To the Graduate Council:

I am submitting herewith a thesis written by Tan, Ting Chow entitled "MODELING OF A NATURAL BIOREMEDIATION PROCESS USING MULTIVARIATE STATISTICAL APPROACH." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Chemical Engineering.

Tse-Wei Wang, Major Professor

We have read this thesis
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor
and Dean of The Graduate School

## STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at The University of Tennessee, Knoxville, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of the source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in her absence, by the Head of Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature 陳丁造

Date July 29, 1994

# MODELING OF A NATURAL BIOREMEDIATION PROCESS USING MULTIVARIATE STATISTICAL APPROACH

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Tan, Ting Chow

August 1994

Copyright © Tan, Ting Chow, 1994
All rights reserved

*To my parents, sisters, and brother.*

# ACKNOWLEDGEMENTS

The author wishes to express his appreciation to the faculty members of the Chemical Engineering Department of the University of Tennessee, Knoxville, for providing him financial support and sharing with him their experience and knowledge in the field of their expertise.

The author is indebted to his adviser, Dr. Tse-Wei Wang, for her guidance and encouragement. Without her help and support, this research would not have been possible. The author also thanks Dr. Duane A. Graves, Dr. Paul R. Bienkowski, Dr. Robert T. Ladd, Dr. John D. Birdwell, and Dr. John T. Wilson of U.S. EPA, Robert S. Kerr Research Laboratory in Ada, Oklahoma, for their guidance and advice.

Appreciation is also extended to Mr. Pramit Sarma, Ms. Bo Huang, Mr. Zhibin Duan, Dr. Roger Horn, Dr. Gaiye Zhou, Mr. Junghui Chen, Mr. Jovan Ilić, Mr. Taisuke Akasaka, Mr. Yongqing Jiang, Mr. Stephen Emge, Mr. Jim Grove, and Mr. Gary L. Grabow for their helps, suggestions, and friendships.

# ABSTRACT

When considering choosing natural bioremediation as the means for cleaning up a contaminated site, the decision makers would require strongly supported data that would allow them to be convinced that natural degradation is indeed taking place, at a fast enough pace to eventually returning the site to a level suitable for reuse in a reasonable amount of time. How does one go about analyzing the site monitoring data in order to provide such data is the subject of this thesis.

Multivariate statistical (MVS) analysis approach has been used to model the natural degradation kinetics of hydrocarbons at a gasoline spill site located at the Sleeping Bear Dunes National Lakeshore area in Michigan. The monitoring data is kindly provided by Dr. John Wilson of the Robert S. Kerr Laboratory of the Environmental Protection Agency at Ada, Oklahoma. The data covers a span of approximately two years in which a total of thirteen sampling data sets were taken, with eleven of which taken within one year of time. The data include the concentrations of the six BTEX compounds (benzene, toluene, ethylbenzene, *o*-xylene, *m*-xylene, and *p*-xylene) at 21 wells scattered over three sites, along with geochemical information, such as pH, temperature, alkalinity, phosphates, and methane concentrations.

The objective of analysis is to take all the concentration data in at one time and apply MVS analysis to the whole set and to attempt to arrive at an overall natural degradation rate for all the contaminant compounds taken as a whole. This would be much more superior to the traditional one-component-at-a-time analysis in that all the information would be used in the analysis and any correlation among the various contaminant compounds is taken explicitly into consideration by the MVS approach and

redundant information is cast away before final analysis is done. As a result, not only can one overall degradation rate constant be calculated that characterize the degradation kinetics of the entire set of contaminant compounds, but the segregation, if any, of the contaminants into distinct classes with different degradation kinetics between classes, can also be detected.

In analyzing the data, the residence time for the BTEX compounds that corresponds to the compounds sampled at each of the sampling points is first calculated from the hydraulic gradient information. Because an inert compound was not included in the monitoring effort until the last two sampling points, the concentration of benzene was used as the reference for correcting for dilution effect of all other compounds, for benzene was found in Wilson's study to have undergone the least amount of bioattenuation over the two years of monitoring time. In preparation for applying the MVS analysis approach, the concentration data is first column mean-centered (i.e. the column mean becomes zero), followed by scaling columnwise, by dividing each column by the variance of the corresponding column at the initial reference time point, such that any subsequent decrease in the concentration of the BTEX compounds can be discerned. Proper choice of the scaling factor is one of the most important, yet at the same time, the most subjective aspect of applying the MVS analysis approach.

MVS analysis performed on the Sleeping Bear Dunes data showed that indeed only one similarly behaving class of compounds is present. Assuming first-order degradation kinetics, an overall rate constant was also derived to be 0.0133/day for the site that is 9.1 meters down gradient from the spill site. This rate constant was arrived at by fitting ten points after MVS analysis by an exponential curve, representing that of first-order kinetics. Therefore, the calculated rate constant is relatively robust against errors in

sampling in any one or several individual concentration values.

Analysis tools such as these that can model the natural degradation rates of the contaminants at a site, can provide valuable information to decision makers as to the feasibility of relying on natural remediation to render the contaminated site suitable for reuse, in a reasonable amount of time. However, it is crucial that suitable monitoring data be taken so that applicable analysis can be done to yield useful information. Among the monitored data, it is imperative that detailed hydraulic gradient information, as well as the concentration of an inert compound that can be used as the reference compound to correct for dilution effect, be available throughout the study period.

# TABLE OF CONTENTS


CHAPTER PAGE

1. INTRODUCTION . . . . . 1

1.1 Motivation . . . . . 1

1.2 Statement of the Problem . . . . . 3

1.3 Advantage of Multivariate Statistical Approach . . . . . 3

2. SOIL SCIENCE PERSPECTIVE . . . . . 5

2.1 Advection and Diffusion . . . . . 6

2.2 Biodegradation Process . . . . . 8

2.3 BTEX Compounds . . . . . 12

3. MULTIVARIABLE METHODOLOGIES . . . . . 15

3.1 Singular Value Decomposition . . . . . 15

3.2 Application of Singular Value Decomposition to Environment Modelling and Analysis . . . . . 18

3.3 Simulation Studies to Show the Role of the Sum of the Square of the Singular Values . . . . . 23

3.4 Principal Component Analysis and Factor Analysis . . . . . 28

3.5 Two Examples of the Application of Principal Component Analysis and Factor Analysis . . . . . 30

3.5.1 Air Pollution Example . . . . . 30

3.5.2 One Independent Varying Factor Example . . . . . 38

4. MVS APPLICATION RESULTS . . . . . 41
4.1 Background Information on the Natural Remediation Data from the Sleeping Bear Dunes National Lakeshore . . . . . 41
4.2 Sleeping Bear Dunes National Lakeshore Case Study Using Multivariate Statistical Analysis Approach . . . . . 48
4.3 Discussion on the Sleeping Bear Dunes National Lakeshore Case Study . . 68

5. SUMMARY & RECOMMENDATIONS FOR FUTURE RESEARCH 75
5.1 Summary . . . . . 75
5.2 Recommendations for Future Research . . . . . 77

BIBLIOGRAPHY . . . . . 81

APPENDICES . . . . . 87
A. Data from the Sleeping Bear Dunes National Lakeshore . . . . . 88
B. Residence Times of TEX Compounds . . . . . 104
C. MATLAB Codes: Part I . . . . . 106
D. MATLAB Codes: Part II . . . . . 111
E. A Numerical Example of Mean-Centering and Variance-Normalizing A Matrix . . . . . 124

VITA . . . . . 126

# LIST OF TABLES


| TABLE | | PAGE |
|---|---|---|
| 3.1 | Singular values of **X** decreased to 80% at three consecutive time points. | 26 |
| 3.2 | Singular values of **X** when $x_1$ and $x_2$ are decreased to 50% and $x_3$, $x_4$, and $x_5$ are decreased to 80%, at three consecutive time points. | 27 |
| 3.3 | Principal components of the air pollution example. | 31 |
| 3.4 | Varimax rotated principal components. | 35 |
| 4.1 | Average residence times of the BTEX compounds. | 54 |
| 4.2 | Dilution correction factors determined by benzene and DMP at 9.1 and 21.3 sites for the TEX compounds. | 58 |
| 4.3 | First three PC's of the 9.1 m and 21.3 m sites. | 67 |

# LIST OF FIGURES


| FIGURE | | PAGE |
|---|---|---|
| 2.1 | Vapor pressure of BTEX compounds. | 13 |
| 3.1 | A scatter diagram of two correlated variables. | 16 |
| 3.2 | Mean-centering and variance-normalizing a matrix. | 20 |
| 3.3 | Shrinking a mean-centered and normalized matrix. | 24 |
| 3.4 | Singular value plots of a matrix decreased to 80% at each successive time points. The singular values are squared and then summed in (a). The individual singular values are plotted in (b). | 25 |
| 3.5 | Singular value plots of $\mathbf{X}$ when $x_1$ and $x_2$ are decreased to 50% and $x_3$, $x_4$, and $x_5$ are decreased to 80%, at each successive time points. The singular values are squared and then summed in (a). The individual singular values are plotted in (b). | 27 |
| 3.6 | Gabriel's plot of the ten air pollutants. | 33 |
| 3.7 | Gabriel's plot of the ten air pollutants (different view). | 33 |
| 3.8 | Gabriel's plot of the ten air pollutants in varimaxed frame. | 36 |
| 3.9 | Gabriel's plot of the ten air pollutants in varimaxed frame (different view). | 36 |
| 3.10 | The plot of the five highly correlated variables (a). Percentage information of the five correlated variables captured by their PC's (b). | 39 |
| 3.11 | Gabriel's plot of a data set that has one dorminant independent variable or PC. (b) is a different view of (a). | 40 |
| 4.1 | Layout of the monitoring boreholes with respect to the plume and the river. | 43 |
| 4.2 | Cross-sectional view of the monitoring boreholes. | 43 |

4.3 Apparent concentrations of benzene, toluene and ethylbenzene in mg/L at the spill, 9.1 m and 21.3 m sites. Each of these concentration profiles is the sum of the concentrations of a BTEX compound at each cluster of the monitoring boreholes. . . . . 44

4.4 Apparent concentrations of *m*-xylene, *o*-xylene, and *p*-xylene in mg/L at the spill, 9.1 m and 21.3 m sites. Each of these concentration profiles is the sum of the concentrations of a BTEX compound at each cluster of the monitoring boreholes. . . . . 45

4.5 Water table levels in terms of elevation above sea level. . . . . 47

4.6 Hydraulic velocity profile. . . . . 47

4.7 General outline of the Sleeping Bear Dunes National Lakeshore Case Study. 49

4.8 Distance travelled by groundwater and contaminants between an arbitrarily chosen initial time point and the indicated time point. The dashed lines are extrapolations. . . . . 51

4.9 Dilution corrected concentration of toluene, ethylbenzene, *m*-xylene, *o*-xylene, and *p*-xylene in mg/L at the 9.1 m and 21.3 m sites. . . . . 59

4.10 Sum of squares of the singular values (ssvd) vs. residence time (t) at 9.1 m site (a), and 21.3 m site (b). Their first-order degradation curve fits are ssvd = 133.8exp(-0.0133t) and ssvd = 15.0exp(-0.0040t) respectively. . 63

4.11 $\frac{\sum_{i=1}^{i=k} \sigma_i^2}{\sum_{j=1}^{j=n} \sigma_j^2}$ vs. $k$th PC at the 9.1 m and 21.3 m sites. . . . . 65

4.12 Gabriel's plots of the BTEX compounds at the 9.1 m and 21.3 m sites. . . 66

A.1 Data format for alkalinity, ammonia, benzene, calcium, ethylbenzene, iron, *m*-xylene, nitrate, *o*-xylene, oxygen, *p*-xylene, pH, phosphates, sulfate, temperature, TOC, and toluene. . . . . 89

# CHAPTER 1

# INTRODUCTION

## 1.1 Motivation

Both the U.S. Department of Energy (DOE) and the U.S. Environmental Protection Agency (EPA) have acknowledged the need for more effective and less costly environmental remediation technologies [19] [38]. The calls for more efficient remediation technologies are prompted by the huge estimated cost to carry out the U.S. environmental regulatory acts. It is estimated that over $30 billion has been spent to remediate Superfund (formally known as CERCLA or Comprehensive Environmental Response, Compensation, and Liability Act) sites during the past thirteen years by EPA and private parties and that another $500 billion could be needed for the remaining work [12]. Furthermore, EPA estimated that under another act, the Resource Conservation and Recovery Act (RCRA), corrective actions at 1500 to 3500 treatment, storage, and disposal facilities and up to 320,000 tank sites could be required [12].

Some of the current environmental remediation techniques include vacuum extraction, bioremediation, low and high temperature thermal treatment, removal/containment, solidification and encapsulation, asphalt incorporation, chemical treatment, soil washing, and vitrification [10]. Removal/containment technique, according to a survey of 169 remedial actions, is the most frequently employed remediation technique [29]. This method which involves the physical transfer of contaminated soil and ground water to a secure site such as a landfill or a treatment facility is very costly when the contam-

ination is extensive or when the contamination is deep underground [28]. Many other remediation techniques, although very effective in achieving their remediation objective, are prohibitively expensive and technically difficult.

Possibly the most cost-effective environmental restoration method is natural bioremediation. Natural bioremediation is also known as intrinsic bioremediation. Natural bioremediation is the name for in-situ biodegradation of contaminants by indigenous microbes at a contaminated site without external enhancement [18]. Natural bioremediation is an effective environment restoration method if the contaminants are biodegradable [28]. In general, the lighter and the more soluble the hydrocarbons are, the more rapidly biodegradable the hydrocarbons [30]. Even though natural bioremediation is a relatively slow remediation technique and is ineffective on heavy metals, toxic or recalcitrant organics [28], it is a very attractive alternative to other remediation techniques in situations where it is applicable due to its relative low cost and simplicity. Such situations could be the low and medium priority contaminated sites where the contaminants do not pose an immediate hazard to humans [18] and the contaminated sites where the technologies of active remediation methods are inapplicable.

Many contaminated sites may be naturally bioremediated given long enough time. Thus for sites where natural bioremediation is applicable, policy makers in the environmental cleanup industry will like to be able to predict the time required to restore their contaminated sites to certain safe levels suitable for reuse. These predictions can be done if a model on the rate of natural biodegradation for each of the contaminated sites is available. Since geochemical data are usually corrupted with noise, interrelated, and incomplete, statistical modeling approach is a very suitable approach to environmental modeling. And since geochemical data are multivariate by nature, this thesis takes the

multivariate statistical approach to modeling, and hopes to provide a tool that would provide some quantitative information on the overall rate of degradation of contaminants and the identification of the classes of compounds undergoing degradation, such that the policy makers would be able to make decisions as to whether bioremediation for the site of interest is feasible.

## 1.2 Statement of the Problem

A modeling technique that models the overall natural biodegradation rate and extrapolates the time required for the contaminants of a contaminated site to be naturally biodegraded to an acceptable level will be valuable to policy makers in the environmental cleanup industry who are pondering whether natural bioremediation or intrinsic bioremediation is the right remedial action for their contaminated sites. Developing such a modeling technique is the objective of this thesis. The problem statement is:

> *Through using all the available field data and through multivariate statistical methodologies, provide a quantitative measure and prediction of the degree of natural bioremediation on a contaminated soil site.*

## 1.3 Advantage of Multivariate Statistical Approach

Traditionally, when available data contain many variables, with possible correlation and noise among the variables, the variables have been analyzed one at a time to assess the rise and fall of individual variable over time. As the number of variables becomes large and when the variables are correlated, such one-at-a-time analysis becomes confusing and inconclusive. An approach that takes into account the interrelatedness of the variables in the presence of noise and uncertainty is the multivariate statistical approach (MVS).

Through MVS, the original data can be reduced and be represented by a fewer number of principal variables. A principal variable is a new variable that is a linear combination of the original variables [2]. The principal variables of a set of data are uncorrelated to one another [2]. The principal variables' rise and fall over time can provide an overall measure of the rise and fall of the entire set of the original variables over time. The relative weighting of the original variables that contribute to the principal variables can also be assessed, so as to give an idea of the contribution of each original variable in the overall change of things.

In summary, the MVS approach analyzes a set of data as a whole in order to assess the overall process trend over time. This is superior to the traditional one-at-a-time analysis in that it gives a coherent, coordinated, and concise indication of the process dynamics.

# CHAPTER 2

## SOIL SCIENCE PERSPECTIVE

During natural degradation, a contaminant degrades into several other lower molecular weight compounds and these compounds and the parent contaminant move and spread in the contaminated site by natural advection and diffusion. The dynamics of these processes complicate the estimation of the rate and extent of the natural degradation process. This chapter provides an understanding on the biogeochemical reactions that control the movement and fate of organic contaminants in soil. Metallic contaminants are not discussed because they are not remediable by natural biodegradation.

The changes in the concentration of contaminants in soil are attributable to migration and biodegradation. A general equation describing this phenomenon is derived in reference [5]. This equation, based on conservation of mass is:

$$\frac{\delta C_A}{\delta t} + (\nu \bigtriangledown C_A) = \mathcal{D}_{Am} \bigtriangledown^2 C_A + R_A \tag{2.1}$$

where

$\mathcal{D}_{Am}$ = effective binary diffusivity of $A$ in a multicomponent mixture

$\bigtriangledown$ = "nabla" or "del" operator

= $\mathbf{i}\frac{\delta}{\delta x_1} + \mathbf{j}\frac{\delta}{\delta x_2} + \mathbf{k}\frac{\delta}{\delta x_3}$ in rectangular coordinates; i, j, k are unit vectors

$\bigtriangledown^2$ = Laplacian operator = $\frac{\delta^2}{\delta x_1^2} + \frac{\delta^2}{\delta x_2^2} + \frac{\delta^2}{\delta x_3^2}$ in rectangular coordinates

$R_A$ = chemical or biological reaction term of chemical $A$

$\nu$ = velocity of the advective migration of contaminant $A$ (m/day)

$C_A$ = concentration of $A$ (mol/L)

$t$ = time (day)

This equation assumes $\mathcal{D}_{Am}$ to be a constant, i.e., it is independent of direction and concentration.

## 2.1 Advection and Diffusion

The migration of contaminants in soil can be characterized as being advective or diffusive. Advection is the process by which groundwater in motion carries dissolved solutes. Diffusion is the process by which dissolved solutes in water move from areas of higher concentration to areas of lower concentration [36]. The migration process of the contaminants is affected by the physicochemical properties of the contaminants [36]. Examples of physicochemical properties are solubility, vapor pressure, partition coefficient, sorption, desorption, and volatilization [36].

One of the most common approaches used to describe advection is Darcy's law [36]:

$$\nu = \frac{K}{n}\frac{\Delta h}{\Delta x}$$

where

$\frac{\Delta h}{\Delta x}$ = hydraulic gradient (m/m)

$\nu$ = average linear velocity of groundwater (m/day)

$n$ = effective porosity

$K$ = hydraulic conductivity (m/day)

Darcy's law is a well tested law and is generally useful [31]. However, its linearity assumption fails when the hydraulic gradient is too high or too low [31]. Darcy's law is

limited to steady state flow [31].

For systems with small advection ($\nu \approx 0$) and no chemical change ($R_A \approx 0$), Eq. 2.1 becomes the Fick's second law of diffusion:

$$\frac{\delta C_A}{\delta t} = \mathcal{D}_{Am} \nabla^2 C_A \tag{2.2}$$

The implicit assumption in Eq. 2.1 and Eq. 2.2 is that the medium of diffusion is continuous. The solution to Eq. 2.2 has to be interpreted in an average sense when it is used to describe a diffusion process in soil because soil is not a continuous medium.

Gravitational and capillary forces are the major forces that cause the migration of contaminants in the general system described by Eq. 2.1 while the contaminants are above the groundwater level. The contaminants descent due to gravity and spread horizontally due to the capillary force.

When the contaminants reach and mix with the groundwater, the movements of the contaminants resemble the movement of the groundwater. However, because contaminants can be sorped onto naturally occurring organic carbon, soil particles, and other materials in an aquifer, contaminants usually move at slower rates than the groundwater [35]. The most commonly utilized and accepted equation that estimates the migration rate of a contaminant in a soil groundwater system is [13]:

$$\nu_c = \nu(1 + K_d \frac{B}{n})^{-1} \tag{2.3}$$

where

$K_d$ = distribution (sorption) coefficient ($cm^3/g$)

$\nu$ = average linear velocity of groundwater (m/day)

$\nu_c$ = average linear velocity of contaminant (m/day)

$B$ = soil bulk density (g/cm$^3$)

$n$ = soil total porosity

The term $(1 + K_d\frac{B}{n})^{-1}$ is known as the retardation factor [13]. The tendency of a chemical to partition between particles containing naturally occurring organic carbon and water is measured by $K_d$ [35]. $K_d$ is defined as the ratio of adsorbed chemical per unit weight of organic carbon to the aqueous solute concentration [35]. $B$, the average soil bulk density, is the mass (weight) of a known soil volume compared to the mass of an equal volume of water [46]. Both the solid and the void space fractions are included in the soil volume [46]. Other characteristics of contaminants that influence the rate of dispersion of the contaminants in water that are not taken into consideration in Eq. 2.3 are solubility, aquifer grain composition, and temperature [48].

## 2.2 Biodegradation Process

In the context of contaminant hydrogeology, $R_A$ in Eq. 2.1 refers to the rate of chemical transformation of species $A$ by oxidation, reduction, hydrolysis, halogenation, dehalogenation, and photochemical breakdown [36]. However, under the environmental conditions found in most aquifers, $R_A$ is dominated by biological reactions [6]. Biodegradation, as defined by reference [18], is the complete mineralization of organic contaminants to water, carbon dioxide, inorganic compounds, and cell protein, or the transformation of organic contaminants to other organic compounds.

Biodegradation is fundamentally an electron transfer or oxidation-reduction process [18]. Electron transfer is catalyzed by enzymes produced by microorganisms [18]. In return, the microorganisms receive energy from the oxidation-reduction process. Biodegra-

dation is typically a stepwise process involving a variety of different enzymes and species of organisms [18]. The terminal electron acceptors for metabolism could be oxygen (aerobic reaction), nitrate (denitrification or nitrate reduction), sulfate (sulfate reduction), ferric iron (ferric iron reduction), and bicarbonate (methanogenesis), depending on whether the organisms have the appropriate enzyme systems [18]. When oxygen is the terminal electron acceptor, the biodegradation is called aerobic. Anaerobic biodegradation occurs at the absence of oxygen.

In the presence of oxygen, aerobic biodegradation is usually the favored biodegradation [18] and almost all hydrocarbons can be biodegraded aerobically [6]. An example of the aerobic biodegradation of toluene can be represented by the theoretical reaction [6]:

$$C_6H_5CH_3 + 9O_2 \longrightarrow 7CO_2 + 4H_2O + Energy$$

Due to the limited solubility of oxygen in water, around 6 to 12 mg/L [41], oxygen in the groundwater is usually used up before the remediation is completed. Because of this, much of natural bioremediation processes is anaerobic. While the ultimate products of aerobic metabolism and methanogenesis are carbon dioxide and water [18], the end products of anaerobic metabolism are carbon dioxide, stable ions (e.g. $HCO_3^-$), methane, and water [8], depending on whether the anaerobic biodegradation is via nitrate reduction, ferric iron reduction, sulfate reduction, or methanogenesis. When the oxygen supply is expended and nitrate is present, the nitrate will be used by some facultatively anaerobic microorganisms as the terminal electron acceptor [6]. Theoretical biodegradation reaction by nitrate reduction of toluene is [6]:

$$C_6H_5CH_3 + 6NO_3^- \longrightarrow 7CO_2 + 4H_2O + 3N_2 + Energy$$

When both oxygen and nitrate are unavailable, oxidized ferric iron may be used by

certain subsurface microorganisms as the terminal electron acceptor [6]. A possible reaction between ferric hydroxide and toluene is [6]:

$$C_6H_5CH_3 + 36Fe(OH)_3 \longrightarrow 7CO_2 + 36Fe^{+2} + 72OH^- + 22H_2O + Energy$$

Certain organics are known to biodegrade by sulfate reduction. At present, our knowledge on the effects of sulfate reduction is still very incomplete [6]. For toluene, theoretical biodegradation by sulfate reduction is [6]:

$$C_6H_5CH_3 + 4.5SO_4^{=} + 3H_2O \longrightarrow 2.25H_2S + 2.25HS^- + 7HCO_3^- + 0.25H^+ + Energy$$

Methanogenesis, which is another type of anaerobic biodegradation, can be approximated by the following theoretical reaction [6]. In this example, water is the terminal electron acceptor.

$$C_6H_5CH_3 + 5H_2O \longrightarrow 4.5CH_4 + 2.5CO_2 + Energy$$

Because the natural bioremediation process used in the case study in Chapter 4 was predominantly due to methanogenesis, methanogenesis is discussed in some detail here. It was observed that methanogens are very sensitive to both pH and temperature [56]. The optimum pH range for methanogenesis is between 7.0 and 7.2 [32]. The production rates of methane drop sharply below 6.6 and stop at 6.2 [32]. The optimal temperature for methanogenesis is 40°C [27] [62] [7]. Many of the nutrients or electron acceptors for methanogenesis become toxic to methanogens when the nutrients are present in high concentration. Ammonia which is the major nitrogen source for methanogens [56] can inhibit methanogenesis. The inhibition starts at concentrations between 100 and 200 mM of ammonia when the pH is greater than 7.4. The ammonium ion becomes toxic regardless of pH when the concentration of ammonia is greater than 200 mM [33].

Although sulfide is a sulfur source to methanogens, it is inhibitory to methanogenesis when present in high concentrations [61]. In addition, according to Atlas [3], nitrate and sulfate are also inhibitors to methanogenesis when present in high concentration. However, nitrate does not usually exist in high enough concentration in nature to be an inhibitor to methanogenesis [3]. Detailed microbiological aspects of methanogenesis can be found in reference [20].

Many factors contribute to the varied rate of biodegradation of contaminants in soils from different sites. These factors are indigenous soil microbial population, contaminant variety and concentration, soil structure, nutrient availability, moisture content, temperature, pH, oxygen availability, and whether the site has been previously contaminated [45] [3]. Even on one site, the rate of degradation does not remain constant. There tend to be more biodegradation after a rain because the high humidity caused by the rain increases the amount of hydrocarbons in aqueous milieu where the microbes live and therefore allows greater contact between the hydrocarbons and the microbes. Therefore, comparing the relative rates of biodegradation across sites, or even at the same site but across time, should be done very carefully and the accompanying environmental factors should be taken into account. Only for sites exhibiting similar environmental profiles, can extrapolation of result from one study to another be valid.

In a study of nine sites, McGill reported that under the assumption of first-order kinetics, between 0.1% and 1% of the oil present at each site biodegraded per day [34]. These rates appear to be reasonable for reasonably biodegradable hydrocarbons or hydrocarbon mixtures that are slowly released to soil under mesophilic temperature [3]. Unfortunately, detailed descriptions of the conditions under which the estimated rates were determined were not given in McGill's report [3].

## 2.3 BTEX Compounds

Because BTEX compounds are common contaminants and are the contaminants in the case study in Chapter 4, this section is devoted to the background information relevent to the bioremediation of BTEX compounds. BTEX, which stands for benzene, toluene, ethylbenzene, *o*-xylene, *m*-xylene, and *p*-xylene, is an important group of aromatic hydrocarbons found in gasoline [45], jet fuel, kerosene and to a lesser extent, diesel. BTEX compounds are important because they are relatively soluble in groundwater, mobile, and are or suspected to be carcinogenics.

Certain physical properties of BTEX compounds facilitate understanding of the behavior of these compounds as groundwater contaminants. BTEX compounds are liquids at 25°C. Their boiling temperatures range from 80.1°C for benzene to 144.4°C for *o*-xylene and their melting temperatures range from -95°C for toluene to 13.3°C for *p*-xylene [35].

The vapor pressures of BTEX compounds from reference [35] are illustrated in Fig. 2.1. The vapor pressures of the isomers of xylene are very similar to one another, thus they are combined in Fig. 2.1. As is shown in the figure, the vapor pressures of xylene and ethylbenzene are almost identical.

Among the BTEX compounds, benzene has the highest water solubility at 1780 ± 45 mg/kg $H_2O$, followed by toluene which has the solubility of 515 ± 17 mg/kg $H_2O$ [45]. Ethylbenzene and the three isomers of xylene have similar water solubility. Their water solubilities are in the range of 146 to 175 ± 17 mg/kg $H_2O$ at 25°C [45].

When BTEX compounds are dissolved in groundwater, benzene disperses more easily than toluene, and toluene disperses more easily than the xylenes [48]. This order of

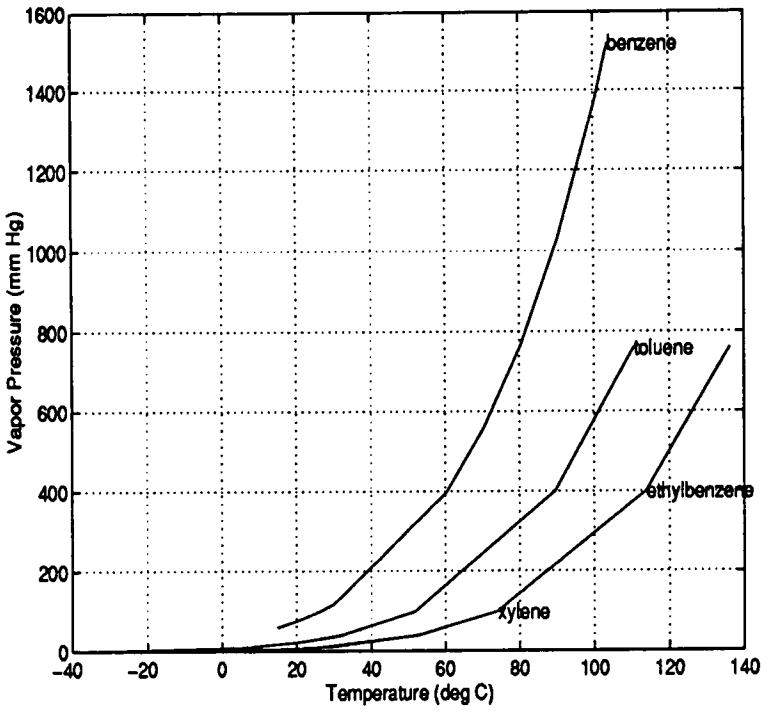



Figure 2.1: Vapor pressure of BTEX compounds.

dispersibility of the BTEX compounds in groundwater are compatible with the experimental values of $K_d$, the distribution sorption coefficient stated in Eq. 2.3. Eq. 2.3 says that the retardation factor to the movement of a contaminant in moving groundwater is proportional to size of $K_d$. The soil/sediment coefficients, $K_{oc}$, of benzene, toluene, ethylbenzene, *m*-xylene, *o*-xylene, and *p*-xylene are 38 $cm^3/g$, 90 $cm^3/g$, 210 $cm^3/g$, 220 $cm^3/g$, 128.8 $cm^3/g$, and 204.2 $cm^3/g$ respectively [45] [44] [1] [35]. The soil/sediment coefficients are related to the distribution coefficients, $K_d$, by [35]:

$$K_{oc} = \frac{K_d}{f_{oc}} \tag{2.4}$$

where $f_{oc}$ is the fraction of naturally occurring organic carbon (e.g. humus) in soil.

Analysis done by Cozzarelli et al. [8] on field data suggests that BTEX compounds are biologically transformed to organic acid intermediates in anoxic (anaerobic) environments via methanogenesis. The oxidized intermediates of benzene, toluene, *o*-xylene, *m*-xylene, and *p*-xylene are phenol, benzoic acid, *o*-toluic acid, *m*-toluic acid, and *p*-toluic acid respectively [8]. These BTEX compounds are first oxidized to phenols or organic acids, then transformed to long-chain volatile fatty acids, which are then metabolized to methane and carbon dioxide [18]. The biodegradability and degradation rates for each of these compounds under anaerobic conditions can vary considerably [18].

Laboratories studies have found that under nitrate reduction conditions, the biodegradation of BTEX compounds happens in sequence. Toluene is the first to biodegrade under these conditions, follow by *p*-xylene, *m*-xylene, ethylbenzene, and *o*-xylene [41]. Mixed results were reported for the biodegradability of benzene by nitrate reduction [41].

Similar to the results of the nitrate reduction studies, experiments done by Edwards et al. [15], [16], and [17] conclude that BTEX biodegrade by sulfate reduction is also a sequential process in order of toluene, *p*-xylene, *o*-xylene, and benzene [41].

# CHAPTER 3

# MULTIVARIABLE METHODOLOGIES

In this chapter, three multivariable methodologies are introduced and briefly discussed. The methodologies are Singular Value Decomposition, Principal Component Analysis, and Factor Analysis.

## 3.1 Singular Value Decomposition

Singular Value Decomposition (SVD) is one of the major matrix factorization methods. Its applications are many and growing [47]. A few examples of the application of SVD, such as in image processing, polar decomposition, determination of the effective rank of a matrix, least square problems, and optimization are cited by Strang [47].

Any matrix $\mathbf{X}$ of dimension $m$ by $n$ can be factored by SVD into the product of three matrices with unique properties:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{T}}$$

where the columns of the $m$ by $m$ $\mathbf{U}$ (the left singular vectors) and the $n$ by $n$ $\mathbf{V}$ (the right singular vectors) are the eigenvectors of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$ respectively. $\mathbf{U}$ and $\mathbf{V}$ are chosen to be orthonormal, i.e., their columns vectors are mutually orthogonal and are of unit length. The $r$ nonzero elements on the diagonal of the $m$ by $n$ $\Sigma$ are the positive square roots of the nonzero eigenvalues of either $\mathbf{X}\mathbf{X}^T$ or $\mathbf{X}^T\mathbf{X}$ [47]. The positive integer $r$ denotes the true rank of $\mathbf{X}$. The diagonal elements in $\Sigma$, which are represented as $\sigma_i$, $i = 1, 2, ..., r$, are called the singular values of $\mathbf{X}$ and they are arranged in descending

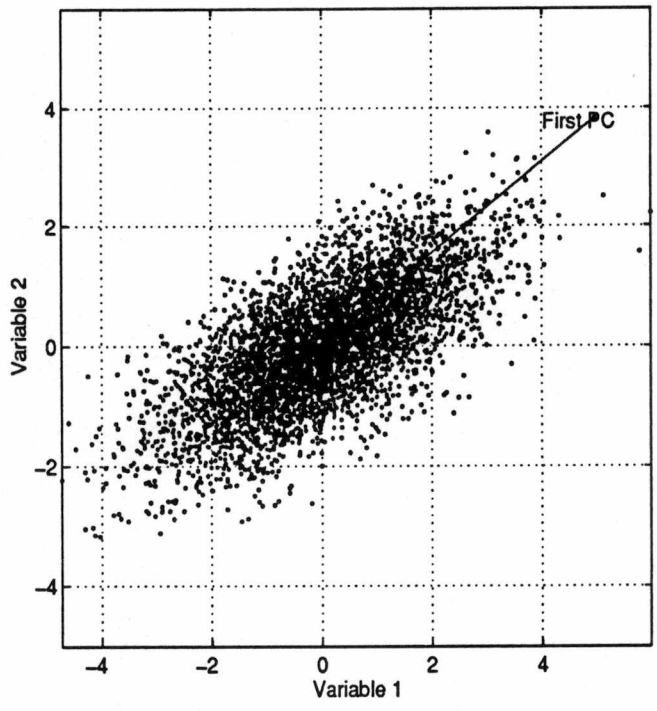



Figure 3.1: A scatter diagram of two correlated variables.

order in $\Sigma$.

The connections between the singular values and singular vectors of a $m$ by $n$ matrix $\mathbf{X}$, whose rows represent observations and whose columns represent variables, is examined through geometric visualization. If all the variables in $\mathbf{X}$ are normally distributed, the scatter diagram of the observations of $n$ variables will form a cluster that can be enclosed by a $n$-dimensional hyperellipsoid [42]. An example of the scatter diagram when $n$ is two is Fig. 3.1. The first right singular vector corresponding to the largest singular value of $\mathbf{X}$ defines the direction of the major axis of the hyperellipsoid. The length of the major axis, i.e., the distance from the center of the hyperellipsoid to the point on the

hyperellipsoid furthest away from its center, is equal to the largest singular value of $\mathbf{X}$ [42]. This distance is the maximum variance of the data points [42]. It can be shown that this variance is the sum of the squared projections of data points onto the first singular vector that is the largest of all possible singular vectors of $\mathbf{X}$. [42]. The right singular vector corresponding to the second largest singular value of $\mathbf{X}$ defines the direction of the second major axis of the $n$-dimensioned hyperellipsoid. The second largest singular value is equal to the second largest variance of the data points along any singular vector. Generalizations of the above apply to the singular values of subsequent magnitude and their corresponding right singular vectors. The major axes are linearly independent and orthogonal of one another. The first singular value, or the maximum singular value, $\sigma_1$, is a measurement of the "size" of the matrix $\mathbf{X}$, in that if $\mathbf{X}$ is considered to be a transformation matrix, then a circle of vectors of unit length centered around the origin are transformed into an elipsoid of vectors with the vector furtherest away from the origin at a distance $\sigma_1$ from the origin. In other words, the singular values of a matrix can be thought of as the extent of the dispersion of the $m$ scattered points represented by the $m$ rows of the data matrix $\mathbf{X}$, as captured by an $n$-dimensional ellipsoid with major and minor axis equal to the singular values of the matrix, $\mathbf{X}$.

The simulated variables used in the visualization above are chosen to be normally distributed for ease of interpretation and visualization. The variables need not be normally distributed for the applications of SVD, Principal Component Analysis (PCA) and Factor Analysis (FA). A formal distribution assumption would be needed if statistical tests or confidence intervals are used [2]. Statistical tests and confidence intervals are not used here.

## 3.2 Application of Singular Value Decomposition to Environment Modelling and Analysis

SVD was applied to a simulated location-versus-contaminant concentration matrix **X**, whose rows represent the sampling locations of the contaminants and whose columns represent the contaminant species, to achieve two objectives:

1. Grouping contaminants with similar behavior;

2. Quantifying the overall changes in each group of contaminants with similar behavior.

Principal Component Analysis and Factor Analysis were used to group contaminants with similar behavior. These methods will be discussed and an example presented in a later section in this chapter. Before **X** is variance-normalized, a singular value of **X** is a measure of the magnitude of a linearly independent factor of **X**. A linearly independent factor of the location-versus-contaminant concentration matrix **X** refers to a group of contaminants with similar behavior. The magnitude of the singular value will be largely dominated by the contaminant with the highest concentration in a group of similarly behaving contaminants. This can be understood by recalling the geometric visualization on singular values and singular vectors in the previous section that the $i$th largest singular value is the length of the $i$th major axis that starts from the center of the hyperellipsoid in the scatter diagram (or the variables versus variables plot) of the matrix to the point on the hyperellipsoid furthest away from its center. After **X** is variance-normalized however, the magnitude of the singular values is no longer directly related to the magnitude of the concentration of the contaminants, but is rather a measure of the dispersion of the

data along the first singular vector.

Each column of the concentration matrix **X** has to be mean-centered and variance-normalized to enhance the ability of SVD to detect groups of contaminants with similar behavior. A mean-centered column vector is a vector with the sum of its elements equal to zero and a variance-normalized column vector is a vector with unit length. Fig. 3.2 demostrates how the ability of SVD to group contaminants with similar behavior is enhanced by preprocessing **X** with mean-centering and variance-normalization. Let **X** be twenty by five. The five variables of **X** are named $x_1$, $x_2$, $x_3$, $x_4$, and $x_5$. In the demostration represented by Fig. 3.2, variables $x_1$ and $x_2$ are a set of two highly correlated random variables, and variables $x_3$, $x_4$, and $x_5$ are another set of highly correlated random variables. The plots in the first column of Fig. 3.2 are the plots of the magnitude of these five variables at twenty observations. The x-axes of the plots in the first column of Fig. 3.2 are labeled 'Row #' because each observation corresponds to a row of **X**. In the plots in the first column of Fig. 3.2, the dashed lines represent $x_1$ and $x_2$, and the solid line lines represent $x_3$, $x_4$, and $x_5$. Only two lines are apparent in these plots because $x_1$ is very similar to $x_2$, and $x_3$, $x_4$, and $x_5$ are very similar to one another. The plots in the second column of Fig. 3.2 are the plots of $\frac{\sum_{i=1}^{i=k}\sigma_i^2}{\sum_{j=1}^{j=5}\sigma_j^2}$ versus the $k$th principal component (PC). The $k$th PC is a new variable that is the linear combination of the original variables and it corresponds to the $k$th right singular vector in the SVD of the matrix **X**. In this example, the original variables are $x_1$, $x_2$, $x_3$, $x_4$, and $x_5$. A PC is linearly independent and orthogonal to other PC's of the same data matrix, i.e. **X** in this example. Each PC is represented by each column of the right singular matrix **V**. The $\frac{\sum_{i=1}^{i=k}\sigma_i^2}{\sum_{j=1}^{j=5}\sigma_j^2}$ versus the $k$th PC plots inform us how much of **X** is represented by the first $k$ PC's. The plots in the last column of Fig. 3.2 are the scattered

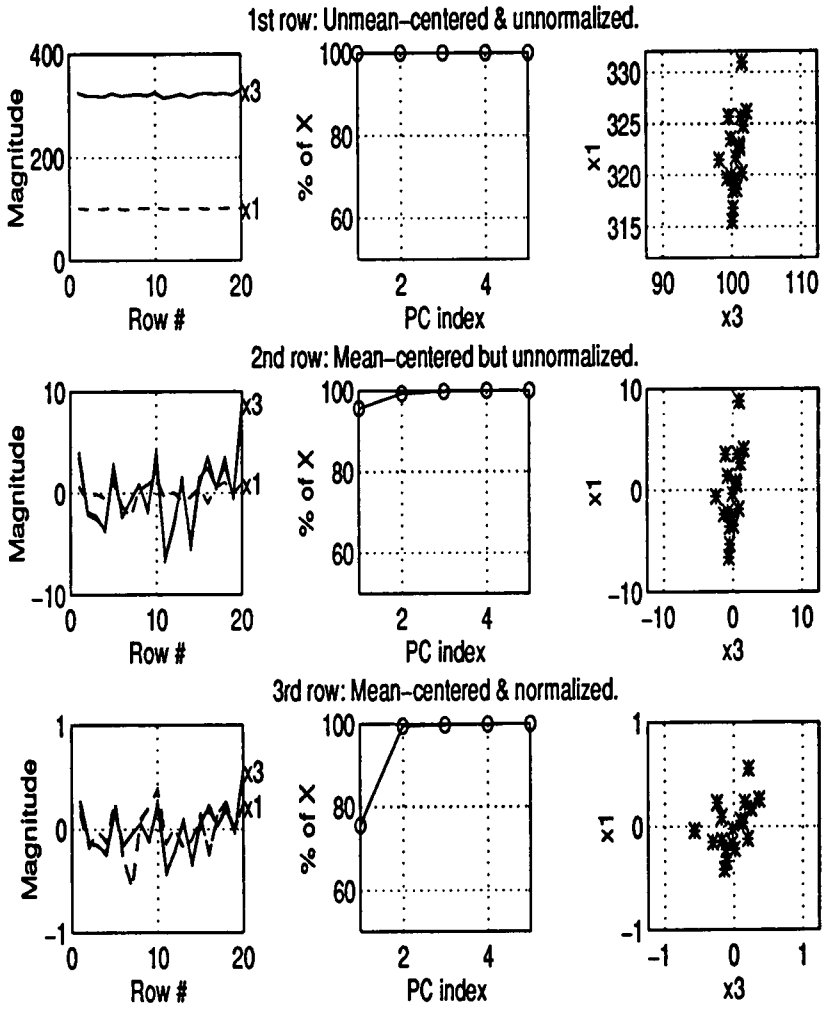



Figure 3.2: Mean-centering and variance-normalizing a matrix.

diagrams or plots of $x_1$ to $x_3$. The first row of plots in Fig. 3.2 are the various plots of the unmean-centered and unvariance-normalized **X**. The second row of plots in the same figure are the various plots of the mean-centered but unvariance-normalized **X**. The plots of both mean-centered and variance-normalized **X** are presented in the third row of Fig. 3.2. The **X** matrix, before and after mean-centering and variance-normalization, illustrated in Fig. 3.2 is shown in Appendix E.

In the first row when the matrix **X** is unmean-centered and unvariance-normalized, it appears that the first principal component (PC) alone contains all the information in **X**, when in reality, there are two independent sets of information contained in **X**, but is masked because of the large difference in the relative scale of the numbers in the two groups. The effects of mean-centering are apparent when we compare the plots in the first and the second rows. Note that after mean-centering, the center of gravity of the set of scattered points is now at the origin. Further examination on the differences in the first and the second rows of plots leads to the conclusion that the difference in the relative magnitude of the two sets of independent variables, i.e. the set $x_1$ and $x_2$, and the set $x_3$, $x_4$, and $x_5$, hides the independent variations in the two sets of variables. It appears that there is only one independent varying class in **X** when in fact there are two. The two order of magnitude difference in scale between the group of $x_1$ and $x_2$ and the group of $x_3$, $x_4$, and $x_5$, masks the independence of the two groups of variables. The plots in the third row show that the variance-normalization of **X** improves the detection of the two sets of independent variables by SVD. This improvement is due to the standardization of the magnitude of the variations of the variables when the variables are variance-normalized. Thus the set of independent variables containing $x_3$ can no longer hide the set of independent variables containing $x_1$ due to the larger magnitude of variations in

the former after variance-normalization. Notice that by variance-normalization, all the data points in the $x_1$ versus $x_3$ plot in the third row of Fig 3.2 are now located within the -1 to 1 range for both the x- and y-axes.

There are many ways other than singular values to determine the magnitude of a matrix. However, singular values is a particularly suitable mean of measuring the magnitude of a matrix of contaminant concentration data collected from the environment because these data are highly correlated and SVD that determines the singular values explicitly identifies this correlation and assigns a magnitude, i.e. a singular value, to each of the linearly independent factor that is in the matrix to reflect its relative importance in accounting for the dispersion of the data points. Another name for the linearly independent factors is principal components (PC). They are represented by the column vectors of the right singular matrix V. When correlation exists among the variables, the original large amount of data can usually be reduced to and be represented by just a few mutually orthogonal principal components without loss of substantial information, thus redundant information is discarded.

Using singular values as a measure of the magnitude of a matrix is not new. Safonov [43] explains the uses of singular values in the context of control system. In control systems, researchers are interested in characterizing bounds of a system's stability, sensitivity to noise, and disturbance attenuation properties. As such, only the maximum and the minimum singular values of the matrices that represent these properties are of interest. When used in environment modeling and analysis it is the first few dominant singular values that are of interest, because these first few singular values represent the amount of information that is contained in the original matrix. Since the square of the $i$th singular value or $\sigma_i^2$ that corresponds to the $i$th PC is equal to the variance of the

data points along the direction of the $i$th PC, the fraction of the amount of information captured by the $i$th PC is $\frac{\sigma_i^2}{\sum_{j=1}^{j=n} \sigma_j^2}$ where the denominator represents the total variance contained in the data set and the numerator represents the dispersion captured by the $i$th PC only, and $n$ is the total number of the original variables. The cumulative amount of information captured by the first $k$ PC's is $\frac{\sum_{i=1}^{i=k} \sigma_i^2}{\sum_{j=1}^{j=n} \sigma_j^2}$ [57], where $n$ is the total number of PC's.

## 3.3 Simulation Studies to Show the Role of the Sum of the Square of the Singular Values

Bioremediation of a site contaminated by five contaminants is simulated. Let **X** be a concentration matrix whose columns represent contaminants and whose rows represent the samples collected at the contaminated site over time. The size of the matrix is 20 by 5.

In the first case, all the concentrations of the contaminants $x_1$, $x_2$, $x_3$, $x_4$, and $x_5$ are decreased to 80% at each succeeding time point (each succeeding row). The first two time points are shown in Fig 3.3. The decrease in the concentrations of the contaminants is discernable in the plots at the third column of Fig 3.3 by the decrease in the dispersion of the variables, and by the singular value plots of **X** in Fig 3.4. The sum of the squares of the singular values is a representation of the overall dispersion of the variables in the scatter diagram.

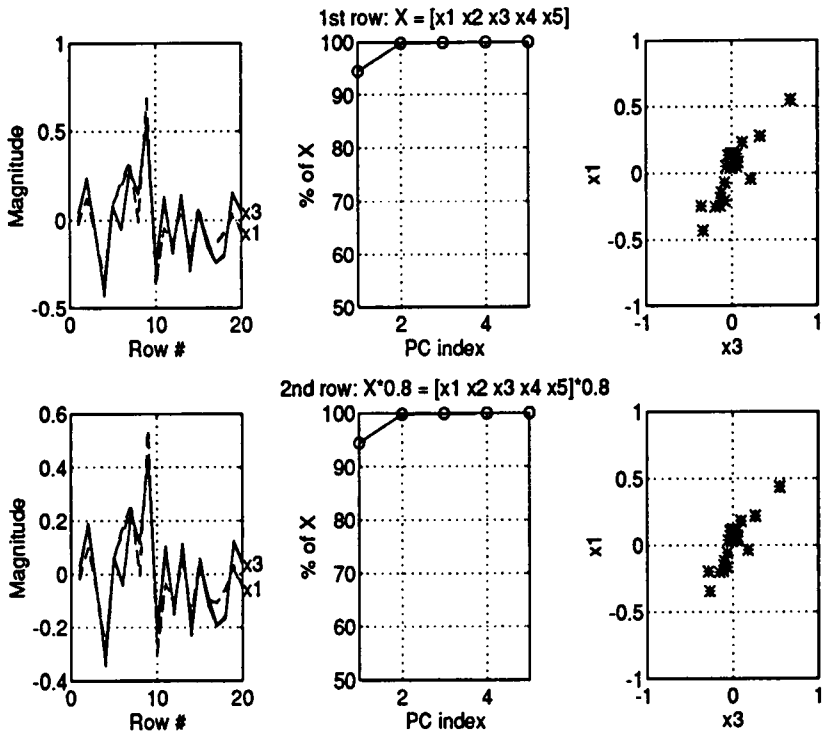



Figure 3.3: Shrinking a mean-centered and normalized matrix.

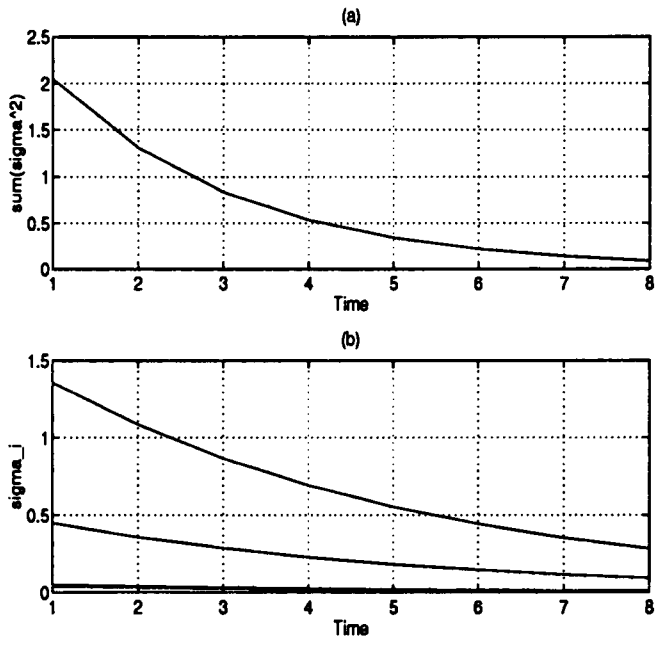



Figure 3.4: Singular value plots of a matrix decreased to 80% at each successive time points. The singular values are squared and then summed in (a). The individual singular values are plotted in (b).

In Table 3.1, $s_1$, $s_2$, and $s_3$ are the sets of singular values of $\mathbf{X}$ that also decreased to 80% at three consecutive time points respectively. Because $\mathbf{X}$ is decreased to 80% at each consecutive time points, $\frac{s_2}{s_1}$ and $\frac{s_3}{s_2}$ are both 0.8.

| $s_1$ | $s_2$ | $s_3$ | $\frac{s_2}{s_1}$ | $\frac{s_3}{s_2}$ | $\frac{s_3}{s_1}$ |
|---|---|---|---|---|---|
| 2.1191 | 1.6953 | 1.3562 | 0.8 | 0.8 | 0.64 |
| 0.7022 | 0.5618 | 0.4494 | 0.8 | 0.8 | 0.64 |
| 0.0824 | 0.0659 | 0.0527 | 0.8 | 0.8 | 0.64 |
| 0.0752 | 0.0602 | 0.0481 | 0.8 | 0.8 | 0.64 |
| 0.0629 | 0.0503 | 0.0403 | 0.8 | 0.8 | 0.64 |

Table 3.1: Singular values of $\mathbf{X}$ decreased to 80% at three consecutive time points.

In the second case, the concentrations of the contaminants $x_1$ and $x_2$ are decreased to 50% and $x_3$, $x_4$, and $x_5$ are decreased to 80% for each succeeding time point. The singular value plots of $\mathbf{X}$ in this case are shown in Fig 3.5. Table 3.2 shows that the values of $\frac{s_2}{s_1}$ and $\frac{s_3}{s_2}$ are between 0.5 and 0.8.

This simulation shows that the sum of the squares of the singular values of a matrix is an indicator of the overall dispersion of the variables in their scatter diagram. When the total dispersion decreases, so does the sum of the squares of the singular values. Thus if contaminants are being degraded naturally by the bacteria present in the soil, the dispersion of the concentration matrix, after some proper pretreatment, over time should also decrease.

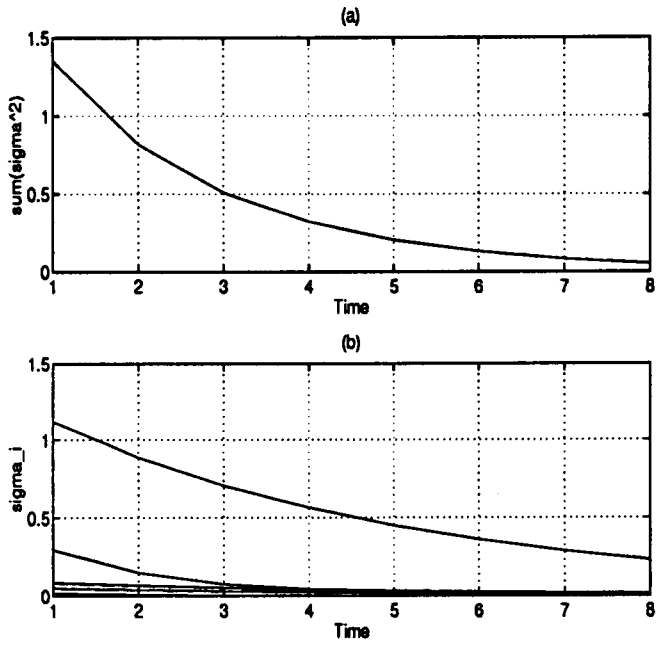



Figure 3.5: Singular value plots of $\mathbf{X}$ when $x_1$ and $x_2$ are decreased to 50% and $x_3$, $x_4$, and $x_5$ are decreased to 80%, at each successive time points. The singular values are squared and then summed in (a). The individual singular values are plotted in (b).

| s1 | s2 | s3 | $\frac{s_2}{s_1}$ | $\frac{s_3}{s_2}$ | $\frac{s_3}{s_1}$ |
|---|---|---|---|---|---|
| 1.9615 | 1.4388 | 1.1207 | 0.7335 | 0.7789 | 0.5713 |
| 1.0583 | 0.5770 | 0.2965 | 0.5452 | 0.5138 | 0.2801 |
| 0.1413 | 0.1068 | 0.0842 | 0.7560 | 0.7883 | 0.5959 |
| 0.0994 | 0.0678 | 0.0506 | 0.6819 | 0.7459 | 0.5087 |
| 0.0517 | 0.0321 | 0.0175 | 0.6208 | 0.5439 | 0.3377 |

Table 3.2: Singular values of $\mathbf{X}$ when $x_1$ and $x_2$ are decreased to 50% and $x_3$, $x_4$, and $x_5$ are decreased to 80%, at three consecutive time points.

## 3.4 Principal Component Analysis and Factor Analysis

Principal Component Analysis (PCA) and Factor Analysis (FA) are very similar to each other in that both of these techniques are used to examine the interrelationships among a set of correlated variables [2]. The major objective of PCA is to select a number of components that explain as much of the total variance as possible [2]. The major purpose of FA on the other hand is to derive from the data easily interpretable common factors [2]. Both PCA and FA are related to SVD.

PCA can be summarized as a method of transforming the original variables into new uncorrelated variables [2]. The new uncorrelated or linearly independent variables are the principal components (PC's). Each PC is a linear combination of the original variables [2] and is represented by a column of the right singular matrix $\mathbf{V}$ of the original matrix. More specifically, the degree of contribution by the original variables to the first PC is shown in the first column of the right singular matrix. The variance of a PC is a measure of the amount of information conveyed by that PC. This variance of a PC is given by the square of the singular value corresponding to this PC. By convention, the PC's are arranged in order of decreasing variance, the most informative PC is the first (represented by the first column of $\mathbf{V}$) and the last PC (represented by the last column of $\mathbf{V}$) is the least informative [2].

We can see how PCA and SVD are related by recalling the structure of SVD. Let $\mathbf{X}_t$ be a $m$ by $n$ matrix, whose rows represent sampling times and whose columns represent contaminant species, and the SVD of $\mathbf{X}_t$ be $\mathbf{X}_t = \mathbf{U}\Sigma\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and are the left and the right sets of singular vectors respectively, and $\Sigma$ is diagonal containing, in decreasing magnitude, the singular values. Recall that the column vectors

of $\mathbf{V}$ are eigenvectors of $\mathbf{X}_t{}^T\mathbf{X}_t$. Because each column of $\mathbf{V}$ represents a PC variable, the PC's of $\mathbf{X}_t$ are the columns of $\mathbf{V}$. The first PC is represented by the first or the left-most column vector of $\mathbf{V}$. The relative magnitude of the original variables in the PC determines the relative importance of the corresponding variables in this PC. It is implicitly assumed here that we are looking for the interrelationships among the variables (represented by the entries of the columns of a matrix) instead of the interrelationships among the observations (represented by the entries of the rows of a matrix).

FA is an extension of PCA [2]. Because the main objective of FA is easy interpretion of the new independent variables, applying a matrix transformation so that this objective is achieved is commonly done. A well accepted matrix transformation that is used to enhance the distinctiveness of the columns in a matrix is Varimax. Varimax does the matrix transformation by an orthogonal rotation on the frame of reference. The new axes are selected such that they go through clusters of the points in the scattered diagram of the independent variables. Computationally, Varimax transformation is done by maximizing the sum of the variances of the squared factor loadings within each factor. Other tranformation methods are Equamax, Quartimax, Promax, and Orthoblique. Two kinds of rotations are possible: orthogonal and oblique. The later may achieve better alignment, but is at the expense of resulting in nonorthogonal PC's, that loss the mutual independence in the data. Orthoblique is an example of the oblique transformation methods.

## 3.5 Two Examples of the Application of Principal Component Analysis and Factor Analysis

### 3.5.1 Air Pollution Example

The first example is from Wolff et al. [59] and is the application of PCA and FA on the air pollution data that were collected in Arizona to determine the possible sources of the air pollutants. PCA and FA were used to group air pollutants that possess similar variations. The raw data is an $333 \times 10$ location-versus-air pollutant concentration matrix. The ten air pollutants were manganese, copper, iron, lead, total particulates, total organics, sulfates, nitrates, chlorides, and carbon monoxide. This matrix was mean-centered and variance-normalized before SVD was applied to determine how many independently varying groups there are.

The element at the $i$th row in the first column in Table 3.3 is the culmulative percent explainable by the first $i$ PC's or factors [59]. It is found by successively adding the percentage of the amount contributed by each PC of the mean-centered and variance-normalized concentration matrix. The first three PC's are shown in Table 3.3. These three columns in Table 3.3 are the variance-normalized first three columns of $\mathbf{V\Sigma}$. The matrices $\mathbf{V}$ and $\Sigma$ are from $\mathbf{X}_t = \mathbf{U\Sigma V}^T$.

Gabriel's plot is a handy illustrative tool for PCA and FA. It is used to illustrate the relationships between the original variables. A Gabriel's plot is the plot of the vectors formed by the rows of $\mathbf{V\Sigma}$, using only the first two or three columns. If only the first two columns of $\mathbf{V\Sigma}$ are considered, the Gabriel's plot is two dimensional. If the first three columns of $\mathbf{V\Sigma}$ are considered, the Gabriel's plot is three dimensional. The objective of doing a Gabriel's plot is to see which of the original variable are correlated

| Summed PC Ratio % | 1st PC | 2nd PC | 3rd PC |
|---|---|---|---|
| 38.9625 | 0.6546 | -0.6318 | 0.1096 |
| 57.5122 | 0.3232 | 0.4160 | 0.5166 |
| 69.7892 | 0.6437 | -0.6464 | -0.1060 |
| 77.9498 | 0.8010 | 0.3965 | -0.1922 |
| 85.2382 | 0.7910 | -0.4932 | 0.0709 |
| 91.3750 | 0.7540 | 0.4200 | -0.2338 |
| 94.1388 | 0.4813 | -0.0347 | 0.1948 |
| 96.7774 | 0.2130 | 0.2039 | 0.7625 |
| 98.7294 | 0.6355 | 0.1982 | 0.2004 |
| 100.0000 | 0.6573 | 0.4539 | -0.4258 |

Table 3.3: Principal components of the air pollution example.

to which other. Correlation among the variables is discernable when the vectors formed by the corresponding variable row vectors of $V\Sigma$ lie close together pointing in similar or opposite directions (positively or negatively correlated respectively). Discussions in the previous sections have indirectly explained that entries in each column of V are the contributions of each of the corresponding original variable to the respective PC. The relative orientation of the vectors that start from the origin to the coordinates specified by the rows of V indicates the relationships between the original variables. Matrix $\Sigma$ functions as a factor to V to incorporate the information on the amount of information conveyed by each PC in a Gabriel's plot. There are two things to look for in a Gabriel's plot: the relative lengths and the relative directions of the vectors. The components of each vector in the plot give the weights of each of the original air pollutant variables in the various principal component respectively. For instance, if an air pollutant vector is aligned mostly with the first PC and the second PC, it means that the first two PC's can capture most of the variation in that air pollutant, and the third PC contains practically no information in regard to that air pollutant. All the Gabriel's plots shown here are the plots of the first three variance-normalized PC's, or the first three columns of $V\Sigma$. Each vector in the Gabriel's plots in Figures 3.6 and Figures 3.7 is labeled with a number that corresponds to an air pollutant. The numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 represents manganese, copper, iron, lead, total particulates, total organics, sulfates, nitrates, chlorides, and carbon monoxide, respectively.

Figures 3.6 and Figures 3.7 suggest that the distributions of air pollutants 4, 6, and 10 (lead, total organics, carbon monoxide) are highly correlated because their respective vectors in the plots point in the same general direction. The distributions of these air pollutants seem to have little correlation with the distributions of air pollutants 1, 3, and

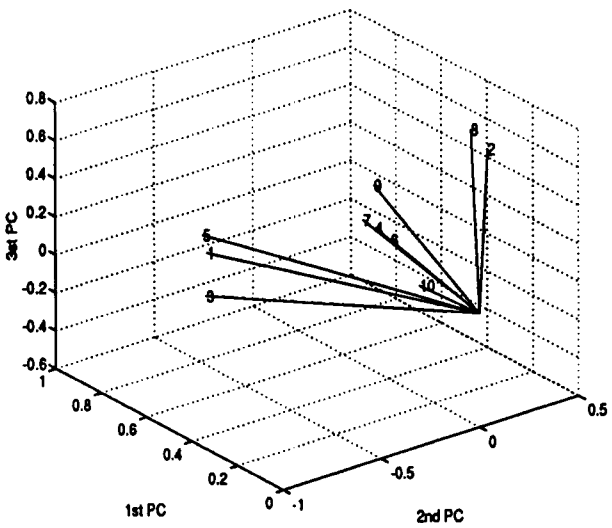



Figure 3.6: Gabriel's plot of the ten air pollutants.

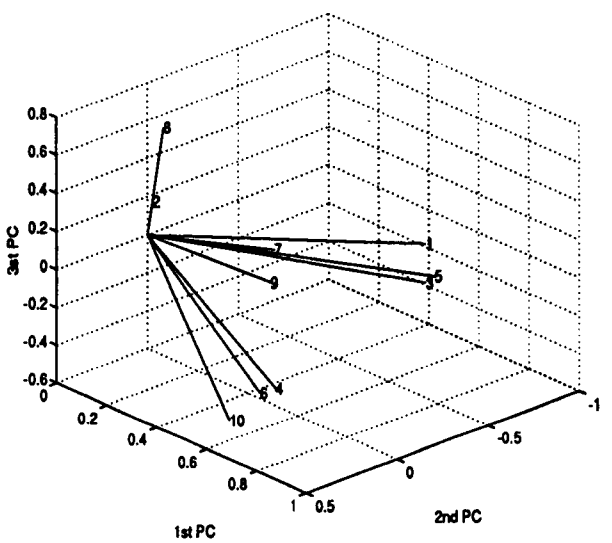



Figure 3.7: Gabriel's plot of the ten air pollutants (different view).

5 (manganese, iron, total particulates) because the cluster of vectors for air pollutants 1, 3, and 5 point in a direction roughly perpendicular to the cluster of vectors for air pollutants 4, 6, and 10. The length of each of these vectors relative to its length in the original $n$-space indicate how well the vectors are represented by these PC's [39]. Here, all the vectors are of about the same length. Thus it is concluded that the three dimensional plots represent the concentration variations of the ten air pollutants about equally well.

Note that in doing PCA, the PC's do not usually correspond to any physically meaningful variable. In order to make interpretation of PCA results more meaningful, a rotation of the PC's can be carried out, such that the rotated PC's align as much as possible to the original set of variables.

A clearer picture of the correlation among the variables is portrayed in Fig. 3.6 and Fig. 3.7, where orthogonal Varimax transformation was used. The first three PC's rotated by a Varimax transformation are shown in Table 3.4. The original variables that weigh much more heavily than other original variables in each of the PC's become more evident after the Varimax transformation. This clustering of variables into more or less mutually independent groups can also be seen in the Gabriel's plot. The Gabriel's plots with the Varimax rotated PC's are plotted as shown in Fig. 3.8 and Fig. 3.9. These figures also show that there are three groups of air pollutant distributions which are highly correlated within each group and only slightly correlated among the groups. The three groups are Group A: 1 (manganese), 3 (iron), and 5 (total particlates), Group B: 4 (lead), 6 (total organics), and 10 (carbon monoxide), and Group C: 2 (copper) and 8 (nitrates). The distributions of air pollutants 7 (sulfates) and 9 (chlorides) seem to be correlated to all three of these groups.

| 1st PC | 2nd PC | 3rd PC |
|---|---|---|
| 0.0482 | 0.9118 | 0.0781 |
| 0.2226 | -0.0659 | 0.7004 |
| 0.1304 | 0.9009 | -0.1218 |
| 0.8640 | 0.2136 | 0.2088 |
| 0.2380 | 0.8942 | 0.1329 |
| 0.8636 | 0.1623 | 0.1657 |
| 0.2250 | 0.3549 | 0.3071 |
| -0.0843 | 0.0382 | 0.8120 |
| 0.4599 | 0.2811 | 0.4391 |
| 0.9029 | 0.0604 | -0.0225 |

Table 3.4: Varimax rotated principal components.

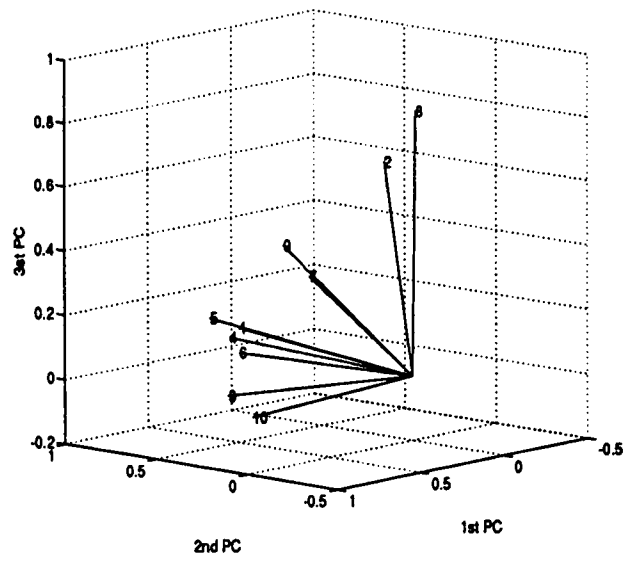



Figure 3.8: Gabriel's plot of the ten air pollutants in varimaxed frame.

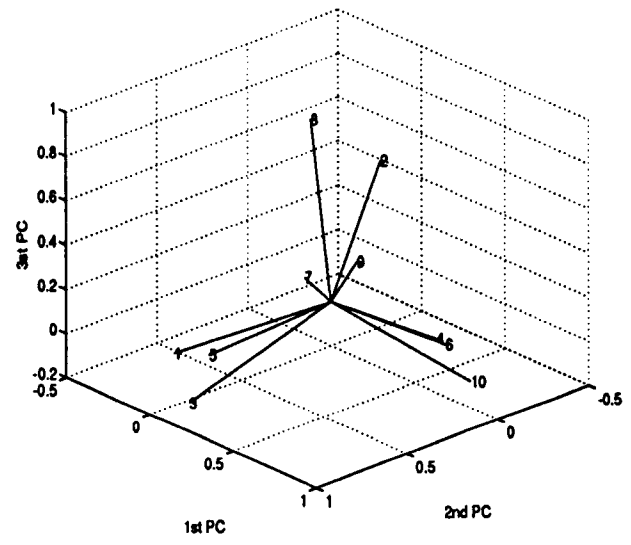



Figure 3.9: Gabriel's plot of the ten air pollutants in varimaxed frame (different view).

The Varimax transformed PC's shown in Fig. 3.8 and Fig. 3.9 indicate that Group A, Group B, and Group C are highly dependent on the 2nd PC, the 1st PC, and the 3rd PC respectively. Since the first three PC's are orthogonal, they represent three underlying distinct factors that determine the distributions of the air pollutants. Here, each PC is interpreted to represent a possible source of the air pollutants. Because automobiles are known to contribute lead, organics, and particulates into air, the source of the air pollutants in Group B is deduced to have come from automobiles. And since copper smelters contribute copper, particulates, and sulfates into the air, the source of the air pollutants in Group C is deduced to have come from the copper smelters. Similarly, since desert dust could add particulates, copper, manganese, and iron into air, the source of the air pollutants in Group A can be possibly attributable to come from desert dust. This example indicates that physical knowledge and experience on the system whose data the PCA and FA are applied to, are very important to the interpretations of the PCA and FA analysis result.

Figures 3.8 and Figures 3.9 suggest that there are at least three independent sources for the pollutants. Although three groups are identified, this does not mean that there are only three sources. The contribution of a particular air pollutant arising from a source whose concentration is significant but mostly constant in time and space could go undetected [26].

As this example shows, MVS has the ability to detect clustering of like variables and treat them as one cluster, and independently from other clusters. The distinct clusters can be tracked over time to see how they change. This approach is essentially what we have taken in modeling the natural remediation at the Sleeping Bear Dunes National Lakeshore site in Chapter 4, except we do not apply orthogonal rotation to the PCA

results, because only one underlying factor exists that accounts for most of the dispersion in the concentration data.

### 3.5.2 One Independent Varying Factor Example

Five variables that are highly correlated to each other are simulated for this example to demostrate that Gabriel's plot is able to detect that there is only one independently varying variable within the five highly correlated variables. Fig. 3.10a shows that although the five variables are highly correlated, their magnitudes and the sizes of their variations are simulated to be different from one another. As shown in Fig. 3.10b, SVD is able to correctly detect that the first PC captures more than 99.9% of the information in the five simulated variables. This means that the five variables can be represented full by only one variable, that is the first PC. SVD is able to correctly detect the high amount of correlation among of the five simulated variables because the five variables are mean-centered and variance-normalized before SVD is applied. The high correlation among the five variables is also detectable by the Gabriel's plots shown in Fig. 3.11. All the vectors in the Gabriel's plots are aligned, indicating that all the variables pointing in similar direction are highly correlated, and that there appears to be only one independently varying factor that accounts for the dispersion in the original data set.

This example is included here because in analyzing the Sleeping Bear Dunes bioremediaton data, only one independently varying factor is also found, as evidenced by the PCA analysis as well as the Gabriel's plot. They will be presented in Chapter 4.

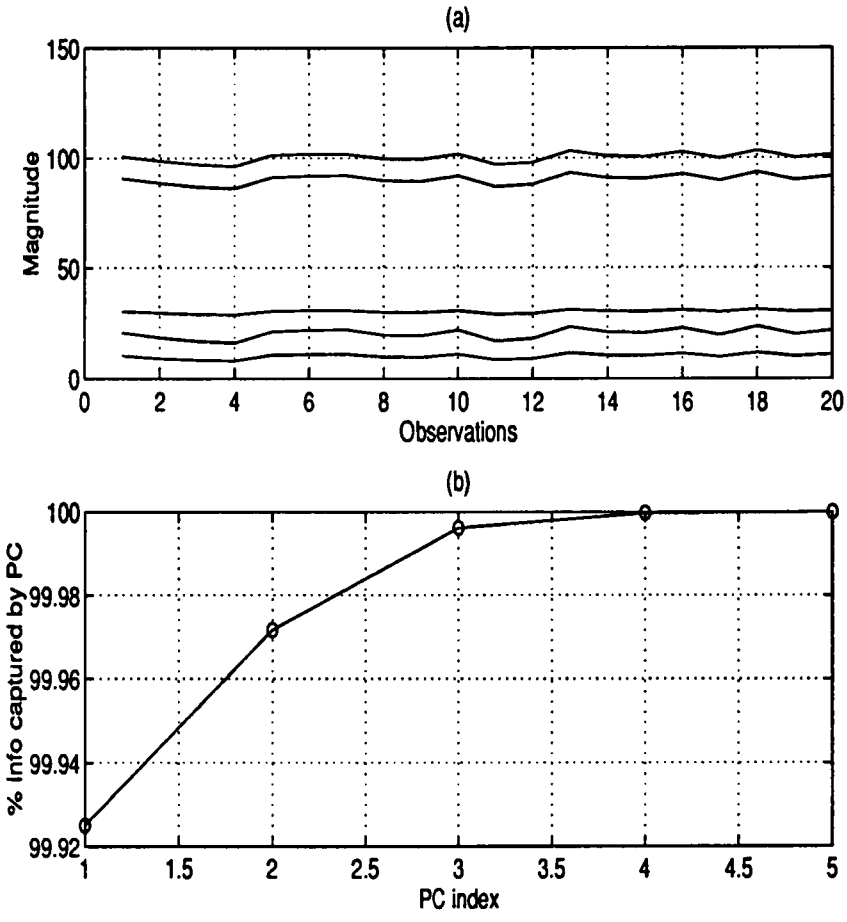



Figure 3.10: The plot of the five highly correlated variables (a). Percentage information of the five correlated variables captured by their PC's (b).

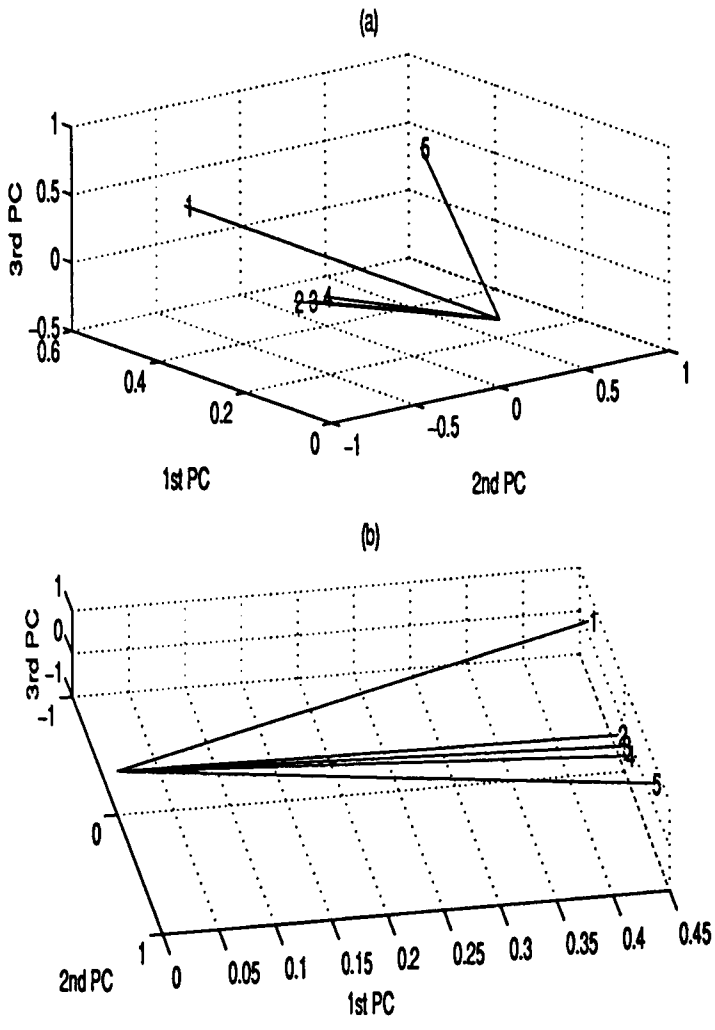



Figure 3.11: Gabriel's plot of a data set that has one dorminant independent variable or PC. (b) is a different view of (a).

# CHAPTER 4

# MVS APPLICATION RESULTS

To assess the viability of in-situ natural or intrinsic bioremediation, a modeling approach to quantify and characterize the likelihood that in-situ natural biodegradation will remediate a contaminated site within a reasonable time period is needed. Field data of BTEX compounds collected during a two-year period from a gasoline contaminated spill site at the Sleeping Bear Dunes National Lakeshore area in Michigan were obtained from Dr. John L. Wilson of U.S. Environmental Protection Agency–Robert S. Kerr Laboratory. These data are valuable because at present there is almost no operating history to judge the effectiveness of in-situ natural bioremediation [6]. The field data are useful in providing a case study for feasibility of developing a degradation model using the Multivariate Statistical analysis approach.

## 4.1 Background Information on the Natural Remediation Data from the Sleeping Bear Dunes National Lakeshore

Field data on the natural bioremediation of the gasoline spilled from three underground storage tanks at the Sleeping Bear Dunes National Lakeshore were obtained from the EPA–Robert S. Kerr Laboratory. The tanks were removed and the excavated soil was backfilled on December 11, 1989 [55]. In February 1989, 21 monitoring wells were installed to monitor the progress of natural bioattenuation [55]. The last samples were taken in November 1992 [55]. The field data contain measurements of the concentrations

of BTEX compounds, oxygen, nitrate, sulfate, and iron II, groundwater pH and temperature, as well as water table level at irregular intervals throughout a two-year period. These field data are shown in Appendix A. Many of the data are presented in matrix format where each row of the matrix represents a sampling location and each column of the matrix represents a sampling time. The BTEX compounds, oxygen, hydraulic conductivity were measured or determined with gas chromatography, an electrode, and the changes in head in passive monitoring wells, respectively [55]. Hydraulic gradients were calculated from the difference in the water table levels. The other data were collected using appropriate means of measurement [55]. These data were almost completely obtained from 21 strategically located monitoring boreholes of different depths, located at three principal sites. Fig. 4.1 shows the arrangement of the monitoring boreholes. The cluster of monitoring boreholes located in the plume will be referred to as the cluster at the spill site, the cluster of monitoring boreholes nearest to the river will be referred to as the cluster at 21.3 m site, and the cluster of monitoring boreholes in between the spill site and the 21.3 m site is referred to as the cluster at the 9.1 m site. Fig. 4.2 provides the cross-sectional view of the three clusters of monitoring boreholes.

The apparent total concentration of each BTEX compound at each cluster of the monitoring boreholes as a function of time is shown in Figures 4.3 and 4.4. Benzene appears to be the only component that shows weathering over time.

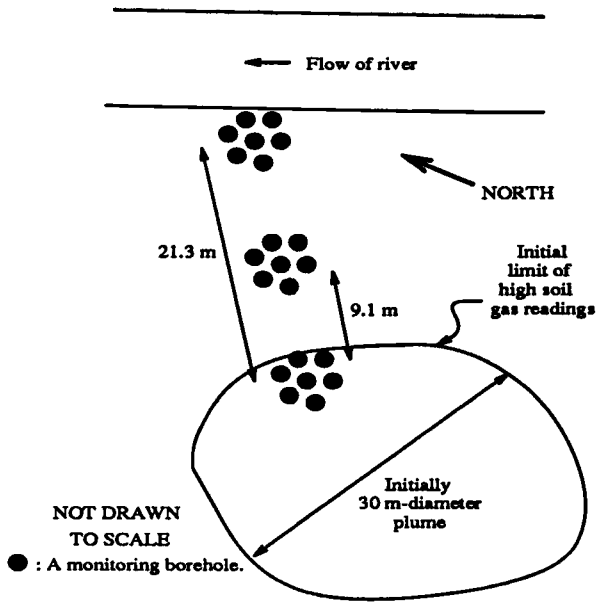



Figure 4.1: Layout of the monitoring boreholes with respect to the plume and the river.

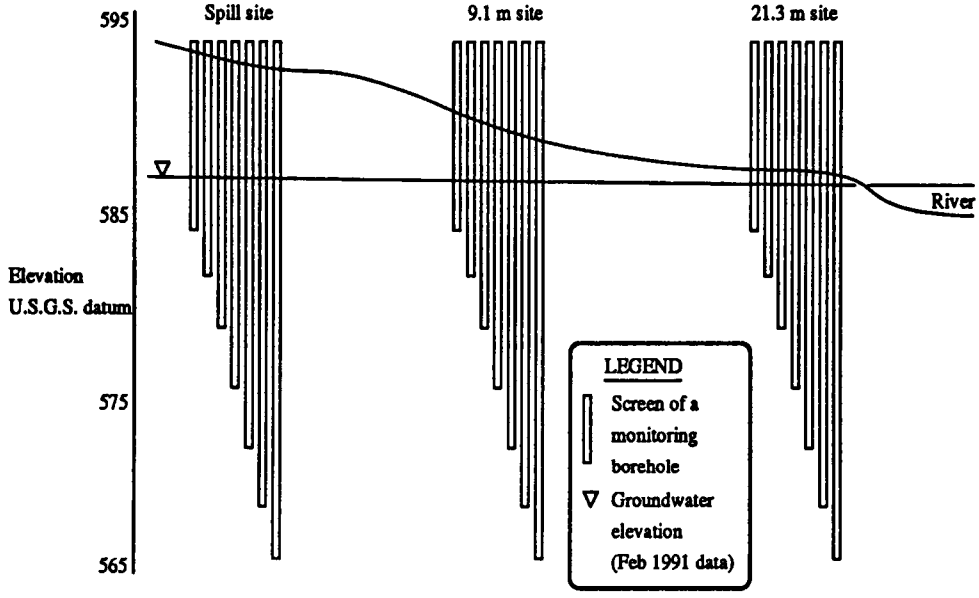



Figure 4.2: Cross-sectional view of the monitoring boreholes.

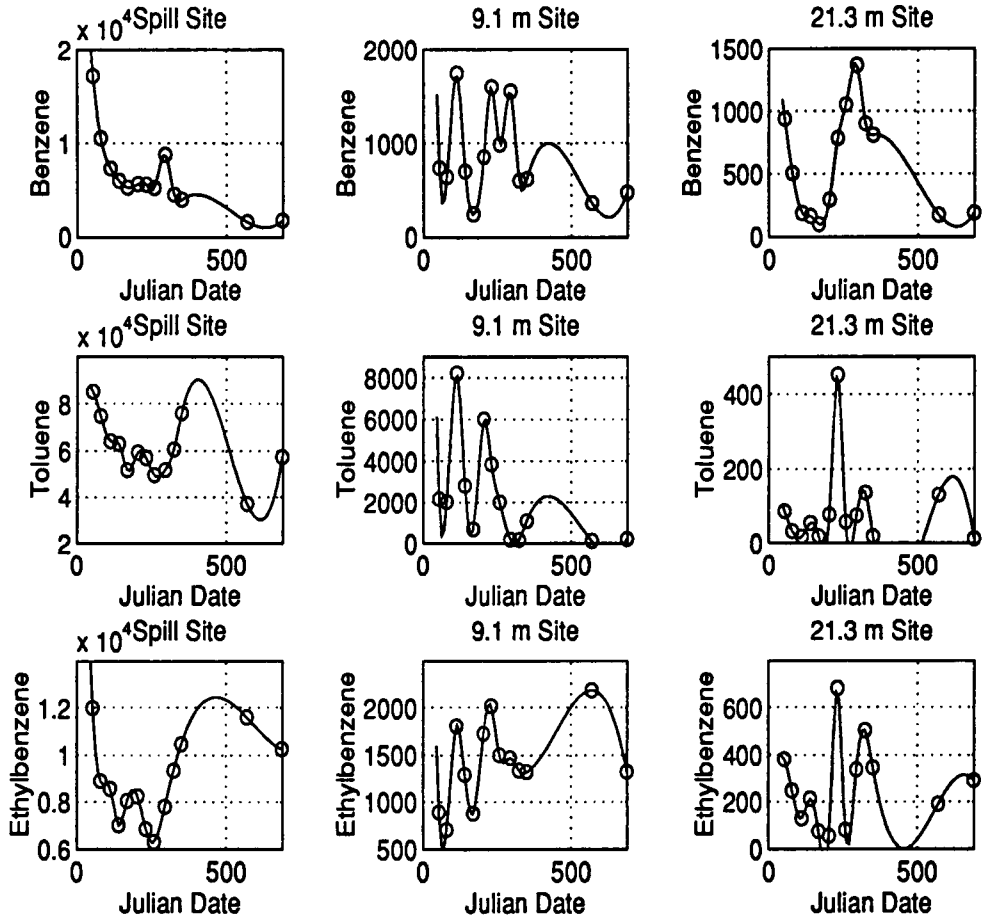



Figure 4.3: Apparent concentrations of benzene, toluene and ethylbenzene in mg/L at the spill, 9.1 m and 21.3 m sites. Each of these concentration profiles is the sum of the concentrations of a BTEX compound at each cluster of the monitoring boreholes.

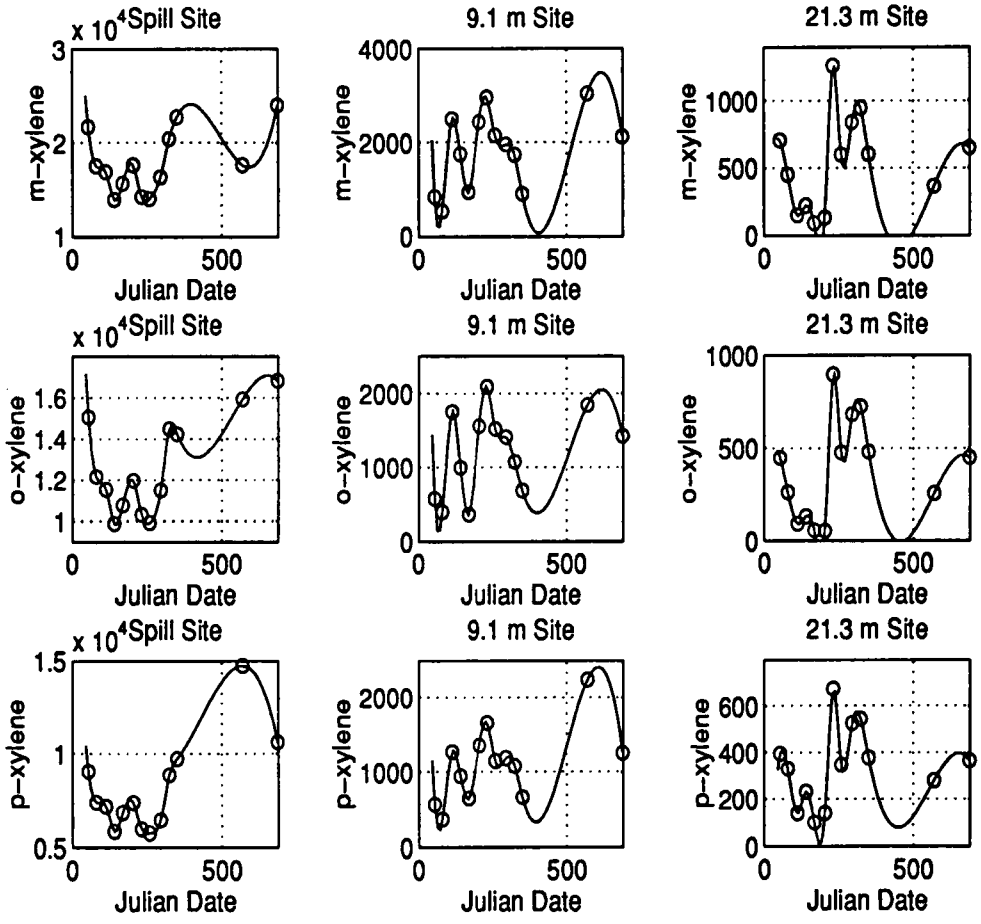



Figure 4.4: Apparent concentrations of *m*-xylene, *o*-xylene, and *p*-xylene in mg/L at the spill, 9.1 m and 21.3 m sites. Each of these concentration profiles is the sum of the concentrations of a BTEX compound at each cluster of the monitoring boreholes.

Figure 4.5 illustrates the water table level throughout the two-year monitoring period. As mentioned earlier, each of the monitoring boreholes in each cluster was screened at a different depth. The screened interval of the monitoring boreholes at each cluster had ranges of 587-584 ft, 584-581 ft, 581-578 ft, 578-575 ft, 575-572 ft, 572-569 ft, and 569-566 ft above sea level [55]. Figure 4.5 shows that the water table level varies in the range of 584.7 to 585.7 ft above sea level in a year, thus the monitoring boreholes with their widely varied depths are adequate in performing their monitoring function in spite of the water table fluctuations.

It was found by Wilson et al. [55] that the hydraulic gradient was sloped in the direction from the gasoline plume to the river for most of the year except during the summer when the direction of the hydraulic gradient was reversed [55]. Wilson et al. determined that a hydraulic gradient of $3\times10^{-3}$ m/m corresponded to an average groundwater flow velocity of 0.4 m/day at this site. The plot of the hydraulic velocity at fifteen sampling time points is shown in Fig. 4.6. This figure will be used later to determine the residence time of each contaminant corresponding to each sampling point.

Toward the end of the monitoring period, Wilson et al. [55] also monitored the concentration of 2,3-dimethylpentane (DMP), a chemical that is known to be inert to degradation in the ambient conditions of the site. The last two sample points include DMP information. Using DMP data to correct for the dilution of the contaminants due to plume dispersion, it was determined that benzene was not biodegraded appreciably [55]. Using the last two sampling points, after correcting for dilution and assuming that the biodegradation of the BTEX compounds followed first-order kinetics, i.e. $\frac{dC}{dt} = -kC$ where $C$ denotes concentration, $t$ time, and $k$ the rate constant, Wilson et al. [53] calculated that the individual first-order degradation rate constants of toluene, ethylbenzene,

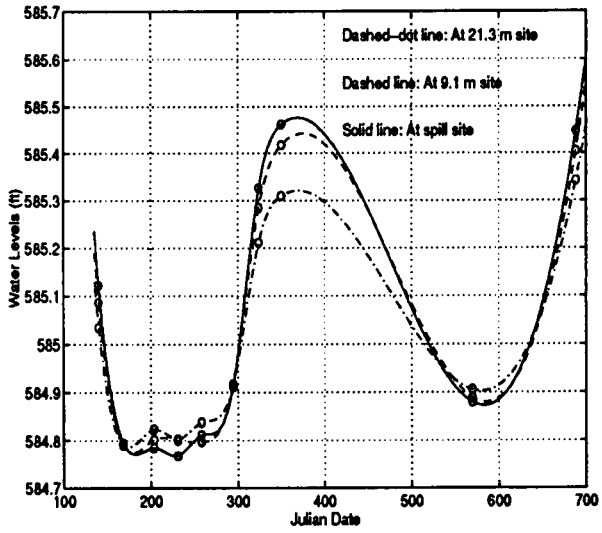



Figure 4.5: Water table levels in terms of elevation above sea level.

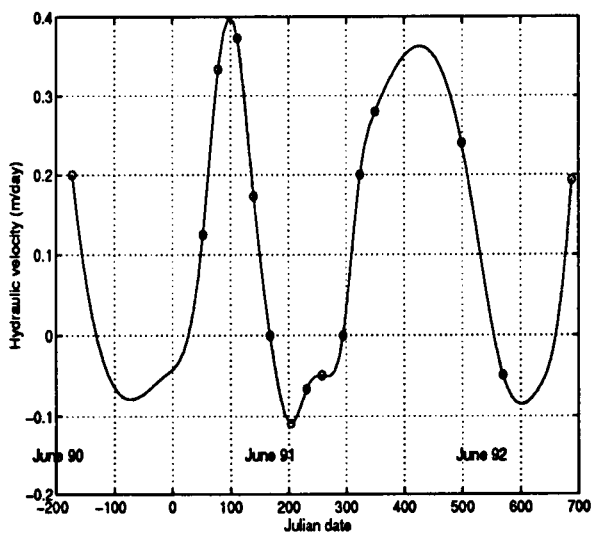



Figure 4.6: Hydraulic velocity profile.

*p*-xylene, *m*-xylene, and *o*-xylene ranged from 0.023 to 0.067 per day, 0.0031 to 0.011 per day, 0.0024 to 0.0096 per day, 0.0037 to 0.014 per day, and 0.0040 to 0.016 per day respectively. There are two numbers for each compound, because two sample points are available. It is evident that the two numbers differ by as much as four fold. Furthermore, from analyzing the electron acceptor species data, Wilson et al. also concluded that the BTEX compounds at the Sleeping Bear Dunes National Lakeshore were predominately biodegraded by methanogenesis, followed by nitrate reduction and sulfate reduction [53]. Aerobic respiration was found to be unimportant [53].

## 4.2 Sleeping Bear Dunes National Lakeshore Case Study Using Multivariate Statistical Analysis Approach

The modeling of the Sleeping Bear Dunes National Lakeshore biodegradation is outlined in Fig. 4.7. Hydraulic gradient data were used to first calculate the residence times of the BTEX compounds sampled at each of the sampling time points. Using the residence times, the initial concentrations of the BTEX compounds corresponding to the residence time prior to the sampling time point before they reached the sampling site were then determined using concentration information as conveyed in Figures 4.3 and 4.4. Out of the thirteen sampling data points available for the BTEX compounds, the first three data points had to be discarded because the initial concentrations of the BTEX for those first three sampling data points were not available. They were not available because they require concentration data at times prior to the start of the monitoring period. Because it did not biodegrade appreciably at the Sleeping Bear Dunes National Lakeshore site [55], benzene was used as the reference compound to correct for dilution

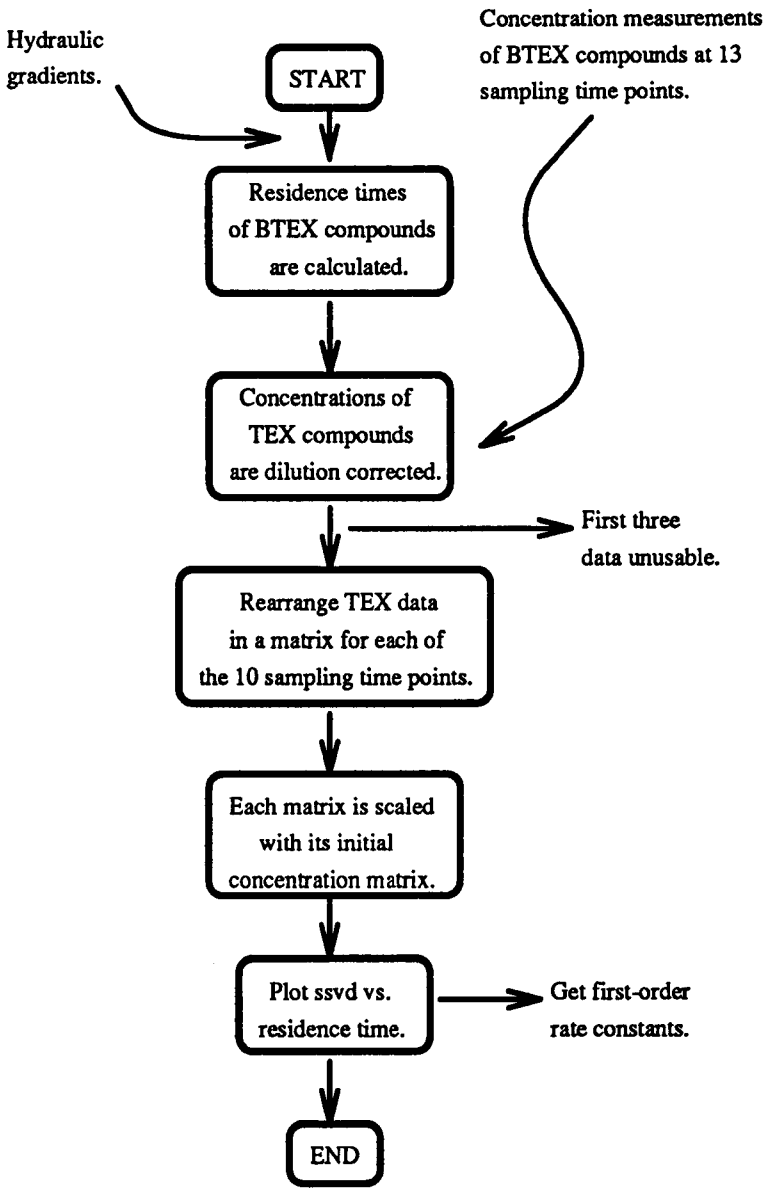



Figure 4.7: General outline of the Sleeping Bear Dunes National Lakeshore Case Study.

for the rest of the BTEX compounds, which are now referred to as the TEX compounds. The measurements on the concentrations of the TEX compounds were put into a matrix form for each of the ten sampling time points. In order to apply SVD on the ten matrices, each of the matrices was first mean-centered columnwise with respect to itself and variance-scaled with respect to the variances of their corresponding initial concentration matrices that correspond to the residence time prior to the sampling time. The initial concentration matrices of TEX compounds were calculable because the residence times of the TEX compounds were known. SVD was then applied to each of the mean-centered and scaled concentration matrices and the sum of the squares of the singular values of each of the concentration matrices were plotted against the average residence times of the TEX compounds. The first-order degradation rate was then determined by fitting the points to a first-order degradation kinetics. This is the general outline for the modeling of the Sleeping Bear Dunes National Lakeshore case study. Details are presented below.

The first task in this case study was to determine the corresponding residence times at each sampling point of each contaminant in groundwater travelling from the spill site to the 9.1 m site and from the 9.1 m site to the 21.3 m site. These residence times can be determined from Figure 4.6, the hydraulic velocity plot. The distance travelled by the groundwater from an arbitrarily chosen initial time point and a later time point is the area under the curve in Figure 4.6 between the two specified time points. The distance travelled by the groundwater as a function of time is illustrated by Figure 4.8. At each of the sampling time points, the curve was traced backward until, say, 9.1 m of distance is covered. The difference between the corresponding date and the respective sampling date is the residence time for the compound that has flowed from the spill site arriving

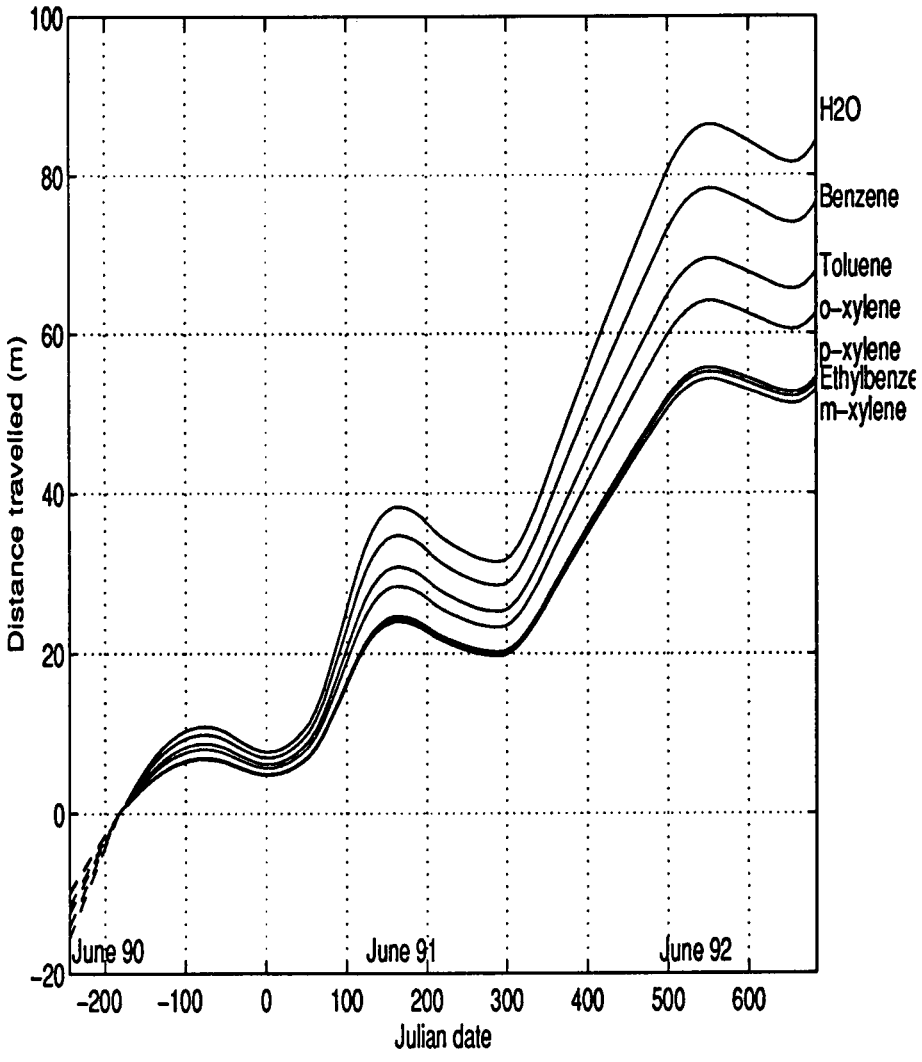



Figure 4.8: Distance travelled by groundwater and contaminants between an arbitrarily chosen initial time point and the indicated time point. The dashed lines are extrapolations.

at the 9.1 m site at that respective sampling time point.

Due to the adsorption of BTEX compounds onto soil and naturally occurring organic carbon, the BTEX compounds that are dissolved in the groundwater do not move as fast as the groundwater. The ratio of the velocity of groundwater to the velocities of the BTEX compounds dissolved in the groundwater has been described by Eq. 2.3. These velocities are linear velocities and are in the direction from the spill site to the 21.3 m site. Variables $B$ (soil bulk density), $n$ (porosity), $K_d$ (distribution coefficient) of Eq. 2.3 at the Sleeping Bear Dunes National Lakeshore are not available. They were estimated from the literature. Since many soils have mean density of particles of 2.6 to 2.75 g/cm$^3$ [46], $B$ was estimated to be 2.7 g/cm$^3$. And since the porosity of coarse-textured soils such as sand tend to be between 30 to 40% and the porosity of fine-textured soils such as clay is usually within the range of 40 to 60% [46], $n$ will be estimated as 40%. The values of $K_d$ for various chemicals are usually reported in their normalized form, $K_{oc}$ [13]. The values of $K_{oc}$ of the BTEX compounds were presented in Chapter 2. Eq. 2.4 shows that $K_d$ and $K_{oc}$ are related through the normalization factor $f_{oc}$. Recall that $f_{oc}$ is the fraction of naturally occurring organic carbon in soil. The reasonable range of $f_{oc}$ for well-drained to moderately well-drained soil is from near 0 to 4% [52]. Since the water table elevations in each of the three clusters of measurement boreholes are within 0.01 m from each other at any sampling day, Wilson et al. concluded that the hydraulic conductivity of the soil at the Sleeping Bear Dunes National Lakeshore was very high [55]. High hydraulic conductivity implies low $f_{oc}$. For the reasons of high hydraulic conductivity in soil, lack of more complete information, and limited amount of data available, $f_{oc}$ is chosen to be very small, 0.04%, so that the retardation factors of the contaminants are close to one. The retardation factor is the inverse term that

appears in the right side of Eq. 2.3. Now that we have an estimate for each of the variables $B$, $n$, $K_{oc}$, and $f_{oc}$, the retardation factors for benzene, toluene, ethylbenzene, m-xylene, o-xylene, and p-xylene were calculated to be 0.91, 0.80, 0.64, 0.63, 0.74, and 0.64, respectively. These retardation factors were used to calculate the velocity of each BTEX compound in the groundwater relative to that of clean groundwater. From these velocities, the distances travelled by the groundwater from an arbitrarily chosen initial time point and a later time points were calculated and are plotted in Figure 4.8.

The residence time of each of the BTEX compound for travelling from the spill site to the 9.1 m site at each of the thirteen sampling time points was determined from Figure 4.8 using the following steps.

1. Determine the distance that corresponds to one of the thirteen sampling times using the curve of a BTEX compound in Figure 4.8.

2. Substract 9.1 m from that distance.

3. Determine the new time point that corresponds to the substracted distance using the curve of that BTEX compound in Figure 4.8.

4. Substract the new time from the original sampling time. This is the residence time corresponding to one of the thirteen sampling times of that BTEX compound which travelled from the spill site to the 9.1 m site.

5. Repeat the steps above for the rest of the thirteen sampling times.

6. Repeat the steps above for the rest of the BTEX compounds.

The time for each BTEX compound for travelling the distance between the 9.1 m to the 21.3 m site at each of the thirteen sampling time points was also determined by following

| Julian date | Spill site to 9.1 m site | 9.1 m site to 21.3 m site |
|---|---|---|
| 139.8 | 47.8 | 62.3 |
| 167.8 | 72.4 | 82.6 |
| 203.5 | 115.9 | 124.6 |
| 230.8 | 149.2 | 159.3 |
| 258.1 | 179.6 | 192.6 |
| 294.5 | 220.2 | 231.8 |
| 323.9 | 237.6 | 246.2 |
| 349.8 | 215.8 | 254.9 |
| 569.6 | 105.7 | 121.7 |
| 687.9 | 227.4 | 243.3 |

Table 4.1: Average residence times of the BTEX compounds.

the same steps above except that 12.2 m (or 21.3 – 9.1) instead of 9.1 m is substracted in Step 2. See Appendix C for the Matlab program that does this calculation. These calculations produce a residence time for each of the BTEX compounds at each of the thirteen sampling time points. The residence times of each BTEX compound at the 9.1 m and 21.3 m sites are shown in Appendix B. The averages of the residence times for the six BTEX compounds at ten sampling time points are shown in Table 4.1. Julian date is the name given for the date which starts from an arbitrarily selected first day. The reason that only ten average residence times are shown in Table 4.1 instead of thirteen average residence times to correspond with the original thirteen sampling times is that

the first three residence times of all the BTEX compounds would correspond to times prior to the start of the two-year sampling period. Due to insufficient data, the first three average residence times could not be determined.

If we assume that the BTEX compounds flowed from the spill site to the 9.1 m site, the residence time of a BTEX compound for travelling from the spill site to the 9.1 m site at each sampling time can be used to determine the initial concentration of a batch of the BTEX compound at the spill site before this particular batch of the BTEX compound travelled, to and arrived at the 9.1 m site at each sampling time, using the concentration versus time data as shown in Figures 4.3 and 4.4. Similarly, if we assume that the BTEX compounds flowed from the 9.1 m site to the 21.3 m site, the initial concentration of a batch of a BTEX compound at the 9.1 m site before that particular batch of the BTEX compound travelled to and arrived at the 21.3 m site at each sampling time can be traced back using the residence time of the BTEX compound for travelling from the 9.1 m site to the 21.3 m site at each sampling time. Thus, the changes in the concentration of the BTEX compounds as the BTEX compounds travel between the sites can be determined.

The changes in the concentration of the BTEX compounds could be caused by dilution due to dispersion of the BTEX plume, fluctuations in the initial concentration of the BTEX compounds, biodegradation of the BTEX compounds, and the rise and fall of the water table. It was assumed that the contaminants had a point source, and thus, as the contaminants moved or dispersed from the source, the contaminants were diluted. The changes in the concentration of the BTEX compounds could also be due to the fluctuations in the initial concentrations of the BTEX compounds. These initial concentrations were the concentrations of the BTEX compounds at the spill or the 9.1 m site that were determined by tracing back the BTEX compounds that flowed from the

spill site to the 9.1 m and from the 9.1 m site to the 21.3 m sites, by using the appropriate residence times of the BTEX compounds. Varying flow velocities of the groundwater and the BTEX compounds could cause these fluctuations in the initial concentrations of the BTEX compounds. The rise and fall of the water table could cause changes in the concentration of the BTEX compounds because most of the BTEX compounds floated on top of the water table, and if the water table has fallen such that what was actually sampled at times may be the nonaqueous top layer of petroleum compounds instead of what are dispersed in groundwater.

In order to model the biodegradation of the BTEX compounds, the changes in the concentration of the BTEX compounds due to plume dispersion have to be nullified. Benzene was used for the purpose of nullifying the effect of dilution on the distribution of the concentrations of ethylbenzene, toluene, *m*-xylene, *o*-xylene, and *p*-xylene (or the TEX compounds for brevity), because according to Wilson et al. [55], benzene was not biodegraded to any significant degree. Thus, any change in the concentration of benzene must be due to causes other than bioremediation. The concentration of a TEX compound at the 9.1 m site was dilution corrected by the following steps:

1. Determine the initial corresponding concentration of benzene at the spill site corresponding to the sampling time under consideration using the appropriate residence times of a TEX compound for each of the ten sampling times.

2. Calculate the ratios of this initial corresponding concentration of benzene to the concentration of benzene at the 9.1 m site at each of the ten sampling time points. These ratios are the dilution correction factors for that TEX compound at the 9.1 m site.

3. Take the products of these ratios with the concentrations of that TEX compound at the 9.1 m site for the ten sampling time points. These products give the corrected TEX concentrations.

These steps are repeated for each TEX compound at the 9.1 m site using the benzene concentration as the reference. The concentrations of the TEX compounds at the 21.3 m site are similarly corrected for dilution except that the initial corresponding concentrations of benzene are traced to the 9.1 m site instead of to the spill site. The dilution correction factors are tabulated in Table 4.2. The dilution correction factors for *m*-xylene and *p*-xylene are not shown in Table 4.2 because they are the same as the correction factors for ethylbenzene. The dilution corrected concentrations of TEX compounds at the 9.1 m and 21.3 m sites vs. time are shown in Fig. 4.9.

Wilson et al. [55] used 2,3-dimetylpentane (DMP) instead of benzene to nullify the dilution due to plume dispersion of the BTEX compounds. DMP being inert under the conditions at the Sleeping Bear Dunes National Lakeshore would not biodegrade. Unfortunately, the concentration of DMP was only measured at the last two sampling time points. The dilution correction factors for all the BTEX compounds calculated by Wilson et al. by DMP are tabulated in Table 4.2 for comparison with the dilution correction factors calculated by benzene as reported here. Because only two data points are available for the DMP, the initial concentration of DMP for each of the 9.1 and 21.3 m sites used in the calculation steps for the dilution correction factor outlined previously is assumed to be a constant that is equal to the average concentration of the two DMP concentration measurements. Because of this, the dilution correction factors determined by DMP are the same for all the BTEX compounds at each of the 9.1 m and 21.3 m sites. Table 4.2 shows that there are as much as 50% difference between the correction

| | Determined by Benzene | | | | | | Determined by DMP [55] | |
|---|---|---|---|---|---|---|---|---|
| | Toluene | | Ethylbenzene | | *o*-xylene | | All | |
| Julian Date | 9.1 m Site | 21.3 m Site | 9.1 m Site | 21.3 m Site | 9.1 m Site | 21.3 m Site | 9.1 m Site | 21.3 m Site |
| 139.8 | 12.6 | 5.7 | 13.8 | 4.1 | 12.6 | 3.9 | – | – |
| 167.8 | 35.4 | 10.0 | 39.0 | 7.1 | 35.4 | 10.0 | – | – |
| 203.5 | 11.0 | 2.5 | 12.0 | 2.3 | 11.0 | 2.5 | – | – |
| 230.8 | 6.4 | 0.8 | 7.3 | 0.9 | 6.4 | 0.8 | – | – |
| 258.1 | 10.2 | 0.6 | 11.6 | 0.7 | 11.6 | 0.6 | – | – |
| 294.5 | 7.1 | 0.5 | 8.5 | 2.5 | 7.1 | 0.5 | – | – |
| 323.9 | 15.9 | 0.8 | 17.3 | 0.7 | 15.9 | 0.8 | – | – |
| 349.8 | 12.6 | 1.6 | 14.0 | 1.2 | 12.6 | 1.6 | – | – |
| 569.6 | 7.4 | 2.8 | 7.9 | 3.0 | 7.6 | 2.9 | 3.6 | 3.8 |
| 687.9 | 5.7 | 2.6 | 6.1 | 2.7 | 5.9 | 2.7 | 5.1 | 3.3 |

Table 4.2: Dilution correction factors determined by benzene and DMP at 9.1 and 21.3 sites for the TEX compounds.

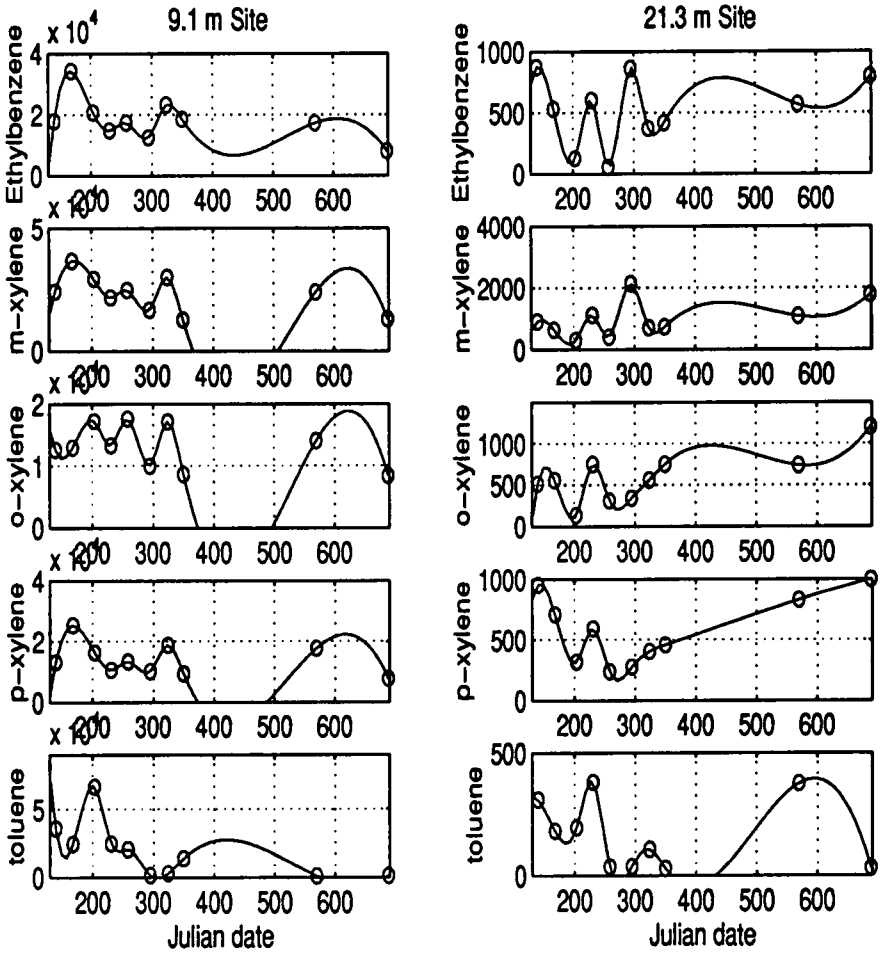



Figure 4.9: Dilution corrected concentration of toluene, ethylbenzene, *m*-xylene, *o*-xylene, and *p*-xylene in mg/L at the 9.1 m and 21.3 m sites.

factors calculated by using benzene as the dilution corrector and DMP as the dilution corrector. In view of the fact that ten benzene measurements are available for dilution correction as opposed to only two measurements for DMP, benzene is a more robust dilution corrector in the sense that any measurement errors in benzene will not affect its integrity in being used as a dilution corrector as much as any measurement errors in DMP will affect DMP as a dilution corrector. Because of this, the dilution correction factors determined by benzene are used in the rest of this case study.

The dilution-corrected concentrations of the TEX compounds at each of the ten sampling time points at each of the 9.1 m and 21.3 m sites were put into a seven-by-five matrix, whose rows represented the sampling locations and whose columns represented the TEX compounds. Let these matrices be named $\{\mathbf{X}\}$. There are two sets of ten such matrices, one set for each of the 9.1 m and the 21.3 m sites. There are also two sets of ten initial concentration matrices that correponded to the TEX concentrations at the 9.1 m and 21.3 m sites, at the corresponding residence time prior to the sampling time. For instance, the concentration matrix at 9.1 m site sampled at the 139.8 Julian date is:

$$\mathbf{X}_1 = \begin{bmatrix} 277 & 235 & 543 & 222 & 101 \\ 9834 & 15375 & 6568 & 7895 & 15282 \\ 3186 & 2078 & 808 & 1939 & 632 \\ 4571 & 6510 & 4673 & 3047 & 19577 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The initial concentration matrix of the BTEX compounds at the spill site at 47.8 days (the residence time corresponding to the 139.8 Julian date sampling time) prior to this

sampling point is determined by interpolations and extrapolations of the BTEX measurements at the spill site using the average residence time of 47.8 days from Table 4.1, and is:

$$\mathbf{X}_1^o = \begin{bmatrix} 3155 & 6462 & 4497 & 2663 & 32754 \\ 2845 & 5733 & 3923 & 2381 & 28643 \\ 1868 & 3692 & 2327 & 1578 & 6474 \\ 934 & 1465 & 1080 & 712 & 1206 \\ 5 & 7 & 4 & 3 & 10 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Each column of the ten concentration matrices for each of the 9.1 m and 21.3 m sites at each sampling time point was mean-centered with respect to itself followed by scaling by the variance of the corresponding column of the ten initial concentration matrices for each of the spill and 9.1 m sites. The variance of a column vector is $\sqrt{\Sigma x_i^2}$ where $x_i$ is the $i$th element of the column vector. The mean-centered and variance-normalized $\mathbf{X_1}$ is:

$$\begin{bmatrix} -0.6760 & -0.4681 & -0.2666 & -0.5810 & -0.1407 \\ 2.1633 & 1.7319 & 1.0123 & 2.1205 & 0.2879 \\ 0.1881 & -0.2004 & -0.2103 & 0.0237 & -0.1257 \\ 0.5996 & 0.4437 & 0.6101 & 0.4138 & 0.4091 \\ -0.7583 & -0.5024 & -0.3818 & -0.6590 & -0.1435 \\ -0.7583 & -0.5024 & -0.3818 & -0.6590 & -0.1435 \\ -0.7583 & -0.5024 & -0.3818 & -0.6590 & -0.1435 \end{bmatrix}$$

These two steps were done so that SVD can correctly identify the PC's and correctly determine the singular values. In addition, the scaling of the columns of the concentra-

tion matrices by the variances of the columns of the corresponding initial concentration matrices nullified the effect of the fluctuations in the initial concentration of the TEX compounds, such that any decrease in the total dispersion of the concentration values over time, as measured by the sum of the squares of the singular values, can be attributable to true biodegradation.

The sum of squares of the singular values (ssvd) of these ten mean-centered and variance-scaled matrices were then plotted against the average residence times of the TEX compounds in Fig. 4.10. Numbers 4 to 13 that appear in Fig. 4.10 represent the sampling order of the data points chronologically. The first three data points were not used because the residence times of most of the TEX compounds for the first three sampling points referred to period before the two-year monitoring work was carried out. The sum of squares of the singular values of a scaled matrix is a measure of the amount of dispersion exhibited by that matrix. If the biodegradation of the TEX compounds follows first-order kinetics, the curves in Fig. 4.10 should exhibit exponential decay kinetic profiles. The dashed lines in Fig. 4.10 are the least square curve fit to exponential decay curves. The first point is left out from the curve fit in Fig. 4.10a. It is regarded as an outlier.

So far, all the TEX compounds are assumed to behave similarly to allow one to focus on the steps of modeling a group of similarly behaved contaminants. However, the utility of the Multivariate Statistical Method is in its ability to detect the clustering of contaminants that behave similarly or the segregation of groups of contaminants that behave differently from each other. If the contaminants can be grouped into more than one group of similarly behaved contaminants, the degradation rate of each group of the contaminants can be modeled separately using similar steps just discussed.

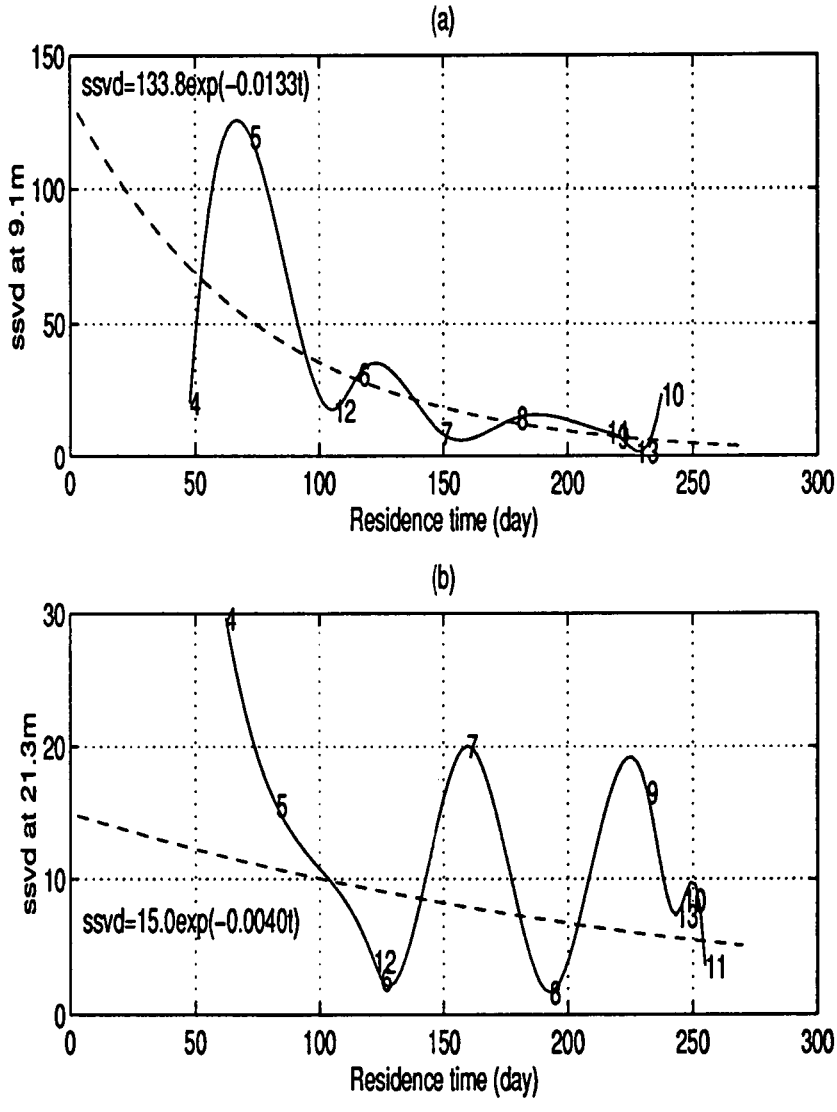



Figure 4.10: Sum of squares of the singular values (ssvd) vs. residence time (t) at 9.1 m site (a), and 21.3 m site (b). Their first-order degradation curve fits are ssvd = 133.8exp(-0.0133t) and ssvd = 15.0exp(-0.0040t) respectively.

The two examples in Chapter 3 explained how PCA and FA are used to detect variables that behave similarly. SVD is applied to the mean-centered and variance-normalized dilution corrected concentration matrix $\mathbf{X}_t$, whose rows represent the ten sampling time points and whose columns represent the TEX compounds, for each of the 9.1 m and 21.3 m sites. The concentration matrix is named $\mathbf{X}_t$ to differentiate it from another kind of concentration matrix $\mathbf{X}$ discussed earlier. The TEX compounds are arranged in the order of ethylbenzene, toluene, *m*-xylene, *o*-xylene, and *p*-xylene in $\mathbf{X}_t$. Fig. 4.11 shows that the first PC explains about 70% and 55% of the information contained in $\mathbf{X}_t$ already for the 9.1 m and 21.3 m sites respectively. The first three PC's together captures more than 95% of $\mathbf{X}$ for both sites. This means that the concentration of the five TEX compounds can be almost completely represented by just two and three PC variables respectively, with one dominating PC. The Gabriel's plots in Fig. 4.12 show that the vectors of the TEX compounds all point at a generally similar direction at both the 9.1 m and the 21.3 m site, indicating that the TEX compounds behave similarly to one another. Notice the subtle difference in the information conveyed by Fig. 4.11 and Fig. 4.12: Fig. 4.11 provide a quantitative measure of the information captured by the PC's, while the Gabriel's plots in Fig. 4.12 provide a more qualitative measure of correlation among the variables. Experience and knowledge on the natural biodegradation of BTEX compounds become important to better interprete Fig. 4.11 and Fig. 4.12. BTEX compounds, due to their similarity in their molecular structure, behave similarly during natural bioremediation. Thus the BTEX compounds can be treated as a group. The roughly 20% information of $\mathbf{X}$ that is captured by each of the second and the third PC's at each of the 9.1 m and 21.3 m sites respectively can be regarded as due to sampling inaccuracies and minor differences among the BTEX compounds. If pressed,

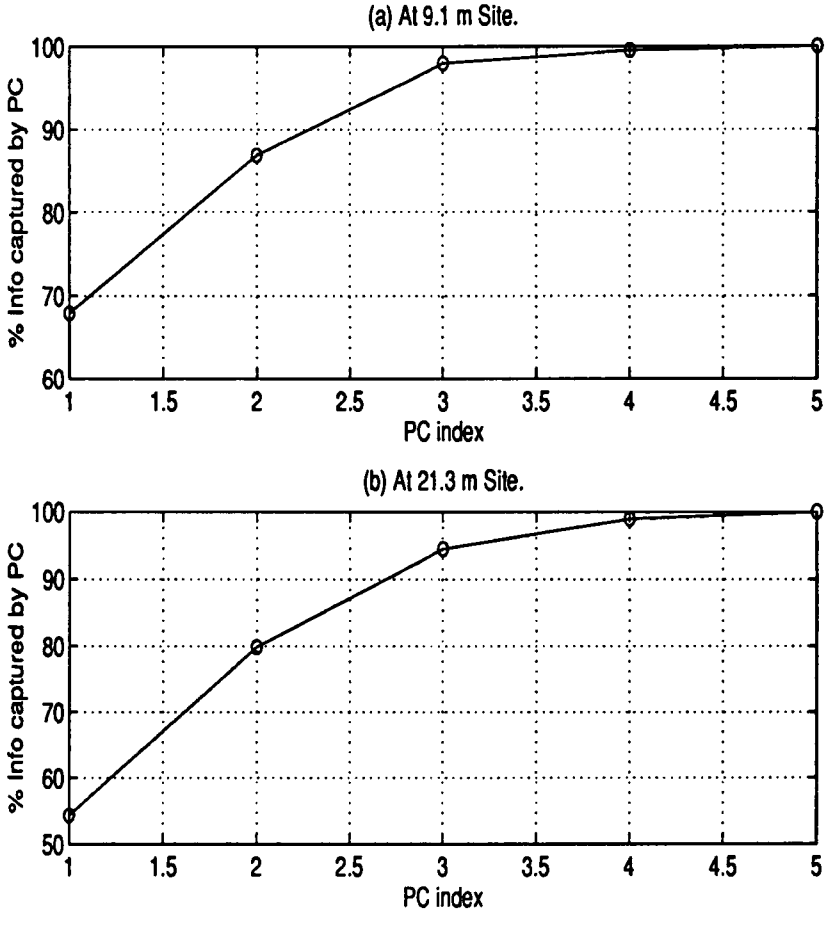



Figure 4.11: $\frac{\sum_{i=1}^{i=k} \sigma_i^2}{\sum_{j=1}^{j=n} \sigma_j^2}$ vs. $k$th PC at the 9.1 m and 21.3 m sites.

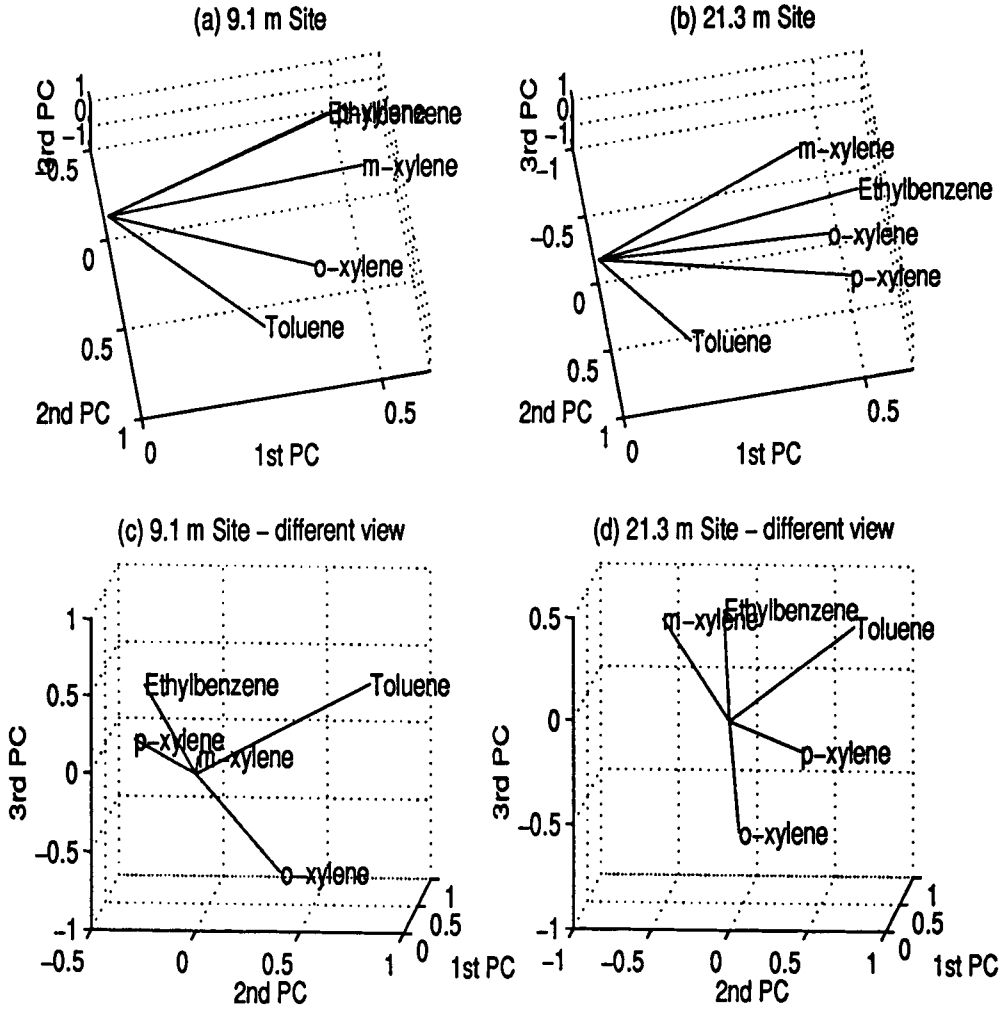



Figure 4.12: Gabriel's plots of the BTEX compounds at the 9.1 m and 21.3 m sites.

| 9.1 m Site | | | 21.3 m Site | | |
|---|---|---|---|---|---|
| 1st PC | 2nd PC | 3rd PC | 1st PC | 2nd PC | 3rd PC |
| 0.4791 | -0.3083 | 0.4050 | 0.5457 | -0.1299 | 0.4040 |
| 0.2688 | 0.8050 | 0.5001 | 0.1516 | 0.7717 | 0.4276 |
| 0.5332 | -0.0581 | -0.0772 | 0.4415 | -0.5007 | 0.3822 |
| 0.4034 | 0.3528 | -0.7605 | 0.4787 | -0.0241 | -0.6568 |
| 0.5011 | -0.3592 | 0.0388 | 0.5051 | 0.3692 | -0.2765 |

Table 4.3: First three PC's of the 9.1 m and 21.3 m sites.

one may claim that toluene perhaps behaves somewhat differently from the rest of the TEX compounds. This is concluded from examining the Gabriel's plot and the entries of the PC's as shown in Table 4.3.

The advantage of the MVS modeling approach over the traditional one-at-a-time approache is more apparent when there are multiple behaving groups of contaminants present at a site. The data from the Sleeping Bear Dunes Lakeshore cannot effectively bring forth this advantage of the MVS modeling approach because there is only one independently varying group of contaminants. Nevertheless, the Sleeping Bear Dunes Lakeshore data will be assumed to be composed of more than one group of similarly behaving contaminants in order to demostrate the advantage of the MVS modeling approach. Recall from Chapter 3 that the first PC is represented by the first column of the $\mathbf{V}$, where $\mathbf{V}$ is the right singular matrix of $\mathbf{X}_t$. The first three columns of $\mathbf{V}$ for each of the 9.1 m and 21.3 m sites are shown in Table 4.3. The squared of the elements in each columns of $\mathbf{V}$ represent the weights of the contributions of the original variables

to each of the new PC variables. Because the TEX compounds are arranged in the order of ethylbenzene, toluene, *m*-xylene, *o*-xylene, and *p*-xylene in $\mathbf{X}_t$, the elements in the first, second, third, fourth, and fifth row of $\mathbf{V}$ corresponds to ethylbenzene, toluene, *m*-xylene, *o*-xylene, and *p*-xylene, in that order. The first PC's shown in Table 4.3 show that toluene has lower weight coefficients, $0.2688^2$ and $0.1516^2$ for the 9.1 m and 21.3 m sites respectively, than the other TEX compounds, whereas it dominants the second PC's with weight coefficients of $0.8050^2$ and $0.7717^2$ for the 9.1 m and 21.3 m sites respectively. This suggests that toluene behaves slightly differently from the rest of the BTEX compounds. Toluene behaves only slightly differently from the rest of the BTEX compounds because $\mathbf{X}_t$ is largely representable by just one PC, i.e. the first PC. When toluene is deleted from the concentration matrices that are used to generate Fig. 4.10, the overall rate constants of the contaminants at the 9.1 m and the 21.3 m sites drop slightly from 0.0133 to 0.0132 per day and from 0.0040 to 0.0037 per day respectively. This suggests that toluene has higher than average rate constants at both the 9.1 m and 21.3 m sites. Indeed, this is the same conclusion arrived by Wilson et al. [55], using two data points, with DMP as the reference compound.

## 4.3 Discussion on the Sleeping Bear Dunes National Lakeshore Case Study

Both plots in Fig. 4.10 show that the sum of squares of the singular values of the scaled concentration matrices $\{\mathbf{X}\}$ at the 9.1 m and 21.3 m sites decrease as the average residence time increases. This decrease is expected since the effects of dilution due to dispersion of the contaminant plume and of the fluctuations in the initial concentration of the contaminants have already been nullified in $\{\mathbf{X}\}$ with the correction and normal-

ization factors that are described previously, leaving biodegradation the sole possible cause for any change in $\{\mathbf{X}\}$. The non-monotonous decrease in the curves of the sum of squares of the singular values of $\{\mathbf{X}\}$ in Fig. 4.10 is acceptable because the ten concentration matrices $\{\mathbf{X}\}$ from each of the 9.1 m and the 21.3 m sites, are representations of different batches of TEX compounds. The rate of decrease of each batch of TEX compounds as represented by each matrix of $\{\mathbf{X}\}$ depends on factors such as oxygen and nutrients level, pH, and temperature, which may be different for each of the ten batchs of TEX compounds.

Fig. 4.10b does not fit the exponential curve well. This poor exponential curve fit is probably due to the wider degree of dispersion and fluctuation associated with the longer time of travelling and to the assumptions made on $\{\mathbf{X}\}$. The assumptions adopted are that:

1. The BTEX compounds flow from the spill site to the 9.1 m site and from the 9.1 m site to the 21.3 m site.

2. $K_{oc}$'s of BTEX compounds are constants. This is an assumption because adsorption processes of various contaminants onto naturally occurring organic matter are typically exothermic reactions [45].

3. Eq. 2.3 is valid. However, studies by Garbarini and Lion [23] showed that oxygen content in soil is also an important factor to the value of $K_d$ for toluene and TOC other than $f_{oc}$, $B$, and $n$. Oxygen content is not taken into account in Eq. 2.3.

4. The estimates of $f_{oc}$, $B$, and $n$ are accurate for the site and are constant.

5. Benzene does not biodegrade.

6. Benzene does not vaporize more quickly after it is dissolved in groundwater than the other BTEX compounds even though benzene has the highest volatility among the BTEX compounds.

7. The measured concentrations reflect those BTEX compounds in the groundwater embedded in the soil matrix and not those BTEX compounds in the dense layer floating on top of the water table.

Assumption 1 probably introduces most uncertainty to the analysis. If the BTEX compounds did not actually flow from the spill site to the 9.1 m site and from the 9.1 m site to the 21.3 m site, then initial concentration of BTEX compounds could not be determined correctly. Assumption 4 also introduces some uncertainty to the case study because the values of $f_{oc}$, $B$, and $n$ were used in the calculations of the residence times of the BTEX compounds. Residence times were the basic information that were used throughtout the case study to nullify effects other than those due to natural biodegradation on the changes in the concentration of BTEX compounds. Assumption 3 should be a moderately good one since Wilson had concluded that benzene did not biodegrade appreciately relative to the other BTEX compounds. Any biodegradation of benzene would cause the dilution correction factors to be incorrectly higher than they should be.

From the exponential curve fits in Figure 4.10, the rate constants of 0.0133 and 0.0040 per day for the 9.1 m and 21.3 m sites respectively are obtained. These rate constants, determined by MVS approach and whose intermediate calculations use benzene as the dilution corrector, are similar in magnitude to the average rate constants calculated from the individual rate constants arrived at by Wilson et al. [55] which uses an one-variable-at-a-time modeling approach and DMP for dilution correction. The degradation

rate constants of ethylbenzene, toluene, *m*-xylene, *o*-xylene, and *p*-xylene calculated by Wilson et al. [55] range from 0.0031 to 0.011 per day, 0.023 to 0.067 per day, 0.0037 to 0.014 per day, 0.0040 to 0.016 per day, and 0.0024 to 0.0096 per day respectively. Now, if we take the average of rate constants of each of the TEX compound calculated by Wilson et al. and multiply each of this average with its corresponding weight (squared of the corresponding element) in the first PC for each of the 9.1 m and 21.3 m sites in Table 4.3, and sum the products, we obtain the values of 0.0105 per day and 0.0087 per day for the 9.1 m and 21.3 m sites respectively. These values are remarkably close to the rate constants of 0.0133 per day for the 9.1 m site and 0.0040 per day for the 21.3 m site which are determined by the MVS approach and whose intermediate calculations use benzene as the dilution corrector. The rate constants from Wilson et al. and the MVS approach for the 21.3 m site are less close to each other (i.e. 0.0087 and 0.0040 per day) as compared to the rate constants from Wilson et al. and the MVS approach for the 9.1 m site (i.e. 0.0105 and 0.0133 per day). This can be explained by the poorer exponential curve fit that is evident in Fig. 4.10b as compare to the better exponential curve fit in Fig. 4.10a. Because the rate constant from the MVS approach is based on ten sampling point and on the entirety of the data set, it is relatively robust against a certain degree of errors in sampling, which is invariable in all cases.

Recall that to nullify the fluctuations in the magnitude of the ten consecutive concentration matrices, $\{\mathbf{X}\}$, due to the fluctuations of the initial concentration of the BTEX compounds, each of the ten consecutive $\{\mathbf{X}\}$ was variance-normalized with the variances of its corresponding initial concentration matrix. Given the conclusion arrived at in Chapter 3 that the magnitude of the ssvd of a matrix is no longer related to the actual magnitudes of the matrix after the matrix is variance-normalized, it might be at first

surprising to see that the rate constants determined by the MVS approach should agree with the rate constants calculated by Wilson et al. [55]. A look at the solution of the first-order kinetic equation will explain why this is the case.

$$ln(\text{ssvd}) = ln(\text{ssvd}_o) - k\tau$$

where

$ln(\text{ssvd}_o)$= natural logarithm of ssvd of the initial concentration matrix

$ln(\text{ssvd})$ = natural logarithm of ssvd of a concentration matrix

$k$ = first-order rate constant (per day)

$\tau$ = residence time (day)

Notice that any normalization factor that is applied to ssvd can be extracted from the logarithm of ssvd and added to $\ln(\text{ssvd}_o)$. Thus any normalization factor that is applied to ssvd will affect $\text{ssvd}_o$ but will not change $k$. In other words, the rate constant $k$ is scale invariant. MVS modeling approach outlined here can quantify the magnitude of the degradation of a group of similarly behaving contaminants because of its ability to treat the group of like compounds as one.

MVS approach appears to be a powerful method to model natural bioremediation because of its ability to detect the clustering of contaminants that behave similarly and quantify the degradation rate of each group of the contaminants independently from the other groups. The MVS approach also allows:

1. The prediction of the time required to naturally bioremediate a contaminated site to a desired low level. With this prediction ability, decision makers in the environmental cleanup industry can allocate funds to sites that take unacceptably

long time to naturally bioremediate and let sites that can be suitably naturally bioremediated to be so-remediated, at a much reduced cost.

2. The monitoring of the few representative contaminants for each similarly behaving group in lieu of all the contaminants in that group at a contaminated site. This can be done only after experience is gained from applying the MVS approach to natural bioremediation of contaminants under various conditions. Sampling costs can be greatly reduced if only the representative compounds from each independently varying group is to be sampled, and the concentrations of the correlated compounds in the group can be inferred from the underlying PCA model.

3. A clearer interpretation of monitoring field data of natural bioremediation because by grouping similarly behaved contaminants, the bigger picture of the dynamics of a natural bioremediation process become more tractable.

Knowledge and experience on the physical processes that are modeled by the MVS approach are paramount in interpreting the MVS modeling and analysis results. Knowledge and experience on the physical processes to be modeled by the MVS approach are needed for the following reasons:

1. Confirm the conclusions on the grouping of contaminants drawn from the Gabriel's plots. Confirmation is needed because Gabriel's plots are only numerical tools, based on mathematical analysis of field data, with no injection of human knowledge concerning the process.

2. Recognize measurement errors and inconsistencies.

An implicit assumption of this MVS approach that has been made is that the mech-

anisms of contaminant remediation in a contaminated site are unchanged once the contaminants are grouped and their first-order rate constants are calculated using the initial monitoring field data of the contaminated site. If the mechanisms of the remediation are radically changed due to, for instance, the sudden addition of nutrients, or depletion of selective indigenious microorganisms, the grouping of the contaminants and their first-order degradation rate constants could also be altered, thus destroying the predictive ability of the MVS approach. Barring active intervension of a contaminated site, after monitoring the natural degradation for a period of time (1-2 years), one can develop a model of the degradation kinetics using multivariate statistical approach, similar to that reported here. One can then predict and extrapolate additional time required to render the site suitable for reuse.

In addition, if one decided to seek active remediation measures, one can still use the MVS approach to help contrasting the contamination scenarios before and after active treatment, and to assess the increased degradation rates due to active treatment.

# CHAPTER 5

# SUMMARY & RECOMMENDATIONS FOR FUTURE RESEARCH

## 5.1 Summary

This thesis provides a concise modeling approach on natural bioremediation processes using the Multivariate Statistical (MVS) approach. While traditional environmental modeling approaches treat each contaminant one-at-a-time and produce many scattered and sometimes conflicting results, MVS approach treats a like group of contaminants that behave similarly yet independently from other like groups, as one group, and tracks its behavior over time. MVS approach results in a clearer understanding of the overall degree of remediation because there are fewer variables to track, and all variables can be grouped into mutually independent acting groups with decreasing importance.

Modeling of bioremediation requires that the field data be processed such that the resultant data solely reflect the effects of bioremediation. A case study was done on modeling the natural bioremediation of BTEX compounds at the Sleeping Bear Dunes National Lakeshores. The changes in the concentration of ethylbenzene, toluene, *m*-xylene, *o*-xylene, and *p*-xylene (TEX compounds) in a result of the composite actions of dilution due to the dispersion of the contaminant plume, fluctuations in the initial concentration of the TEX compounds, biodegradation of the BTEX compounds, and the rise and fall of the water table. The TEX concentration profiles were first dilution corrected by multiplying the TEX concentrations with the ratio of the initial concentration of benzene to the concentration of benzene at each sampling time point. Benzene

is chosen as the dilution corrector because it was found to not biodegrade to any significant degree [55]. The effect of the fluctuations in the intensity of the source of TEX compounds was nullified by normalizing the TEX concentrations with the variances of their corresponding initial concentrations.

Principal Component Analysis was then used to detect and group contaminants that behave similarly over a period of observation and to track its degradation kinetics. At each sampling time point, the measurements of similarly behaved contaminants were combined into a matrix and the sum of squares of the singular values (ssvd) of the matrix was calculated. Assuming first-order kinetics, the first-order rate constant was then determined from the curve fit of the points in the plot of ssvd versus average residence time. Our approach yielded an overall degradation first-order rate constant of 0.0133 per day at the 9.1 m site.

Wilson et al. [55] used 2,3-dimethylpentane (DMP) instead of benzene to nullify the dilution of the BTEX compounds due to the dispersion of the contaminant plume. However since DMP was measured only at the last two sampling time points, Wilson's calculation of rate constant is not robust. Further, the traditional one-at-a-time analysis approach would yield many calculated rate constants, one for each compound, even though there may only be one class of compound. Our analysis approach is more robust in the sense that any errors that may be present in the concentration measurements would not affect appreciably the results, since the entire set of data over a two-year monitoring period is used in the model building.

Like all modeling approaches, MVS modeling approach requires the knowledge and experience on the processes to be modeled. Knowledge and experience on the physical processes to be modeled by the MVS approach are needed to confirm numerical conclu-

sions on the grouping of contaminants drawn from the PCA and Gabriel's plots analysis and to recognize measurement errors and inconsistencies.

It is emphasized that the MVS approach takes all the data in together, analyzes them coordinatedly, and arrives at one overall kinetic analysis. The calculated rate constant by the MVS approach is similar to the weighted average of the individual rate constants determined by Wilson et al. [55]. The latter would show a degree of scattering since individual rate constants are calculated one for each compound. The MVS approach surpasses the traditional one-at-a-time approach in terms of analysis and interpretation of the results when many different types of contaminants are present at a contaminated site. The MVS approach also has the ability to contrast two distributions of concentration data in the overall sense, extracting the dominating similarities and differences. The traditional one-at-a-time approach cannot do these.

## 5.2 Recommendations for Future Research

The layout of the monitoring boreholes at the Sleeping Bear Dunes National Lakeshore are recommended for sites where natural bioremediation processes are monitored. Two important aspects of the layout of the monitoring boreholes are:

1. The monitoring boreholes are clustered to a few groups so that each cluster of monitoring boreholes straddles an interval of varying depths. The monitoring boreholes at each cluster should have depths that are both shallower and deeper than the main body of the contaminants in the aquifer, in addition to the depths where the main body of the contaminants is, to ensure that the monitoring boreholes are monitoring the main bulk of the contaminants.

2. The clusters of monitoring boreholes are located such that the contaminants flow from one cluster of monitoring boreholes to the next cluster of monitoring boreholes directly down gradient as much as possible. This enables one to calculate the initial concentrations of the contaminants, necessary for achieving proper scaling, such that any real change over time is discernable.

Establishing the residence times of contaminants dissolved in and moving with groundwater is a cornerstone of the modeling effort and require detailed and accurate information on the water table level, percentage of naturally occurring organic carbon in the soil at the contaminated site, the soil porosity, the soil bulk density, and the contaminants' soil/sediment adsorption coefficients. In future natural bioremediation monitoring studies, these measurements must be recorded accurately, at close time intervals, and over a period of time that is commensurate with the hydraulic gradient at the contaminated site. The measurements must be done at close intervals such that the seasonal changes of the quantities measured are followed. The smaller the hydraulic gradient at a contaminated site, the longer the period of monitoring time is required in order to back-calculate the initial concentration of the contaminants at the corresponding residence time prior to the sampling time. The exact period of monitoring time required varies from site to site depending on the particular site's hydraulic gradient. These measurements directly affect the value of the dilution correction factors, thus the calculation of the degradation rate constant.

Future research should examine how to best generalize the model of natural bioremediation from one contaminated site to another contaminated site with similar physical and chemical characteristics. Being able to have a reliable model to predict the time required to naturally bioremediate a contaminated site to an acceptable degree, without

first having a monitoring field data of the contaminated site for a certain period of time, as that is required for the MVS approach discussed in this thesis, to group like contaminants and to determine the first-order rate constants of the groups of contaminants, will further reduce the cost of the site remediation effort. A regression analysis can also be done to relate the rate constants of the groups of contaminants in a contaminated site to the environmental factors such as nutrient levels and oxygen content, temperature, and pH at the contaminated site. The resulting regression coefficients are indicators of the degree of importance of these factors towards the natural bioremediation of the contaminants under study. Grouping and modeling contaminants that behave similarly and thus reducing the number of variables that characterize a contaminated site, as are done in this thesis, are the positive steps towards achieving this goal of creating a generalizable model.

MVS modeling approach helps policy makers in environmental cleanup industry in deciding whether natural bioremediation is an appropriate remediation technique to their contaminated sites by providing the policy makers concise and meaningful information, i.e. a generalized rate constant for each group of similarly behaving contaminants. With the MVS models, substantial site remediation costs can be expected to be avoided by the environmental cleanup industry by employing natural bioremediation as the remediation technique for all suitable contaminated sites. When enough experience is gained, the sampling cost of monitoring the contaminated site can be further reduced by sampling only those few representative contaminants of each group at the site. The concentration of other unsampled contaminants can be inferred from the measurements of the few representative contaminants. The benefits of using the MVS modeling approach become evident when the number of contaminant species in a contaminated site is large and the

interactions among the contaminants are complex and not precisely known. For a site such as this, traditional one-at-a-time approaches to natural bioremediation modeling could become so complex that the overall picture of the degree of remediation of the site could be obscured by the details and become intractable. The MVS approach is superior to the traditional one-at-a-time approach because all the information is utilized and treated as a whole and the interrelationships between the contaminant measurements are analyzed in a coordinated manner, with the redundant information discarded first. It not only provides a reasonable approach for data reduction, but it also effectively clusters variables into just a few independently varying families of variables whose kinetics can be tracked much more easily over time.

It is recommended that addition sets of natural bioremediation data be obtained and subjected to MVS analysis in order to further validate the feasibility of such modeling approach. Especially desirable are data sets that contain several different types of contaminant hydrocarbons such that segregation into more than one independent acting clusters by the MVS analysis approach is expected. The advantages of using multivariate statistical approach to analyze data of such complexity should then become self-evident.

# BIBLIOGRAPHY

# BIBLIOGRAPHY


[1] Abdul, S. A., T. L. Gibson, and D. N. Rai. "Statistical Correlations for Predicting the Partition Coefficient for Nonpolar Orgainc Contaminants Between Aquifer Organic Carbon and Water." *Haz. Waste Haz. Mater.* 4.3(1987):211-222.

[2] Afifi, A. A. and Virginia Clark. *Computer-Aided Multivariate Analysis.* 2nd ed. New York:Van Nostrand Reinhold, 1990.

[3] Ronald M. Atlas. *Petroleum Microbiology.* New York:Macmillan, 1984.

[4] Baross, J. A., M. D. Lilley, and L. I. Gorden. "Is the $CH_4$, $H_2$, and CO venting from submarine hydrothermal systems produced by thermophilic bacteria?" *Nature.* London, 1982.

[5] Bird, R. B., Warren E. Stewart, and Edwin N. Lightfoot. *Transport Phenomena.* Singapore:Wiley, 1960.

[6] Borden, Robert C. "Natural Bioremediation of Hydrocarbon-Contaminated Ground Water." Ed. John E. Matthews. *Handbook of Bioremediation.* Boca Raton:Lewis Publishing, 1994.

[7] Chen, R. L., D. R. Keeney, and A. G. Hashimoto, A. G. "Effect of chemistry on methane fermentation kinetics of beef-cattle manure." *Anaerobic Digestion.* Ed. D. A. Stafford. Cardiff:Scientific, 1(1980).

[8] Cozzarelli, Isabelle M., Robert P. Eganhouse, Mary Jo Baedecker. "Transformation of Monoaromatic Hydrocarbons to Organic Acids in Anoxic Groundwater Environment." *Environmental Geology and Water Science.* 16.2(1990):135-141.

[9] Crank, J. *The Mathematics of Diffusion.* 2nd ed. Oxford:Clarendon Press, 1975.

[10] *CRC Catalog: Environmental Sciences.* Boca Raton, FL:Lewis Publishers, 1994.

[11] Davis, John C. *Statistics and Data Analysis in Geology.* 2nd ed. New York:John Wiley & Sons, 1986.

[12] Delong, Lois Anne. "Site Remediation." *Chemical Engineering Progress.* Ed. Mark D. Rosenzweig. May 1994:29.

[13] Dragun, James. *The Soil Chemistry of Hazardous Materials.* Greenbelt, MA:Hazardous Materials Control Resources Institute, 1988.

[14] Duewer, D. L., B. R. Kowalski, and T. F. Schatzki. "Source Identification of Oil Spills by Pattern Recognition Analysis of Natural Elemental Composition." *Analytical Chemistry.* 47(1975):1573-1583.

[15] Edwards, E. A., L. E. Wills, D. Grbic-Galic, and M. Reinhard. "Anaerobic Degradation of Toluene and Xylene – Evidence for Sulfate as the Terminal Electron Acceptor." In: *In-situ Bioreclamation, Applications and Investigations for Hydrocarbon and Contaminated Site Remediation.* Eds. R. E. Hinchee and R. F. Olfenbuttel. Butterworth-Heinemann:Boston, MA, 1991.

[16] Edwards, E. A., L. E. Wills, M. Reinhard, and D. Grbic-Galic. "Anaerobic Degradation of Toluene and Xylene by Aquifer Microorganisms Under Sulfate-reducing Conditions." *Appl. Environ. Microbiol.* 58.3(1992):794-800.

[17] Edwards, E. A., and D. Grbic-Galic. "Complete Mineralization of Benzene by Aquifer Microorganisms Under Strictly Anaerobic Conditions." *Appl. Environ. Microbiol.* 58.8(1992):2663-2666.

[18] Wisconsin Department of Natural Resources Emergency and Remedial Response Section. *Release News: ERRP Issues Guidance on Natural Biodegradation.* Vol. 3, No. 1, February 1993.

[19] *Federal Register.* Vol. 58, No. 46, March 11,1993.

[20] Ferry, James G., ed. *Methanogenesis.* New York:Chapman & Hall, 1993.

[21] Freeze, D. A. and J. A. Cherry. *Groundwater.* Englewood Cliffs, NJ:Prentice Hall, 1974.

[22] Gaarenstroom, Phillip D., Sam P. Perone, and Jarvis L. Moyers. *Application of Pattern Recognition and Factor Analysis for Characterization of Atmospheric Particulate Composition in Southwest Desert Atmosphere.* 11 (1977): 795-800.

[23] Garbarini, Doug R. and Leonard W. Lion. "Influence of the Nature of Soil Organics on the Sorption of Toluene and Trichloroethylene." *Environmental Science and Technology.* Vol. 20, No. 12, 1986, 1263-1269.

[24] Gelb, Arthur, ed. *Applied Optimal Estimation.* Cambridge:MIT Press, 1974.

[25] Hadley, Paul W. and Richard Armstrong. " 'Where's the Benzene?' – Examining California Ground-Water Quality Surveys." *Ground Water.* 29.1(Jan-Feb 1991):35:40.

[26] Hopke, Philip K., Ernest S. Gladney, Glen E. Gordon, William H. Zoller, and Alun G. Jones. "The Use of Multivariate Analysis to Identify Sources of Selected Elements in the Boston Urban Aerosol." *Atmospheric Environment.* Vol. 10, 1976.

[27] Komaya, T. "Gaseous metabolism in lake muds and paddy soils." *Advances in Organic Geochemistry.* Oxford:Pergaman Press, 1963.

[28] Lee, M. D., J. T. Wilson, and C. H. Ward. "In Situ Restoration Techniques for Aquifers Contaminated with Hazardous Wastes." *Journal of Hazardous Materials.* 14 (1987) 71-82.

[29] Neely, N. S., J. S. Walsh, D. P. Gillespie, and F. J. Schauf. "Remedial Actions at Uncontrolled Waste Sites." Ed. D. W. Shultz. *Land Disposal: Hazardous Waste.* Proceedings of 7th Annual Res. Symp., Philadelphia, 1981, U. S. Environmental Protection Agency EPA-600/9-81-002b, Cincinnati, pp. 312-319, 1988

[30] Norris, Robert D. "In-Situ Bioremediation of Soils and Ground Water Contaminated with Petroleum Hydrocarbons." Ed. John E. Matthews. *Handbook of Bioremediation.* Boca Raton:Lewis Publishing, 1994.

[31] Marsily, Ghislain de. *Quantitative Hydrogeology: Groundwater Hydrology for Engineers.* Trans. Gunilla de Marsily. Florida:Academic Press, 1986.

[32] McCarty, P. L. "The methane fermentation." *Principles and Applications in Aquatic Microbiology.* Eds. H. Henkelekian and N. C. Dondero. New York:John Wiley, 1964.

[33] McCarty, P. L. "Anaerobic waste treatment fundamentals. III Toxic materials and their control." Public Works, 1964.

[34] McGill, W. B., M. J. Rowell, and D. W. S. Westlake. "Biochemistry, ecology, and microbiology of petroleum components in soil." *Soil Biochemistry.* Eds. E. A. Paul and J. N. Ladd. New York:Marcel Dekker. 5(1981):229-296.

[35] Montgomery, John H. and Linda M. Welkom. *Groundwater Chemicals Desk Reference.* Chelsea, MI:Lewis Publishers, 1990.

[36] Palmer, Christopher M. *Principles of Contaminant Hydrogeology.* Chelsea, MI:Lewis Publishing, 1992.

[37] Piovoso, Michael J. "Process Data Chemometrics." *IEEE Transactions on Instrumentation and Measurement.* 41(1992):262-268.

[38] Powell, Daniel M. "Selecting Innovative Cleanup Technologies: EPA Resources." *Chemical Engineering Progress.* Ed. Mark D. Rosenzweig. 90.5(May 1994):33-41.

[39] Rawlings, John O. *Applied Regression Analysis: A Research Tool.* Pacific Grove: Wadsworth & Brooks/Cole, 1988.

[40] Razack, M., and J. Dazy. "Hydrochemical Characterization of Groundwater Mixing in Sedimentary and Metamorphic Reservoirs with Combined use of Piper's Principle and Factor Analysis." *Journal of Hydrology.* 114(1990):371-393.

[41] Reinhard, Martin. "In-situ Bioremediation Technologies for Petroleum-Derived Hydrocarbons Based on Alternative Electron Acceptors (Other Than Molecular Oxygen)." Ed. John E. Matthews. *Handbook of Bioremediation.* Boca Raton:Lewis Publishing, 1994.

[42] Reyment, Richard, and K. G. Jöreskog. *Applied Factor Analysis in the Natural Sciences.* New York:Cambridge University Press, 1993.

[43] Safonov, Michael G., Alan J. Laub, and Gary L. Hartmann. "Feedback Properties of Multivariable Systems: The Role and Use of the Return Difference Matrix." *IEEE Transaction on Automatic Control.* Vol. AC-26, No. 1, February 1981.

[44] Schwarzenbach, R. P. and J. Westall. "Transport of Nonpolar Organic Commpounds from Surface Water to Groundwater. Laboratory Sorption Studies." *Environ. Sci. Technol.* 15(11):1360-1367 (1981).

[45] Smith, Lawrence A. and Robert E. Hinchee. *In Situ Thermal Technologies for Site Remediation.* Boca Raton, FL:Lewis Publishers, 1992.

[46] Stolzy, Lewis H. and William A. Jury. "Soil Physics." *Handbook of Soils and Climate Agriculture.* Ed. Victor J. Kilmer. Boca Raton, FL:CRC Press, 1982.

[47] Strang, Gilbert. *Linear Algebra and its Applications.* 3rd ed. San Diego:Hancourt, 1988.

[48] Testa, Stephen M. and Duane L. Winegardner. *Restoration of Petroleum Contaminated Aquifers.* Chelsea, MI:Lewis Publishers, 1991.

[49] Troy, M. A., McGinn, S. R., Greenwald, B. P., Jerger, D. E., and Allen, L. B. "Bioremediation of Diesel Fuel Contaminated Soil at a Former Railroad Fueling Yard." *Soil Contamination: Diesel Fuel.* Eds. P. T. Kostecki and E. J. Calabrese. Chelsea:Lewis.

[50] Vestal, J. Robie. "The Metabolism of Gaseous Hydrocarbons by Microorganisms." *Petroleum Microbiology.* Ed. Ronald M. Atlas. New York:Macmillan, 1984.

[51] Viessmann, Warren, Jr., John W. Knapp, Gary L. Lewis, and Terrence E. Harbaugh. *Introduction to Hydrology.* 2nd ed. New York:Harper & Row, 1977.

[52] Volk, Bob G. and Richard H. Loeppert. "Soil Organic Matter." *Handbook of Soils and Climate in Agriculture.* Ed. Victor J. Kilmer. Boca Raton, FL:CRC Press, 1982.

[53] Wilson, J. T., D. H. Kampbell, and J. Armstrong. "Natural Bioreclamation of Alkylbenzenes (BTEX) from a Gasoline Spill in Methanogenic Groundwater." *In the proceddings of the Second International Symposium on In Situ and On Site Bioreclamation.* San Diego, California, 1993.

[54] Wilson, B. H., J. T. Wilson, D. H. Kampbell, B. E. Bledsoe, and J. M. Armstrong. "Biotransformation of Monoaromatic and Chloromated Hydrocarbons at an Aviation Gasoline Spill Site." *Geomicrobiology Journal.* 8(1990):225-240.

[55] Wilson, J. T., D. H. Kampbell, and J. Armstrong. "Natural Bioreclamation of Alkylbenzenes (BTEX) from a Gasoline Spill in Methanogenic Groundwater." In: *Proceedings, Second International Symposium on In Situ and on Site Bioreclamation.* San Diego, CA, April 5-8, 1993.

[56] Winfrey, Michael R. "Microbial Production of Methane." *Petroleum Microbiology.* Ed. Ronald M. Atlas. New York:Macmillan, 1984.

[57] Wise, B. M., N. L. Ricker, D. F. Veltkamp, and B. R. Kowalski. "A Theoretical Basis For The Use Of Principle Component Models for Monitoring Multivariate Processes." *Process Control and Quality.* 1(1990).

[58] Wold, Svante, Kim Esbensen, and Paul Geladi. "Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems.* 2 (1987): 37-52.

[59] Wolff, Diane D. and Michael L. Parsons. *Pattern Recognition Approach to Data Interpretation.* New York:Plenum, 1983.

[60] Zehnder, A. J. B., K. Ingvorsen, and T. Marti. "Microbiology of methane bacteria." *Anaeronic Disgestion.* Eds. D. E. Hughes, D. A. Staford, B. I. Wheatley, W. Baader, G. Lettinga, E. J. Nyns, W. Verstraette, and R. L. Wentworth. Elsevier Biomedical:Amsterdam, 1982.

[61] Zehnder, A. J. B. and K. Wuhrmann. "Physiology of Methanobacterium strain AZ." *Arch. Microbiol.* 1977.

[62] Zeikus, J. G., M. R. Winfrey. "Temperature limitation of methanogenesis in aquatic sediments." *Appl. Environ. Microbiol.* 1976.

# APPENDICES

# APPENDIX A

## Data from the Sleeping Bear Dunes National Lakeshore

t_day_c' = Sampling time points in Julian dates.
Note that measurements of alkalinity, ammonia, calcium, iron, nitrate, oxygen, pH, phosphates, sulfate, temperature, and TOC are available for the first 11 sampling time points only. Their measurements at 569.6 and 687.9 Julian dates are not available.
H20level_213' = Water level at 21.3 m Site.
H20level_9.1' = Water level at 9.1 m Site.
H20level_spill' = Water level at Spill Site.

| t_day_c' | H20level_spill' | H20level_91' | H20level_213' |
|---|---|---|---|
| 53.0000 | 584.9550 | 584.9330 | 584.8950 |
| 78.9000 | 585.2430 | 585.1670 | 585.0520 |
| 111.8000 | 585.4980 | 585.4350 | 585.3270 |
| 139.8000 | 585.1230 | 585.0870 | 585.0350 |
| 167.8000 | 584.7950 | 584.7920 | 584.7900 |
| 203.5000 | 584.7850 | 584.8020 | 584.8240 |
| 230.8000 | 584.7680 | 584.8000 | 584.8030 |
| 258.1000 | 584.8120 | 584.7970 | 584.8380 |
| 294.5000 | 584.9120 | 584.9180 | 584.9130 |
| 323.9000 | 585.3270 | 585.2850 | 585.2120 |
| 349.8000 | 585.4620 | 585.4180 | 585.3100 |
| 569.6000 | 584.8800 | 584.8900 | 584.9060 |
| 687.9000 | 585.4470 | 585.4050 | 585.3420 |

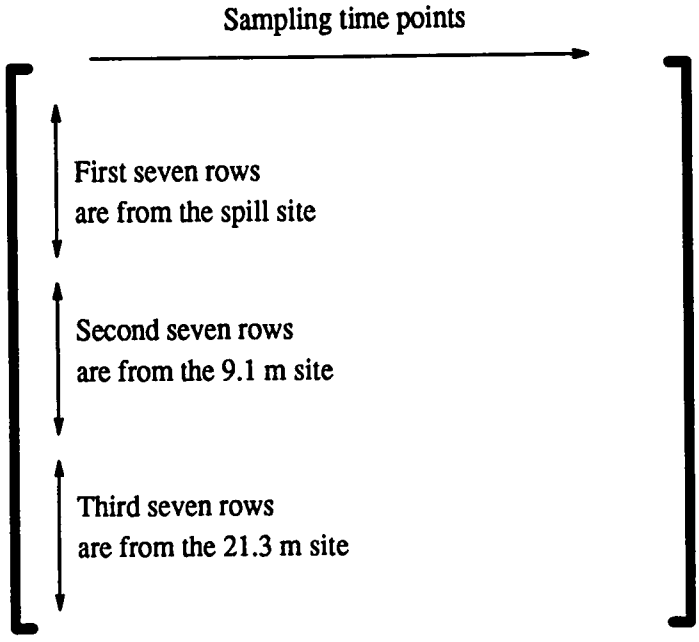



Figure A.1: Data format for alkalinity, ammonia, benzene, calcium, ethylbenzene, iron, *m*-xylene, nitrate, *o*-xylene, oxygen, *p*-xylene, pH, phosphates, sulfate, temperature, TOC, and toluene.

```
alkalinity' =
   650   368   320   259   294   336   356   367   354   344   334
   702   394   378   339   430   366   373   367   423   326   354
   440   331   278   289   314   306   340   275   303   340   379
   408   277   272   254   278   267   280   281   288   288   315
   200   157   172   172   142   139   163   161   213   173   182
   174   127   166   167   171   138   171   150   148   154   166
   164   143   146   138   136   128   140   130   150   143   139
   126    78   109   144   159   258   172   151   123   108   110
   376   272   354   337   427   252   482   227   347   306   301
   550   450   453   399   436   495   494   444   430   410   458
   522   356   404   349   290   270   371   311   238   302   336
   344   237   225   169   219   176   188   153   172   158   210
   256   175   183   182   180   165   196   188   172   155   136
   210   178   174   151   151   129   153   141   144   126   180
   320   240   242   244   241   290   288   284   270   234   271
   451   352   313   295   369   341   242   195   289   239   309
   356   259   216   198   219   300   283   324   268   279   293
   318   215   187   168   188   191   197   223   283   217   277
   248   194   180   151   187   157   144   144   201   172   186
   218   141   153   152   161   134   134   115   120   126   152
   169   138   114   106   127   118   126   129   121   110   129
```

```
ammonia' =
  Columns 1 through 7
   12.2900    5.7100    1.4000    1.4000    5.6000    1.2000    0.8400
    8.7800    5.0600    2.2000    2.0000    6.2000    2.0000    0.8600
    3.1700    6.0000    1.1000    0.8400    7.6000    0.8200    0.9800
    3.0500    2.5300    0.4600    0.4700    4.7000    0.4000    0.9200
    2.6300    1.1000    0.6400    0.6100    6.6000    0.6500    0.8000
    1.2000    0.7800    0.6000    0.6100    6.0000    0.6500    0.9600
    0.4900    3.6700    0.5600    0.3200    2.6000    0.3900    0.8400
   11.2500    8.6200    0.5900    0.2700    5.4000    0.8200    1.1000
    4.7800   10.8700    2.1000    1.4000    4.1000    0.8800    2.1000
    4.6100    4.6900    0.9000    0.8300    5.2000    0.4300    1.0000
    6.4300    2.0900    1.3000    1.0000    1.2000    0.8300    0.0200
    6.5200    1.1800    0.7000    0.7100    1.2000    0.9400    0.5100
    1.2900    2.7800    0.3200    0.4400    2.9000    0.3900    0.9800
    5.4800    0.9100    0.4500    0.3300    1.3000    1.1000    0.2900
    8.4200    5.6000    1.4000    1.4000    2.7000    1.7000    2.0000
    4.8700    3.6800    2.1000    2.4000    2.8000    2.5000    1.9000
    5.4800    1.2400    0.6100    0.8900    1.4000    0.7200    1.3000
    6.9500    3.4400    0.2900    0.5700    0.8700    0.2600    0.4300
    3.0900    4.6600    0.3700    0.8400    2.2000    0.2500    0.2900
    0.6700    0.3400    0.1000    0.3300    0.5200    0.1600    0.4200
    4.5700    0.5400    3.5000    0.2200    0.3400    0.1400    0.1900

  Columns 8 through 11
    1.9000    0.9600    0.9600    2.2800
    2.4000    0.9800    0.9600    2.2000
    0.8600    0.7500    0.4200    1.9000
    0.4300    0.5400    0.3000    1.0500
    0.5000    0.5700    0.4000    0.0050
    0.4500    0.4800    0.3800    0.3500
    0.2900    0.3600    0.3000    0.2400
    0.6700    0.2200    0.4000    0.3400
    1.8000    1.3000    1.2000    1.3700
    1.7000    0.9900    0.8900    0.9100
    0.0600    0.6900    0.5000    0.6900
    0.5200    0.5600    0.4800    0.6900
    0.4200    0.3600    0.3200    0.2100
    0.2800    0.2700    0.2400    0.1700
    2.0000    1.7000    0.6500    1.0600
    1.6000    1.4000    1.0000    1.2600
    1.3000    0.7200    0.6400    0.6300
    0.6000    0.5100    0.3600    0.4000
    0.3200    0.4500    0.3200    0.3400
    0.2200    0.0800    0.1000    0.1300
    0.0900    0.2700    0.0050    0.0050
```

```
benzene' =
   1.0e+03 *
  Columns 1 through 7
    6.9300    4.6500    3.7300    3.2200    2.8460    2.7100    2.5700
    7.0600    4.1800    2.5700    2.5000    1.5360    2.1800    2.2900
    2.2000    1.1700    0.6600    0.0710    0.7330    0.6000    0.6600
    0.9700    0.5500    0.3600    0.2200    0.1920    0.1700    0.0880
    0.0040    0.0070    0.0010         0         0         0         0
    0.0130         0         0         0         0         0         0
    0.0040         0         0         0         0    0.0020         0
         0         0         0         0    0.0020    0.0140    0.0270
    0.0320    0.1300    0.7700    0.2000    0.1340    0.1600    0.5100
    0.0460    0.1100    0.2400    0.1100    0.0580    0.5200    0.5100
    0.6500    0.4000    0.7200    0.3900    0.0500    0.1600    0.5500
    0.0070    0.0020    0.0120         0         0         0         0
         0         0         0         0         0         0         0
    0.0020         0         0         0         0         0         0
    0.0460    0.0840    0.0250    0.0240    0.0140    0.0240    0.0340
    0.7400    0.3500    0.1600    0.1400    0.0860    0.2700    0.7400
         0         0         0         0         0         0    0.0120
    0.0440    0.0520    0.0010         0         0         0         0
    0.1100    0.0190    0.0020         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0

  Columns 8 through 13
    2.7800    1.6200    1.9300    1.3800    0.9300    1.0660
    2.1500    5.9000    0.7800    1.2800    0.4700    0.2490
    0.2400    0.7300    1.1000    1.0100    0.1680    0.2930
    0.0730    0.4200    0.7300    0.3400    0.0460    0.1680
    0.0070    0.1300    0.0020         0         0         0
    0.0120    0.0060         0         0         0         0
    0.0020         0         0         0         0         0
    0.0270         0         0         0         0    0.0013
    0.1500    1.0700    0.2000    0.2800    0.0970    0.1590
    0.3300    0.3100    0.1700    0.1900    0.0540    0.1240
    0.4800    0.1700    0.2300    0.1500    0.2100    0.1880
         0         0         0         0         0         0
         0         0         0         0         0         0
         0         0         0         0         0         0
    0.2300    0.3000    0.1200    0.2400    0.0117    0.0052
    0.5100    0.9100    0.6600    0.3900    0.1600    0.1800
    0.2000    0.0360    0.0080         0    0.0019    0.0048
    0.1100    0.0650    0.1100    0.1600         0         0
    0.0040    0.0550    0.0040    0.0220         0         0
         0         0         0         0         0         0
         0         0         0         0         0         0
```

```
calcium' =
   1.0e+03 *
  Columns 1 through 7
    2.3372    0.5050    0.2600    0.3312    0.3488    0.5055    0.2486
    1.0735    0.5030    0.3172    0.3542    0.3382    0.2074    0.1916
    0.6270    0.5000    0.1305    0.3494    0.3324    0.1976    0.1319
    0.7668    0.5600    0.3020    0.3418    0.3449    0.1234    0.2654
    1.4176    0.5310    0.1769    0.3485    0.3382    0.1181    0.0976
    0.5943    0.4460    0.1858    0.3437    0.1290    0.1199    0.2093
    0.0688    0.5100    0.1113    0.1206    0.1302    0.0627    0.0624
    0.1759    0.4610    0.1098    0.0181    0.0136    0.0071    0.0200
    1.4904    0.5180    0.3347    0.3398    0.3370    0.1278    0.1744
    1.2835    0.5050    0.3162    0.7490    0.3468    0.1146    0.1639
    0.5287    0.4910    0.3107    0.1154    0.3414    0.2049    0.1209
    0.5492    0.5180    0.1395    0.1771    0.1499    0.0814    0.0834
    0.7275    0.4950    0.1121    0.3418    0.1270    0.0735    0.0787
    0.6489    0.5570    0.1538    0.1296    0.1760    0.0507    0.1925
    0.8633    0.4920    0.1293    0.1765    0.3488    0.2000    0.1424
    0.9416    0.5160    0.3221    0.5000    0.3353    0.3464    0.1781
    0.7492    0.5160    0.1269    0.3408    0.3295    0.1490    0.1277
    0.6230    0.5410    0.1359    0.3398    0.1235    0.1225    0.1438
    0.8738    0.5570    0.1063    0.1312    0.1239    0.1014    0.0711
    0.0566    0.4730    0.0886    0.1275    0.0855    0.0539    0.0561
    0.8663    0.5280    0.1624    0.0822    0.0712    0.0786    0.0655
  Columns 8 through 11
    0.1994    0.1733    0.2148    0.1494
    0.1500    0.1775    0.2000    0.1473
    0.0996    0.0974    0.1261    0.1469
    0.1492    0.1414    0.1565    0.1453
    0.0985    0.1288    0.0807    0.1222
    0.0710    0.0744    0.0997    0.0940
    0.0495    0.0770    0.0481    0.0995
    0.0347    0.0266    0.0958    0.0949
    0.0949    0.1238    0.2556    0.1411
    0.1500    0.1713    0.3679    0.1441
    0.1514    0.0898    0.1598    0.1003
    0.1009    0.0688    0.1614    0.1441
    0.1000    0.0638    0.1926    0.1003
    0.0743    0.0717    0.1573    0.0993
    0.1445    0.0863    0.1056    0.1215
    0.0985    0.1030    0.2333    0.1441
    0.1012    0.0873    0.2222    0.1437
    0.0672    0.0807    0.1898    0.0948
    0.0931    0.0818    0.1492    0.0995
    0.0545    0.0485    0.1228    0.0471
    0.0931    0.0586    0.0958    0.1225
```

```
ethylbenzene' =
   1.0e+03 *
  Columns 1 through 7
    4.3900    3.1500    3.1700    3.0700    3.5070    3.6500    2.7800
    3.9500    2.8500    2.8300    3.2700    2.7060    2.9300    2.6300
    2.3200    1.9800    1.5600    0.1300    1.4050    1.2100    1.1800
    1.2700    0.9100    1.0000    0.5600    0.4280    0.4500    0.2700
    0.0070    0.0060    0.0020         0    0.0010         0         0
    0.0420         0         0         0         0         0         0
    0.0130         0         0         0         0    0.0050         0
    0.0060    0.0170    0.0240    0.0200    0.0180    0.0050    0.1200
    0.0210    0.1200    0.8900    0.7100    0.6630    0.5700    0.6000
    0.2600    0.2700    0.3600    0.2300    0.1400    1.0000    0.8400
    0.6000    0.3000    0.5300    0.3300    0.0570    0.1500    0.4600
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
    0.0040         0         0         0         0         0         0
    0.0240    0.0060         0    0.0020    0.0010         0    0.0030
    0.2800    0.2100    0.1300    0.2100    0.0740    0.0560    0.6800
         0         0         0         0         0         0         0
    0.0010    0.0020         0         0         0         0         0
    0.0770    0.0290         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
  Columns 8 through 13
    3.3400    2.8200    3.3300    3.7000    4.9000    3.5500
    2.2800    3.0700    2.5100    3.4700    5.1300    3.2700
    0.4800    1.2000    2.2000    2.2100    1.1300    2.6400
    0.2000    0.5900    1.3000    1.0900    0.4390    0.8010
    0.0050    0.1200    0.0030         0         0         0
    0.0120    0.0070         0         0         0         0
         0         0         0         0         0         0
    0.1200    0.0110    0.0160    0.0200    0.0210    0.0220
    0.2100    0.7100    0.4000    0.4300    1.6500    0.3950
    0.6600    0.5800    0.4100    0.3900    0.1870    0.3270
    0.5100    0.1700    0.5100    0.4800    0.3300    0.5740
         0         0         0         0         0         0
         0         0         0         0         0    0.0011
         0         0         0         0         0    0.0027
    0.0110    0.0270    0.0180    0.0320    0.0010    0.0017
    0.0520    0.2900    0.4400    0.2600    0.1900    0.2890
    0.0130    0.0020         0         0         0         0
    0.0060    0.0050    0.0440    0.0540         0         0
         0    0.0140    0.0010         0         0         0
         0         0         0         0         0         0
         0         0         0         0         0         0
```

```
iron' =
  Columns 1 through 7
   12.0000   16.5000   14.0000   13.0000   13.0000   18.0000   17.5000
   24.0000   21.5000   20.0000   18.5000   17.5000   23.0000   18.5000
   17.5000   11.5000    9.2500   11.0000   10.0000   11.0000   15.0000
   16.5000   14.0000    8.2500    8.0000    7.0000    6.0000    9.0000
    9.0000    7.0000    4.0000    5.0000    2.5000    2.0000    2.0000
    8.0000    4.0000    5.0000    8.0000    2.0000    7.0000    4.7500
    6.0000   13.0000    4.0000    1.5000    0.7500    0.7500    2.0000
   55.0000   15.0000   16.0000   30.0000   12.0000   23.0000   17.5000
   13.0000   12.0000   10.0000    7.5000    7.0000    7.0000    2.5000
   15.0000   15.0000    8.2500   23.2500    8.0000    9.5000    9.5000
   16.5000   17.0000   10.0000   10.0000    6.0000    3.5000   11.0000
   14.0000   13.0000    7.0000    3.2500    3.5000    3.7500    2.5000
   11.0000   10.0000    4.0000    4.5000    2.0000    2.2500    2.0000
    9.0000   10.0000    3.7500    2.5000    1.0000    4.5500    5.5000
   22.5000   13.0000   10.0000   10.0000   12.0000   14.0000   11.5000
   15.0000   10.5000   10.0000    9.2500    8.0000    7.0000    7.0000
   16.0000    9.0000    6.0000    1.2500    4.0000    9.5000    6.0000
   10.1000   10.0000    8.0000    3.5000    3.0000    2.4500    4.5000
    9.0000   12.5000    0.1500    2.0000    3.5000    1.6000    4.5000
    9.0000  110.0000    0.1000    3.5000    0.4000    0.6000    4.5000
   10.5000   10.0000    4.0000    2.0000    1.5000    2.2500    1.2500
  Columns 8 through 11
   12.2500   16.5000   15.0000   14.0000
   17.5000   23.0000   17.0000    4.7500
    8.2500    9.5000   11.2500   12.0000
    6.2500   11.2500   13.2500    9.5000
    1.7500    8.5000    1.7500    4.2500
    2.0000    6.5000    3.0000    3.0000
    1.5000    4.7500    1.2500    2.7500
   12.5000   14.5000    5.5000    7.2500
    7.2500   17.5000    9.0000    7.5000
   12.5000   20.0000    9.0000   10.2500
   10.0000   14.0000   11.2500   11.5000
    3.5000    6.2500    3.2500    6.2500
    1.7500    7.5000    2.2500    4.7500
    1.0000    4.2500    2.0000    3.5000
    8.0000   10.5000    8.0000    9.5000
    9.2500    9.5000    7.4000    7.0000
    7.0000   10.5000    9.0000    9.7500
    2.7500    5.0000    3.2500    8.5000
    3.0000    9.0000    1.5000    5.5000
    1.0000    0.7500    3.5000    4.2500
    1.5000    2.0000    2.2500    3.7500
```

metaxylene' =

1.0e+03 *

Columns 1 through 7

| | | | | | | |
|---|---|---|---|---|---|---|
| 8.8500 | 6.4300 | 6.5500 | 6.2600 | 7.2760 | 8.6600 | 6.3000 |
| 6.9500 | 5.8000 | 5.5500 | 6.5100 | 5.4140 | 5.9900 | 5.4200 |
| 4.5300 | 3.9100 | 3.0900 | 0.2500 | 2.4400 | 2.3900 | 2.2800 |
| 1.3700 | 1.3800 | 1.7000 | 0.9200 | 0.5430 | 0.5800 | 0.2400 |
| 0.0030 | 0.0090 | 0.0030 | 0.0010 | 0.0010 | 0 | 0 |
| 0.0040 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0.0020 | 0.0110 | 0.0020 |
| 0.0060 | 0.0070 | 0.0160 | 0.0170 | 0.0150 | 0.0380 | 0.0780 |
| 0.2800 | 0.0400 | 1.4400 | 1.1100 | 0.8660 | 0.8300 | 0.9500 |
| 0.0420 | 0.0860 | 0.3000 | 0.1500 | 0.0180 | 1.3800 | 1.2700 |
| 0.5100 | 0.4000 | 0.7400 | 0.4700 | 0.0430 | 0.1900 | 0.6600 |
| 0.0010 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0.0060 | 0.0020 | 0 | 0.0010 | 0.0030 | 0.0020 | 0.0100 |
| 0.7000 | 0.4500 | 0.1500 | 0.2200 | 0.0870 | 0.1300 | 1.2500 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0.0030 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Columns 8 through 13

| | | | | | |
|---|---|---|---|---|---|
| 8.0800 | 6.9300 | 8.5000 | 9.1000 | 6.7000 | 9.5600 |
| 4.9300 | 6.3800 | 5.8600 | 7.2000 | 6.6000 | 8.1000 |
| 0.8300 | 2.0700 | 4.3700 | 4.3500 | 3.6000 | 4.9200 |
| 0.1600 | 0.7700 | 1.6700 | 2.0900 | 0.7300 | 1.4100 |
| 0.0050 | 0.1800 | 0.0040 | 0.0010 | 0 | 0 |
| 0.0160 | 0.0100 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0.0010 | 0 | 0 | 0 |
| 0.0780 | 0.0070 | 0.0120 | 0.0140 | 0.0100 | 0.0099 |
| 0.2600 | 0.9300 | 0.6000 | 0.3500 | 2.4000 | 0.6630 |
| 0.9000 | 0.7100 | 0.4100 | 0.2600 | 0.1600 | 0.4950 |
| 0.9100 | 0.3100 | 0.7100 | 0.2800 | 0.4700 | 0.9430 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0.0031 |
| 0 | 0 | 0 | 0 | 0 | 0.0073 |
| 0.0390 | 0.0530 | 0.0190 | 0.0550 | 0.0051 | 0.0033 |
| 0.5100 | 0.7700 | 0.8600 | 0.5000 | 0.3600 | 0.6480 |
| 0.0500 | 0.0100 | 0.0020 | 0 | 0 | 0 |
| 0.0010 | 0 | 0.0680 | 0.0490 | 0 | 0 |
| 0 | 0.0090 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |

```
nitrate' =
  Columns 1 through 7
    0.0100    0.0100    2.5000    0.8000    0.0100    0.3000    2.4000
    0.0100    0.0100    1.6000    1.5000    0.4000    2.0000    2.4000
    0.0100    0.0100    1.2000    1.5000    0.0100    1.8000    1.2000
    0.0100    0.0100    2.1000    0.8000    0.0100    1.5000    0.8000
    0.0300    0.0100    0.4000    0.0100    0.0100    0.0100    0.5000
    0.0100    0.0100    1.1000    1.9000    0.3000    1.5000    1.7000
    0.0100    0.0100    0.6000    0.0100    0.6000    1.5000    0.5000
    0.0500    0.0100    2.1000    0.3000    0.0100    0.5000    1.0000
    0.0200    0.0100    2.1000    0.4000    0.0100    1.4000    1.9000
    0.1000    0.0300    0.6000    0.8000    0.0100    1.8000    1.0000
    0.0100    0.0100    0.8000    0.0100    0.0100    1.8000    0.7000
    0.0100    0.0100    0.3000    0.0100    0.0100    0.6000    1.0000
    0.0100    0.0100    0.6000    0.5000    0.0100    1.1000    0.3000
    0.0700    0.0200    0.8000    0.6000    0.4000    0.9000    0.3000
    0.0100    0.0100    1.1000    0.0100    0.0100    0.0100    1.5000
    0.0200    0.0100    1.1000    1.0000    0.0100    1.8000    1.7000
    0.0300    0.0100    0.8000    0.8000    0.0100    1.7000    0.7000
    0.0300    0.0100    0.3000    0.4000    0.0100    0.0100    0.3000
    0.0100    0.0100    0.3000    0.3000    0.3000    0.3000    0.3000
    0.0100    0.0100    0.4000    0.0100    0.0100    0.9000    0.3000
    0.0300    0.0100    0.8000    1.3000    0.4000    1.2000    0.3000
  Columns 8 through 11
    3.3000    0.0100    3.6000    5.6000
    3.4000    0.0100    1.7000    2.6000
    1.5000    1.2000    0.4000    0.7000
    1.4000    0.0300    0.0100    0.7000
    0.3000    0.0100    0.0100    0.0100
    1.2000    0.0100    0.0100    0.0100
    4.0000    0.0100    0.3000    0.0100
    2.9000    0.0100    2.4000    1.1000
    2.2000    0.3000    0.0100    0.5000
    1.9000    0.0100    0.0100    1.1000
    1.2000    0.3000    0.0100    0.9000
    1.5000    1.0000    0.0100    0.7000
    1.2000    3.1000    0.0100    0.5000
    0.9000    0.0100    0.0100    0.7000
    1.1000    0.0100    0.0100    0.0100
    0.9000    0.0100    0.0100    0.3000
    0.8000    0.0100    0.0100    0.0100
    0.6000    1.5000    0.0100    0.3000
    1.9000    3.1000    0.0100    0.5000
    0.6000    0.9000    0.9000    0.0100
    0.6000    2.0000    0.0100    0.0100
```

orthoxylene' =

1.0e+03 *

Columns 1 through 7

| | | | | | | |
|---|---|---|---|---|---|---|
| 6.1000 | 4.4700 | 4.5200 | 4.4700 | 5.0110 | 5.9100 | 4.5900 |
| 4.8400 | 4.0500 | 3.8100 | 4.6400 | 3.6990 | 4.2300 | 3.9300 |
| 3.1300 | 2.6500 | 2.0400 | 0.1700 | 1.7290 | 1.5000 | 1.6300 |
| 0.9900 | 0.9900 | 1.1600 | 0.6000 | 0.3360 | 0.3300 | 0.1700 |
| 0.0030 | 0.0060 | 0.0020 | 0 | 0 | 0 | 0 |
| 0.0020 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0.0010 | 0.0060 | 0.0010 |
| 0.0230 | 0.0220 | 0.0440 | 0.0430 | 0.0200 | 0.1700 | 0.0920 |
| 0.1600 | 0.0230 | 0.9200 | 0.5200 | 0.3160 | 0.4800 | 0.6400 |
| 0.0510 | 0.0620 | 0.1900 | 0.0640 | 0.0060 | 0.7900 | 0.8700 |
| 0.3400 | 0.2900 | 0.6000 | 0.3700 | 0.0230 | 0.1200 | 0.4900 |
| 0.0010 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0.0080 | 0 | 0.0010 | 0.0010 | 0.0030 | 0.0010 | 0.0060 |
| 0.4400 | 0.2600 | 0.0900 | 0.1300 | 0.0530 | 0.0520 | 0.8900 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0.0020 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Columns 8 through 13

| | | | | | |
|---|---|---|---|---|---|
| 5.7000 | 4.8500 | 5.9400 | 6.8600 | 6.6000 | 6.7500 |
| 3.4700 | 4.4500 | 4.0200 | 5.4100 | 6.4000 | 5.5000 |
| 0.6200 | 1.4600 | 3.1700 | 0.3500 | 2.5000 | 3.4900 |
| 0.1100 | 0.5900 | 1.3600 | 1.6100 | 0.4400 | 1.0900 |
| 0.0050 | 0.1400 | 0.0030 | 0.0010 | 0 | 0 |
| 0.0120 | 0.0080 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0.0920 | 0.0110 | 0.0280 | 0.0330 | 0.0330 | 0.0291 |
| 0.1900 | 0.7300 | 0.3500 | 0.2900 | 1.4000 | 0.4090 |
| 0.5500 | 0.4500 | 0.2500 | 0.1800 | 0.1100 | 0.2850 |
| 0.6900 | 0.2200 | 0.4500 | 0.1800 | 0.3000 | 0.6950 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0.0015 |
| 0 | 0 | 0 | 0 | 0 | 0.0044 |
| 0.0280 | 0.0420 | 0.0140 | 0.0560 | 0.0044 | 0.0024 |
| 0.4100 | 0.6200 | 0.6700 | 0.4100 | 0.2500 | 0.4470 |
| 0.0380 | 0.0090 | 0.0020 | 0 | 0 | 0 |
| 0 | 0 | 0.0390 | 0.0140 | 0 | 0 |
| 0 | 0.0120 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |

```
oxygen' =
  Columns 1 through 7
    0.5000    0.0100    0.0100    0.0100    0.1000    0.0100    0.0100
    0.0100    0.0100    1.0000    0.0100    0.0100    0.0100    0.4500
    0.0100    0.0100    0.5400    0.0100    0.6000    0.0500    0.8000
    1.9000    0.0100    0.0100    0.0100    0.0100    0.4000    0.2000
    0.1500    0.0100    1.6000    1.4000    0.7000    1.0000    0.0100
    1.5000    0.0100    1.6500    0.0100    0.6500    2.8000    0.0100
    1.1000    0.0100    0.0100    1.2500    0.6500    1.4000    0.0100
    1.1000    0.0100    0.0100    0.0100    0.0100    0.0100    0.0100
    0.9500    0.0100    0.0100    0.0100    0.0100    0.0100    0.0200
    0.1000    0.0100    0.0100    0.0100    0.0100    0.0100    0.0100
    0.1500    0.0100    0.5100    0.0100    0.0100    0.0100    1.0000
    0.7500    0.0100    1.3000    1.1500    0.0100    3.9000    1.7500
    1.4000    0.0100    1.5000    0.1000    0.3500    1.4500    0.1500
    1.5000    0.0100    1.3500    0.1000    0.3500    2.5000    0.0100
   10.5000    0.0100    0.5900    0.0100    0.0100    0.4000    1.0500
    0.8000    0.0100    1.1000    0.0100    0.1500    0.0100    1.3000
    1.2000    0.0100    1.8000    1.8000    1.9500    1.2000    1.1500
   10.5000    0.0100    1.3000    1.0000    2.0000    3.1500    1.2000
    0.9000    0.0100    0.0100    2.4000    1.5000    3.8000    0.0100
    0.9000    0.0100    0.5000    0.0100    1.0500    3.1500    0.5000
    1.1000    0.0100    0.0100    0.8500    0.0100    6.0000    0.2500
  Columns 8 through 11
    0.5000    0.0100    0.1000    0.5000
    0.0100    0.1500    0.1000    0.1500
    1.7500    0.3000    0.1000    0.3000
    0.3500    0.0100    0.0100    0.0100
    0.0100    0.0100    1.1000    1.1000
    0.4000    0.0100    0.7000    1.3000
    0.0100    0.5000    0.4000    0.3000
    0.0100    0.0100    0.0100    0.1500
    0.0100    0.0100    0.0100    0.0100
    1.2000    0.0100    0.2000    0.1000
    1.6500    1.2000    0.1000    0.2500
    0.0100    0.4000    0.0100    0.0100
    0.0100    0.0100    0.6000    0.0100
    1.0000    0.2000    2.2000    1.8500
    0.4500    0.5000    0.7000    0.8000
    1.5000    0.6000    0.4000    0.5000
    0.2000    0.0100    0.2000    1.1000
    0.2500    0.5000    1.4000    0.0100
    0.7500    0.0100    1.0000    0.0100
    0.0100    0.3000    0.3000    1.0500
    0.0100    1.1000    0.7000    1.6000
```

```
paraxylene' =
   1.0e+03 *
  Columns 1 through 7
    3.6600    2.6500    2.7000    2.5900    3.0820    3.6200    2.5900
    2.8700    2.4000    2.3300    2.7100    2.2550    2.4300    2.2400
    1.9100    1.6600    1.3500    0.1100    1.2000    1.0200    1.0000
    0.6000    0.6800    0.8000    0.4400    0.3080    0.2900    0.1400
    0.0010    0.0040    0.0010         0         0         0         0
    0.0020         0         0         0         0         0         0
         0         0         0         0         0    0.0060    0.0010
    0.0090    0.0100    0.0180    0.0160    0.0140    0.0280    0.0600
    0.2100    0.0580    0.6900    0.5700    0.5530    0.4500    0.4700
    0.0750    0.1000    0.2200    0.1400    0.0520    0.7800    0.6700
    0.2700    0.2000    0.3400    0.2200    0.0300    0.1000    0.4600
    0.0010         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
    0.0050    0.0010    0.0001    0.0020    0.0050    0.0100    0.0210
    0.3900    0.3300    0.1400    0.2300    0.0950    0.1300    0.6500
         0         0         0         0         0         0    0.0040
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
  Columns 8 through 13
    3.3100    2.7500    3.4500    3.8500    6.5000    3.9800
    1.9500    2.3000    2.3500    3.0400    6.2000    3.4000
    0.3800    0.9000    1.9400    1.8900    1.7000    2.3900
    0.1000    0.4100    1.1100    0.9400    0.3500    0.8540
    0.0030    0.0810    0.0020         0         0         0
    0.0070    0.0050         0         0         0         0
         0         0         0         0         0         0
    0.0600    0.0070    0.0120    0.0130    0.0130    0.0165
    0.1400    0.5500    0.3300    0.2400    1.8000    0.3440
    0.5300    0.4800    0.3100    0.1900    0.1500    0.2950
    0.4200    0.1500    0.4300    0.2200    0.2800    0.5940
         0         0         0         0         0         0
         0         0         0         0         0    0.0014
         0         0         0         0         0    0.0031
    0.0430    0.0440    0.0220    0.0520    0.0198    0.0042
    0.2600    0.4600    0.4700    0.2900    0.2600    0.3600
    0.0420    0.0110    0.0020         0         0    0.0016
    0.0020    0.0030    0.0510    0.0360         0         0
         0    0.0080         0         0         0         0
         0         0         0         0         0         0
         0         0         0         0         0         0
```

ph' =

Columns 1 through 7

| | | | | | | |
|---|---|---|---|---|---|---|
| 6.4000 | 6.4000 | 6.3000 | 6.6000 | 6.4000 | 6.6000 | 6.6000 |
| 6.4000 | 6.5000 | 6.7000 | 6.7000 | 6.5000 | 6.7000 | 6.7000 |
| 6.4000 | 6.5000 | 6.3000 | 6.9000 | 6.7000 | 6.8000 | 6.7000 |
| 6.6000 | 6.7000 | 6.8000 | 7.0000 | 7.0000 | 7.0000 | 6.9000 |
| 7.1000 | 7.1000 | 6.9000 | 7.8000 | 7.8000 | 7.7000 | 7.6000 |
| 7.1000 | 7.3000 | 7.0000 | 7.6000 | 7.7000 | 7.5000 | 7.5000 |
| 7.0000 | 7.2000 | 7.0000 | 7.6000 | 7.9000 | 7.7000 | 7.7000 |
| 6.1000 | 6.1000 | 6.7000 | 6.5000 | 6.2000 | 7.6000 | 6.5000 |
| 6.5000 | 6.6000 | 6.8000 | 6.8000 | 6.6000 | 6.9000 | 6.7000 |
| 6.3000 | 6.6000 | 6.7000 | 6.9000 | 6.5000 | 6.8000 | 6.6000 |
| 6.5000 | 6.7000 | 6.7000 | 6.9000 | 7.0000 | 7.1000 | 7.0000 |
| 6.8000 | 6.9000 | 6.9000 | 7.4000 | 7.3000 | 7.7000 | 7.4000 |
| 6.8000 | 7.0000 | 7.0000 | 7.5000 | 7.6000 | 7.3000 | 7.3000 |
| 6.9000 | 7.1000 | 7.1000 | 7.5000 | 7.7000 | 7.4000 | 7.3000 |
| 6.7000 | 6.6000 | 6.8000 | 7.0000 | 6.6000 | 7.0000 | 6.9000 |
| 6.6000 | 6.7000 | 6.8000 | 7.0000 | 6.7000 | 6.8000 | 6.9000 |
| 6.8000 | 6.6000 | 6.9000 | 7.2000 | 7.2000 | 7.1000 | 7.0000 |
| 6.9000 | 6.7000 | 6.8000 | 7.3000 | 7.5000 | 7.5000 | 7.3000 |
| 6.8000 | 6.8000 | 6.3000 | 7.0000 | 7.6000 | 7.6000 | 7.2000 |
| 7.5000 | 7.4000 | 6.9000 | 7.5000 | 7.7000 | 7.6000 | 7.4000 |
| 7.1000 | 6.8000 | 6.8000 | 7.1000 | 7.9000 | 7.5000 | 7.4000 |

Columns 8 through 11

| | | | |
|---|---|---|---|
| 6.5000 | 6.4000 | 6.7000 | 6.6000 |
| 6.4000 | 6.3000 | 6.6000 | 6.7000 |
| 6.6000 | 6.4000 | 6.7000 | 6.6000 |
| 6.7000 | 6.5000 | 6.7000 | 6.7000 |
| 6.9000 | 7.0000 | 7.3000 | 7.2000 |
| 6.9000 | 7.2000 | 7.3000 | 7.0000 |
| 7.1000 | 7.0000 | 7.3000 | 7.1000 |
| 6.5000 | 6.3000 | 6.1000 | 7.4000 |
| 6.7000 | 6.6000 | 6.7000 | 7.0000 |
| 6.4000 | 6.4000 | 6.5000 | 6.5000 |
| 6.7000 | 6.7000 | 6.7000 | 6.9000 |
| 7.1000 | 7.1000 | 7.1000 | 7.2000 |
| 6.9000 | 6.9000 | 7.2000 | 6.9000 |
| 7.1000 | 7.0000 | 7.2000 | 7.3000 |
| 6.8000 | 6.7000 | 6.8000 | 7.0000 |
| 6.9000 | 6.7000 | 6.8000 | 7.0000 |
| 6.6000 | 6.8000 | 6.8000 | 7.0000 |
| 7.0000 | 6.9000 | 6.9000 | 7.0000 |
| 7.1000 | 7.0000 | 7.0000 | 7.3000 |
| 7.1000 | 7.1000 | 7.2000 | 7.3000 |
| 7.2000 | 7.1000 | 7.2000 | 7.3000 |

phosphates' =

Columns 1 through 7

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.0030 | 2.0800 | 0.0030 | 0.1230 | 0.0400 | 0.0870 | 0.0060 |
| 1.0500 | 1.7100 | 0.0030 | 0.2430 | 0.0020 | 0.0030 | 0.0100 |
| 0.7400 | 1.7900 | 0.0030 | 0.2830 | 0.0030 | 0.0030 | 0.0100 |
| 1.8500 | 1.2500 | 0.0100 | 0.1850 | 0.0030 | 0.0350 | 0.0140 |
| 1.4400 | 1.1500 | 0.0080 | 0.5170 | 0.0360 | 0.0190 | 0.0170 |
| 1.3800 | 0.6000 | 0.0120 | 0.3240 | 0.0250 | 0.0670 | 0.0120 |
| 0.8300 | 2.9000 | 0.0090 | 0.3030 | 0.0290 | 0.0670 | 0.0150 |
| 0.0030 | 0.0030 | 0.0030 | 1.3630 | 0.1050 | 0.4220 | 0.1600 |
| 1.6100 | 4.7100 | 0.0340 | 3.5000 | 0.2330 | 1.0470 | 0.0150 |
| 2.4600 | 2.1300 | 0.0030 | 0.4070 | 0.0810 | 0.0030 | 0.0170 |
| 1.6000 | 0.8700 | 0.0030 | 0.3480 | 0.0030 | 0.0970 | 0.0110 |
| 2.5000 | 1.1800 | 0.0090 | 0.6120 | 0.0250 | 0.0760 | 0.0240 |
| 1.0100 | 2.5100 | 0.0090 | 0.9120 | 0.0710 | 0.0020 | 0.0130 |
| 2.6800 | 1.9200 | 0.0120 | 0.3300 | 0.0420 | 0.0470 | 0.0130 |
| 0.0030 | 0.0030 | 0.0060 | 0.2590 | 0.0140 | 0.1010 | 0.0130 |
| 2.8400 | 1.4100 | 0.0050 | 0.3890 | 0.1900 | 0.1100 | 0.0160 |
| 2.5600 | 0.6900 | 0.0180 | 0.0830 | 0.0130 | 0.0560 | 0.0180 |
| 2.6000 | 1.7400 | 0.0090 | 0.0220 | 0.0670 | 0.0400 | 0.0130 |
| 1.9000 | 2.3600 | 0.0100 | 0.2170 | 0.0320 | 0.0600 | 0.0130 |
| 1.2800 | 0.5900 | 0.0090 | 0.1240 | 0.0170 | 0.0440 | 0.0080 |
| 3.4300 | 1.3900 | 0.0120 | 0.2470 | 0.0420 | 0.0830 | 0.0130 |

Columns 8 through 11

| | | | |
|---|---|---|---|
| 0.0030 | 0.0340 | 0.0030 | 0.1120 |
| 0.0850 | 0.0290 | 0.0030 | 0.0680 |
| 0.0280 | 0.0160 | 0.0030 | 0.0060 |
| 0.2060 | 0.1300 | 0.0400 | 0.0140 |
| 0.0840 | 0.0870 | 0.0290 | 0.0840 |
| 0.0530 | 0.1300 | 0.0290 | 0.0740 |
| 0.0590 | 0.1080 | 0.0400 | 0.0600 |
| 0.1530 | 0.0750 | 0.1130 | 0.1790 |
| 0.5400 | 0.6890 | 1.2800 | 0.6920 |
| 0.0320 | 0.0270 | 0.0670 | 0.0630 |
| 0.2670 | 0.1220 | 0.0820 | 0.0650 |
| 0.1920 | 0.0040 | 0.0010 | 0.0950 |
| 0.0110 | 0.0230 | 0.0380 | 0.0180 |
| 0.0340 | 0.0700 | 0.0460 | 0.0820 |
| 0.0420 | 0.0620 | 0.1230 | 0.0430 |
| 0.1040 | 0.0300 | 0.1040 | 0.0880 |
| 0.0540 | 0.0020 | 0.0030 | 0.0450 |
| 0.1230 | 0.0340 | 0.0180 | 0.0420 |
| 0.0500 | 0.0160 | 0.0540 | 0.0260 |
| 0.0310 | 0.0110 | 0.0010 | 0.0780 |
| 0.0370 | 0.0400 | 0.0680 | 0.0550 |

sulfate' =

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 7 | 18 | 10 | 5 | 3 | 9 | 34 | 33 | 29 | 35 |
| 11 | 6 | 18 | 8 | 4 | 3 | 4 | 17 | 25 | 10 | 25 |
| 3 | 4 | 8 | 8 | 3 | 2 | 2 | 7 | 12 | 25 | 7 |
| 2 | 31 | 6 | 10 | 3 | 2 | 2 | 6 | 6 | 10 | 1 |
| 26 | 31 | 50 | 35 | 15 | 18 | 20 | 20 | 16 | 27 | 29 |
| 23 | 24 | 5 | 10 | 15 | 19 | 16 | 15 | 18 | 23 | 28 |
| 26 | 20 | 36 | 32 | 12 | 1 | 11 | 16 | 24 | 24 | 28 |
| 61 | 34 | 33 | 31 | 10 | 7 | 13 | 48 | 27 | 26 | 24 |
| 70 | 36 | 28 | 30 | 4 | 19 | 3 | 34 | 10 | 16 | 31 |
| 8 | 18 | 34 | 30 | 31 | 2 | 4 | 11 | 7 | 11 | 2 |
| 2 | 2 | 6 | 4 | 3 | 2 | 2 | 14 | 6 | 8 | 1 |
| 14 | 24 | 45 | 32 | 34 | 21 | 23 | 23 | 21 | 26 | 21 |
| 29 | 30 | 64 | 43 | 31 | 19 | 25 | 27 | 28 | 31 | 32 |
| 27 | 26 | 30 | 31 | 12 | 7 | 19 | 15 | 34 | 25 | 22 |
| 31 | 30 | 31 | 31 | 7 | 6 | 4 | 9 | 7 | 7 | 3 |
| 2 | 4 | 8 | 7 | 4 | 2 | 3 | 7 | 28 | 6 | 1 |
| 26 | 21 | 40 | 36 | 14 | 2 | 2 | 7 | 7 | 5 | 21 |
| 27 | 35 | 44 | 40 | 14 | 18 | 18 | 18 | 6 | 23 | 16 |
| 28 | 31 | 38 | 37 | 23 | 13 | 19 | 24 | 33 | 30 | 28 |
| 20 | 17 | 34 | 30 | 12 | 5 | 13 | 9 | 14 | 23 | 23 |
| 13 | 13 | 20 | 20 | 12 | 4 | 8 | 9 | 14 | 100 | 11 |

temperature' =

Columns 1 through 7

| | | | | | | |
|---|---|---|---|---|---|---|
| 6.9000 | 6.3000 | 9.3000 | 10.2000 | 12.5000 | 16.0000 | 17.3000 |
| 7.6000 | 7.1000 | 8.4000 | 9.9000 | 12.1000 | 14.7000 | 15.7000 |
| 8.3000 | 7.4000 | 8.9000 | 9.3000 | 11.8000 | 12.7000 | 13.7000 |
| 8.5000 | 8.2000 | 9.5000 | 9.7000 | 11.6000 | 11.3000 | 13.6000 |
| 9.6000 | 9.4000 | 8.1000 | 9.3000 | 10.0000 | 11.9000 | 11.7000 |
| 9.6000 | 9.6000 | 10.0000 | 9.9000 | 10.8000 | 10.9000 | 11.4000 |
| 9.9000 | 10.0000 | 10.5000 | 10.1000 | 10.5000 | 11.1000 | 10.7000 |
| 3.1000 | 3.0000 | 7.2000 | 13.0000 | 16.1000 | 0 | 19.0000 |
| 4.4000 | 4.5000 | 7.2000 | 12.6000 | 13.7000 | 6.3000 | 15.6000 |
| 6.1000 | 5.8000 | 7.4000 | 9.7000 | 12.2000 | 14.4000 | 14.5000 |
| 6.4000 | 6.0000 | 8.6000 | 9.2000 | 11.9000 | 13.3000 | 13.6000 |
| 7.8000 | 6.5000 | 9.3000 | 9.3000 | 11.1000 | 11.8000 | 12.4000 |
| 8.7000 | 7.7000 | 9.5000 | 9.5000 | 0.4000 | 11.2000 | 11.9000 |
| 9.2000 | 8.8000 | 9.6000 | 9.7000 | 10.2000 | 10.5000 | 11.0000 |
| 3.4000 | 4.0000 | 8.0000 | 12.6000 | 16.4000 | 18.5000 | 17.8000 |
| 4.6000 | 4.7000 | 8.0000 | 11.0000 | 14.2000 | 16.4000 | 16.1000 |
| 6.9000 | 5.7000 | 8.1000 | 9.3000 | 12.3000 | 14.3000 | 14.9000 |
| 6.8000 | 6.3000 | 7.8000 | 9.2000 | 10.7000 | 12.6000 | 13.6000 |
| 7.6000 | 7.4000 | 8.8000 | 11.2000 | 10.6000 | 12.2000 | 12.8000 |
| 8.8000 | 8.2000 | 9.4000 | 9.5000 | 10.3000 | 11.1000 | 11.8000 |
| 9.2000 | 9.2000 | 9.3000 | 9.3000 | 10.8000 | 10.2000 | 10.6000 |

Columns 8 through 11

| | | | |
|---|---|---|---|
| 17.3000 | 14.3000 | 11.6000 | 9.8000 |
| 16.1000 | 14.3000 | 12.0000 | 10.4000 |
| 14.6000 | 14.1000 | 12.2000 | 10.1000 |
| 13.1000 | 13.5000 | 12.7000 | 11.5000 |
| 12.3000 | 12.4000 | 12.3000 | 11.3000 |
| 12.3000 | 12.1000 | 12.2000 | 11.5000 |
| 14.3000 | 11.5000 | 11.6000 | 10.6000 |
| 18.5000 | 0 | 0 | 5.5000 |
| 16.1000 | 12.5000 | 10.1000 | 8.2000 |
| 15.4000 | 12.9000 | 10.8000 | 8.7000 |
| 15.4000 | 12.9000 | 11.3000 | 9.3000 |
| 14.7000 | 12.8000 | 11.8000 | 10.3000 |
| 13.9000 | 12.8000 | 11.9000 | 10.2000 |
| 12.5000 | 11.6000 | 11.4000 | 10.9000 |
| 18.7000 | 12.4000 | 9.3000 | 6.6000 |
| 17.1000 | 12.9000 | 10.6000 | 8.5000 |
| 17.5000 | 13.6000 | 11.4000 | 8.7000 |
| 15.0000 | 13.6000 | 12.1000 | 10.1000 |
| 14.1000 | 13.2000 | 12.1000 | 10.5000 |
| 12.8000 | 12.2000 | 11.9000 | 8.5000 |
| 11.5000 | 11.4000 | 11.5000 | 10.9000 |

toc' =

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 280 | 62 | 5 | 130 | 170 | 29 | 150 | 43 | 107 | 190 | 149 |
| 160 | 192 | 120 | 96 | 100 | 180 | 160 | 150 | 180 | 87 | 192 |
| 68 | 87 | 48 | 46 | 54 | 32 | 60 | 49 | 47 | 110 | 40 |
| 29 | 128 | 26 | 27 | 26 | 26 | 21 | 25 | 40 | 52 | 39 |
| 53 | 110 | 7 | 8 | 5 | 7 | 5 | 8 | 12 | 7 | 9 |
| 8 | 58 | 13 | 9 | 7 | 9 | 5 | 7 | 10 | 7 | 7 |
| 8 | 54 | 8 | 13 | 6 | 6 | 7 | 1 | 1 | 1 | 7 |
| 89 | 170 | 170 | 160 | 220 | 145 | 130 | 59 | 120 | 160 | 122 |
| 92 | 85 | 110 | 100 | 110 | 58 | 56 | 52 | 6 | 97 | 90 |
| 120 | 34 | 28 | 35 | 30 | 40 | 46 | 43 | 44 | 55 | 42 |
| 42 | 29 | 33 | 32 | 15 | 21 | 23 | 32 | 24 | 33 | 36 |
| 17 | 38 | 13 | 8 | 11 | 9 | 6 | 9 | 13 | 9 | 12 |
| 18 | 12 | 11 | 11 | 7 | 9 | 10 | 9 | 12 | 7 | 8 |
| 26 | 19 | 7 | 12 | 1 | 1 | 1 | 6 | 7 | 6 | 1 |
| 78 | 76 | 70 | 66 | 110 | 52 | 53 | 37 | 37 | 50 | 40 |
| 34 | 27 | 32 | 46 | 40 | 49 | 24 | 17 | 28 | 28 | 36 |
| 34 | 14 | 11 | 15 | 11 | 15 | 21 | 25 | 23 | 11 | 16 |
| 22 | 5 | 8 | 7 | 7 | 7 | 9 | 12 | 15 | 14 | 18 |
| 21 | 35 | 6 | 5 | 1 | 1 | 1 | 1 | 11 | 7 | 12 |
| 8 | 25 | 8 | 6 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| 59 | 16 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 1 |

```
toluene' =
   1.0e+04 *
  Columns 1 through 7
    3.6840    3.3560    3.2040    3.1420    2.5270    2.8830    2.6860
    3.5300    3.1300    2.6290    3.0740    2.3230    2.5380    2.4500
    1.1550    0.8950    0.4280    0.0430    0.3200    0.4890    0.5650
    0.1360    0.0930    0.1450    0.0380    0.0132    0.0160    0.0050
    0.0019    0.0017    0.0003         0         0         0         0
    0.0180         0         0         0         0         0         0
    0.0029         0         0         0    0.0003    0.0016    0.0003
    0.0002    0.0004    0.0010    0.0008    0.0016    0.0065    0.0110
    0.0044    0.0120    0.4010    0.1210    0.0624    0.3950    0.0440
    0.0020    0.0150    0.0730    0.0050    0.0009    0.1680    0.2010
    0.2090    0.1760    0.3460    0.1550    0.0039    0.0310    0.1300
    0.0002    0.0002    0.0002         0         0         0         0
    0.0001         0         0         0         0         0         0
    0.0007         0         0         0         0         0         0
    0.0016         0         0         0    0.0001         0    0.0002
    0.0069    0.0032    0.0017    0.0054    0.0017    0.0077    0.0450
         0         0         0         0         0         0         0
    0.0001         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
  Columns 8 through 13
    2.8600    2.0460    2.9980    2.8780    1.7000    3.1400
    1.9490    2.4720    1.5200    3.0830    1.7000    1.5300
    0.1470    0.5430    1.2920    1.3840    0.3100    1.0400
    0.0120    0.0790    0.2500    0.2530    0.0110    0.0269
    0.0026    0.0480    0.0005    0.0003         0         0
    0.0005    0.0004         0         0         0         0
         0         0         0         0         0         0
    0.0110    0.0003    0.0005    0.0008    0.0006    0.0004
    0.0083    0.0098    0.0098    0.0640    0.0082    0.0097
    0.0150    0.0046    0.0020    0.0310    0.0004    0.0027
    0.1670    0.0039    0.0045    0.0130    0.0018    0.0046
         0         0         0         0         0    0.0001
         0         0         0         0         0    0.0004
         0         0         0         0         0    0.0017
    0.0006    0.0008    0.0004    0.0005    0.0001         0
    0.0026    0.0024    0.0130    0.0011    0.0130    0.0012
    0.0005    0.0002         0         0         0         0
    0.0020    0.0003    0.0003    0.0002         0         0
         0    0.0038         0         0         0         0
         0         0         0         0         0         0
         0         0         0         0         0         0
```

# APPENDIX B

## Residence Times of TEX Compounds

For benzene:

| residencetime_b_91 | residencetime_b_213 | t_day |
|---|---|---|
| 43.4545 | 60.8364 | -172.4 |
| 234.6545 | 252.0364 | 53.0 |
| 69.5273 | 243.3455 | 78.9 |
| 34.7636 | 43.4545 | 111.8 |
| 43.4545 | 52.1455 | 139.8 |
| 60.8364 | 78.2182 | 167.8 |
| 104.2909 | 112.9818 | 203.5 |
| 139.0545 | 147.7455 | 230.8 |
| 173.8182 | 182.5091 | 258.1 |
| 208.5818 | 217.2727 | 294.5 |
| 225.9636 | 234.6545 | 323.9 |
| 52.1455 | 243.3455 | 349.8 |
| 43.4545 | 60.8364 | 498.9 |
| 95.6000 | 104.2909 | 569.6 |
| 217.2727 | 225.9636 | 687.9 |

For toluene:

| residencetime_t_91 | residencetime_t_213 |
|---|---|
| 52.1455 | 69.5273 |
| 243.3455 | 252.0364 |
| 225.9636 | 252.0364 |
| 34.7636 | 52.1455 |
| 43.4545 | 52.1455 |
| 69.5273 | 78.2182 |
| 112.9818 | 121.6727 |
| 147.7455 | 156.4364 |
| 173.8182 | 182.5091 |
| 217.2727 | 225.9636 |
| 234.6545 | 243.3455 |
| 243.3455 | 252.0364 |
| 52.1455 | 69.5273 |
| 95.6000 | 112.9818 |
| 217.2727 | 234.6545 |

For ethylbenzene:

| residencetime_e_91 | residencetime_e_213 |
|---|---|
| 60.8364 | 78.2182 |
| 252.0364 | 269.4182 |
| 243.3455 | 269.4182 |
| 43.4545 | 69.5273 |
| 52.1455 | 69.5273 |
| 78.2182 | 86.9091 |
| 121.6727 | 130.3636 |
| 156.4364 | 165.1273 |
| 182.5091 | 199.8909 |
| 225.9636 | 243.3455 |
| 243.3455 | 252.0364 |
| 252.0364 | 260.7273 |
| 60.8364 | 78.2182 |
| 112.9818 | 130.3636 |
| 234.6545 | 252.0364 |

For m-xylene:

| residencetime_m_91 | residencetime_m_213 |
|---|---|
| 60.8364 | 86.9091 |
| 252.0364 | 269.4182 |
| 243.3455 | 269.4182 |
| 43.4545 | 69.5273 |
| 52.1455 | 69.5273 |
| 78.2182 | 86.9091 |
| 121.6727 | 130.3636 |
| 156.4364 | 165.1273 |
| 182.5091 | 199.8909 |
| 225.9636 | 243.3455 |
| 243.3455 | 252.0364 |
| 252.0364 | 260.7273 |
| 60.8364 | 78.2182 |
| 112.9818 | 130.3636 |
| 234.6545 | 252.0364 |

For o-xylene:

| residencetime_o_91 | residencetime_o_213 |
|---|---|
| 52.1455 | 69.5273 |
| 243.3455 | 260.7273 |
| 234.6545 | 260.7273 |
| 34.7636 | 52.1455 |
| 43.4545 | 60.8364 |
| 69.5273 | 78.2182 |
| 112.9818 | 121.6727 |
| 147.7455 | 156.4364 |
| 182.5091 | 191.2000 |
| 217.2727 | 225.9636 |
| 234.6545 | 243.3455 |
| 243.3455 | 252.0364 |
| 52.1455 | 69.5273 |
| 104.2909 | 121.6727 |
| 225.9636 | 243.3455 |

For p-xylene:

| residencetime_p_91 | residencetime_p_213 |
|---|---|
| 60.8364 | 78.2182 |
| 252.0364 | 269.4182 |
| 243.3455 | 269.4182 |
| 43.4545 | 69.5273 |
| 52.1455 | 69.5273 |
| 78.2182 | 86.9091 |
| 121.6727 | 130.3636 |
| 147.7455 | 165.1273 |
| 182.5091 | 199.8909 |
| 225.9636 | 234.6545 |
| 243.3455 | 252.0364 |
| 252.0364 | 260.7273 |
| 60.8364 | 78.2182 |
| 112.9818 | 130.3636 |
| 234.6545 | 252.0364 |

# APPENDIX C

## MATLAB Codes: Part I

```
% ************************************************************** %
% Written by: Ting Chow Tan (May 29 - Jun 11, 1994               %
% At: /apple/tan/thesis/sleepbear/XIII/tan6.m                    %
% Purpose: 1) Show hydraulic gradient vs time plot.              %
%          2) Determining residence times at both 9.1 and        %
%             21.3 meter sites.                                  %
%          3) Generate distance vs time plot.                    %
% ************************************************************** %
clear, clg
data
% ************************************************************** %
% PLOT THE AVAILABLE HYDRAULIC GRADIENT VS TIME DATA.            %
% ************************************************************** %
t_day = t_week *7 + (-62.6*7+53);
hydro_velocity = hydro_grad*.4/3;
plot(t_day,hydro_velocity,'o'), hold on, grid
xlabel('Julian date'), ylabel('Hydraulic velocity (m/day)')
% ************************************************************** %
% EYEBALL EXTRAPOLATION OF THE PREVIOUS PLOT.                    %
% ************************************************************** %
th = [
 -172.4000    0.2000

  -39.2      -0.0667
  -11.90     -0.0493

   53.0000    0.1253
   78.9000    0.3333
  111.8000    0.3733
  139.8000    0.1733
  167.8000         0

  203.5000   -0.1093
  230.8000   -0.0667
  258.1000   -0.0493

  294.5000         0
  323.9000    0.2000
  349.8000    0.2800
  498.9000    0.2400

  569.6000   -0.0493
  630.8000   -0.0667
  658.1000   -0.0093

  687.9000    0.1933
  688.0000    0.1943];
t_day_e = th(:,1);
hydro_velocity_e = th(:,2);
n = .4;        % Effective soil porosity
B = 2.7;       % Bulk soil density g/cm^3
foc = 0.0004; % Frac of naturally occurring organic carbon in soil.
```

```
% ************************************************************ %
% ON THE SAME PLOT, PLOT THE INTERPOLATION. THE INTERVAL       %
% ON t_day_fine IS THE RANGE THAT IS USED IN THE NEXT PLOT.    %
% ************************************************************ %
t_day_fine = linspace(-172.4,688,100);
hydro_velocity_fine = ...
  interp1(t_day_e,hydro_velocity_e,t_day_fine,'spline');
plot(t_day_fine,hydro_velocity_fine)
text(125*7 - 730 + (-62.6*7+53),-.15,'June 90')
text(125*7 - 365 + (-62.6*7+53),-.15,'June 91')
text(125*7       + (-62.6*7+53),-.15,'June 92')
% print hyvelocity
pause, clg
for k = 1:7
  if     k==1, BTEX_velocity_e = hydro_velocity_e/(1+foc*Koc(k)*B/n);
  elseif k==2, BTEX_velocity_e = hydro_velocity_e/(1+foc*Koc(k)*B/n);
  elseif k==3, BTEX_velocity_e = hydro_velocity_e/(1+foc*Koc(k)*B/n);
  elseif k==4, BTEX_velocity_e = hydro_velocity_e/(1+foc*Koc(k)*B/n);
  elseif k==5, BTEX_velocity_e = hydro_velocity_e/(1+foc*Koc(k)*B/n);
  elseif k==6, BTEX_velocity_e = hydro_velocity_e/(1+foc*Koc(k)*B/n);
  elseif k==7, BTEX_velocity_e = hydro_velocity_e;
  end
  BTEX_velocity_fine = ...
    interp1(t_day_e,BTEX_velocity_e,t_day_fine,'linear');
  % ************************************************************ %
  % NEXT PLOT: DISTANCE TRAVELLED VS TIME.                       %
  % ************************************************************ %
  t_day_fine_gap = t_day_fine(2) - t_day_fine(1);
  area = t_day_fine_gap*ones(size(BTEX_velocity_fine)).* ...
    BTEX_velocity_fine;
  cumsum_area = cumsum(area);
  plot(t_day_fine,cumsum_area), hold on
  if k==7
    xlabel('Julian date'), ylabel('Distance travelled (m)'), grid
    text(125*7 - 730 + (-62.6*7+53),-20,'June 90')
    text(125*7 - 365 + (-62.6*7+53),-20,'June 91')
    text(125*7       + (-62.6*7+53),-20,'June 92')
  end
  % ************************************************************ %
  % DATA INSUFFICIENT => DUPLICATE A SECTION OF CURVE AND ADD THAT %
  % AS AN EXTRAPOLATION TO THE ORIGINAL CURVE.                   %
  % ************************************************************ %
  j = find(t_day_fine<470 & t_day_fine>0);
  t_day_fine_t = [t_day_fine(j)-645 t_day_fine];
  cumsum_area_t=[cumsum_area(j)'-max(cumsum_area(j)) cumsum_area'];
  plot(t_day_fine_t,cumsum_area_t,'--')
  mx = max(size(cumsum_area_t));
  if     k==1, text(max(t_day_fine_t),cumsum_area_t(mx),'Benzene')
  elseif k==2, text(max(t_day_fine_t),cumsum_area_t(mx),'Toluene')
  elseif k==3,text(max(t_day_fine_t),cumsum_area_t(mx),'Ethylbenzene')
  elseif k==4, text(max(t_day_fine_t),cumsum_area_t(mx)-3,'m-xylene')
  elseif k==5, text(max(t_day_fine_t),cumsum_area_t(mx),'o-xylene')
  elseif k==6, text(max(t_day_fine_t),cumsum_area_t(mx)+3,'p-xylene')
```

```
  elseif k==7, text(max(t_day_fine_t),cumsum_area_t(mx)+3,'H2O')
  end
  % ************************************************************ %
  % READ & USE THE SECOND PLOT TO FIND THE RESIDENCE TIME.       %
  % FIRST: FIND THE distance AT THE TIME OF MEASUREMENT.         %
  %        SUBTRACT 9.1 AND 12.2 FROM THAT distance.             %
  % ************************************************************ %
  distance = table1([t_day_fine_t' cumsum_area_t'],t_day);
  distance91 = distance - 9.1;
  distance213 = distance - 12.2;
  % ************************************************************ %
  % SECOND: USE THE TWO SUBTRACTED distance SETS TO GET THEIR     %
  %         CORRESPONDING TIME => GET RESIDENCE TIME.            %
  % ************************************************************ %
  for j = 1:max(size(t_day))
    for i = t_day(j):-t_day_fine_gap:min(t_day_fine_t)
      distance_marker = table1([t_day_fine_t' cumsum_area_t'],i);
      if distance_marker<distance91(j)
t_day_91(j) = i; break
      end
    end
    for i = t_day(j):-t_day_fine_gap:min(t_day_fine_t)
      distance_marker = table1([t_day_fine_t' cumsum_area_t'],i);
      if distance_marker<distance213(j)
t_day_213(j) = i; break
      end
    end
  end
  residencetime91 = t_day - t_day_91;
  residencetime213 = t_day - t_day_213;
  if k==1,
    t_day_b_91 = t_day_91;
    t_day_b_213 = t_day_213;
    residencetime_b_91 = residencetime91;
    residencetime_b_213 = residencetime213;
  elseif k==2,
    t_day_t_91 = t_day_91;
    t_day_t_213 = t_day_213;
    residencetime_t_91 = residencetime91;
    residencetime_t_213 = residencetime213;
  elseif k==3,
    t_day_e_91 = t_day_91;
    t_day_e_213 = t_day_213;
    residencetime_e_91 = residencetime91;
    residencetime_e_213 = residencetime213;
  elseif k==4,
    t_day_m_91 = t_day_91;
    t_day_m_213 = t_day_213;
    residencetime_m_91 = residencetime91;
    residencetime_m_213 = residencetime213;
  elseif k==5,
    t_day_o_91 = t_day_91;
    t_day_o_213 = t_day_213;
    residencetime_o_91 = residencetime91;
    residencetime_o_213 = residencetime213;
  elseif k==6,
    t_day_p_91 = t_day_91;
    t_day_p_213 = t_day_213;
```

```
  residencetime_p_91 = residencetime91;
  residencetime_p_213 = residencetime213;
elseif k==7,
  t_day_h_91 = t_day_91;
  t_day_h_213 = t_day_213;
  residencetime_h_91 = residencetime91;
  residencetime_h_213 = residencetime213;
end
% ************************************************************** %
% PRINTOUT.                                                      %
% ************************************************************** %
disp(' ')
if k==1
  disp('  t_day_b_91 t_day_b_213    t_day ')
  disp('  -------- --------- ---------')
  disp([t_day_b_91' t_day_b_213' t_day'])
  disp('residencetime_b_91 residencetime_b_213')
  disp('---------------- ----------------')
  disp([residencetime_b_91' residencetime_b_213'])
elseif k==2
  disp('  t_day_t_91 t_day_t_213    t_day ')
  disp('  -------- --------- ---------')
  disp([t_day_t_91' t_day_t_213' t_day'])
  disp('residencetime_t_91 residencetime_t_213')
  disp('---------------- ----------------')
  disp([residencetime_t_91' residencetime_t_213'])
elseif k==3
  disp('  t_day_e_91 t_day_e_213    t_day ')
  disp('  -------- --------- ---------')
  disp([t_day_e_91' t_day_e_213' t_day'])
  disp('residencetime_e_91 residencetime_e_213')
  disp('---------------- ----------------')
  disp([residencetime_e_91' residencetime_e_213'])
elseif k==4
  disp('  t_day_m_91 t_day_m_213    t_day ')
  disp('  -------- --------- ---------')
  disp([t_day_m_91' t_day_m_213' t_day'])
  disp('residencetime_m_91 residencetime_m_213')
  disp('---------------- ----------------')
  disp([residencetime_m_91' residencetime_m_213'])
elseif k==5
  disp('  t_day_o_91 t_day_o_213    t_day ')
  disp('  -------- --------- ---------')
  disp([t_day_o_91' t_day_o_213' t_day'])
  disp('residencetime_o_91 residencetime_o_213')
  disp('---------------- ----------------')
  disp([residencetime_o_91' residencetime_o_213'])
elseif k==6
  disp('  t_day_p_91 t_day_p_213    t_day ')
  disp('  -------- --------- ---------')
  disp([t_day_p_91' t_day_p_213' t_day'])
  disp('residencetime_p_91 residencetime_p_213')
  disp('---------------- ----------------')
  disp([residencetime_p_91' residencetime_p_213'])
elseif k==7
```

```
    disp('  t_day_h_91 t_day_h_213    t_day ')
    disp('  -------- --------- ---------')
    disp([t_day_h_91' t_day_h_213' t_day'])
    disp('residencetime_h_91 residencetime_h_213')
    disp('---------------- ----------------')
    disp([residencetime_h_91' residencetime_h_213'])
  end
end
% ********************************************************************* %
% PRINTING THE RETARDATION FACTORS.                                     %
% ********************************************************************* %
disp('                 Retardation factors of:')
disp('    Benzene  Toluene Ethylbenzene m-Xylene o-Xylene p-Xylene')
disp('    -------  ------- ------------ -------- -------- --------')
disp( 1./(1+foc*Koc*B/n))
% ********************************************************************* %
% SAVE VARIABLES FOR tan7.m.                                            %
% ********************************************************************* %
save frtan6 t_day t_day_fine t_day_b_91 t_day_b_213  t_day_t_91
     t_day_t_213 t_day_e_91 t_day_e_213 t_day_m_91 t_day_m_213
     t_day_o_91 t_day_o_213 t_day_p_91 t_day_p_213 t_day_h_91 t_day_h_213
     residencetime_b_91 residencetime_b_213 residencetime_t_91
     residencetime_t_213 residencetime_e_91 residencetime_e_213
     residencetime_m_91 residencetime_m_213 residencetime_o_91
     residencetime_o_213 residencetime_p_91 residencetime_p_213
     residencetime_h_91 residencetime_h_213
% ***************************************E N D - P R O G R A M.
```

# APPENDIX D

## MATLAB Codes: Part II

```
% ************************************************************** %
% Written by: Tan, Ting Chow (29 May-30 June, 1994)              %
% Located at:/apple/tan/thesis/sleepbear/XIII/tan7.m             %
% Purpose: 1) Determine rate constants.                          %
% ************************************************************** %
clear, clg, clc
data, load frtan6
% ************************************************************** %
% SUMMING UP THE CONCENTRATION MEASUREMENTS AT EACH OF THE THREE  %
% CLUSTERS OF BOREHOLE.                                          %
% ************************************************************** %
[n,m] = size(benzene);
sum_bz_spill = sum(benzene(1:n,1:7)');
sum_bz_91 = sum(benzene(1:n,8:14)');
sum_bz_213 = sum(benzene(1:n,15:21)');
% ************************************************************** %
% BENZENE CONCENTRATION AT THE 1ST AND THE 13RD TIME POINTS ARE NOT %
% AVAILABLE.                                                     %
% ************************************************************** %
t_day_c     = t_day;
t_day_c(1)  = [];
t_day_c(12) = [];
t_day_c_b_91     = t_day_b_91;
t_day_c_b_91(1)  = [];
t_day_c_b_91(12) = [];
t_day_c_t_91     = t_day_t_91;
t_day_c_t_91(1)  = [];
t_day_c_t_91(12) = [];
t_day_c_e_91     = t_day_e_91;
t_day_c_e_91(1)  = [];
t_day_c_e_91(12) = [];
t_day_c_m_91     = t_day_m_91;
t_day_c_m_91(1)  = [];
t_day_c_m_91(12) = [];
t_day_c_o_91     = t_day_o_91;
t_day_c_o_91(1)  = [];
t_day_c_o_91(12) = [];
t_day_c_p_91     = t_day_p_91;
t_day_c_p_91(1)  = [];
t_day_c_p_91(12) = [];
t_day_c_b_213     = t_day_b_213;
t_day_c_b_213(1)  = [];
t_day_c_b_213(12) = [];
t_day_c_t_213     = t_day_t_213;
t_day_c_t_213(1)  = [];
t_day_c_t_213(12) = [];
```

```
t_day_c_e_213     = t_day_e_213;
t_day_c_e_213(1)  = [];
t_day_c_e_213(12) = [];
t_day_c_m_213     = t_day_m_213;
t_day_c_m_213(1)  = [];
t_day_c_m_213(12) = [];
t_day_c_o_213     = t_day_o_213;
t_day_c_o_213(1)  = [];
t_day_c_o_213(12) = [];
t_day_c_p_213     = t_day_p_213;
t_day_c_p_213(1)  = [];
t_day_c_p_213(12) = [];
% ************************************************************** %
% t_day_c_fine IS CREATED TO LIMIT THE EXTRAPOLATION SHOWN IN THE  %
% NEXT PLOT.                                                      %
% ************************************************************** %
t_day_c_fine = t_day_fine;
t_day_c_fine(1:20) = [];
t_day_cc_fine = t_day_fine;
t_day_cc_fine(1:25) = [];
t_day_ccc = t_day_c;
% ************************************************************** %
% BTEX CONCENTRATION VS TIME PLOT.                                %
% ************************************************************** %
fine = interp1(t_day_c,sum(benzene(:,1:7)'),t_day_cc_fine,'spline');
subplot(3,3,1), plot(t_day_c,sum(benzene(:,1:7)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Benzene'), title('Spill Site'), axis([0 690 0 2e4])
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(benzene(:,8:14)'),t_day_cc_fine,'spline');
subplot(3,3,2), plot(t_day_c,sum(benzene(:,8:14)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Benzene'), title('9.1 m Site'), axis([0 690 0 2000])
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(benzene(:,15:21)'),t_day_cc_fine,'spline');
subplot(3,3,3), plot(t_day_c,sum(benzene(:,15:21)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Benzene'), title('21.3 m Site'), axis([0 690 0 1500])
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(toluene(:,1:7)'),t_day_cc_fine,'spline');
subplot(3,3,4), plot(t_day_c,sum(toluene(:,1:7)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Toluene'), axis([0 690 2e4 10e4]), title('Spill Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(toluene(:,8:14)'),t_day_cc_fine,'spline');
subplot(3,3,5), plot(t_day_c,sum(toluene(:,8:14)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Toluene'), axis([0 690 0 9000]), title('9.1 m Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(toluene(:,15:21)'),t_day_cc_fine,'spline');
subplot(3,3,6), plot(t_day_c,sum(toluene(:,15:21)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Toluene'), axis([0 690 0 500]), title('21.3 m Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(ethylbenzene(:,1:7)'),t_day_cc_fine,'spline');
subplot(3,3,7), plot(t_day_c,sum(ethylbenzene(:,1:7)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Ethylbenzene'), axis([0 690 .6e4 1.4e4]), title('Spill Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(ethylbenzene(:,8:14)'),t_day_cc_fine,'spline');
```

```
subplot(3,3,8), plot(t_day_c,sum(ethylbenzene(:,8:14)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Ethylbenzene'), axis([0 690 500 2500]), title('9.1 m Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(ethylbenzene(:,15:21)'),t_day_cc_fine,'spline');
subplot(3,3,9), plot(t_day_c,sum(ethylbenzene(:,15:21)'),'o',t_day_cc_fine,fine)
  grid, ylabel('Ethylbenzene'), axis([0 690 0 800]), title('21.3 m Site')
  xlabel('Julian Date')
% print f7uncorrflowa
pause, clg
fine = interp1(t_day_c,sum(metaxylene(:,1:7)'),t_day_cc_fine,'spline');
subplot(3,3,1), plot(t_day_c,sum(metaxylene(:,1:7)'),'o',t_day_cc_fine,fine)
  grid, ylabel('m-xylene'), axis([0 690 1e4 3e4]), title('Spill Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(metaxylene(:,8:14)'),t_day_cc_fine,'spline');
subplot(3,3,2), plot(t_day_c,sum(metaxylene(:,8:14)'),'o',t_day_cc_fine,fine)
  grid, ylabel('m-xylene'), axis([0 690 0 4000]), title('9.1 m Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(metaxylene(:,15:21)'),t_day_cc_fine,'spline');
subplot(3,3,3), plot(t_day_c,sum(metaxylene(:,15:21)'),'o',t_day_cc_fine,fine)
  grid, ylabel('m-xylene'), axis([0 690 0 1400]), title('21.3 m Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(orthoxylene(:,1:7)'),t_day_cc_fine,'spline');
subplot(3,3,4), plot(t_day_c,sum(orthoxylene(:,1:7)'),'o',t_day_cc_fine,fine)
  grid, ylabel('o-xylene'), axis([0 690 .9e4 1.8e4]), title('Spill Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(orthoxylene(:,8:14)'),t_day_cc_fine,'spline');
subplot(3,3,5), plot(t_day_c,sum(orthoxylene(:,8:14)'),'o',t_day_cc_fine,fine)
  grid, ylabel('o-xylene'), axis([0 690 0 2500]), title('9.1 m Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(orthoxylene(:,15:21)'),t_day_cc_fine,'spline');
subplot(3,3,6), plot(t_day_c,sum(orthoxylene(:,15:21)'),'o',t_day_cc_fine,fine)
  grid, ylabel('o-xylene'), axis([0 690 0 1000]), title('21.3 m Site')
  xlabel('Julian Date')
fine = interp1(t_day_c,sum(paraxylene(:,1:7)'),t_day_cc_fine,'spline');
subplot(3,3,7), plot(t_day_c,sum(paraxylene(:,1:7)'),'o',t_day_cc_fine,fine)
  grid, ylabel('p-xylene'), xlabel('Julian Date'), axis([0 690 0.5e4 1.5e4])
  title('Spill Site')
fine = interp1(t_day_c,sum(paraxylene(:,8:14)'),t_day_cc_fine,'spline');
subplot(3,3,8), plot(t_day_c,sum(paraxylene(:,8:14)'),'o',t_day_cc_fine,fine)
  grid, ylabel('p-xylene'), xlabel('Julian Date'), axis([0 690 0 2500])
  title('9.1 m Site')
fine = interp1(t_day_c,sum(paraxylene(:,15:21)'),t_day_cc_fine,'spline');
subplot(3,3,9), plot(t_day_c,sum(paraxylene(:,15:21)'),'o',t_day_cc_fine,fine)
  grid, ylabel('p-xylene'), xlabel('Julian Date'), axis([0 690 0 800])
  title('21.3 m Site')
% print f7uncorrflowb
pause, clg
```

```
% ************************************************************** %
% THE FIRST THREE DATA POINTS ARE NOT USABLE BECAUSE THE DILUTION  %
% CORRECTION FACTOR OF THOSE POINTS ARE NOT AVAILABLE/CALCULABLE.  %
% ************************************************************** %
t_day_c_b_91(1:3)  = [];
t_day_c_b_213(1:3) = [];
t_day_c_t_91(1:3)  = [];
t_day_c_t_213(1:3) = [];
t_day_c_e_91(1:3)  = [];
t_day_c_e_213(1:3) = [];
t_day_c_m_91(1:3)  = [];
t_day_c_m_213(1:3) = [];
t_day_c_o_91(1:3)  = [];
t_day_c_o_213(1:3) = [];
t_day_c_p_91(1:3)  = [];
t_day_c_p_213(1:3) = [];
% ************************************************************** %
% ONE EXTRAPOLATION POINT IS ADDED FOR THE BENZENE CONCENTRATION SO %
% THAT CERTAIN DATA POINTS THAT OTHERWISE CANNOT BE USED CAN BE     %
% USED. THE EXTRAPOLATION SEEM REASONABLE. HOWEVER, FIRST 3 DATA    %
% POINTS HAVE TO BE DISCARDED BECAUSE THEY'RE LOCATED TOO FAR TO BE %
% EXTRAPOLATED. SEE THE RESIDENCE TIME DATA.                        %
% ************************************************************** %
sum_bzt_spill_91 = table1([[40 t_day_c]' [2e4 sum_bz_spill]'],t_day_c_t_91) ;
sum_bzt_91_213   = table1([[40 t_day_c]' [2e4 sum_bz_91]'],   t_day_c_t_213);
sum_bze_spill_91 = table1([[40 t_day_c]' [2e4 sum_bz_spill]'],t_day_c_e_91) ;
sum_bze_91_213   = table1([[40 t_day_c]' [2e4 sum_bz_91]'],   t_day_c_e_213);
sum_bzm_spill_91 = table1([[40 t_day_c]' [2e4 sum_bz_spill]'],t_day_c_m_91) ;
sum_bzm_91_213   = table1([[40 t_day_c]' [2e4 sum_bz_91]'],   t_day_c_m_213);
sum_bzo_spill_91 = table1([[40 t_day_c]' [2e4 sum_bz_spill]'],t_day_c_o_91) ;
sum_bzo_91_213   = table1([[40 t_day_c]' [2e4 sum_bz_91]'],   t_day_c_o_213);
sum_bzp_spill_91 = table1([[40 t_day_c]' [2e4 sum_bz_spill]'],t_day_c_p_91) ;
sum_bzp_91_213   = table1([[40 t_day_c]' [2e4 sum_bz_91]'],   t_day_c_p_213);
% ************************************************************** %
% THE FIRST FOUR DATA POINTS ARE NOT USABLE BECAUSE THE DILUTION   %
% CORRECTION FACTOR OF THOSE POINTS ARE NOT AVAILABLE/CALCULABLE.  %
% ************************************************************** %
sum_bz_spill_c = sum_bz_spill;
sum_bz_spill(1:3) = [];
sum_bz_91(1:3)    = [];
sum_bz_213(1:3)   = [];
% ************************************************************** %
% CALC THE DILUTION CORRECTION FACTORS.                            %
% ************************************************************** %
correction91_t  = sum_bzt_spill_91./sum_bz_91';
correction213_t = sum_bzt_91_213  ./sum_bz_213';
correction91_e  = sum_bze_spill_91./sum_bz_91';
correction213_e = sum_bze_91_213  ./sum_bz_213';
correction91_m  = sum_bzm_spill_91./sum_bz_91';
correction213_m = sum_bzm_91_213  ./sum_bz_213';
correction91_o  = sum_bzo_spill_91./sum_bz_91';
correction213_o = sum_bzo_91_213  ./sum_bz_213';
correction91_p  = sum_bzp_spill_91./sum_bz_91';
```

```
correction213_p = sum_bzp_91_213  ./sum_bz_213';
% ************************************************************** %
% CHECKING THE DILUTION CORRECTION FACTORS AS DETERMINED BY THE    %
% 2,3-DIMETHYLPENTANE (DMP). SINCE ONLY TWO POINTS ARE AVAILABLE,  %
% THE AVERAGE OF THE TWO BENZENE CONCENTRATION IS USED FOR THE     %
% SPILL SITE AND THE 9.1 M SITE WHEN THE CONCENTRATION ARE USED AS %
% THE INITIAL CONCENTRATION.                                       %
% ************************************************************** %
dmp_spill = [(70+69+17.8+7.3+81.8+79.2+36.3+11.4)/2 ...
     (70+69+17.8+7.3+81.8+79.2+36.3+11.4)/2];
dmp_91  = [38+7.4+5.7   17.3+10.8+8.14];
dmp_213 = [2.1+7.5+1.9  11.8+1.5];
dmp_spill_fine = interp1(t_day(14:15),dmp_spill,t_day_fine(60:100),'spline');
dmp_spill_91 = table1([t_day_fine(60:100)' dmp_spill_fine],t_day_c_b_91(9:10));
correction91_dmp = dmp_spill_91./dmp_91';
dmp_91 = [(38+7.4+5.7+17.3+10.8+8.14)/2 (38+7.4+5.7+17.3+10.8+8.14)/2];
dmp_91_fine = interp1(t_day(14:15),dmp_91,t_day_fine(60:100),'spline');
% subplot(2,1,2), plot(t_day(14:15),dmp_91,'o',t_day_fine(60:100),dmp_91_fine)
%   xlabel('Julian Date'), ylabel('DMP ppb'), grid, pause, clg
dmp_91_213 = table1([t_day_fine(60:100)' dmp_91_fine], t_day_c_b_213(9:10));
correction213_dmp = dmp_91_213./dmp_213';
% ************************************************************** %
% COMPARE THE DILUTION CORRECTION FACTORS AS CALCULATED BY BENZENE %
% AND THE DILUTION CORRECTION FACTORS AS CALCULATED BY DMP.        %
% ************************************************************** %
disp('')
disp('correction91_t correction213_t correction91_e correction213_e')
disp('-------------- -------------- -------------- --------------')
disp([correction91_t correction213_t correction91_e correction213_e])
disp('')
disp('correction91_m correction213_m correction91_o correction213_o')
disp('-------------- -------------- -------------- --------------')
disp([correction91_m correction213_m correction91_o correction213_o])
disp('')
disp('correction91_p correction213_p')
disp('-------------- --------------')
disp([correction91_p correction213_p])
disp('')
disp('correction91_dmp correction213_dmp')
disp('---------------- ----------------')
disp([correction91_dmp correction213_dmp])
disp('')
% ************************************************************** %
% GET THE INITIAL CONCENTRATIONS OF TEX AT 9.1 M SITE FROM THE     %
% SPILL SITE. THIS IS NEEDED FOR RATE CONSTANT CALCULATED AND      %
% NORMALIZATION CALCULATION.                                       %
% ************************************************************** %
for i = 1:7
  ethylbenzenespill91(:,i) = table1([t_day_c' ethylbenzene(:,i)], t_day_c_e_91);
  metaxylenespill91(:,i)   = table1([t_day_c' metaxylene(:,i)],   t_day_c_m_91);
  orthoxylenespill91(:,i)  = table1([t_day_c' orthoxylene(:,i)],  t_day_c_o_91);
  paraxylenespill91(:,i)   = table1([t_day_c' paraxylene(:,i)],   t_day_c_p_91);
  toluenespill91(:,i)      = table1([t_day_c' toluene(:,i)],      t_day_c_t_91);
```

```
end
% ********************************************************************* %
% GET THE INITIAL CONCENTRATIONS OF TEX AT 21.3 M SITE FROM THE         %
% 9.1 M SITE. THIS IS NEEDED FOR RATE CONSTANT CALCULATION AND          %
% NORMALIZATION CALCULATION. ONE ARTIFICIAL POINT IS ADDED IN ORDER     %
% TO USE THE TWO DATA POINTS THAT OTHERWISE CANNOT BE USED.             %
% ********************************************************************* %
for i = 8:14
  ethylbenzene91213(:,i-7) = table1([[0 t_day_c]' ...
    [ethylbenzene(1,i); ethylbenzene(:,i)]],t_day_c_e_213);
  metaxylene91213(:,i-7) = table1([[0 t_day_c]' ...
    [metaxylene(1,i); metaxylene(:,i)]],t_day_c_m_213);
  orthoxylene91213(:,i-7) = table1([[0 t_day_c]' ...
    [orthoxylene(1,i); orthoxylene(:,i)]],t_day_c_o_213);
  paraxylene91213(:,i-7) = table1([[0 t_day_c]' ...
    [paraxylene(1,i); paraxylene(:,i)]],t_day_c_p_213);
  toluene91213(:,i-7) = table1([[0 t_day_c]' ...
    [toluene(1,i); toluene(:,i)]],t_day_c_t_213);
end
% ********************************************************************* %
% THE FIRST FOUR DATA POINTS ARE NOT USABLE BECAUSE THE DILUTION        %
% CORRECTION FACTOR OF THOSE POINTS ARE NOT AVAILABLE/CALCULABLE.       %
% ********************************************************************* %
t_day_c(1:3)         = [];
ethylbenzene(1:3,:) = [];
metaxylene(1:3,:)    = [];
orthoxylene(1:3,:)   = [];
paraxylene(1:3,:)    = [];
toluene_c = toluene;
toluene(1:3,:)       = [];
% ********************************************************************* %
% CORRECT THE DATA AT 9.1 AND 21.3 M SITES.                             %
% ********************************************************************* %
for i = 8:14
  ethylbenzene91(:,i-7) = ethylbenzene(:,i).*correction91_e;
  metaxylene91(:,i-7)   = metaxylene(:,i).*correction91_m;
  orthoxylene91(:,i-7)  = orthoxylene(:,i).*correction91_o;
  paraxylene91(:,i-7)   = paraxylene(:,i).*correction91_p;
  toluene91(:,i-7)      = toluene(:,i).*correction91_t;
end
for i = 15:21
  ethylbenzene213(:,i-14) = ethylbenzene(:,i).*correction213_e;
  metaxylene213(:,i-14)   = metaxylene(:,i).*correction213_m;
  orthoxylene213(:,i-14)  = orthoxylene(:,i).*correction213_o;
  paraxylene213(:,i-14)   = paraxylene(:,i).*correction213_p;
  toluene213(:,i-14)      = toluene(:,i).*correction213_t;
end
% ********************************************************************* %
% CONCENTRATION VS TIME PLOT OF BENZENE AND TOLUENE FOR ABSTRACT.       %
% ********************************************************************* %
% abstr_plot
```

```
% ************************************************************** %
% DILUTION CORRECTED BTEX CONCENTRATION VS TIME PLOT.            %
% ************************************************************** %
t_day_cc_fine(1:10) = [];
subplot(5,2,1)
  sum_et_91 = sum(ethylbenzene91');
  sum_et_91fine = interp1(t_day_c,sum_et_91,t_day_cc_fine,'spline');
  plot(t_day_c,sum_et_91,'o',t_day_cc_fine,sum_et_91fine)
  grid, ylabel('Ethylbenzene'), title('9.1 m Site')
  axis([130 690 0 40000])
subplot(5,2,2)
  sum_et_213 = sum(ethylbenzene213');
  sum_et_213fine = interp1(t_day_c,sum_et_213,t_day_cc_fine,'spline');
  plot(t_day_c,sum_et_213,'o',t_day_cc_fine,sum_et_213fine)
  grid, ylabel('Ethylbenzene'), title('21.3 m Site')
  axis([130 690 0 1000])
subplot(5,2,3)
  sum_me_91 = sum(metaxylene91');
  sum_me_91fine = interp1(t_day_c,sum_me_91,t_day_cc_fine,'spline');
  plot(t_day_c,sum_me_91,'o',t_day_cc_fine,sum_me_91fine)
  grid, ylabel('m-xylene')
  axis([130 690 0 50000])
subplot(5,2,4)
  sum_me_213 = sum(metaxylene213');
  sum_me_213fine = interp1(t_day_c,sum_me_213,t_day_cc_fine,'spline');
  plot(t_day_c,sum_me_213,'o',t_day_cc_fine,sum_me_213fine)
  grid, ylabel('m-xylene')
  axis([130 690 0 4000])
subplot(5,2,5)
  sum_or_91 = sum(orthoxylene91');
  sum_or_91fine = interp1(t_day_c,sum_or_91,t_day_cc_fine,'spline');
  plot(t_day_c,sum_or_91,'o',t_day_cc_fine,sum_or_91fine)
  grid, ylabel('o-xylene')
  axis([130 690 0 20000])
subplot(5,2,6)
  sum_or_213 = sum(orthoxylene213');
  sum_or_213fine = interp1(t_day_c,sum_or_213,t_day_cc_fine,'spline');
  plot(t_day_c,sum_or_213,'o',t_day_cc_fine,sum_or_213fine)
  grid, ylabel('o-xylene')
  axis([130 690 0 1500])
subplot(5,2,7)
  sum_pa_91 = sum(paraxylene91');
  sum_pa_91fine = interp1(t_day_c,sum_pa_91,t_day_cc_fine,'spline');
  plot(t_day_c,sum_pa_91,'o',t_day_cc_fine,sum_pa_91fine)
  grid, ylabel('p-xylene')
  axis([130 690 0 40000])
subplot(5,2,8)
  sum_pa_213 = sum(paraxylene213');
  sum_pa_213fine = interp1(t_day_c,sum_pa_213,t_day_cc_fine,'spline');
  plot(t_day_c,sum_pa_213,'o',t_day_cc_fine,sum_pa_213fine)
  grid, ylabel('p-xylene')
  axis([130 690 0 1000])
subplot(5,2,9)
```

```
  sum_to_91 = sum(toluene91');
  sum_to_91fine = interp1(t_day_c,sum_to_91,t_day_cc_fine,'spline');
  plot(t_day_c,sum_to_91,'o',t_day_cc_fine,sum_to_91fine)
  grid, ylabel('toluene'), xlabel('Julian date')
  axis([130 690 0 90000])
subplot(5,2,10)
  sum_to_213 = sum(toluene213');
  sum_to_213fine = interp1(t_day_c,sum_to_213,t_day_cc_fine,'spline');
  plot(t_day_c,sum_to_213,'o',t_day_cc_fine,sum_to_213fine)
  grid, ylabel('toluene'), xlabel('Julian date')
  axis([130 690 0 500])
% print f8corrflow
pause, clg
% ********************************************************************* %
% sigma^2/sum(sigma^2) OF TIME VS VARIABLES PLOT AT 9.1 M & 21.3 M. %
% ********************************************************************* %
AA = [sum_et_91' sum_to_91' sum_me_91' sum_or_91' sum_pa_91'];
[A,mA,stdA] = auto(AA);
[U,S1,VA] = svd(A,0);
[m n] = size(A);
S(:,1) = diag(S1).^2;
total = sum(S);
for i  = 1:n, ratio(i,:) = S(i,:)./total; end
cumratio(1,:) = ratio(1,:);
for i  = 2:n, cumratio(i,:) = sum(ratio(1:i,:)); end
cumratio = cumratio*100;
subplot(2,1,1), plot([1:n],cumratio), hold on, plot([1:n],cumratio,'o'), grid
ylabel('% Info captured by PC'), xlabel('PC index'), title('(a) At 9.1 m Site.')
BB = [sum_et_213' sum_to_213' sum_me_213' sum_or_213' sum_pa_213'];
[B,mB,stdB] = auto(BB);
[U,S1,VB] = svd(B,0);
[m n] = size(B);
S(:,1) = diag(S1).^2;
total = sum(S);
for i  = 1:n, ratio(i,:) = S(i,:)./total; end
cumratio(1,:) = ratio(1,:);
for i  = 2:n, cumratio(i,:) = sum(ratio(1:i,:)); end
cumratio = cumratio*100;
subplot(2,1,2), plot([1:n],cumratio), hold on, plot([1:n],cumratio,'o'), grid
ylabel('% Info captured by PC'), xlabel('PC index'), title('(b) At 21.3 m Site.')
% print f9sigma
pause, clg
% ********************************************************************* %
% GABRIEL'S PLOTS FOR 9.1 M & 21.3 M.                                   %
% ********************************************************************* %
subplot(2,2,1)
  for i  = 1:5
    plot3([VA(i,2) 0],[VA(i,1) 0],[VA(i,3) 0]), hold on
  end, hold off, grid
  text(VA(1,2),VA(1,1),VA(1,3),'Ethylbenzene')
  text(VA(2,2),VA(2,1),VA(2,3),'Toluene')
  text(VA(3,2),VA(3,1),VA(3,3),'m-xylene')
  text(VA(4,2),VA(4,1),VA(4,3),'o-xylene')
  text(VA(5,2),VA(5,1),VA(5,3),'p-xylene')
```

```
  xlabel('2nd PC'), ylabel('1st PC'), zlabel('3rd PC')
  view(80,80)
  title('(a) 9.1 m Site')
subplot(2,2,3)
  for i  = 1:5
    plot3([VA(i,2) 0],[VA(i,1) 0],[VA(i,3) 0]), hold on
  end, hold off, grid
  text(VA(1,2),VA(1,1),VA(1,3),'Ethylbenzene')
  text(VA(2,2),VA(2,1),VA(2,3),'Toluene')
  text(VA(3,2),VA(3,1),VA(3,3),'m-xylene')
  text(VA(4,2),VA(4,1),VA(4,3),'o-xylene')
  text(VA(5,2),VA(5,1),VA(5,3),'p-xylene')
  xlabel('2nd PC'), ylabel('1st PC'), zlabel('3rd PC')
  view(5,10)
  title('(c) 9.1 m Site - different view')
subplot(2,2,2)
  for i  = 1:5
    plot3([VB(i,2) 0],[VB(i,1) 0],[VB(i,3) 0]), hold on
  end, hold off, grid
  text(VB(1,2),VB(1,1),VB(1,3),'Ethylbenzene')
  text(VB(2,2),VB(2,1),VB(2,3),'Toluene')
  text(VB(3,2),VB(3,1),VB(3,3),'m-xylene')
  text(VB(4,2),VB(4,1),VB(4,3),'o-xylene')
  text(VB(5,2),VB(5,1),VB(5,3),'p-xylene')
  xlabel('2nd PC'), ylabel('1st PC'), zlabel('3rd PC')
  view(80,80)
  title('(b) 21.3 m Site')
subplot(2,2,4)
  for i  = 1:5
    plot3([VB(i,2) 0],[VB(i,1) 0],[VB(i,3) 0]), hold on
  end, hold off, grid
  text(VB(1,2),VB(1,1),VB(1,3),'Ethylbenzene')
  text(VB(2,2),VB(2,1),VB(2,3),'Toluene')
  text(VB(3,2),VB(3,1),VB(3,3),'m-xylene')
  text(VB(4,2),VB(4,1),VB(4,3),'o-xylene')
  text(VB(5,2),VB(5,1),VB(5,3),'p-xylene')
  xlabel('2nd PC'), ylabel('1st PC'), zlabel('3rd PC')
  view(5,10)
  title('(d) 21.3 m Site - different view')
% print gabrielplots
pause
% ********************************************************************* %
% VARIMAXED GABRIEL'S PLOTS FOR 9.1 M & 21.3 M.                         %
% ********************************************************************* %
[VAm,T,refstr,phi]=vrotate(VA(:,1:3),1,1,1);
[VBm,T,refstr,phi]=vrotate(VB(:,1:3),1,1,1);
disp('                9.1 m Site                              21.3 m Site')
disp('     1st PC       2nd PC      3rd PC     1st PC       2nd PC      3rd PC')
disp('-----------------------------------    -----------------------------------')
disp([VAm VBm])
disp('-----------------------------------    -----------------------------------')
```

```
% ************************************************************** %
% REARRANGE 4 MATRICES OF 11 BOREHOLE VS VARIABLE SUBMATRICES IN  %
% STACK FORMAT.                                                  %
% ************************************************************** %
for i = 1:10
  snapshotspill91(i*7-6:i*7,:) = [ethylbenzenespill91(i,:)' ...
                  metaxylenespill91(i,:)' ...
                  orthoxylenespill91(i,:)'...
                  paraxylenespill91(i,:)'...
                  toluenespill91(i,:)'];
  snapshot91213(i*7-6:i*7,:)   = [ethylbenzene91213(i,:)' ...
                                  metaxylene91213(i,:)' ...
                                  orthoxylene91213(i,:)'...
                                  paraxylene91213(i,:)'...
                                  toluene91213(i,:)'];
  snapshot91(i*7-6:i*7,:) =  [ethylbenzene91(i,:)' ...
      metaxylene91(i,:)' ...
      orthoxylene91(i,:)'...
      paraxylene91(i,:)'...
      toluene91(i,:)'];
  snapshot213(i*7-6:i*7,:) = [ethylbenzene213(i,:)' ...
                              metaxylene213(i,:)' ...
                              orthoxylene213(i,:)'...
                              paraxylene213(i,:)'...
                              toluene213(i,:)'];
end
% ************************************************************** %
% MEAN-CENTER THE ABOVE MATRICES. NORMALIZE THE 9.1 AND          %
% 21.3 M MATRICES WITH THEIR CORRESPONDING SPILLED SITE MATRICES.%
% NEXT, FIND THE SUM OF THE SVD OF THE RESULTANT MEAN-CENTERED AND%
% AND NORMALIZED MATRICES.                                       %
% ************************************************************** %
for i = 1:10
  dsnapshotspill91(i*7-6:i*7,:) = dtrend(snapshotspill91(i*7-6:i*7,:));
  dsnapshot91213(i*7-6:i*7,:)   = dtrend(snapshot91213(i*7-6:i*7,:));
  dsnapshot91(i*7-6:i*7,:)      = dtrend(snapshot91(i*7-6:i*7,:));
  dsnapshot213(i*7-6:i*7,:)     = dtrend(snapshot213(i*7-6:i*7,:));
  for j = 1:5
    devsnapshotspill91(j)          = sqrt(sum(dsnapshotspill91(i*7-6:i*7,j).^2));
    sdsnapshot91(i*7-6:i*7,j)      = dsnapshot91(i*7-6:i*7,j)/devsnapshotspill91(j);
    sdsnapshotspill91(i*7-6:i*7,j)=dsnapshotspill91(i*7-6:i*7,j)/devsnapshotspill91(
    devsnapshot91213(j)            = sqrt(sum(dsnapshot91213(i*7-6:i*7,j).^2));
    sdsnapshot213(i*7-6:i*7,j)     = dsnapshot213(i*7-6:i*7,j)/devsnapshot91213(j);
    sdsnapshot91213(i*7-6:i*7,j)   = dsnapshot91213(i*7-6:i*7,j)/devsnapshot91213(j)
  end
  sssvdsdsnapshotspill91(i) = sum(svd(sdsnapshotspill91(i*7-6:i*7,:)).^2);
  sssvdsdsnapshot91213(i)   = sum(svd(sdsnapshot91213(i*7-6:i*7,:))  .^2);
  sssvdsdsnapshot91(i)      = sum(svd(sdsnapshot91(i*7-6:i*7,:))     .^2);
  sssvdsdsnapshot213(i)     = sum(svd(sdsnapshot213(i*7-6:i*7,:))    .^2);
end
```

```
% ************************************************************ %
% FIRST: THE AVERAGE OF THE RESIDENCE TIME IS DETERMINED.      %
% SECOND: BENZENE CONCENTRATION AT THE 1ST AND THE 13RD TIME POINTS %
% ARE NOT AVAILABLE. FURTHERMORE THE FIRST FOUR DATA POINTS ARE NOT %
% USABLE BECAUSE THE DILUTION CORRECTION FACTOR OF THOSE POINTS %
% ARE NOT AVAILABLE/CALCULABLE => THESE POINTS'RE DELETED.     %
% ************************************************************ %
residencetime91_c = (residencetime_b_91 + residencetime_t_91 + ...
                     residencetime_e_91 + residencetime_m_91 + ...
                     residencetime_o_91 + residencetime_p_91)/6;
residencetime91_c(1)    = [];
residencetime91_c(12)   = [];
residencetime91_c(1:3)  = [];
residencetime213_c = (residencetime_b_213 + residencetime_t_213 + ...
                      residencetime_e_213 + residencetime_m_213 + ...
                      residencetime_o_213 + residencetime_p_213)/6;
residencetime213_c(1)   = [];
residencetime213_c(12)  = [];
residencetime213_c(1:3) = [];
% ************************************************************ %
% SUM OF SQUARE OF THE SINGULAR VALUES VS RESIDENCE TIME PLOTS  %
% AT 9.1 AND 21.3 M.                                           %
% ************************************************************ %
subplot(2,1,1)
  [sortedrestime91,Indexrestime91] = sort([residencetime91_c]);
  sortedrestime91fine = linspace(min(sortedrestime91),max(sortedrestime91));
  sssvdsdsnapshot91fine = interp1(sortedrestime91, ...
    sssvdsdsnapshot91(Indexrestime91),sortedrestime91fine,'spline');
  plot(sortedrestime91fine,sssvdsdsnapshot91fine)
  text(residencetime91_c(1),sssvdsdsnapshot91(1),'4')
  text(residencetime91_c(2),sssvdsdsnapshot91(2),'5')
  text(residencetime91_c(3),sssvdsdsnapshot91(3),'6')
  text(residencetime91_c(4),sssvdsdsnapshot91(4),'7')
  text(residencetime91_c(5),sssvdsdsnapshot91(5),'8')
  text(residencetime91_c(6),sssvdsdsnapshot91(6),'9')
  text(residencetime91_c(7),sssvdsdsnapshot91(7),'10')
  text(residencetime91_c(8),sssvdsdsnapshot91(8),'11')
  text(residencetime91_c(9),sssvdsdsnapshot91(9),'12')
  text(residencetime91_c(10),sssvdsdsnapshot91(10),'13')
  title('(a)'), grid, hold on
  ylabel('ssvd at 9.1m'), xlabel('Residence time (day)')
subplot(2,1,2)
  [sortedrestime213,Indexrestime213] = sort([residencetime213_c]);
  sortedrestime213fine = linspace(min(sortedrestime213),max(sortedrestime213));
  sssvdsdsnapshot213fine = interp1(sortedrestime213, ...
    sssvdsdsnapshot213(Indexrestime213),sortedrestime213fine,'spline');
  plot(sortedrestime213fine,sssvdsdsnapshot213fine)
  text(residencetime213_c(1),sssvdsdsnapshot213(1),'4')
  text(residencetime213_c(2),sssvdsdsnapshot213(2),'5')
  text(residencetime213_c(3),sssvdsdsnapshot213(3),'6')
  text(residencetime213_c(4),sssvdsdsnapshot213(4),'7')
  text(residencetime213_c(5),sssvdsdsnapshot213(5),'8')
  text(residencetime213_c(6),sssvdsdsnapshot213(6),'9')
  text(residencetime213_c(7),sssvdsdsnapshot213(7),'10')
```

```
  text(residencetime213_c(8),sssvdsdsnapshot213(8),'11')
  text(residencetime213_c(9),sssvdsdsnapshot213(9),'12')
  text(residencetime213_c(10),sssvdsdsnapshot213(10),'13')
  title('(b)'), grid, hold on
  ylabel('ssvd at 21.3m'), xlabel('Residence time (day)')
% ************************************************************** %
% ASSUMING 1ST ORDER BIODEGRADATION, THE RATE CONSTANTS ARE:     %
% ************************************************************** %
  sortedrestime91(1) = [];
  sssvdsdsnapshot91(1) = [];
  Indexrestime91(1) = [];
  X = pinv([ones(size(sortedrestime91')) sortedrestime91'])* ...
      log(sssvdsdsnapshot91(Indexrestime91-1)')
  subplot(2,1,1), plot(linspace(3,270),exp(X(1))*exp(X(2)*linspace(3,270)),'--')
  text(5,140,'ssvd=133.8exp(-0.0133t)')
  Xb = pinv([ones(size(sortedrestime213')) sortedrestime213'])* ...
      log(sssvdsdsnapshot213(Indexrestime213)')
  subplot(2,1,2), plot(linspace(3,270),exp(Xb(1))*exp(Xb(2)*linspace(3,270)),'--')
  text(5,7,'ssvd=15.0exp(-0.0040t)')
% print f5sumsvd
% pause, clg
% ************************************************************** %
% WATER LEVELS VERSUS TIME PLOT.                                 %
% ************************************************************** %
H20level_spill(1:3) = [];
H20level_91(1:3) = [];
H20level_213(1:3) = [];
t_day_h2o_fine = linspace(135,700);
H20level_spill_fine = interp1(t_day_c,H20level_spill,t_day_h2o_fine,'spline');
H20level_91_fine    = interp1(t_day_c,H20level_91,   t_day_h2o_fine,'spline');
H20level_213_fine   = interp1(t_day_c,H20level_213,  t_day_h2o_fine,'spline');
plot(t_day_c,H20level_spill,'o',t_day_h2o_fine,H20level_spill_fine), hold on
plot(t_day_c,H20level_91   ,'o',t_day_h2o_fine,H20level_91_fine,'--')
plot(t_day_c,H20level_213  ,'o',t_day_h2o_fine,H20level_213_fine,'-.')
ylabel('Water Levels (ft)'), xlabel('Julian Date'), grid
text(420,585.45,'Solid line: At spill site')
text(420,585.55,'Dashed line: At 9.1 m site')
text(420,585.65,'Dashed-dot line: At 21.3 m site')
% print f6h2olev
% ************************************************************** %
% PARAMETERS VS TIME PLOT.                                       %
% ************************************************************** %
t_day_c     = t_day;
t_day_c(1)  = [];
t_day_c(12) = [];
t_day_c(12:13) = [];
subplot(6,3,1), plot(t_day_c,  alkalinity(:,1:7)), grid, ylabel('Alkalinity')
  title('Spill Site')
subplot(6,3,2), plot(t_day_c, alkalinity(:,8:14)), grid, ylabel('Alkalinity')
  title('9.1 m Site')
subplot(6,3,3), plot(t_day_c,alkalinity(:,15:21)), grid, ylabel('Alkalinity')
  title('21.3 m Site')
```

```
subplot(6,3,4), plot(t_day_c,  ammonia(:,1:7)), grid, ylabel('Ammonia')
subplot(6,3,5), plot(t_day_c, ammonia(:,8:14)), grid, ylabel('Ammonia')
subplot(6,3,6), plot(t_day_c,ammonia(:,15:21)), grid, ylabel('Ammonia')
subplot(6,3,7), plot(t_day_c,  calcium(:,1:7)), grid, ylabel('Calcium')
subplot(6,3,8), plot(t_day_c, calcium(:,8:14)), grid, ylabel('Calcium')
subplot(6,3,9), plot(t_day_c,calcium(:,15:21)), grid, ylabel('Calcium')
subplot(6,3,10), plot(t_day_c,  iron(:,1:7)), grid, ylabel('Iron')
subplot(6,3,11), plot(t_day_c, iron(:,8:14)), grid, ylabel('Iron')
subplot(6,3,12), plot(t_day_c,iron(:,15:21)), grid, ylabel('Iron')
subplot(6,3,13), plot(t_day_c,  methane(:,1:7)), grid, ylabel('Methane')
subplot(6,3,14), plot(t_day_c, methane(:,8:14)), grid, ylabel('Methane')
subplot(6,3,15), plot(t_day_c,methane(:,15:21)), grid, ylabel('Methane')
subplot(6,3,16), plot(t_day_c,  nitrate(:,1:7)), grid, ylabel('Nitrate')
  xlabel('Julian Date')
subplot(6,3,17), plot(t_day_c, nitrate(:,8:14)), grid, ylabel('Nitrate')
  xlabel('Julian Date')
subplot(6,3,18), plot(t_day_c,nitrate(:,15:21)), grid, ylabel('Nitrate')
  xlabel('Julian Date')
% print f11others_a
pause, clg
subplot(6,3,1), plot(t_day_c,  oxygen(:,1:7)), grid, ylabel('Oxygen')
  title('Spill Site')
subplot(6,3,2), plot(t_day_c, oxygen(:,8:14)), grid, ylabel('Oxygen')
  title('9.1 m Site')
subplot(6,3,3), plot(t_day_c,oxygen(:,15:21)), grid, ylabel('Oxygen')
  title('21.3 m Site')
subplot(6,3,4), plot(t_day_c,  ph(:,1:7)), grid, ylabel('pH')
subplot(6,3,5), plot(t_day_c, ph(:,8:14)), grid, ylabel('pH')
subplot(6,3,6), plot(t_day_c,ph(:,15:21)), grid, ylabel('pH')
subplot(6,3,7), plot(t_day_c,  phosphates(:,1:7)), grid, ylabel('Phosphates')
subplot(6,3,8), plot(t_day_c, phosphates(:,8:14)), grid, ylabel('Phosphates')
subplot(6,3,9), plot(t_day_c,phosphates(:,15:21)), grid, ylabel('Phosphates')
subplot(6,3,10), plot(t_day_c,  sulfate(:,1:7)), grid, ylabel('Sulfate')
subplot(6,3,11), plot(t_day_c, sulfate(:,8:14)), grid, ylabel('Sulfate')
subplot(6,3,12), plot(t_day_c,sulfate(:,15:21)), grid, ylabel('Sulfate')
subplot(6,3,13), plot(t_day_c,  temperature(:,1:7)), grid, ylabel('Temperature')
subplot(6,3,14), plot(t_day_c, temperature(:,8:14)), grid, ylabel('Temperature')
subplot(6,3,15), plot(t_day_c,temperature(:,15:21)), grid, ylabel('Temperature')
subplot(6,3,16), plot(t_day_c,  toc(:,1:7)), grid, ylabel('TOC')
  xlabel('Julian Date')
subplot(6,3,17), plot(t_day_c, toc(:,8:14)), grid, ylabel('TOC')
  xlabel('Julian Date')
subplot(6,3,18), plot(t_day_c,toc(:,15:21)), grid, ylabel('TOC')
  xlabel('Julian Date')
% print f11others_b
pause, clg
% ******************************************E N D - P R O G R A M.
```

# APPENDIX E

## A Numerical Example of Mean-Centering and Variance-Normalizing A Matrix

**X** is a twenty observations by five variables matrix used in an example in Chapter 3.

X(original) =

| | | | | |
|---|---|---|---|---|
| 21.2478 | 21.2203 | 68.3070 | 68.4047 | 68.4475 |
| 20.9505 | 20.9424 | 67.1023 | 67.1614 | 67.2061 |
| 21.0838 | 21.0877 | 67.0102 | 67.0807 | 67.1542 |
| 21.0017 | 20.9886 | 66.8053 | 66.7759 | 66.8470 |
| 21.3257 | 21.3487 | 68.1280 | 68.0564 | 68.0617 |
| 20.8305 | 20.8414 | 67.0495 | 67.0473 | 67.0041 |
| 20.5935 | 20.6069 | 67.4215 | 67.3963 | 67.3201 |
| 21.1892 | 21.1710 | 67.7409 | 67.7308 | 67.6564 |
| 21.2860 | 21.2814 | 67.1600 | 67.2004 | 67.0927 |
| 21.4339 | 21.4460 | 68.4067 | 68.2076 | 68.3024 |
| 20.9805 | 20.9766 | 66.1727 | 66.2039 | 66.2824 |
| 21.1193 | 21.1119 | 66.8294 | 66.9101 | 66.8511 |
| 20.9464 | 20.9193 | 67.8495 | 67.7752 | 67.7581 |
| 21.0111 | 21.0189 | 66.4389 | 66.5081 | 66.3005 |
| 21.2055 | 21.2276 | 67.7122 | 67.8816 | 67.7857 |
| 20.8868 | 20.8812 | 68.3018 | 68.0739 | 67.9410 |
| 21.2218 | 21.2311 | 67.6623 | 67.6505 | 67.3860 |
| 21.3312 | 21.3335 | 68.1199 | 68.2778 | 68.1327 |
| 21.0904 | 21.0878 | 67.4803 | 67.4489 | 67.4482 |
| 21.2946 | 21.3128 | 69.4045 | 69.2535 | 69.3350 |

X(mean-centered) =

| | | | | |
|---|---|---|---|---|
| 0.1463 | 0.1185 | 0.7519 | 0.8525 | 0.9318 |
| -0.1510 | -0.1594 | -0.4528 | -0.3909 | -0.3095 |
| -0.0177 | -0.0141 | -0.5450 | -0.4715 | -0.3615 |
| -0.0998 | -0.1132 | -0.7498 | -0.7763 | -0.6686 |
| 0.2242 | 0.2469 | 0.5728 | 0.5041 | 0.5461 |
| -0.2710 | -0.2603 | -0.5056 | -0.5049 | -0.5115 |
| -0.5080 | -0.4948 | -0.1336 | -0.1559 | -0.1956 |
| 0.0876 | 0.0693 | 0.1857 | 0.1785 | 0.1408 |
| 0.1844 | 0.1797 | -0.3952 | -0.3519 | -0.4230 |
| 0.3324 | 0.3442 | 0.8515 | 0.6554 | 0.7868 |
| -0.1210 | -0.1252 | -1.3825 | -1.3484 | -1.2332 |
| 0.0178 | 0.0101 | -0.7257 | -0.6422 | -0.6646 |
| -0.1551 | -0.1825 | 0.2943 | 0.2229 | 0.2425 |
| -0.0904 | -0.0828 | -1.1162 | -1.0442 | -1.2152 |
| 0.1040 | 0.1259 | 0.1570 | 0.3294 | 0.2701 |
| -0.2147 | -0.2205 | 0.7466 | 0.5217 | 0.4254 |
| 0.1203 | 0.1293 | 0.1071 | 0.0982 | -0.1297 |
| 0.2297 | 0.2318 | 0.5648 | 0.7256 | 0.6171 |
| -0.0111 | -0.0140 | -0.0749 | -0.1034 | -0.0675 |
| 0.1931 | 0.2111 | 1.8494 | 1.7012 | 1.8194 |

X(mean-centered and variance-normalized) =

| | | | | |
|---|---|---|---|---|
| 0.0344 | 0.0275 | 0.0470 | 0.0565 | 0.0605 |
| -0.0355 | -0.0370 | -0.0283 | -0.0259 | -0.0201 |
| -0.0042 | -0.0033 | -0.0341 | -0.0312 | -0.0235 |
| -0.0234 | -0.0263 | -0.0469 | -0.0514 | -0.0434 |
| 0.0527 | 0.0573 | 0.0358 | 0.0334 | 0.0355 |
| -0.0637 | -0.0604 | -0.0316 | -0.0335 | -0.0332 |
| -0.1193 | -0.1148 | -0.0084 | -0.0103 | -0.0127 |
| 0.0206 | 0.0161 | 0.0116 | 0.0118 | 0.0091 |
| 0.0433 | 0.0417 | -0.0247 | -0.0233 | -0.0275 |
| 0.0781 | 0.0799 | 0.0532 | 0.0434 | 0.0511 |
| -0.0284 | -0.0290 | -0.0864 | -0.0894 | -0.0801 |
| 0.0042 | 0.0023 | -0.0454 | -0.0426 | -0.0432 |
| -0.0364 | -0.0423 | 0.0184 | 0.0148 | 0.0158 |
| -0.0212 | -0.0192 | -0.0698 | -0.0692 | -0.0789 |
| 0.0244 | 0.0292 | 0.0098 | 0.0218 | 0.0175 |
| -0.0504 | -0.0512 | 0.0467 | 0.0346 | 0.0276 |
| 0.0283 | 0.0300 | 0.0067 | 0.0065 | -0.0084 |
| 0.0540 | 0.0538 | 0.0353 | 0.0481 | 0.0401 |
| -0.0026 | -0.0032 | -0.0047 | -0.0069 | -0.0044 |
| 0.0454 | 0.0490 | 0.1156 | 0.1127 | 0.1182 |

# VITA

Tan, Ting Chow was born on August 19, 1970 in Johor Bahru, Johor, Malaysia. He entered the University of Tennessee, Knoxville, as an undergraduate on January 1989 and received a Bachelor of Science degree in Chemical Engineering with a minor in Business Administration from the same university on May 1992. In August, 1992, he became a graduate student at the University of Tennessee, Knoxville. He received a Master of Science degree in Chemical Engineering from the university on August 1994.