INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

UNIVERSITY OF CALIFORNIA

Los Angeles

Classification of Stormwater and Landuse using Neural Networks

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Civil Engineering

by

Haejin Ha

2002

UMI Number: 3070094

UMI®

UMI Microform 3070094

Copyright 2003 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© Copyright by

Haejin Ha

2002

The dissertation of Haejin Ha is approved.

Irwin H. Suffet

Keith D. Stolzenbach

Thomas C. Harmon

M. Indira Venkatesan

Venkatesan M. Indira

Michael K. Stenstrom, Committee Chair

University of California, Los Angeles

2002

To my parents-in-law who helped both materially and morally,

to my lovely daughters Terry and Tammy,

to my husband Simon, and

to my classmate Leehyung;

without their love and understanding,

this dissertation would not have been completed

TABLE OF CONTENTS

List of Figures

List of Tables

VITA

November 24, 1966	Born, Kwangju, South Korea
1990	B.S. Chemical Engineering Chonnam National University Kwangju, South Korea
1992	M.S. Chemical Engineering Chonnam National University Kwangju, South Korea
1998	M.S. Civil and Environmental Engineering University of California, Los Angeles Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

Ma, J-S., Khan, S., Li, Y-X., Kim, L-H., Ha, H., Lau, S-L., Kayhanian, M., and Stenstrom, M.K. (2002) Implication of oil and Grease measurement in stormwater management systems. To be presented at the 9th ICUD, Portland, Oregon, September 2002.

Stenstrom, M.K. and Ha, H. (2001). Final Report AB-538 Study, September, 2001.

Stenstrom, M.K., Kim, L-H. and Ha, H. (2000). Inlet Deflector Feasibility Study, Research report for Department of California Transportation, May, 2000

Stenstrom, M.K., S-L. Lau, H-H. Lee, J-S. Ma, H. Ha, L-H. Kim, S. Khan and M. Kayhanian (2001). First Flush Stormwater runoff from Highways, *Conference proceedings of Environmental and Water Resources Institute's (EWRI's) World Water & Environmental Resource Congress*, May, Orlando, Florida

Stenstrom, M.K., S-L. Lau, H-H. Lee, J-S. Ma, H. Ha, L-H. Kim, S. Khan and M. Kayhanian (2001). First Flush Stormwater runoff from Highways, *Research report for Department of California Transportation*, September, 2000.

Stenstrom, M.K., S-L. Lau, J-S. Ma, H. Ha, L-H. Kim, S. Lee, S. Khan and M. Kayhanian (2001), First Flush Stormwater runoff from Highways, *Research report for Department of California Transportation*, September, 2001.

ABSTRACT OF THE DISSERTATION

Classification of Stormwater and Landuse

using Neural Networks

by

Haejin Ha

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2002

Professor Michael K. Stenstrom, Chair

Stormwater runoff is a major contributor to the pollution of coastal waters in the United States. Differences in landuse patterns result in different pollutant concentrations, and therefore landuse-related control strategies are essential to control storm water pollution effectively. An approach that can differentiate landuse types in stormwater could provide opportunities for better landuse management to minimize stormwater pollution.

A neural network model was applied to examine the relationship between stormwater water quality and various types of landuse. The neural model can be used to identify landuse types for future known and unknown cases. The neural model uses a Bayesian network and has ten water quality input variables, four neurons in the hidden layer, and

five landuse target variables. The neural model correctly classified 92.3 % of test files. Simulations were performed to predict the landuse type of a known data set, and accurately described the behavior of the new data set. This study demonstrates that a neural network can effectively classify landuse types with water quality data.

A similar approach was applied to two local stormwater monitoring programs, which use human activity as measured by landuse or standard industrial classification code to describe stormwater, in the hopes that these classifications will be useful to planners and regulators in abating stormwater pollution. Data sets produced by the landuse based program were successfully identified by the neural network, and the monitoring program is successful in accomplishing its goals. The industrial stormwater monitoring program is not successful; standard industrial classification code is not related to stormwater quality. Improvements are suggested, which include sample type, parameters and timing.

Beach water quality monitoring programs were next evaluated using examples from Southern California. Data collection that is useful in minimizing fecal contamination from human and animal waste is their major objective. Lack of specificity of indicator organisms is the major problem, and better, real-time indicators are needed. Beach monitoring programs seldom collect water quality or other data that might be used with neural network techniques to identify pollutant sources. Methods to detect human fecal pollution and differentiate it from other sources such as animals are reviewed. Microbial methods, especially those using molecular biology, and chemical methods are reviewed.

At present there is no easy, low cost or rapid method for differentiating between human and non-human fecal contamination. It is much more likely that a combination of methods can be used to accurately identify human fecal pollution.

Chapter 1.

Introduction

California coastal waters are important recreational and economic resources, which make the safety of coastal waters of concern to both state and county health departments and beachgoers (Jiang et al., 2001). The completion of wastewater treatment plants mandated by the Clean Water Act has reduced conventional water pollution to California's beaches and bays. As a result, non-point source pollution such as stormwater runoff is now a major contributor to the pollution of the coastal water including Santa Monica Bay, which is among the most severely polluted Bays in the United States (Wong et al., 1997).

Stormwater pollution starts with precipitation falling on the ground. When the precipitation exceeds the capacity of the land to retain the rainfall, stormwater runoff is produced. The stormwater runoff picks up natural and human-made contaminants that accumulate on the ground during dry days and carries them directly into the receiving waters without any treatment. These contaminants may include heavy metals, pesticides, and other organic compounds, inorganic phosphates and nitrates, radionuclides, ammonia, and sediments. Stormdrains /urban runoff accounted 34% of the Beach Mile-Day (one linear mile of beach closed for one day, or BMD) closures and warnings, and are a main cause of permanent beach postings at many California beaches (State Water Resources Control Board, 2000 and 2001). The problem of stormwater pollution is becoming worse because of population growth, which results in increased impermeable area.

It is generally recognized that different human activities will create different types and varying concentrations of stormwater contaminants (Stenstrom et al. 1984). For example, runoff from transportation-associated landuse is a primary source of metals and hydrocarbons (LACDPW and Woodward-Clyde, 1998). Vehicles release hydrocarbons from leaks, engine byproducts and unburned fuel and various metals from corrosion, fuel combustion and wearing surfaces such as brake pads (Rogge et al. 1993; Sansalone and Buchberger, 1997). Therefore landuse-related control strategies are essential to control storm water pollution effectively.

An approach that can differentiate stormwater from different landuse types could help to better understand the system, and eventually provide opportunities for better landuse management to control stormwater pollution. To examine the relationship between water quality variables and various types of landuse, a neural network was applied. The neural network model can then be used to identify landuse type for future known and unknown cases. If the network cannot confirm the known landuse, it suggests that something else is occurring such a spill, leakage, illegal discharge, or that monitoring data are suspect.

A neural network (NN), more specifically an artificial neural network, is a computational tool that operates similarly to the biological processes of human brain. NN is in the "black-box" class of models. These models do not require detailed knowledge of the internal functions of a system in order to recognize relationships between inputs and outputs (El-Din and Smith, 2002). Because of this feature, NN modeling is being increasingly applied in various aspects of science and engineering, including the environmental field. Neural networks (NNs) have been used to predict wastewater inflow rate (El-Din and Smith, 2002), sanitary sewer flows (Djebbar and Kadota, 1998), the flux during ultrafiltration and after backwashing (Teodosiu et al., 2000), peak *Cryptosporidium* and *Giardia* concentrations (Neelakantan et al., 2001), and metal bioleaching in municipal sludge (Laberge et al., 2000). In addition, NNs have been applied to simulate nitrate leaching (Kaluli et al., 1998), model solid transport in sewers (Gong et al., 1996) and to identify non-point sources of fecal contamination (Brion and Lingireddy, 1999). In this study, a NN was used to the identify landuse types as a function of stormwater quality data.

Public agencies are responding by requiring stormwater monitoring to satisfy the National Pollutant Discharge Elimination System (NPDES) Stormwater Permit as authorized by the Clean Water Act. For example, the Los Angeles County Department of Public Works (LACDPW) has been monitoring stormwater under the 1990 NPDES Municipal Permit (No. CA0061654) and later 1996 Municipal Permit (No. CAS614001) since the 1994-1995 wet seasons. Addition sampling is required by other agencies, such as the City of Los Angeles and the California Department of Transportation. Similar programs are underway in other areas of California and the United States.

The existence of stormwater monitoring programs should represent progress towards achieving clean water goals; however, studies have not yet been performed to understand the utility of the current programs or to improve their usefulness. Several monitoring programs were evaluated to determine if the results will be helpful to planners and regulators in abating stormwater pollution. In this study, datasets from a major municipal program, several research projects, beach monitoring, and a large self-monitoring program were used.

Study Objectives

The primary objective of this study was to develop a neural network model to examine the relationships between stormwater quality variables and landuse types. Chapter 2 describes a model developed using a Bayesian network, which was then used to identify landuse types for known cases. Based on this experience, the industrial stormwater permit monitoring was evaluated to determine if monitoring will be helpful to planners and regulators in abating stormwater pollution. Chapter 3 focused on the evaluation with the other monitoring programs. Chapter 4 is a review of indicators of human fecal contamination of surface waters including beach waters, and their ability to differentiate human contamination from other sources, such as animals and soil. The review includes microbial methods, especially those using molecular biology, and chemical methods. An early objective of this research was to use neural networks with fecal indicators to identify contamination sources, but too little data are available and indicators are not robust.

References

Brion, G.M. and Lingireddy, S. (1999) A neural network approach to identifying non-point sources of microbial contamination. *Water Research* **33**(14), 3099 – 3106.

Djebbar, Y. and Kadota, P.T. (1998) Estimating sanitary flows using neural networks. *Water Science and Technology* **38**(10), 215-222.

El-Din, A.G. and Smith, D.W. (2002) A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Research* **36**, 1115-1126.

Gong, N., Denoeux, T. and Bertrand-Krajewski, J-L. (1996) Neural networks for solid transport modeling in sewer systems during storm events. *Water Science and Technology* **33**(9), 85-92.

Jiang, S., Noble, R., and Chu, W. (2001) Human adenoviruses and coliphages in urban runoff impacted coastal waters of Southern California. *Applied and Environmental Microbiology,* **67**(1), 179-184.

Kaluli, J.W., Madramootoo, C.A. and Djebbar, Y. (1998) Modeling nitrate leaching using neural networks. *Water Science and Technology* **38**(7), 127 – 134.

Laberge, C., Cluis, D. and Mercier, G. (2000) Metal bioleaching prediction in continuous processing of municipal sewage with *Thiobacillus Ferrooxidans* using neural networks. *Water Research* **34**(4), 1145-1156.

LACDPW and Woodward-Clyde (1998) Los Angeles County 1997-1998 Stormwater Monitoring Report. Prepared by Los Angeles County Department of Public Works, Alhambra, CA, Woodward-Clyde, Oakland, CA and Southern California Coastal water Research Project, Westminster, CA, July 1998.

Neelakantan, T.R., Brion, G.M. and Lingireddy, S. (2001) Neural network modeling of *Cryptosporidium* and *Giardia* concentrations in the Delaware River, USA. *Water Science and Technology* **43**(12),125-132.

Rogge, W.F., Hildemann, L.M., Mazurek, M.A., and Cass, G.R. (1993) Sources of fine organic aerosol. 3. Road dust, tire debris, and organometallic break lining dust: Roads as sources and sinks. *Environmental Science and Technology,* **27**(9), 1892-1904.

Sansalone, J.J. and Buchberger, S.G. (1997) Partitioning and first flush of metals in urban roadway storm water. *Journal of Environmental Engineering,* **123**(2), 134-143.

State Water Resources Control Board, California Environmental Protection Agency (2000) *California Beach Closure Report 1999.*

State Water Resources Control Board, (2001) California Beach Closure Report 2000. Prepared by the State Water Resources Control Board, California Environmental Protection Agency, July 2001.

Stenstrom, M. K., Silverman, G.S., and Bursztynsky, T.A. (1984) Oil and Grease in Urban Stormwaters, *Journal of the Environmental Engineering Division,* ASCE, **110**(1), 58-72.

Teodosiu, C., Pastravanu, O. and Macoveanu, M. (2000) Neural network models for ultrafiltration and backwashing. *Water Research* **34**(18), 4371-4380.

Wong, K., Strecker, E.W. and Stenstrom, M.K. (1997) A Geographic Information System to estimate stormwater pollutant mass loadings. *Journal of the Environmental Engineering*, ASCE, **123**(8), 737-745.

Chapter 2.

Identification of land use with water quality data in stormwater using a neural network

Abstract

To control stormwater pollution effectively, development of innovative, landuse-related control strategies will be required. An approach that could differentiate landuse types from stormwater quality would be the first step to solving this problem. We propose a neural network approach to examine the relationship between stormwater water quality and various types of landuse. The neural network model can be used to identify landuse types for future known and unknown cases. The neural model uses a Bayesian network and has ten water quality input variables, four neurons in the hidden layer, and five landuse target variables (commercial, industrial, residential, transportation, and vacant). We obtained 92.3 percent of correct classification and 0.157 root-mean-squared error (RMSE) on test files. Based on the neural model, simulations were performed to predict the landuse type of a known data set, which was not used when developing the model. The simulation accurately described the behavior of the new data set. This study demonstrates that a neural network can be effectively used to produce landuse type classification with water quality data.

Introduction

Stormwater is now a major non-point source contributor to coastal water pollution including Santa Monica Bay, which is among the most severely polluted Bays in the United States (Wong et al., 1997). Storm drains and urban runoff accounted for 34 percent of the beach mile-day beach closures and warnings (California State Water Resources Control Board, 2000). The problem of stormwater pollution is becoming worse because of population growth, which results in increased impermeable surfaces. Stormwater runoff picks up natural and human-made contaminants that accumulated on surfaces during the dry days and transports them to the coastal waters.

The forms and concentrations of contaminants from runoff are closely related to various types of landuse because human activity is different according to landuse. To control stormwater pollution effectively, development of innovative landuse-related control strategies will be required. An approach that could differentiate landuse types in stormwater would be a first step to solving this problem. We propose a neural network approach to examine the relationship between water quality variables and various types of landuse. The neural network model could then be used to identify landuse type for future cases where only the inputs are known. We could also apply the neural model to water quality data sets from known landuses. If the network cannot confirm the known landuse, it suggests that something else is occurring such a spill, leakage, illegal discharge, or that monitoring data are suspect. Such information could provide opportunities for better landuse management to control stormwater pollution.

A neural network (NN), more specifically an artificial neural network, is a computational tool that operates similarly to the biological processes of human brain. Many researchers have discussed the history, capability, kinds, structure, and learning algorithm of neural networks (Marther and Shaw, 1993; Basheer et al., 1996; Zhao et al., 1997; Walley and Fontama, 1998; Lek et al., 1999; Loke et al., 1997; Laberge et al., 2000; Gob et al., 2001). NN is in the "black-box" class of models. These models do not require detailed knowledge of the internal functions of a system in order to recognize relationships between inputs and outputs (El-Din and Smith, 2002). Because of this feature, NN modeling is being increasingly applied in various aspects of science and engineering, including the environmental field. Neural networks (NNs) have been used to predict wastewater inflow rate (El-Din and Smith, 2002), sanitary sewer flows (Djebbar and Kadota, 1998), the flux during ultrafiltration and after backwashing (Teodosiu et al., 2000), peak *Cryptosporidium* and *Giardia* concentrations (Neelakantan et al., 2001), and metal bioleaching in municipal sludge (Laberge et al., 2000). In addition, NNs have been applied to simulate nitrate leaching (Kaluli et al., 1998), model solid transport in sewers (Gong et al., 1996) and to identify non-point sources of fecal contamination (Brion and Lingireddy, 1999). In this study, a NN was applied to the identification of landuse types as a function of stormwater quality data.

Our objectives were to develop a neural network model to examine the relationships between stormwater quality variables and landuse types. We evaluated three approaches

: multi-layer perceptron, radian basis function and bayesian network. A model was developed using the bayesian network, which was then used to identify landuse types for known cases. Future applications will investigate unknown cases where only the inputs are available.

Overview of Bayesian Networks

A Bayesian Network (BN) is the application of probability theory in Bayesian statistics to a multi-layer perceptron (MLP), the most common supervised neural network. Therefore, a BN has many common features with MLP. The major difference is the way error is measured during training. In an MLP, only the weights between neurons are adjusted during training. In a BN, both these weights and a set of parameters can be altered to reduce error measure during training. This enables BN to produce a generalized model without a validation data file, which is particularly useful for relatively small data sets such as ours.

The BN used here is a three-layered, supervised feed-forward neural network with back-propagation algorithm. For supervised learning, the network can be trained using both input and target values. After training, when we present only the input values to the neural model, it will compute an output value that approximates the target value. In a feed-forward neural network, feed back loops are absent. Information is processed in a forward manner only from input to output, and thus it always gives the same output result for the same input. In a back-propagation algorithm, the prediction error is generated at the output layer of neurons and then propagates backwards through the network. It consists of an input layer. output layer and a hidden layer between the input and output layers. The number of neurons in the input layer is usually equal to the number of input variables. The number of output layer neurons is usually the same as the target variable number. The number of neurons in the hidden layer is determined to optimize performance.

Data Collection

Development data

Development data are used for building a neural network model. We selected the Los Angeles County Department of Public Works (LACDPW) landuse stormwater monitoring data as development data because they are representative of this study purpose and are produced by a well-regarded agency. The LACDPW has been monitoring landuse stormwater in the County of Los Angeles since 1996. They provide flow and water quality data for various types of landuse (commercial, educational, industrial, high-density residential, multi-family residential, mixed residential, transportation, and vacant) for every storm event. Flow-weighted composite samples were analyzed for many water quality variables including indicator bacteria, general minerals, nutrients, metals, semi-volatile organic compounds, oil and grease and pesticides. Records of the stormwater quality data in various types of landuse were obtained online (http://www.ladpw.org/wmd/NPDES/report_directory.cfm) for the 1998-

2001 seasons, and from the Los Angeles County Stormwater Monitoring Report, for the 1996-1998 seasons (LACDPW, 1997; LACDPW, 1998).

The development data set was divided into three data files: a training file, a validation file, and a test file. Validation data were used to monitor the neural model's performance during training to prevent problems such as overtraining. Test data were used to measure the performance of the trained neural model.

Run data

Run data, similar to test data, but not used when developing a neural network model, were used to test the developed model's suitability for identifying landuse types. Highway monitoring data being collected in our laboratory were used as a run data set. Three sites from interstate highways near UCLA have been monitored since 1999 (Stenstrom et al., 2000; Stenstrom et al., 2001). The sites in Los Angeles are adjacent to the 101 and 405 freeways, which are among the busiest freeways in the United States (280,000 to 330,000 average daily traffic). Runoff samples during storm events were analyzed for a large suite of parameters such as indicator bacteria, general minerals, nutrients, metals, polycyclic aromatic hydrocarbons, and oil and grease.

Data Screening and Preprocessing

Target variables and input variables

Commercial, industrial, residential, transportation, and vacant landuse types were chosen as the target variables because they cover most types of landuse in urban areas. For residential landuse, we selected high-density residential sites because they have been monitored longer than the multi-family residential and mixed residential sites. Table 2.1 shows the landuse distribution in their sampling sites by LACDPW. The Santa Monica Pier site, representative of commercial landuse, was not monitored in the 1999-2001 storm seasons because the City of Santa Monica was constructing a stormwater treatment plant.

Among approximately 90 water quality variables in stormwater, 42 candidate variables were initially selected because they were detected in more than 25 percent of the events. After clipping the data cases that contained missing values, 184 samples or cases remained. We next considered the maximum number of input variables within the available data set to develop a general neural network model. The following rough rule was used: "The minimum number of training data cases should be ten times the total number of input variables and target variables" (SPSS, 1997). Taking into account the five target variables and 85 percent of the available data for the training file, the maximum number of input variables was ten. Among the 42 water quality variables, the ten input variables were selected using discriminant analysis as a preliminary classification approach (Systat 9, SPSS Inc., Chicago, IL). Discriminant analysis was used initially to determine the most useful variables for classifying the target classes. We selected the top ten most useful variables among 42 water quality variables. The selected ten input variables were Potassium, Sulfate, Alkalinity, Dissolved Phosphorus, Nitrite-N, Total Dissolved Solids, Volatile Suspended Solids, Suspended Solids, Dissolved Copper, and Dissolved Zinc.

Clipping outliers

In a preliminary test, poor classification was obtained between target variables and the output because of data cases that contained unusually high values in the input variables. These high values were identified as outliers and the data cases with outliers we removed (clipped) from the data sets. Outliers were defined as any observation exceeding ten times its median value in the whole range of the data set. We found 11 data cases that contained outliers. After clipping, 173 storm events from the five landuse monitoring stations remained for the 1996-2001 seasons. The descriptive statistics of the input variables according to their landuse after clipping outliers are shown in Table 2.2.

Data file and multiple data sets

The 173 cases were divided into three data files. The training file contained 121 cases or 70 % of the total cases; the validation file contained 26 cases or 15% of the total cases, which left 26 cases or 15 % of the total cases for the test file. The cases for each file were selected randomly from the overall data set. For the BN, which does not require a validation file, the data that would have been used for validation was grouped with the training data.

Lack of data is a common problem in developing neural networks and data sets with a relatively small number of data cases such as ours are a typical. To partially overcome this problem, additional combinations were created by randomly selecting different new test data sets from the original data set. The remaining 85% of the data were then used as training and validation files. Three data sets were created, called A, B and C, from the original data and were used to measure the overall performance of the neural network model.

Building a neural network model

Selection of neural network

In this study, Neural Connection 2.1 (SPSS Inc. and Recognition Systems Inc, Chicago, IL) was used to build the neural model. Three supervised, feed-forward neural networks, namely, Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), and Bayesian Network (BN) were used. In the early work, the neural networks were trained at their default settings. Both MLP and BN used eight neurons in their hidden layer. The BN automatically normalizes the input data. The RBF network started with 5 centers, and increased in steps of 5 centers to 50 centers using a spline radial function. The performance of the networks was measured using correct classification in percentage and the Root-Mean-Square Error (RMSE), which is the mean square difference between the target and the actual output value. An ideal neural network model would correctly classify all cases and have zero RMSE. The overall performance of the neural networks, as the average over each of the three data sets, is presented in Fig.2.1. The best performing network in the preliminary evaluation was BN. For this reason, BN was selected for further development.

Architecture of the neural model

For all cases the number of neurons in the input layer was equal to the number of water quality input variables and the number of neurons in the output layer was equal to the number of landuse types. In many problems, a second hidden layer does not produce a large improvement in performance, and varying the number of hidden neurons in the one hidden layer is usually sufficient (El-Din and Smith, 2002). We evaluated the benefits of one to nine hidden neurons. The model had a significant improvement in the RMSE and classification accuracy as the neurons were increased from one to four. After four neurons, additional neurons had no benefit (see Fig. 2.2). In general, the number of neurons in the hidden layer should be as low as possible to make a generalized neural model. Therefore, the model that utilizes four neurons in the hidden layer was chosen to be the final model. The BN model had nineteen total neurons: ten in the input layer, four in the hidden layer, and five in the output layer.

Figure 2.3 shows the error reduction as a function of the number of training epochs. A stopping criterion was established based upon error improvement. If the reduction in

RMSE over a fixed number of epochs was less than 0.01, training was stopped. The fixed number of epochs was varied from 10 to 100, and 80 epochs was selected.

Results of the final model

The performance of the final neural model on the training and test data files in RMS error and correct classification is shown in Table 2.3. In general, the statistical quality of the neural model was high. We obtained 92.3 and 94.5 percent overall correct classifications and 0.157 and 0.154 overall RMSE on the test and training files, respectively.

The overall performance on each target landuse type in RMSE and correct classification, measured as the average over each of the three data sets, is shown in Fig.2.4. and Fig. 2.5. respectively. Vacant landuse type had the lowest value of RMSE among the five landuse types and 100 percent in overall correct classification on test and training files. The result indicates that vacant land is the most distinct among the five landuse types based on water quality data. This might be expected since there is almost no human activity to create contaminants.

Analysis of the final model

Sensitivity of input variables

To evaluate the sensitivity of the model to different water quality variables, each variable was omitted ("leave-one-out method") and the network was retrained with all other conditions as before. In this way the impact of each variable on the final result can be evaluated. The percent difference in RMSE between ten and the nine input variables is shown in Table 2.4. If the error increases after removing a variable, it means that the removed variable was helpful in improving classification. If the error is unchanged, it means that the variable was not helpful for classification. A negative percentage in Table 2.4 suggests that the variable was harmful and might not be related to land use. All cases in Table 2.4 are positive except sulfate in dataset A. The neural model was the most sensitive to the level of dissolved copper, and the least sensitive to the level of sulfate. The values shown in Table 2.4 are useful in that they suggest ways of differentiating landuse in future monitoring programs.

Simulation

The utility of a neural network model, and our reason for developing it, is to confirm landuses from a large number of datasets. An agency seeking to reduce stormwater pollution can review monitoring data from known land uses and compare them to benchmark data used in developing the neural network. If the network cannot confirm the known landuse, it suggests that something else is occurring or that monitoring data are suspect. Such a tool could be useful watershed-based approaches to stormwater management. Large data sets could be screened to identify opportunities for water quality improvement.

To demonstrate this concept, simulations were carried out on a known data set (motor vehicle highways) that was not used in developing the model. Table 2.5 shows the monitoring data for the selected input variables. All data were from flow-weighted composite samples for 2000-2001 wet season, and collected by our laboratory. Two variables present in the LACDPW, potassium and total dissolved solids (TDS), were not collected as part of the highway monitoring program. The two missing variables provide an opportunity to evaluate classification sensitivity by substituting hypothetical data for the missing values.

The model correctly classifies the land use if TDS and potassium concentrations typically associated with highway land use are substituted into the data set for the missing values for all samples except S3-E5 and S3-E6. By substituting a large range of possible values, the sensitivity of classification can be observed. Figure 2.6 shows the changes in classification caused by changing TDS and potassium concentrations. Low values of TDS and potassium are associated with transportation land use. If potassium concentrations increase, the classification changes to commercial and finally residential landuse. The TDS concentration "pushes" the classification towards commercial at higher concentrations.

Samples S3-E5 ad S3-E6 in Table 2.5 did not classify correctly with typical values of TDS and potassium; they were incorrectly classified as industrial, commercial or residential. The most likely reason is the low values of copper and zinc. To test this suggestion, the values of copper and zinc were increased. The classification changes to transportation as the copper, or copper and zinc concentrations are changed to 22.5 μg/l and 112.7 μg/l, respectively. The classification changes to industrial if only the zinc concentration is changed.

The simulation demonstrates the utility of the model for understanding why land use classifications change as a function of water quality variables. Further usage of the model may be helpful in identifying "problem" sites or developing a better classification system – a system that creates landuse types based upon water quality as opposed to observable human activities or the needs of public agencies (e.g. tax records).

Conclusions

A neural network model for identifying the various types of landuse with stormwater quality data was successfully developed using LADPW stormwater monitoring data, collected during 1996-2001. A Bayesian Network model was best and had ten water quality input variables, four neurons in the hidden layer and five landuse target variables. The statistical quality of the neural model was high. We obtained 92.3 and 94.5 percent of correct classification, and 0.157 and 0.154 in the RMSE on the test and training files (173 cases). The model was used as a simulation tool to predict landuse type from highways stormwater monitoring data that was not used to develop the model. The simulations showed the sensitivity to classification and demonstrated a method to identify water quality variables that affect classification.

This research has demonstrated that a neural network can be used to classify landuses from water quality data, and that the technique can be automated. An approach for identifying opportunities for water quality improvement could be developed using this concept. Such information could provide opportunities for better management to control stormwater pollution.

Table 2.1. Land Use distribution of the monitored stations by LACDPW

Station Name	Drainage Area (Square km)	Land Use Distribution [a] (%)
Santa Monica Pier	0.33	**Commercial/Retail** (53.6)
		Other (36.7)
		Multi-Family Residential (5.1)
		Transportation (4.6)
Sawpit Creek	13.41	**Vacant** (98)
		Other (2)
Project 620	0.49	**High-density Residential** (100)
Dominguez Channel	3.65	**Transportation** (75.2)
		Light Industrial (17)
		Other (7.1)
		High Density Residential (0.6)
		Retail/Commercial (0.1)
Project 1202	2.77	**Light Industrial** (67.1)
		Other (26.9)
		Transportation (4.7)
		Vacant (1)
		Retail/Commercial (0.3)

[a] Bold letter indicates the major landuse in the station.

Table 2.2. Continued

Land Use		Potassium (mg/l)	Sulfate (mg/l)	Alkalinity (mg/l)	Dissolved Phosphorus (mg/l)	Nitrite-N (mg/l)	Total Dissolved Solids (mg/l)	Suspended Solids (mg/l)	Volatile Suspended Solids (mg/l)	Dissolved Copper (ug/l)	Dissolved Zinc (ug/l)
Transport	n	56	56	56	56	56	56	56	56	56	56
	Min.	0	1.59	7.7	0	0	18	15	4	0	0
	Max.	5.66	37	58.3	1.08	0.359	180	384	136	95	724
	Median	1.98	8.29	16.3	0.318	0.081	56	59	25	28.8	199.5
	Mean	2.2	9.52	19.76	0.37	0.1	61.75	83.04	32.14	32.93	210.56
	Std. Dev.	1.13	6.97	9.36	0.26	0.07	30.24	71.46	24.4	21.58	137.98
Vacant	n	30	30	30	30	30	30	30	30	30	30
	Min.	1.05	10.9	148	0	0	206	3	0	0	0
	Max.	3.48	31.1	200	0.316	0.043	264	567	76	1	0
	Median	2.385	16.2	175	0	0	242	28.5	15.5	0	0
	Mean	2.4	19.19	174.8	0.03	0.01	241.27	113.1	19.77	0.07	0
	Std. Dev.	0.55	7.08	12.71	0.06	0.02	14.3	155.56	18.06	0.25	0

Table 2.2. Descriptive statistics of the input variables according to their land use after clipping outliers
(0 indicated level below detection limit)

Land Use		Potassium (mg/l)	Sulfate (mg/l)	Alkalinity (mg/l)	Dissolved Phosphorus (mg/l)	Nitrite-N (mg/l)	Total Dissolved Solids (mg/l)	Suspended Solids (mg/l)	Volatile Suspended Solids (mg/l)	Dissolved Copper (ug/l)	Dissolved Zinc (ug/l)
Detection Limit		1	0.1	4	0.05	0.1	2	2	1	5	50
Commercial	n	20	20	20	20	20	20	20	20	20	20
	Min.	0.96	3.59	15.4	0.083	0	48	14	9	0	0
	Max.	9.3	89.3	122	0.66	0.283	514	170	66	22.7	450
	Median	3.195	16	28.1	0.186	0.076	113	70	35.5	7.7	145
	Mean	3.59	29.06	42.19	0.26	0.11	187	77.85	34.5	9.63	144.9
	Std. Dev.	2.22	27.39	30.50	0.17	0.08	153.10	45.05	16.02	6.75	113.52
Residential	n	23	23	23	23	23	23	23	23	23	23
	Min.	1.29	2.26	7.7	0.055	0	22	14	12	0	0
	Max.	12.5	34	127.2	0.845	0.505	302	531	200	26.2	127
	Median	3.08	5.19	15.9	0.31	0.052	46	74	42	6.97	0
	Mean	4.30	7.51	25.99	0.37	0.09	70.44	119.74	57.7	7.22	13.5
	Std. Dev.	3.32	7.43	27.43	0.22	0.12	64.64	122.68	47.17	7.78	32.59
Industrial	n	44	44	44	44	44	44	44	44	44	44
	Min.	0	1.9	5.3	0	0	28	16	6	0	0
	Max.	7.66	69.9	164	0.73	0.36	438	596	157	39.7	1128
	Median	2.205	7.945	20.15	0.201	0.062	74	129	45	9.665	301.5
	Mean	2.61	11.4	26.5	0.22	0.09	88.27	172.48	46.59	11.27	333.91
	Std. Dev.	1.57	11.6	24.58	0.17	0.08	69.21	109.52	27.74	9.47	240.94

Table 2.3. Result of the neural model on the test and training files.

Data set	Correct Classification (%)		RMS error	
	Test file	Training file	Test file	Training file
Data set A	92.3	95.0	0.171	0.160
Data set B	96.2	92.6	0.131	0.152
Data set C	88.5	95.9	0.169	0.149
Average	92.3	94.5	0.157	0.154

Table 2.4. Leave-one-out impact analyses based on all 10 input variables

Input variables	Impacts (%)			
	Data set A	Data set B	Data set C	Average
Potassium	7.6	31.8	13.9	17.8
Sulfate	-5.4	32.1	1.7	9.4
Alkalinity	7.1	20.7	22.7	16.8
Dissolved Phosphorus	5.4	24.9	7.9	12.7
Nitrite - N	9.6	12.8	9.0	10.5
Total Dissolved Solids	8.8	20.5	16.2	15.2
Suspended Solids	11.0	25.1	16.8	17.7
Volatile Suspended Solids	10.1	12.1	8.0	10.1
Dissolved Copper	24.5	33.8	31.2	29.8
Dissolved Zinc	11.2	29.9	22.4	21.2

Table 2.5. Water quality results from the highway monitoring

Input Variables	Storm date and station number and storm event number					
	2/9/2001	2/18/2001	2/23/2001	4/6/2001	4/6/2001	4/19/2001
	S2 - E4 [a]	S3 - E5	S3 - E6	S2 - E8	S3 - E8	S1 - E9
Potassium (mg/l)	NA [b]	NA	NA	NA	NA	NA
Sulfate (mg/l)	24.82	3.84	1.73	373.16	4.78	8.66
Alkalinity (mg/l)	17	9.5	11	112.1	10	28.5
Dissolved Phosphorus (mg/l)	0.243	0.161	0.061	0.126	0.152	0.154
Nitrite - N (mg/l)	0.162	0.119	0.048	0.103	0.075	0.151
Total Dissolved Solids (mg/l)	NA	NA	NA	NA	NA	NA
Suspended Solids (mg/l)	86.87	80.57	62.7	127.1	56.4	33.69
Volatile Suspended Solids (mg/l)	37.6	25.80	18.3	38.5	21.3	14.27
Dissolved Copper (ug/l)	20.55	6.98	5.4	25.8	16.2	27.35
Dissolved Zinc (ug/l)	90.5	54.85	42	100.8	107.4	151.99

[a.] S indicates station number and E indicates storm event number.

[b.] NA indicates sample was not analyzed.

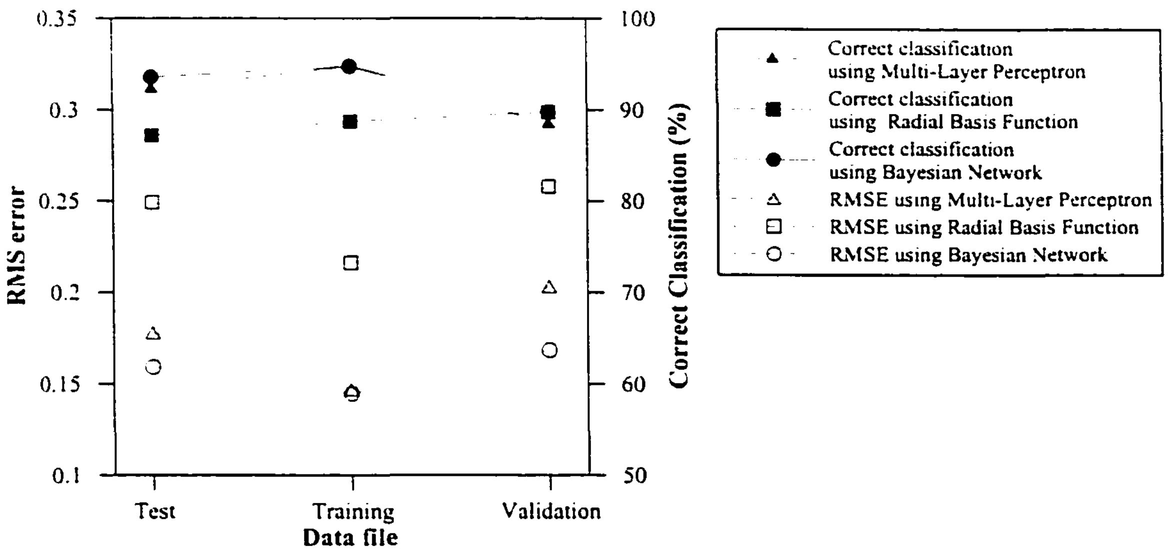


Fig. 2.1. The overall performance of the various neural networks in the preliminary tests. The overall performance measured average over each of the three data set A, B, and C.

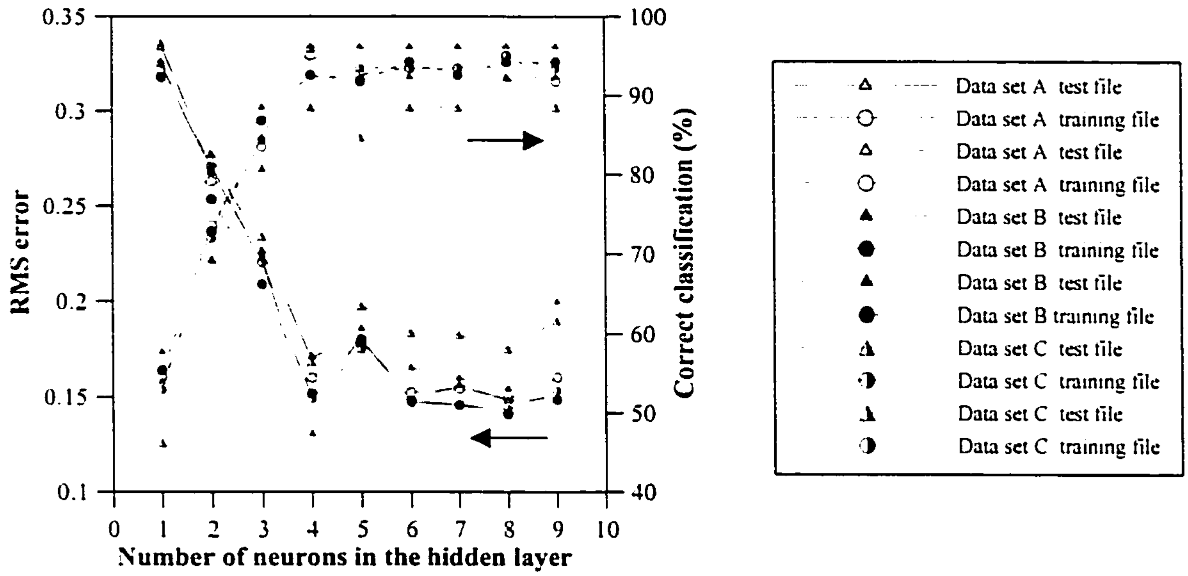


Fig. 2.2. The effect of neuron number in the hidden layer

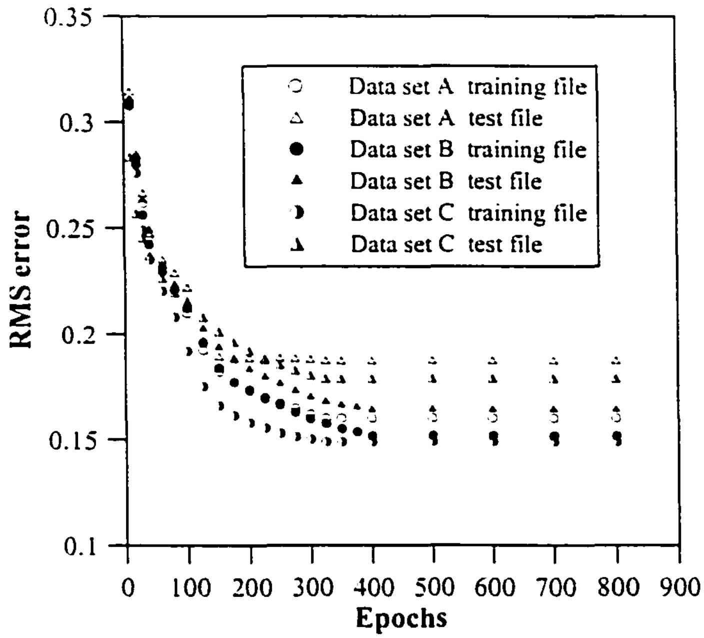


Fig. 2.3. RMS error on the training and validation file as a function of the number of epochs

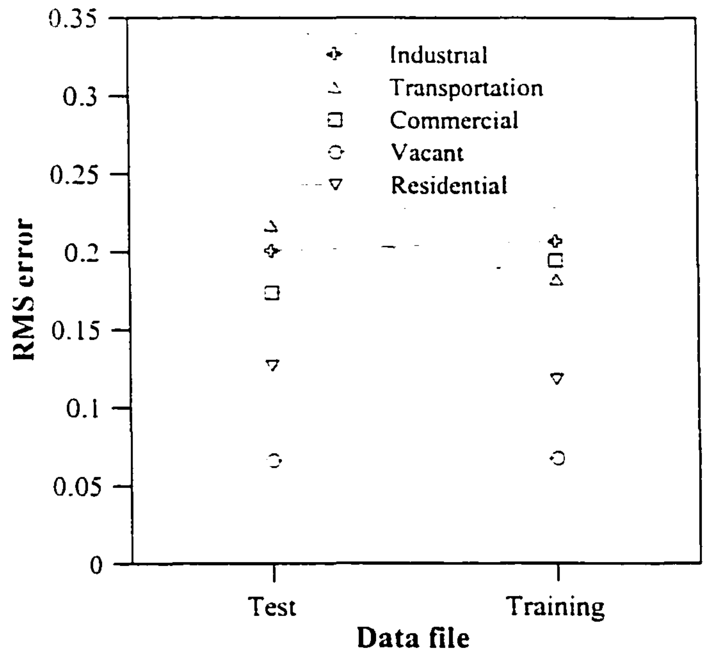


Fig. 2.4. The overall performance in RMSE on each target landuse type.

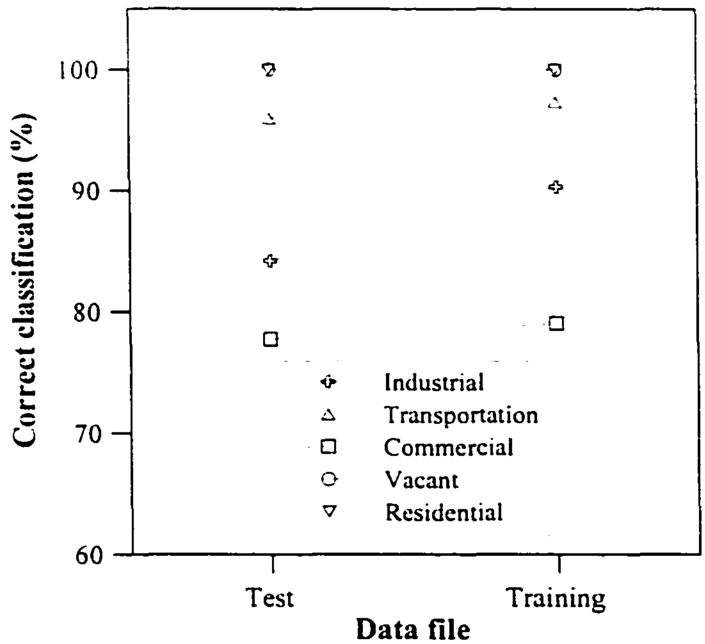
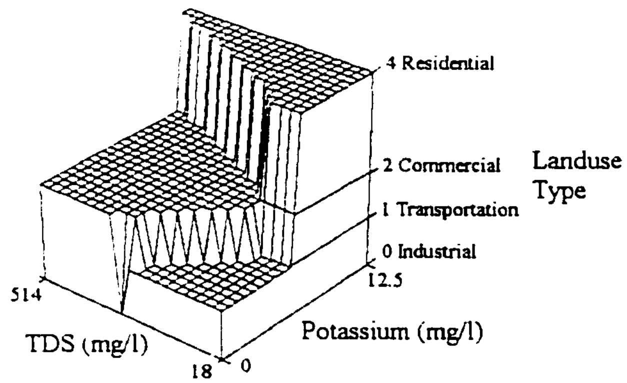


Fig. 2.5. The overall performance in correct classification on each target landuse type.



Landuse Type
4- Residential
3- Vacant
2- Commercial
1- Transportation
0- Industrial

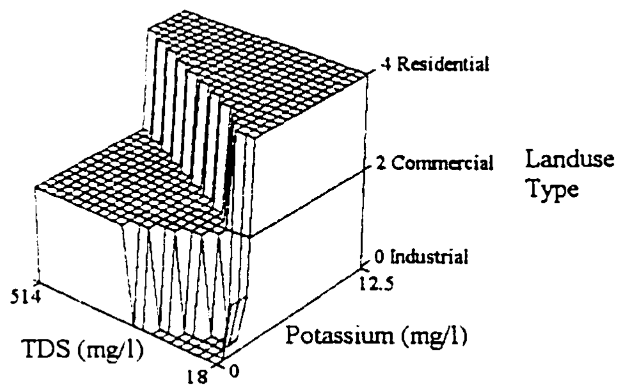


Fig. 2.6. Land Use classification as a function of Potassium and Total Dissolved Solids concentration for the case of S1-E9 (the top) and S3-E5 (the bottom), calculated by the neural model using data set B.

References

Basheer, I.A., Najjar, Y.M., and Hajmeer, M.N. (1996) Neuronet modeling of VOC adsorption by GAC. *Environmental Technology* **17**, 795-806.

Brion, G.M. and Lingireddy, S. (1999) A neural network approach to identifying non-point sources of microbial contamination. *Water Research* **33**(14), 3099 – 3106.

California State Water Resources Control Board, California Environmental Protection Agency (2000) *California Beach Closure Report 1999.*

Djebbar, Y. and Kadota, P.T. (1998) Estimating sanitary flows using neural networks. *Water Science and Technology* **38**(10), 215-222.

El-Din, A.G. and Smith, D.W. (2002) A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Research* **36**, 1115-1126.

Gob. S., Oliveros, E., Bossmann, S.H., Braun, A.M., Nascimento, C.A.O. and Guardani, R. (2001) Optimal experimental design and artificial neural networks applied to the photochemically enhanced Fenton reaction. *Water Science and Technology* **44**(5), 339-345.

Gong, N., Denoeux, T. and Bertrand-Krajewski, J-L. (1996) Neural networks for solid transport modeling in sewer systems during storm events. *Water Science and Technology* **33**(9), 85-92.

Kaluli, J.W., Madramootoo, C.A. and Djebbar, Y. (1998) Modeling nitrate leaching using neural networks. *Water Science and Technology* **38**(7), 127 – 134.

Laberge, C., Cluis, D. and Mercier, G. (2000) Metal bioleaching prediction in continuous processing of municipal sewage with *Thiobacillus Ferrooxidans* using neural networks. *Water Research* **34**(4), 1145-1156.

LACDPW (1997) Los Angeles County 1996-1997 Stormwater Monitoring Report. Los Angeles County Department of Public Works and Southern California Coastal water Research Project and URS Greiner Woodward Clyde.

LACDPW (1998) Los Angeles County 1997-1998 Stormwater Monitoring Report. Los Angeles County Department of Public Works and Southern California Coastal water Research Project and URS Greiner Woodward Clyde.

Lek, S., Guiresse, M. and Giraudel, J-L. (1999) Predicting stream nitrogen concentration from watershed features using neural networks. *Water Research* **33**(16), 3469-3478.

Loke, E., Warnaars, E.A., Jacobsen, P., Nelen, F., and do Ceu Almeida, M. (1997) Artificial neural networks as a tool in urban storm drainage. *Water Science and Technology* **36**(8-9), 101-109.

Mather, A.J. and Shaw, I.S. (1993) Alternative methods for the control of balancing tank at a wastewater treatment plant. *Water Science and Technology* **28**(11-12), 523-530.

Neelakantan, T.R., Brion, G.M. and Lingireddy, S. (2001) Neural network modeling of *Cryptosporidium* and *Giardia* concentrations in the Delaware River, USA. *Water Science and Technology* **43**(12),125-132.

SPSS (1997) Neural Connection 2.0 User's Guide. SPSS Inc. and Recognition Systems Inc.

Stenstrom, M.K., Lau, S-L., Lee, H-H., Ma, J-S., Ha, H., Kim, L-H., Khan, S. and Kayhanian, M. (2000) First Flush Stormwater runoff from Highways, *Research report for California Department of Transportation*, September, 2000.

Stenstrom, M.K., Lau, S-L., Ma, J-S., Ha, H., Lim, L-H., Lee, S., Khan, S. and Kayhanian, M. (2001), First Flush Stormwater runoff from Highways, *Research report for Department of California Transportation*, September, 2001.

Teodosiu, C., Pastravanu, O. and Macoveanu, M. (2000) Neural network models for ultrafiltration and backwashing. *Water Research* **34**(18), 4371-4380.

Walley, W.J. and Fontama, V.N. (1998) Neural network predictors of average score per taxon and number of families at unpolluted river sites in Great Britain. *Water Research* **32**(3), 613-622.

Wong, K., Strecker, E.W. and Stenstrom, M.K. (1997) A Geographic Information System to estimate stormwater pollutant mass loadings," *Journal of the Environmental Engineering*, ASCE, **123**(8), 737-745.

Zhao, H., Hao, O.J., McAvoy, T.J. and Chang, C-H. (1997) Modeling nutrient dynamics in sequencing batch reactor. *Journal of Environmental Engineering* **123**(4), 311-319.

Chapter 3.

Utility of stormwater monitoring

Abstract

Stormwater runoff is now a major contributor to the pollution of coastal waters in the United States. Public agencies are responding by requiring stormwater monitoring to satisfy the National Pollutant Discharge Elimination System (NPDES) Stormwater Permit. However, studies to understand the utility of the current programs or to improve their usefulness have not yet been performed. In this paper, we evaluate the landuse based program, the industrial stormwater permit program, and beach water quality monitoring in the County of Los Angeles to determine if the results will be helpful to planners and regulators in abating stormwater pollution. The utility of the program has been assessed based upon the programs' ability to accurately estimate the emissions for different classes of landuses. The land use program appears successful while the industrial monitoring program does not. Beach water quality monitoring suffers from a lack of real time monitoring techniques. We also provide suggested improvements such as sampling method and time, and parameter selection.

Introduction

California coastal waters are important recreational and economic resources, which make the safety of coastal waters of concern to both state and county health departments and beachgoers (Jiang et al., 2001). The completion of wastewater treatment plants mandated by the Clean Water Act has reduced conventional water pollution to California's beaches and bays. As a result, non-point source pollution such as stormwater runoff is now a major contributor to the pollution of the coastal water including Santa Monica Bay, which is among the most severely polluted Bays in the United States (Wong et al., 1997). Storm drains entering the ocean are a main cause of permanent beach postings at many California beaches (State Water Resources Control Board, 2001). The problem of stormwater pollution is becoming worse because of population growth, which results in increased impermeable area.

Public agencies are responding by requiring stormwater monitoring to satisfy the National Pollutant Discharge Elimination System (NPDES) Stormwater Permit as authorized by the Clean Water Act. For example, the Los Angeles County Department of Public Works (LACDPW) has been monitoring stormwater under the 1990 NPDES Municipal Permit (No. CA0061654) and later 1996 Municipal Permit (No. CAS614001) since the 1994-1995 wet seasons. Addition sampling is required by other agencies, such as the City of Los Angeles and the California Department of Transportation. Similar programs are underway in other areas of California and the United States.

The existence of stormwater monitoring programs should represent progress towards achieving clean water goals; however, studies have not yet been performed to understand the utility of the current programs or to improve their usefulness. In this paper we evaluate several monitoring programs to determine if the results will be helpful to planners and regulators in abating stormwater pollution. Datasets from a major municipal program, several research projects, beach monitoring, and a large self-monitoring program were used. The results suggest that parts of the current monitoring programs will not be helpful to regulators and planners, and we make proposals for improvement, along with projected cost increases

Background

The Los Angeles County Department of Public Works has been monitoring stormwater since early 1970s. In 1994 they began an improved program, which was designed to determine total pollutant emissions to Santa Monica Bay as well as determine landuse specific discharges (Stenstrom and Strecker, 1993). Total emissions are estimated from flow-weighted composite samples that are collected at five sampling stations (four stations are required under the 1996 Permit and one station remains from an earlier permit.). These stations are "mass emission" stations in that they were selected to sample the greatest runoff mass with the least number of stations. The stations are equipped with flow monitoring equipment and operate unattended in secure facilities. Samples from specific landuses are also required by the 1996 Municipal Permit and are collected with composite samplers at engineered sampling stations. A large suite of water quality parameters is measured, including indicator organisms, general minerals, nutrients, metals, semi-volatile organic compounds and pesticides.

Additional monitoring is being conducted by other agencies to satisfy regulations or for research. The California Department of Transportation (CalTrans) has a large monitoring program for their highways. Our laboratory has monitored three highway locations near UCLA (adjacent to the 101 and 405 freeways) since 1999 (Stenstrom et al., 2000; Stenstrom et al., 2001). The study is also sponsored by Caltrans and an extensive suite of parameters is measured, including indicator bacteria, general minerals, nutrients, metals, polycyclic aromatic hydrocarbons, and oil and grease.

The previous programs monitor discharges to the Bay, but there are also programs that monitor coastal waters. The California Assembly passed Bill 411(chapter 765 of Statutes of 1997) to address the problem of declining beach water quality and restore confidence in the healthful of beach swimming. Three types of indicators organisms are monitored and retesting in the event of an exceedence is also required. The more restrictive procedures by the bill have increased the frequency of beach postings and closures. The closure of Huntington Beach in Orange County, CA, during the summer of 1999, was the first example of beach closures caused by the new regulations (Orange County Sanitation District, 1999). Many organizations are monitoring the microbiological water quality of Southern California coastal waters (Noble et. al., 2000)

An example of a new monitoring activity is the Industrial Activities Stormwater General Permit (General Permit), which mandates all industrial stormwater permittees to analyze stormwater samples twice per year for at least four analytical parameters. The industries are classified by Standard Industrial Classification (SIC) code (www.swrcb.ca.gov/rwqcb4/html/programs/stormwater/sw_industrial.htmlreference). The monitored analytical parameters are pH, total suspended solids (TSS), specific conductance (SC), and total organic carbon (TOC). Oil and Grease (O&G) may be substituted for TOC. In addition, the permittees must monitor any other pollutants, which they believe to be present in their stormwater discharge as a result of industrial activity (www.swrcb.ca.gov/stormwtr/docs/induspmt.doc). Permittees in some cases may be required to sample at more than one location.

It is natural to ask if the monitoring programs are valuable. Is the resulting water quality database useful to planners and regulators to identify acute problems, improve long-term water quality, and understand landuse/water quality relationships? An improved understanding of the relationship of landuse to stormwater quality is an expected result since landuse specific sampling is required by the NPDES permit. The original purpose of the monitoring programs was to identify larger sources (e.g., "hot spots") as well as to create a database to help develop total mass daily loads (TMDLs) and other management tools. To answer this question, we reviewed the current industrial stormwater permit program. We also comment on other monitoring programs, and suggest improvements in sampling strategies and water quality parameter selection, with their anticipated cost increases.

Monitoring Program Utility

It is generally recognized that different human activities will create different types and varying concentrations of stormwater contaminants (Stenstrom et al. 1984). For example, runoff from transportation-associated landuse is a primary source of metals and hydrocarbons (LACDPW and Woodward-Clyde, 1998). Vehicles release hydrocarbons from leaks, engine byproducts and unburned fuel and various metals from corrosion, fuel combustion and wearing surfaces such as brake pads (Rogge et al. 1993; Sansalone and Buchberger, 1997). Differences in landuse patterns will likely result in different pollutant concentrations, and therefore landuse-related control strategies are essential to control storm water pollution effectively.

Landuse monitoring data

The landuse-based program administered by the LACDPW is a useful example. The landuse monitoring program required by the 1996 Municipal Permit was examined to determine if different landuses produce different stormwater quality. If the monitoring program is successful, landuses should be identifiable from the collected data. We developed a neural network approach to identify the various types of landuse (commercial, residential, industrial, transportation, and vacant) as a function of stormwater quality data (Ha and Stenstrom, 2002). The neural model uses a Bayesian network, and was trained using LACDPW data collected during 1996-2001 wet seasons. The model was successful at classifying 92 percent of the cases. Ten water quality parameters were used: potassium, sulfate, alkalinity, dissolved phosphorus, nitrite-N, total dissolved solids, volatile suspended solids, total suspended solids, dissolved copper, and dissolved zinc. The model was useful in that a data set could be manipulated by changing various water quality parameters, and the changes in classifications could be observed. It is also possible to determine which parameters are most sensitive for the classification, and which are most active in a particular case. The model will eventually be useful to automatically examine many datasets to identify abnormally high or low parameters for a particular landuse, and label these as opportunities for investigation or improvement.

Industrial stormwater monitoring

Based on this experience, a similar approach was applied to the industrial stormwater discharge data for the 1998-2001 wet seasons. This dataset contains approximately 14,000 cases. Neural networks were trained to differentiate between several industrial categories based on SIC code and water quality data. It was hoped that the trained model, would be help to identify industrial "hot sources" or outliers. Eight industrial categories were selected based their prevalence in Los Angels County, which means some SIC codes have many more cases than others. The selected eight industrial categories and each category's case number for the three years are shown in Table 3.1. The data cases that contain both the mandatory water quality parameters (pH, TSS, SC, TOC, and Oil and Grease) and metals are limited. Because of the reason, a neural model trained separately with the water quality data and metal data. Outliers in this study were defined as the upper two percent of the whole range of the data set for each parameter, and these cases were removed.

In this study, Neural Connection 2.1 (SPSS Inc. and Recognition Systems Inc, Chicago, IL) was used to build the neural models. Three supervised, feed-forward neural networks, namely, Multi-Layer Perceptron, Radial Basis Function, and Bayesian Network were used to differentiate the various types of industries. The neural models were extensively trained with various architectures; however, the performance of all models was very poor. This indicates a weak or almost no relationship between the industrial categories based on the SIC code and the available water quality data.

To further seek a relationship between water quality data and various landuses of industries, an unsupervised Kohonen neural network was used. The goal of Kohonen network is to map the spatial relationships among clusters of data points into hyperdimensional space (Aguilera et al., 2001). Once trained successfully, it may be used to identify unknown data patterns, and it was hoped that useful patterns between water quality and landuse would be identified.

A Kohonen neural model with two dimensions in the Kohonen layer was trained with different node sizes of 3×3, 5×5, and 7×7. The method performs square normalization, which normalizes the original input data patterns to zero mean and unit variance. The results were generally unsatisfactory and it was difficult to make a decision to cluster from the activation maps by the neural model. Fig. 3.1 shows an activation map having 3×3 neurons obtained by a Kohonen model that was trained with four parameters: pH, TSS, SC and O&G. The shading intensity indicates the degree of similarity to their neighbor nodes; lighter shades indicate similar characteristics as the neighboring node, and darker patterns indicate greater differences.

There are two possible clusters. The first contains nodes 1, 4, 5, and 7, and the second contains nodes 2, 3 and 6. Figure 3.2 shows the number of cases assigned to the various nodes. Nodes 4 and 5 contain most of the cases, and the majority (82%) would be assigned to the first cluster. A classification system that assigns such a large fraction to a single cluster is not useful; basically, the classification system is saying that it can find no difference in the available water quality parameters among the majority of the SIC codes. Nodes 4 and 5 tend to have the lowest pollutant concentrations, but the members are not distinguished by SIC codes. Similar results were obtained using 5×5 or 7×7 neurons and with different set of input parameters. The conclusion from this analysis is that stormwater quality is not distinguishable by SIC code using the current water quality parameters.

To further investigate possible relationships, the water quality data were transformed into a three member fuzzy set with categories of low, medium and high. Each of the resulting data sets except for pH was examined using a Kohonen neural model. Each model had three nodes in the one dimensional Kohonen layer and was trained for each parameter separately. The output result of the model was assigned a specific node number, from 1 to3, for every case. The nodes were reordered so that higher node numbers always indicated greater pollutant concentration, with the number 3 representing the highest pollutant concentration for each parameter. For the case of pH, a qualitative score was assigned manually, based upon deviation from neutrality. An overall qualitative score was created by summing the fuzzy states. Figure 3.3 shows the overall process. For example, when we used four parameters, the possible minimum and maximum overall qualitative score is 4 and 12 respectively, with 12 representing the worst water quality.

Figure 3.4 shows the distribution of the overall qualitative score for the various types of industries with different sets of input parameters. The upper figure used all five water quality parameters (pH, TSS, SC, TOC, and O&G). The middle part shows the classifications when TOC is left out. The bottom shows the classification using the metal analysis. In general, no distinguishing differences were found among industrial categories. When four or five water quality parameters were used (Fig. 3.4, top and middle), food and kindred product facilities have the least abundance of low scores (4 or 5), suggesting that it is the land use with the worst stormwater quality. The wholesale trade-durable goods category has least abundance of small scores if scores up to 6 are considered. For metals (lower portion of Figure 3.4), primary metal facilities have least abundance of low scores (3 and 4). This suggests that this industry has the worst stormwater quality with respect to metals. The statistical significance of these findings has not been evaluated and it all likelihood, a new method would need to be developed or an existing method adapted.

The industrial data set was also examined to determine if a seasonal first flush could be identified. Los Angeles has two distinct rainfall seasons. The late spring to late fall or early winter is usually dry. Most rainfall occurs in winter and early spring. This rainfall pattern creates a long period for pollutant build-up and the first storm of the season usually has abnormally high pollutant concentrations, which is called a seasonal first flush. The industrial permit requires the first storm to be sampled and one later storm to be sampled, which was required in order to identify the seasonal first flush.

To determine if the industrial stormwater monitoring program was successful in identifying the seasonal first flush, the data (for 2000-2001 season only) were divided into first and second sample datasets. In some cases the first sample does not represent the first rainfall event. In cases when there were more than two samples collected, the later samples were ignored. Cases with only one sample were also ignored. The comparison of the first to second sample for the 2000-2001 season are shown in Figure 3.5 using notched bar plots and in Table 3.2. Concentration for all parameters were higher in the first sample than the second sample by 0 to120 % for the median and 20 to 85 % for the mean. TOC showed the greatest difference between first and second samples; oil and grease showed the smallest difference. Statistically significant differences can be observed in the notched bar plot.

Beach monitoring

Assembly Bill 411 created improved beach water quality monitoring requirements. The improved monitoring was mandated after an epidemiological study of Santa Monica Bay swimmers suggested increased health risk associated with swimming near storm drains (Haile, et al. 1999). Daily samples for total coliforms, fecal coliforms and enterococcus were mandated with new, lower levels that trigger a beach posting or closure. Leecaster and Weisberg (2001) examined sampling data from 24 sites in Los Angeles County between 1995 and 1999. They report that over 70% of the water quality exceedences were for only 1 day.

The time required to analyze indicator organism data is generally more than 24 hours. This created a chronology as follows: day 1 – a sample is collected and analysis begins; day 2 – the sample is analyzed and an exceedence is noted, with a beach posting and a new sample is collected and analysis begins; day 3 – the second sample result is negative 70% of the time, and the beach posting is removed. The chronology creates a situation that beaches are posted when the samples do not exceed standards and open when they do. Clearly the problem is a monitoring program that cannot be implemented with current technology. Rapid indicators are needed. Furthermore the utility of conventional indicator organisms for fecal contamination in beach waters is in question.

Discussion and Recommendations

Three stormwater monitoring programs were discussed. The landuse monitoring program was generally successful and showed the anticipated differences in water quality based upon land use. This program used automatic, flow-weighted composite samplers with trained personnel. The second program, the industrial monitoring program, used grab samples collected at various times for two or more storms with SIC codes as landuse or industrial-use descriptors. The program is generally unsuccessful in identifying relationships between water quality and landuse. It was successful in showing a seasonal first flush, and its utility for identifying acute problems is questionable, based upon outliers, to be discussed later. The third is the beach water quality monitoring program, which uses grab samples and analyses that are not real-time. This creates problems of beach postings, which are out of phase with exceedences.

In this section, we discuss possible reasons for less successful program and suggest ways to improve monitoring. Some suggestions will require new technology.

Sampling method

For the industrial stormwater permit monitoring, grab samples are allowed, and facility operators are instructed to collect the sample during the first hour of discharge from the first event for the wet season (October to May), and at least one other storm event in the wet season. A grab sample is a discrete sample taken within a short period of time, usually less than 15 minutes. Flow-weighted composite samples were collected in the landuse program, which requires instrumentation and perhaps site preparation to create a channel for flow measurement and security for equipment. The flow-weighted samplers collect a composite by combining a series of discrete samples of specific volume, collected at specific flow- weighted intervals over the duration of a storm event (LACDPW and Woodward-Clyde, 1998).

It is useful to compare the results of the two programs. They are analogous in that both programs attempt to measure the emissions from a particular human activity, although the industrial program also attempts to identify high dischargers. The results of the 1998 – 2001 industrial permit using grab samples were compared to the 1996- 2001 landuse monitoring that used flow-weighted composites. Figure 3.6 is a notched bar plot that shows the differences.

There are a large number of outliers among the grab samples and almost no outliers among the composite samples. The number of outliers suggests the need for a quality assurance program, and is helpful in understanding why the neural networks could not identify significant differences in stormwater from SIC codes.

The standard deviations of the concentrations are much lower among the composite samples (Table 3.3). For example, the standard deviation for TOC is 174 for grab samples and 9.7 for composite samples, or a ratio of 18. The other parameters have ratios of standard deviations from 2.3 (pH) to 66 (zinc). With this large range of differences, one has to question to the utility of such a monitoring program for any purpose. The application of any normalization method of the original data is not useful to generalize for use with a neural model. In addition, there are too many upper and lower outliers in the data set, which results in excessive clipping.

A flow-weighted composite sample for a storm event is generally better represents of the storm event than a single grab sample that may be biased due to the collection time. The result is an event mean concentration, which can be multiplied by the flow rate to calculate overall mass emissions. This is useful for spreadsheet load models (Wong, et al., 1997), which are finding wide spread use for planners and TMDL development.

A grab sample suffers from a variety of errors and biases, but one that has not been fully explored is the effect of first flush. Many parameters exhibit a first flush, which is typified by a declining concentration from storm beginning to storm end (Ma, et al. 2002a). When the grab sample is collected early in the storm, it will be higher than the EMC; conversely, if collected too late, it will be lower than the EMC. The industrial monitoring program suggests collecting a sample within the first hour. The best time for sampling oil and grease from highway landuse is between 2 and 3 hours and is related to cumulative rainfall and duration (Ma, et al., 2002b). There might be some improvement in the existing program with better definition of collection times.

It is almost universally recognized that composite samplers are better for stormwater monitoring; however, to collect a flow-weighted composite sample, an automatic sampler must be installed and operated properly before a storm event. It would be a burden to all industrial permittees to construct composite sampling facilities. Additionally, several water quality parameters such as oil and grease and indicator bacteria are not easily measured by a composite sample.

To improve sampling, it might be reasonable to randomly select a small subset of industrial users for composite sampling. This might be funded by fee permittees or by allowing a reduced number of grab samples to be collected. A trained team would also increase quality assurance to eliminate outlier. Such an approach might be a better and/or less expensive method of determining stormwater emissions on receiving waters.

Parameter selection

A variety of metal-related industries are included among the SIC codes in the industrial monitoring program. Many industries should be sources of metals such as chromium, copper, lead, nickel and zinc (Woodward Clyde, 1992). Fig. 3.7 shows the mean concentration of the basic analytical parameters and metals as a function of their industrial categories for 1998-2001 seasons. Outliers defined as before have been removed. The numbers of cases for all parameters vary with as many as 800 for conventional parameters and only about 80 for metals. The conventional water quality parameters show much less relation to industrial category than metals. The mean concentrations of lead, zinc, and nickel were highest for the primary metal industries category, and copper was highest at the transportation equipment facilities category. Mean concentrations of oil and grease and total organic carbon were highest for whole trade-durable good industries and mean concentration of conductivity and suspended solids were highest for Electric/Gas/Sanitary service facilities.

The addition of metals to the basic permit's requirement for basic water quality parameters would be a useful way of adding information to the dataset. A neural model trained with both metals and basic parameters will perform better than that trained existing water quality parameters or metals alone. The addition of metals will increase monitoring cost. Table 3.4 shows the current costs for laboratory analysis. The addition of metals to the permit will approximately double or triple the laboratory costs. The cost of collecting the samples should be quite similar. Cost increases are probably inevitable, but this approach may be less expensive that other approaches.

Conclusions

This paper has examined three stormwater monitoring programs. The utility of the programs have been assessed based upon the programs' ability to accurately estimate the emissions for different classes of landuses, as well as other obvious benefits. The following conclusions are made:

1. Data collected by grab samples had much higher variability than composite samplers. The coefficients of variation (standard deviation divided by the mean) for the same parameters were generally 2 to 9 times higher for the grab samples.

The variability suggests that composite samples should be collected, even if it means a reduction in the total number of samples or facilities that can be monitored.

2. The time required to analyze a sample must be commensurate with the intended use of the results. Beach water quality monitoring suffers from analysis time for indicator organisms. The data suggests that 70% of the beach postings are out of phase with the water quality parameter exceedence.
3. Metals (zinc, copper, lead, nickel) are potentially more useful to distinguish landuse patterns. Adding them to existing permits might double or triple the cost, but will add value to the resulting monitoring database.

Managing stormwater is a developing technology and much remains to be done. This paper has shown that even with the limited experience we have so far, that there are improvements that can be made.

Table 3.1. The selected eight major industries and its case number according to the different sets of input parameters after clipping outliers for the 1998-2001 seasons

Major industries	SIC code	Input parameters: pH, TSS, SC, TOC, and O&G	Input parameters: pH, TSS, SC, and O&G	Input parameters: Lead, copper, and Zinc
Food and kindred products (FKP)	20	184	472	10
Chemical and allied products (CAP)	28	305	850	35
Primary metal industries (PMI)	33	144	773	100
Fabricated metal products, except machinery and transportation equipment (FMP)	34	417	1325	155
Transportation equipment (TE)	37	193	601	187
Motor freight transportation and warehousing (MFTW)	42	263	731	76
Electric, gas, and sanitary services (EGSS)	49	182	505	198
Wholesale trade-durable goods (WT)	50	120	723	471
Number of the total cases		1808	5980	1232

Table 3.2. Comparison of first to second sample for the 2000/2001 wet season

Water quality parameters		First sample	Second sample
Total Suspended Solids (mg/l)	Number	1058	1035
	Minimum	0.464	0
	Maximum	25160	7860
	Median	69.5	42
	Mean	236.93	145.39
	Standard Dev.	1129	406
Specific Conductance (umhos/cm)	Number	1057	1052
	Minimum	0.017	0.018
	Maximum	15000	32200
	Median	160	99
	Mean	431.31	310.54
	Standard Dev.	1017	1226
Oil and Grease (mg/l)	Number	823	820
	Minimum	0	0
	Maximum	1000	650
	Median	5	5
	Mean	14.38	10.365
	Standard Dev.	60	29.013
Total Organic Carbon (mg/l)	Number	443	465
	Minimum	0.05	0.05
	Maximum	3150	1890
	Median	28.1	13
	Mean	83.98	39.81
	Standard Dev.	230	119.0
Lead (mg/l)	Number	211	208
	Minimum	0	0.005
	Maximum	90	50
	Median	0.08	0.05
	Mean	1.21	0.656
	Standard Dev.	8.4	4.903

Table 3.2. Continued.

Water quality parameters		First sample	Second sample
Copper (mg/l)	Number	209	208
	Minimum	0.009	0.006
	Maximum	6.24	21.9
	Median	0.11	0.067
	Mean	0.414	0.351
	Std. Dev.	0.91	1.73
Zinc (mg/l)	Number	349	341
	Minimum	0.001	0.01
	Maximum	56	28.2
	Median	0.66	0.407
	Mean	2.09	1.12
	Standard Dev.	5.3	2.7
Nickel (mg/l)	Number	99	96
	Minimum	0.006	0.004
	Maximum	4.63	15.1
	Median	0.06	0.042000
	Mean	0.226	0.2900
	Standard Dev.	0.554	1.546

Table 3.3. Comparison of grab to composite sample (0 indicates level below detection limit)

Water quality parameters		Grab Industrial stormwater permit sample 1998-2001	Flow-weighted composite Landuse monitoring sample 1996-2001 (Industrial site alone)
pH	Number	8584	51
	Minimum	0.1	6.04
	Maximum	12.7	8.32
	Median	6.88	6.82
	Mean	6.91	6.83
	Standard Dev.	0.96	0.41
Total Suspended Solids (mg/l)	Number	8424	49
	Minimum	0	16
	Maximum	101000	1865
	Median	48	140
	Mean	219.11	232.55
	Standard Dev.	1693	298
Specific Conductance (umhos/cm)	Number	8297	47
	Minimum	0.017	48.9
	Maximum	71000	691
	Median	121	126
	Mean	365.17	150.06
	Standard Dev.	1555	111
Oil and Grease (mg/l)	Number	6685	not analyzed
	Minimum	0	
	Maximum	6640	
	Median	5	
	Mean	13.63	
	Standard Dev.	95	
Total Organic Carbon (mg/l)	Number	3404	50
	Minimum	0	2.4
	Maximum	3700	45.62
	Median	18	9.85
	Mean	56.01	12.67
	Standard Dev.	174	9.7
Lead (mg/l)	Number	171	low detection frequency
	Minimum	0	
	Maximum	90	
	Median	0.06	
	Mean	0.402	
	Standard Dev.	3.5	

Table 3.3. Continued.

Water quality parameters		Grab Industrial stormwater permit sample 1998-2001	Flow-weighted composite Landuse monitoring sample 1996-2001 (Industrial site alone)
Copper (mg/l)	Number	1917	54
	Minimum	0	0.0053
	Maximum	49.5	0.99
	Median	0.084	0.0185
	Mean	0.337	0.047
	Std. Dev.	1.6	0.13
Zinc (mg/l)	Number	2917	54
	Minimum	0	0.079
	Maximum	2200	5.97
	Median	0.6	0.36
	Mean	4.86	0.63
	Standard Dev.	64.4	0.97
Nickel (mg/l)	Number	803	54
	Minimum	0	0
	Maximum	15.1	0.0804
	Median	0.05	0.005995
	Mean	0.196	0.0082
	Standard Dev.	0.76	0.013

Table 3.4. Comparison of the current requirement parameters to the complementary parameters on the costs of laboratory analysis (data source: Los Angeles County Department of Agricultural Commissioner/Weights and Measures Environmental Toxicology Laboratory)

Water quality parameters	Cost per sample (S)	Current requirement parameters		Complementary parameters	
		pH, TSS, SC, and TOC (1)	pH, TSS, SC, and O&G (2)	pH, TSS, Pb, Cu, Zn, and Ni (3)	pH, TSS, SC, O&G, Pb, Cu, Zn, and Ni (4)
pH	3.5	3.5	3.5	3.5	3.5
Total Suspended Solids (TSS)	7.68	7.68	7.68	7.68	7.68
Specific Conductance (SC)	6.4	6.4	6.4		6.4
Total Organic Carbon (TOC)	23.46	23.46			23.46
Oil and Grease (O&G)	36.25		36.25		36.25
Lead (Pb)	20.29			20.29	20.29
Copper (Cu)	20.29			20.29	20.29
Zinc (Zn)	20.29			20.29	20.29
Nickel (Ni)	20.29			20.29	20.29
Total cost per sample (S)		41.04	53.83	92.34	158.45

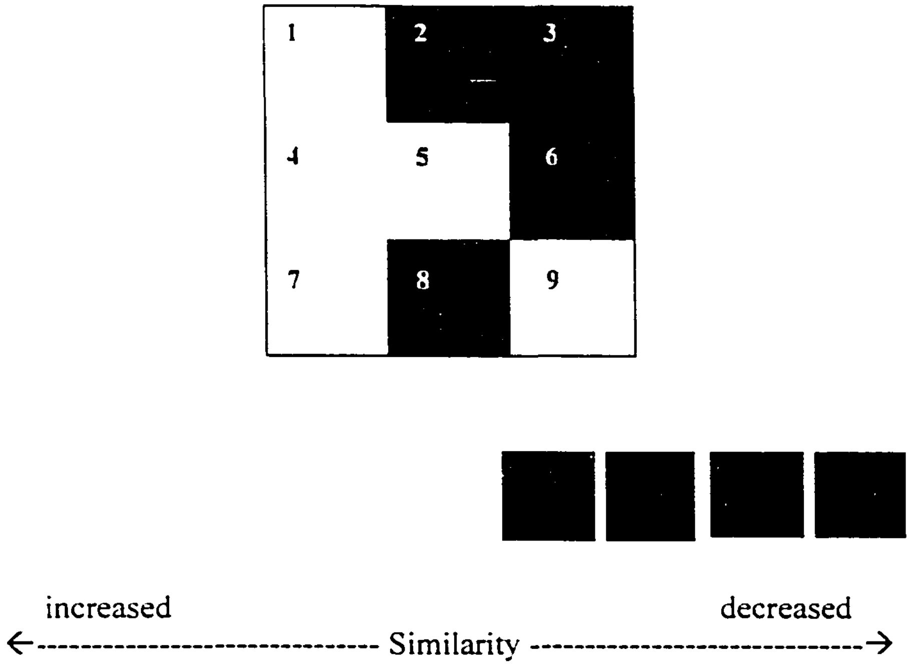


Fig. 3.1. Activation map having 3×3 neurons obtained by a Kohonen neural model that was trained with four input parameters (pH, TSS, SC and O&G). The shading intensity indicates the degree of similarity to their neighbor nodes. Numbers indicate node in the Kohonen layer.

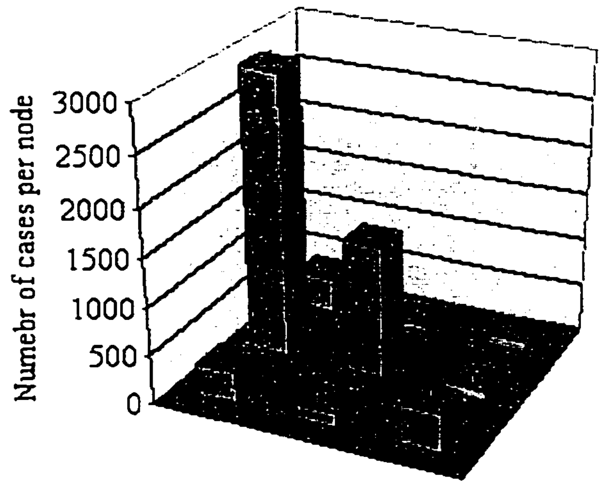


Fig. 3.2. Number of cases per node obtained by the Kohonen neural model that explained in Figure 3.1.

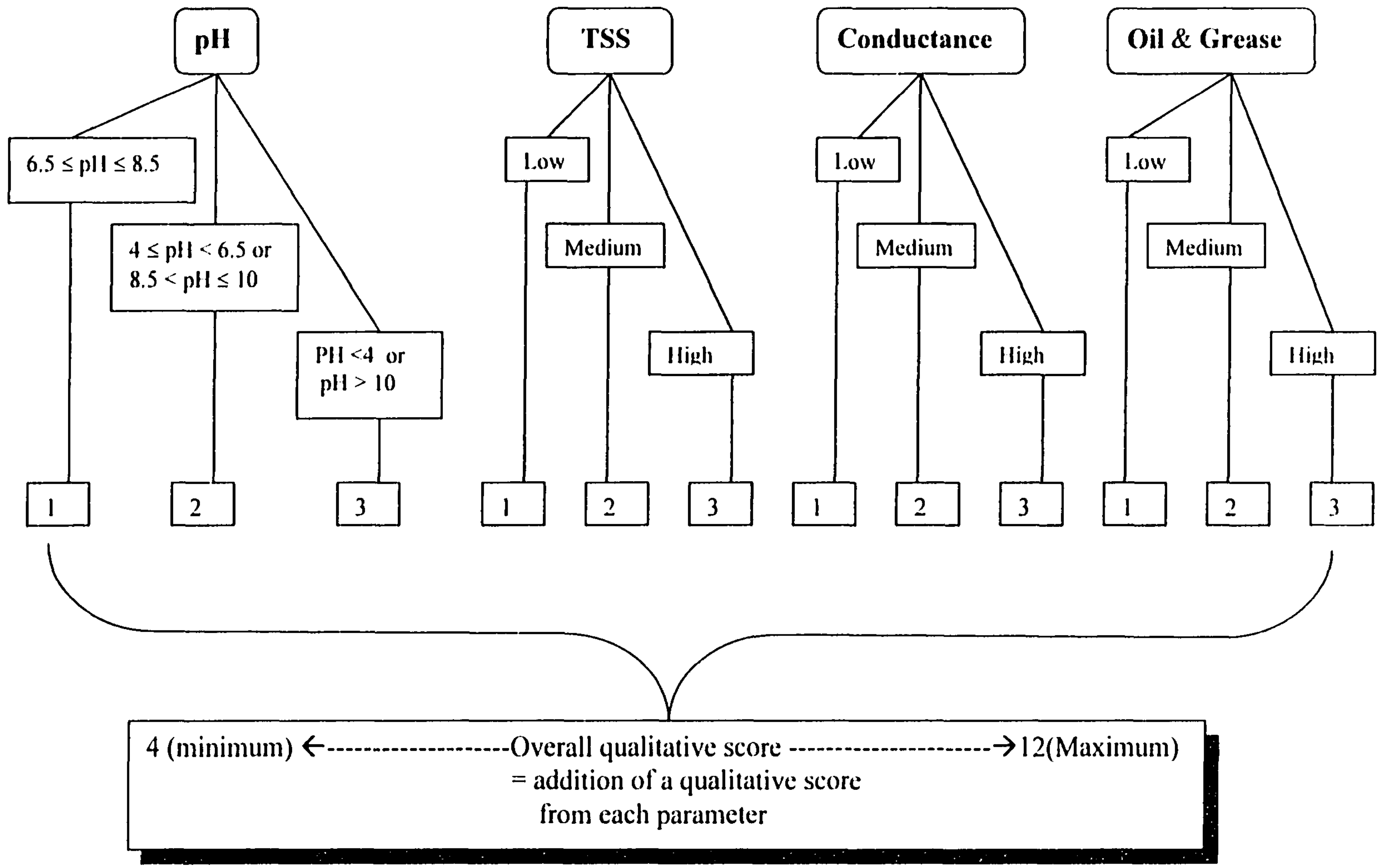


Fig. 3.3. Overall process of producing an overall qualitative score with four parameters (Shade area: A Kohonen network having three nodes in the Kohonen layer was trained for each parameter)

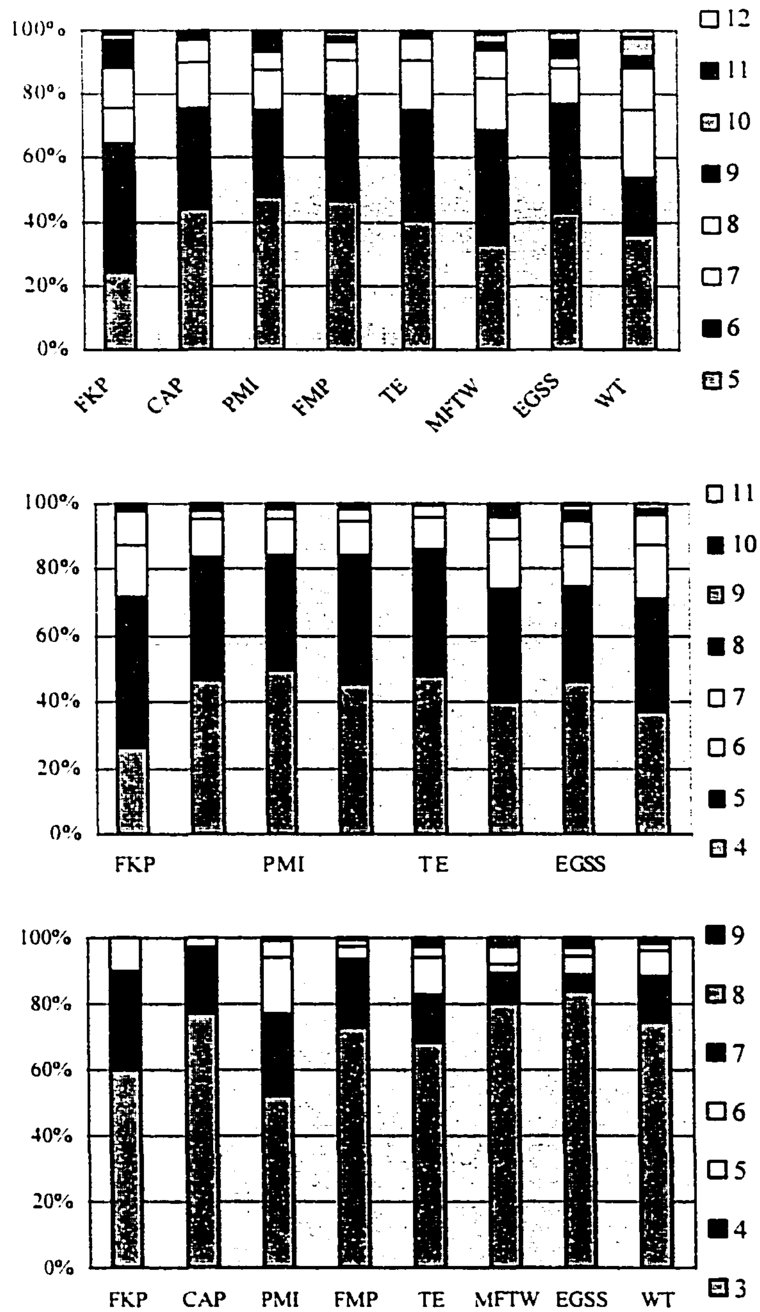


Fig. 3.4. Distribution of the overall qualitative score for each industrial category. (Top: five parameters, pH, TSS, SC, TOC, and O & G, were used; Middle: four parameters, pH, TSS, SC and O & G were used; Bottom: three parameters, Pb, Cu, and Zn, were used.) Number of cases per category with different sets of parameters was shown in Table 3.1.

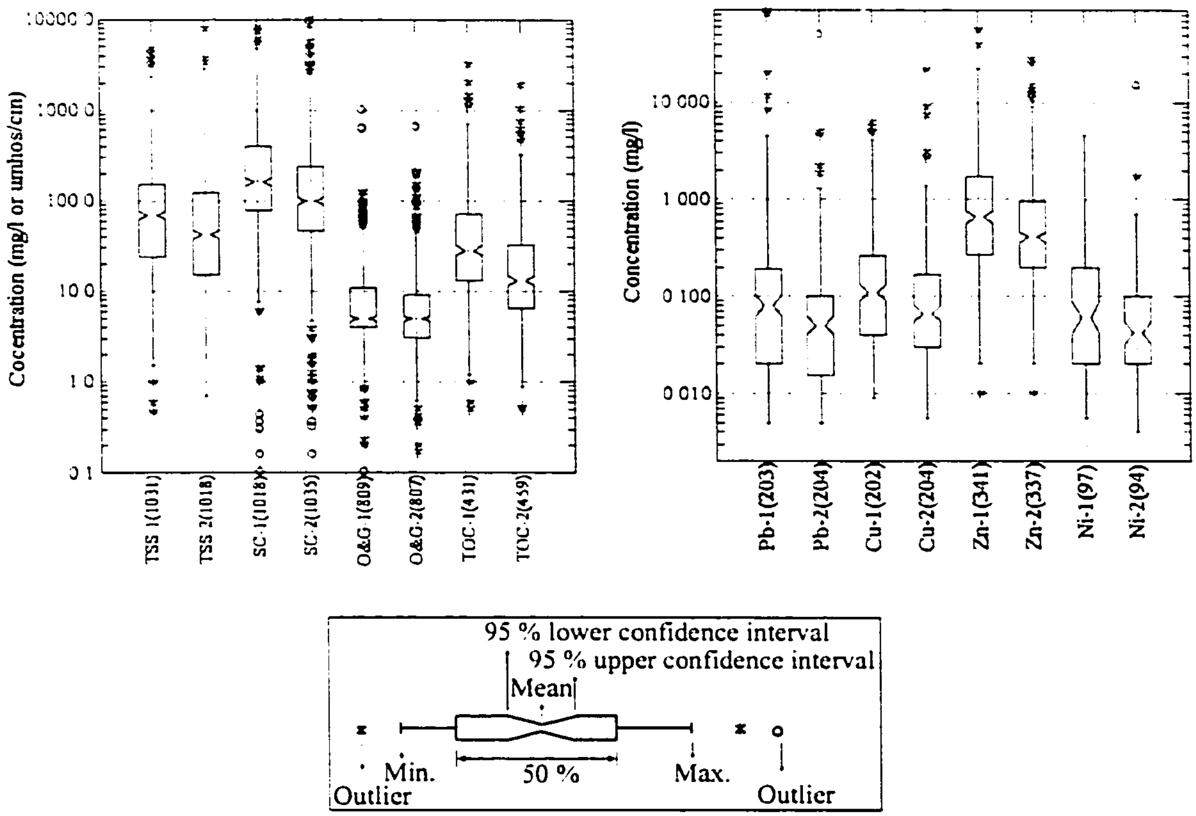


Fig.3.5 Comparison of first to second sample for the 2000-2001 wet season. All outliers are now shown. 1 in the x-axis indicates the first sample and 2 indicates the second sample. Number in a parenthesis in x-axis indicates number of cases.

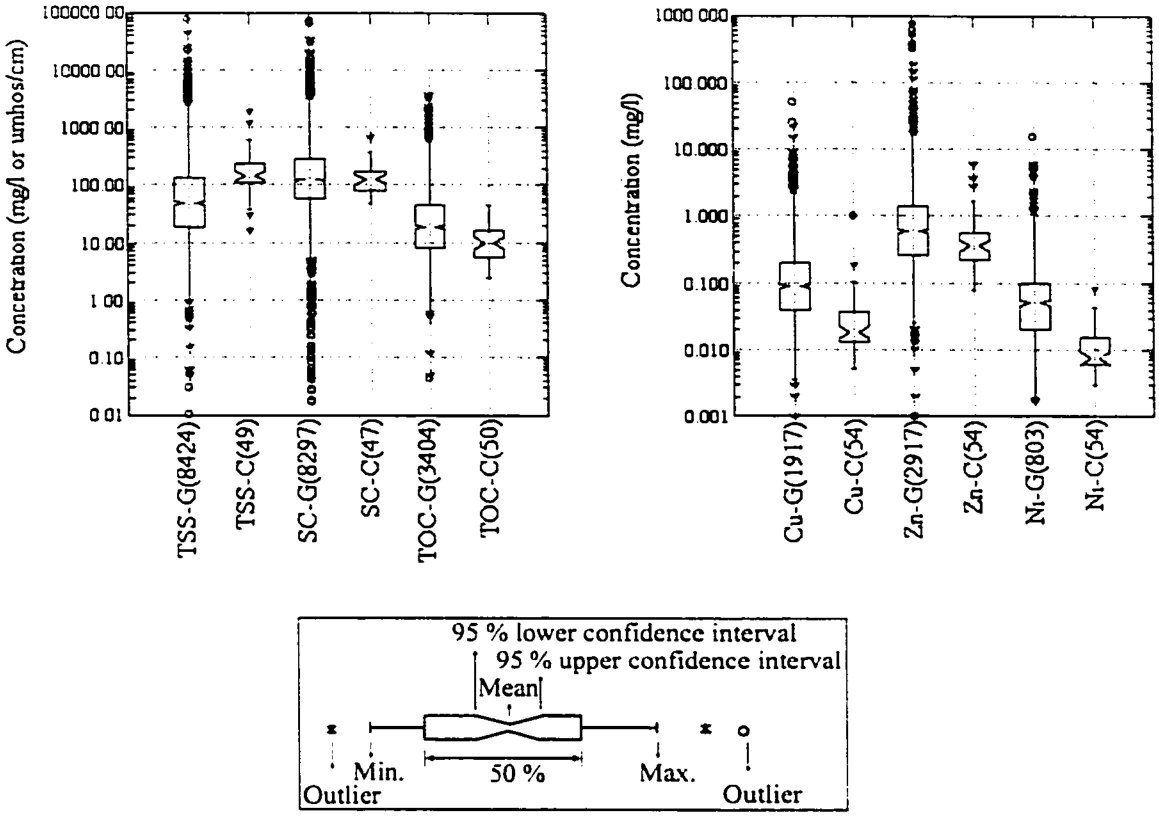


Fig.3.6 Comparison of grab sample from the industrial stormwater discharge data for 1998-2001 to flow-weighted composite sample from the landuse monitoring data (industrial landuse alone) for 1996-2001.G in x-axis indicates a grab sample and C indicates composite sample. Number in a parenthesis in x-axis indicates number of cases.

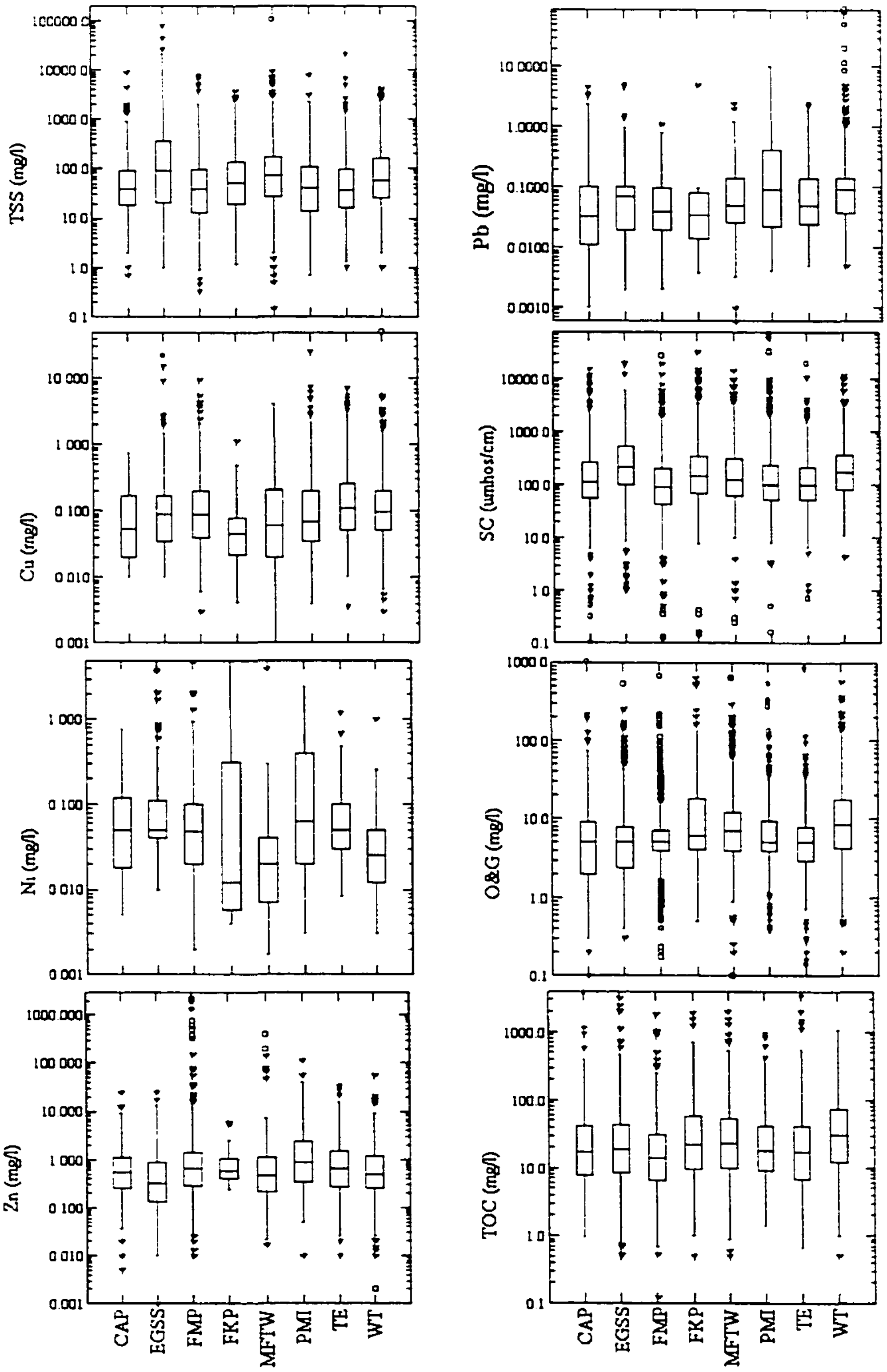


Fig. 3.7 Distribution of the analytical parameters and metals for each industrial category. The number of cases is varied.

References

Aguilera, P.A., Garrido Frenich, A., Torres, J.A., Castro, H., Martinez Vidal, J.L., and Canton, M. (2001) Application of the Kohonen neural networks in coastal water management: Methodological development for the assessment and prediction of water quality. *Water Research*, **35**(17), 4053-4062.

Ha, H. and Stenstrom, M.K. (2002) Identification of land use with water quality data in stormwater. Submitted to the *Water Research.*

Haile, R. W., Witte, J.S., Gold, M., Cressey, R., Mcgee, C., Millikan, R.C., Glasser, A., Harawa, N., Ervin, C., Harmon, P., Harper, J., Dermand, J., Alamillo, J., Barrett, K., Nides, M., and Wang, G. (1999). The health effects of swimming in ocean water contaminated by storm drain runoff. *Epidemiology* **10**(4), 355-363.

Jiang, S., Noble, R., and Chu, W. (2001) Human adenoviruses and coliphages in urban runoff impacted coastal waters of Southern California. *Applied and Environmental Microbiology*, **67**(1), 179-184.

LACDPW and Woodward-Clyde (1998) Los Angeles County 1997-1998 Stormwater Monitoring Report. Prepared by Los Angeles County Department of Public Works, Alhambra, CA, Woodward-Clyde, Oakland, CA and Southern California Coastal water Research Project, Westminster, CA, July 1998.

Leecaster, M and Weisberg, S. (2001) Effect of Temporal Sample Frequency on Shoreline Microbiology Assessments. SCCWRP Annual Report, 1999-2000, Westminster, CA.

Ma, J-S., Khan, S., Li, Y-X., Kim, L-H., Ha, S., Lau, S-L., Kayhanian, M., and Stenstrom, M.K. (2002a) First flush phenomena for highways: How it can be meaningfully defined. To be presented at the 9th ICUD, Portland, Oregon, September 2002.

Ma, J-S., Khan, S., Li, Y-X., Kim, L-H., Ha, H., Lau, S-L., Kayhanian, M., and Stenstrom, M.K. (2002b) Implication of oil and Grease measurement in stormwater management systems. To be presented at the 9th ICUD, Portland, Oregon, September 2002.

Noble, R.T., Dorsey, J.H., Leecaster, M., Orozco-Borbon, V., Reid, D., Schiff K. and Weisberg, S.B. (2000) A regional survey of the microbiological water quality along the shoreline of the Southern California Bight. *Environmental Monitoring and Assessment*, **64**, 435-447.

Orange County Sanitation District (1999). Huntington Beach Closure Investigation-Final Report. Prepared by Orange County Sanitation District, Fountain Valley, CA, December 1999.

Rogge, W.F., Hildemann, L.M., Mazurek, M.A., and Cass, G.R. (1993) Sources of fine organic aerosol. 3. Road dust, tire debris, and organometallic break lining dust: Roads as sources and sinks. *Environmental Science and Technology*, **27**(9), 1892-1904.

Sansalone, J.J. and Buchberger, S.G. (1997) Partitioning and first flush of metals in urban roadway storm water. *Journal of Environmental Engineering*, **123**(2), 134-143.

State Water Resources Control Board, (2001) California Beach Closure Report 2000. Prepared by the State Water Resources Control Board, California Environmental Protection Agency, July 2001.

Stenstrom, M.K and Strecker, E. (1993) Assessment of storm drain sources of contaminants to Santa Monica Bay, Vol. III, Surface Drainage Water Quality Monitoring Program Plan. UCLA-ENG-93-64, May 1993, Vol. III, 1-59.

Stenstrom, M. K., Silverman, G.S., and Bursztynsky, T.A. (1984) Oil and Grease in Urban Stormwaters, *Journal of the Environmental Engineering Division*. ASCE, **110**(1), 58-72.

Stenstrom, M.K., Lau, S-L., Lee, H-H., Ma, J-S., Ha, H., Kim, L-H., Khan, S. and Kayhanian, M. (2000) First Flush Stormwater runoff from Highways. *Research report for California Department of Transportation*, September 2000.

Stenstrom, M.K., Lau, S-L., Ma, J-S., Ha, H., Lim, L-H., Lee, S., Khan, S. and Kayhanian, M. (2001), First Flush Stormwater runoff from Highways. *Research report for Department of California Transportation*, September 2001.

Wong, K., Strecker, E.W. and Stenstrom, M.K. (1997) A Geographic Information System to estimate stormwater pollutant mass loadings. *Journal of the Environmental Engineering*, ASCE, **123**(8), 737-745.

Woodward Clyde (1992) Source identification and control report. Prepared for Santa Clara Valley Urban Pollution Program, December 1992.

Chapter 4.

Methods to Identify Human and Animal Fecal pollution in Water: A Review

Abstract

In this chapter methods to detect human fecal pollution and differentiate it from other sources such as animals are reviewed. This review includes microbial methods, especially those using molecular biology, and chemical methods. The conclusion is that no single method can provide definitive answers, as least not with our current understanding or experience. Additional testing with some of the reviewed methods may provide the required experience and confidence. It is much more likely that a combination of methods can be used to accurately identify human fecal pollution. Unfortunately, a combination of procedures will be more expensive and most likely be no faster than existing techniques.

Introduction

Fecal bacterial contamination from human and animal waste is a major cause of deteriorating water quality in receiving waters and has direct economic impacts to coastal communities through the loss of shellfisheries and restrictions of recreational uses. The possible sources of fecal contamination are point sources, such as industrial and municipal effluents, or nonpoint sources, such as surface runoff, direct animal and human input, failing or inadequate septic systems, and sewer overflows. In recent years nonpoint pollution has surpassed point sources as the major source of Fecal contamination to surface water. Management of this problem depends on knowing which sources of Fecal matter are the cause. A method that could distinguish sources would be the first step to solving this problem.

Methods for distinguishing between human and animal fecal pollution are necessary for assessing the overall protection of water supplies and implementing effective remediation for epidemiological studies, and even for legal purposes when it is necessary to determine the source of environmental contamination. Animal Fecal pollution is not without risks, and many of the risks are unknown, but it is generally thought that animal sources pose less risk. Furthermore, knowing the source will help in identifying and eliminating the problems.

Information on the human or animal origin of Fecal pollution gives an indication of the types of pathogens that maybe expected, the risk of infection, and the treatment that may be required to control the transmission of disease. Many waterborne pathogens are difficult to detect and quantify, and specific methodology to detect them in environmental water samples has still to be developed.

Bacterial indicator organisms such as Fecal coliforms have been used to test water samples for Fecal pollution, but such indicators do not provide specific information on the specific source of pollution. These bacteria may be found in a variety of warm-blooded animals and are not unique to the human intestinal flora.

Since the early 1900s there have been various attempts to develop methods that differentiate the source of Fecal pollution. Traditionally, efforts have concentrated on determining Fecal pollution of human origin. It is now is also important to distinguish between animal sources of Fecal pollution as well as human source because animals can carry potentially harmful human pathogens. If animals are the source of indicator organisms, control measures and management practices will be different. Ribotyping analysis is one encouraging method that may be able to differentiate sources.

This review paper discusses some of the current methods to identify human sources from nonhuman sources of Fecal contamination in surface water. This review is divided into sections on microbiological and chemical approaches for identifying sources of Fecal contamination. Microbiological approaches cover bacterial and viral indicators found in the intestines of warm-blooded animals. Chemical approaches cover natural byproducts of human metabolism or human activity. Microbiological approaches include the measurement of the ratio of Fecal coliforms to Fecal streptococci or total coliforms, the detection of bacteriophages of bacteroides fragilis HSP40 and some serotypes of F-specific RNA coliphages, antibiotic resistance analysis, ribotype analysis, rep-PCR DNA technique, and use of human enteric viruses. Chemical approaches include Fecal sterol fingerprinting technique and the presence of contaminants normally associated with sewage, such as detergents. This review provides a short description of each method, some examples of studies that used the method, and a discussion of advantage and disadvantage of each method if possible.

Microbiological Methods

The ratio of fecal coliforms to fecal streptococci or total coliforms

Human fecal material may be distinguishable from animal fecal material using an old method, the ratio of fecal coliforms to fecal streptococci (FC/FS). Fecal streptococci have received widespread acceptance as useful indicators of Fecal pollution in natural aquatic ecosystem. Fecal streptococci are more abundant in animal feces than in humans; in contrast, Fecal coliforms are more abundant in human feces than in animals (see Table 4.1). Therefore, Fecal coliform to Fecal streptococci ratio have been used to differentiate human Fecal contamination from that of other warm-blooded animals (Geldreich and Kenner, 1969; Feachem, 1975). The ratio of Fecal coliforms to Fecal streptococci (FC/FS) greater than four were associated with human Fecal sources while a ratio of less than 0.7 was associated with animal Fecal sources.

If this ratio were reliable it would an inexpensive and practical method. However, the application of this method is now considered unreliable due to the variable survival rates of Fecal streptococci species. Furthermore, the ratio is affected by the methods for enumerating Fecal streptococci and by disinfection of wastewater (APHA, 1998).

Therefore, to use this method to provide information on possible Fecal pollution source we have to consider its limits: (1) Sampling needs to occur soon after waste contamination (within 24 hours if possible) because the Fecal bacteria may die off at different rates; (2) it becomes difficult to distinguish Fecal streptococci in waters from Fecal streptococci that are naturally present in soil and water when fewer than 100 Fecal streptococci/100ml are present, and (3) the water pH needs to be between 4 and 9 because Fecal coliforms die off quicker than Fecal streptococci in more acid or alkaline water (Geldreich and Kenner,1969; Coyne and Howell,1994).

Many attempts have been made to use the ratio to determine the sources of Fecal bacteria. For example, Jagals et al. (1995) showed that the ratio of Fecal coliforms to Fecal streptococci was close to unity in streams and rivers, which were upstream of the settlements, and were exposed to Fecal pollution predominantly of domestic animal origin. However, downstream of settlements which were exposed predominantly to human Fecal pollution, the ratio increased to 3.5 to 4.7.

Coyne and Howell (1994) measured FC/FS from two watersheds typical of agricultural use in Kentucky with some success. They concluded that the FC/FS ratio suggests the probable source of Fecal contamination, but considered their conclusions tentative. This method is an inexpensive and moderately complicated laboratory procedure (Sargeant, 1999). The result of this method taken alone must be quite carefully evaluated. If the method were used with some other method, such as the detection of bacteriophages, the result will be more reliable.

Fecal (thermotolerant) coliforms constitute a subset of total coliforms. These bacteria conform to all the criteria used to define total coliforms, but in addition they grow and ferment lactose with production of gas and acid at 44.5 ± 0.2 C within the first 48 hours of incubation. The ratio of Fecal coliforms to total coliforms (FC/TC) is used to show the percentage of the total coliforms comprised of Fecal coliforms, i.e., coming from the guts of warm-blooded animals. If the Fecal coliforms to total coliforms ratio exceeds 0.1 (fecal coliforms comprise10% or more than the total coliform group) suggests the presence of human fecal contamination.

Hiraishi et al. (1984) measured TC, FC, and BOD from the Tamagawa River and its tributaries in Tokyo. Geometric means of the fecal coliforms to total coliforms ratios ranged from 0.007 to 0.069 in streams, which were located on the upstream of human contamination sources, but downstream of human sources, the ratio ranged from 0.21 to 0.26. Noble et al. (2000) measured FC/TC in a regional survey of the microbiological water quality along the shoreline of the Southern California. Although they used two total/fecal ratio criteria, more than 0.1 and more than 0.2, the results were very similar. Poor water quality was found at the point-zero freshwater outlets sites where 21.8 % of the shoreline-miles exceeded 0.1 at the ratio of Fecal coliforms to total coliforms.

This method roughly shows the possibility of Fecal pollution but this method is not for distinguishing human from animal–derived Fecal matter. One of this method's shortcomings is the potential growth of Fecal coliforms in soils in tropical areas. As a result, its application in tropical areas is questionable (Bartram and Rees, 2000). The method should not be discarded for tropical areas, since it maybe useful in conjunction with other methods.

Bacteroides fragilis (strain HSP40)

Bacteroides fragilis is one of about 11 species, which are loosely placed together in the 'B. fragilis' group. They are gram-negative, anaerobic, pleomorphic rods. Tartera and Jofre (1987) tested twelve strains of different Bacteriodes species and found that one B. fragilis strain, HSP40, was detected in feces of 10 % of 40 human Fecal samples and was never detected in feces of other animal species. They suggested that the detection of Bacteriophages by strain HSP40 of B. fragilis could be used to distinguish between Fecal pollution of human and animal origin. This observation confirmed by Grabow et al. (1995). They investigated the Fecal excretion of somatic and male-specific coliphages and phages of B. fragilis strain HSP40 by human and a variety of animals. Bacteriodes fragilis phages were detected in only 13 % of 90 human stool samples but not in any animal or birds feces.

Many researchers have investigated the detection of bacteriophages infecting strain HSP40 of bacteroides fragilis. Table 4.2 summarizes levels of bacteriophages of B.

fragilis HSP40 found in different countries. Tartera et al. (1989) reported that phage infecting B. fragilis HSP40 have the same origin as human viruses and were able to multiply under anaerobic conditions, but did not replicate significantly in the environment. Jofre et al. (1989) found a significant correlation between the numbers of B. fragilis phages and human enteric viruses. Jagals et al. (1995) investigated a stream and river exposed to predominately Fecal pollution of domestic animal origin and to run-off. B. fragilis HSP40 phages were not detected by direct plaque assays in any of their samples. They concluded that more sensitive detection methods were required for the phages. Sun et al. (1997) reported that bacteriophages of Bateroides fragilis have been proven as specifically present in human feces and have relationships with water contamination by eterovirus. The researcher reported that the MPN method appeared to be more sensitive than that of PFU as reported previously Ajaujo et al. (1993) and Tartera et al. (1988).

Puig et al. (1997) tested 115 strains of B. fragilis isolated from humans and 6 of the strains were examined in feces from various animal species and in slaughterhouse wastewater. The strain HSP 40 and RYC4023 were similar in number of phages in urban sewage, but phages were not present in animal feces. Their study was performed in different countries (Puig et al., 1999). Strain HSP40 was not detected in slaughterhouse wastewater of the different geographical areas. Counts ranging from 0 to 4.5×10^4 pfu per 100ml were found in urban sewage from different geographical areas (see Table 4.2). Bradley et al. (1999) reported that the numbers of B. fragilis bacteriophages, were higher than the other bacteriophages, including F+ bacteriophages in their sampling site but they failed to isolate B. fragilis HSP40. They pointed to a lack of these bacteriophages in sewage in their study area and a need to concentrate the samples before assay.

The use of bacteriophages of B.fragilis HSP40 has the advantage of their high specificity of human Fecal pollution. Strain HSP40 detects numbers of phages up to 10^5 per 100ml of urban sewage and polluted water in some areas. However they are present in low or zero concentrations both in sewage and in natural polluted water in some countries (Jagals et al., 1995; Bradley et al., 1999; Puig et al., 1999). Therefore the use of B. fragilis HSP40 for phages detection may limit their usefulness as a universal method.

F-specific RNA coliphages subgroups

The use of male-specific RNA (FRNA) coliphages has been proposed as potential sewage pollution indicator (Havelaar and Hogeboom 1984; Havelaar et al., 1990; Furuse, 1987). The gastrointestinal tract of warm-blooded animals and domestic sewage are major habitats for these viruses (Furuse et al., 1978). FRNA phages may also be source-specific. FRNA phages fall into four distinct subgroups (groups I, II, III, and IV). Groups I and II are related, and together form major group A. Subgroups III and IV are very similar and are together called major group B. The subgroup identity of FRNA phages from environmental samples may help distinguish between human and animal waste.

Either serotyping or genotyping may achieve identification of the FRNA phage subgroups. Usually, identification of FRNA phages as members of one of the subgroups is achieved by serotyping (Osawa et al., 1981; Furuse, 1987; Havelaar et al., 1990).

Osawa et al. (1981) showed that FRNA phages belonging to group I were only detected in feces or gastrointestinal contents of domestic farm and feral zoo animals. FRNA phages isolates from pigs belonged to group I and II and those from humans groups II and III. Phages belonged to group III were exclusive to humans. Furuse (1987) found that subgroup II and III tend to be isolated from human faeces; subgroup I is usually isolated from the faces of non-human mammals and subgroup IV phages are mixed origin. Havelaar et al. (1990) serotyped 178 FRNA phage strains from faeces and 206 from wastewater. FRNA phages occur rarely in faces. FRNA phages strain from Fecal source belonged to either group I or IV with one exception. The group I and IV also predominated in wastewater samples in particular from slaughterhouse wastewater and gray water. Domestic and hospital wastewater samples sometimes yield group II and III phages. Subgroup II phages were abundant in wastewater of human origin but rare in feces. They suggested that FRNA phage should be considered as indicators of sewage pollution rather than Fecal pollution.

However, serotyping is ambiguous and too time consuming for routine assay since it requires propagation of individual plaques, preparation, titration and maintenance of antiphage sera as well as neutralization assays (Beekwilder et al., 1996). Some researchers investigated genotyping method as an alternative approach to distinguishing the four groups of FRNA phages (Hsu et al., 1995; Beekwilder et al., 1996; Griffin et al., 2000). The method employs specific gene probes to differentiate between the four subgroups of FRNA phages.

Hsu et al. (1995) developed a genotyping methods to group F-specific coliphages by nucleic acid hybridization with nonradioactive oligonucleotide probes, and compared this method with serotyping. Of the 203 isolates of FRNA phages from environmental samples, wastewater and shellfish, 99.5 and 96.6% could be classified into each group by serotyping and genotyping, respectively. Beekwilder et al. (1996) reported that identification of organisms by nucleic acid hybridization is genome-targeted and therefore has a high probability of exposing true relationships between organisms. Furthermore, it is easily performed and it appears to be quantitative and highly specific. Griffin et al. (2000) demonstrated that F-specific RNA coliphages genotyping provide confirmatory data to determine the sources of Fecal contamination. In their study FRNA phage analysis indicated that Fecal contamination at a park and surrounding areas in Florida, influenced by animal and non-human sources and 86% of the isolated FRNA phages from wastewater treatment plants were subgroup II - human in origin.

The use of FRNA phages is limited because FRNA phages are found in low occurrence in humans, although FRNA phages occur at reasonably high rates in sewage. (These phages may be poor indicators of human contamination in nonpoint source areas.) Sharing serotype between human and animals such as pigs is also a problem.

Clostridium perfringens

Spores of Clostridium Perfringens are largely Fecal in origin (Sorensen et al., 1989). They are ubiquitous in sewage sludge at concentrations several orders of magnitude higher than in soil or sediments (Fujioka and Shizumura, 1985). Contamination of deep-water disposal sites has been confirmed by the distribution of Clostridium perfringens in sediments (Hill et al., 1993 and 1996). The concentration of Clostridium perfringens as well as fecal sterols (discussed in detail in Chemical method section) in the Antarctic sediments have been used to investigate the contamination from untreated sewage outfall (Edwards et al., 1998).

A combined C. perfringens and fecal coliforms data can be used to differentiate between native birds and domestic pets (Leeming et al., 1997). Dog and cat faeces contain roughly equal and higher numbers (10^6 - 10^8 cfu/g) of both fecal coliforms and C. perfringens pore, whereas the feces of native birds (seagulls, swans, rosellas, magpies and ducks) contained 10^6 - 10^8 cfu/g of fecal coliforms and generally less than 10^2 cfu /g of C. perfringens spores. Therefore, the relatively higher ratio of Clostridium perfringens spores found in dog and cat feces may be a useful indicator when fresh fecal contamination is being investigated (Leeming et al., 1998). Furthermore, C. perfringens was found to significantly correlate to other pathogens like Giardia and Aeromonas from sewage-impacted waters (Ferguson et al., 1996).

Use of Enterococci diversity

The Enterococcus group is a subgroup of the fecal streptococci. The enterococci portion of the fecal streptococcus groups are being used as bacterial indicator for detecting the extent of fecal contamination of recreational surface waters (APHA, 1998). Water quality guidelines based on enterococci were incorporated in State of Califomnia, Assembly Bill 411 (35 enterococci/100ml based on 30 day geometric mean). In the closure of Huntington Beach in Orange County, CA, during the summer of 1999, enterococci were the most frequent bacterial indicator that exceeded the thresholds. Enterococci are found inboth human and animal faeces and vegetation. Therefore identifying the specific sources of enterococci may be a good approach to indentify fecal pollution.

In New Zealand, enterococci were specified and characterized to a sub-specific level to address the possible influence of these non-Fecal sources enterococci in a beach environment (seawater, sand, seaweed, stream water, sediment)(Anderson and Lewis, 2001). The genotypic diversity of Ent. faecium and Ent. Fecalis, and the phenotypic diversity of Ent. casseliflavus were examined using randomly amplified polymorphic DNA (RAPD)-PCR fingerprinting and biochemical screening, respectively. In their initial study, calculation of similarity coefficients from the sub-species revealed a complexity of associations between beach environmental sources. There were limited relationships among specific enterococci strains and specific environments. Similarity coefficients from Ent. Fecalis and Ent. casseliflavus were found. For example, seaweed: sand; marine water: stream water; seaweed: marine water. A high similarity value suggests but does not confirm a biological or ecological association.

Further research on the using enterococci species and sub-species is required to use identification to identify the specific sources of fecal pollution (Anderson and Lewis, 2001). This technique may be a promising method and deserves future development.

Multiple Antibiotic Resistance (MAR) Analysis

The patterns of antibiotic resistance have been used to identify sources of Fecal pollution in water. This approach is based on the fact that bacteria from wildlife species are generally lacking in antibiotic resistance, while strains from humans and domestic animals exhibit varying multiple antibiotic resistance (Sargeant, 1999). For this procedure either Escherichia coli or Fecal streptococci from different animal species are analyzed to determine the resistance pattern for several different types and strengths of antibiotics.

There have been several reports of the use of antibiotic resistance profiles to determine sources of E. coli. Krumperman (1983) showed that the multiple antibiotic resistance index of E. coli from wild animals was generally low, while human and poultry isolates had higher MAR indices. Kaspar and Burgess (1990) reported that there were larger multiple antibiotic resistance of E. coli isolated in urban areas than from rural areas, and postulated that human isolates are may present. Knudtson and Hartman (1993) measured antibiotic resistance of Fecal enterococci isolates from humans, pigs, and natural waters but found only slight differences among the various sources. Although the studies have measured antibiotic resistance of Fecal isolates from various sources, it has been difficult to use that information to identify the sources of Fecal pollution (Wiggins, 1996).

Wiggins (1996) demonstrated that Discriminant Analysis (DA) of antibiotic resistance patterns of Fecal streptococci is a useful tool for differentiating human and animal sources of Fecal pollution in water. Discriminant function analysis is a variation of multivariate analysis of variance and can be used to classify individuals into groups on the basis of the values of several classification variables (Tabachnick and Fidell, 1983; Hair et al., 1998). In the study an average of 74% of the known isolates were correctly classified into one of six possible sources (beef, chicken, dairy, human, turkey, or wild). 92% of human isolates were correctly classified. Human versus animal isolates were correctly classified at an average rate of 95%. In their recent study, more than 10,000 Fecal streptococci isolates were obtained from 236 samples of human sewage and septage, cattle feces, poultry feces, and pristine waters (Wiggins et al.1999). The average rates of correct classification into one of four possible groups (human, cattle, poultry, and wildlife) ranged 64 to 78 %. Parveen (1998) reported that DA of MAR profiles of E. coli isolates from the Apalachicola National Estuarine Research Reserve, Florida, classified 82 % and 68% of human sources and nonhuman isolates, respectively

Hagedorn et al. (1999) affirmed the work by Wiggins (1996) with the addition of Cluster Analysis. Patterns of antibiotic resistance in Fecal streptococci were analyzed in a rural Virginia watershed. They used 13 antibiotics and more than 7000 isolates from 147 samples obtained from humans, dairy cattle, beef cattle, chickens, deer, and waterfowl. Correct classification into one of the six groups was above 87 %. Fecal streptococci from their study site were classified as being predominantly from cattle (>78% of isolates)

The MAR method for differentiating between Fecal sources is promising. This method may successfully differentiate between human Fecal pollution and animals and even differentiate between animals. However, this method is time intensive for the field and laboratory work and its laboratory procedure is complicated and costly (Sargeant, 1999). Parveen et al. (1999) said the antibiotic resistance patterns of bacteria are influenced by selective pressure and thus may be different in other geographical areas and may vary over time.

DNA-based approach

More recently DNA-based approaches have been evaluated to determine whether they can be used to differentiate among sources of Fecal contamination of water. Genotype is considered to be more reliable than phenotypic biochemical reaction. Genotypic approaches differ with respect to the level of resolution of individual bacterial species or strains into distinct categories (Versalovic et al., 1998). Genotypic bacterial typing methods are shown in Table 4.3. We will provide short description and some work of Rep-PCR technique, and Ribotyping method only.

Ribotype Analysis

Genetic testing has been found to be very effective in matching DNA patterns in microorganisms to their sources. Genetic fingerprinting uses collections of E. coli, which are easily modified and adapt to various host environments, leading to changes in genetic material that are thought to be specific to these host environments. As such, the genetic variability of E. coli can be used to identify their host organisms. The DNA patterns from each of these isolates, known as a ribotype, are used to match specific strains of E. coli from a contaminated site to potential sources.

Ribotyping has been used to determine the sources of E. coli contaminating Little Soos Creek in Washington State (Samadpour and Chechowitz, 1995). In this study, 71% of the source matches belonged to 57 identified strains, leaving 29% unmatched. Samadpour also conducted DNA analysis for an investigation of the sources of fecal contamination of four San Diego beaches (CSDDEH, 1999) and the Agua Hedionda watershed (URS, 1999). From the 489 total isolates collected from San Diego beaches, storm drains and a river during wet and dry weather, 353(72.2%) were matched to 12 source groups; 179 isolates were reported unknown. Human isolates were responsible for the highest percentage of matches during dry weather conditions but were completely absence in wet weather samples: dog and bird isolates were generally the most abundant groups in wet weather samples. In the Agua Hedioda watershed, San Diego, the water samples provided bacteria for 656 E. coli isolates, 417 (63.6%) of which could be matched to a known source among a variety of warm-blooded animals (URS, 1999). The three dominant groups of source organisms were domestic pets (dogs and cat), birds, and human.

Parveen et al. (1999) analyzed Ribotype profiles of 238 E. coli isolates from human sources and nonhuman sources. Human and nonhuman source isolates showed 41 and 61 RT profiles, respectively. Ribotyping profiles with discriminant analysis showed that 97% of the nonhuman source isolates and 100% of the animal Fecal isolates were correctly classified.

Dombek et al. (2000) said ribotyping methods tend to require extensive manipulation of DNA and the use of labeled gene probes. Grouping may be influenced by a strain's prior exposure to antibiotics. Sargeant (1999) concluded this is an excellent method for determining some of the sources of Fecal contamination in a watershed, but laboratory analysis is expensive. Only a portion of receiving water isolates can be identified, leaving a significant percentage of unknown origin.

Rep-PCR DNA technique

Repetitive sequence-based polymerase chain reaction (rep-PCR) was introduced by Versalovic et al. (1991) and yields DNA fingerprints comprised of multiple, differently-sized DNA amplicons. The rep-PCR method has been useful for DNA fingerprinting of a large variety of prokaryotic and eukaryotic microorganisms (Versalovic et al., 1994; de Bruijn et al., 1995; Louws et al., 1996). Versalovic et al. (1998) reported that key advantages of rep-PCR based chromosomal typing include its speed, reproducibility, convenience, and modest resource requirements. The required equipment is often available in molecular biology laboratories.

Dombek et al. (2000) investigated the rep-PCR DNA fingerprint technique, which uses repetitive intergenic DNA sequences, to differentiate E. coli strains obtained from human and animal sources (geese, ducks, cows, pigs, chicken, and sheep). BOX and REP primers were used to generate DNA fingerprints. Their studies revealed that DNA fingerprints obtained with the BOX primer were more effective for grouping E. coli strains than with REP primers. Jackknife analysis of the similarity coefficients revealed that 100% of the chicken and cow isolates, 83 % of the human isolates and between 78 and 90 % of the other animal isolates were assigned to the correct source groups. Genotypic analysis such as rep-PCR is considered less subject to environmental effects than phenotypic analysis. Other advantages of rep-PCR are its simplicity, accuracy, and speed, which are desirable for high-throughput analysis (Versalovic et al., 1994; Dombek et al., 2000). In addition, in the rep-PCR analyses performed in their study, DNA fingerprints were generated by using whole cell suspensions, which eliminated the need for DNA purification.

Use of human enteric viruses

Human enteric viruses can be used to confirm the presence of human Fecal material. Human enteric virus groups include Norwalk virus, rotavirus, hepatitis A virus, adenovirus, and enterovirus. Adenoviruses are the only human enteric viruses that contain DNA rather than RNA and are substantially more stable than either poliovirus or hepatitis A virus in tap water and seawater (Enriquez et al., 1995). They are potentially good human source indicators. However, the methodologies involved in their detection and enumeration tend to be intensive, costly and time consuming (Sinton, 1998). Many researchers are trying to develop less expensive and reliable ways of finding viruses in seawater.

Worldwide, only a few laboratories are capable of tracking viruses in ocean water. Sassaroli et al. (2000) collected 43 samples of raw sewage and sewage-polluted creek water near San Paulo, Brazil. Adenoviruses were detected in 25(58.1%) samples by PCR using the primers hexAA1885 and hexAA1913. Jiang et al. (2001) tested human adenovirus with Coliphages and bacterial indicators (TC, FC, enterococci) in coastal waters of Southern California. Human adenoviruses were successfully detected in 4 of the 12 samples using the nested PCR method. They suggested that the detection of adenovirus could be used as an index for human fecal pollution, and the presence of other human viruses.

Host-specific molecular markers

Unlike antibiotic resistance and ribotyping analysis, which require culturing indicator organisms, detection of host-specific molecular markers does not require culturing and holds promise as a precise, rapid method for identifying sources of fecal contamination (Bernhard and Field, 2000b). Several researchers have suggested that members of the genera Bacteroides could be used an alternative fecal pollution indicator (Allsop and Stickler, 1985 and Kreader, 1995). Members of these genera are strict anaerobes, are restricted to warm-blooded animals, and make up a significant portion of fecal bacteria. The use of these organisms as indicators however, has been limited because strict anaerobes are often difficult to grow. Using molecular methods rather culture-based methods to detect them can circumvent the difficulty of growing strict anaerobes.

Bernhard and Field (2000a) identified host-specific Bacteroides-Prevotella 16S rDNA markers for human and cows by screening fecal DNAs by length heterogeneity PCR (LH-PCR) and terminal restriction fragment length polymorphism (T-RFLP) analysis. LH-PCR (Suzuki et al., 1998) and T-RFLP (Bruce, 1997;Clement et al., 1998; Liu et al., 1997) are methods that are used to analyze differences in the lengths of gene fragments due to insertions and deletions and to estimate the relative abundance of each fragment. Following the study, they identified additional clones, recovered from water samples, and developed cluster-specific primers that can discriminate between human and ruminant feces using the sequences from fecal and water clones (Bernhard and Field, 2000b). They believe that these PCR assays provide a promising diagnostic tool for identifying nonpoint sources of fecal pollution, although extensive field-testing is required to determine the efficiency of the assays and the geographic distribution of the host-specific markers.

In the past, most of the DNA related technologies involved characterization of individual isolates. Since we are limited in the number of isolates we can test, there is always a statistical problem if a high abundance of target organism present in the contaminated water (Jiang, 2001). This method could overcome conventional culture-dependent technique's problem that mentioned above, it is promising, and further investigation of this approach is required.

Chemical Methods

Fecal sterols

The term ''Fecal sterol''is a broad term covering the various A-C27, C28, and C29 cholestane-based sterols found in Fecal material (Sinton et al., 1998). Using Fecal sterol such as coprostanol has been proposed as an alternative measure of Fecal pollution by a large number of researchers (Walker et al., 1982). Coprostanol is formed in the gut of human and higher mammals by enzymatic hydrogenation of choleserol or by stereo-specific bacterial reduction of cholesterol (MacDonald et al., 1983). Therefore it is present in sewage effluent and sewage contaminated waters. Several studies have highlighted the usefulness of coprostanol for examining sewage pollution in many diverse environments (Venkatesan and Kaplan, 1990; Venkatesan and Mirsadeghi, 1992). Fecal sterols analysis has been extended to differentiate human and animals sources of pollution. The distribution of sterols found in faeces and hence their source specificity is largely determined by the following three factors (Leeming et al., 1994): (1) The animal's diet. For example, humans, cows, and dogs, respectively, are omnivorous or herbivorous or carnivorous. Each type of diet contains a different sterol profile and the proportions of sterol precursors entering the digestive tract are different; (2) irrespective of dietary habits, many animals can biosynthesize sterols, and (3) anaerobic bacteria in the digestive tract of some animals biohydrogenate sterols to stanols of different isomeric configurations. This is probably the major factor in dictating the composition and characteristics of the sterols fingerprint. Because of this feature, Fecal sterols offer an advantageous approach that can help distinguish among sources of Fecal pollution.

Fecal sterol such as coprostanol, which constitutes ~60% of the total sterols found in human faces, has been successfully used to trace sewage in many countries. Coprostanol is produced in the intestine of humans and some higher mammals by bacterial biohydrogenation of cholesterol to the 5β(H)-stanol. Coprostanol profiles from a wide variety of animals show difference in the presence/absence or relative amounts of individual sterols (Venkatesan, 1995; Leeming et al, 1996; DNRP, Broward County, 1998). Figure 4.1 shows coprostanol content (ug /g) in Fecal matter from various sources.

Venkatesan (1995) analyzed Fecal sterols from humans, several animals and influent and effluent samples of Hyperion Plant, and four storm drains. The relative and absolute amounts of coprostanol were much higher in human feces compared to the animals and avian species. For example, the human specimen contained at least 10 times as much coprostanol than the specimens from carnivores and 20-100 times the specimens from herbivores. She suggested that coprostanol in conjunction with other specific sterols parameters can be used to distinguish input of humans from domestic animals and birds.

Leeming et al. (1996) examined the Fecal sterols from humans and 14 species of animals common to rural or urban environments. Human faeces contained ten times more coprostanol on a dry weight basis than faeces from cats and pigs. Herbivores such as cows, sheep and horse feces contained some coprostanol, but their sterol profiles were dominated by C_{29} sterols (24-ethylcoprostanol and 24-ethylcholesterol). They concluded the 'Sterol fingerprints' of the feces of humans and animals are sufficiently distinctive to be of diagnostic value in determining whether Fecal pollution in water samples are of human or animal origin. They also studied the use of a wider range of Fecal sterols, in combination with conventional bacterial indicators, to distinguish the source of Fecal pollution in Lake Tuggerah, Australia, and found that native birds were a major source of the Fecal pollution using an empirical basis and sterol ratio (Leeming et al., 1997).

A Fecal sterol study was conducted in Florida to determine coprostanol, epicoprostanol, cholestanol and epicholesterol from surface water, feces and sediment (DNRP, Broward county, 1998). From the initial data they suggested that it was possible to tell the difference between fresh Fecal samples of human and nonhuman origin based upon the concentration ratios of two of the Fecal sterols. The coprostanol/cholestanol concentration ratio was shown to be greater than 1.0 in human sources and less than 1.0 in non-human feces.

More study is needed on this method if it is to be used for nonpoint sources. This method requires expensive gas chromatography and requires up to 10 liters of samples to be filtered through a glass fiber filter to concentrate particulate stanols. Nevertheless, it is an appropriate method for specific studies investigating the proportion of human and animal Fecal contamination (Bartram and Rees, 2000).

Long-Chain Alkylbenzenes

Long-chain alkylbenzens (LABs) having C_{10}–C_{14} normal alkyl chains are sulfonated in the industrial production of linear alkylbenzene sulfonates. They are widely used as anionic surfactants in commercial detergents (Eganhouse et al., 1983). A number of studies have found LABs in the waters and sediments exposed to sewage. Observations have been made worldwide and especially in Southern California. Table 4.4 shows the occurrence of LABs in different California locations. LABs are purely synthetic and are derived solely from direct industrial discharges and domestic wastes (Eganhouse, 1986). They are therefore strongly indicative of human sources. However, they may not be related to sewage such as industrial pollution (Bartram and Rees, 2000). They are also generally present up to one order of magnitude lower than the corresponding Fecal sterol in human derived wastes (Sinton et al.,1998). They are therefore regarded as complimentary to the Fecal sterols as domestic sewage pollution.

Caffeine

Caffeine is a compound that is present in several beverages, coffee, tea, and carbonated drinks, and in pharmaceutical products. Caffeine and its metabolites are excreted in the urine of individuals who have consumed beverages and pharmaceuticals containing caffeine. It has been speculated that caffeine has promise as an indicator of human fecal pollution if the population being studied uses caffeine, is uniquely and unambiguously associated with human activity. Caffeine has been detected in domestic wastewater effluent, environmental surface water samples, ground water and finished drinking water in several locations worldwide. Caffeine was present in municipal wastewaters of populations that use caffeine at levels between 20 and 300 ug/L (Rogers et al., 1986). C affeine has also been detected in Los Angels County wastewater treatment plant effluent samples, 1980-1981, at 40 ug/L (Spectrum Laboratories, 1998). Concentrations of caffeine in water along the main stem of the Mississippi River, from Minneapolis, Minnesota, to New Orleans, Louisiana, ranged from 0.01 to 0.07 ug/L (Barber et al., 1995). Caffeine was present in samples collected from wastewater treatment plant effluent, and agricultural and urban runoff in Canada and United States, in both dissolved and particulate phases at concentrations up to 0.115 ug/L and 0.044ug/L, respectively (Standley et al., 2000). Although caffeine has been extensively detected in environments exposed to human wastes, there are only a small number of studies that can be used to estimate the probable concentrations of caffeine that might result from sewage spills. Also, because caffeine is extensively metabolized; only 3 percent of ingested caffeine is excreted unmetabolized in the urine (Tang-Liu et al., 1983), its sensitivity as a marker of human fecal pollution is unknown. Therefore, further investigations are required.

Conclusions

There is no easy, low cost method for differentiating between human and non-human sources of bacterial contamination. No single indicator or approach is likely to represent all the facets and issues associated with contamination of waterways with Fecal matter.

At present, the best hope of distinguishing Fecal pollution of human and animal origin is an appropriate combinations of indicators. Statistical analyses of appropriate groups of methods offer the best possibility of identifying human sources. Unfortunately, relying on a combination of methods will probably require a longer period of analysis than relying on a single method. A combination of methods may be useful to determine sources in chronic situations as opposed to episodic events.

Many promising methods have been identified in this review. None, at least at the time of this writing, has been demonstrated in a full scale monitoring program. Most techniques have been limited to research laboratories. Such demonstrations are needed to demonstrate the utility of the methods. Also, commercial laboratories will need assurances that the required investments in training and equipment are justified and recoverable.

A good course of action to further this technology would be to conduct several, long-term investigations, where an advanced method is used in parallel with existing monitoring techniques. Monitoring agencies will need to be involved, in order to evaluate the required retraining and adjustments in their current procedures.

Table 4.1. Bacterial densities in warm-blooded animals feces.
(Sources: Pitt, 1998; Godfrey, 1992;Geldrich et al., 1962)

Source	Fecal coliform (density/gram)	Fecal streptococci (density/gram)	Ratio FC/FS
Human	1.3×10^7	3.0×10^6	4.33
Cats	7.9×10^6	2.7×10^7	0.29
Dogs	2.3×10^7	9.8×10^8	0.02
Rats	1.6×10^3	4.6×10^7	0.003
Cows	2.3×10^3	1.3×10^7	0.02
Ducks	3.3×10^7	5.4×10^7	0.61

Table 4.2. Levels of active bacteriophages against B. fragilis HSP40 in wastewater and surface water.

Samples	Country	Range (mean) pfu or cfu or MPN/100 ml	% positive samples	Reference
Sewage	Spain	(8.9×10^3)	100	Tartera et al. 1988
Highly polluted river		(4.8×10^3)	100	
Moderately polluted river		(6.5×10^3)	59	
Pollted sea water		(7.3×10)	44	
Non-sewage polluted water		0	0	
Raw Sewage1[a]	Spain	2.1×10^2 - 4.6×10^4 (5.3×10^3)	100	Tartera et al. 1989
Raw Sewage2[b]		2.3×10^2 - 4.6x103 (1.3×10^3)	100	
Slaughterhouse wastewater1		0 - 1, 2	20	
Slaughterhouse wastewater2		0	0	
River		0 - 43 (6.7)	46	
Non-sewage polluted water		0	0	
Sewage	Spain	1.0×10^3 - 2.6×10^5 (2.5×10^4)	100	Comaz et al. 1991
Seawater				
10 m from sewage discharge		2.9×10^3 - 1.0×10^5 (1.8×10^4)	100	
100 m from sewage disch.		<10 - 1.0×10^4 (3.4×10^2)	70	
1000 m from sewage disch.		< 10 - 7.9×10^2 (< 10)	80	
Non-sewage polluted seawater		< 10 (< 10)	100	
Highly polluted river	Spain	10 - 1.6×10^3 (1.2×10^2)	100	Joffe et al. 1995
Low-pollute river		BLDc - 1.0×10^2 (10)	36.4	
Stream and river water	S. Africa		0	Jagals et al. 1997
Sewage (influent)	France	$\geq 4.4 \times 10^4$	100	Sun et al. 1997
Sewage (effluent)		$\geq 4.4 \times 10^3$	75	
River (downstream treatment)		$\geq 1.6 \times 10^3$	100	
River (upstream treatment)		< 40	50	
Sewage, river water	Spain	(2.3×10^3)	100	Araujo et al. 1997
River water		(2.3×10^1)	83.3	
Stream	PA, US	(2.21×10^2 - 3.9×10^2)		Brenner et al. 1999
Marine bathing water	UK		0	Bradley et al. 1999

Table 4.2. Continued.

Samples	Country	Range (mean) pfu or cfu or MPN/100 ml	% positive samples	Reference
Sewage (different countries)	Netherlands	1.0×10^2 - 2.6×10^2	0	Puig et al. 1999
	Ireland	1.4×10^2 - 1.6×10^2		
	Austria	0.5×10^2 - 8.5×10^2		
	Portugal	0 - 0.4×10^2		
	Germany	1.3×10^2 - 2.2×10^2		
	Sweden	0.9×10^2		
	France	3.1×10^3		
	S. Africa			
Slaughterhouse wastewater	different countries[d]	1.1z10^4 - 4.5×10^4		

[a] Samples were colleted from Colector Prim which receives mainly domestic sewage.
[b] Samples were collected from de Levante, which receives a mixture of domestic and industrial waste water
[c] BDL, below detection limit
[d] The countries are Netherlands, Ireland, Denmark, Portugal, Germany, South Africa

Table 4.3. List of bacterial genotypic typing methods according to ability to distinguish genus/species or subspecies/strains (Versalovic et al., 1998).

Genus/Species	*Subspecies/Strain*
Ribotyping	ARDRA
tRNA-PCR	Chromosomal RFLP
ITS-PCR	ITS Sequencing
16S rRNA sequencing	Plasmid RFLP
	Pulsed-field gel electrophoresis (PFGE)
	Randomly amplified polymorphic DNA (RAPD)
	Rep-PCR

ARDRA: Amplified ribosomal DNA Restriction Analysis
ITS: Internal Transcribed Spacer
RFLP: Restriction fragment length polymorphisms
PCR: Polymerase chain reaction

Table 4.4. Levels of LABs in the state of California.

Samples	Area of study	Characteristic	Concentration[a]	Reference
Treated wastewater effluent (ug/L)	CSDOC[b]		8.2 ± 1.8	Phillips et al., 1997
	PLWTP[c]	Particulates	1.92 - 2.76	Zeng et al., 1997
River water	Tijuana river runoff, San Diego	Particulates	0.057-0.714	Zeng et al., 1997
Sewage sludge (ng/g)	JWPCP[d]	Primary-secondary	200,000	Bayona et al., 1997
Sediments (ng/g)	Santa Monica basin	Box-core (0-1, 0-4 cm)	236 ± 124	Bayona et al., 1997
	Santa Ana river		10 ± 7	Phillips et al., 1997
	Newport bay		18 ± 12	Phillips et al., 1997
	Off the coast of San Diego		ND[e] - 39.2	Zeng et al., 1997

[a] Arithmetic mean of total concentrations ± standard deviation or range of concentration.
[b] CSDOC: County Sanitation Districts of Orange County.
[c] PLWTP : Point Loma Wastewater Treatment Plant.
[d] JWPCP: Joint Water Pollution Control Plan.
[e] ND: not detectable.

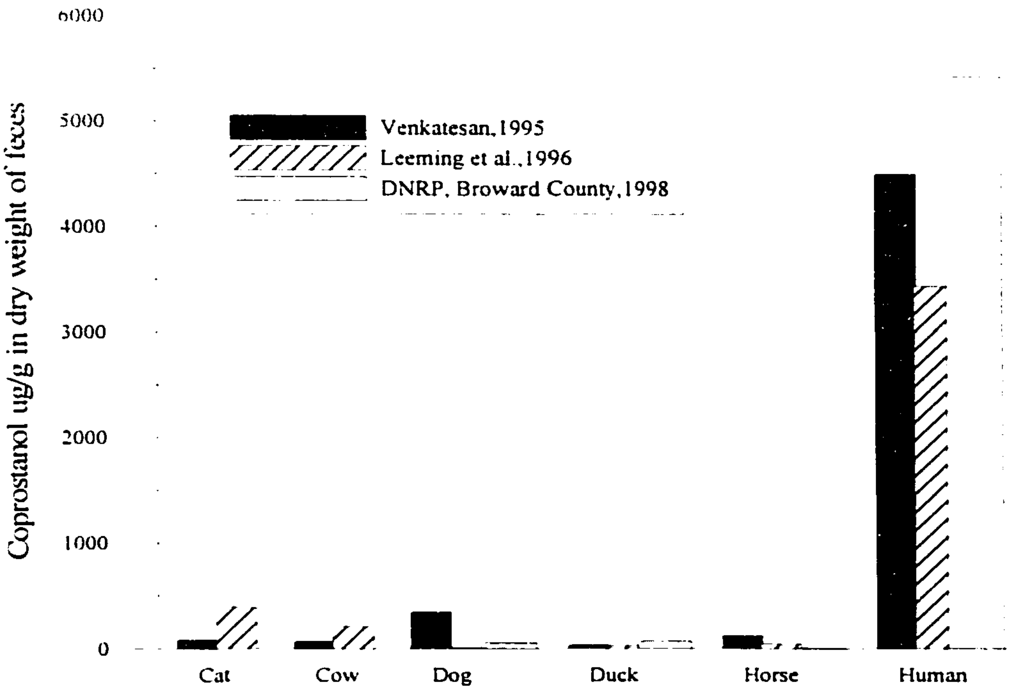


Figure 4.1. Comparison of coprostanol content in various Fecal samples, ug/g

References

Allsop, K., and Stickler, J.D. (1985). " An assessment of Bacteroides fragilis group organisms as indicators of human Fecal pollution." *Journal of Applied Bacteriology* **58**: 95-99.

Anderson, S. A., and Lewis, G. D. (2001). " Evaluation of enterococci diversity and persistence in a marine bathing beach environment," University of Auckland, Auckland, New Zealand.

APHA. (1998). *Standard Methods for the examination of water and wastewater*, American Public Health Association, American Water Works Association and Water Environmental Federation.

Araujo, R. M., Lasobras, J., Lucena, F., and Jofre, J. (1993). "Methodological improvements for their recovery of Bacteroides fragilis phages and coliphages from environmental samples." *Water Science and Technology* **27**(3-4): 119-122.

Araujo, R., Lasobras, J., Puig, A., Lucena,F., and Jofre, J. (1997). "Abundance of Bacteriophages of enteric bacteria in different freshwater environment." *Water Science and Technology* **35**(11-12): 125-128.

Barber, L.B.II., Leenheer, J.A., Pereira, W.E., Noyes, T.I., Brown, G.K., Tabor, C.F., and Writer, J.H.(1995). " Organic compounds and sewage-derived contaminants" In *Contaminants in the Mississippi River*, 1987-1992, ed. Meade, R.H., U.S. Geological Survey Circular 1133:115-135 (water.usgs.gov/pubs/circ1133/organic.html)

Bartram, J., and Rees, G. (2000). Monitoring Bathing Water, E & FN SPON.

Bayona, J.M., Chalaux, N., Dachs, J., Maldonado, C., Venkatesan, M.I., and Albaiges, J. (1997) " Use of Trialkylamines as marker of sewage addition into the marine environment" *ACS Symposium Series 671*:261-275. Eganhouse, R.P. (Ed).

Beekwilder, J., Nieuwenhuizen, R., Havelaar, A.H., and van Duin, J. (1996). "An oligonucletide hybridization assay for the identification and enumeration of F-specific RNA phages in surface water." *Journal of Applied Bacteriology* **80** (179-186).

Bernhard, A.E., and Field, K.G. (2000a). " Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes." *Applied and Environmental Microbiology* **66**(4): 1587-1594.

Bernhard, A.E., and Field, K.G. (2000b). " A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA." *Applied and Environmental Microbiology* **66**(10): 4571-4574.

Bradley, G., Carter, J., Gaudie, D., and King, C. (1999). "Distribution of the human Fecal bacterium Bacteroides fragilis, its bacteriophages and their relationship to current sewage pollution indicators in bathing water." *Journal of Applied Microbiology Symposium Supplement* **85**: 90s-100s.

Brenner, F. J., Brenner, E.K., and Schwartz, T.E. (1999). "Use of plaque assay to detect enteric viruses in a rural watershed." *Journal of Environmental Quality* **28**: 845-849.

Bruce, K.D. (1997). " Analysis of mer gene subclasses within bacterial communities in soils and sediments resolved by Fluorescent-PCR-Restriction Fragment Length Polymorphism profiles." *Applied and Environmental Microbiology* **63**(12): 4914-4919.

Clement, B.G., Kehl, L.E., DeBord, K.L., and Kitts, C.L. (1998). " Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities." *Journal of Microbiological Methods* **31**: 135-142.

Cornax, R., Morinigo, M. A., Balebona, M.C., Castro, D., and Borrego,J.J. (1991). "Significance of several Bacteriophages groups as indicators of sewage pollution in marine waters." *Water Research* **25**(6): 673-678.

Coyne, M. S., and Howell, J.M. (1994). "The fecal coliform/fecal streptococci ratio(FC/FS) and water quality in the bluegrass region of Kentucky." *Soil and Science News & Views* **15**(9).

CSDDEH (1999). An investigation of the source of fecal contamination to four San Diego beaches, County of San Diego Department of Environmental Health

de Bruijn, F. J., Schneider, M., Rossbach, U., and Louws, F. J. (1995). Automated florescent and conventional rep-PCR genomic fingerprinting and multiplex PCR to classify bacteria and track genes. Proceedings of the Seventh International Symposium on Microbial ecology, Brazil.

DNRP, Broward County (1998). Implementation of a chemical method for differentiating human and animal fecal impacts in surface waters and sediment, Broward County Department of Natural Resource Protection.

Dombek, P. E., Johnson, L.K., Zimmerley, S.T., and Sadowsky. M.J. (2000). "Use of repetitive DNA sequences and the PCR to differentiate Escherichia coli isolates from human and animal sources." *Applied and Environmental Microbiology* **66**(6): 2572-2577.

Edwards, D. D., McFeters, G.A., and Venkatesan, M. I. (1998) " Distribution of Clostridium perfringens and fecal sterols in a Benthic coastal marine environment influenced by the sewage outfall from McMurdo station, Antarctica." *Applied and Environmental Microbiology* **64**(7): 2596-2600.

Eganhouse, R.P. (1986) "Long-chain benzenes: their analytical chemistry, environmental occurrence and fate." *International Journal of Environmental Chemistry* **26**:241-263

Eganhouse, R.P., Blumfield, D.L., Kaplan, I.R. (1983). " Long-chain alkylbenzenes as molecular tracers of domestic wastes in the marine environment." *Environmental Science and Technology* **17**:523-530

Enriquez, C. E., Hurst, C. J., and Gerba, C. P. (1995). " Survival of the enteric adenoviruses 40 and 41 in tap, sea, and wastewater." *Water Research* **29**:2548-2553.

Feachem, R. G. (1975). "An improved role for Fecal coliform to fecal streptococci ratios in the differentiation between human and non-human pollution sources(Note)." *Water Research* **9**: 689-690.

Ferguson, C. M., Coote, B.G., Ashbolt, N. J., and Stevenson, I. M. (1996) " Relationships between indicators, pathogens and water quality in an estuarine system." *Water Research* **30**(9): 2045-2054.

Fujioka, R. S., and Shizumura, L. K. (1985) "Clostridium perfringens, a reliable indicator of stream water quality." *Journal of the Water pollution Control Federation* **57**: 986-992.

Furuse, K. (1987). *Distribution of coliphages in the environment :General consideration. Phage ecology.* S. M. Goyal, Gerba,C.P.,and Bitton, G., Wiley Interscience: 87-124.

Furuse, K., Sakurai,T., Hirashima,A., Katsuki,M., Ando,A., and Watanabe, I. (1978). "Distribution of ribonucleic acid coliphages in South and East Asia." *Applied and Environmental Microbiology* **35**: 995-1002.

Geldreich, E. E., and Kenner, B.A. (1969). "Concepts of fecal streptococci in stream pollution." *Journal of the Water Pollution Control Federation* **41**(8): R336-R352.

Geldrich, E., Borden,R., Huff,C., Clark,H., and Kabler,P. (1962). "Type distribution of coliform bacteria in the feces of warm-blooded animals." *Journal Water Pollution Control Federation* **34**(295).

Godfrey, A. (1992). Sources and fate of microbial contaminants. Recreational Water Quality Management Volume2:Freshwaters. D. K. a. R. Hanbury. New York. Ellis Horwood: 137-154.

Grabow, W. O. K., Neubrech, T. E., Holtzhausen, C. S., and Jofre, J. (1995). "Bacteroides fragilis and Escherichia coli bacteriophages: Excretion by human and animals." *Water Science and Technology* **31**(5-6): 223-230.

Griffin, D. W., Stokes, R., Rose, J.B., and Paul III, J.H. (2000). "Bacterial indicator occurrences and the use of an F+specific RNA coliphages assay to identify fecal sources in Homosassa springs, Florida." *Microbial Ecology* **39**(1): 56-64.

Hagedorn, C., Robinson, S.L., Filtz, J. R., Grubbs, S. M., Angier, T.A., and Reneau, Jr., R.B. (1999). "Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in Fecal Streptococci." *Applied and Environmental Microbiology*, **65**(12): 5522-5531.

Hair, J. F., Jr., Anderson, R. E., Tatham, R.L., and Black, W.C. (1998). *Multivariate data analysis*, 5th ed. Upper Saddle River, N. J., Prentice-Hall, Inc.

Havelaar, A. H., and Hogeboom, W.M. (1984). "A method for the enumeration of male – specific bacteriophages in sewage." *Journal of Applied Bacteriology*, **56**: 439-447.

Havelaar, A. H., Pot-Hogeboom, W.M., Furuse, K. , Pot, R., and Hormann, M.P. (1990). "F-specific RNA bacteriophages and sensitive host strains in faces and wastewater of human and animal origin." *Journal of Applied Bacteriology*, **69**: 30-37.

Hill, R.T., Knight, I.T., Anikis, M.S., and Colwell, R.R. (1993) " Benthic distribution of sewage sludge indicated by Clostridium perfringens at a deep-ocean dump site." *Applied and Environmental Microbiology* **59**(1): 47-51.

Hill, R.T., Straube, W.L., Palmisano, A.C., Gibson, S.L. and Colwell R.R. (1996) " Distribution of sewage indicated by Clostridium perfringens at a deep-water disposal site after cessation of sewage disposal." *Applied and Environmental Microbiology* **62**(5): 1741-1746.

Hiraishi, A., Saheki, K., and Horie, S. (1984). " Relationships of total coliform, fecal coliform, and organic pollution levels in the Tamagawa River." *Bulletin of the Japanese Society of Scientific Fisheries*, **50**(6): 991-997.

Hsu, F., Carol Shieh, Y. S., van Duin, J., Beekwilder, M.J., and Sobsey, M.D. (1995). "Genotyping male-specific RNA coliphages by hybridization with Oligonucleotide probes." *Applied and Environmental Microbiology*, **61**(11): 3960-3966.

Jagals, P., Grabow, W.O.K., and de Villiers, J.C. (1995). "Evaluation of indicators for assesment of human and animal Fecal pollution of surface run-off." *Water Science and Technology*, **31**(5-6): 235-241.

Jiang, S. (2001). Personal communication, Assistant Professor of Environmental Analysis and Design at University of California, Irvine, CA.

Jiang, S., Noble, R., and Chu, W. (2001). " Human adenoviruses and coliphages in urban runoff impacted coastal waters of Southern California." *Applied and Environmental Microbiology*, **67**(1): 179-184.

Jofre, J., Blasi, M., Bosch, A. and Lucena, F. (1989). "Occurrence of bacteriophages infecting Bacteroides fragilis and other virus in polluted marine sediments." *Water Science and Technology*, **21**: 15-19.

Jofre, J., Olle,E., Lucena, F. and Ribas, F. (1995). "Bacteriophages removal in water treatment plants." *Water Science and Technology*, **31**(5-6): 69-73.

Kaspar, C. W., and Burgess, J.L. (1990). "Antibiotic resistance indexing of Escherishia coli to identify sources of fecal contamination in water." *Canadian Journal of Microbiology*, **36**: 891-894.

Knudtson, L. M., and Hartman, P.A. (1993). "Antibiotic resistance among enterococcal isolates from environmental and clinical sources." *Journal of Food Protection*. **56**: 489-492.

Kreader, C.A. (1995). " Design and evaluation of Bacteroides DNA probes for the specific detection of human fecal pollution." *Applied and Environmental Microbiology* **61**:1171-1179.

Krumperman, P. H. (1983). "Multiple antibiotic resistance indexing of Escherichia coli to Identify high-risk sources of fecal contamination of foods." *Applied and Environmental Microbiology*, **46**: 165-170.

Leeming, R., Ball, A., Ashbolt, N., Jones, G., and Nichols, P.(1994) " Distinguishing between human and animal sources of Fecal pollution." *Chemistry in Australia*, 62:434-435.

Leeming, R., Ball, A., Ashbolt, N. and Nichols, P. (1996). "Using Fecal sterols from humans and animals to distinguish Fecal pollution in receiving waters." *Water Research* **30**(12): 2893-2900.

Leeming, R., Latham, V., Rayner, M., and Nichols, P. (1997). "Detecting and Distinguishing sources of sewage pollution in Australian inland and coastal waters and sediments." *ACS Symposium Series* 671: 306-319. Eganhouse, R.P. (Ed).

Leeming, R., Bate, N., Hewlett, R., and Nichols, P. D. (1998) " Discriminating Fecal pollution: A case study of stormwater entering port Phillip Bay, Australia." *Water Science and Technology* **38**(10):15-22.

Liu, W., Marsh, T.L., Cheng, H, and Forney, L.J. (1997). " Characterization of microbial diversity by determining Terminal Restriction Fragment Length Polymorphisms of genes encoding 16S r RNA." *Applied and Environmental Microbiology* **63**(11): 4516-4522.

Louws, F. J., Schneider, M., and de Bruijn, F.J. (1996). Assessing genetic diversity of microbes using repetitive sequence-based PCR (rep-PCR). *Environmental Applications of Nucleic Acid Amplifications Techniques*. G. Toranzos, Technomic Publishing Co., Lancaster, PA: 63-64.

MacDonald, I. A., Bokkenheuser, V.D., Winter, J., McLernon, A.M. and Mosbach, E.H. (1983). "Degradation of sterols in the human gut." *Journal of Lipid Research*, **24**: 675-694.

Noble, R. T., Dorsey, J.H., Leecaster, M., Orozco-Borbon, V., Reid, D., Schiff, K., and Weisberg, S.B. (2000). " A regional survey of the microbiological water quality along the shoreline of the Southern California bight" *Environmental Monitoring and Assessment* **64**:435-447.

Osawa, S., Furuse, K., and Watanabe, I. (1981). "Distribution of ribonucleic acid coliphages in animals." *Applied and Environmental Microbiology* **41**(1): 164-168.

Parveen, S., and Tamplin, M.L. (1998). Methods for measuring levels of human and nonhuman sources of fecal pollution in water. *Abstracts of the 98th General Meeting of the American Society for Microbiology* 1998, American Society for Microbiology, Washington, D.C.

Parveen, S., Portier, K.M., Robinson, K., Edmiston, L., and Tamplin, M.L. (1999). "Discriminant analysis of ribotype profiles of Escherichia coli for differentiating Human and Nonhuman sources of fecal pollution." *Applied and Environmental Microbiology*, **65**(7): 3142-3147.

Pitt, R. (1998). Epidemiology and stormwater management. *Stormwater Quality Management*. New York,, CRC/Lewis publishers.

Phillips, C.R., Venkatesan, M.I., and Bowen, R. (1997). " Interpretations of contaminant sources to San Pedro Shelf sediments using molecular markers and principal component analysis." *ACS Symposium Series 671*:242-260. Eganhouse, R.P. (Ed).

Puig, A., Jofre, J., and Araujo, R. (1997). "Bacteriophages infecting various bacteroides fragilis strains differ in their capacity to distinguish human from animal Fecal pollution." *Water Science and Technology* **35**(11-12): 359-362.

Puig, A., Queralt, N., Jofre, J., and Araujo, R. (1999). "Diversity of Bacteroides fragilis strains in their capacity to recover phages from human and animal wastes and from fecally polluted wastewater." *Applied and Environmental Microbiology*, **65**(4): 1772-1776.

Rogers, I.H., Birtwell, I.K. and Kruznski, G.M. (1986). "Organic extractables in municipal waste water" *Canadian Journal of Water Pollution Research* **21**(2),187-204.

Samadpour, M., and Chechowitz, N. (1995). Little Soos Creek microbial source tracking: a survey, University of Washington Department of Environmental Health.

Sargeant, D. (1999). Fecal contamination source identification methods in surface water, Washington State Department of Ecology.

Sassaroli, A., Santos, F.M., Harsi, C.M., Vieira, M. J., Garrafa, P., and Mehnert, D.U. (2000). " One-year survey of hepatitis A and adenoviruses in wastewater samples on Sao Paulo city, Brazil" *1st World Water Congress of the International Water Association : Book 7 Health-Related water Microbiology*: 158.

Sinton, L. W., Finlay, R. K. and Hannah, D.J. (1998). "Distinguishing human from animal Fecal contamination in water :a review." *New Zealand Journal of Marine and Freshwater* **32**: 323-348.

Sonsino, J., Torrijos, R.L., and Zimmerman, M. E. (1999). "Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution." *Applied and Environmental Microbiology* **65**(8): 3483-3486.

Sorensen, D.L., Eberl, S.G. and Diksa, R.A. (1989) "Clostridium perfringens as a point source indicator in non-point polluted streams." *Water Research* **23**; 191-197.

Spectrum Laboratories. (1998). Caffeine Fact Sheet (www.speclab.com/compound/c58082)

Standley, L.J., Kaplan, L.A., and Smith, D., (2000). " Molecular tracer of organic matter sources to surface water resources." *Environmental Science and Technology* **34**(15): 3124-3130.

Sun, Z. P., Levi, Y., Kiene,L., Dumoutier, N., and Lucena, F. (1997). "Quantification of bacteriophages of bacteroides fragilis in environmental water samples of Seine river." *Water, Air, and Soil Pollution* **96**(1-4): 175-183.

Suzuki, M., Rappe, M. S., and Giovannoni, S.J. (1998). " Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity." *Applied and Environmental Microbiology* **64**(11): 4522-4529.

Tabachnick, B. G., and Fidell, L.S. (1983). *Using multivariate statistics*, Harper & Row, Publishers, Inc., New York.

Tang-Liu, D.D., Williams, R.L., and Riegelman, S. (1983). "Disposition of caffeine and its metabolites in man." *Journal of Pharmacology and Experimental Therapeutics* **24**(1): 180-185.

Tartera, C., and Jofre, J. (1987). "Bacteriophages active against Bacteroides fragilis in sewage-polluted waters." *Applied and Environmental Microbiology* **53**: 1632-1637.

Tartera. C., Jofre. J., and Lucena. F. (1988). "Relationship between numbers of enteroviruses and bacteriophages infecting Bacteroides fragilis in different environmental samples." *Environmental Technology Letters* **9**: 407-410.

Tartera, C., Lucena,F., and Jofre, J. (1989). "Human origin of Bacteroides fragilis bacteriophages present in the environment." *Applied and Environmental Microbiology* **55**(10): 2696-2701.

URS Greiner Woodward Clyde (1999). Agua Hedionda watershed bacteriological study 1998-1999. San Diego, URS Greiner Woodward Clyde.

Venkatesan, M. I. (1995). Coprostanol and other sterols as innovative indicators for human pathogen, Institute of Geophysics and Planetary Physics, University of California, Los Angeles., A report to Santa Monica Bay Restoration Project, Award # 3-040-140-0.

Venkatesan, M. I., and Kaplan, I.R. (1990). "Sedimentary coprostanol as an index of sewage addition in Santa Monica Basin, South California." *Environmental Science and Technology* **24**: 208-214.

Venkatesan, M. I., and Mirsadeghi, F.H. (1992). " Coprostanol as sewage tracer in McMurdo Sound, Antarctica." *Marine Pollution Bulletin* **25**(9-12): 328-333.

Versalovic, J., de Bruijn, F.J., and Lupski, J.R. (1998). Repetitive sequence-based PCR(rep-PCR) DNA fingeprinting of bacterial genomes. *Baterial Genomes:physical Structure and analysis*, Chapman and Hall, New York: 437-454.

Versalovic, J., Koeuth, T., and Lupski, J.R. (1991). "Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes." *Nucleic Acids Research* **19**: 6823-6831.

Versalovic, J., Schneider, F., de Bruijn, F.J., and Lupski, J.R. (1994). "Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction." *Methods in Molecular and Cellular Biology*. **5**: 25-40.

Walker, R.W., Wun, C.K., and Litsky, W. (1982). " Coprostanol as an indicator of fecal pollution." *CRC Critical Reviews in Environmental Control* **12**(2):91-112.

Wiggins, B. A. (1996). "Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of Fecal pollution in natural waters. " *Applied and Environmental Microbiology* **62**(11): 3997-4002.

Wiggins, B. A., Andrews, R.W., Conway, R.A., Corr, C.L., Dobratz, E.J., Dougherty, D.P., Eppard, J.R., Knupp, S.R., Limjoco, M.C., Mettenburg, J.M., Rinehardt. J.M., Sonsino, J., Torrijos, R.L., and Zimmerman, M.E. (1999). "Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution." *Applied and Environmental Microbiology* **65**(8): 3483-3486.

Zeng, E.Y., Khan, A.R. and Tran, K. (1997). " Organic pollutants in the coastal marine environment off San Diego, California. 3. Using Linear alkylbenzenes to trace sewage-derived organic materials." *Environmental Toxicology and Chemistry* **16**(2): 196-201

Chapter 5.

Conclusions

The primary objective of this study was to develop a neural network model to classify landuses from water quality data. For non-point source such as stormwater, load allocations are often based on landuse types. Identifying landuse with water quality data will help set the Total Maximum Daily Loads level and eventually provide opportunities for better management to control stormwater pollution.

The existence of stormwater monitoring programs should represent progress towards achieving clean water goals. However, studies have not yet been performed to understand the utility of the current program or to improve their usefulness. Several monitoring programs were evaluated to determine if the results will be helpful to planners and regulators in reducing stormwater pollution.

Fecal bacterial contamination from human and animal wastes is a major cause of deteriorating water quality in receiving waters. A method that can distinguish sources of fecal matter would be the first step to solving this problem. The methods to detect human fecal pollution and differentiate it from other sources such as animals and soil are reviewed. The review includes microbial methods, especially those using molecular biology, and chemical methods.

The conclusions from each study are summarized.

Identification of land use with water quality data in stormwater using a neural network

A neural network model for identifying the various types of landuse with stormwater quality data was successfully developed using LADPW stormwater monitoring data, collected during 1996-2001. A Bayesian Network model was best and had ten water quality input variables, four neurons in the hidden layer and five landuse target variables. The statistical quality of the neural model was high. We obtained 92.3 and 94.5 percent of correct classifications, and 0.157 and 0.154 in the RMSE on the test and training files (173 cases). The model was used as a simulation tool to predict landuse type from highways stormwater monitoring data that was not used to develop the model. The simulations showed the sensitivity to classification and demonstrated a method to identify water quality variables that affect classification.

This research has demonstrated that a neural network can be used to classify landuses from water quality data, and that the technique can be automated. An approach for identifying opportunities for water quality improvement could be developed using this concept.

Utility of stormwater monitoring

The utility of three stormwater monitoring programs has been assessed based upon the programs' ability to accurately estimate the emissions for different classes of landuses, as well as other obvious benefits. The following conclusions are made:

1. Data collected by grab samples had much higher variability than composite samplers. The coefficients of variation (standard deviation divided by the mean) for the same parameters were generally 2 to 9 times higher for the grab samples. The variability suggests that composite samples should be collected, even if it means a reduction in the total number of samples or facilities that can be monitored.
2. The time required to analyze a sample must be commensurate with the intended use of the results. Beach water quality monitoring suffers from analysis time for indicator organisms. The data suggests that 70% of the beach postings are out of phase with the water quality parameter exceedence.
3. Metals (zinc, copper, lead, nickel) are potentially more useful to distinguish landuse patterns. Adding them to existing permits might double or triple the cost, but will add value to the resulting monitoring database.

Managing stormwater is a developing technology and much remains to be done. This research has shown that even with the limited experience we have so far, that there are improvements that can be made.

Methods to Identify Human and Animal Fecal pollution in Water: A Review

There is no easy, low cost method for differentiating between human and non-human sources of bacterial contamination. No single indicator or approach is likely to represent all the facets and issues associated with contamination of waterways with fecal matter. At present, the best hope of distinguishing fecal pollution of human and animal origin is an appropriate combination of indicators and water quality parameters. Statistical analyses of appropriate groups of methods offer the best possibility of identifying human sources. Unfortunately, relying on a combination of methods will probably require a longer period of analysis than relying on a single method. A combination of methods may be useful to determine sources in chronic situations as opposed to episodic events.