respectively, and the electricity consumption for pumping in the TWTNS was 100.4% greater than that in the CWTS. This was because the increase of the pipe length in the TWTNS required the higher power capacity of pumps to transfer wastewater to distributed WTPs networked. However, the electricity consumption for wastewater treatment in the TWTNS was just 8.2% less than that in the CWTS, because of the closing of the three distributed WTPs. This small difference was because the mixing power of agitators and flocculators is related to the volume of coagulation and flocculation basins but not to the flowrate of wastewater.

Tradeoffs among the environmental effect scores of the contributors to the environmental burdens were incurred in the synthesis of the distributed and terminal WTPs, as shown in Table 5. The electricity consumption for pumping was a principal contributor to the environmental burdens of the CWTS and TWTNS. The proportions of the effect scores from the electricity consumption for pumping to the total effect scores summed over the life cycle were from 32.8% to 69.6% for the CWTS and from 44.2% to 82.9% for the TWTNS. The effect scores of the electricity consumption for pumping in the TWTNS were 100.4% higher than those in the CWTS. The effect scores of electricity consumption for wastewater treatment in the TWTNS were 8.2% lower than those in the CWTS. Most of the effect scores of the other contributors in the TWTNS were higher than those in the CWTS. It should be noted that the decrease of the effect scores of the electricity consumption for wastewater treatment caused the increase of effect scores of the other contributors as a result of the synthesis of the distributed and terminal WTPs.

Most of environmental burdens were incurred in the O&M stage, which was in accordance with the fact that the electricity consumption for pumping was a principal contributor to environmental burdens. The proportions of the effect scores in the O&M stage to the total effect

scores summed over the life cycle were from 54.7% to 96.1% for the CWTS and from 58.7% to 97.5% for the TWTNS. The total effect scores in each life cycle stage in the TWTNS were higher than those in the CWTS, as shown in Table 5.

The environmental burdens from the TWTNS throughout the life cycle were greater than those from the CWTS. Table 5 shows that the total effect scores throughout the life cycle in the TWTNS were from 29.6% to 68.3% higher than those in the CWTS. Therefore, the TWTNS was not more environmentally friendly than the CWTS. This was because the decrease of the environmental impacts from the electricity consumption for wastewater treatment did not outweigh the increase of the environmental burdens from the other contributors in the TWTNS.

Tradeoffs among the economic costs of the contributors to the life cycle cost occurred in the synthesis of the distributed and terminal WTPs. Fig. 3 shows the results of the cost estimation in the construction stage. All the costs of the contributors in the TWTNS were higher than those in CWTS. The piping cost was a principal contributor in the construction stage because the proportions of the piping costs to the construction costs were 66.0% for the CWTS and 65.9% for the TWTNS. Fig. 4 shows the result of the cost estimation carried out on an annual basis in the O&M stage. The labor cost in the TWTNS was 29.0% lower than that in the CWTS because of the closing of the three distributed WTPs. The electricity cost for pumping in the TWTNS was 100.7% higher than that in the CWTS, while the electricity cost for wastewater treatment in the TWTNS was 8.2% lower than that in the CWTS. The M&R cost in the TWTNS was 5.3% higher than that in the CWTS, which was in accordance with the results of the piping and equipment costs. It should be noted that the decrease of the labor cost outweighed the increases of the electricity cost for pumping and the M&R cost in the O&M stage. The high proportion of the labor cost in the O&M cost was because all the WTPs were operated in four shifts. Fig. 5 shows all the costs incurred in the disposal stage. The pipe decommissioning cost was a principal contributor to the disposal cost. The results of the cost estimation in the disposal stage were in accordance with those in the construction stage. Higher revenue was obtained in the TWTNS from the recycling of copper,

Table 5
Results of the life cycle impact assessment for the conventional wastewater treatment system (CWTS) and total wastewater treatment network system (TWTNS)

Construction	Piping	CWTS	ADP	AP	EP	FAETP	GWP	HTP	MAETP	ODP	POCP	RAD	TETP
Construction	Piping	CWTS	3.3E+03	1.7E+03	3.6E+02	2.2E+05	4.1E+05	1.4E+06	2.6E+08	4.1E-02	4.7E+02	2.6E-03	3.3E+03
		TWTNS	3.4E+03	1.8E+03	3.7E+02	2.3E+05	4.3E+05	1.5E+06	2.7E+08	4.3E-02	4.9E+02	2.8E-03	3.4E+03
	Equipment	CWTS	1.2E+01	2.8E+01	4.8E-01	4.8E+01	1.8E+03	1.6E+03	3.9E+05	1.2E-04	1.9E+00	4.6E-05	4.7E+00
		TWTNS	1.2E+01	2.9E+01	5.0E-01	4.9E+01	1.8E+03	1.6E+03	4.1E+05	1.2E-04	2.0E+00	4.8E-05	4.9E+00
O&M	Total	CWTS	3.3E+03	1.7E+03	3.6E+02	2.2E+05	4.1E+05	1.4E+06	2.6E+08	4.1E-02	4.7E+02	2.7E-03	3.3E+03
		TWTNS	3.5E+03	1.8E+03	3.8E+02	2.3E+05	4.3E+05	1.5E+06	2.7E+08	4.3E-02	4.9E+02	2.8E-03	3.5E+03
	Electricity (pumping)	CWTS	2.7E+04	7.3E+03	7.5E+02	3.8E+05	4.6E+06	1.5E+06	2.1E+09	1.2E-01	6.9E+02	4.6E-02	6.3E+04
		TWTNS	5.5E+04	1.5E+04	1.5E+03	7.6E+05	9.3E+06	3.1E+06	4.2E+09	2.5E-01	1.4E+03	9.1E-02	1.3E+05
Disposal	Electricity (treatment)	CWTS	9.7E+03	2.6E+03	2.7E+02	1.4E+05	1.7E+06	5.4E+05	7.4E+08	4.4E-02	2.5E+02	1.6E-02	2.2E+04
		TWTNS	8.9E+03	2.4E+03	2.5E+02	1.2E+05	1.5E+06	5.0E+05	6.8E+08	4.0E-02	2.3E+02	1.5E-02	2.1E+04
	Maintenance & repairs	CWTS	1.5E+03	8.1E+02	1.9E+02	9.8E+04	2.0E+05	6.3E+05	1.2E+08	1.9E-02	2.1E+02	1.2E-03	1.5E+03
		TWTNS	1.6E+03	8.4E+02	2.0E+02	1.0E+05	2.0E+05	6.6E+05	1.2E+08	2.0E-02	2.2E+02	1.3E-03	1.6E+03
Total life cycle	Total	CWTS	3.8E+04	1.1E+04	1.2E+03	6.2E+05	6.5E+06	2.7E+06	2.9E+09	1.9E-01	1.1E+03	6.3E-02	8.7E+04
		TWTNS	6.5E+04	1.8E+04	2.0E+03	9.9E+05	1.1E+07	4.2E+06	5.0E+09	3.1E-01	1.8E+03	1.1E-01	1.5E+05
	Scrap recycling	CWTS	6.9E+03	3.2E+03	2.5E+02	2.0E+05	7.3E+05	4.5E+05	2.4E+08	5.1E-02	4.8E+02	3.5E-03	2.3E+02
		TWTNS	1.2E+04	5.3E+03	4.2E+02	1.2E+05	1.2E+06	7.5E+05	3.6E+08	8.5E-02	8.0E+02	5.7E-03	3.9E+02
Total life cycle	Total	CWTS	6.9E+03	3.2E+03	2.5E+02	2.0E+05	7.3E+05	4.5E+05	2.4E+08	5.1E-02	4.8E+02	3.5E-03	2.3E+02
		TWTNS	1.2E+04	5.3E+03	4.2E+02	1.2E+05	1.2E+06	7.5E+05	3.6E+08	8.5E-02	8.0E+02	5.7E-03	3.9E+02
	Total life cycle	CWTS	4.9E+04	1.6E+04	1.8E+03	1.0E+06	7.6E+06	4.5E+06	3.4E+09	2.8E-01	2.1E+03	6.9E-02	9.0E+04
		TWTNS	8.0E+04	2.5E+04	2.7E+03	1.3E+06	1.3E+07	6.4E+06	5.6E+09	4.4E-01	3.1E+03	1.2E-01	1.5E+05

The CML 2001 methodology was employed for classification and characterization. ADP: abiotic depletion potential (kg Sb-equivalents); AP: acidification potential (kg SO₂-equivalents); EP: eutrophication potential (kg phosphate-equivalents); FAETP: freshwater aquatic ecotoxicity potential (kg DCB-equivalents); GWP: global warming potential (100 years) (kg CO₂-equivalents); HTP: human toxicity potential (kg DCB-equivalents); MAETP: marine aquatic ecotoxicity potential (kg DCB-equivalents); ODP: ozone layer depletion potential (steady state) (kg R11-equivalents); POCP: photochemical ozone creation potential (kg ethene-equivalents); RAD: radioactive radiation (DALY); TETP: terrestrial ecotoxicity potential (kg DCB-equivalents); DCB: 1,4-dichlorobenzene.

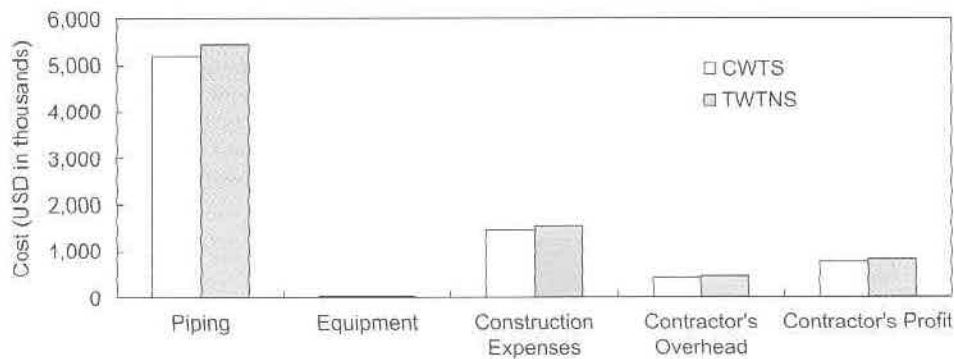


Fig. 3. Cost estimation of all the contributors in the construction stage (CWTS: conventional wastewater treatment system, TWTNS: total wastewater treatment network system).

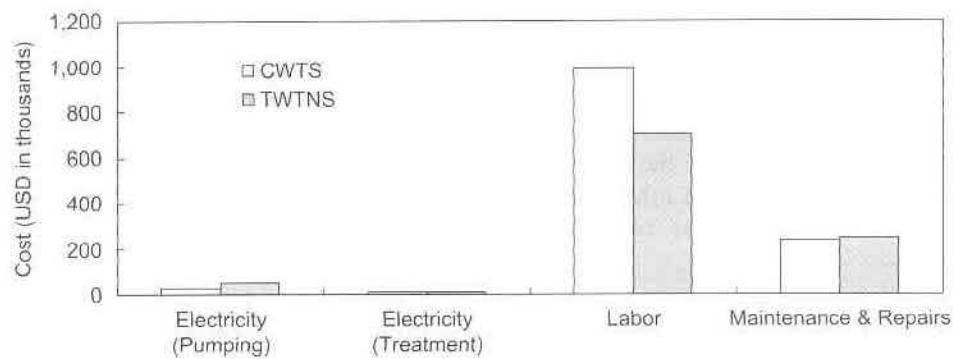


Fig. 4. Cost estimation of all the contributors on an annual basis in the operations and maintenance (O&M) stage (CWTS: conventional wastewater treatment system, TWTNS: total wastewater treatment network system).

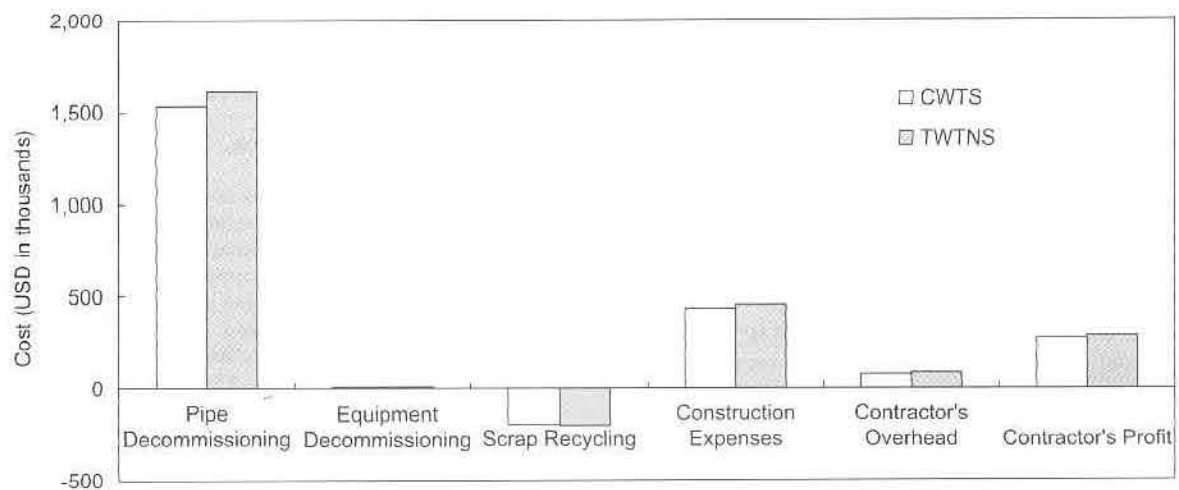


Fig. 5. Cost estimation of all the contributors in the disposal stage (CWTS: conventional wastewater treatment system, TWTNS: total wastewater treatment network system).

steel and iron scrap derived from the disposal of the pipes and equipment.

The synthesis of the distributed and terminal WTPs resulted in the reduction of the total life cycle cost throughout the life cycle. Fig. 6 shows the results of the

cost estimation in each life cycle stage and the life cycle costs. The costs of the design and supervision, construction and disposal in the TWTNS were from 5.2% to 5.4% higher than those in the CWTS, while the O&M cost in the TWTNS was 19.7% lower than that in the CWTS. The

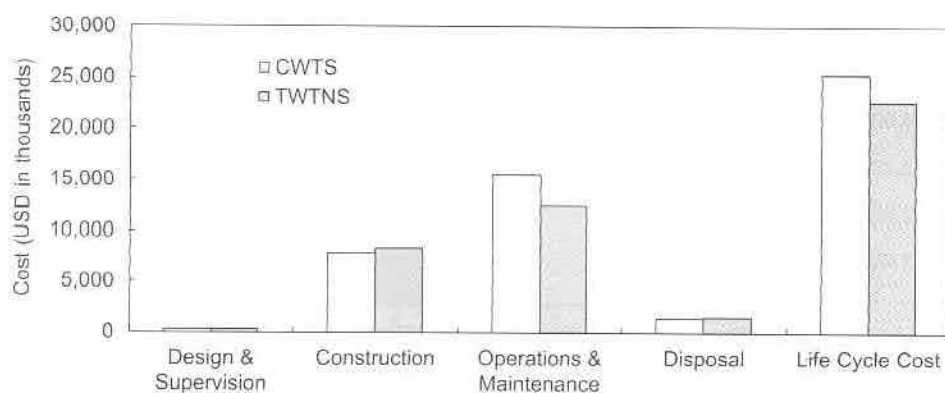


Fig. 6. Cost estimation of each life cycle stage and the life cycle cost. Future costs were discounted to present values (CWTS: conventional wastewater treatment system, TWTNS: total wastewater treatment network system).

proportions of the O&M cost to the life cycle costs were 61.5% for the CWTS and 55.0% for the TWTNS and were greatest among the costs in the life cycle stages. The life cycle cost in TWTNS was 10.1% lower than that in the CWTS. This was because the decrease of the O&M cost outweighed the increase of the costs in the other life cycle stages. Therefore, the TWTNS was more economically friendly than the CWTS.

4. Conclusions

The environmental and economic feasibility of the TWTNS was estimated using the LCA and LCC methods. The TWTNS was not more environmentally friendly than the CWTS, but was more economically friendly. When the ratios of the total environmental effect scores and economic costs throughout the life cycle in the TWTNS to those in the CWTS were equally compared, it turned out that the TWTNS was not eco-efficient from the viewpoint of eco-design, which is to comprehensively enhance environmental and economic performance. This was because the drawback in the environmental performance of the TWTNS overrode the enhancement of its economic performance.

In many industrial plants, wastewater generation rates have been decreased by the increasing efforts to reduce the consumption of freshwater: water reuse, recycling and recirculation. Consequently, many WTPs have been treating lower wastewater flowrates than their design capacities. Therefore, opportunities for the synthesis of existing distributed and terminal WTPs have been increasing. The methods and results of this environmental and economic feasibility study are expected to play important roles in the decision-making process used to determine whether to implement a TWTNS.

Acknowledgments

This work was financially supported in part by the Korean Science and Engineering Foundation (R11-2003-

006-01001-1) through the Advanced Environmental Biotechnology Research Center at Pohang University of Science and Technology and in part by the Korean Ministry of Commerce, Industry, and Energy through the Korean National Cleaner Production Center. This work was also supported by the program for advanced education of chemical engineers (2nd stage of BK21).

References

- Carpentieri, M., Corti, A., Lombardi, L., 2005. Life cycle assessment (LCA) of an integrated biomass gasification combined cycle (IBGCC) with CO₂ removal. *Energy Conversion and Management* 46, 1790–1808.
- Friedrich, E., 2002. Life-cycle assessment as an environmental management tool in the production of potable water. *Water Science and Technology* 20, 29–36.
- Galan, B., Grossmann, I.E., 1998. Optimal design of distributed wastewater treatment networks. *Industrial and Engineering Chemistry Research* 37, 4036–4048.
- GAMS Development Corporation, 2005. *GAMS: A User Guide*. Washington, DC.
- Gluch, P., Baumann, H., 2004. The life cycle costing (LCC) approach: a conceptual discussion of its usefulness for environmental decision-making. *Building and Environment* 39, 571–580.
- Hernandez-Suarez, R., Castellanos-Fernandez, J., Zamora, J., 2004. Superstructure decomposition and parametric optimization approach for the synthesis of distributed wastewater treatment networks. *Industrial and Engineering Chemistry Research* 43, 2175–2191.
- Huang, C., Chang, C.T., Ling, H., Chang, C.C., 1999. A mathematical programming model for water usage and treatment network design. *Industrial and Engineering Chemistry Research* 38, 2666–2679.
- IKP, PE Europe GMBH, 2004. *Gabi 4.0 Software*.
- ISO, 1997. *Environmental management—life cycle assessment—principles and framework*. ISO 14040.
- Korea National Statistical Office, 2006. <<http://www.nso.go.kr>> (accessed January 2006).
- Korea Price Information Corp., 2006a. *The Cost Estimation*. Seoul, South Korea.
- Korea Price Information Corp., 2006b. *The Price Information*. Seoul, South Korea, vol. 422.
- Korean Standards Information Center, 2006. <<http://www.standard.go.kr>> (accessed January 2006).
- Kuo, W.-C., Smith, R., 1997. Effluent treatment system design. *Chemical Engineering Science* 52 (23), 4273–4290.

- Mann, J.G., Liu, Y.A., 1999. Industrial Water Reuse and Wastewater Minimization. McGraw-Hill, New York.
- McGhee, T.J., 1991. Water Supply and Sewerage. McGraw-Hill, New York.
- Peters, M.S., Timmerhaus, K.D., 1991. Plant Design and Economics for Chemical Engineers. McGraw-Hill, New York.
- Spitzley, D.V., Grande, D.E., Keoleian, G.A., Kim, H.C., 2005. Life cycle optimization of ownership costs and emissions reduction in US vehicle retirement decisions. *Transportation Research Part D* 10, 161–175.
- Swiss Center for Life Cycle Inventories, 2005. Ecoinvent data version 1.2.
- The Bank of Korea, 2006. <<http://www.bok.or.kr>> (accessed January 2006).
- Wang, Y.P., Smith, R., 1994. Design of distributed effluent treatment systems. *Chemical Engineering Science* 49 (18), 3127–3145.

K. Atreya, S. Sharma, R.M. Bajracharya and N.P. Rajbhandari	Developing a sustainable agro-system for central Nepal using reduced tillage and straw mulching	547
H.-w. Sun and Q.-s. Yan	Influence of pyrene combination state in soils on its treatment efficiency by Fenton oxidation	556
S.-R. Lim, D. Park and J.M. Park	Environmental and economic feasibility study of a total wastewater treatment network system	564

Journal of
Environmental
Management

August 2008 Volume 88 Number 3

Contents

Y. Aşçı, M. Nurbaş and Y. Sağ Açikel	A comparative study for the sorption of Cd(II) by K-feldspar and sepiolite as soil components, and the recovery of Cd(II) using rhamnolipid biosurfactant	383
M. Geza and J.E. McCray	Effects of soil data resolution on SWAT model stream flow and water quality predictions	393
C. Mata, I. Hervás, J. Herranz, F. Suárez and J.E. Malo	Are motorway wildlife passages worth building? Vertebrate use of road-crossing structures on a Spanish motorway	407
D.M. Johnson, D.R. Hokanson, Q. Zhang, K.D. Czupinski and J. Tang	Feasibility of water purification technology in rural areas of developing countries	416
C. Song, J. Zhang, Y. Wang, Y. Wang and Z. Zhao	Emission of CO ₂ , CH ₄ and N ₂ O from freshwater marsh in northeast of China	428
H.E. Muga and J.R. Mihelcic	Sustainability of wastewater treatment technologies	437
S. Gokgoz-Kilic and M.M. Aral	Probabilistic fugacity analysis of Lake Pontchartrain pollution after Hurricane Katrina	448
R. Zhao and S. Chen	Fuzzy pricing for urban water resources: Model construction and application	458
L. Scholes, D.M. Revitt and J.B. Ellis	A systematic approach for the comparative assessment of stormwater pollutant removal potentials	467
M.A.B. Promentilla, T. Furuichi, K. Ishii and N. Tanikawa	A fuzzy analytic network process for multi-criteria evaluation of contaminated site remedial countermeasures	479
R.H. Crawford	Validation of a hybrid life-cycle inventory analysis method	496
A. Elkamel, E. Fatehifar, M. Taheri, M.S. Al-Rashidi and A. Lohi	A heuristic optimization approach for Air Quality Monitoring Network design with the simultaneous consideration of multiple pollutants	507
L. Ortega, S. Barrington and S.R. Guiot	Thermophilic adaptation of a mesophilic anaerobic sludge for food waste treatment	517
E. Dimitriou, I. Karaouzas, K. Sarantakos, I. Zacharias, K. Bogdanos and A. Diapoulis	Groundwater risk assessment at a heavily industrialised catchment and the associated impacts on a peri-urban wetland	526
S.M. Zalba, Y.A. Cuevas and R.M. Boó	Invasion of <i>Pinus halepensis</i> Mill. following a wildfire in an Argentine grassland nature reserve	539

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(200808)88:3;1-B

Available online at



ScienceDirect

www.sciencedirect.com