PARAMETER ESTIMATION AND UNCERTAINTY
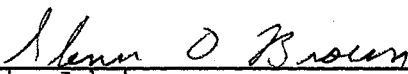
FOR VOLATILE ORGANIC TRANSPORT

By

JIAN YUE

Bachelor of Science
Chengdu Science and Technology
University
Chengdu, P.R. China
1984

Master of Science
Beijing Graduate College
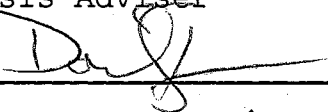Beijing, P.R. China
1987

Submitted to the faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirement for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1994

PARAMETER ESTIMATION AND UNCERTAINTY

FOR VOLATILE ORGANIC TRANSPORT

Thesis Approval:

Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGMENTS

With the assistance of a very special group of people, this endeavor has finally come through. I am deeply indebted to Dr. Glenn Brown, my major advisor, whose intelligent guidance, continuous encouragement, and generous financial assistance were invaluable in the accomplishment of this dissertation. Dr. Brown has been a true mentor throughout my doctoral program, with understanding and patience. It was his advice, trust, and encouragement that pulled me through those valleys of *The Objective Function*. To him, I offer my heartfelt thanks.

I am grateful to Dr. Tom Haan who graciously agreed to chair my committee. His expertise on stochastics has been invaluable in the progress of this dissertation. For all his help, I am very appreciative. I wish to extend my appreciation also to Dr. Dan Storm, Dr. Nick Basta, and Dr. Jack Stone for serving on my committee. Additionally, special thanks are also given to Ross Fenske for spending so many hours with me on the editing of the dissertation, and to Ming Yu for her hard work in the lab to make data available for this study. Finally, I wish to acknowledge Arthur Baehr of the U.S.G.S. who supplied this model for my use.

A special note of thanks is given to my personal friends:

TABLE OF CONTENTS

vi

LIST OF TABLES

LIST OF FIGURES

Figure   Page

viii

CHAPTER I

INTRODUCTION

Problem Statement

Ground water is the most abundant supply of fresh water lying beneath the Earth's land surface. In the United States, groundwater supplies about 25 percent of the nation's domestic, agricultural, and industrial water, with about half of all U.S. homes depending on it for drinking water supply (Moseley, 1988). Moreover, as part of the hydrologic cycle, it serves as a reservoir to rivers, streams, and lakes during dry seasons.

Unfortunately, ground water has been negatively affected by human activities. Chemicals from pesticides, fertilizers, sewage waste, and hazardous wastes from industrial dumping sites are finding their way into more and more aquifers. Likewise, through the natural discharges of an aquifer, such as springs and seeps, ground water contaminants can return to pollute surface water. Therefore, in a sense, pollution of ground water is pollution of water everywhere.

In many years people thought that ground water was of adequate quality for drinking due to the natural process of

filtration where soil and rocks filter out contaminants. It wasn't until Rachel Carson (1962) portrayed the results of environmental contamination by pesticides, in her book *Silent Spring*, that ground water pollution was brought into public view. In fact, ground water contamination has become an environmental issue as increasing incidents of ground water contamination were reported during the past two decades. For example, wide-spread aquifer contamination by overuse of a pesticide was found in the potato-growing region of Long Island, New York (Carcel, 1987).

Researchers have labored to develop mathematical models to predict the mobility and persistence of chemicals in the unsaturated zone of the soil profile. More recently, the presence of gaseous volatile organic compounds in the unsaturated zone has caused new concern in ground water research. For example, in the United States, the EPA has estimated that five to 15 percent of underground storage tanks that hold approximately 14 billion gallons of gasoline are leaking (Moseley, 1988). The leaked gasoline can pose a serious threat to health and safety. The common chemicals found in gasoline, such as benzene, toluene, and lead, are toxic. Benzene is a proven human carcinogen, toluene is one of the toxic chemicals regulated by the Clean Air Act 1990, and lead causes disorders of the central nervous system. To clean up all the leaks nationwide, the EPA estimates that it will cost $7.5 billion, excluding possible health or property

damage awards due to law suits (Moseley, 1988).

Gasoline and other volatile organics, such as crude oil, are used in abundance all over the world in vehicles and industry. In connection with ground water, the problem is not in the use of these organics, but rather in the possible disasters that can occur when storing and transporting them. Because they are handled in such large amounts, when a disaster happens, the ecological effect on fresh water can be catastrophic.

The most challenging problem in dealing with ground water contamination is to predict the movement and fate of the contaminants. In order to clean up when the pollution occurs, one needs to locate the high concentration area needing remediation, and to discern the pollutants' migratory path and extent. The ability to predict the movement of contaminants can also contribute to the prevention of pollution by forecasting possible disasters.

Even though gasoline and other volatile organic contaminants are immiscible in both air and water, they dissolve into the water phase within the saturated zone and also volatilize into the soil gas phase within the unsaturated zone. Baehr (1987) used a mathematical model of vapor and solute transport in the unsaturated zone to show that significant amounts of gasoline hydrocarbons can partition into the water in the unsaturated zone. They can percolate into the ground and leave zones of the residual contaminant in

3

the pore structure. Such residuals form a long-lasting source of contaminants that will migrate into the vapor phase and eventually dissolve into the ground water (Mendoza and Frind, 1990).

Numerical models have contributed to the understanding of multiphase transport of volatile organics in the unsaturated zone. However, since it is very difficult to observe and measure most transport processes, several problems are still ahead for researchers to explore. Jury et al. (1987) pointed to the lack of detailed and accurate experimental data as factors limiting further development of computational models. Brown (1991) summarized some main aspects that remain unaddressed in existing research, namely: (1) the small amounts of experimental data available to verify the theory, (2) difficulties of independent parameter measurement, and (3) the limitations of chemical linear adsorption and phase equilibrium assumptions.

Model usefulness depends on the accuracy of determining model parameters and the validation of model performance. On one hand, the accuracy of model predictions should ideally depends on the accuracy of model parameters. Unfortunately, most model parameters are not measurable. Some investigators have used inverse techniques in solute transport studies, however the application of these methods to volatile organic transport modeling is still new. On the other hand, any environmental fate and transport model must be tested for

4

predictive performance before it can generally be useful as an aid to regulatory decision making. Thus the purpose of assessing model performance is to quantify the uncertainty of model predictions, which results from two primary sources: uncertainty in the model structure and uncertainty in the model parameters.

The assessment of model performance from parameter uncertainty and experimental data verification for volatile organic transport has received little attention. Few articles about the evaluation methodology of general solute transport models have been found in the literature. According to Loague (1990), well-defined procedures for testing models are not yet available, despite tremendous model development efforts. Pennell (1990) also mentioned that an accepted method or systematic approach for the validation of pesticide simulation models does not exist.

Study Objectives

Up until now, the attempts at solving inverse problems and quantifying uncertainty in model predictions for volatile organic transport has not been seen in published works. Therefore, the objective of this research is to develop a systematic methodology which combines well established methods to pursue the following two goals:

1. Estimate parameters for volatile organic transport.
2. Quantify uncertainty in model predictions due to

uncertainty in these parameter estimates.

General Procedure

A compositional multiphase model developed by Baehr (1987) was selected for this study. This model describes the multiphase transport of petroleum products composed of different components. The numerical solution of this model assumes that only diffusive transport is a significant transport mechanism for gas and immiscible phases.

Although there are many parameters involved in this model, my interest is focused on the following four parameters: (1) the partition coefficient between the gas and the aqueous phases, (2) the partition coefficient between the aqueous and solid phases, (3) the tortuosity in the gas phase, and (4) the tortuosity in the aqueous phase.

Bayesian statistical theory was used to estimate the optimal values and distributions of these parameters. Monte Carlo simulation was employed to determine the uncertainty of model output.

This dissertation is organized as follows: Chapter 2 reviews the related literature; Chapter 3 describes the volatile organic transport model employed in this research; Chapter 4 contains estimation and uncertainty theories and develops a methodology to use these theories; results and analyses are presented in Chapters 5 and 6, and finally, Chapter 7 summarizes the results and presents the conclusions.

CHAPTER II

LITERATURE REVIEW

Volatile organic transport modeling involves basic principles of hydrogeology, chemistry, soil science, and mathematics. The results achieved by investigators from these and other branches of science are extensive and scattered. Thus it is necessary to narrowly focus this review on topics of immediate interest. This research relates to three major subjects: modeling of volatile solute transport, the estimation of model parameters, and the uncertainty analysis of estimated parameters. Accordingly, this chapter summarizes present knowledge and identifies problems for future research within these topics.

Modeling of Volatile Solute Transport

Modeling of volatile solute transport combines mathematical representations of several transport processes to predict the behavior and fate of contaminants. This section reviews the transport processes of volatile organics, the governing equations, and specific applications.

The Mechanisms of Vapor Phase Organic Transport

The transport of contaminants in the vapor phase may occur due to both advection and diffusion and is influenced by phase partitioning and degradation (Falta et al., 1989). Vapor phase advection may result from vapor density gradients or vapor pressure gradients. In the published literature, there is not a clear understanding of which condition favors each driving force (Mendoza and Frind, 1990). Density-driven flow may be of special concern for many organic chemicals found at contaminated sites. When organic liquids evaporate, the density of the vapor in contact with the liquid changes with respect to the ambient soil gas. Diffusion is of greater significance in the gas phase than in the aqueous phase, because gas-phase diffusion coefficients are much larger than aqueous-diffusion coefficients. Both dispersive and diffusive transport may contribute to total flux for the gas flow system. However, experimental data showed that diffusion predominates over mechanical dispersion for gas-phase transport unless velocities are very high (Brusseau, 1991). Usually, diffusive flux in the gas phase, J, is modeled by Fick's First Law (Baehr, 1990):

$$J = -c\omega D^b \theta_g \xi_g \frac{dX}{dz} \qquad (1)$$

where c is the vapor-phase molar density ($c = 4.46 \times 10^{-5}$ mol/cm^3 for an ideal gas at standard temperature and pressure), ω is

the molecular weight of the constituent, D^b is the diffusion coefficient of the organic constituent in air, θ_g is air filled porosity, X is the mole fraction of the constituent in the vapor phase, Z is vertical distance, and ζ_g is Fick's First Law estimate of tortuosity. Tortuosity represents the internal geometry of the porous media, and may be a function of θ_g.

Recent research has shown that Fick's First Law is not appropriate under some circumstances of interest. According to Thorstenson and Pollock (1989), the accuracy of Fick's First Law depends primarily on the relative magnitudes of the viscous and diffusive flux components and Fick's First Law is not adequate to deal with stagnant gases. Baehr and Bruell (1990) also pointed out that equations based on Fick's First Law are not appropriate for some systems, for example, systems where the concentration in the gas phase is not dilute or where significant evaporative fluxes occur. They recommended the Stefan-Maxwell equations to provide a more comprehensive model for quantifying steady-state transport when a vapor phase is composed of arbitrary proportions of its constituents.

When a volatile liquid organic chemical is spilled on the soil or leaks from a tank into the soil, it will begin to partition into the liquid and vapor phases, and then become dissolved in soil moisture and adsorbed onto the surfaces of soil minerals and organic matter (Silka, 1988). Accurate

description of contaminant transport in a subsurface multiphase system requires the interphase partitioning of individual chemical components among all phases present. Even though such partitioning between phases is complicated, the common assumption that an equilibrium condition exists among the phases present can simplify the problem.

Raoult's Law is usually employed to quantify the equilibrium between the nonaqueous liquid and vapor phases (Corapcioglu and Baehr, 1987),

$$H_{gi} = \frac{\omega}{RT} p^*$$ (2)

$$C_g = H_{gi} \chi_i \gamma_i$$ (3)

where subscript g represents the gas phase, subscript i represents the nonaqueous phase, p^* is the vapor pressure over the pure constituent, ω is the molecular weight of the chemical, R is the universal gas constant, T is temperature; H_{gi} is an equilibrium partition coefficient, C_g is the concentration of the chemical in the gas-phase, χ_i is the mole fraction of the chemical in the nonaqueous phase, and γ_i is the activity coefficient for the chemical which adjusts for nonideality.

This analysis will be limited to a single volatile solute that does not exist in a separate phase. Under these restrictions, Henry's Law is applied to express the equilibrium between the air and water phases as:

10

$$K_h = \frac{C_g}{C_l} \qquad \qquad (4)$$

where K_h is Henry's Law constant at a specified temperature, C_g is the concentration of the vapor phase, and C_l is the concentration of the aqueous phase.

Besides the partitioning between the vapor and aqueous phases, soil organic and mineral materials can adsorb some volatile organics to a lesser extent. Usually, soil solids are surrounded by water layers of at least several molecules thickness. The process of partitioning between the vapor phase and the solid phase then becomes a two-step process from the vapor into the water and subsequently from the water onto the soil solids (Silka, 1988). At equilibrium, the degree of partitioning between the soil solids and the soil moisture is expressed as:

$$K_d = \frac{C_s}{C_l} \qquad \qquad (5)$$

where K_d is the partition coefficient or distribution coefficient, C_s is the mass of chemical adsorbed per unit dry mass of soil solids, and C_l is the concentration of the chemical in the soil moisture.

The Governing Equation

A governing equation is the mathematical representation or model which combines various transport processes through

the principle of mass conservation to describe the actual transport system. Brown and McWhorter (1990) derived a one-dimensional multiple phase transport equation which is general in nature and can be applied to any one-dimensional volatile solute transport case in a homogeneous porous media. The equation is formed as:

$$C_t = R_s C_s = R_l C_l = R_g C_g \tag{6}$$

$$R\frac{\partial C_t}{\partial t} = \frac{\partial}{\partial x}\left(\left(\frac{D_l}{R_l}+\frac{D_g}{R_g}\right)\frac{\partial C_t}{\partial x}\right) - \left(\frac{q_l}{R_l}+\frac{q_g}{R_g}\right)\frac{\partial C_t}{\partial x} + \left(\frac{\rho_g}{\rho_l}\left(\frac{1}{R_l}-\frac{1}{R_g}\right)\frac{\partial}{\partial x}(q_g+q_l)\right)C_t \tag{7}$$

where $R = V/R_s + \theta_l/R_l + \theta_g/R_g$, θ_l is the volumetric solution content, V is the solid phase volume, θ_g is the gas phase volume, C_t is the total solute concentration, C_s, C_l, and C_g are the solute concentrations in the three phases, R_s, R_l, and R_g are relationships that equate the individual concentrations to the total, ρ_l and ρ_g are densities of solution and gas, and q_l and q_g are volume fluxes of solution and gas phases.

Application and Unaddressed Problems

Many mathematical models can simulate diffusive transport in the vapor phase. Jury et al. (1984) introduced a screening model for describing pesticide volatilization, leaching, and degradation in the soil. They performed tests on 35 chemicals to determine the diffusive mobility and general persistence in the soil. Corapcioglu and Baehr (1987) developed a

12

compositional multiphase model to describe the fate of hydrocarbon constituents of petroleum products introduced to soils as an immiscible liquid. Their results showed that diffusive transport in the unsaturated zone is a significant transport mechanism in ground water. Silka (1988) presented a two-dimensional diffusive transport model. He computed an effective diffusion coefficient that incorporates the effects of tortuosity, moisture content, and the organic carbon content of soil.

Several investigators also addressed the importance of including density-driven advection as a transport mechanism in the unsaturated zone. Falta et al. (1989) suggested that significant advective gas flow will result from the evaporation of volatile liquids in soils giving a high permeability. Mendoza and McAlary (1990) studied the potential effects of density-driven vapor advection in the saturated zone. Their results showed that advection becomes increasingly important as the soil permeability increases. However, diffusion still dominates near the periphery of the vapor plume because the density gradient diminishes at lower concentration. They concluded that neglecting advection may underestimate the rate of vapor transport. They also suggested that experimental data are necessary to validate the theoretical basis for the density-driven advection of vapors in further research. In another paper, Mendoza and Frind (1990) used breakthrough curves obtained experimentally to

13

determine whether density gradients are likely to play a role. Their results showed that the advective mechanism can be highly effective in mobilizing organic vapors and thus in accelerating the contamination of a ground water system.

Gierke et al. (1990) presented a comprehensive model that considered air and water advection and dispersion in the direction of flow, mass transfer resistance at the air-water interfaces, partitioning between the air-water phase, and sorption to soil organic matter from aqueous solution. The validation compared the breakthrough curves obtained from model simulations and experiments. Their results indicated that both liquid dispersion and diffusion in immobile water are important. Vapor diffusion is not important when the average pore water velocities are greater than 0.02 cm/s. The rates of mass transfer across the air-water and the mobile-immobile water interfaces are rapid.

All the studies discussed above assumed homogeneous media and phase partitioning equilibrium. However, many studies have shown that the assumption of homogeneity of porous media does not reflect reality. Observations of pesticide residues in agricultural fields long after application also support the invalidity of the local equilibrium assumption about sorption (Brusseau, 1991).

Models that incorporate heterogeneity and nonequilibriate mass transfer are beginning to appear in the literature. Abriola and Pinder (1985) developed a multiphase approach to

describe the simultaneous transport of a chemical contaminant in three physical forms: as a nonaqueous phase, as a soluble component of an aqueous phase, and as a mobile fraction of a gaseous phase. They used a heterogeneous porous medium and incorporated diffusion, dispersion, and interphase mass exchange into the transport model. Interphase mass exchange is considered as local equilibrium.

Brusseau (1991) presented a model that incorporates the effects of physical heterogeneity and rate-limited sorption on gas-phase advection and dispersion. He assessed performance of the model by comparing simulations to data obtained from the literature. However, he simulated the model only under steady flow conditions. Sleep and Sykes (1989) accounted for nonequilibrium conditions with respect to interphase mass transfer, but they considered the saturated zone only.

Although research continues and much work has been reported in the related literature, there exist many important areas that need further research. Brown (1991) summarized the existing problems as:

1) There are minor amounts of experimental data available to verify the theory. As stated by Jury et al. (1987), "further development of such computational models is limited by a dearth of detailed and accurate experimental data."

2) The distributed parameters make independent measurement of the parameters difficult.

3) The assumptions of phase equilibrium are not well justified.

Estimation of Model Parameters

Improvements in the precision of model predictions depend on the ability to accurately determine the model parameters. For ground water flow and solute transport models, most parameters are distributed. The response of the system is governed by a partial differential equation, and parameters embedded in the equation are spatially dependent (Yeh, 1986). Unfortunately, these parameters are seldom measurable. Even if one can measure some of them in the laboratory, there still exists no well-defined correlation between laboratory and field values. To deal with this problem, investigators since the mid-70's have developed various parameter estimation techniques that optimize parameters from observations of dependent variables along with initial and boundary conditions. The following sections review the relevant literature dealing with the parameter estimation, along with specific applications to solute transport in the vadose zone.

General Parameter Estimation Procedures

The most popular statistical methods applied are least squares, maximum likelihood, and Bayesian theory. The following sections discuss the theory, application, evolution, and the evaluation of these three methods.

Least Squares Method   Originally, the least squares

method was developed to estimate regression coefficients. It also provides the basic idea and basic optimal objective function in a mathematical form (Yan, 1990). The idea is to estimate parameters by minimizing the sum of error squares between model output and observed output responses as:

$$Min\sum_{n=1}^{N} [Y_n - \eta(X_n, P)]^2 \qquad (8)$$

where Y_n is observed output, η is model output, X_n is model input, N is the number of observations, and P is a parameter vector.

This simple form is called the ordinary least squares method (OLS). Kool and Parker (1988) stated that the OLS formulation has probably been the most popular for parameter estimation problems due to its simplicity and the minimum amount of information required. However, it requires assumptions of uncorrelated errors and constant variance to provide unbiased and minimum variance estimate.

When the assumptions of the constant variance and uncorrelated errors are violated, weighted least squares (WLS), also known as generalized least squares (GLS), can be used to satisfy the assumptions. The WLS can be formulated as:

$$Min\sum_{n=1}^{N} [Y_n - \eta(X_n, P)]^T W(Y_n - \eta(X_n, P)) \qquad (9)$$

where W is a symmetric weighting matrix that corrects for

17

unequal error variances.

Jacquard et al. (1965) reported the first application of least squares to the inverse problem. They divided a petroleum reservoir into zones of constant permeability and used a variational method to minimize the sum of squared head residuals. Their method was not sufficient to get a stable and unique solution. Korganoff (1970) improved this method by imposing penalty criteria (based on computed parameters) to residual errors which reduced unwarranted oscillation.

Maximum Likelihood Method The maximum likelihood method considers model parameters as unknown but deterministic. The objective of this method is to find the parameters that maximize the likelihood of obtaining the measured data, given the joint probability density function of all measurements. This method assumes that the errors are normally distributed (Kool and Parker, 1988).

Carrera and Neuman (1986) employed this method to estimate aquifer parameters along with the prior information. They examined the prior errors affecting the solutions and concluded that, after transforming some parameters logarithmically, the error distributions should not be too far from a normal distribution. Since not all the factors which contributed to the prior errors can be quantified statistically, Carrera and Neuman expressed the covariance matrices of these errors in terms of several parameters which can be estimated jointly with the hydraulic parameters. They

also pointed out that the maximum likelihood concept might be useful for selecting the best ground water models.

Bayesian Estimates Bayesian estimates that incorporate prior information were first applied to aquifer parameter estimation by Gavalas et al. (1976). This method considered model parameters as random variables with a defined probability distribution and introduces a statistically based smoothing criterion.

In Bayesian theory, parameter vector P is a random variable with probability distribution $f(P)$, and the probability distribution of observation Y depending on P can be expressed as:

$$f(Y,P) = f(Y/P) f(P) = f(P/Y) f(Y) \qquad (10)$$

and

$$f(P/Y) = \frac{f(Y/P) f(p)}{f(Y)} \qquad (11)$$

where $f(P)$ is the prior probability distribution of parameters, and $f(P/Y)$ is the posterior distribution of P given Y. The parameters are estimated by maximizing $f(P/Y)$.

The prior information required for Bayesian estimation includes the mean and covariance matrix of the parameters. Gavalas et al. (1976) have shown that Bayesian estimation reduces to a quadratic minimization problem, provided the parameters and the measurement errors are normally distributed and the model is linear in parameters.

Critical Evaluation The maximum likelihood method allows more flexible assumptions than least squares and produces smaller parameter variances when errors have non-constant variances and are correlated (Yan, 1990). The Bayesian estimate is a convenient method to evaluate parameter uncertainty which also considers the penalties that can arise in the action space due to incorrect specification of the unknown parameter value (Yan, 1990). However, the Bayesian estimation can be impractical if the required prior information is not reliable because inaccurate prior information can make the estimates worse instead of reducing uncertainty (Yeh, 1986). Box and Tiao (1973), though, did provide a noninformative form of Bayesian estimates to avoid the negative effect of inaccurate prior information. This method considers a prior distribution as a uniform distribution which reflects minimal knowledge of a parameter, thus relies more on observed data.

Ill-Posedness in Parameter Estimation

Carrera and Neuman (1986) defined ill-posedness as a functional relationship, $h(x,t)=F(p_i(x))$, between a set of spatially varying parameters. The problem is properly posed if and only if the following three conditions are satisfied: (1) to every $h(x,t)$ there corresponds a solution, $p_i(x)$, (2) the solution is unique for any given $h(x,t)$, and (3) the solution depends continuously on $h(x,t)$ (the solution is

stable). An inverse problem is ill-posed if it fails to satisfy one or more of these three requirements.

As a misbehavior of the inverse solution, the cause of ill-posedness is not always well understood. However, researchers characterize it as nonidentifiability, nonuniqueness, and instability. According to Yeh (1986), identifiability addresses the question of whether it is possible to obtain unique solutions of the inverse problem, which means if different parameter sets can lead to a given output, the parameters are unidentifiable. Carrera and Neuman (1986) pointed out that identifiability refers to the forward relationship and uniqueness refers to the inverse relationship or minimization process. Stability means that small errors in the observed data must not result in large changes in the computed parameters. Therefore, instability manifests itself as spatially oscillating parameters.

Kool and Parker (1986) pointed out that correlation among parameters often causes ill-posedness. This is especially true when parameters are negatively correlated because a change in one parameter will balance a corresponding change in the correlated parameter and can lead to the same model prediction.

He also pinpointed observed data as a cause of ill-posedness in two different ways. On one hand, insufficient experimental data may cause an objective function insensitive to one or more of the parameters. This might also result in

large estimation variance for parameters. On the other hand, if the estimated parameters are too sensitive to observed data, instability will occur because small measurement errors can cause significant errors in parameter estimates.

Up until now, there are no definite ways to solve ill-posedness problems because solutions vary with different circumstances and some problems might even be unsolvable. However, when ill-posedness occurs, checking model structure, observed data, and dimensionality of parameter space can better help solve problems. Model structure should be checked because the nonlinearity in a model and the insensitivity of model predictions to model parameters could lead to ill-posedness. Observed data should be checked for its sufficiency and accuracy, and prior information about parameters can help solve the problem. High dimensionality of parameter space always complicates problems because of more interactions among parameters. Therefore, parameters that do not have much impact on model predictions should be excluded to reduce the dimensionality.

Application to Solute Transport

Solute transport is such a complex phenomenon that the relevant transport processes are hard to identify. As a result, there are few examples of inverse models in the literature involving solute transport (Keidser and Rosbjerg, 1991). Volatile organic transport is even more complex due to

the existence of multiple phases, thus examples dealing with parameter estimation of this particular interest have not been seen in publications. The following refers to single phase solute transport only.

Murty and Scott (1977) estimated the dispersion coefficient from observed data for solute concentrations. Their results showed that the accuracy of parameter estimation depends on both the accuracy of the solution to the transport model and the measurements of the concentration values. Umari et al. (1979) also estimated the dispersion coefficient from observations in the field. They used a general nonlinear program to minimize the discrepancy between calculated and observed values of the concentration profile. Kool and Parker (1988) reviewed the status of parameter estimation techniques and their utility for determining key parameters affecting water flow and solute transport in the vadose zone. They pointed out that efforts are needed to extend parameter estimation methods to more complex field conditions. These conditions usually require models that can fit soil heterogeneity, variable and uncertain boundary conditions, simultaneous flow and transport, and complex biochemical processes and other phenomena.

Jury and Sposito (1985) used least squares, maximum likelihood, and the method of moments estimation procedures for field-scale validation. They found that these three procedures gave different parameter estimates for a given set

of data.

Wagner and Gorelick (1986) combined a contaminant transport simulation with a weighted least squares procedure to estimate parameters that characterize the transport of contaminants. They showed the importance of using Monte Carlo analysis to quantify the reliability of parameter estimates. They concluded that nonlinear regression technique can provide accurate and reliable estimates of the nonlinear parameters when large random errors are present in the data.

Knopman and Voss (1989) developed a multiobjective sampling design that addressed model discrimination, parameter estimation, and cost of field sampling. They estimated parameters by minimizing some measure related to variance and covariance of parameters. They also indicated that sensitivity of solute concentrations to a change in a parameter contributes information to the relative variance of a parameter estimate.

Since contaminants are primarily transported as dissolved components in the water phase, modeling of contaminant transport and fluid flow is strongly coupled. Thus, optimizing both flow and transport parameters simultaneously has recently received more attention. Strecker and Chu (1986) first estimated both flow and transport parameters in a two-stage approach. In the first stage, they estimated transmissivity controlling the flow process. In the second stage, they estimated dispersivity representing the solute

transport process. Keidser and Rosbjerg (1991) modified the two-stage approach. In the first stage, the transmissivity field was estimated using both head and concentration data. Transferring the estimated transmissivity field to the second stage, the transport parameters were optimized based on the concentration measurements. The second stage repeated the stage one estimation to adjust the transmissivity parameters using the optimized parameters.

Mishra and Parker (1989) used a combined simulation-optimization method to deal with the estimation of soil hydraulic and transport parameters from transient unsaturated flow. They used a nonlinear weighted least squares algorithm to estimate unknown model parameters by minimizing deviations between concentrations, water content, and pressure heads obtained from hypothetical experiments. They found that simultaneous estimation of hydraulic and transport properties yields smaller estimation errors for model parameters than a stage-wise method.

Uncertainty Analysis of Model Predictions

Uncertainty analysis procedures quantify the range or the probability distribution of model predictions. Considering parameters involved in a solute transport model random variables makes model predictions random variables as well. According to Haan (1977), a variable that is a function of other random variables is also a random variable. The

probability that a random variable equals a fixed value is zero. Therefore, a model prediction is meaningless unless its uncertainty represented by a range or a probability distribution is quantified.

Uncertainty in model predictions results from natural uncertainty, inadequacy in the model structure, and errors in the model parameters. To date, the evaluation of uncertainty in model performance has focused more on parameter uncertainty than on the other sources. The following sections review the relevant literature dealing with uncertainty analysis, along with a critical evaluation.

General Uncertainty Analysis Procedures

The current approaches to uncertainty analysis include deterministic, simulation, and nonparametric approaches. Each of these methods are summarized here in a general sense.

<u>Deterministic Approach</u>   The deterministic approach is an analytic method based on a Taylor series expansion about a fixed point, usually the mean of the input variables. Only first or second-order terms of the Taylor series are typically used (Doctor, 1989).

Considering a univariant random function $Y=f(X)$, the Taylor series expansion of Y can be written as:

$$Y = f(\mu_x) + \frac{\partial f}{\partial X}(X-\mu_x) + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(X-\mu_x)^2 + \dots \qquad (12)$$

where μ_x is the mean of the variable X.

Neglecting the terms that include second-order or higher partial derivatives, the first-order method estimates the mean and variance of Y by:

$$E(Y) \approx f(\mu_x) \qquad \qquad \text{(13)}$$

and

$$Var(Y) \approx (\frac{\partial f}{\partial X})^2 (Var(X))^2. \qquad \qquad \text{(14)}$$

In a multivariate case, first-order analysis can be stated as:

$$Y \approx f(M_x) + (X-M_x) b^T \qquad \qquad \text{(15)}$$

where M_x is a vector of means, and b^T is the transposition of a vector of partial derivatives (Zhang, 1990). The variance of this estimate can be written as:

$$Var(Y) \approx b^T C_x b \qquad \qquad \text{(16)}$$

where C_x is the covariance matrix of the functionally dependent variables X. The second-order method is similar to the first-order method, but the former is more accurate since the mean of Y is conditioned on the mean and variance of X.

Several researchers have employed the first and second-order analysis methods based on Taylor series expansions in hydrologic research. Mishra and Parker (1988) applied first-order analysis to assess the reliability of unsaturated flow model predictions subject to parameter uncertainty. Andrews et al. (1987) also used the first-order analysis to evaluate the uncertainty of ground water travel time.

Simulation Approach The simulation approach is often called the Monte Carlo method. This method uses random or pseudorandom numbers for solution of a model. For uncertainty analysis, this method requires a known probability distribution for each model parameter so that pseudorandom samples can be generated from the distribution. This method then runs the model at a large number of points in the input parameter space, and produces a probability density function (pdf) for the output variable. Because of its simplicity and the ability to deal with complex systems, the Monte Carlo method has been commonly used in many fields.

Nonparametric Approach   A nonparametric approach, known as the "bootstrap", is often used to estimate the reliability of model prediction. The bootstrap constructs an empirical distribution output by resampling a set of N independent observations rather than makes prior assumptions about the shape of the output distribution. Such assumption has been one of the limiting factors for statistical theory (Zhang, 1990). Willmott et al. (1985) described the application of the bootstrap in calculating the reliability of model prediction. Suppose the N observations $(X_1, X_2, \ldots\ldots, X_N)$ are from a distribution D. A bootstrap sample (D^*) of size N is randomly chosen one element at a time from D with replacement. Once a D^* has been selected, a bootstrap measure of the accuracy of model prediction may be calculated. If this process is repeated B times, it yields an empirically derived

frequency distribution that approaches the true distribution as B becomes large. The standard error of the mean is given by:

$$\sigma = [\frac{1}{B-1} \sum_{i=1}^{B} (X_i - \mu)^2]^{1/2}. \tag{17}$$

Critical Evaluation

The advantage of the first and second-order methods is that they are simple to use. They evaluate only the model and the partial derivatives at the mean value of the input variables. The disadvantage is that they are applicable only to some simple models assuming that the uncertainty in the model output can be completely described by a mean and variance.

There are some apparent drawbacks associated with the simulation method. For example, it requires intensive computation, assumes complete representation of the population distribution by the available sample, and becomes complex when the input variables are dependent. However, the simulation method is simple to use and is powerful for dealing with complex models. Its intensive computational requirement is becoming less important as computers are becoming more powerful and faster. Its concern about assumed probability distributions can be solved by incorporating stochastic parameter estimation procedures.

According to Willmott et al. (1985), the bootstrap method

has two advantages over its parametric counterparts: (1) assumptions about the underlying but unknown frequency distribution of output do not affect the method's validity; (2) confidence can readily be established for any accuracy measure of interest even if its distributional characteristics previously have not been derived and cataloged. However, the bootstrap is limited when few sample observations are available because it assumes the observed data represent the true population.

Recommendation

This review demonstrates the principle of basic volatile solute transport processes, the development of solute in vapor phase transport modeling, the application of various techniques in parameter estimation, and ways to quantify errors by uncertainty analysis. Accordingly, the growth and challenge in the field of ground water pollution control are also illustrated. There exist many important areas that need further research to reliably predict the behavior of volatile contaminants in the subsurface system.

As this review demonstrates, the volatile solute transport processes are so complicated that the scientific understanding of the importance of various processes is still inadequate. Moreover, even if a perfect model can be established to reliably describe the behavior of contaminants in the subsurface, accuracy of the model prediction still

depends on the accuracy of model parameters.

Existing parameter estimation procedures have their own advantages and drawbacks. The application in solute transport shows that investigators have the tendency to use the simplest method, like the least squares method. Unfortunately, the least squares method is not always applicable, since its assumptions about constant variance and normally distributed errors may not always reflect reality. Therefore, application of other parameter estimation procedures need to be investigated.

Estimated parameters always contain errors that will affect the model predictions. Therefore, uncertainty analysis quantifying the uncertainty of estimated parameters is essential for a model to be useful. However, the application of parameter uncertainty analysis in ground water pollution control is quite new, and many more thorough investigations in this field are expected in the future.

Apparently, each of the three areas surveyed in this chapter reveals much room for researchers to explore. This study will develop a systematic methodology investigating parameter estimation and uncertainty analysis. This methodology will employ Bayesian estimates to study parameters involved in a volatile organic transport model developed by Baehr (1987) in depth. It will also use Monte Carlo analysis to quantify uncertainty in model predictions due to errors associated in parameter estimation.

CHAPTER III

VOLATILE ORGANIC TRANSPORT MODEL

This research used a compositional multi-phase-model for volatile organic transport developed by Arthur L. Baehr of the U.S. Geologic Survey (Baehr, 1987). This model is referred to as Baehr's model throughout this dissertation. The following sections describe the model formulation, its application to this research, and the characteristics of parameters of interest.

Model Formulation

Baehr's model is a two-dimensional mathematical model developed to deal with multiphase transport of petroleum contaminants. Petroleum products like gasoline involve different kinds of constituents such as benzene, toluene, and xylene. These hazardous hydrocarbons can be dissolved and can enter an aquifer through the unsaturated zone, where each constituent can either migrate as a solute in the water phase, a vapor in the air phase, and an immobile constituent in the oil phase, or be adsorbed in the solid phase. The total quantity of chemical per unit soil volume can be written as:

$$C_t = \rho_b C_s + \theta_l C_l + \theta_i C_i + \theta_g C_g \qquad (18)$$

where C_s is the adsorbed chemical concentration, C_l is the dissolved chemical concentration, C_i is the immiscible chemical concentration, C_g is the vapor chemical concentration, ρ_b is soil bulk density, and θ_l, θ_g, and θ_i represents volumetric water content, air content, and nonaqueous liquid content respectively.

The governing equation starts with the mass conservation equation:

$$\frac{\partial C_t}{\partial t} + \nabla \cdot J = Sources - Sinks \qquad (19)$$

where J is the total mass flux, the source is the total chemical mass gain (which equals zero for the total system), and the sink is the total chemical mass loss which eventually equals the total rates of molecular transformation due to microbial and abiotic reactions.

The total mass flux is quantified by the advective-dispersive model as:

$$J_i = q_i C_i - D^d_i \nabla C_i \qquad (20)$$

$$J_l = q_l C_l - D^d_l \nabla C_l \qquad (21)$$

$$J_g = q_g C_g - D^d_g \nabla C_g \qquad (22)$$

where q_i, q_l, and q_g are the specific discharge for the oil, water, and air phases, respectively, D_i^d, D_l^d, and D_g^d are the hydrodynamic dispersion tensor for the chemical in each phase.

The hydrodynamic dispersion tensor can be decomposed into functions of physical properties of the porous media, the moving fluid, and the chemical constituent:

$$D^d_f = \delta_f + d_f \theta_f \xi_f \qquad (23)$$

where δ_f is the mechanical dispersion coefficient, d_f is the molecular diffusion constant, ξ_f is the tortuosity of each phase, and f represents different phases.

When combining these transport processes together, the compositional multi-phase diffusive model is defined as:

$$\frac{\partial}{\partial t} [C_g \theta_g + C_l \theta_l + C_s \rho_b + C_i \theta_i] + \nabla \cdot [J_l + J_g + J_i] = -R_{bio} \qquad (24)$$

where R_{bio} denotes the total rate of microbial and abiotic degradation.

In its application, Baehr's (1987) numerical model assumed that the immiscible phase is at residual saturation and neglected C_i and J_i. The model also assumed that the air phase is at atmospheric pressure and the porous media is homogeneous, isotropic, and isothermal.

Baehr's model employed equilibrium approximations to

partition among the air, water, and adsorbed phases. The partitioning between the air and water phase is modeled by the approximation to Henry's law:

$$C_g = K_h C_l \qquad (25)$$

where K_h is the air/water partition coefficient of the chemical. This equilibrium relationship provides a K_h that is independent of the porous media.

The partitioning between the aqueous phase and solid phase is modeled by the linear isotherm,

$$C_s = K_d C_l \qquad (26)$$

where K_d is an adsorption isotherm constant.

In conjunction with Henry's Law, the following relationship is obtained:

$$C_s = \frac{K_d}{K_h} C_g. \qquad (27)$$

Thus when neglecting biodegradation, the governing equation can be written in terms of either the air or water phase concentration in the unsaturated zone. For the gas phase, neglecting gas advection, the governing equation will be simplified as:

$$a\frac{\partial C_g}{\partial t} + \nabla \cdot [b\nabla C_g + cC_g] = 0 \qquad (28)$$

$$a = \theta_g + (\theta_l + K_d\rho_b)/K_h \qquad (29)$$

$$b = -((\delta_l + D_l)/K_h + D_g) \qquad (30)$$

$$c = q_l/K_h \qquad (31)$$

$$D_g = d_g\theta_g\xi_g \qquad (32)$$

$$D_l = d_l\theta_l\xi_l. \qquad (33)$$

The numerical solution to this system was obtained for a radially symmetrical geometry (Baehr, 1987), where the unsaturated porous media is assumed to be isothermal and air/water partition coefficients were assumed constant. Porosity, water content, tortuosity, hydrodynamic dispersion coefficients, adsorption coefficients, and the volumetric water flux were also assumed constant.

The Model Application in This Research

This research applied the transport model described above

36

to the experimental data measured by Yu (1995). This experiment measured toluene gas phase concentration profiles at different times by running a vertical soil column with 45 cm diameter and 25 cm height. Toluene gas was sampled by syringe at various depths and times. Concentrations were measured immediately with a gas chromatograph with a flame ionization detector. The soil was uniformly packed with toluene so that the initial concentration was constant. The bottom of the soil column was sealed and the top of the column was open to the atmosphere. The water content was uniform and well below saturation. While some drying at the top occurred during the test, the modeling ignored water transport.

Conditions under which the transport model was run are consistent with this column test. Therefore, the boundary condition at the bottom was:

$$\frac{dC_g}{dz}\Big|_{bottom,\,t} = 0\,. \tag{34}$$

This equation implies that the bottom of the soil column is impervious to vapors and that no mass can escape from it. The boundary condition at the surface is:

$$C_g\big|_{surface,\,t} = 0\,. \tag{35}$$

The initial condition is:

$$C_g\big|_{t=0,\,z} = constant\,. \tag{36}$$

Parameter inputs involved in this model can be classified

into three groups: geometry parameters, soil property parameters, and chemical property parameters. Table I lists the values used for each group (consistent with the experimental conditions). This research focused on tortuosities in the air phase and water phase, and partition coefficients between the water-air phase and between the water-solid phase.

Characteristics of the Model Parameters

Tortuosity is a measure of the added resistance to diffusion imposed by the structure of the medium (Kreamer et al., 1988). As a major component that determines the rate of diffusion of a given chemical, it is independent of the chemical properties and is dependent on the pore geometry.

Although many experimental methods have been developed to measure tortuosity either in labs or in fields, they have not been able to give reliable results. For lab methods, coring and repacking samples can substantially change the structure of the medium, thus causing a variance between lab results and field values. On the other hand, field tests are time consuming and expensive, and require skilled people to analyze the data. Moreover, Kreamer et al. (1988) reported that the tortuosity they measured may be in error by as much as 40%. There are also empirical equations that can estimate tortuosity. However, significant discrepancies among these estimates have been reported. Kreamer et al (1988) summarized

Table I

The Description of Parameter Inputs for Baehr's Model

Geometry Parameters	Soil Properties	Chemical Properties (Toluene)
Column Diameter 45 cm	Porosity 0.4	Specific Volume 1.14 g/cm^3
Column Height 25 cm	water content 0.17	Molecular Weight 92
	bulk density 1.59 gm/cm^3	Solubilities 0.515 E-3 cm^3/g
	water recharge 0.0	Molecular Diffusion Coefficient in Water 10^{-5} cm^2/s
	Longitudinal Mechanic Dispersion Coefficient 0.0 cm^2/s	Molecular Diffusion Coefficient in Air 0.1 cm^2/s
	Transverse Mechanical Dispersion Coefficient 0.0 cm^2/s	Henry's Constant 0.26
	Air Phase Tortuosity 0.34	Adsorption Coefficient 0.43
	Water Phase Tortuosity 0.1	

that under the same condition, the tortuosity estimated by Lai et al. (1976) is 0.13 while that estimated by Marshall (1959) is 0.47. These large variations suggest that the empirical equations tend to be applicable only to the materials and conditions for which they were developed.

The adsorption coefficient is the ratio of the amount of chemical adsorbed per unit weight of soil to the concentration of the chemical in solution at equilibrium. It represents the extent to which an organic chemical partitions itself between the solid and solution phases. It is determined by several physical and chemical properties of both the chemical and the soil. However, for soils with high organic carbon content, basing the adsorption coefficients on soil organic carbon (K_{oc}) rather than on total mass (K_d) can eliminate some influence of soil properties. Even so, studies show that the spread of values obtained from a number of different soils generally results in an uncertainty ranging from 10% to 140% (Lyman, 1990). Errors also arise from the use of simple adsorption isotherms, such as the linear adsorption isotherm when the isotherms could be nonlinear (Villeneuve et al., 1988).

The Henry's law constant is conventionally defined as a ratio of partial pressure in the vapor to the concentration in the liquid (Mackay et al., 1981). However, it is more convenient to express it as a dimensionless ratio of concentration in vapor phase and concentration in water phase. Mackay et al. (1981) summarized three general methods that can

be used to measure the Henry's law constant: 1) measurement of the ratio of vapor pressure and solubility, 2) direct measurement of air and aqueous concentrations in a system at equilibrium, and 3) measurement of relative changes in concentration within one phase, while effecting a near-equilibrium exchange with the other phase. According to Gossett (1987), the first method suffers from the lack of reliable solubility data, the second method is difficult to carry out where concentrations are low, and the third method suffers if equilibrium is hard to reach. He proposed a modified equilibrium partitioning in closed system method which achieved 3-4% C_v in measured Henry's constant. However, he concluded that the precision deteriorated dramatically for compounds with very low Henry's constant.

A Comment On Model Form

Equations 28 to 33 form a second order, partial differential equation for the air phase solute concentration in terms of time and space. The coefficients, a and b are lumped parameters, which combine phase partition coefficients (K_h and K_d) and soil transport properties (ξ_g and ξ_l) (the other input parameters are usually well defined). Equations such as these are notorious for both their difficulty of solution, and even more importantly, the difficulty of estimating their parameters from experimental data. This inverse problem of parameter estimation is traditionally

41

approached by trial and error procedures which simply match simulated model output to measured data. Of course with a lump parameter model, an error in the estimate of one parameter can be compensated by the appropriate error in another. This problem is made worst by the uncertainties introduced by measurement errors. Thus, the utility of any model, or the benefit of laboratory measurements, is questionable unless some procedure is available to quantify both the uncertainty associated with input parameters and the accuracy of any model prediction based on them. Unless that can be done, process based models such as this have little advantage over pure empiricalism.

CHAPTER IV

ESTIMATION AND UNCERTAINTY THEORY

This chapter will discuss the basic Bayes' theorem underlying this research and procedures used in parameter estimation and uncertainty analysis for volatile organic transport.

Basic Theory

Bayes' Theorem

Consider a random variable Y with a vector of n observations $y=(y_1,\ldots,y_n)$. It has a joint probability distribution $p(y/\theta)$, which depends on the values of k parameters $\theta=(\theta_1,\ldots,\theta_k)$. Suppose that θ is also a random variable and has a probability distribution $p(\theta)$, then from the definition of conditional probability:

$$p(y/\theta)\,p(\theta)=p(y,\theta)=p(\theta/y)\,p(y). \tag{37}$$

When given the observed y, the conditional distribution of θ is:

43

$$p(\theta/y) = \frac{p(y/\theta)\,p(\theta)}{p(y)} \,. \qquad\qquad \textbf{(38)}$$

Note that

$$p(y) = \int p(y,\theta)\,d\theta = \int p(y/\theta)\,p(\theta)\,d\theta = constant. \qquad \textbf{(39)}$$

Therefore $p(y)$ is only a constant which assures $p(\theta/y)$ of integrating to 1. This leads to:

$$p(\theta/y) = cp(y/\theta)p(\theta) \,. \qquad\qquad \textbf{(40)}$$

In this expression, $p(\theta)$ is the prior distribution of θ and it represents the known information about θ before observing y; $p(\theta/y)$ is the posterior distribution of θ given y and it tells the information of θ after knowing y. c is a normalizing constant to ensure that $p(\theta/y)$ integrates to 1. $P(y/\theta)$ will be explained in the following section.

The Likelihood function Given the observation of y, the probability distribution $p(y/\theta)$ may be regarded as a function of θ rather than y. This function is called the likelihood function of θ for given y. When the observation y is independent and identically distributed, the likelihood function is:

$$l(\theta/y) = p(y_1, y_2, \ldots, y_n/\theta) = \prod_{i=1}^{n} p(y_i/\theta) \,. \qquad \textbf{(41)}$$

where $\prod$ represents product. This leads to another form of

44

Bayes' theorem:

$$p(\theta/y) \propto l(\theta/y)\,p(\theta).\tag{42}$$

The likelihood function here is a function through which the observed y modifies prior knowledge of θ.

<u>Sequential Nature of Bayes' Theorem</u> One important aspect of Bayes' theorem is that it allows updating information on θ when taking more observations of y. Therefore for an initial sample of observations y_1:

$$p(\theta/y_1) \propto p(\theta)\,l(\theta/y_1)\tag{43}$$

When we have a second sample of observations y_2 distributed independently of the first sample, then:

$$p(\theta/y_1,y_2) \propto p(\theta)\,l(\theta/y_1)\,l(\theta/y_2) \propto p(\theta/y_1)\,l(\theta/y_2)\tag{44}$$

Apparently, the posterior distribution for θ given y_1 serves as the prior distribution for the second sample. If we have n independent observations, the posterior distribution can be recalculated after each new observation:

$$p(\theta/y_1,\ldots,y_m) \propto p(\theta/y_1,\ldots,y_{m-1})\,l(\theta/y_m)\,,m=2,\ldots,n\tag{45}$$

This provides a process of learning from experience. The interesting thing is that the advantage of Bayes' Theorem is also its disadvantage. Incorporating prior information can improve the results. However, questionable prior information might lead to faulty results. This is the reason that people

may hesitate to choose this method. Box and Tiao (1973) provided noninformative prior distributions to address this problem.

Noninformative Prior Distributions

A noninformative prior distribution is a uniform distribution that reflects minimal prior knowledge of the parameters. Noninformative prior distribution will have virtually no effect on the resulting posterior probability (Edwards 1988). That is, the prior distribution provides little information relative to what is provided by the intended experiment.

Mathematically, a non-informative prior is defined as:

$$f(\theta/x) = \frac{f(\theta) \, L(\theta/x)}{\int_{-\infty}^{\infty} f(\theta) \, L(\theta/x) \, d\theta} \approx \frac{L(\theta/x)}{\int_{-\infty}^{\infty} L(\theta/x) \, d\theta} = l(\theta/x) \, . \tag{46}$$

This concept means that the noninformative prior is a uniform distribution: $f(\theta)$=constant.

In the case of the Normal mean with n and σ known, the likelihood function can be written as (Box and Tiao, 1973):

$$l(\theta/\sigma, y) \propto \exp\left[-\frac{n}{2\sigma^2} (\theta - y_m)^2\right] \tag{47}$$

where y is a random variable, y_m is the sample mean of y, and θ is the mean of the population. Apparently, the data enters the likelihood only via the sample mean. Therefore, when the

46

likelihood is expressed in terms of θ, the sample mean y_m affects only the location of the likelihood curve. That is, the likelihood function is completely determined a priori except for its location. This is called the data translated likelihood function.

However, it is possible that the immediate interest is not θ itself but the reciprocal $k=\theta^{-1}$. In this case, the noninformative prior in terms of k can be evaluated as (Wilson, 1990):

$$f(k) = f(\theta) \left| \frac{d\theta}{k} \right| = k^{-2} f(\theta) = k^{-2} c \qquad (48)$$

Here it is obvious that the prior distribution for k is not constant. Therefore, the standardized likelihood function will not only change locations but also spread with different sets of data. The spread of the distribution is then biased by the selection of the variable and is in conflict with the original goal of selecting a uniform prior. When this happens, it is necessary to derive a parameter transformation that produces the data-translated likelihood function. The noninformative prior for the normal distribution of σ introduces this concept.

Consider a standardized likelihood function with an unknown standard deviation and a known mean θ for n observed values of x:

$$l(\sigma/x,\theta) = K\sigma^{-n}\exp\left[-\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\sigma^2}\right] \qquad (49)$$

where K is the normalizing constant. This equation can be evaluated using s^2 defined as:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)^2 \qquad (50)$$

and

$$l(\sigma/x,\theta) = K\sigma^{-n}\exp\left[-\frac{s^2}{2\sigma^2}\right]. \qquad (51)$$

This is not data translated. Now consider the transformed variable: $k=log\ \sigma$ and $dk=\sigma^{-1}d\sigma$. Box and Tiao (1973) showed the resulting transformed distribution is data translated. Therefore the appropriate locally uniform distribution can be written as $f(k)$=constant, then this prior can be written as:

$$f(\sigma) = f(k)\left|\frac{dk}{d\sigma}\right| = \frac{c}{\sigma}. \qquad (52)$$

For an independent variable, the joint probability density function can be written as $f(\theta,\sigma)=f(\theta)f(\sigma)$. For a normal-gamma function, the appropriate noninformative priors for each marginal distribution can be written as $f(\theta)=c_1$ and $f(\sigma)=c_2/\sigma$. Therefore the joint probability can be written as $f(\theta,\sigma)=c/\sigma$ for $\sigma \geq 0$ and $f(\theta,\sigma)=0$ for $\sigma<0$, where c incorporates c_1 and c_2.

Posterior Distribution

Based on the above Bayes Theorem, suppose that we have a model $f(x, \theta)$ which is used to simulate the output $(y_1, \ldots, y_p)^T$ as a function of inputs $(x_1, \ldots, x_m)$ and parameters $(\theta_1, \ldots, \theta_n)$, where T represents transposition. The observed model outputs can be expressed as:

$$y = f(x, \theta) + \epsilon \tag{53}$$

where ϵ is the residual.

Box and Tiao (1973) used independent and exponential-power distributions with zero mean to describe the stochastic nature of the residuals. According to them, the probability density function for each of these residuals can be expressed as:

$$f(\epsilon) = \frac{\omega(\beta)}{\sigma} \exp\left[-c(\beta)\left|\frac{\epsilon}{\sigma}\right|^{2/(1+\beta)}\right] \tag{54}$$

where β is a parameter between -1 and 1 and $\omega(\beta)$ is defined as:

$$\omega(\beta) = \frac{[\Gamma(1.5(1+\beta))]^{1/2}}{(1+\beta)[\Gamma(0.5(1+\beta))]^{3/2}} \tag{55}$$

and $c(\beta)$ is defined as:

$$c(\beta) = \left[\frac{\Gamma(1.5(1+\beta))}{\Gamma(0.5(1+\beta))}\right]^{1/(1+\beta)}. \tag{56}$$

The parameter β is a measure of kurtosis that describes the

non-normality of the observation data. The observation data have a normal distribution when $\beta=0$, a double exponential distribution when $\beta=1$, and a rectangular distribution when $\beta=-1$.

Assuming prior independence between the vector of the model parameter and the standard deviation of the residuals, an appropriate noninformative prior probability density function can be given as:

$$p'(\theta, \sigma) \propto 1/\sigma. \tag{57}$$

By Bayes' Theorem, the posterior probability density function is proportional to the product of the prior probability density function and the likelihood function:

$$p''(\theta, \sigma / \beta, \epsilon) \propto \frac{1}{\sigma^{n+1}} \exp\left[-c(\beta) \sum_{i=1}^{n} \left|\frac{\epsilon_i}{\sigma}\right|^{2/(1+\beta)}\right]. \tag{58}$$

By integrating the above equation with respect to σ, the distribution of θ can be obtained as:

$$p''(\theta / \beta, \epsilon) \propto \left[\sum_{i=1}^{n} |\epsilon_i|^{2/(1+\beta)}\right]^{-n(1+\beta)/2}. \tag{59}$$

In terms of observed y_i, the probability density function of θ can be written as:

$$p''(\theta / \beta, y) = \frac{\left[\sum_{i=1}^{n} |y_i - f(x_i, \theta)|^{2/(1+\beta)}\right]^{-n(1+\beta)/2}}{\int \left[\sum_{i=1}^{n} |y_i - f(x_i, \theta)|^{2/(1+\beta)}\right]^{-n(1+\beta)/2} d\theta}. \tag{60}$$

50

The point estimate of θ is taken as the mode of the posterior probability density function of θ (the mode is the most frequently occurring value) and can be found by searching:

$$\min\left[\sum_{i=1}^{n} |y_i - f(x_i, \theta)|^{2/(1+\beta)}\right]. \tag{61}$$

In this research, y_i was substituted by a measured concentration profile, $f(x_i, \theta)$ was the simulated concentration profile produced by Baehr's model, and θ represented the four parameters specified in Chapter III. β was assumed to be zero which represents a normal distribution for residuals.

We have so far considered only observations made from a single model response. However, contaminant transport models often produce several outputs. Incorporating more observed information into parameter estimation should help the problem become better posed. Suppose that the model produces m outputs and each output has n observations, then the residual between the model simulation and observation will be an $n \times m$ matrix. Assume that the error vector is distributed as the m-variate Normal $M_m(0, \Sigma)$, where Σ is the $m \times m$ covariance matrix of residuals. The joint distribution of the n vectors of error $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$ is:

$$p(\epsilon|\Sigma, \theta) = \prod_{u=1}^{n} p(\epsilon_u|\Sigma, \theta). \tag{62}$$

Expanding Equation 62 yields:

$$p(\epsilon|\Sigma,\theta) = (2\pi)^{-mn/2}|\Sigma|^{-n/2}\exp\left(-\frac{1}{2}\sum_{u=1}^{n}\epsilon_u\Sigma^{-1}\epsilon_u\right). \tag{63}$$

Letting

$$S(\theta) = [S_{ij}(\theta_i,\theta_j)] = [\sum_{u=1}^{n}\epsilon_{ui}\epsilon_{uj}] \tag{64}$$

then the exponent in the previous equation can be expressed as:

$$\sum_{u=1}^{n}\epsilon_u\Sigma^{-1}\epsilon_u = trS(\theta)\Sigma^{-1} = \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma^{ij}S_{ij}(\theta_i,\theta_j) \tag{65}$$

where tr $S(\theta)$ means the trace of the matrix $S(\theta)$. Given these observations, the likelihood function can be written as:

$$l(\theta,\Sigma|y) \propto p(\epsilon|\Sigma,\theta) \propto |\Sigma|^{-n/2}\exp\left[-\frac{1}{2}tr\Sigma^{-1}S(\theta)\right]. \tag{66}$$

Now for the prior distribution of the parameters (θ,Σ), assume that θ and Σ are approximately independent so that

$$p'(\theta,\Sigma) \doteq p'(\theta)p'(\Sigma). \tag{67}$$

When taking θ as locally uniform and applying Jeffreys' Rule for multiple parameters to the covariance matrix of the residuals Σ,

$$p'(\theta) \propto constant \tag{68}$$

and

52

$$p'(\Sigma) \propto |\Sigma|^{-\frac{1}{2}(m+1)} . \tag{69}$$

Now the posterior joint probability density function is proportional to the product of the likelihood function and the prior joint probability density function. When integrating out Σ, the marginal distribution of θ will simply be

$$p(\theta|y) \propto |S(\theta)|^{-n/2} \tag{70}$$

and the "most probable" value of θ will be

$$\min|S(\theta)| . \tag{71}$$

Note that this is a general derivation for multiple model responses cases. When applied in this research, only two model responses were considered. Therefore, $S(\theta)$ was a 2 x 2 matrix, y represented the two sets of measured toluene gas concentration profiles at different times, n represented the number of data points involved in a concentration profile, and θ represented parameters specified in Chapter III.

Parameter Estimation Procedure

The foregoing part of this chapter has provided a method for solving parameter estimation problems. This section will present the procedure that applies the present methodologies to solve Equations 60 and 61 or 70 and 71. This will lead to the point estimates (the estimation of an optimal set of parameters) and the marginal distribution estimate.

The Point Estimates

The point estimates can be obtained by using optimization techniques to solve Equation 61 or Equation 71. A vast number of optimization methods exist. There seems to be no answer to the question of which is the best strategy. This research chose the simplex method over other methods because of simplicity. Moreover, for only a few variables, the simplex method is robust and reliable (Schwefel, 1981).

Nelder and Mead (1964) developed the basic concept. A simplex has N+1 vertices, where N is the number of parameters. For two variables, there will be three vertices arranged as an equilateral triangle. The objective function is evaluated at all the vertices. The vertex with the largest objective function value is replaced by its reflection in the midpoint of the other vertices, or the expansion of the reflection, and/or the contraction of the reflection, depending on which is the best.

The criterion for ending the search is to test whether the variance of the objective function values at the vertices of the simplex is less than a prescribed limit. The following steps describe how the algorithm works:

(1). Select initial vertices $(X_1, \ldots, X_{n+1})$. The coordinates of each vertex are a set of n-dimensional parameters $(x_{1i}, \ldots, x_{ni})$.

(2). Calculate objective function values for initial vertices.

(3). Determine the point corresponding to the largest value.

(4). Find the center of mass of these points by

$$x_{i,c} = \frac{1}{n} \left[\left(\sum_{j=1}^{n+1} x_{i,j} \right) - x_{i,R} \right] \tag{72}$$

where $x_{i,R}$ is the rejected point and $x_{i,c}$ is the center of mass of the remaining points.

(5). Determine the tentative new point to replace the rejected point by using $\theta = 1$ with

$$x_{i,N} = x_{i,R} + (1+\alpha)(x_{i,c} - x_{i,R}) \tag{73}$$

where α equals 1 for regular simplex, 2 for expanding simplex, and 0.5 for contracting simplex.

(6). Decide whether a different point using contraction or expansion should be obtained using the following criteria:

a. Expanding simplex: if the tentative new point gives a value that is better than the current best value, then calculate a new point and its function value using this expansion equation ($\alpha = 2$). If the expanded point is better than the tentative point, then use the expanded point, otherwise, use the tentative point.

b. Contracting simplex: if the tentative new point gives a value that is worse than the second worst point, then calculate a new point and its function value using the contraction equation ($\alpha = 0.5$). If the contracted point is

55

better than the tentative point, then use it, otherwise, use the tentative point.

(7). Repeat steps (2) through (5) until the tolerance value is acceptable.

The weak point of this method is that the fixed searching parameter α limits the advancement of the searching. Marsili-Libelli and Castelli (1987) modified this method by making the searching parameter adaptive. Their modification enables the determination of a local minimum in the search direction each time an expansion is performed. This feature allows the algorithm to adapt the pattern search parameters to the particular shape of the objective function.

Marsili-Libelli (1992) applied this modified method to parameter estimation of ecological models. His results proved that this method well fit cases where the minimum lies in a narrow trough in the parameter space. This research employed the modified method to search for the optimal set of parameters.

Marginal Distribution

The other goal of parameter estimation is to obtain the marginal probability distribution for each parameter by integrating Equation 60 or Equation 70. This research employed Monte Carlo integration because of its simplicity and its ability to deal with multidimensional problems.

The detailed theory can be found in Davis and Rabinowitz

(1975). Suppose we want to compute

$$I = \int_a^b f(x)\,dx \qquad (74)$$

then the mean value of $f(x)$ over the interval $[a,b]$ is $I/(b-a)$. If we sample $f(x)$ so that x is from a random uniform distribution, then the sample mean would be

$$f = \frac{1}{n} \sum_{i=1}^n f(x_i). \qquad (75)$$

Therefore, the sample mean f could be an approximation of the mean value $I/(b-a)$, which leads to:

$$\int_a^b f(x)\,dx \approx \frac{b-a}{n}[f(x_1) + \ldots + f(x_n)]. \qquad (76)$$

The variance of this estimate is $O(1/N)$, where N is the sample size. Rubinstein (1981) described the weighted Monte Carlo integration for variance reduction:

1. Generate $X_1, \ldots, X_n$ from $U(0,1)$.

2. Arrange $X_1, \ldots, X_n$ in the increasing order.

3. Estimate the integral by

$$\int_a^b f(x)\,dx \approx \frac{1}{2}\left[\sum_{i=0}^N (f(x_i) + f(x_{i+1}))(x_{i+1} - x_i)\right]. \qquad (77)$$

where $X_0 = 0$ and $X_{N+1} = 1$.

In two dimensional case, the variance of this estimation is $O(1/N^2)$ which gives a standard error of $O(1/N)$. Therefore, in order to achieve the two significant figures of accuracy

57

(with a standard error less than 0.005), we need to sample x about 2500 times (with a standard error=1/2500=0.004).

Assumptions The goal of parameter estimation is to get unbiased and consistent estimates. Since parameter estimation that relies on mathematical and statistical inferences requires assumptions, it is important to state and check assumptions carefully. If assumptions are violated, the estimated parameters and the predicted output are biased.

The following assumptions are used in most parameter estimation methods:

- Errors have zero mean.

- Errors have a constant variance.

- Errors are uncorrelated.

- Errors are normally distributed.

These assumptions are often violated in inverse problems. More serious difficulties arise due to violation of the constant variance and uncorrelated errors assumptions which often occur in practical problems. For instance, error variances are commonly found to increase with the magnitude of the property being measured. Unequal error variances also result when the observation vector contains different types of measurement expressed in different units.

If any assumptions are violated, the general procedure is to transform the data so that the transformed data satisfy the assumptions. Commonly, transformation of data will overcome violations of the assumptions. However, non-constant variance

and non-normality cannot always be eliminated by transformations. Some commonly used techniques for data transformation include:

- standardization,
- ARMA model transformation to eliminate the autocorrelation,
- reciprocal transformation,
- power transformation, and
- Logarithmic transformation.

The last four transformations can often change the error properties in several aspects simultaneously, such as eliminating non-constant variance, and non-normal distribution. Another commonly used method for correcting the non-constant variance is the weighted least-squares method mentioned in Chapter II (Equation 9).

Uncertainty Analysis Procedure

Chapter II discussed the idea that the prediction of uncertainty ultimately relates to three basic sources: natural uncertainty, parameter uncertainty, and model structure uncertainty. In fact, to date, the evaluation of uncertainty has focused more on parameter uncertainty than on the other two sources. The estimation of model parameters is subject to greater errors when few measured data are available to form the estimates. Also, the performance measures are influenced more by some parameters than by others. Therefore, when we

quantify model uncertainty, we are looking at two aspects:

1. Sensitivity analysis that studies each variable to understand its relative importance in the model.

2. Joint uncertainty analysis that quantifies the uncertainty of model response influenced by the uncertainty of all parameters.

Sensitivity analysis

There are two ways to conduct sensitivity analyses. One basic approach is to introduce small perturbations in the various processes and parameters of the model and to study their relative effects on the output variable of interest. This method requires intensive computation for accurate calculation, however, it is sufficient for a rough analysis.

The other method is to consider a sensitivity as a partial derivative, which represents the change in model prediction resulting from a change in a model parameter. If $Y=f(X_1,\ldots,X_n)$, then the relation is:

$$S = \frac{\partial Y}{\partial X_i} \cdot \frac{X_i}{Y} \qquad (78)$$

where S is the sensitivity index of Y with respect to change in X_i.

Determining the sensitivity of the variable Y to each of the input variables $X_1,\ldots,X_n$ at the point $(x_1,\ldots,x_n)$ requires the calculation of n partial derivatives. Solute transport models are too complex to calculate the partial derivatives

60

directly.

Usually, both deterministic and statistical approaches can be used to deal with this problem. Deterministic sensitivity analysis is a numerical estimate of the partial derivative of Y with respect to X_1 at the point $x_1, \ldots, x_n$. Statistical method evaluates the model at many points in the input space and then fits a response surface of the input variables to the output variables.

The problem with deterministic sensitivity analysis is that it has difficulty dealing with correlated input variables (Doctor, 1989). Thus, this research will use the statistical method of partial regression techniques incorporated with a Monte Carlo simulation to complete the sensitivity analysis.

The response surface can be represented by a linear model (Doctor, 1989):

$$Y = \alpha + \sum \beta_i X_i + \epsilon \tag{79}$$

where β_i is the partial regression coefficient which is the estimate of the sensitivity of Y to the input variable X_i.

Equation 79 is a linear approximation to the nonlinear model. Following Tiscareno-Lopez et al (1993), it is assumed that "the linear model is able to assess unbiased estimates of sensitivity indices of model parameters of a complex nonlinear model when a large number of model simulations are performed." This equation can be standardized as:

$$Y^* = \sum \beta_i^* X_i^* + \epsilon \qquad (80)$$

where $Y^* = (Y - Y_m)/\sigma_y$, $X^* = (X - X_m)/\sigma_x$, Y_m is the mean of Y, X_m is the mean of X, σ_y and σ_x are standard deviations of Y and X, and β^* is the standardized partial regression coefficient. The value of X can be simulated from the estimated probability distribution.

For the standardized partial regression coefficients to be reasonable estimates of the sensitivity, the response surface must give an adequate representation of the function. For lack of a better test, this adequacy is measured by the multiple correlation coefficient, R^2, defined as:

$$R^2 = 1 - \frac{\sum (Y^* - \hat{Y}^*)^2}{\sum (Y^* - Y_m^*)^2} \qquad (81)$$

where $\hat{Y}^*$ is the estimated value of Y^*, Y_m^* is the mean of Y^*. In a strict sense, R^2 does not directly measure how precisely β^* can estimate S in Equation 78, but high values of R^2 must imply a reasonable estimate.

Uncertainty Analysis

This research employed the Monte Carlo simulation method to accomplish uncertainty analysis. The procedure includes the following steps:

1. Generate parameter samples from the estimated distribution achieved in the foregoing procedure.

2. Run the transport model on these parameters to get a

sample of model outputs.

3.   Analyze the model output sample to obtain its distribution.

4.   Quantify its uncertainty by the coefficient of variation C_v and other statistics.

Generating parameters samples is always the critical step of the Monte Carlo simulation method.   If parameters are independent, the sample can be directly generated from the marginal pdf's because the joint probability density function (pdf) of the parameters is simply the product of the univariate pdfs.   However, many model parameters are not independent which means that the joint pdf is not the product of the univariate pdf's.   In this case, the generated parameter samples have to preserve the covariance.   This research chose the procedure given by Haan (1977) which uses principle components to generate multivariate normal parameter samples, along with empirical modification suggested by Taylor and Bender (1988).   Chapter VI will explain this further.

CHAPTER V

PARAMETER ESTIMATION RESULTS

Preliminary Sensitivity Analysis

The difficulty of parameter estimation is at least proportional to the number of parameters. The purpose of the preliminary sensitivity analysis is to qualitatively investigate the impact of each parameter on model output. Insensitive parameters should be excluded from the estimation process. The sensitivity analysis method described in Chapter IV was not used here (Chapter VI will discuss the application of this method). The qualitative method used here is to introduce small perturbations in a parameter of the model while fixing the others as constants to study its impact on the output variable.

The volatile organic transport model (Baehr, 1987) used in this research produces two output variables: the concentration profile (the concentration distribution along the soil column at different times) and total mass that escaped from the soil surface during a certain time interval. The change in output corresponding to the change of each input parameter while keeping the others constant was investigated. Modeling conditions were described in Chapter III with nominal

parameter values listed in Table I. Results are compared for 5 and 21 hours of volatilization to compare long and short term sensitivity.

Figures 1 and 2 plot the simulated toluene gas phase concentration versus depth for five and 21 hours in the soil column, with different water phase tortuosities (ξ_l). The plots show that as the tortuosity in the water phase was changed from 0.2 to 1 by increments of 0.2, the concentration profiles remained approximately the same. Likewise, Figures 3 and 4 depict the effect of water phase tortuosity on the mass transfer out of column after five hours and 21 hours.

Apparently, tortuosity in the water phase is not an important factor under the conditions of this research. This is due to the molecular diffusion coefficient in the water phase being very small (1×10^{-5}), relative to the vapor diffusion coefficient of 0.1. Therefore, it is not necessary to include water phase tortuosity in the estimation process used here.

Figures 5 and 6 show the concentration profiles after five hours and 21 hours with different tortuosities in the air phase (ξ_g). Tortuosity in the air phase was increased from 0.2 to 1 by increments of 0.2. The toluene concentration shows dramatic corresponding change, especially at the bottom of the column. Comparing Figures 5 and 6, it appears that the concentration profile is more sensitive to air phase tortuosity later in the simulation.

65

Figures 7 and 8 plot the total toluene mass escaped from the column surface and the mass remaining in the column versus tortuosity in the air phase after five hours and 21 hours, respectively. Both plots show a significant sensitivity to air phase tortuosity. Again, the sensitivity to air tortuosity is greater in later hours. However, the sensitivity becomes smaller as air phase tortuosity reaches its higher range, which is consistent with the results of the concentration profiles.

Figures 9 and 10 are the concentration profiles after five hours and 21 hours, respectively, with different values of Henry's constants (K_h). Unlike the effect of air phase tortuosity, the sensitivity increases with Henry's constant. On the other hand, the sensitivity is larger early in the simulation. Figures 11 and 12 plot the corresponding toluene mass escaped. Henry's constant is again sensitive in both cases and more so for the early time.

Figures 13 through 16 demonstrate the concentration and mass changes corresponding to the change of the solid phase adsorption coefficient (K_d) within the range of 0.2 to 1.0. Notice that, in Figures 15 and 16, the mass remaining in the column shows an increase as the adsorption coefficient is increased. On the contrary, Figures 13 and 14 show that the concentration profile decreases as the adsorption coefficient increases, which is due to plotting the gas phase concentration, and not the total concentration. When the

adsorption coefficient increased, the mass partitioned into the gas phase decreased, which reduced the gas phase concentration. Figures 13 and 14 illustrate that the concentration profile is more sensitive to the adsorption coefficient in early time. It is interesting that the adsorption coefficient in its lower range affects the concentration profile dramatically early, while it has little effect at later times.

This analysis has shown that the tortuosity in air phase, Henry's constant, and the adsorption coefficient all have an important impact on model output. However, each parameter behaves differently in terms of time and its range. Air phase tortuosity is more sensitive in its lower range and at a later time. Henry's constant is more sensitive at its higher range and at an earlier time. The adsorption coefficient is more sensitive in its lower range and at an earlier time.

To better demonstrate combined sensitivity, Figures 17 through 21 present the total toluene mass escaping the column as a function of Henry's constant (K_h) and air phase tortuosity with the solid phase adsorption coefficient (K_d) held constant. Generally speaking, the accumulated mass that escaped increases as air tortuosity and Henry's constant increases and decreases as the adsorption coefficient increases. The interaction among these three parameters affects model output as well. These five graphs (Figures 17 through 21) show that Henry's constant has less impact on

model output with large tortuosity. In Figure 17, when air tortuosity is 0.1, the increase of Henry's constant by 0.8 almost doubles the mass escaping. When air tortuosity is 1.0, the same increase of Henry's constant increases mass escaping by less than 10 percent. On the other hand, an increase of the adsorption coefficient reduces the impact of air tortuosity while it increases the impact of Henry's constant slightly. Curves in Figure 21 are flatter than in Figure 17, and the curve spread range is broader. Figures 18, 19, and 20 show an uniform transition between the extremes. Thus, while these are complex relations, they are well behaved.

The Optimal Estimates of Parameters

The Bayesian methodology described in Chapter IV combined with an adaptive simplex method was used here to find a set of optimal parameters for either Equation 61 or Equation 71. To verify the estimation procedure, hypothetical simulated data were used first to test the procedure, followed by the experimental data. As for the hypothetical data, model simulation results for given parameters were used as observations fed into the estimating algorithm. Parameters were taken from literature and experience. Henry's constant for toluene is 0.26 (Baehr, 1987), air phase tortuosity was calculated from the Millington and Quirk model (Brown and McWhorter, 1990) as 0.34, and the adsorption coefficient was taken from Yu (1995) as 0.43.

The simulated results fed into the Bayesian algorithm were corrupted with additive noise to better model real data. The noise was added by:

$$Y_i = Y_{mi} + \epsilon_i, \ i = 1, 2, \ldots, n \qquad (82)$$

where Y_i is the value fed into the Bayesian algorithm, Y_{mi} is the simulated result and ϵ is an uncorrelated random noise sample drawn from a Gaussian distribution with zero mean and standard deviation σ. Various values of σ were used to study the effect of measurement error on parameter estimates.

If parameters estimated by the algorithm are close to the given parameters within an acceptable tolerance, the algorithm is considered feasible. Before getting into an optimal search, it is necessary to study the characteristics of the objective function because of the apparent nonlinearity in model parameter structure.

Objective Function Response Surface and Contours

The volatile organic transport model employed in this research produces concentration profiles across the plume at different times. These outputs were used as observations in the objective function equations 61 and 71 described in chapter IV.

Using One Concentration Profile Equation 61 was used here as the objective function to find optimal estimates for air phase tortuosity, Henry's Law constant, and the adsorption coefficient. Figures 22 through 27 demonstrate the three-

dimensional response surfaces and two-dimensional contours in two-parameter space with the third parameter fixed. The response surfaces show the objective function over the entire domain, while the contours are best suited for locating the minimum.

It is apparent from the contour maps that the minima of the objective function fall within an optimal range and there are many local minima. For Henry's constant and adsorption coefficient space, there are two optimal ranges. These characteristics reflect that the three parameters are correlated due to the model structure. Apparently, the data used in the objective function are not sufficient to overcome the interrelationship among parameters, therefore it is necessary to use more information.

<u>Using Two Concentration Profiles</u> Equation 71 was employed to bring in two concentration profiles at different times. Figures 28 through 33 show the resulting response surfaces and contour maps. The difficulty due to the correlation between air phase tortuosity and the adsorption coefficient has been reduced. Similarly, the situation for air phase tortuosity and Henry's constant has been improved by bringing in more information. The global minimum exists within a narrow trough. However, Figure 33 shows that the correlation between the adsorption coefficient and Henry's constant has not been overcome because there are still many local minima.

The straight contour lines in Figure 33 reflect the linear relationship between Henry's constant and the adsorption coefficient. From the governing equation in Chapter III, we can see that the adsorption coefficient K_d and Henry's constant K_h form the coefficient additively:

$$a = \theta_g + \frac{K_d}{K_h}\rho_b + \frac{\theta_l}{K_h} . \qquad (83)$$

Therefore,

$$K_d = \frac{a - \theta_g}{\rho_b}K_h - \frac{\theta_l}{\rho_b} . \qquad (84)$$

K_d is linearly related to K_h in the model formulation. This is more clear if the physical meaning of these two parameters is considered. Both parameters are phase partitioning coefficients. For a certain mass amount, the change in phase partitioning coefficients must offset one another to assure the same total mass. Therefore, these two parameters are not identifiable if both of them are considered uncertain.

Chapter II discussed that reducing the dimensionality of the parameter field can overcome some identifiability problems associated with spatially varying parameters. Since Henry's constant is easier to measure than adsorption coefficient, it will not receive further consideration in this analysis and it will be set to its nominal value for the remaining analysis.

The Optimal Results

The objective contour map plotted in Figure 29 shows that a global minimum exists. However, due to the interaction between the two parameters, the objective function does exhibit elongated regions around the minimum. It is well known that many direct search methods perform poorly or fail altogether for this kind of situation. Therefore, an adaptive search algorithm based on the simplex search method of Nelder and Mead (1964) and modified by Marsili-Libelli and Caslelli (1987) was used here. Chapter IV described the theory of the search technique. The only constraints placed on the optimal parameters were that air tortuosity is less than unity and greater than zero and the adsorption coefficient is greater than zero.

This analysis intends to emphasize how the error in observation measurement would affect the parameter estimation results. Fifteen sets of hypothetical observations were constructed with Equation 82, thus, adding noise with different standard deviations (σ) to the simulated data with given model parameters. Fifteen values of σ were used from 0.1 to 0.8 by an increment of 0.05. Even for the same variance, different realizations of observations would have different results. Therefore, 15 randomly chosen realizations of observations were used for each noise variance. This implies that for each noise variance, there is a distribution of optimal parameters.

72

Figures 34 and 35 present the behavior of the estimated parameters as increasing noise is added to the simulated results. Clearly, as the noise variance increases, the mean of each estimated parameter tends to exhibit larger oscillations around the true values. The 95% confidence interval of the mean increases as well. When observation error is large, not only does the optimal estimate drift more from the true value, but the estimate itself is also less reliable. Such solution instability is not rare with inverse problems. Carrera and Neuman (1986) summarized several studies that used the hydraulic head data to estimate hydraulic conductivity. He stated that when the head data are corrupted by noise, the computed conductivity values exhibit uncontrolled spatial oscillations due to some "improperly posed" problems in the governing partial differential equations. He also suggested that smoothing observed head data can overcome this instability. Under the conditions specified in this research, Figures 34 and 35 clearly show that when the standard deviation of the noise is less than 0.35 mg/1, the optimal estimates can be considered stable. This result could serve as a criterion to judge the quality of measured data.

To provide a qualitative feel for this level of noise, Figures 36 and 37 plot the simulated results of two realizations with noise standard deviations of 0.3 and 0.4. Apparently, this range of measurement error is not too

difficult to achieve. In fact, many data measured in the lab could even be less noisy than this (Roll, 1995).

Application to Experimental Data

The procedure tested above is now applied to lab data described in Chapter III. The data were obtained by the experiment reported by Yu (1995) under the conditions listed previously. Figure 38 shows the concentration profiles measured at different times. Two profiles measured at 13 and 21 hours were chosen to construct the objective function (Equation 71). All the data inputs are the same as those used in the above testing procedure. The search algorithm reached the optimal estimates of 0.42 for air phase tortuosity and 0.39 for the adsorption coefficient. These two optimal estimates were put back in the program to produce a new simulation. Figures 39 and 40 compare the simulated concentration profiles using estimated parameters and measured data. The simulated data fit lab data very well. Similar comparisons for lab data obtained at other times appear in Figures 41 through 45. Not too surprisingly, since these data were not used as the criteria of the optimal search, they did not fit as well as the data measured at 13 and 21 hours.

One purpose of a point estimator for model parameters is to make judgements regarding the stochastic nature of the associated residuals (Edwards, 1988). Chapter III stated the least squares assumptions that this estimation algorithm must

meet to justify the results being unbiased and consistent optimal estimates. If the estimates obtained here result in residuals that violate those assumptions, Equation 72 cannot be used to estimate the joint probability density functions of the model parameters directly. A data transformation must be conducted first until those assumptions are satisfied.

Figures 46 and 47 are the plots of concentration residuals at 13 and 21 hours. The residuals do not show any obvious trends or non-constant variance. Therefore, the necessary assumptions underlying the estimation procedure are considered to be satisfied and Equation 72 can be used to estimate the joint probability density function of model parameters.

Marginal Distribution of Parameters

We have obtained the optimal estimates for each model parameter. However, the algorithm considers each model parameter as a random variable that is best represented by a probability distribution. Therefore, the ultimate goal of the estimation algorithm developed in this research is to obtain the probability distribution for each model parameter.

The marginal probability density function of a model parameter was obtained by integrating the joint probability distribution of parameters described by Equation 70, respect to other parameters. When the number of parameters involved in the joint probability is greater than 2, this integral is

multidimensional. Monte Carlo integration described in Chapter IV (Equation 77) was employed here to deal with possible multidimensional integration. The simulation sample size N was chosen as 3000 to achieve the two significant figures of accuracy (with a standard error less than 0.005).

To investigate how observed error would affect the estimation, the hypothetical observations described previously were used here first as well. Figures 48 and 49 depict the probability density function of air phase tortuosity for two groups of observation errors. Likewise, Figures 50 and 51 depict the probability density function of the adsorption coefficient for two groups of observation errors. These Figures clearly show that as the uncertainty in observations increases, the uncertainty in parameter estimation increases. In a more explicit way, Figures 52 and 53 depict the relationship of the half height width of each distribution, the mode, and the observation error. The half height width is the width of the distribution at one half the maximum height. As the standard deviation of observation error increases from 0.05 to 0.55, the half height width of the distribution increased from 0.011 to 0.101 for air tortuosity. For the adsorption coefficient, the half height width increases from 0.0195 to 0.185. The mode also showed an increasing tendency departing from the true value.

Table II and Table III summarize the detail statistics of each distribution for air tortuosity and the adsorption

coefficient respectively. Table II shows that the standard deviation of estimated air tortuosity distribution increases from 0.00429 to 0.047 while the standard deviation of observation error increases from 0.05 to 0.55.

Table II

Statistics of Estimated Air Tortuosity for Synthetic Data

Synthetic Data Error	Mean	Mode	Standard Deviation	Skewness	Kurtoisis
0.05	0.339	0.34	0.00429	-0.659	6.78
0.1	0.332	0.33	0.00837	-0.441	3.93
0.15	0.334	0.34	0.01	-0.230	3.58
0.2	0.343	0.34	0.013	-0.247	3.54
0.25	0.313	0.32	0.022	-0.492	3.46
0.3	0.339	0.34	0.020	-0.268	4.47
0.35	0.339	0.34	0.022	-0.408	3.61
0.4	0.342	0.35	0.027	-0.524	3.68
0.45	0.288	0.29	0.037	-0.255	2.87
0.5	0.336	0.34	0.042	-0.094	2.98
0.55	0.358	0.37	0.047	-0.536	3.66

Similarly, Table III shows that the standard deviation of the estimated adsorption coefficient increases from 0.011 to 0.084.

Table III

Statistics of Estimated Adsorption Coefficient

for Synthetic Data

Synthetic Data Error	Mean	Mode	Standard Deviation	Skewness	Kurtoisis
0.05	0.436	0.43	0.011	0.496	5.14
0.1	0.446	0.45	0.017	0.304	3.75
0.15	0.44	0.44	0.02	0.126	3.55
0.2	0.423	0.42	0.03	0.103	3.52
0.25	0.471	0.47	0.038	0.127	3.17
0.3	0.399	0.39	0.042	0.46	4.1
0.35	0.42	0.42	0.048	0.175	3.44
0.4	0.424	0.42	0.051	0.218	3.38
0.45	0.499	0.5	0.069	-0.114	3.0
0.5	0.518	0.52	0.069	-0.165	3.3
0.55	0.43	0.42	0.084	0.378	3.52

The measured data were then used. Figures 54 and 55 illustrate the marginal probability distribution for air tortuosity and the adsorption coefficient respectively, based on lab data. Table IV summarizes the statistics of model parameters estimated from measured data. The covariance of the two parameters then was determined by integrating their joint probability density function $P(\theta/Y)$ described by Equation 70. Since θ here is a vector of these two parameters, $P(\theta/Y)$ can be written as $P(\theta_1, \theta_2/Y)$, thus the covariance between these two parameters can be calculated by

$$COV(\theta_1, \theta_2) = \iint (\theta_1 - \mu_1)(\theta_2 - \mu_2) P(\theta_1, \theta_2/Y) \, d\theta_1 d\theta_2 \qquad \textbf{(85)}$$

where θ_1 and θ_2 represent air tortuosity and the adsorption coefficient, respectively, μ_1 is the mean of θ_1, μ_2 is the mean of θ_2, and Y represents the two sets of measured toluene gas concentration profiles. The correlation coefficient listed in Table IV was derived from the covariance and standard deviation of each parameter.

Table IV

Statistics of Estimated Parameters for Lab Data

Statistics	Air Tortuosity	Adsorption Coefficient
Mean	0.417	0.392
Mode	0.42	0.39
Standard Deviation	0.014	0.039
Skewness	-0.498	0.275
Kurtosis	4.21	3.67
Coeffcient of Variation	0.033	0.1
Correlation	-0.707	

CHAPTER VI

UNCERTAINTY ANALYSIS RESULTS

This chapter applies the Monte Carlo simulation described in Chapter III to study the uncertainty of model prediction caused by the uncertainty of model parameter. Two questions will be addressed: 1) What significant effect does a model parameter have on model output and what is their relative importance? 2) What relationship exists between the distributions of model input and output?

Data Sampling From the Estimated Distribution

Chapter V has provided the marginal probability distribution for each model parameter. However, in uncertainty analysis we need to use actual parameter values drawn from their distribution. The following paragraphs will present the sampling process and assess the quality of the sampled data.

The results in Chapter V showed that the correlation coefficient between air tortuosity and the adsorption coefficient is -0.707. Therefore, these two parameters cannot be considered independent. The generated data should preserve the means, variances, covariance, and correlations between

these two parameters. Haan (1977) described a procedure which uses principal components to generate a multivariate normal distribution. Figures 56 and 57 show the comparisons of distributions of the two estimated parameters and normal distribution. Apparently, they are not exactly normal distributions because their kurtosises are greater than 3 which is the kurtosis of a normal distribution. Air tortuosity especially, has a relatively greater probability concentration near the mean than does the normal distribution. So the adjusted procedure includes the following steps (Taylor and Bender, 1988):

1. Use the means, variances, and correlation matrix to generate a multivariate normally distributed data X according to the procedure given by Haan (1977).

2. Calculate the cumulative probability level for data X.

3. For each data value of X, use its probability level as a reference and apply it in the distribution estimated from lab data to get Y with the same probability level. Y is then the data generated from the estimated distribution.

10,000 samples were generated for each parameter using this method. They were analyzed again to see if they preserve the means, variances and covariance of the population. All parameter statistics were matched within 0.1%.

Sensitivity Analysis

This section will answer the first question posed at the beginning of this chapter. A statistical sensitivity analysis described in Chapter III was performed to rank model parameters in terms of their contribution to overall error in model predictions.

The 10,000 samples of each parameter generated from the preceding paragraphs were put back in the volatile organic transport model to produce a corresponding model output, the total toluene mass escaping the column. In fact, by testing, a sample size around 3000 is enough to produce stable results. A multiple linear regression analysis was then performed using the standardized parameter samples and model output. If the response surface approximation is sufficiently close to the model over the region of interest evaluated by sufficiently large R^2, standardized partial regression coefficients can be used as estimates of sensitivity.

It has been established that there is interaction between the two model parameters (ξ_g and K_d), which this research has been studying. Doctor (1989) recommended the use of a step-wise regression procedure to deal with mutually dependent input variables. Step-wise regression approximates the model output by sequentially adding or deleting variables to the response surface until adding more variables cannot improve the R^2 criterion substantially.

Since there are only two parameters involved, the first

step conducted the regression by including either parameter, while the second step included both parameters. Table V summarizes the regression results. When the regression includes air tortuosity alone, its R^2 is 0.74 while the standardized regression coefficient is 0.86. When the adsorption coefficient was included alone, R^2 is 0.93 and the regression coefficient is -0.96. When air tortuosity and the adsorption coefficient were both included in the regression, R^2 is 0.999, while the partial regression coefficient is 0.37 for air tortuosity, and -0.71 for the adsorption coefficient.

Table V

Results of Step-wise Regression on Ranks

Variables Included	Standardized Regression Coefficient	R^2
Air Tortuosity	0.86	0.74
Adsorption Coeff.	-0.96	0.93
Air Tortuosity & Adsorption Coeff.	0.37 -0.71	0.999

Including two parameters in the regression apparently does improve the R^2 criterion. A sufficiently large R^2 also demonstrates that the partial regression coefficients are reasonable estimates of the sensitivity index.

These results clearly showed that the adsorption coefficient has the greatest effect on the total mass leaving the column. Therefore, the uncertainty in the adsorption coefficient would contribute more to model output uncertainty than the uncertainty in air tortuosity. The sensitivity of the adsorption coefficient is almost twice that of air tortuosity's sensitivity index. Specifically, one standard deviation change in the adsorption coefficient will lead to a 0.71 standard deviation change in the model prediction. The same degree of change in air tortuosity will lead to a 0.37 standard deviation change in the model prediction.

Notice that the sensitivity index of the adsorption coefficient appears with a negative sign. A positive index means that an increase in the input variable increases the predicted model variable in proportion to the sensitivity index. A negative index means that an increase in the input variable decreases the model prediction in proportion to the index.

These differences will be more apparent when considering the physical meanings of these parameters and the model predictions. The model prediction here is the total mass escaped from the column to the atmosphere. Soil with a larger adsorption coefficient will retain more organic compound in the column. Similarly, when air tortuosity increases, the effective diffusion coefficient increases, which speeds up the diffusive transport process making the organic compound leave

the column faster.

Uncertainty Analysis

This section answers the second question posed at the beginning of this chapter. The Monte Carlo simulation method described in Chapter III was employed to provide information on the variability in model output as a function of uncertainty in the input variables. In other words, the volatile organic model was evaluated at a large number of points in the input parameter space. An empirical pdf for the output variable was constructed from the results.

As described in the sensitivity analysis, the volatile organic transport model was run on the generated 10,000 parameter samples. The total mass leaving the column at 21 hours serves as the output for the uncertainty calculations. Figure 58 shows the probability distribution of model output due to parameter variability. The squares represent the distribution of model output, and the solid line is normal distribution. It appears that the normal distribution describes the distribution of model output well.

Table VI summarizes statistics of the empirical distribution. The minimum value of model output is 439 mg and the maximum value is 616 mg, which gives a range of 177 mg. The mean is 563 mg and the standard deviation is 24 mg. The 95% confidence interval of the mean is 1.50, which indicates the probability is 95% that the interval 525 to 528 contains

the mean.

Uncertainty is usually characterized by the coefficient of variation (C_v) which measures the dispersion. C_v is the standard deviation divided by the mean. In Table IV, the coefficient of variation for parameter air tortuosity is 0.033 and for the adsorption coefficient is 0.1. Table VI shows the coefficient of variation for model output is 0.05. One can discern from these results that with a known range of model parameters, a specific range for model output can be expected.

Table VI

Properties of Total Mass Leaving Column at 21 Hours

Due to Model Parameter Uncertainty

Total Mass Leaving Column (mg)	
Mean	527
Standard Deviation	24
Minimum	439
Maximum	616
Skewness	-0.034
Kurtosis	3.30
Confidence Interval	1.50
Coeff. of Variation (C_v)	0.05
Number of Cases	10000

However, it would be more conclusive to quantify how the change of parameters' uncertainties would affect the uncertainty of model output. A simple way to quantify the relationship between parameter uncertainty and model output uncertainty is to run the transport model on more parameter profiles. These parameter profiles will have the same mean but different standard deviations.

Standard deviations were chosen so that the C_v varies between 0.1 and 0.6. As the standard deviation increases, negative numbers occur in the generated parameter profiles. Since the negative sign contradicts the physical meaning of model parameters, the insignificant tails of parameter distributions were cut off to assure positive numbers for parameter profiles. However, when standard deviation is large enough, negative occurrences increase effectively and the tail becomes significant, which is why the C_v was chosen under 0.6.

The transport model was run on these new parameter profiles. Figures 59 through 68 show the probability distributions of new parameter profiles and corresponding model output resulting from these new runs. Since the same standard deviations were used to generate both air phase tortuosity and the adsorption coefficient, distributions of these two parameters look very much the same except the location of the mean. Thus only distributions of air phase tortuosity were plotted. As the coefficients of variation of model parameters increase, the probability distribution of

model output disperses as well, which shows increasing uncertainty. Table VII summarizes the coefficient of variation of model output and corresponding C_v of the parameters. As the C_v of the parameter increases from 0.1 to 0.6, the model output's C_v increases from 0.037 to 0.293. Figure 69 illustrates this relationship. The solid line represents the corresponding regression equation:

$$MC_v = 0.574 \times PC_v - 0.00824 \qquad\qquad \textbf{(86)}$$

where MC_v is the coefficient of variation for model output, PC_v is the coefficient of variation for model parameters. PC_v is the combined factor for both air phase tortuosity and the adsorption coefficient. Figure 69 shows the regression line fits the calculated data very well with a R^2 of the regression of 0.986.

Table VII

Impact of Model Parameter Uncertainty

on the Uncertainty Of Model Output

C_v of Model Parameters	C_v of Model Output
0.1	0.037
0.2	0.111
0.3	0.172
0.4	0.230
0.5	0.268
0.6	0.293

Chapter VII

SUMMARY AND CONCLUSIONS

Summary

From the dawn of the industrial revolution, pollution has been an increasingly major environmental concern. Water pollution is especially disastrous not only because of health factors but also because contaminants in water can easily migrate to surrounding areas. Cleaning up water pollutants can be a very difficult and costly task. This is particularly true when concerning ground water pollution because of its proximity. Therefore, it is very beneficial for decision makers to have scientific suggestions in order to identify areas of high concentrations and movement of pollutants so that water can be pumped out for treatment or insitu remediations can be implemented. This is where the importance of volatile organic transport modeling can be clearly seen, because mathematical modeling is an important tool for predicting the fate and movement of pollutants in ground water.

Volatile organic transport models often involves many parameters controlling various transport and phase transferring processes. Apparently, the usefulness of a model

91

depends on the precision of its predictions about contaminant movement and the confidence of this precision. The precision of model predictions depends on the ability to determine model parameters precisely. The confidence depends on the ability to quantify model uncertainty. However, since volatile organic transport in subsurface water is a complex multiphase phenomenon, attention has been focused on understanding its transport processes by establishing mathematical models, thus leaving plenty of room for researchers to investigate model application. It was my intention to explore this area by studying model parameter estimation and model uncertainty quantification.

A multiphase compositional organic transport model (Equations 28 to 33) developed by Baehr (1987) was used in this research. Of the many parameters involved in this model, four coefficients that control major transport and phase transferring processes were of interest. They included air phase tortuosity (ξ_g), water phase tortuosity (ξ_l), Henry's constant (K_h), and the adsorption coefficient (K_d). These parameters are lumped together in the model to form two other parameters (a and b). The distributive nature of these parameters along with the lumping make them very difficult to estimate from experimental data. Moreover, because they are lumped together, the change in one parameter can be offseted by the change in another parameter. Estimation would be even less reliable when measurement errors are introduced.

Therefore, methodologies which can estimate these parameters and quantify the accuracy of model predictions are in need in order to make models such as this useful.

A systematic methodology which employs well established mathematical techniques has been developed to achieve this goal. The purpose of this methodology is to first determine each model parameter's optimal estimate along with its probability distribution, and then quantify the uncertainty of model predictions due to the errors in parameter estimation, using the estimated parameter distributions.

A typical procedure for estimating a parameter is to collect field or laboratory data and then analyze the error between that data and model outputs under certain values of the required parameters, so that the desired parameter minimizes the error. This research used lab data measured by Yu (1995), which included toluene vapor concentration profiles cross the column at different times as the gas escaped from the top.

The first part of this methodology, parameter estimation algorithm employed Bayesian statistical inferences to accomplish two tasks. First, it has provided the optimal estimates for each parameter which minimizes the errors between model outputs and observed data, using the modified adaptive simplex method. For example, it was found in Chapter V that under the experimental condition used in this research, the optimal estimate of air tortuosity is 0.42, and the

optimal estimate of the adsorption coefficient is 0.39. These estimates can then be used in this model to predict the movement of contaminants when applied to other similar cases. Secondly, this algorithm has produced a probability distribution for each parameter, which reflects reality more than the optimal estimates alone. With this probability distribution, we can tell the mean, the standard deviation, and the possible range of a parameter. For example, we can say, from Table IV in Chapter V, that under the experimental condition used in this research, the mean of air phase tortuosity estimate is 0.42 with a standard deviation of 0.01 which tells how confident this estimate is. Likewise, the mean of the adsorption coefficient estimate is 0.39 with a standard deviation of 0.04. The marginal distribution of each parameter was obtained by integrating the joint distribution of these two parameters using Monte Carlo integration. This method was used because of its simplicity.

Along with these estimates, several relevant concerns have been addressed, such as reducing dimensionality of parameter space, evaluating the effect of observed data on the behavior of objective function, and quantifying the uncertainty of parameter estimation induced by errors in observed data. The first concern needs to be addressed because high dimensional parameter space increases the chance of ill-posedness as well as the difficulty of computation. The other two concerns were addressed because they serve as a

guide to experiment about how much information is sufficient and what quality the measured data should meet, in order to get meaningful parameter estimates.

A preliminary sensitivity analysis was conducted to describe the impact that the change of each parameter had on model outputs. It was found that water phase tortuosity had little impact on model outputs, thus it was excluded from the estimation process to reduce dimensional difficulty.

The question about how much observed data are sufficient to construct a well behaved objective function has been addressed by analyzing the characteristics of error function using both single (one set of measured toluene gas concentration profile) and two model outputs (two sets of measured toluene gas concentration profiles). When single model output was used, the error contours plotted on every two-parameter space show valleys containing many local minima which indicate high correlation among the parameters. Therefore, information brought in the analysis was not sufficient to identify the optimal parameter.

When two model outputs were used, the contour map for air phase tortuosity and the adsorption coefficient shows a global minimum which means the correlation between air phase tortuosity and the adsorption coefficient was reduced by bringing in the second set of measured data. However, Henry's constant and the adsorption coefficient remained correlated. This is because these two parameters are negatively correlated

due to their physical definition. Since the adsorption coefficient is more uncertain and harder to measure than Henry's constant, this research eliminated Henry's constant to reduce correlation difficulty, leaving two parameters for further analysis.

To find out how experimental error would affect parameter estimation, hypothetical data with different error standard deviation were used. It was then discovered when the error standard deviation in observed data was less than 0.35 mg/l, parameter estimates are stable and reliable. This result serves as a guide to analyze and decide if lab data are usable.

The second part of the methodology developed in this research (i.e. uncertainty analysis algorithm) has accomplished two tasks as well. First, through sensitivity analysis, it was found that the sensitivity coefficients of the adsorption coefficient and air phase tortuosity are 0.7 and 0.3 respectively. These numbers indicate that the adsorption coefficient has more effect on model outputs than air phase tortuosity does. Such a result can help a modeler to pay more attention to the more sensitive parameter when he does modeling. Secondly, through joint uncertainty analysis, this algorithm produced an empirical probability distribution corresponding to the parameter distributions estimated earlier. For example, when the standard deviations of the estimated adsorption coefficient and air phase tortuosity are

0.01 and 0.04 respectively, the standard deviation of model prediction is 24 mg. Further more, it provides a linear regression equation (Equation 69 in Chapter VI) to quantify the relationship between model prediction uncertainty and parameter uncertainty.

Specifically, 10,000 samples for each parameter were generated from the estimated marginal distributions and were put back in the transport model to produce corresponding model outputs. Parameter sensitivity analysis was conducted by performing a multiple linear regression using these generated parameter samples as independent variables and produced model outputs as dependent variables. The regression coefficients of this multiple linear regression approximate sensitivity index of parameters. An empirical probability distribution of model outputs obtained from these generated parameter samples was analyzed to quantify its uncertainty (C_v) due to model parameter uncertainty. Parameters were resampled from the hypothetical distribution with different C_v's to quantify the relationship between parameter uncertainty and model output uncertainty.

Conclusions

Of the four parameters studied, water phase tortuosity was found insensitive to model output because of the small molecular diffusion coefficient. The other three parameters were highly interactive because of the way the model was

structured. The numerical difficulty induced by parameter interaction could be reduced to some extent by bringing in more information, such as using two model outputs.

Errors in observed data affect the accuracy of parameter estimation significantly. For the point estimates, the increasing variance of observation error tends to cause the optimal estimates to drift more from the true value, and the confidence interval to increase. It was concluded that when the standard deviation of observation error was less than 0.35 mg/l, the estimated results could be considered reliable. However, an interesting point is that we can only refer to a tendency here, because different realization of data will result in different accuracy. It is possible that observations with a larger error variance will result in a more accurate estimation. On the other hand, the marginal distributions of parameters demonstrated a steadily increasing variance which directly corresponds with the variance of observation error. Different realization of data affects the mode of the distribution but not the variance of the distribution.

Sensitivity analysis showed that the model output is more sensitive to the parameter controlling the adsorption process than to the parameter controlling the diffusion process. Uncertainty analysis showed that the model output distributes normally when the estimated parameter distribution was used. Surprisingly, its uncertainty is less than the parameters'

uncertainty which is quantified by C_v. Finally, A linear relationship between parameter uncertainty and model output uncertainty was quantified.

Recommendations for Further Studies

This research is the first study for parameter estimation and uncertainty quantification in volatile organic transport modeling. The modeling conditions were simplified to keep consistent with experimental conditions being carried out by others. The following are suggestions for future investigation:

1. Only one multiphase organic transport model was used in this study. Therefore, the results of this study might be model specific. Similar studies conducted for other models will still be meaningful. Subsurface multiphase contaminant transport is such a complex phenomenon that it is not realistic to expect a numerical model to represent its every attribute. If several models which favor different driving forces are available, the methodology developed in this research can be applied to them to show how model parameters behave under different circumstances.

2. More complicated conditions could be included, such as two dimensional flow, transient water flow, and gas advection. Of particular interest in terms of gas advection would be the analysis of conditions when vapor and liquid transport were equal in magnitude.

3.   This study emphasized the impact of model parameter uncertainty on model prediction uncertainty, not a validation of the transport model.   Verifying model predictions with experimental data could be an interesting topic.   The exact form of the vapor tortuosity coefficient for both very dry and almost saturated soil deserves additional consideration.

4. Since parameter estimation process repeatedly uses the numerical transport model, the time efficiency of the numerical model is critical to the time efficiency of parameter estimation.   Therefore, improvement on this aspect is worth mentioning.
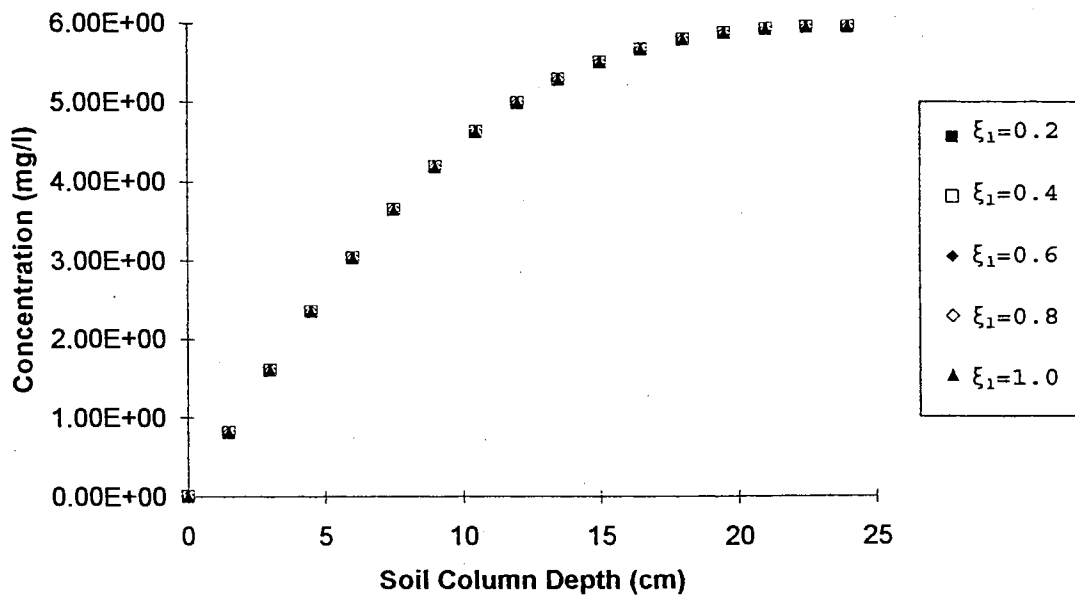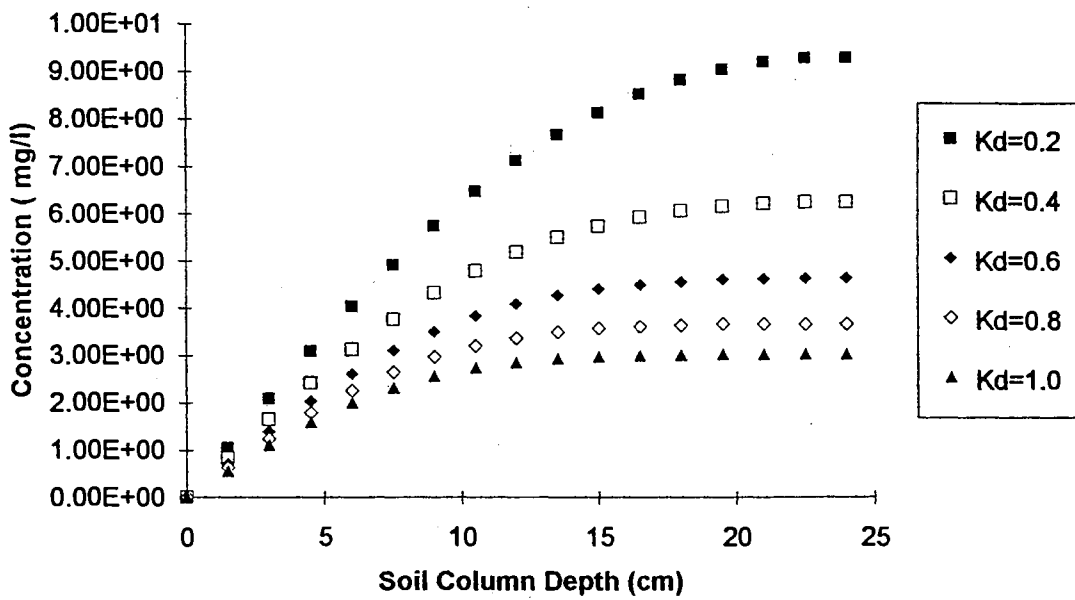
Figure 1. Simulated toluene air phase concentration
profiles for different water phase
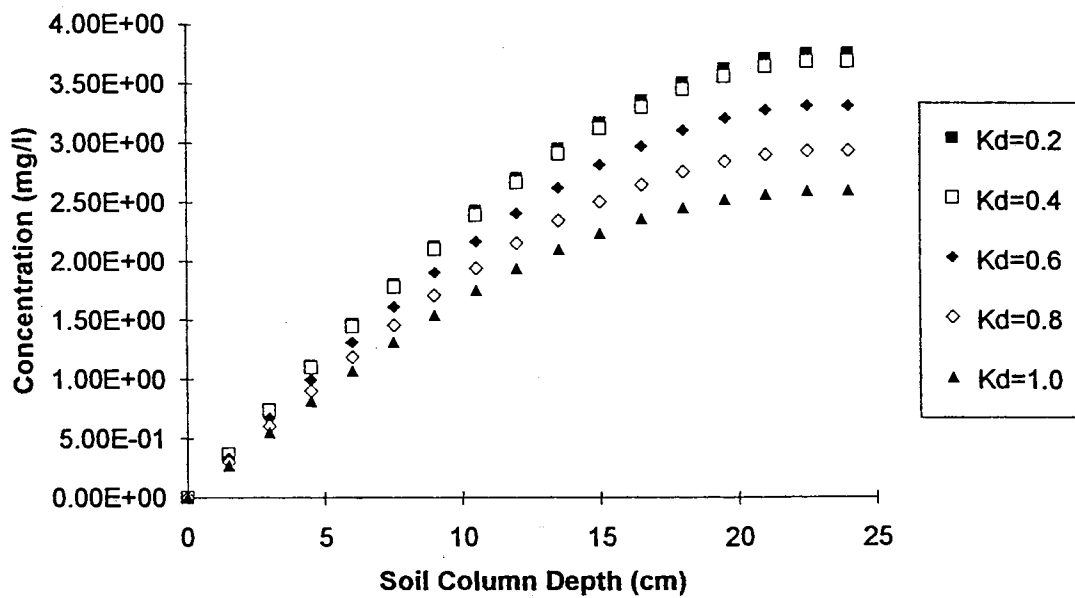tortuosity (ξ_1) at five hours.
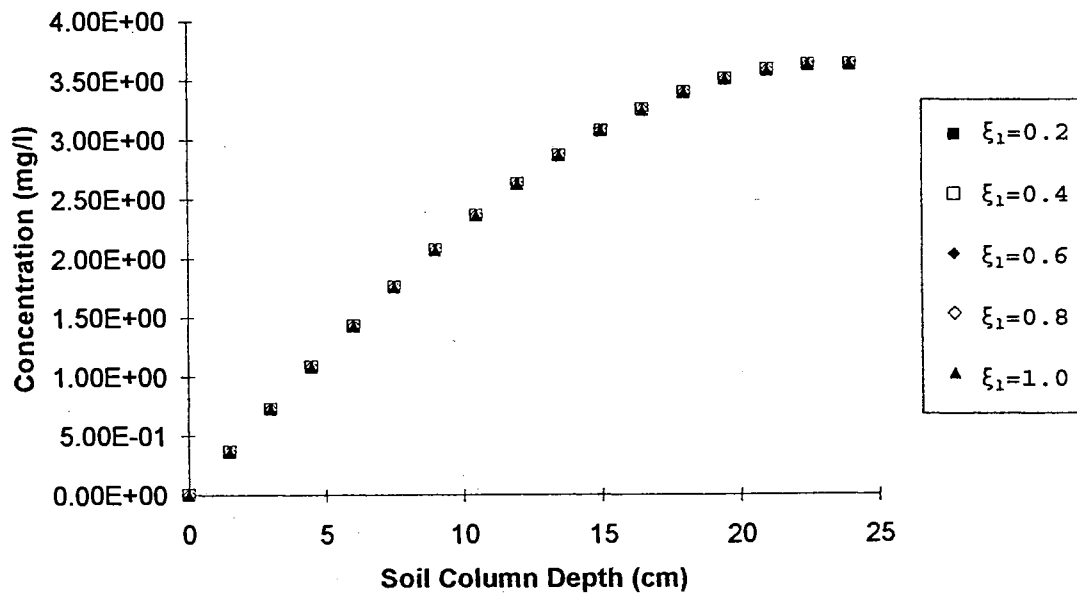
Figure 2. Simulated toluene air phase concentration profiles for different water phase tortuosity (ξ_1) at 21 hours.
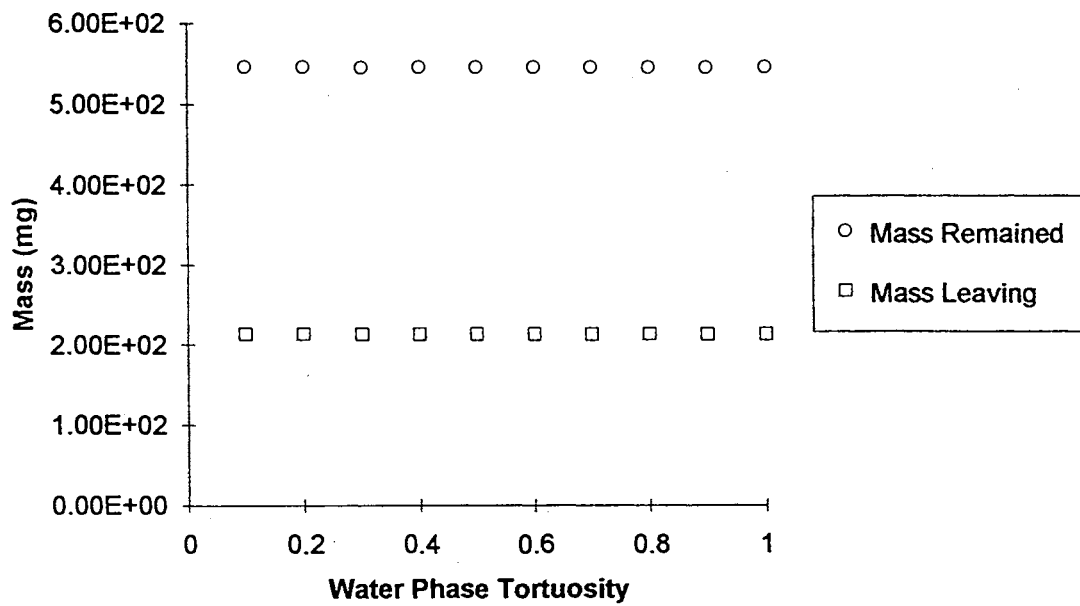
Figure 3. Simulated total toluene mass escaped from
the soil column surface with different
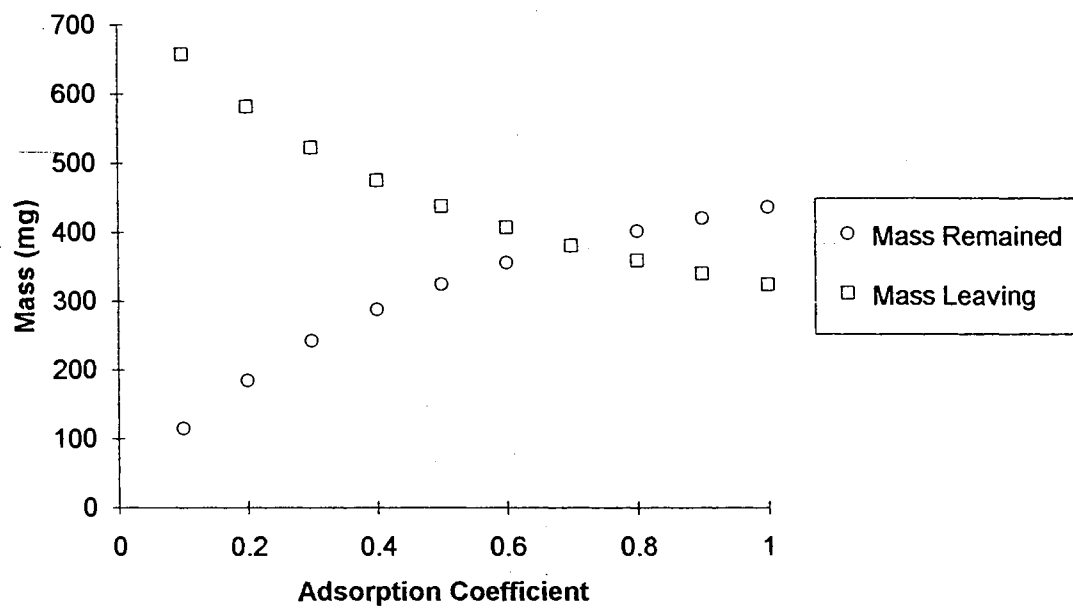water phase tortuosity (ξ_1) at five
hours.
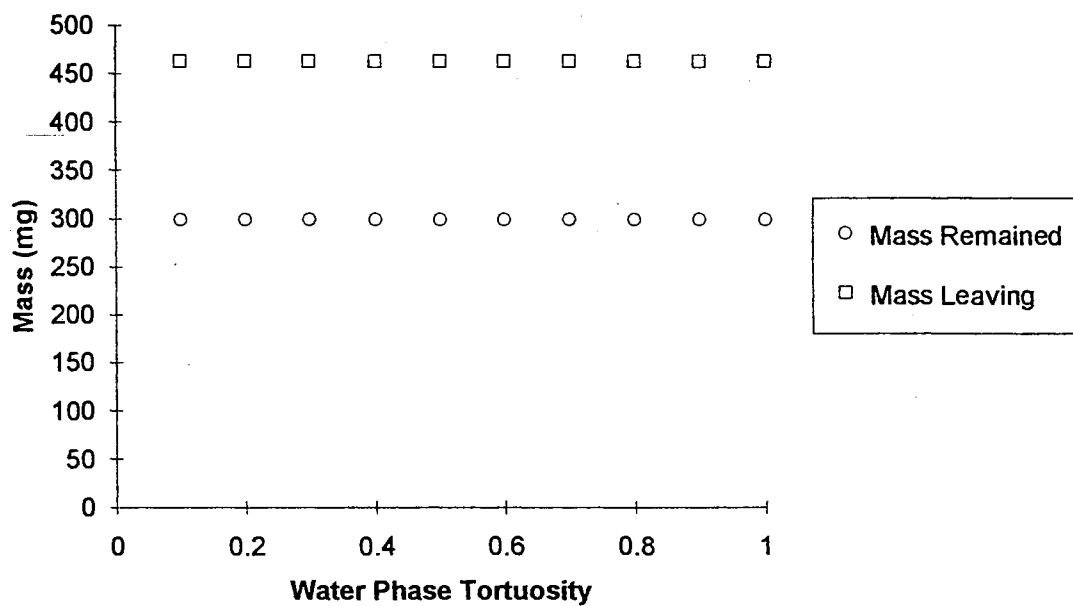
Figure 4. Simulated total toluene mass escaped from
the soil column surface with different
water phase tortuosity (ξ_1) at 21 hours.
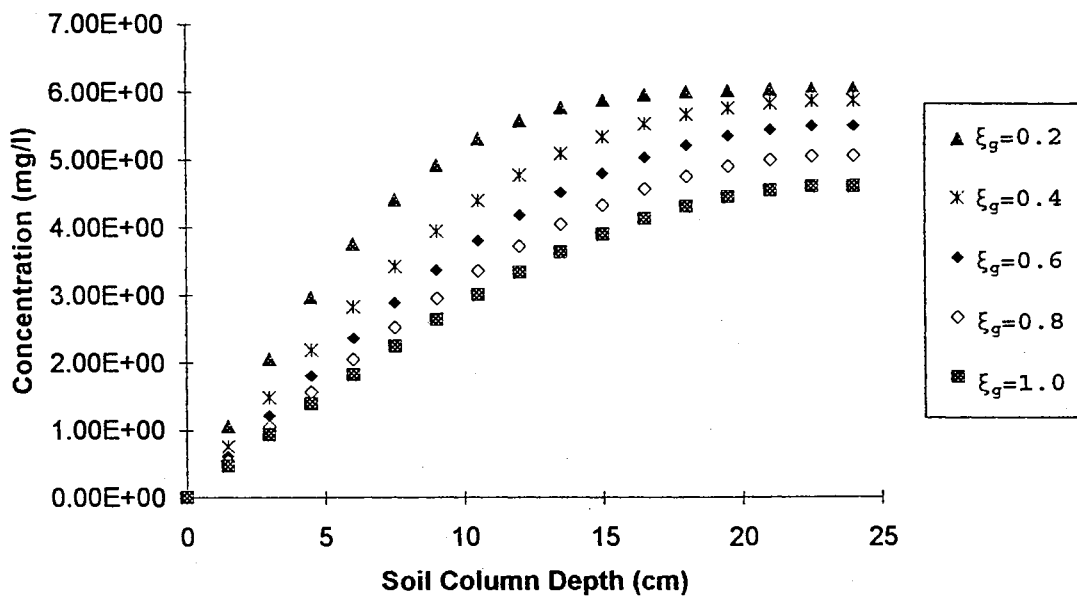
Figure 5. Simulated toluene air phase concentration profiles for different air phase tortuosity (ξ_g) at five hours.
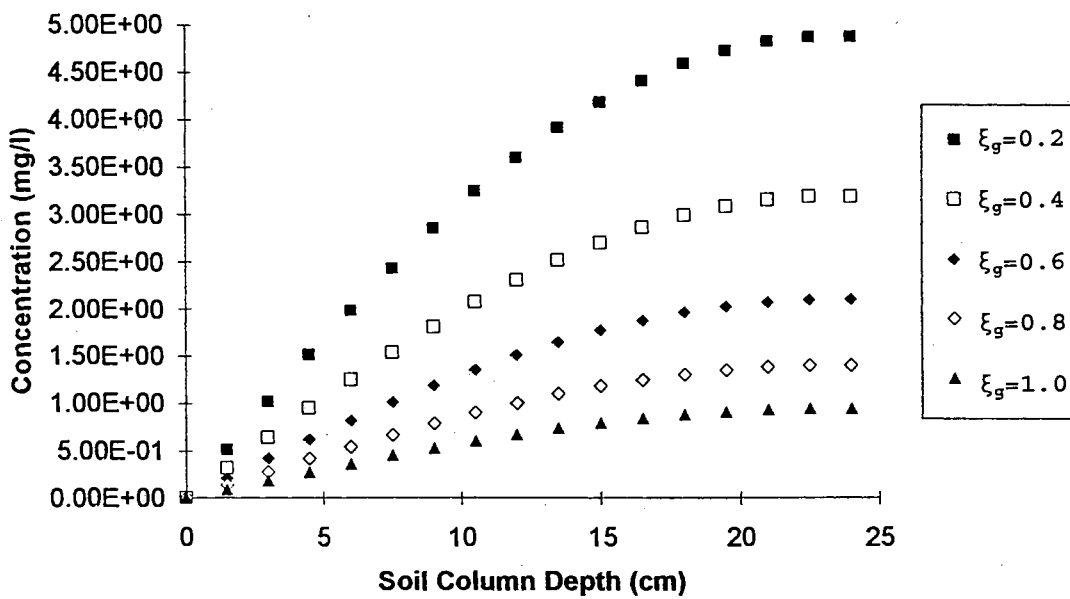
Figure 6. Simulated toluene air phase concentration
profiles for different air phase
tortuosity (ξ_g) at 21 hours.
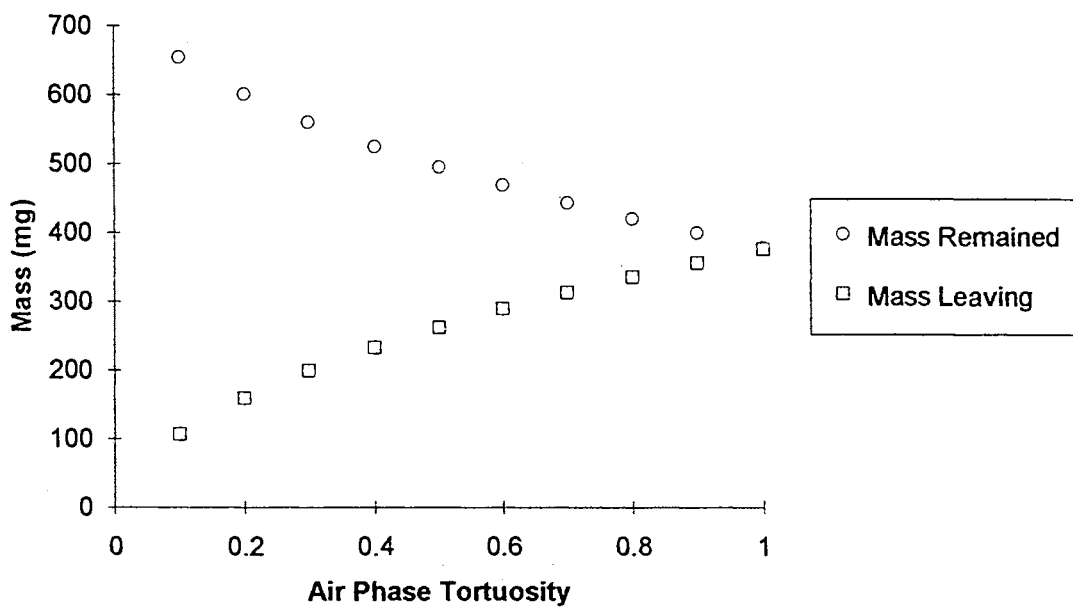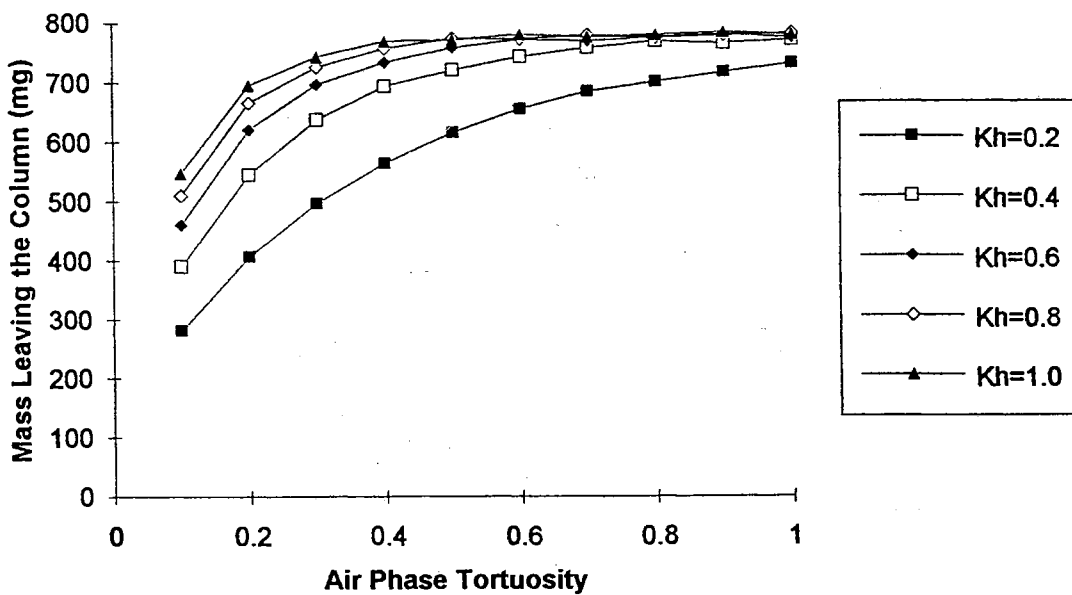
Figure 7.   Simulated total toluene mass escaped from
            the soil column surface with different
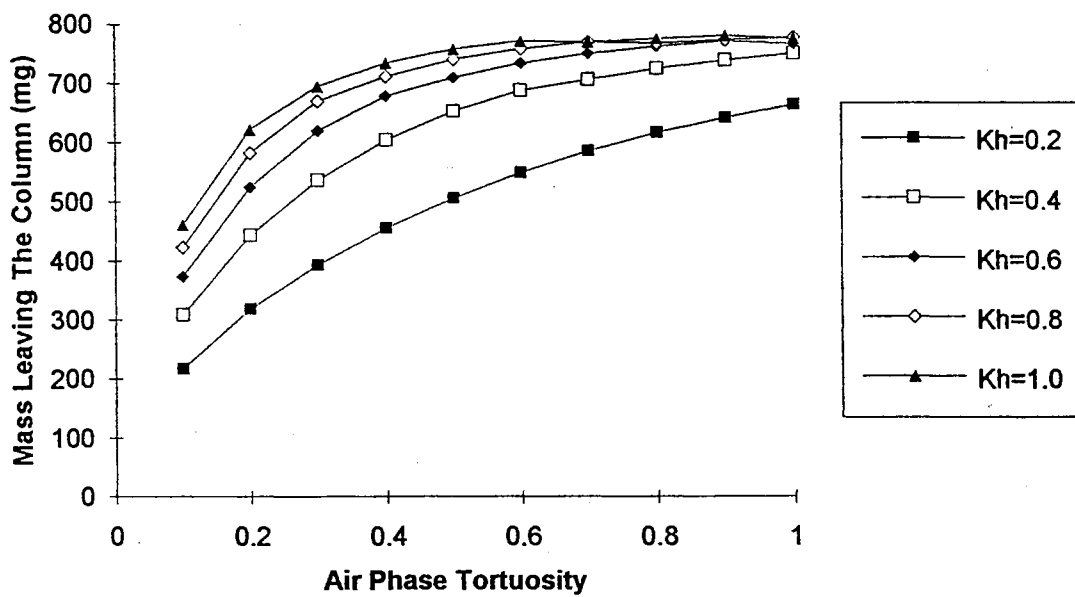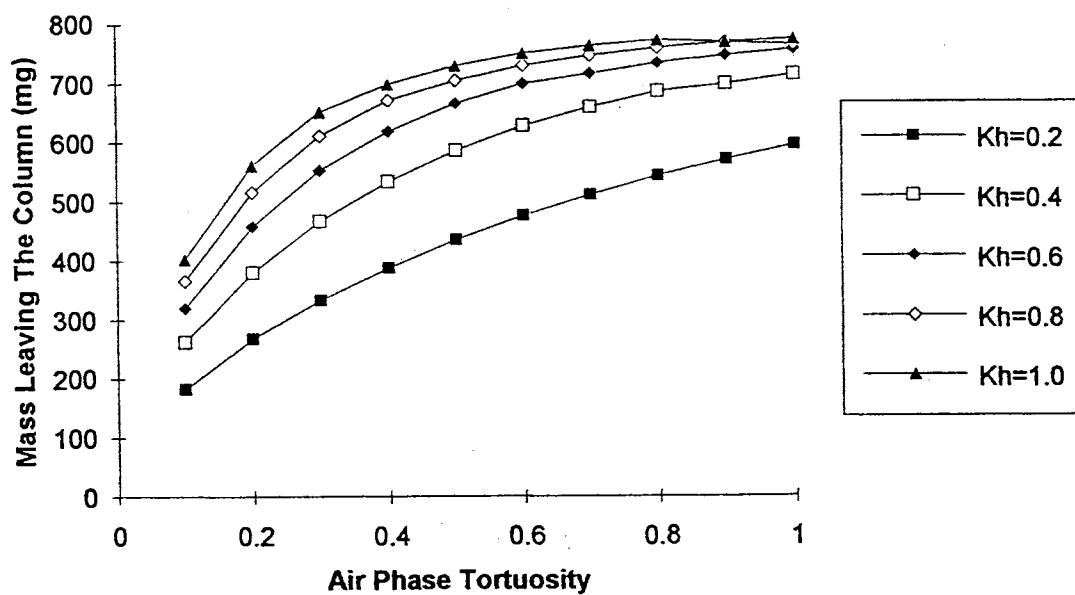   air phase tortuosity (ξ_g) at five hours.
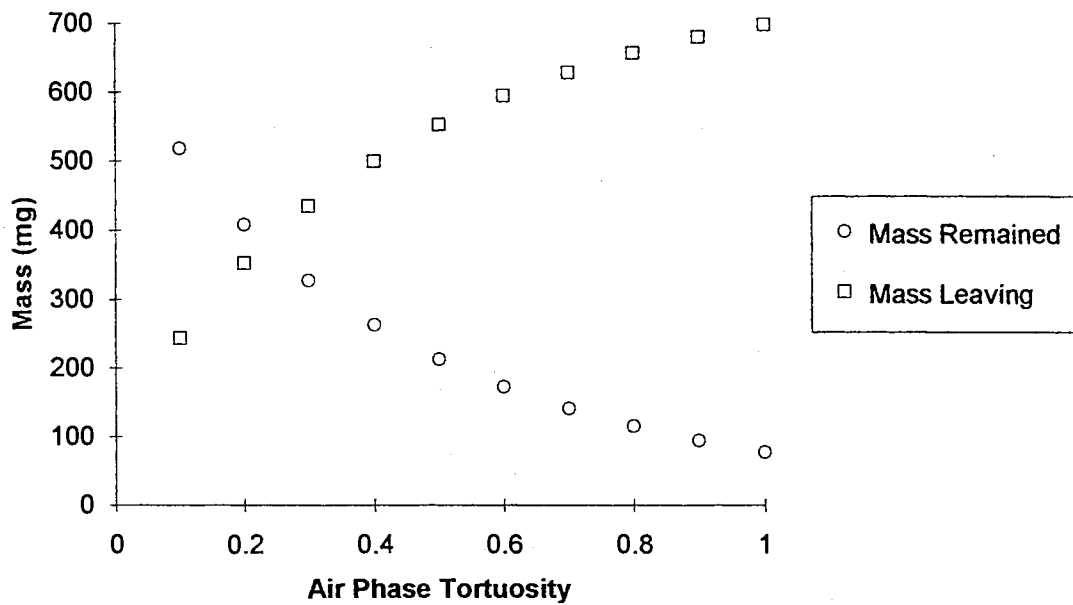
Figure 8. Simulated total toluene mass escaped from
the soil column surface with different
air phase tortuosity (ξ_g) at 21 hours.
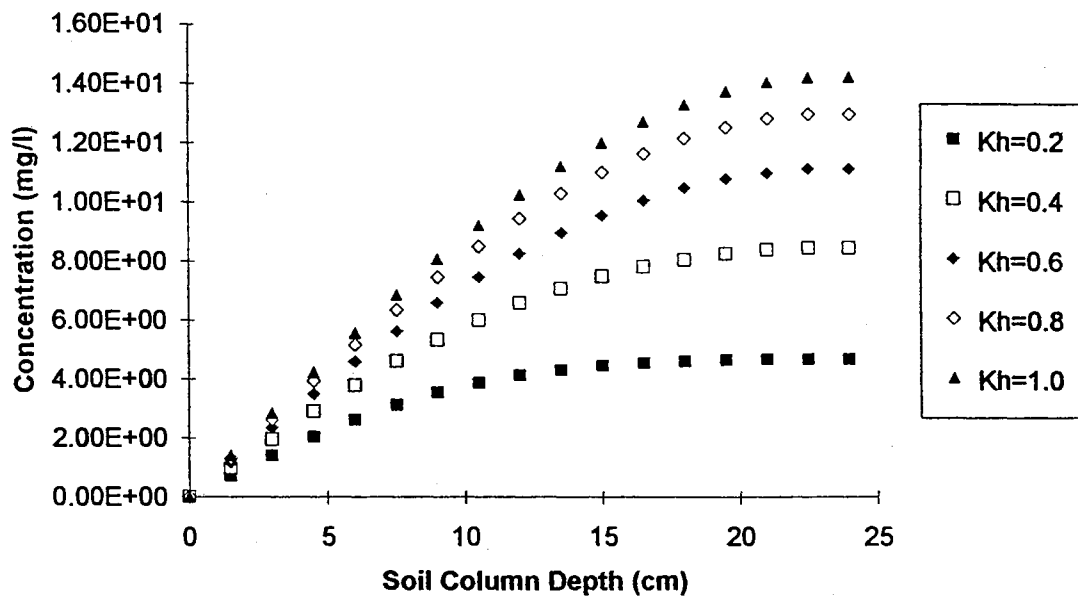
Figure 9.   Simulated toluene air phase concentration
profiles for different Henry's law
constant (K_h) at five hours.
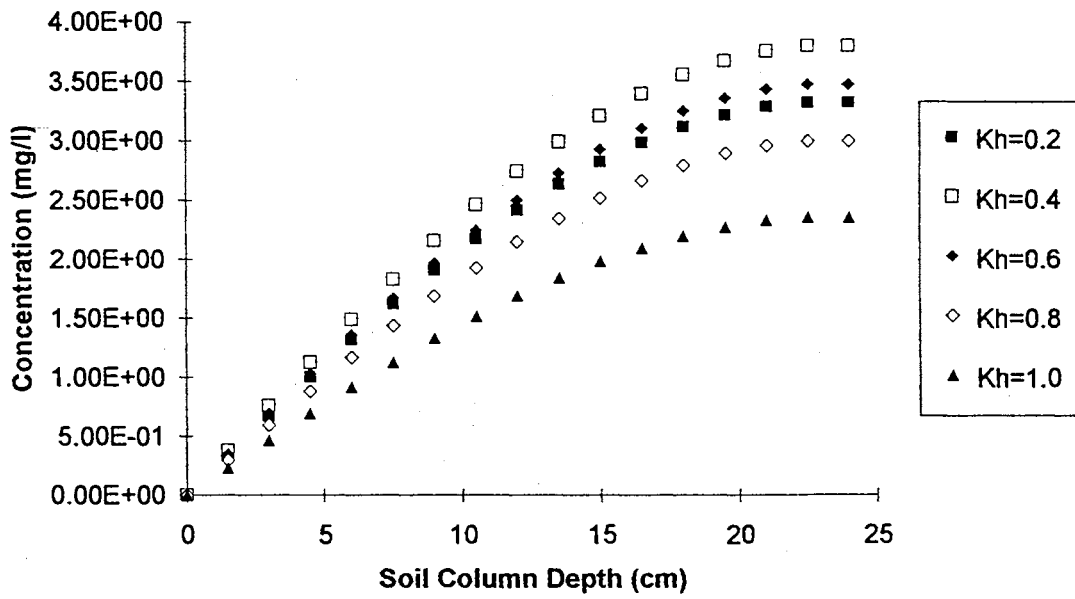
Figure 10. Simulated toluene air phase concentration
          profiles for different Henry's law
          constant (K_h) at 21 hours.

110
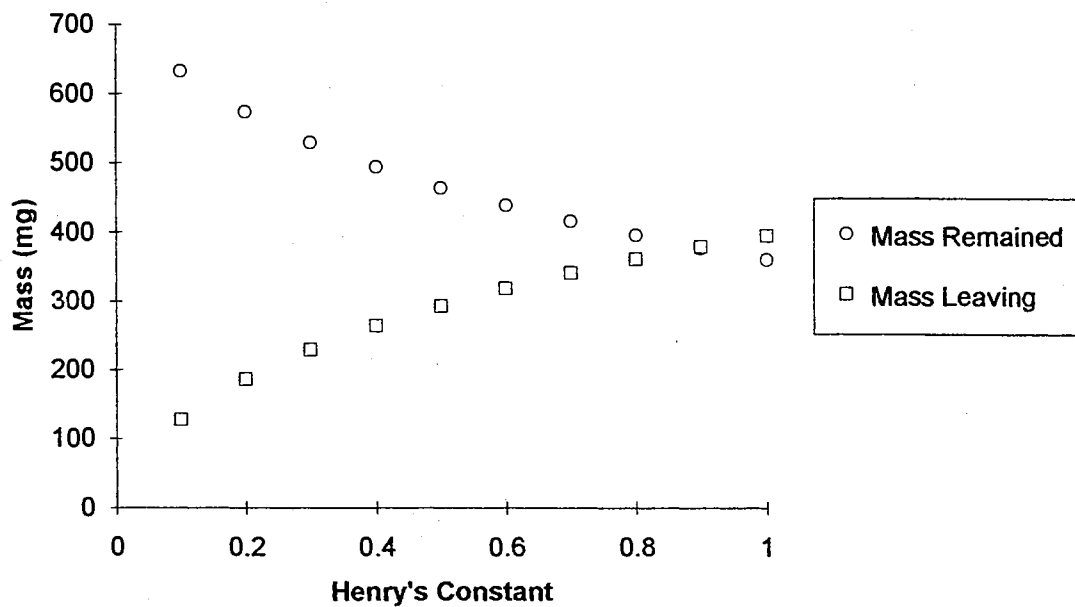
Figure 11.   Simulated total toluene mass escaped from
             the soil column surface for different
             Henry's law (K_h) constant at 5 hours.
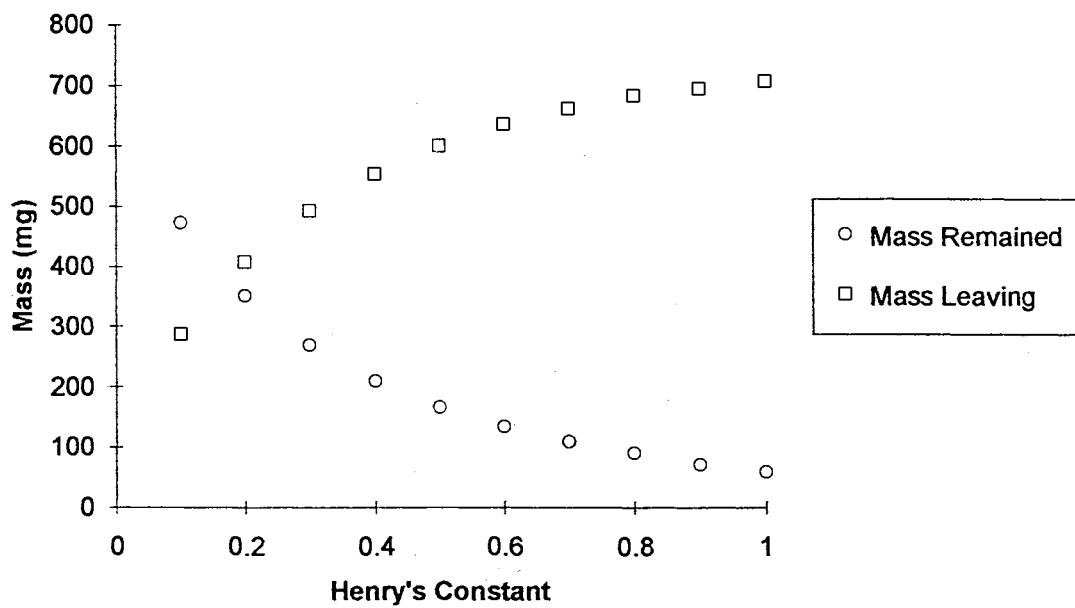
Figure 12. Simulated total toluene mass escaped from
           the soil column surface for different
           Henry's law (K_h) constant at 21 hours.

Figure 13. Simulated toluene air phase concentration profiles for different adsorption coefficient (K_d) at five hours.

Figure 14.   Simulated toluene air phase concentration
             profiles for different adsorption
             coefficient (K_d) at 21 hours.
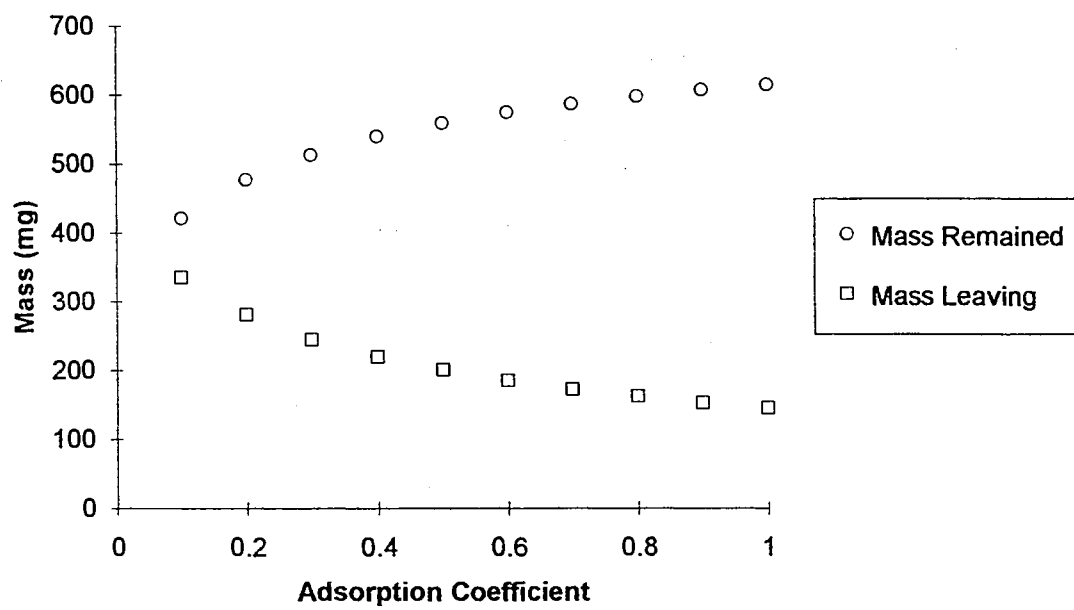
Figure 15. Simulated total toluene mass escaped from the soil column surface with different adsorption coefficient (K_d) at 5 hours.

Figure 16.   Simulated total toluene mass escaped from
             the soil column surface with Different
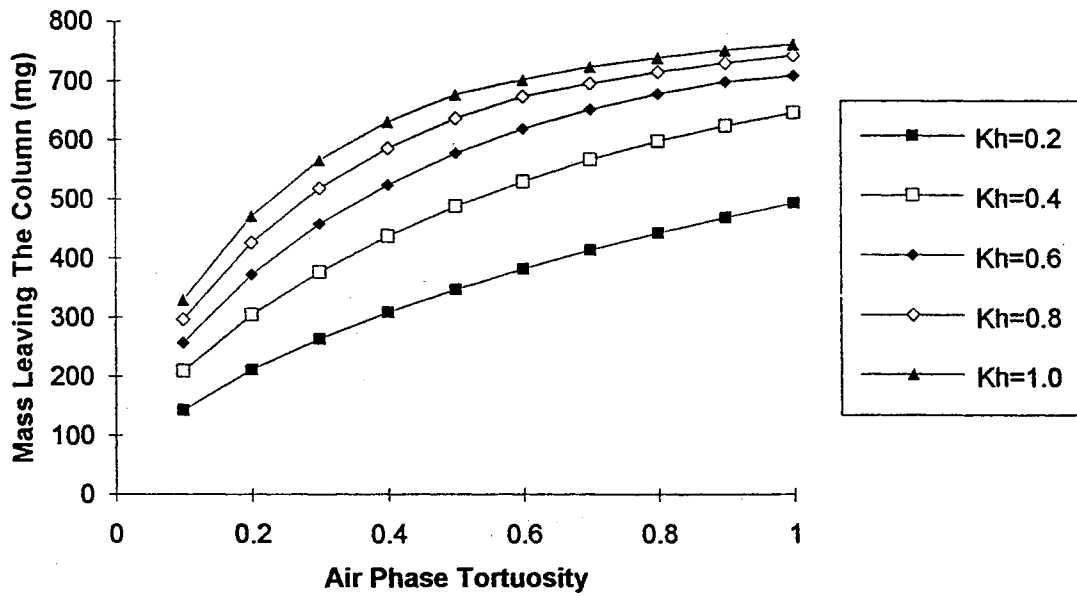   adsorption coefficient (K_d) at 21 hours.

Figure 17.   Simulated total toluene mass escaped from
             the soil column surface at 21 hours with
             different air phase tortuosity (ξ_g) and
             different Henry's constant (K_h) when
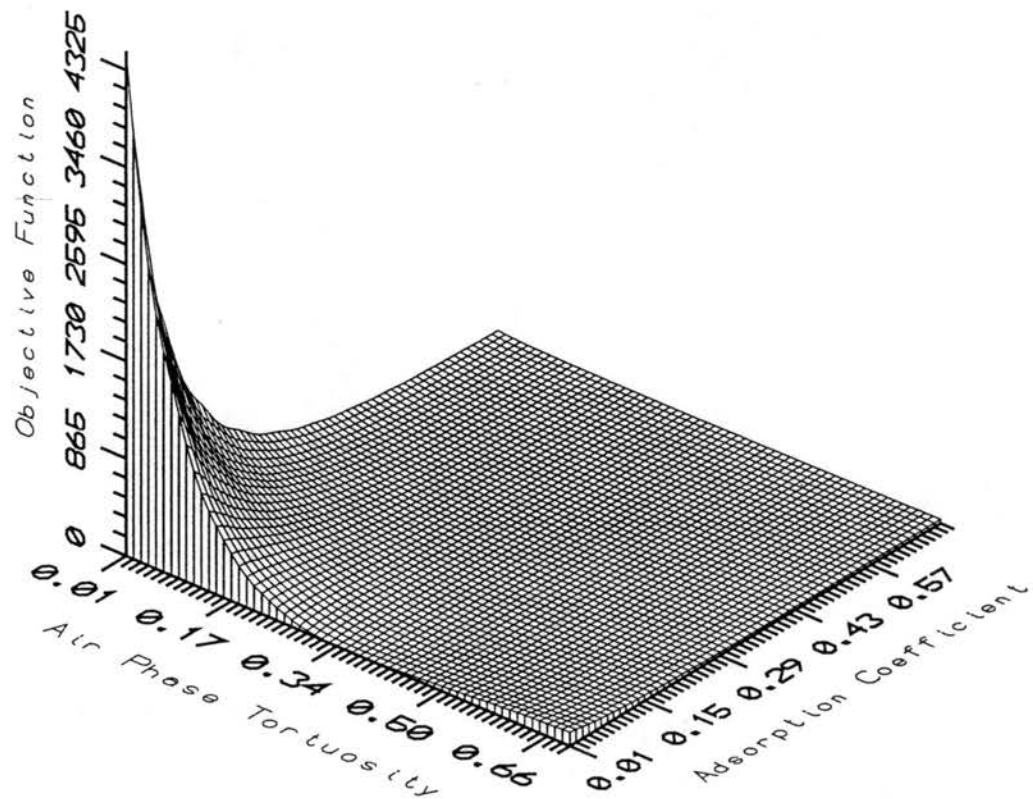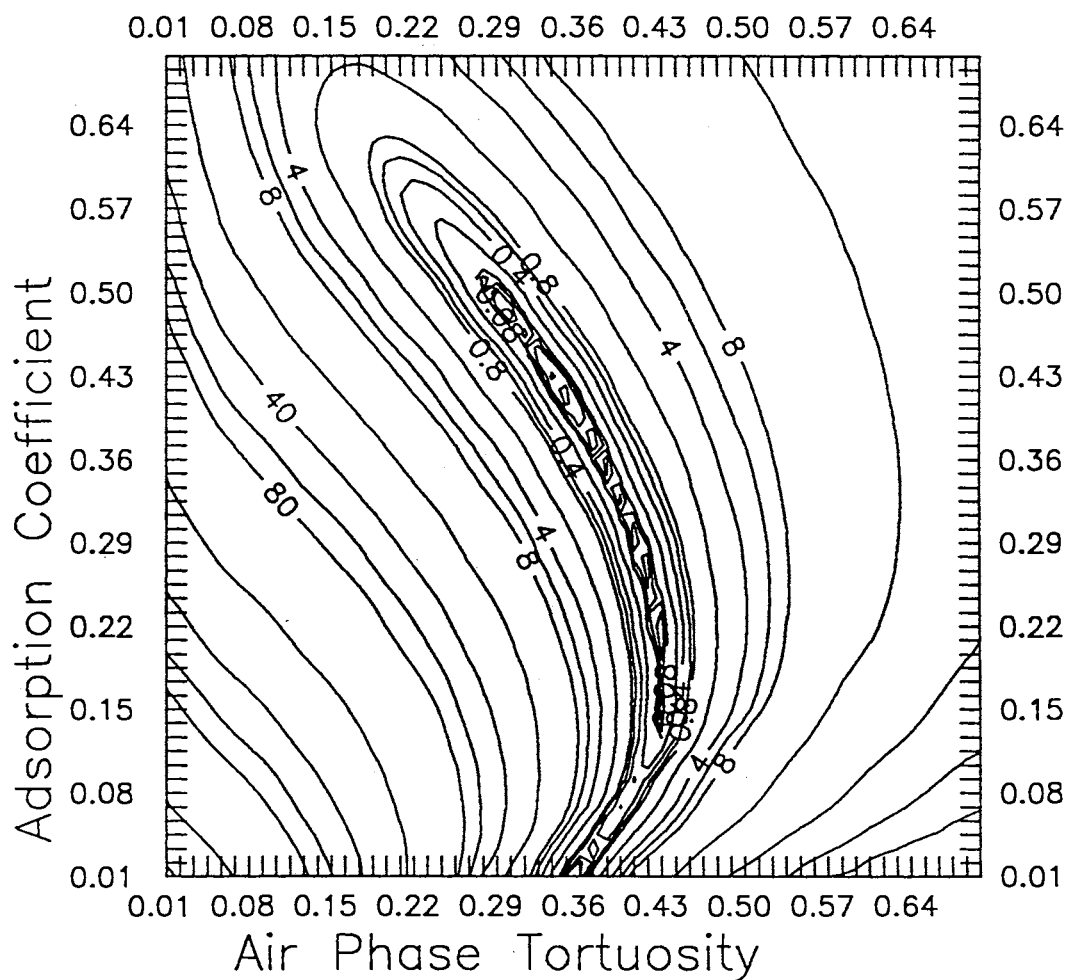   adsorption coefficient (K_d) is 0.2.

Figure 18.   Simulated total toluene mass escaped from
            the soil column surface at 21 hours with
            different air phase tortuosity (ξ_g) and
            different Henry's constant (K_h) when
            adsorption coefficient (K_d) is 0.4.

118

Figure 19.   Simulated total toluene mass escaped from
             the soil column surface at 21 hours with
             different air phase tortuosity (ξ_g) and
             different Henry's constant (K_h) when
             adsorption coefficient (K_d) is 0.6
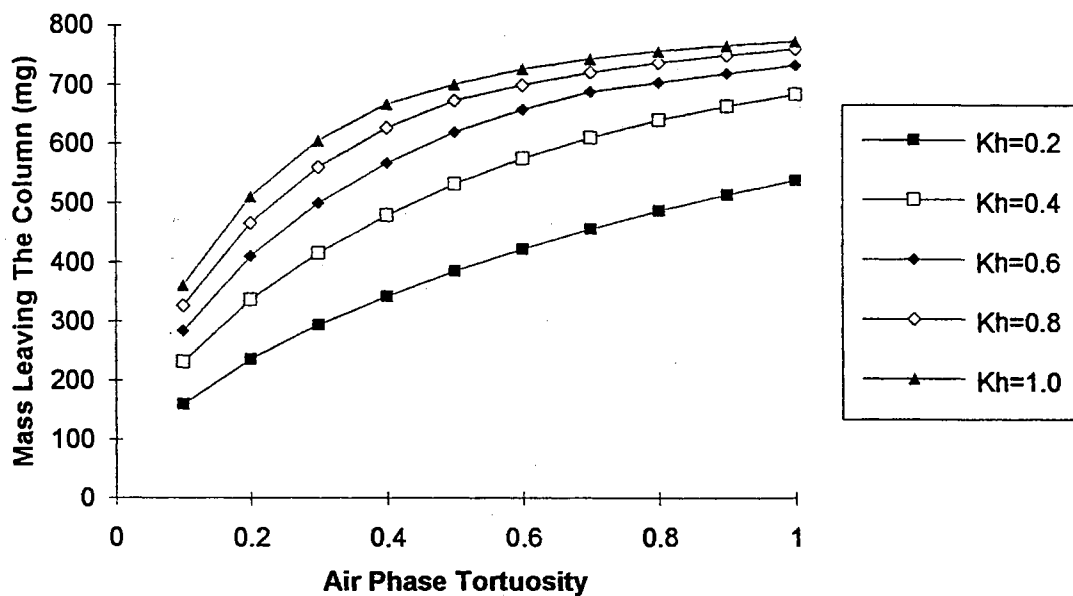
Figure 20.   Simulated total toluene mass escaped from
the soil column surface at 21 hours with
different air phase tortuosity (ξ_g) and
different Henry's constant (K_h) when
adsorption coefficient (K_d) is 0.8.

Figure 21.   Simulated total toluene mass escaped from
            the soil column surface at 21 hours with
            different air phase tortuosity (ξ_g) and
            different Henry's constant (K_h) when
            adsorption coefficient (K_d) is 1.0.

Figure 22. Objective function response surface in air
phase tortuosity (ξ_g) and adsorption
coefficient (K_d) space using one model
output (Henry's constant K_h set to
0.26).

Figure 23. Objective function contours in air phase
tortuosity (ξ_g) and adsorption
coefficient (K_d) space using one model
output (Henry's constant K_h set to 0.26
and nonuniform contour lines selected
for clarity).
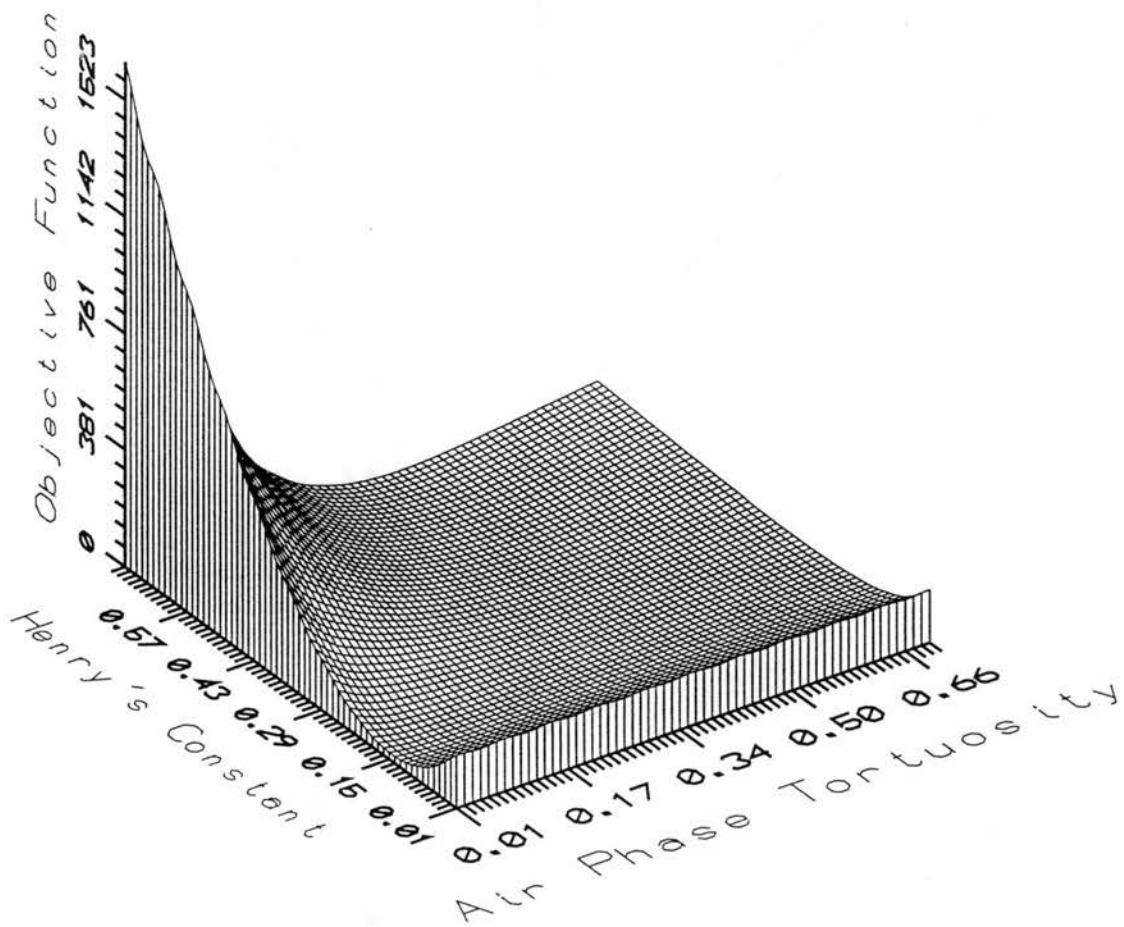
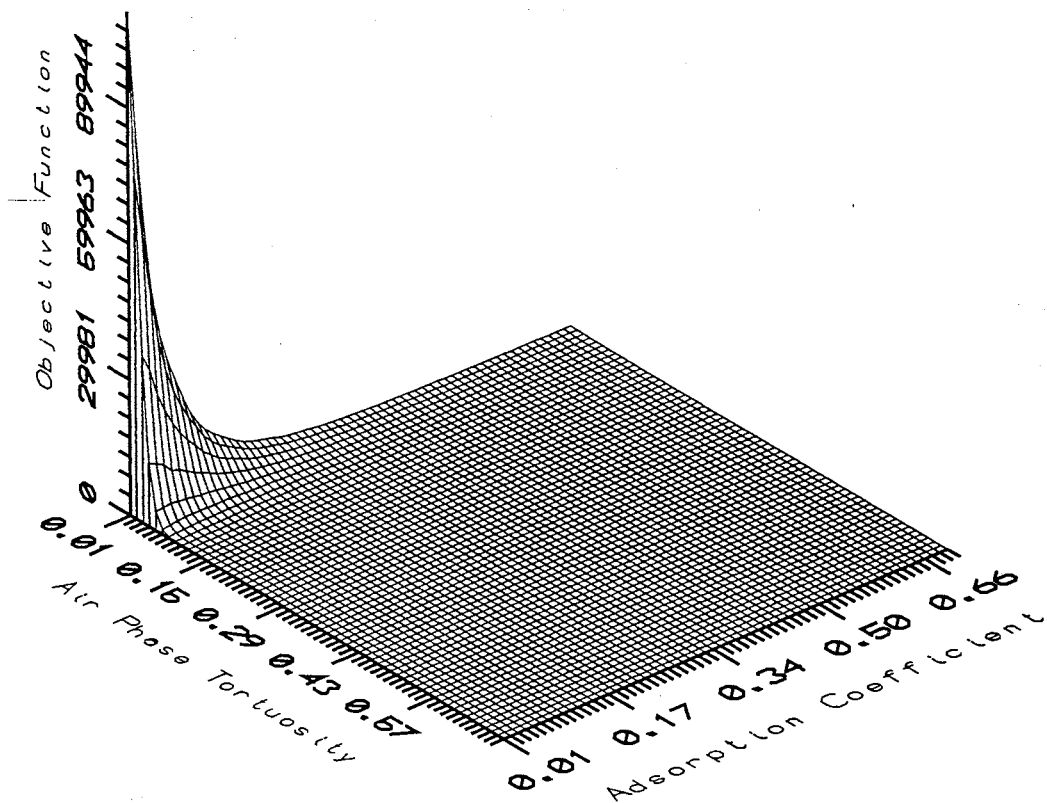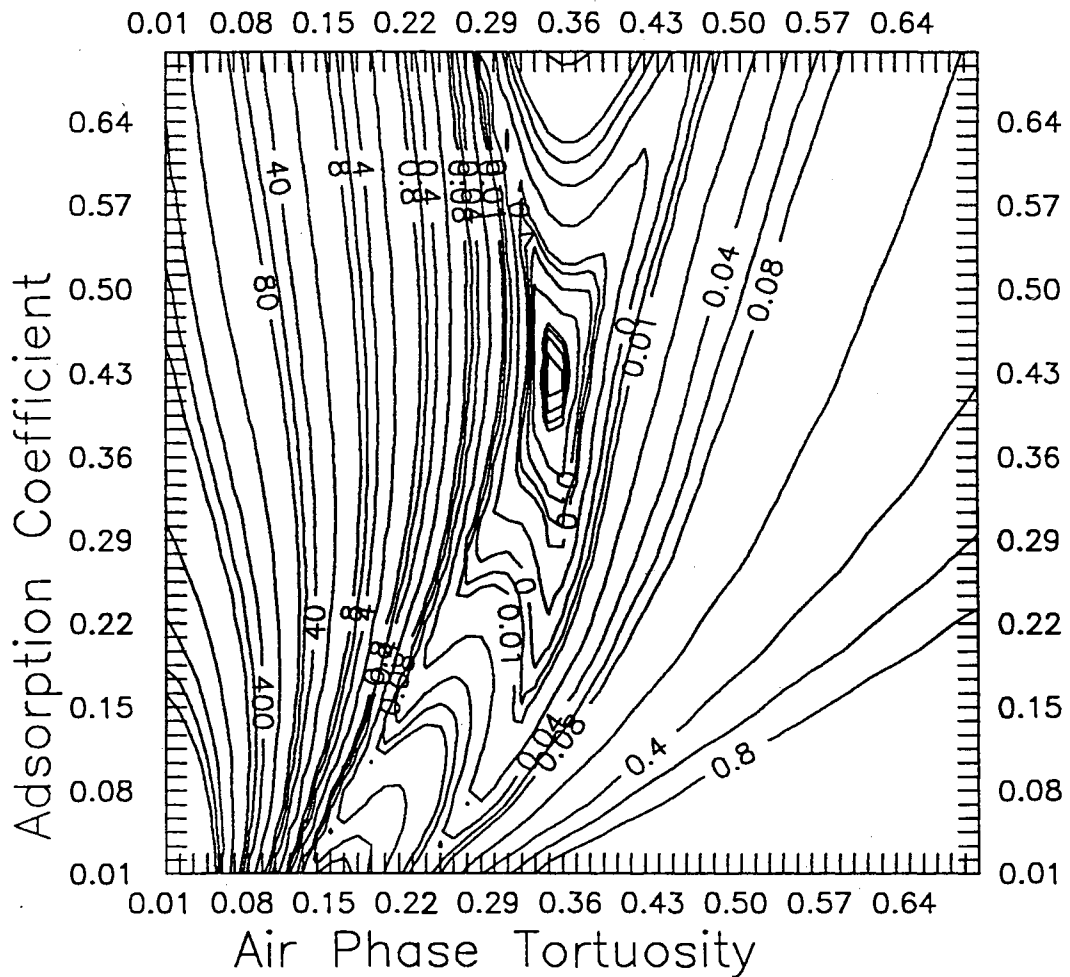Figure 24. Objective function response surface in air phase tortuosity (ξ_g) and Henry's constant (K_h) space using one model output (adsorption coefficient K_d set to 0.43).
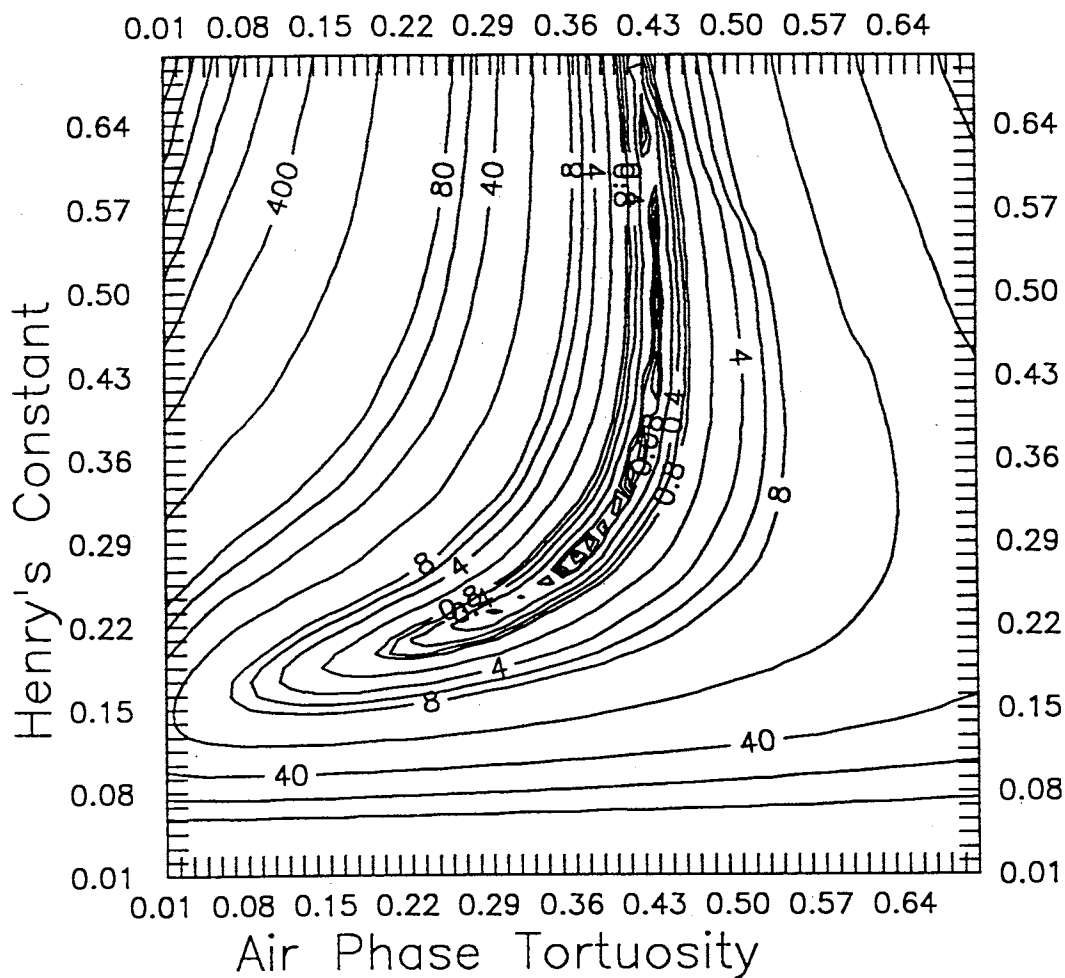
Figure 25.   Objective function contours in air phase
             tortuosity (ξ_g) Henry's constant (K_h)
             space using one model output (adsorption
             coefficient K_d set to 0.43 and
             nonuniform contour lines selected for
             clarity).
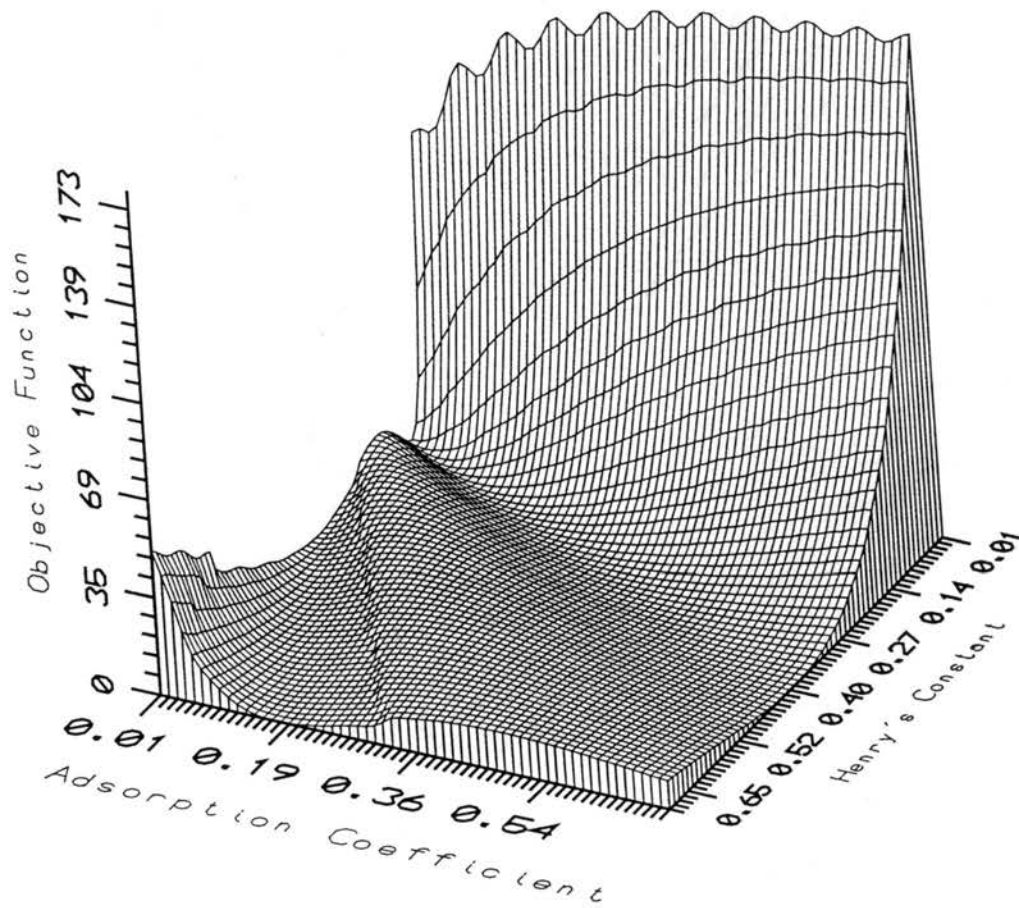
Figure 26.   Objective function response surface in adsorption coefficient (K_d) and Henry's constant (K_h) space using one model output (air phase tortuosity ξ_g set to 0.34).
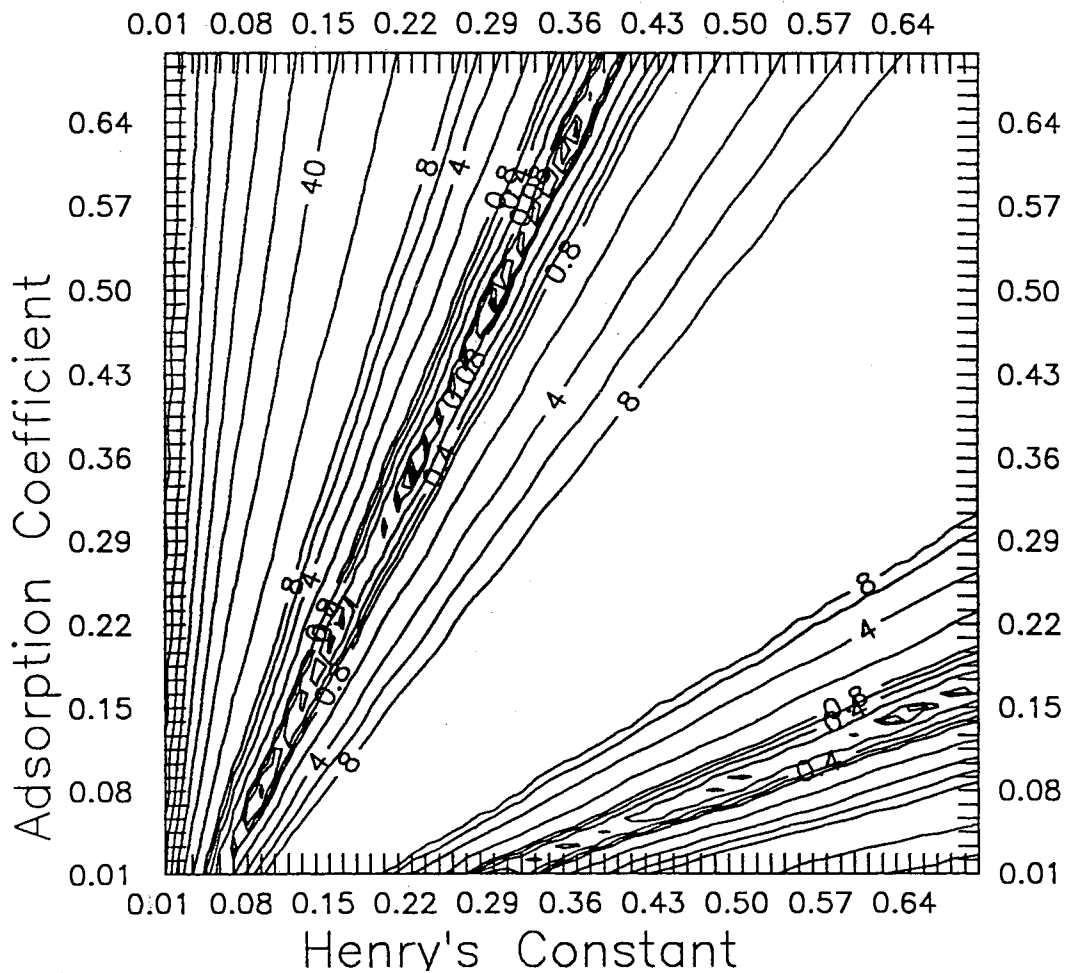
Figure 27. Objective function contours in adsorption
coefficient (K_d) and Henry's constant
(K_h) space using one model output (air
tortuosity ξ_g set to 0.34 and nonuniform
contour lines selected for clarity).

Figure 28. Objective function response surface in air
phase tortuosity (ξ_g) and adsorption
coefficient (K_d) space using two model
outputs (Henry's constant K_h set to
0.26).

128

Figure 29. Objective function contours in air phase
            tortuosity (ξ_g) and adsorption
            coefficient (K_d) space using two model
            outputs (Henry's constant K_h set to 0.26
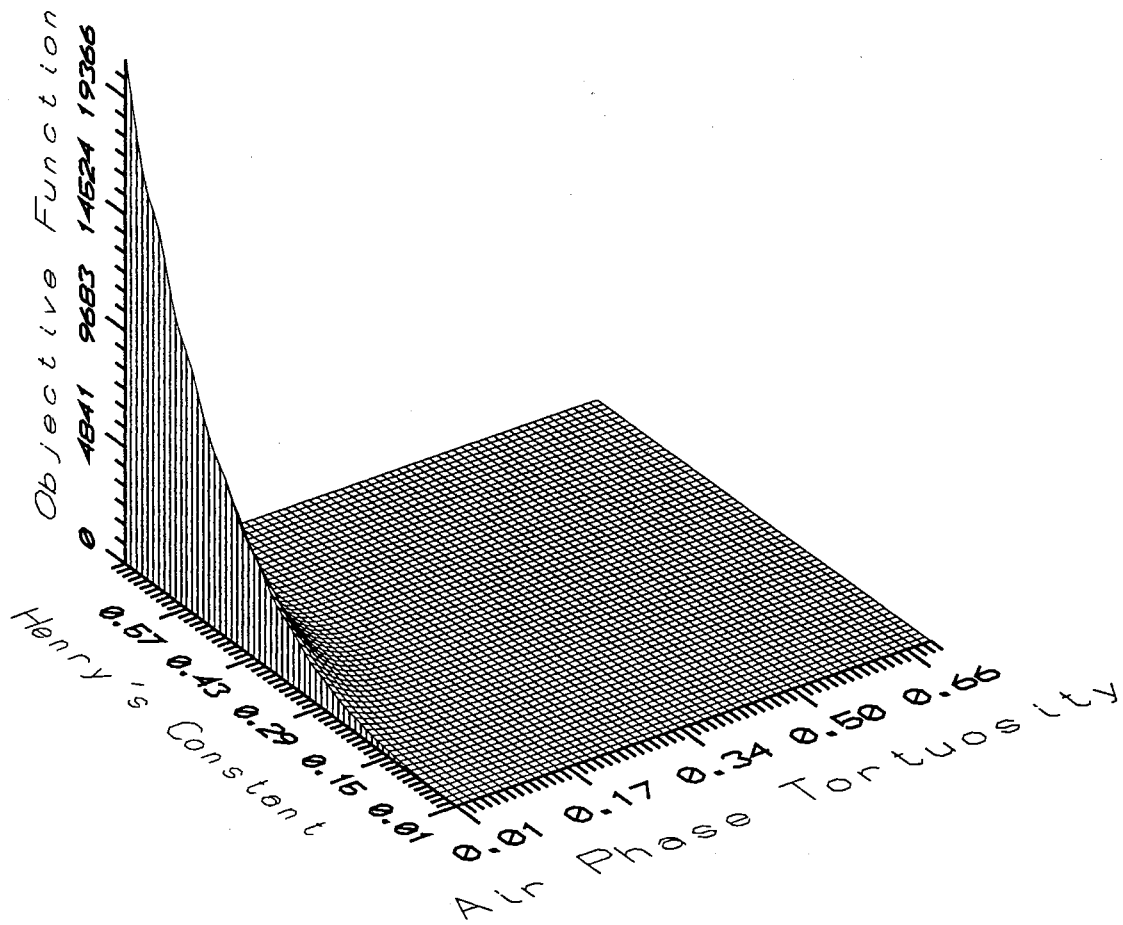   and nonuniform contour lines selected
            for clarity).

Figure 30. Objective function response surface in air
phase tortuosity (ξ_g) and Henry's
constant (K_h) space using two model
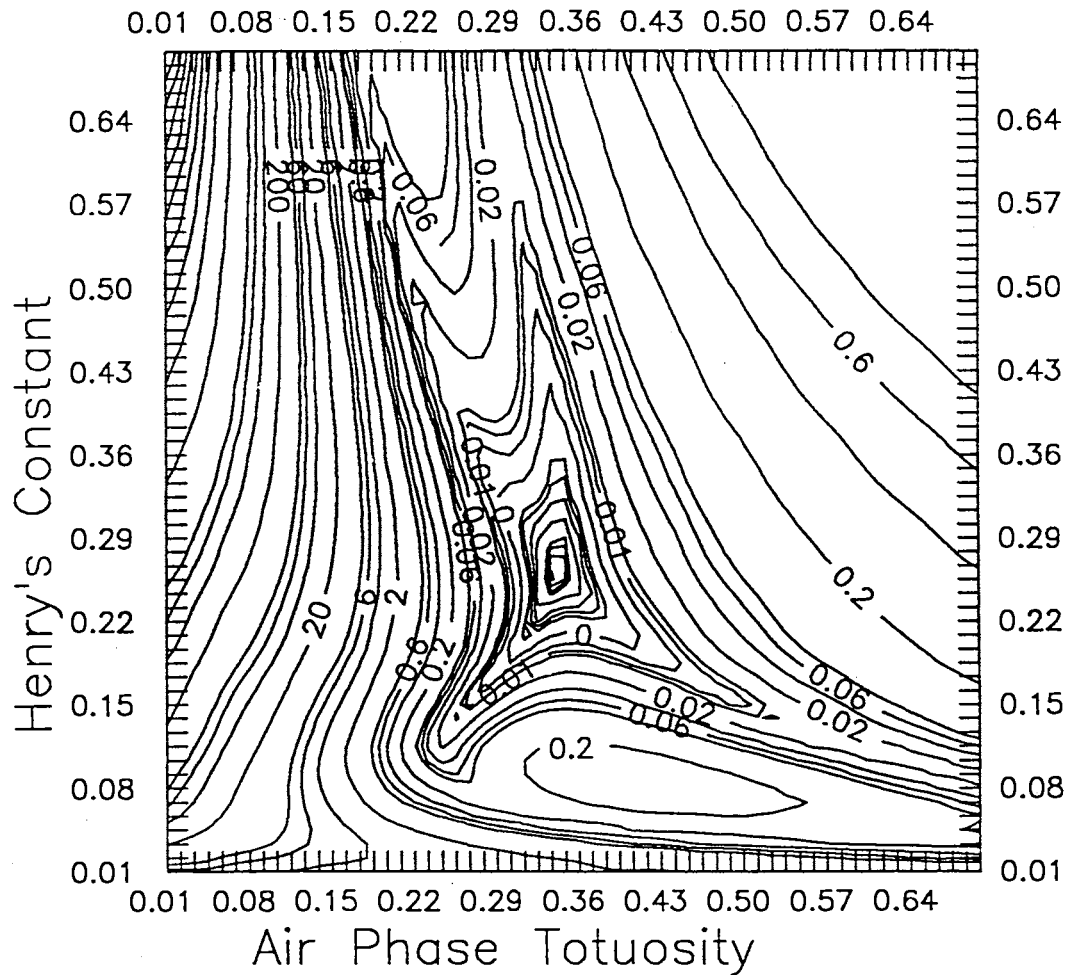outputs (adsorption coefficient set to
0.43).

Figure 31. Objective function contours in air phase tortuosity (ξ_g) and Henry's constant (K_h) space using two model outputs (adsorption coefficient K_d set to 0.43 and nonuniform contour lines selected for clarity).
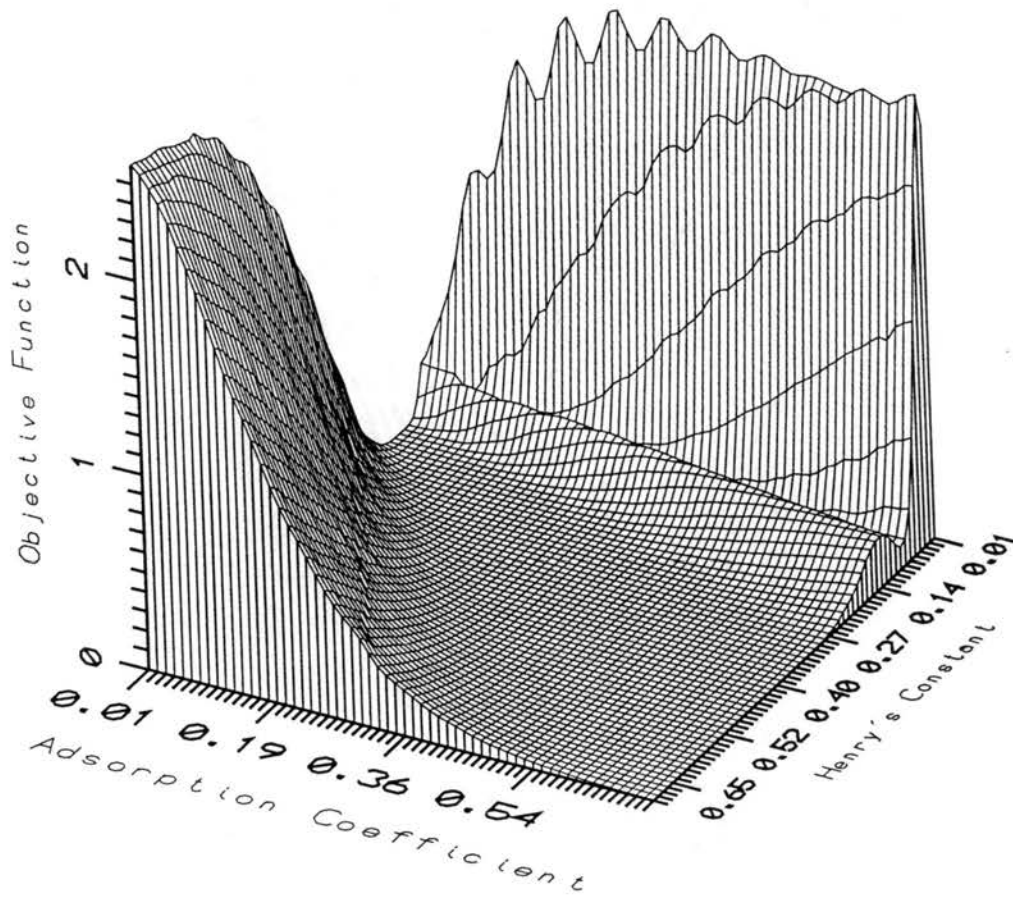
131

Figure 32. Objective function response surface in adsorption coefficient (K_d) and Henry's constant (K_h) space using two model outputs (air phase tortuosity ξ_g set to 0.34).
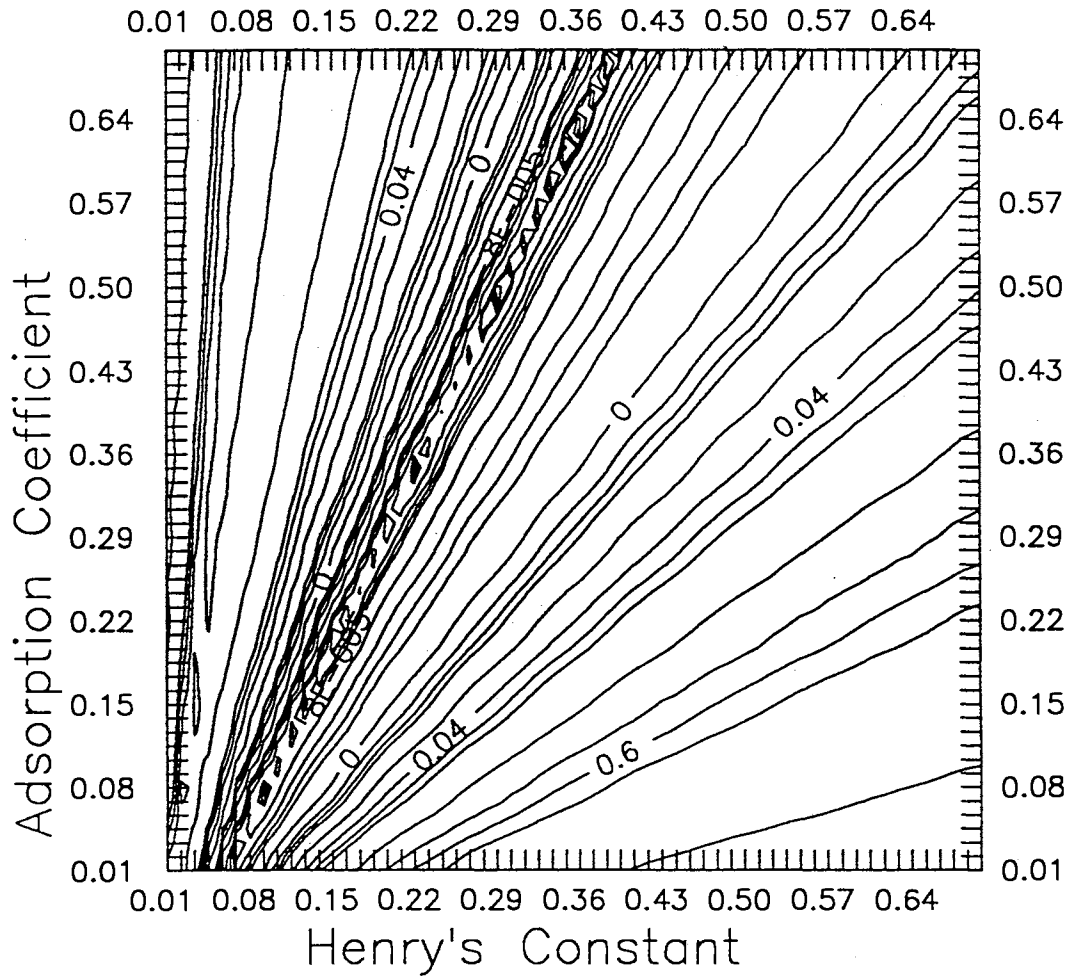
Figure 33. Objective function contours in adsorption coefficient (K_d) and Henry's constant (K_h) space using two model outputs (air phase tortuosity ξ_g set to 0.34).
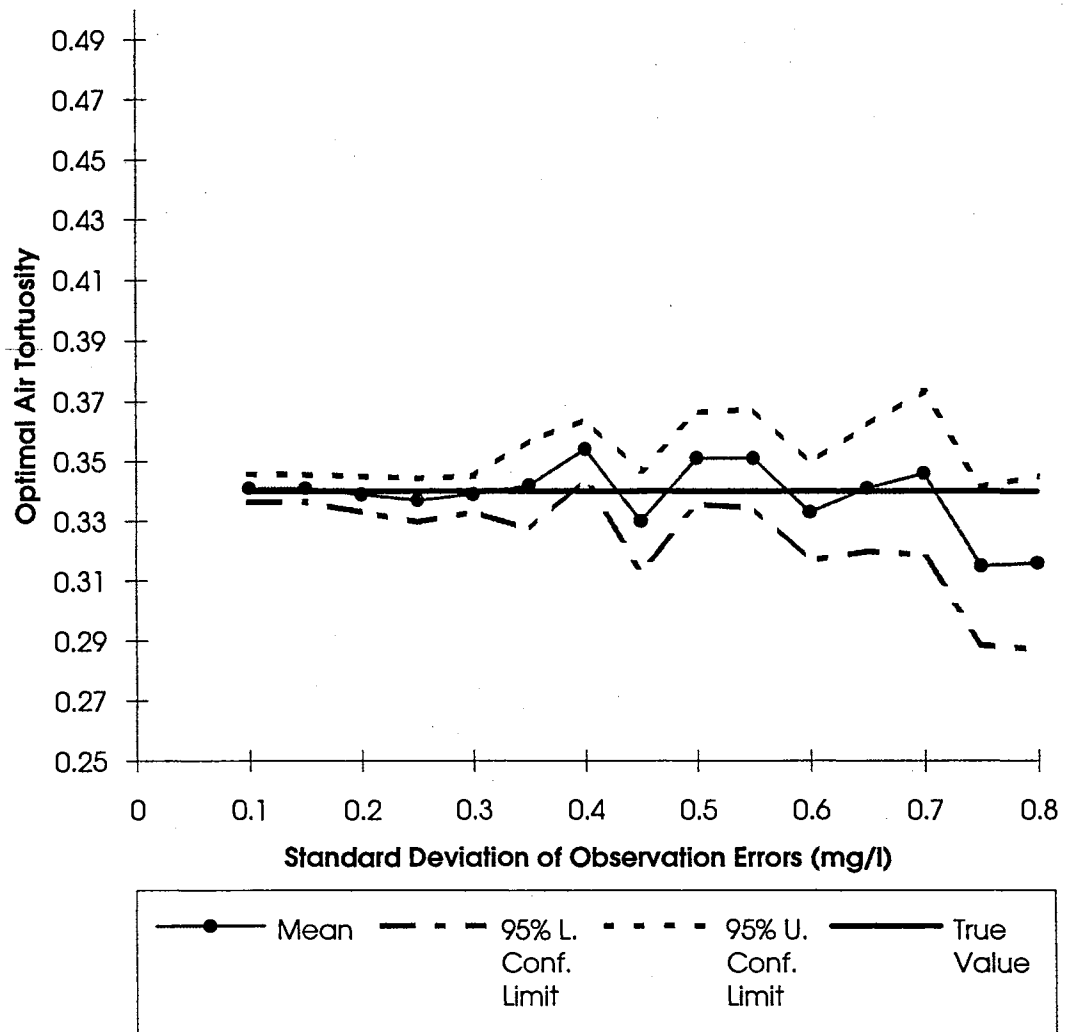
Figure 34.   Impact of observation error on the
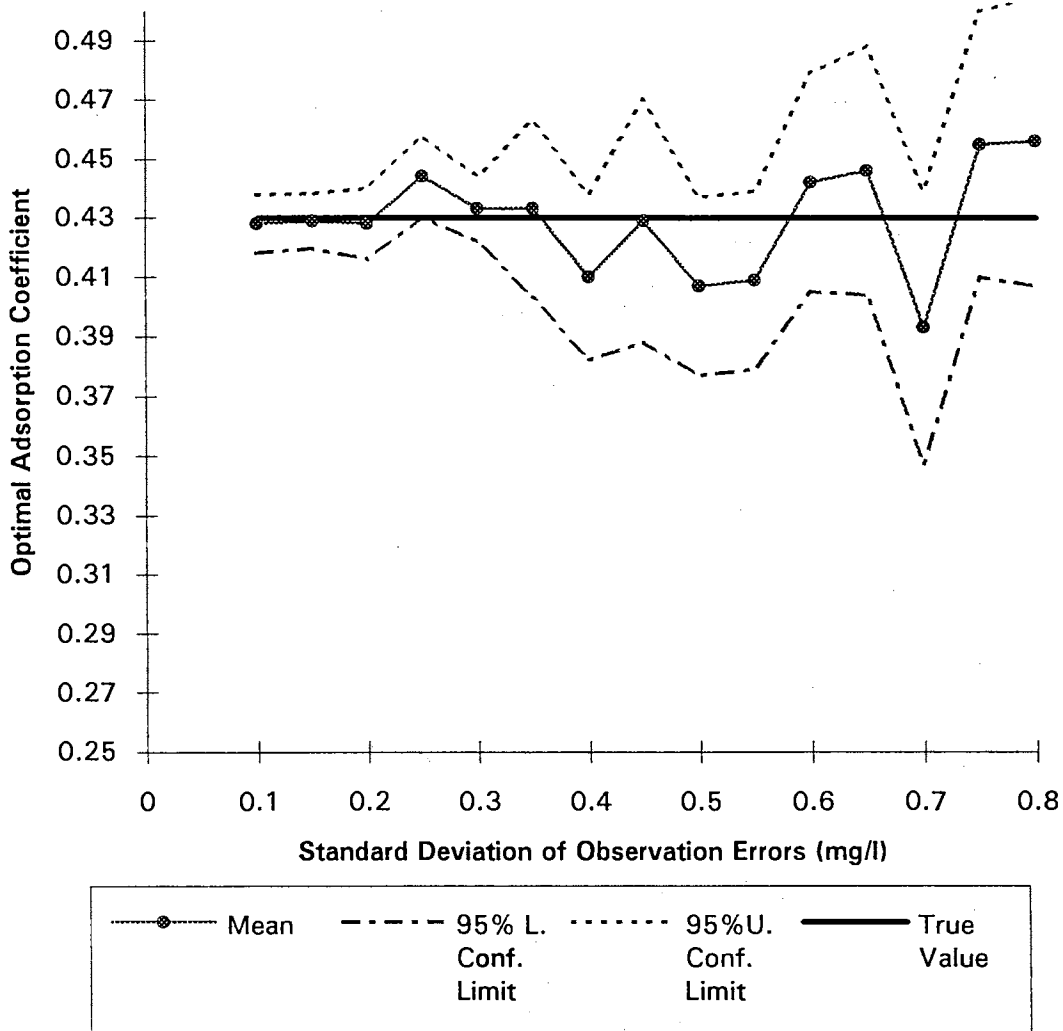   estimates of air phase tortuosity (ξ_g).

Figure 35. Impact of observation error on the estimates of adsorption coefficient (K_d).
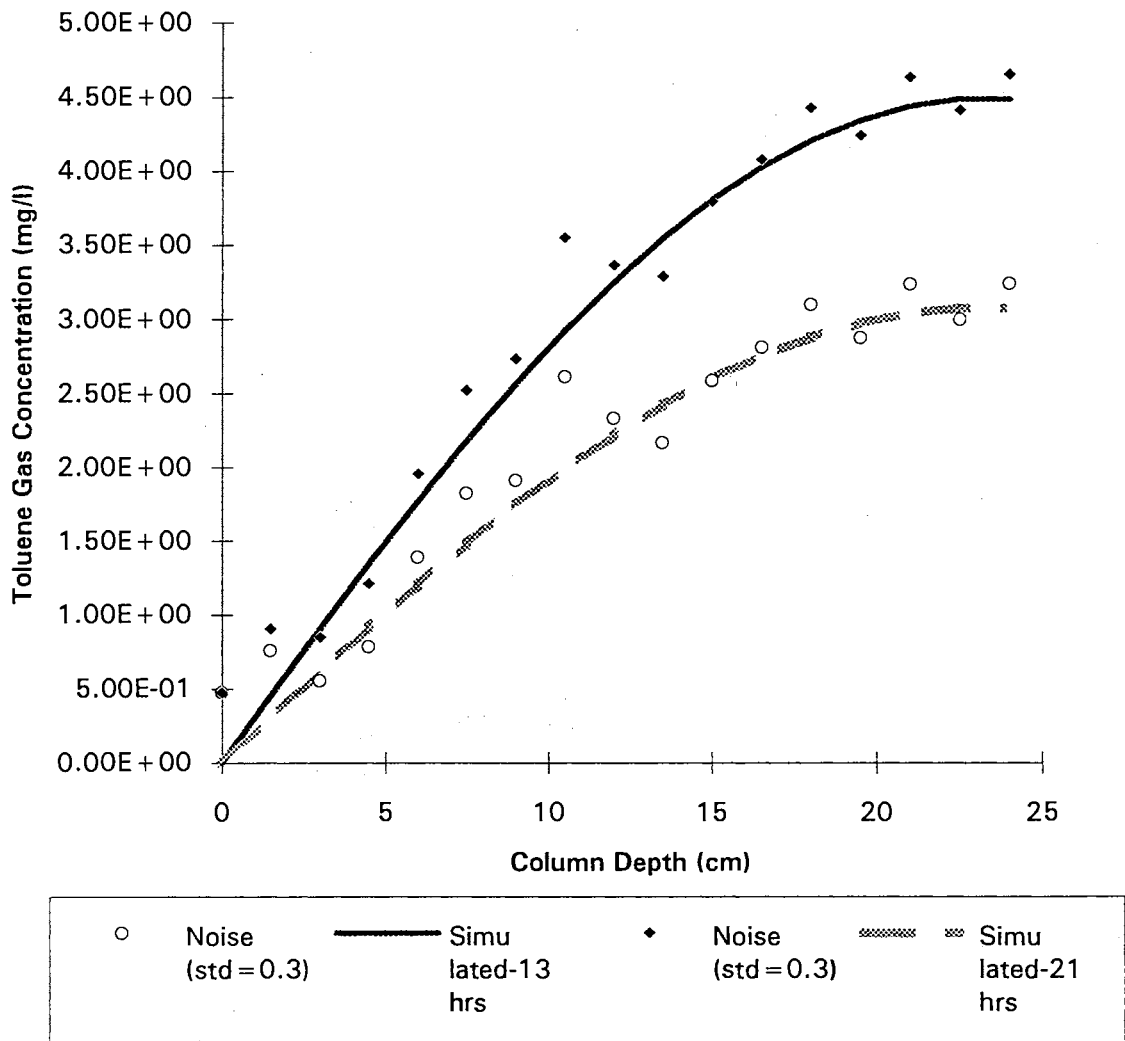
Figure 36. Demonstration of observation error with a
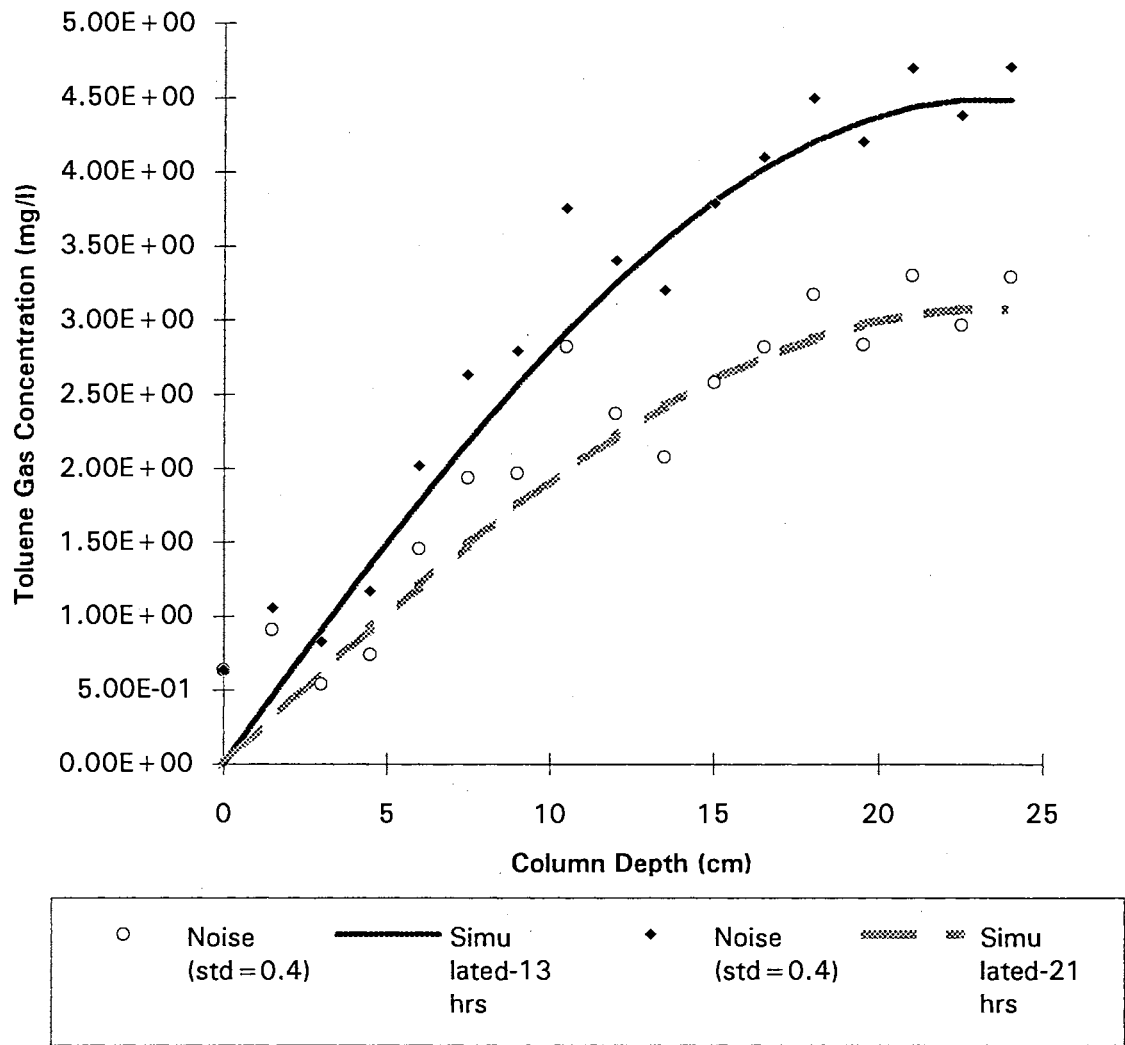standard deviation of 0.3 mg/l.

Figure 37.   Demonstration of observation error with a
standard deviation of 0.4 mg/l.
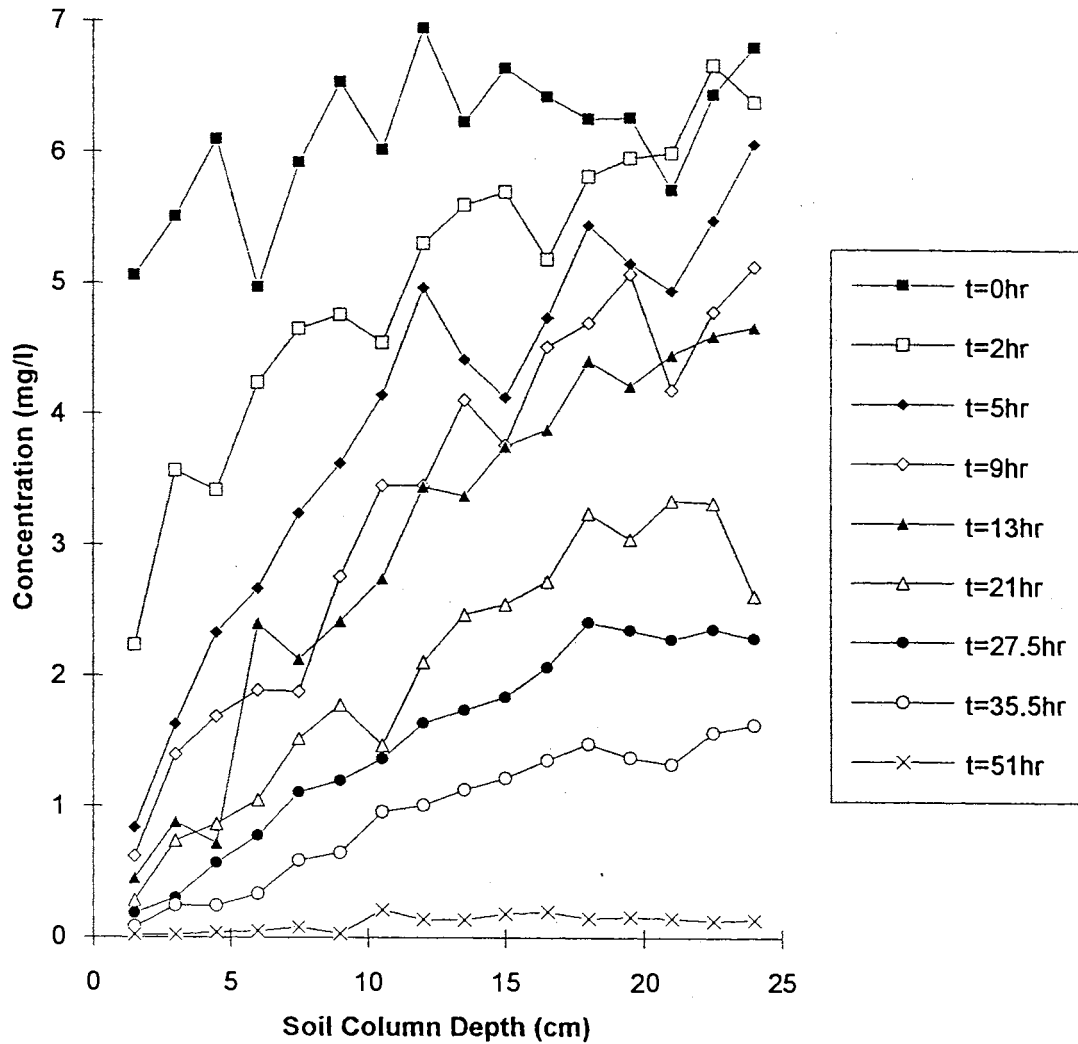
Figure 38. Measured toluene air phase concentration profiles at different times (Yu, 1995).

Figure 39. Comparison of measured air phase toluene concentrations and calibrated model simulation with estimated parameters at 13 hours ($K_d=0.39$ and $\xi_g=0.42$).

139

Figure 40.   Comparison of measured air phase toluene
             concentrations and calibrated model
             simulation with estimated parameters at
             21 hours (K_d=0.39 and ξ_g=0.42).

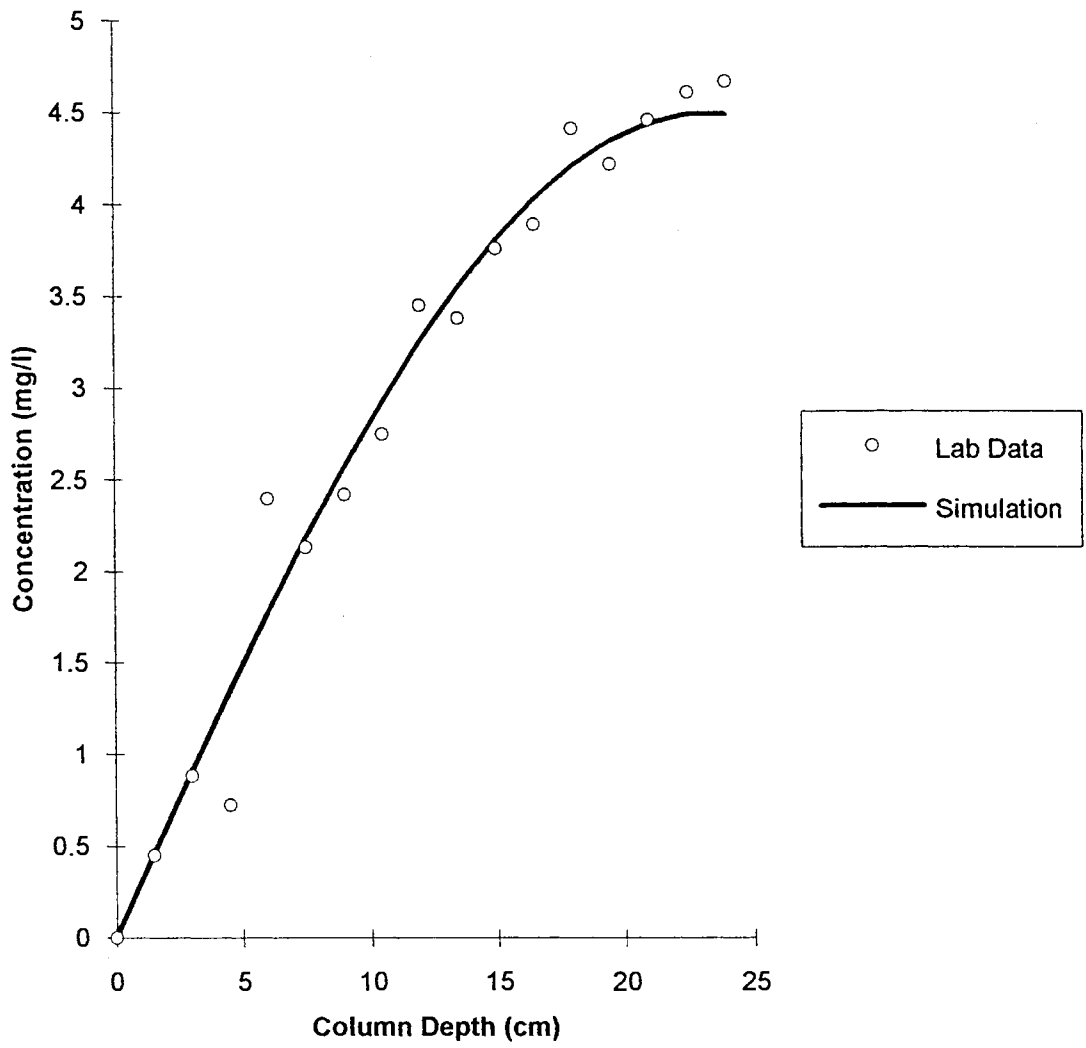Figure 41. Verification of measured air phase toluene
concentrations and model simulation with
estimated parameters at 2 hours (K_d=0.39
and ξ_g=0.42).

Figure 42. Verification of measured air phase toluene
concentrations and model simulation with
estimated parameters at 5 hours (K_d=0.39
and ξ_g=0.42).
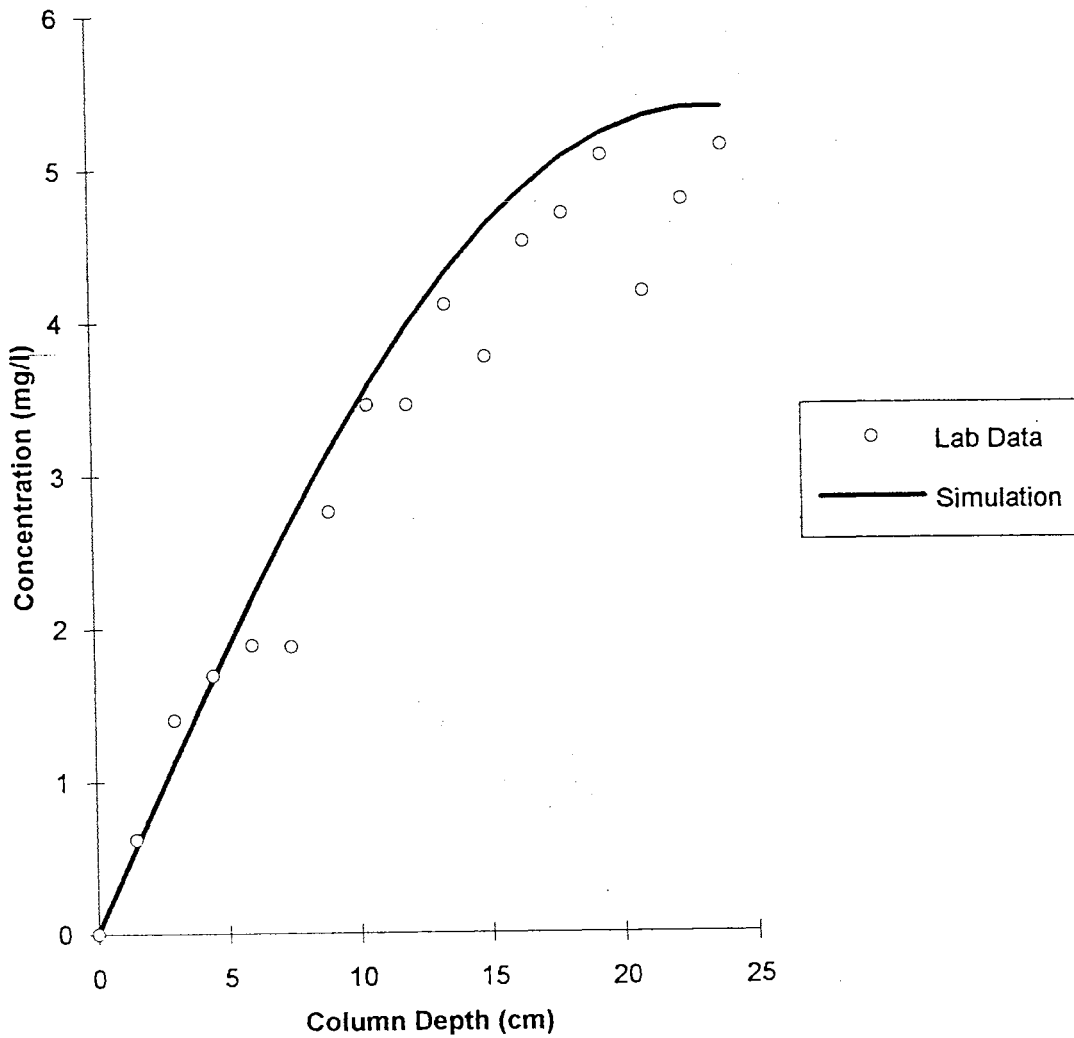
Figure 43.   Verification of measured air phase toluene
              concentrations and model simulation with
              estimated parameters at 9 hours (K_d=0.39
              and ξ_g=0.42).
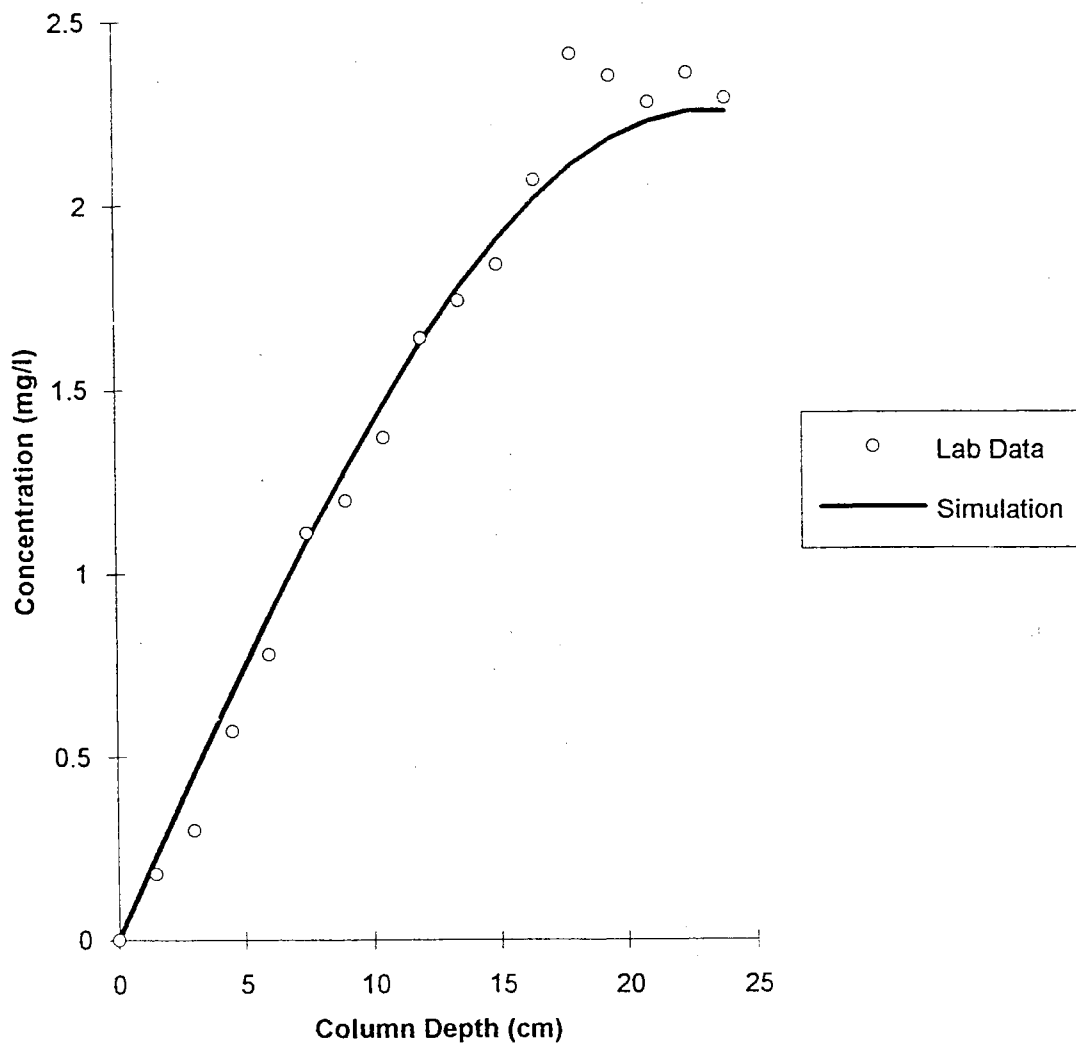
Figure 44.   Verification of measured air phase toluene
             concentrations and model simulation with
             estimated parameters at 27.5 hours
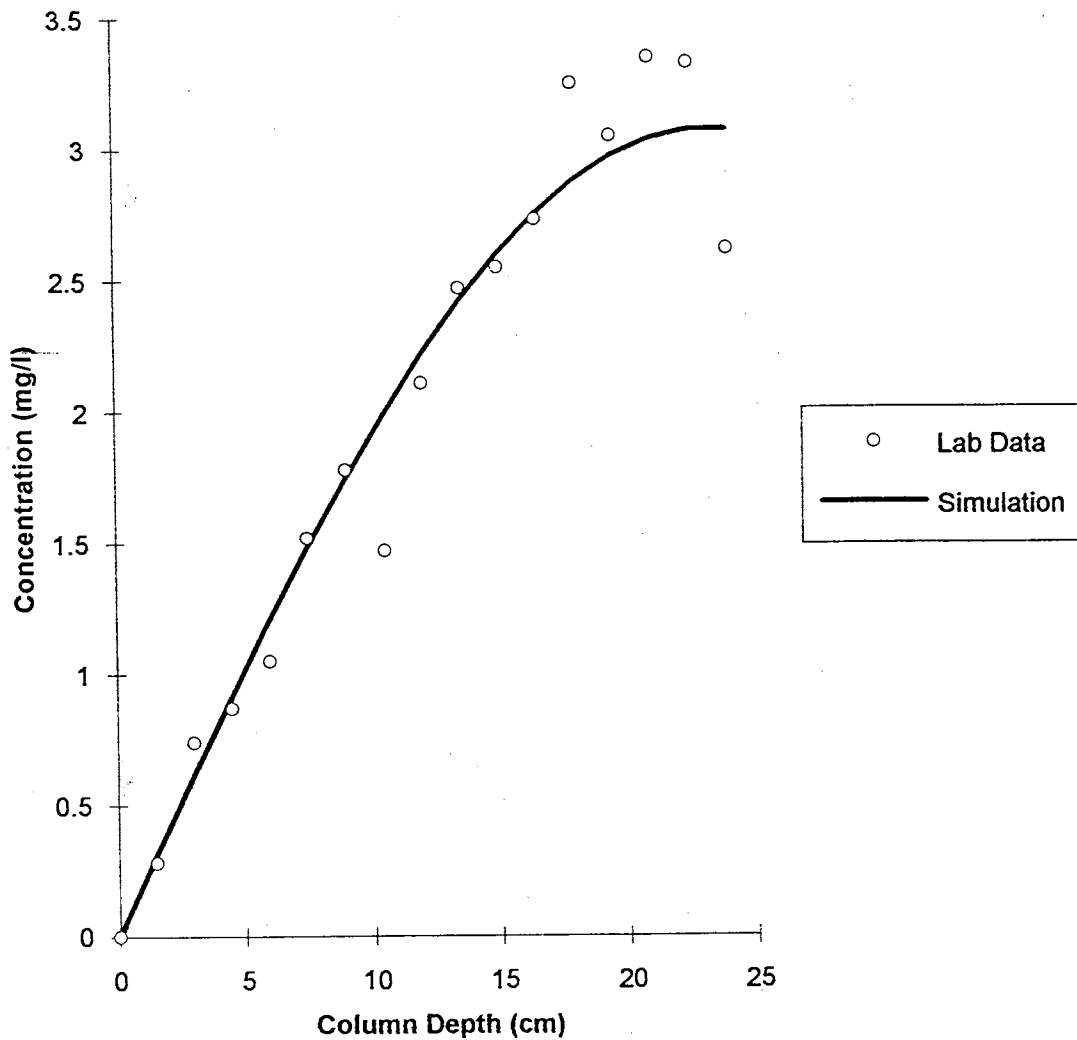   (K_d=0.39 and ξ_g=0.42).
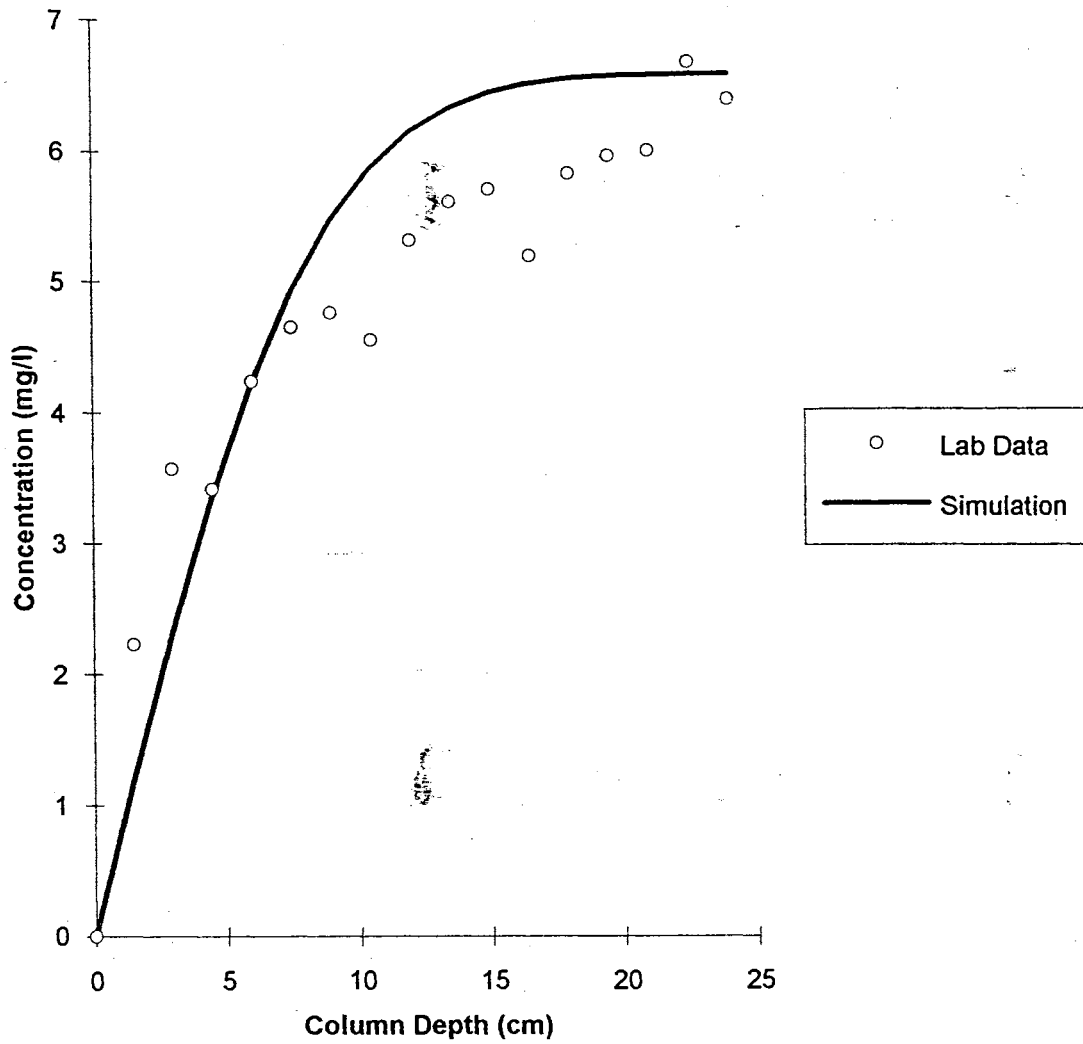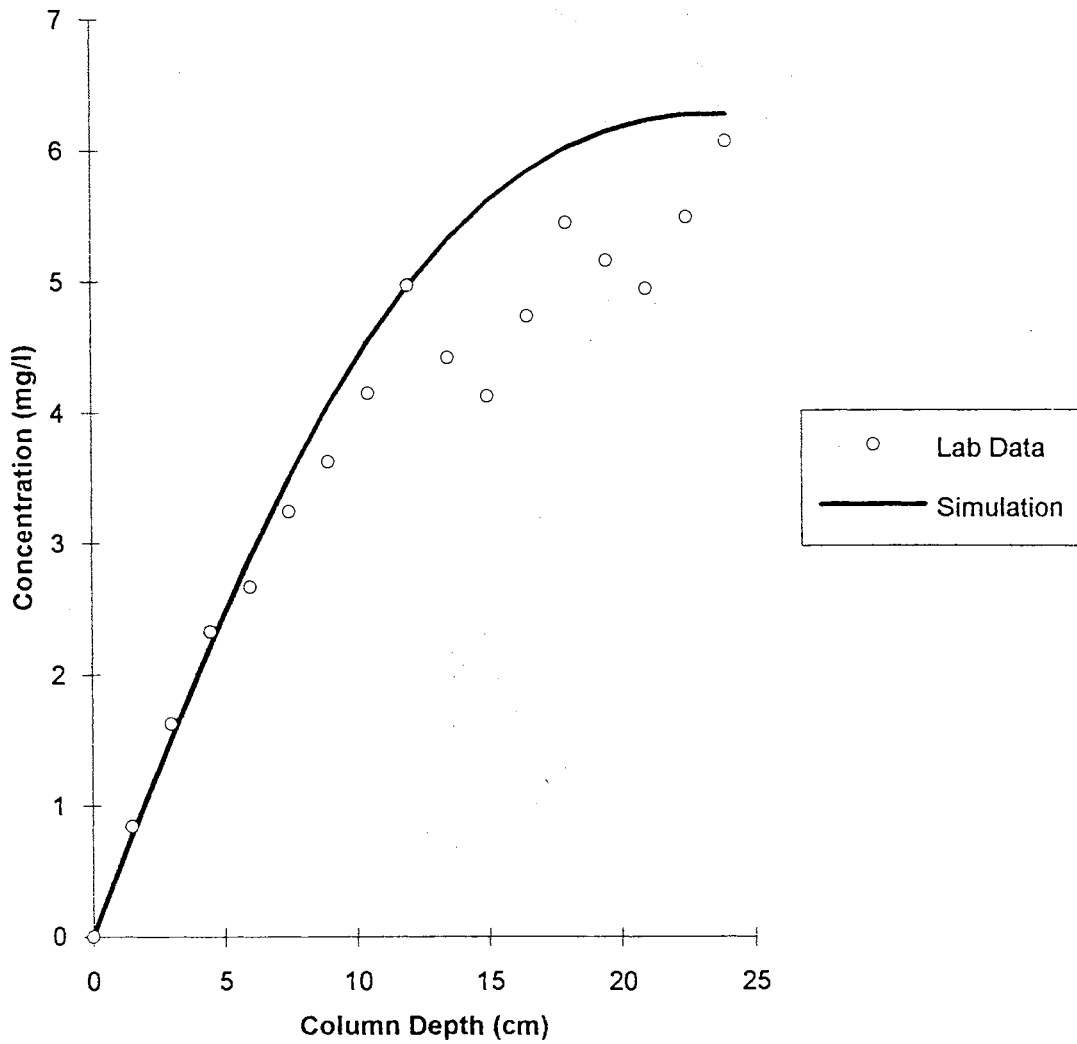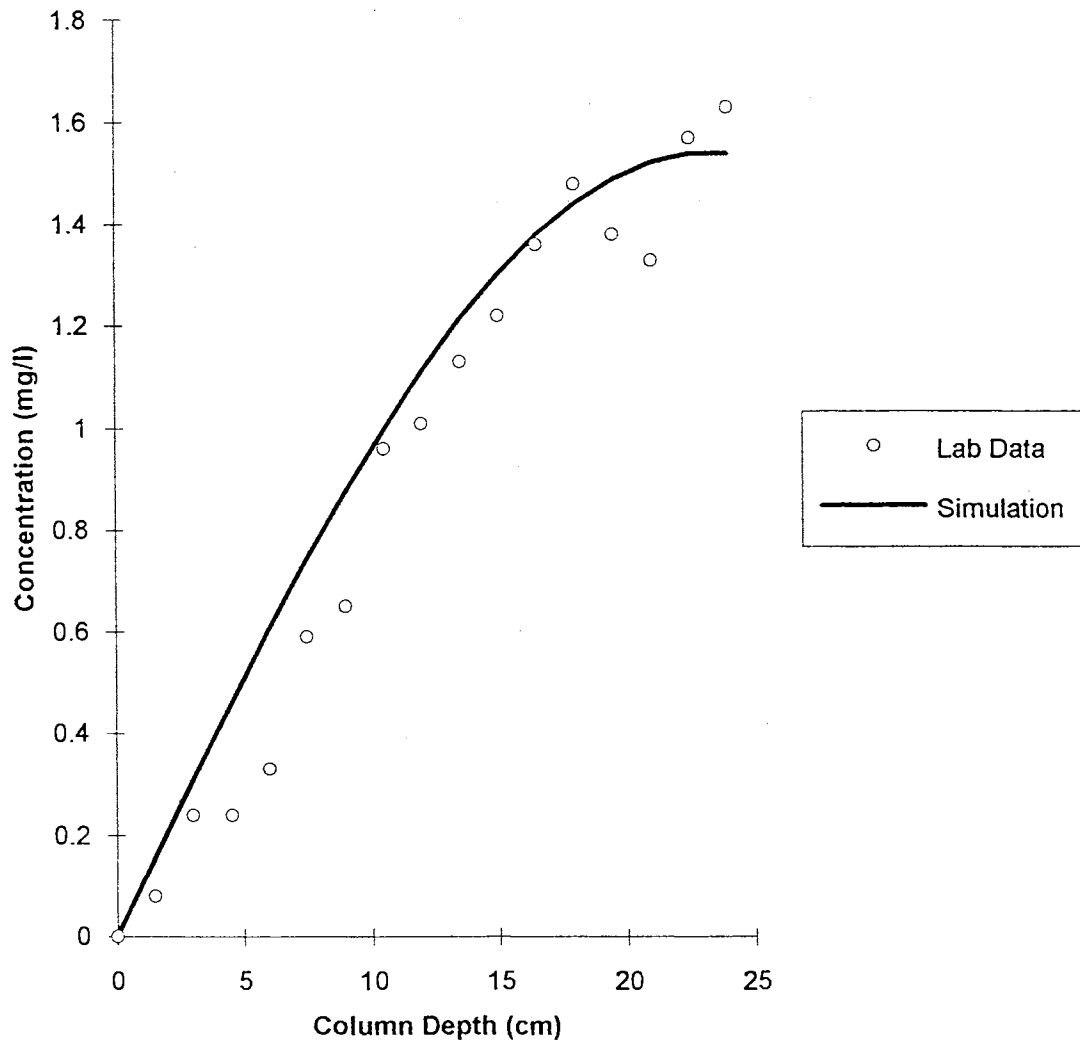
Figure 45. Verification of measured air phase toluene
           concentrations and model simulation with
           estimated parameters at 35.5 hours
           (K_d=0.39 and ξ_g=0.42).

145

Figure 46.   Residuals between simulated and measured
             toluene concentration profiles at 13
             hours.
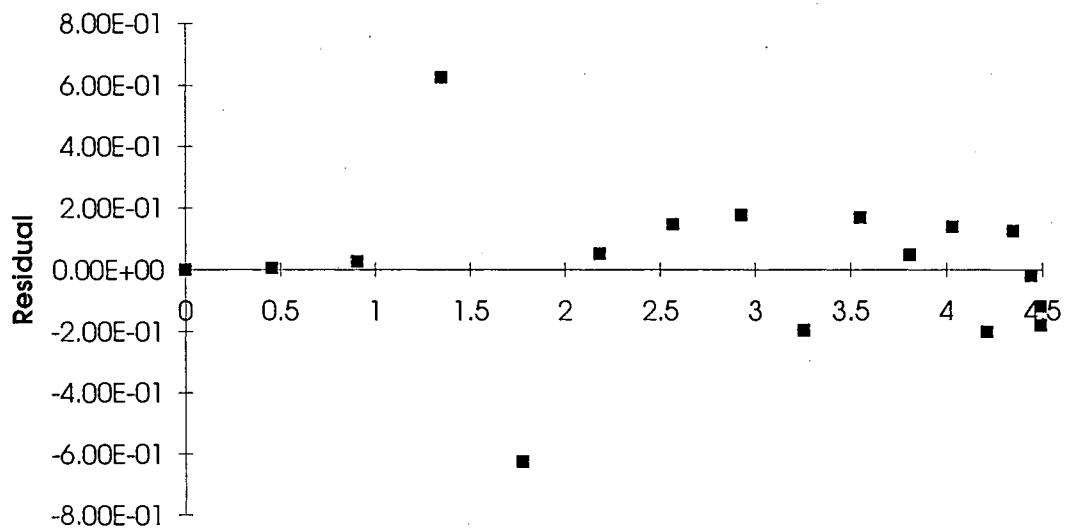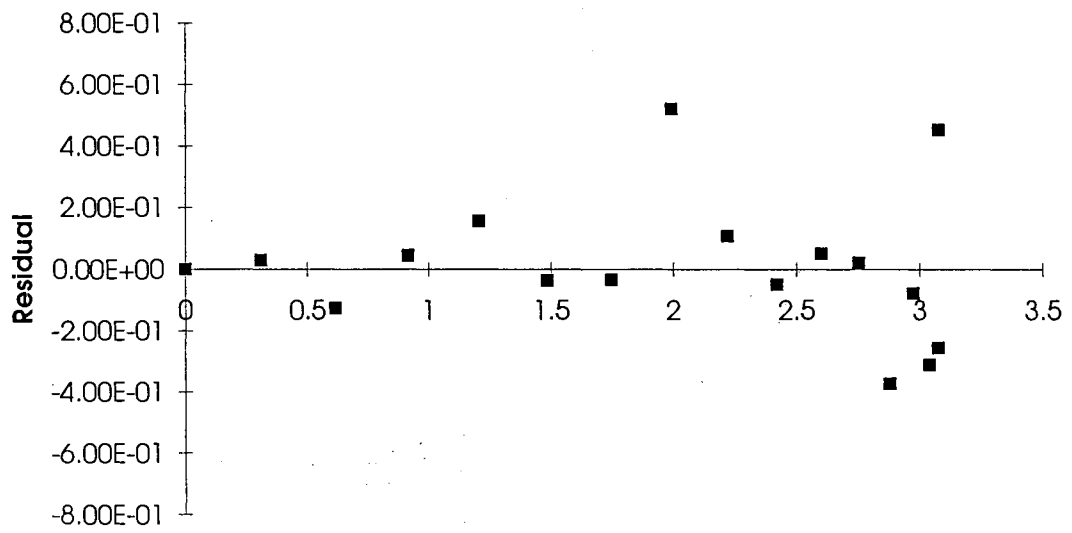
Simulated Toluene Gas Concentration Profile at 13 Hours (mg/l)

Figure 47. Residuals between simulated and measured
            toluene concentration profiles at 21
            Hours.
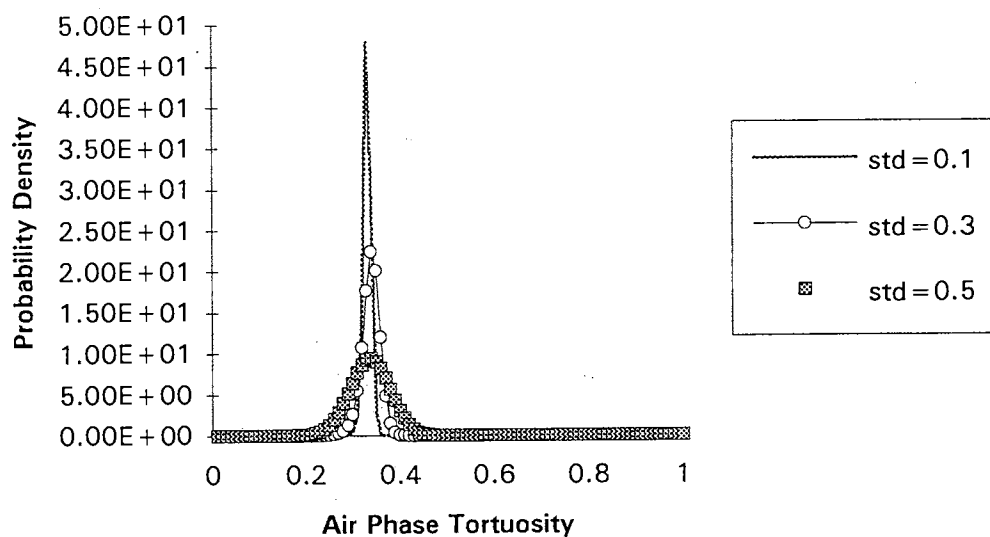
Figure 48. Marginal probability distributions of air
            phase tortuosity (ξ_g) estimated from
            observations with different error
            standard deviations.

Figure 49. Marginal probability distributions of air phase tortuosity (ξ_q) estimated from observations with different error standard deviations.
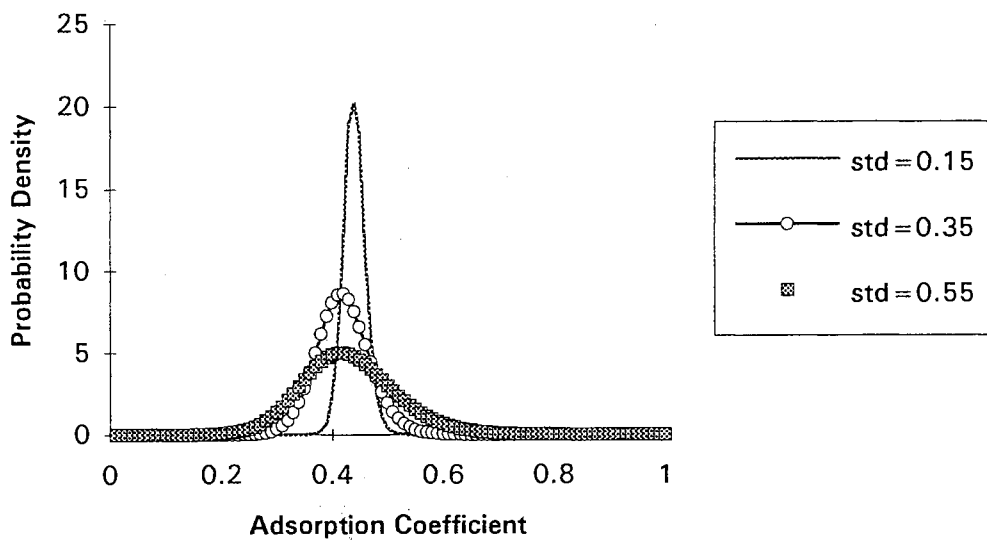
Figure 50. Marginal probability distributions of the
adsorption coefficient (K_d) estimated
from observations with different error
standard deviations.
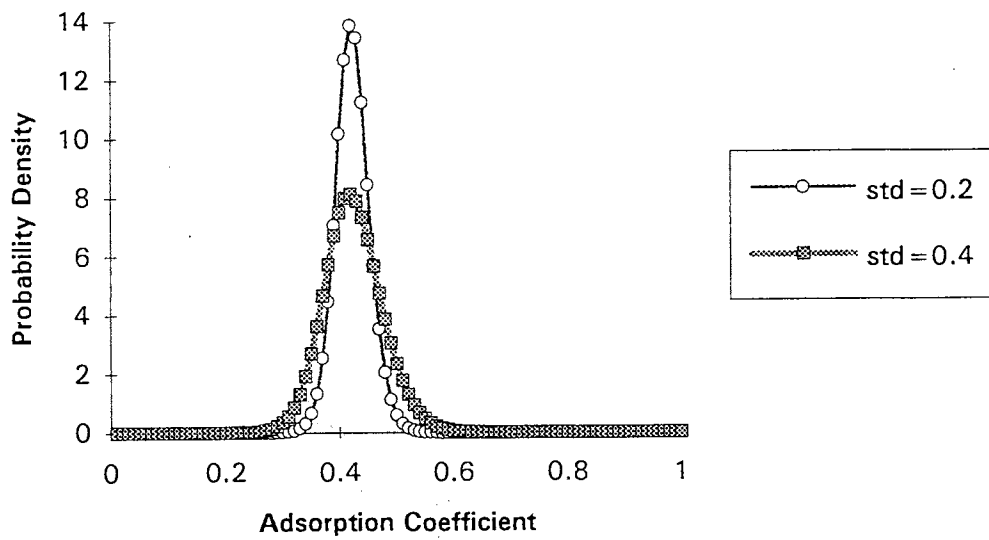
Figure 51. Marginal probability distributions of the adsorption coefficient (K_d) estimated from observations with different error standard deviations.
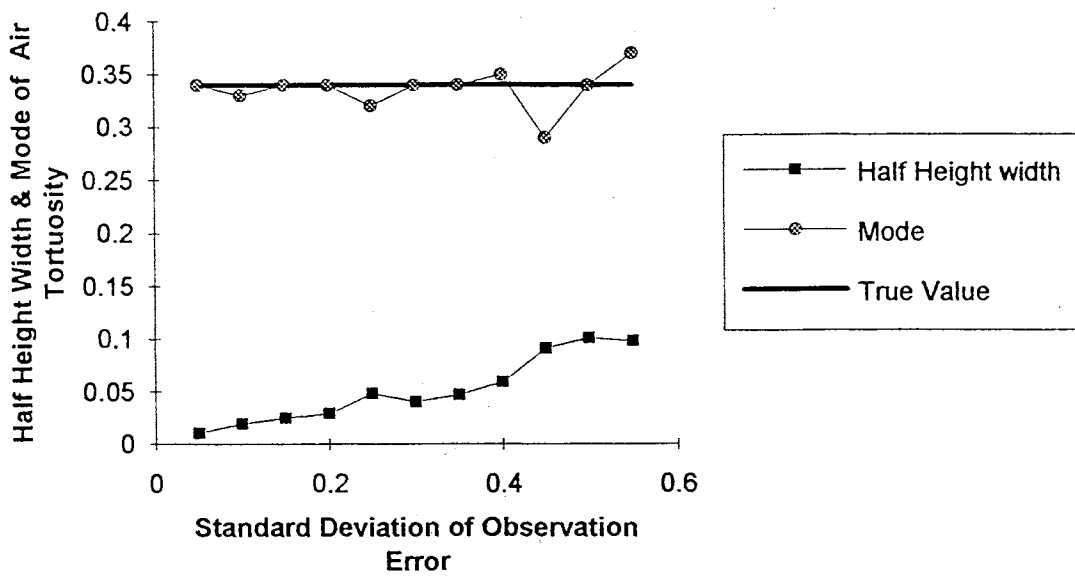
Figure 52.   Distributions of the half height width and
             modes of air phase tortuosity
             distributions due to the change of the
             standard deviation of observation
             errors.

152
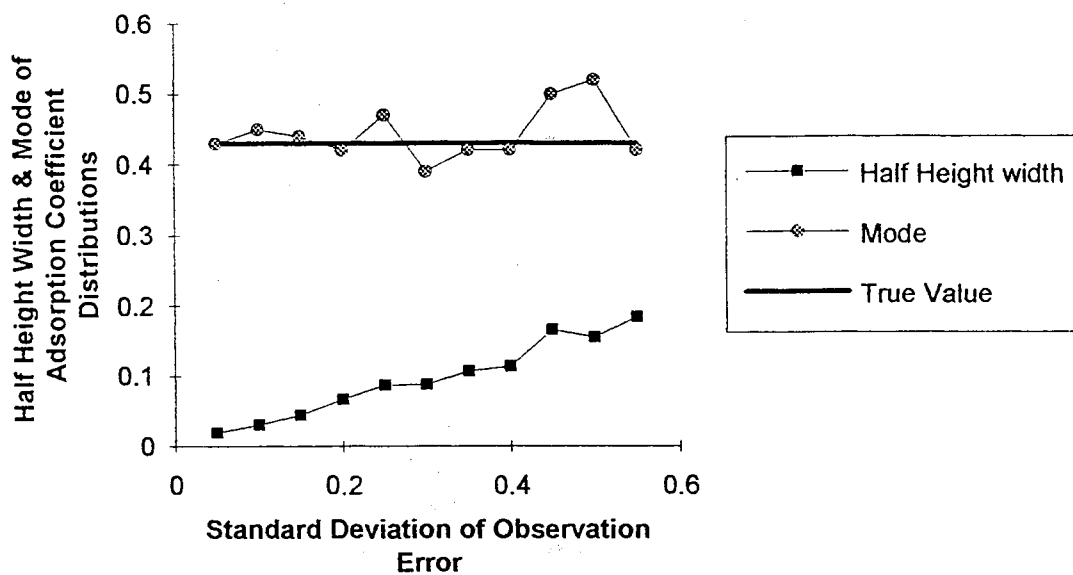
Figure 53. Distributions of the half height width and
modes of adsorption coefficient (K_d)
distributions due to the change of the
standard deviation of observation
errors.
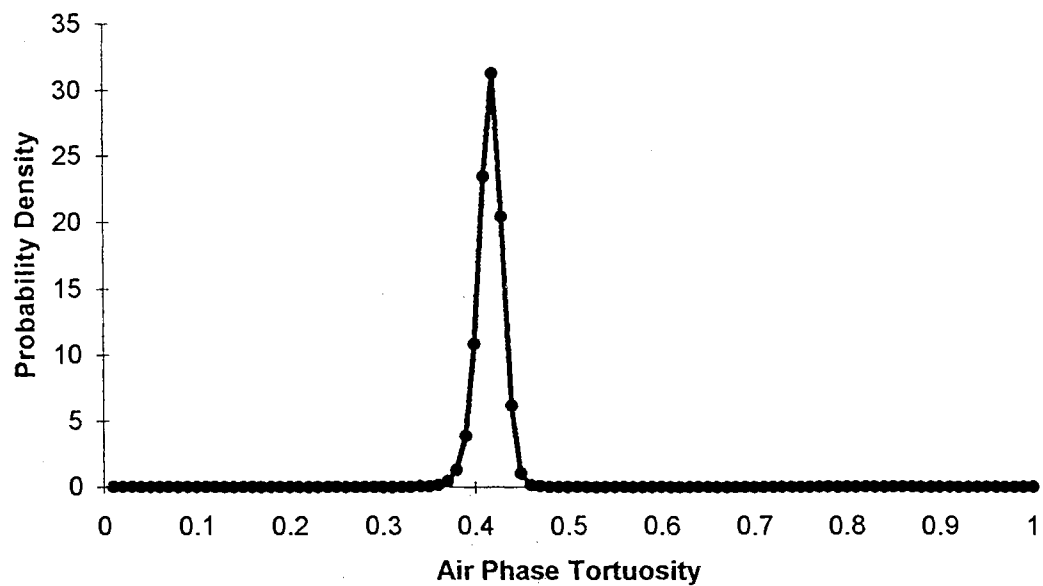
Figure 54. Marginal probability distribution of air
            phase tortuosity (ξ_g) estimated from
            measured toluene gas phase concentration
            profiles at 13 hours and 21 hours.
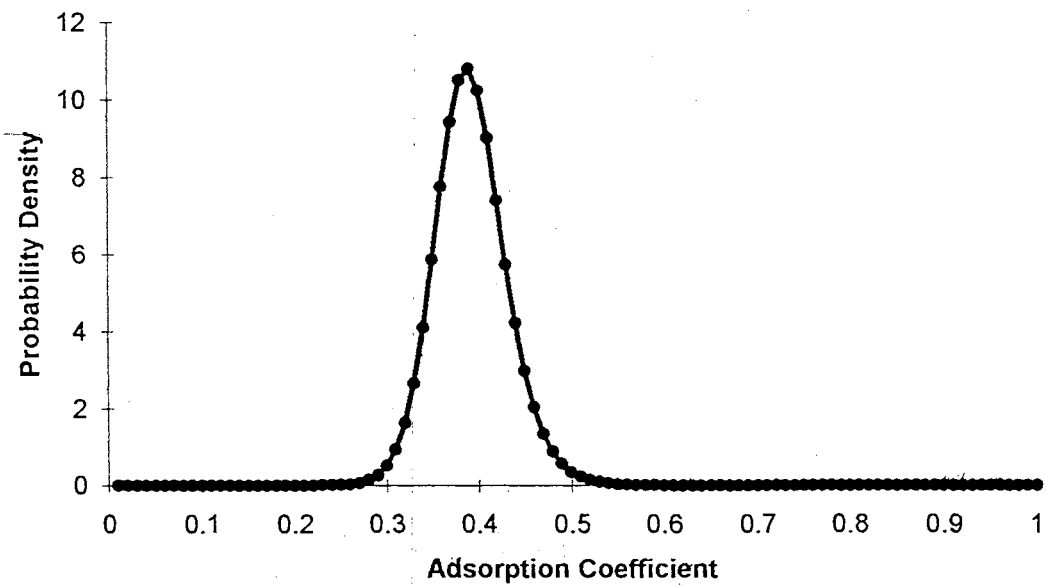
Figure 55. Marginal probability distribution of
adsorption coefficient (K_d) estimated
from measured toluene gas phase
concentration profiles at 13 hours and
21 hours.
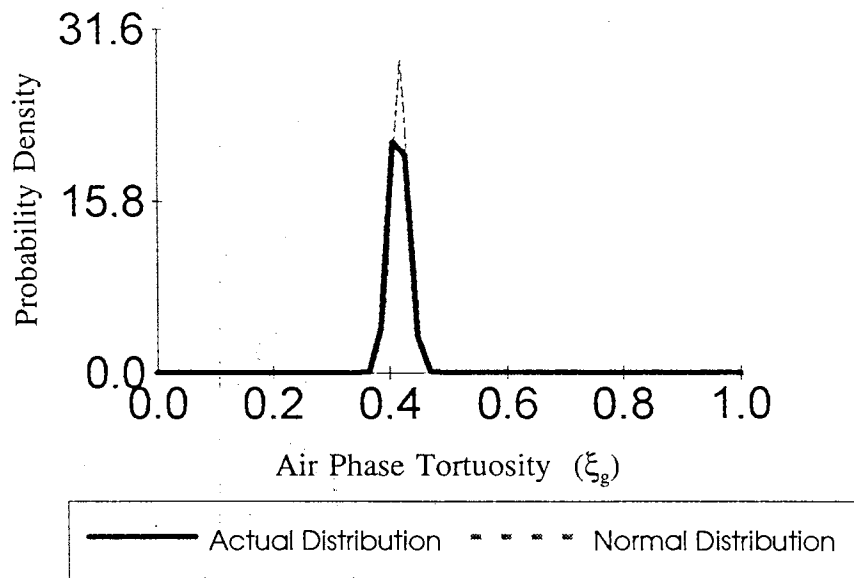
Figure 56. Comparison of generated air phase
tortuosity (ξ_g) distribution and Normal
(0.39,0.02).
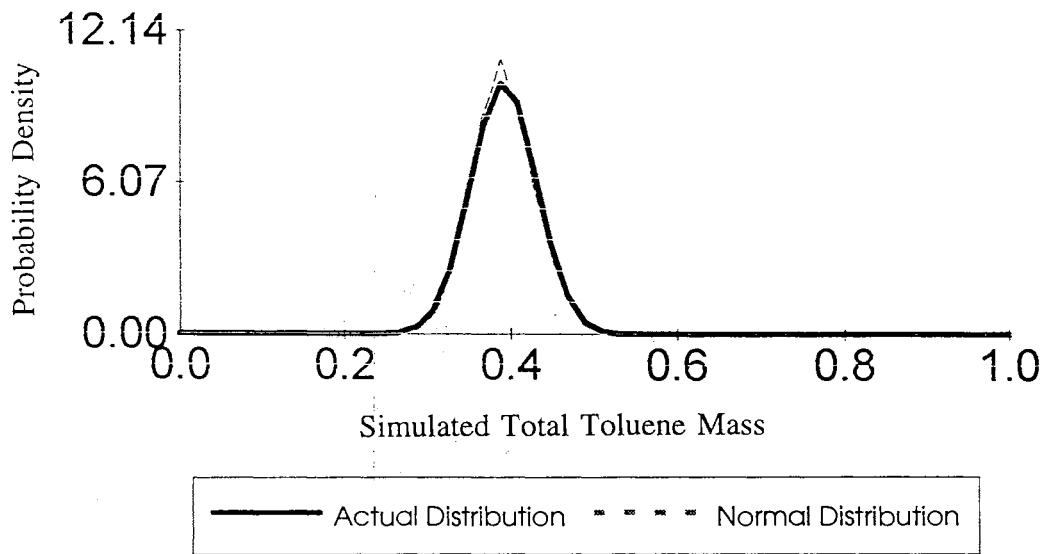
Figure 57. Comparison of generated adsorption
coefficient (K_d) distribution and Normal
(0.39,0.04).
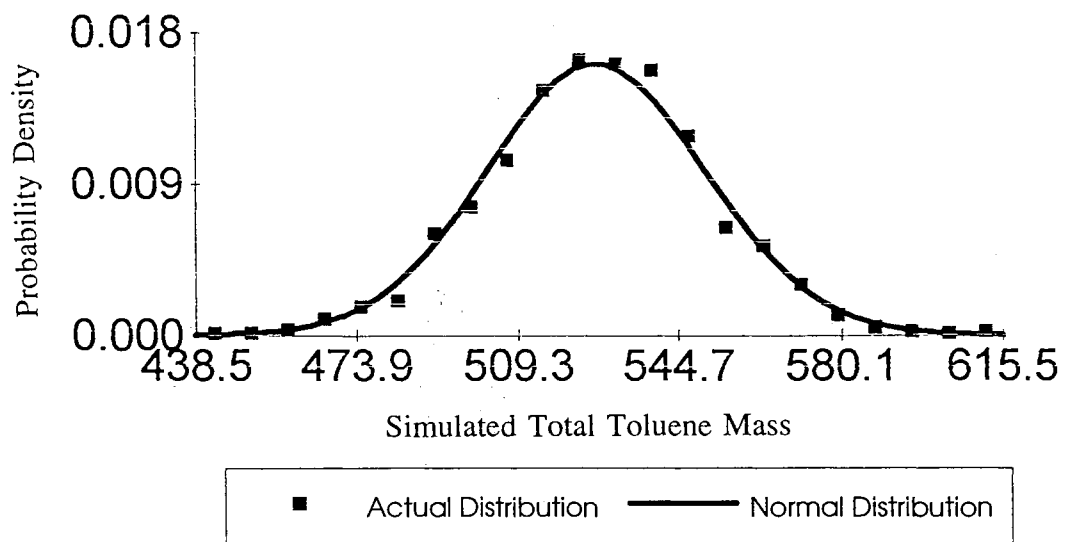
Figure 58. Comparison of the distribution of simulated total toluene mass and normal (527, 24.14).
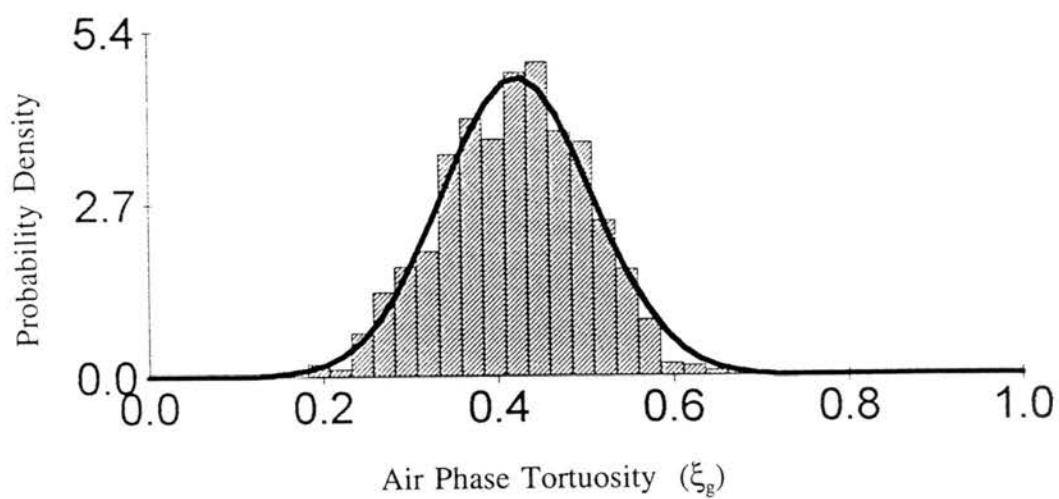
Figure 59.   Distribution of air phase tortuosity (ξ_g) with C_v=0.2 (The bars represent the actual distribution and the solid line is the normal distribution N (0.42, 8.62e-2)).
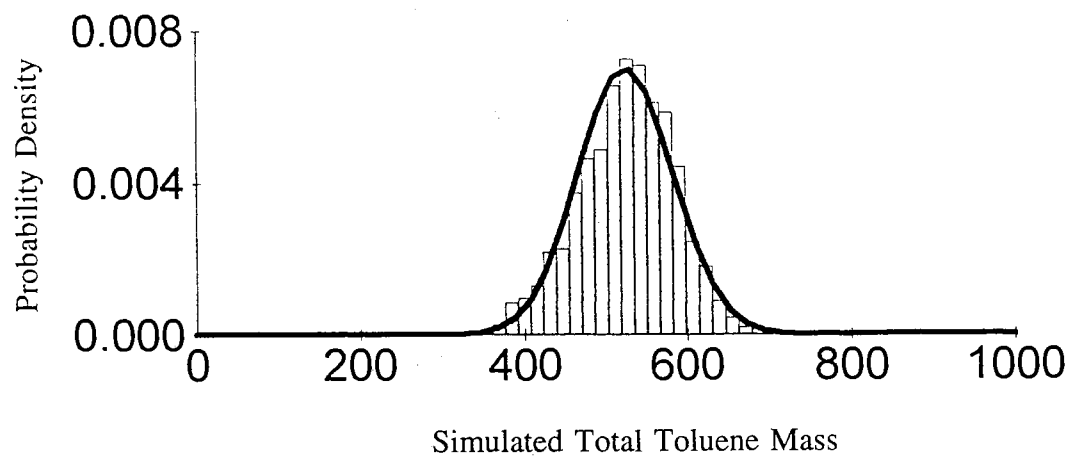
Figure 60.   Distribution of simulated total toluene
             mass (The bars represent the actual
             distribution and the solid line is the
             normal distribution N (5.26e+2, 58.24)).
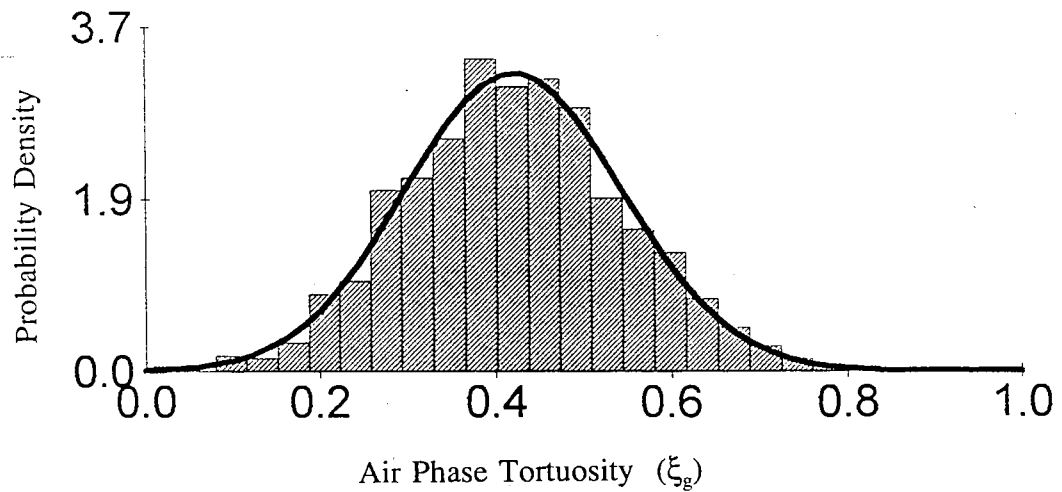
Figure 61. Distribution of air phase tortuosity (ξ_g)
with C_v=0.3 (The bars represent the
actual distribution and the solid line
is the normal distribution N(0.42,
0.12)).
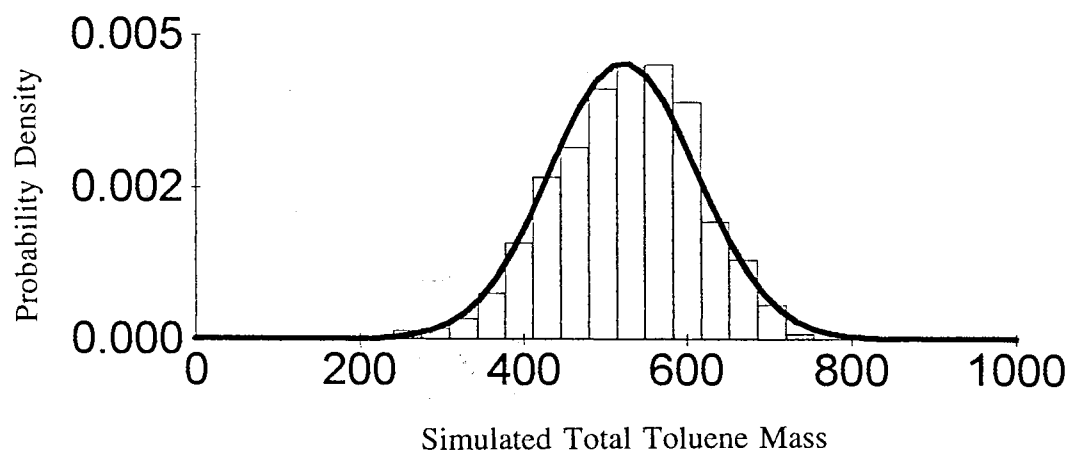
Figure 62. Distribution of simulated total toluene
mass (The bars represent the actual
distribution and the solid line is the
normal distribution N(5.23e+2, 89.95)).
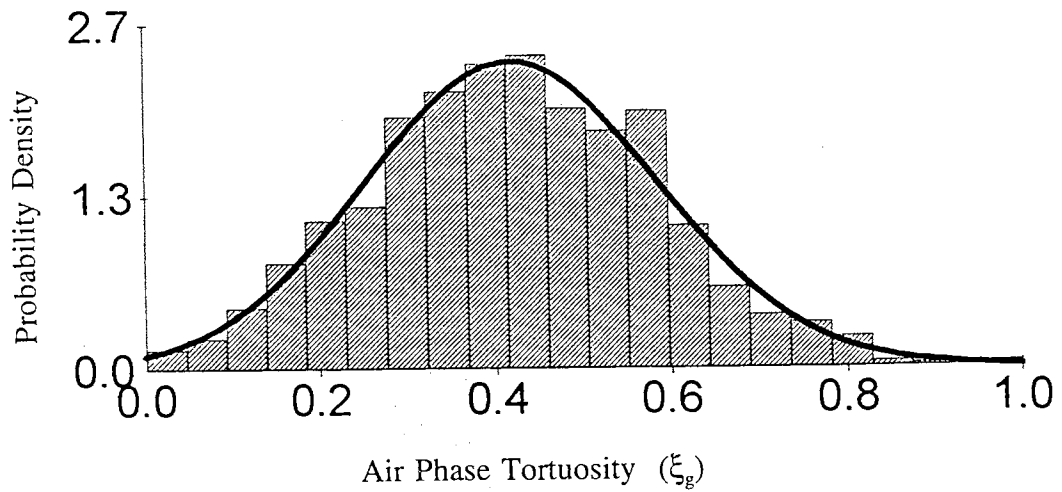
Figure 63.   Distribution of air phase tortuosity (ξ_g)
with C_v=0.4 (The bars represent the
actual distribution and the solid line
is the normal distribution N(0.42,
0.17)).

163
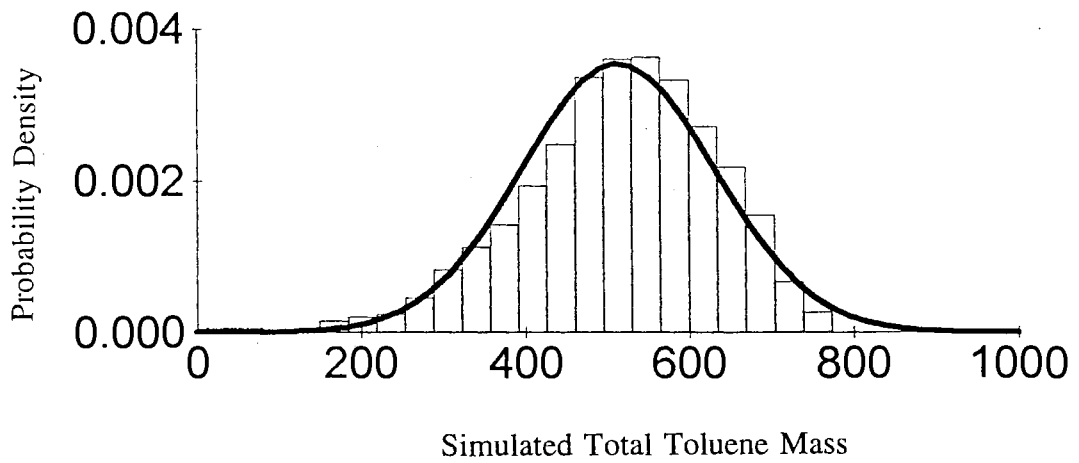
Figure 64.   Distribution of simulated total toluene
             mass (The bars represent the actual
             distribution and the solid line is the
             normal distribution N (5.14e+2, 118))
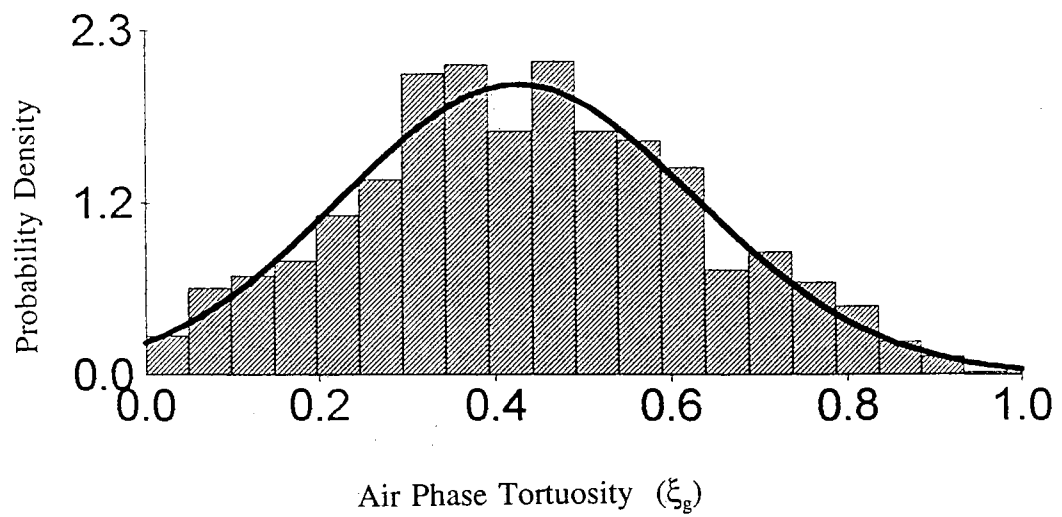
Figure 65. Distribution of air phase tortuosity (ξ_g)
            with $C_v=0.5$(The bars represent the
            actual distribution and the solid line
            is the normal distribution N (0.42,
            0.2)).
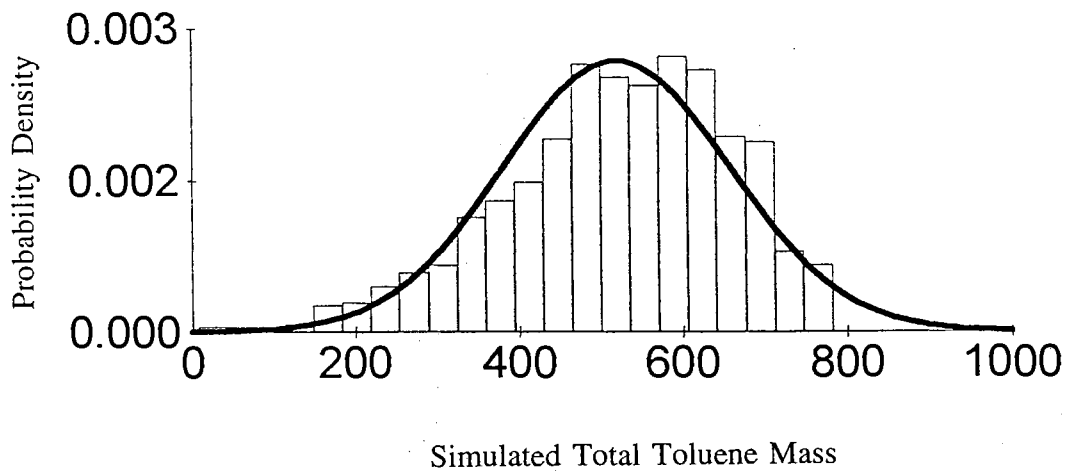
Figure 66. Distribution of simulated total toluene mass (The bars represent the actual distribution and the solid line is the normal distribution N (520, 139)).

166
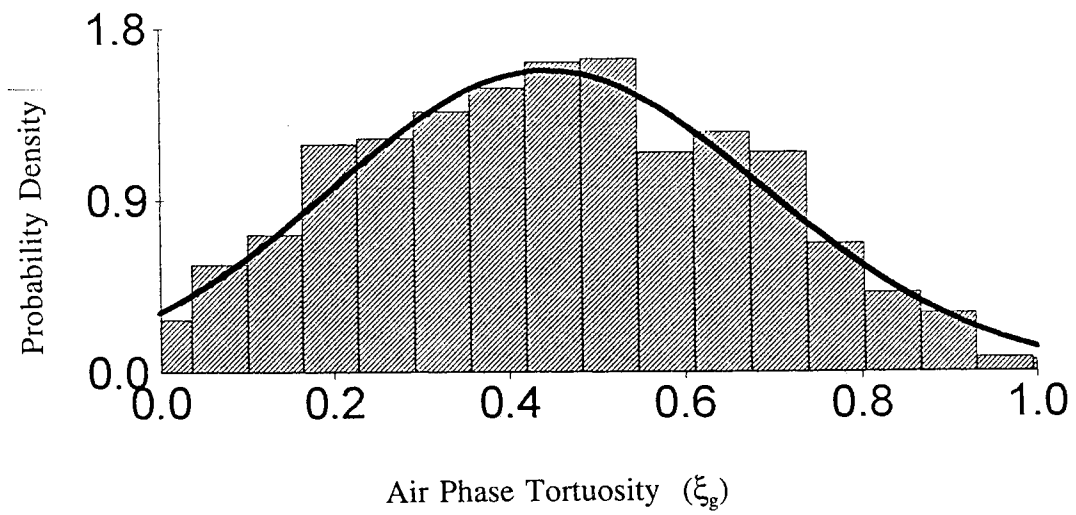
Figure 67.   Distribution of air phase tortuosity (ξ_g)
with $C_v=0.6$ (The bars represent the
actual distribution and the solid line
is the normal distribution N (0.42,
0.25)).
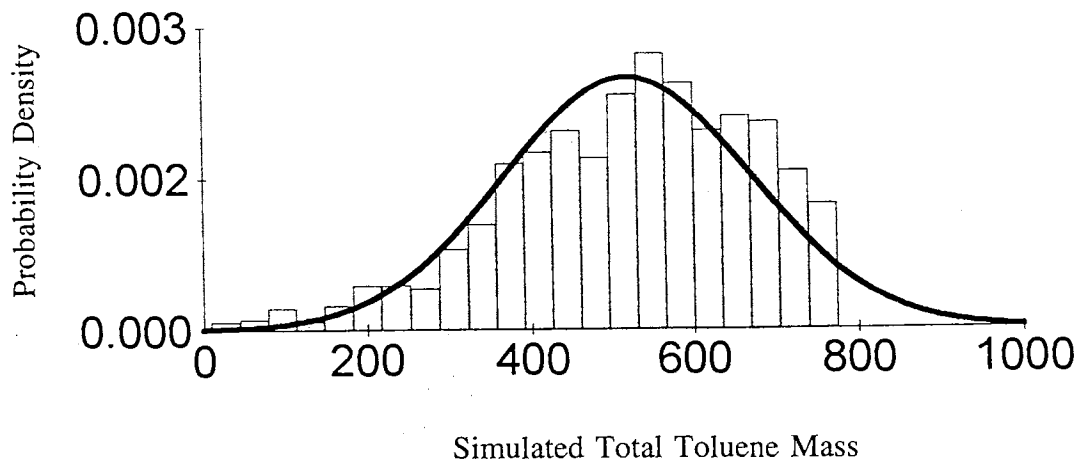
Figure 68. Distribution of simulated total toluene
(The bars represent the actual
distribution and the solid line is the
normal distribution N(519, 152)).

168
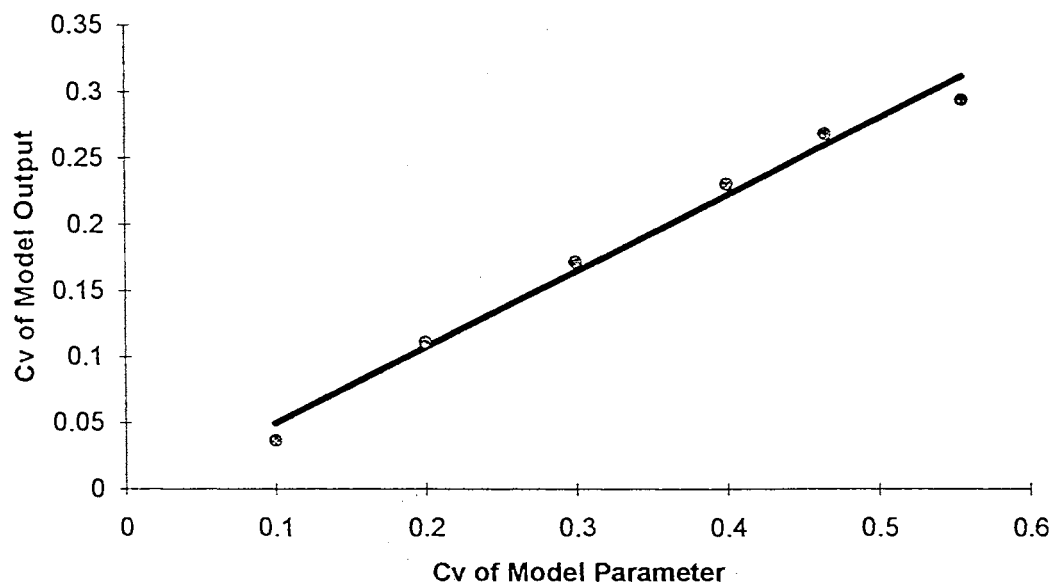
Figure 69. The relationship between uncertainty in
            simulated total toluene mass and
            uncertainty in parameters (K_h and K_d)
            (The solid line is regression equation
            and filled circles are the actual
            relationship).

169

BIBLIOGRAPHY

Abriola, L.M. and Pinder, G.F. (1985). A multiphase approach to the modeling of porous media contamination by organic compounds, 1. equation development, 2. numerical simulation. Water Resources Research, 21(1),11-26.

Andrews, R., LaVenne, A.M., and McNeish, J.A. (1987). Evaluating the effect of sampling and spatial correlation on groundwater travel time uncertainty coupling geostatistical, stochastic, and first order, second moment methods. Proceedings of the conference on geostatistical, sensitivity, and uncertainty methods for groundwater flow and radionuclide transport modeling. San Francisco, California, September 15-17, 839-559.

Ashworth, R.A., Howe, G.B., Mullins, M.E., and Rogers, T.N. (1988). Air-water partitioning coefficients of organics in dilute aqueous solutions. Journal of Hazardous MaterialS, 18, 25-36.

Baehr, A.L. and M.Y. Corapcioglu. (1987). A compositional multiphase model for groundwater contamination by petroleum products: 2. Numerical solution. Water Resources Research, 23(1):201-213.

Baehr, A.L. (1987). Selective transport of hydrocarbons in the unsaturated zone due to aqueous and vapor phase partitioning. Water Resources Research, 23(10):1926-1938.

Baehr, A.L. and Bruell, C.J. (1990). Application of the Stefan-Maxwell equations to determine limitations of Fick's Law when modeling organic vapor transport in sand columns. Water Resources Research, 26(6):1155-1163.

Baehr, A., Hoag, G. E., and Marley, M. C. (1989). Removing volatile contaminants from the unsaturated zone by inducing advective air-phase transport. Elsevier Science Publishers B. V.

Box, G.E.P. and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley Publishing Co., Reading, MA.

Brown, G.O. (1987). Solute transport by a volatile solvent. Ph.D. Dissertation, Department of Civil Engineering, Colorado State University, Fort Collins, Colorado.

Brown, G.O. (1991). Volatile organic transport project funding proposal. Oklahoma State University, Stillwater, Oklahoma.

Brown, G.O. and Mcwhorter (1990). Solute transport by a volatile solvent. Journal of Contaminant Hydrology. 5, 387-402.

Browne, T. E. and Cohen, Y. (1990). Aqueous-phase adsorption of trichloroethane and chloroform onto polymeric resins and activated carbon. Ind. Eng. Chem. Res., 29:1338-1345.

Brusseau, M.L. (1991). Transport of organic chemicals by gas advection in structured or heterogeneous porous media: development of a model and application to column experiments. Water Resources Research, 27(12):3189-3199.

Carcel, R.F. and Smith, C.N. (1987). Impact of pesticides on ground water contamination. Silent Spring Revised. American Chemical Society, Washington DC, 71-84.

Carrera, J. and Neuman, S.P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 1.Maximum likelihood method incorporating prior information. Water Resources Research, 22(2):199-210.

Carrera, J. and Neuman, S.P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 2.Uniqueness, stability, and solution algorithms. Water Resources Research, 22(2):2117-227.

Carrera, J. and Neuman, S.P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 3.Application to synthetic and field data. Water Resources Research, 22(2):228:242.

Carson, R. (1962). Silent Spring. Fawcett Crest, New York.

Cooley, R.L. (1983). Incorporation of prior information on parameters into nonlinear regression groundwater flow models 2. Applications. Water Resources Research, 19(3):662-676.

Corapcioglu, M.Y. and Baehr, A.L. (1987). A Compositional multiphase-Model for groundwater contamination by petroleum products, 1. Theoretical considerations,2. Numerical solution. Water Resources Research, 23(1):191-213.

Corapcioglu, M.Y. and Hossain, M.A. (1989). Ground water contamination by high-density immiscible hydrocarbon slugs in gravity-driven gravel aquifers. Ground Water, 28(3):403-412.

Crittenden, J.C., Hutzler, N.J., Geyer, D.G., Oravitz, J.L., and Friedman, G. (1986). Transport of organic compounds with saturated groundwater flow: Model development and parameter sensitivity. Water Resources Research, 22(3):271-284.

Cukier, R.I., Levine, H.B., and Shuler, K.E. (1978). Nonlinear sensitivity analysis of multiparameter model systems. Journal of Computational Physics, 26,1-42.

Davis, P.J. and Rabinowitz, P. (1975). Methods of Numerical Integration. Academic Press, Inc. New York.

Dilks, D.W., Canale, R.P., and Meier, P.G. (1992). Development of Bayesian Monte Carlo techniques for water quality model uncertainty. Ecological Modelling, 62,149-162.

Doctor, P.G. (1989). Sensitivity and uncertainty analyses for performance assessment modeling, Engineering Geology, 26,411-429.

Edwards, D.R. (1988). Incorporating parametric uncertainty into flood estimation methodologies for ungaged watersheds and for watersheds with short records. Ph.D. thesis. Agricultural Engineering Department, Oklahoma State University, Stillwater, Oklahoma.

Falta, R.W., Javandel, I., Pruess, K., and Witherspoon, P.A. (1989). Density-driven flow of gas in the unsaturated zone due to the evaporation of volatile organic compounds. Water Resources Research, 25(10), 2159-2169.

Garbarini, D.R. and Lion, L.W. (1985). Evaluation of sorptive partitioning of nonionic pollutants in closed systems by headspace analysis. Environ.Sci.Technol, 19, 1122-1128.

Gavalas, G.R., Shah, P.C., and Seinfeld, T.H. (1976). Reservoir history matching by Bayesian estimation. Soc. Pet. Eng. J., 16, 337-350.

Gierke,J.S., Hutzler, N.J., and Crittenden, J.C. (1990). Modeling the movement of volatile organic chemicals in columns of unsaturated soil. Water Resources Research, 26(7), 1529-1547.

Gossett, J.M. (1987). Measurement of Henry's Law constants for C1 and C2 chlorinated hydrocarbons. Environ. Sci. Thechnol., 21, 202-208.

172

Haan, C.T. (1977). _Statistical Methods in Hydrology_. Ames. Iowa State Press.

Hunt, J. R. and Sitar, N. (1988). Nonaqueous phase liquid transport and cleanup, 1. Analysis of mechanisms, 2. Experimental studies. _Water Resources Research_, 24, 1247-1269.

Ince, N. and Inel, Y. (1991). A Semi-empirical approach to relate the volatilization rates of organic chemicals to their physical properties. _Water Research_, 25(8), 903-910.

Jacquard, P. and Jain, C. (1965). Permeability distribution from field pressure data. _Soc. Pet. Eng. J._, 281-294.

Jones, R.L., Black, G.W., and Estes, T.L. (1986). Comparison of computer model predictions with unsaturated zone field data for aldicarb and aldoxycarb. _Environmental Toxicology and Chemistry_, 5, 1027-1037.

Jury, W. A., Dyson, J. S., and Butters, G. L. (1990). Transfer function model of field-scale solute transport under transient water flow. _Soil Sci. Soc. Am. J._, 54, 327-332.

Jury, W.A., Farmer, W.J., and Spencer, W.F. (1984). Behavior assessment model for trace organics in soil: II.Chemical classification and parameter sensitivity, III. Application of screening model, IV. Review of experimental evidence. _J. Environ. Qual._, 13(4), 567-586.

Jury, W.A., Russo, D., Streile, G., and Abd, H.E. (1990). Evaluation of volatilization by organic chemicals residing below the soil surface. _Water Resources Research_, 26(1),13-20.

Jury, W.A., A.M Winer, W.F. Spencer, and D.D. Focht. (1987). Transport and transformations of organic chemicals in the soil-air-water ecosystem. _Reviews of Enviromental Contamination and Toxicology_, 99, 119-165.

Jury, W.A. and Sposito, G. (1985). Field calibration and validation of solute transport models for the unsaturated zone. _Soil Sci. Soc. Am. J._, 49,1331-1341.

Keidser, A. and Rosbjerg, D. (1991). A comparison of four inverse approaches to groundwater flow and transport parameter identification. _Water Resources Research_, 27(9), 2219-2232.

Khaleel, R., Reddy, R., and Overcash, M.R. (1980). Transport of potential pollutants in runoff water from land areas receiving animal wastes: a review. _Water Research_, 14, 421-436.

Knighton, R. E. and Wagenet, R.J. (1987). Simulation of solute transport using a continuous time Markov process, 1. Theory and steady state application, 2. Application to Transient field conditions. Water Resources Research, 23, 1917-1925.

Knopman, D.S. and Voss, C.I. (1989). Multiobjective sampling design for parameter estimation and model discrimination in groundwater solute transport. Water Resources Research, 25(10), 2245-2258.

Kool, J.B. and Parker, J.C. (1988). Analysis of the inverse problem for transient unsaturated flow. Water Resources Research, 24(6), 817-830.

Kool, J.B., Parker, J.C., and Van Genuchten, M.Th. (1986). Parameter estimation for unsaturated flow and transport models - A review. Journal of Hydorlogy, 91, 255-293.

Korganoff, A. (1970). Sur la resolution de problems "inverse" en hydrogeologie, Bull. Int. Assoc. Sci. Hydrol., 15(2), 67-78.

Kreamer, D.K., Weeks, E.P., and Thompson, G.M. (1988). A field technique to measure the tortuosity and sorption-affected porosity for gaseous diffusion of materials in the unsaturated zone with experimental results from near Barnwell, South Carolina. Water Resources Research, 24(3), 331-341.

Lai, S.H., Tiedje, J.M. and Erickson, A.E. (1976). In situ measurement of gas diffusion coefficient in soil. Soil Sci. Soc. Am. Proc. 40(1), 3-6.

Liu, K-H., Enfield, C.G., and Mravik, S.C. (1991). Evaluation of sorption models in the simulation of Naphthalene transport through saturated soils. Ground Water, 29(5), 685-692.

Loague, K. and Green, R.E. (1990). Statistical and graphical methods for evaluating solute transport models: Overview and application. Journal of Contaminant Hydrology, 7, 51-73.

Lyman, W.J. (1990). Adsorption coefficient for soils and sediments. Handbook of Chemical Property Estimation Methods, American Chemical Society, Washington DC. pp. 4-1 to 4-33.

Mackay, D. and Shiu, W.Y. (1981). A critical review of henry's Law constants for chemicals of environmental interest. J. Phys. Chem. Ref. Data. 10(4), 1175-1683.

Mackay, D., Shiu, W.Y., Maijanen, A., and Feenstra, S. (1991). Dissolution of non-aqueous phase liquids in groundwater. J. of Contaminant Hydrology, 8, 23-42.

MacQuarrie, K.T.B., Sudicky, E.A., and Frind, E.O. (1990). Simulation of biodegradable organic contaminants in groundwater, 1. Numerical formulation in principal directions, 2. Plume behavior in uniform and random flow fields. <u>Water Resources Research</u>, <u>26</u>(2), 207-239.

Marshall, T.J. (1959). The diffusion of gases through porous media, <u>J. Soil Sci.</u> <u>10</u>,79-82.

Marsili-Libelli, S. (1992). Parameter estimation of ecological models. <u>Ecol. Modelling.</u> <u>62</u>, 233-258.

Marsili-Libelli, S. and Castelli, M. (1987). An adaptive search algorithm for numerical optimization. <u>Ecological Modeling</u>, <u>62</u>, 341-357.

Melancon, S.M., Pollard, J.E., and Hern, S.C. (1986). Evaluation of Sesoil, Przm and Pestan in a laboratory column leaching experiment. <u>Environmental Toxicology and Chemistry</u>, <u>5</u>, 865-878.

Mendoza, C.A. and Frind, E.O. (1990). Advective-dispersive transport of dense organic vapors in the unsaturated zone, 1: Model development. <u>Water Resources Research</u>, <u>26</u>(3), 379-387.

Mendoza, C.A. and McAlary, T.A. (1990). Modeling of ground-water contamination caused by organic solvent vapors. <u>Ground Water</u>, <u>28</u>(2), 199-206.

Metcalfe, D.E. and Farquhar, G.J. (1986). Modeling gas migration through unsaturated soils from waste disposal sites. <u>Water, Air, and Soil Pollution</u>. <u>32</u>, 247-259.

Mishra, S. and Parker, J.C. (1989). Effects of parameter uncertainty on predictions of unsaturated flow. <u>Journal of Hydrology</u>, <u>108</u>, 19-33.

Moseley, F. and Leger, C. (1988). <u>Groundwater Cleanup and Protection: Selected National Issues</u>. North West-Midwest Institute, Washington, D. C.

Murty, V.V.N. and Scott, V.H. (1977). Determination of transport model parameters in groundwater aquifers. <u>Water Resources Research</u>, <u>13</u>(6), 941-947.

Nelder, A.J. and Mead, R. (1964). A Simplex method for function minimization. <u>Computer Journal</u>. pp. 308.

Neuman, S.P. and Uakowitz, S. (1979). A statistical approach to the inverse problem of aquifer hydrology, 1. Theory. <u>Water Resources Research</u>. <u>15</u>(4), 845-860.

Pennell, K.D., Hornsby, A.G., Jessup, R.E., and Rao, P.S.C. (1990). Evaluation of five simulation models for predicting Aldicarb and Bromide behavior under field conditions. Water Resources Research, 26(11), 2679-2693.

Peterson, M.S., Lion, L.W., and Shoemaker, C.A. (1988). Influence of vapor-phase sorption and diffusion on the fate of trichloroethylene in an unsaturated aquifer system. Environ. Sci. Technol., 22, 571-578.

Roll, J. (1995). Transient multiphase volatile organic transport in soils. Ph.D dissertation in preparation. Oklahoma State University, Stillwater, Oklahoma.

Rose, K.A., Cook, R.B., Brenkert, A.L., Gardner, R.H., and Hettelingh, J.P. (1991). Systematic comparison of ILWAS,MAGIC,and ETD watershed acidification models, 1. Mapping among model inputs and deterministic results, 2. Monte Carlo analysis under regional variability. Water Resources Research, 27(10), 2257-2603.

Robinstein, R.Y. (1981). Simulation and the Monte Carlo Method. John Wiley & Sons, Inc. New York.

Russo, D. (1991). Stochastic analysis of simulated vadose zone solute transport in a vertical cross section of heterogeneous soil during nonsteady water flow. Water Resources Research, 27(3), 267-283.

Sardin, M., Schweich, D., Leij, F.J., and van Genuchten, M.T. (1991). Modeling the nonequilibrium transport of linearly interacting solutes in porous media: a review. Water Resources Research, 27(9), 2287-2307.

Schwefel, H.P. (1981). Numerical Optimization of Computer Models. Chichester, New York.

Shoemaker, C.A., Culver, T.B., Lion, L.W., and Petersn, M.G. (1990). Analytical models of the impact of two-phase sorption on subsurface transport of volatile chemicals. Water Resources Research, 26(4), 745-758.

Silka, L.R. (1988). Simulation of vapor transport through the unsaturated zone-interpretation of soil-gas surveys. GWMR, 115-123.

Sleep, B.E. and Sykes, J.F. (1989). Modeling the transport of volatile organics in variably saturated media. Water Resources Research, 25(1), 81-92.

Strecker, E.W. and Chu, W.S. 1986. Parameter identification of a groundwater contaminant transport model. Groundwater, 24(1),56-62.

Taylor, D.A. and Bender, D.A. (1988). Simulating correlated lumber properties using a modified multivariate normal approach. Trans. ASAE 31(1), 182-186.

Thorstenson, D. C. and Pollock, D.W. (1989). Gas transport in unsaturated zones: Multicomponent systems and the adequacy of Fick's Laws. Water Resources Research, 25(3),477-507.

Tiscareno-Lopez, M., Lopes, V.L., Stone, J.J., and Lane, L.J. (1993). Sensitivity analysis of the WEPP watershed model for rangeland applications I: Hillslope processes. ASAE, 36(6), 1659-1672.

Umari, A., Willis, R., and Liu, P. (1979). Identification of aquifer dispersivities in two-dimensional transient groundwater contaminant transport: An optimization approach. Water Resources Research, 15(4), 815-831.

van der Zee, S.E.A.T.M. (1990). Analysis of solute redistribution in a heterogeneous field. Water Resources Research, 26(2),273-278.

Vecchia, A.V. and Cooley, R.L. (1987). Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. Water Resources Research, 23(7), 1237-1250.

Villeneuve, J-P., Lafrance, P., Banton, O., Frechette, P., and Robert, C. (1988). A sensitivity analysis of adsorption and degradation parameters in the modeling of pesticide transport in soils. Journal of Contaminant Hydrology, 3, 77-96.

Wagner, B.J. and Gorelick, S.M. (1986). A statistical methodology for estimation transport parameters: Theory and applications to one dimensional advective-dispersive systems. Water Resources Research, 22(8), 1303-1315.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., and Rowe, C.M. (1985). Statistics for the evaluation and comparison of models. Journal of Geophysical Research, 90(c5), 8995-9005.

Wilson, B. (1990). Experimental data analysis class notes. Oklahoma State University. Stillwater, Oklahoma.

Yamaguchi, T., Moldrup, P., and Yodosi, S. (1989). Using breakthrough curves for parameter estimation in the convection-dispersion model of solute transport. <u>Soil Sci. Soc. Am. J.</u>, <u>53</u>, 1635-1641.

Yan, J.S. (1990). Parameter estimation for multipurpose hydrologic models. Ph.D. Thesis. Agricultural Engineering Department, Oklahoma State University, Oklahoma.

Yeh, W.W-G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. <u>Water Resources Research</u>, <u>22</u>(2), 95-108.

Yu, M. (1995). Boundary layor affects volatilization of organic solutes. Ph.D dissertation in preparation. Oklahoma State University, Stillwater, Oklahoma.

Zhang, H. (1990). Assessment of uncertainties of estimated solute transport to ground water. Ph.D. Thesis. Environmental Science, Oklahoma State University, Oklahoma.

APPENDIX

Computer Program for Optimal Search

```
C*****************************************************************
C*    This program searches the minimum using modified          *
C*              adaptive Simplex method                         *
C*--------------------------------------------------------------*
C*              Function of Subroutines                         *
C*                                                              *
C* Subroutine Baehr:                                            *
C*    This is the multiphase compositional model developed by*
C*    Baehr (1987). All the inputs this model requires are      *
C*    contained in this subroutine except parameters to be      *
C*    estimated, which are wired in from the main program.     *
C*    The computation of objective function was added to this*
C*    subroutine and the result was fed back to the main        *
C*    program.                                                  *
C*                                                              *
C* Subroutine Expansion:                                        *
C*    This subroutine was designed to expand a simplex until   *
C*    the local minimum has been reached.                       *
C*                                                              *
C* Subroutine GRN:                                              *
C*    This subroutine generates random numbers from N(0,1)     *
C*    (Zhang, 1990)                                            *
C*--------------------------------------------------------------*
C*              Definition of Variables                         *
C*                                                              *
C*    NHZ  : Number of data points in a concentration profiles *
C*    NP   : Number of parameters                               *
C*    MP   : Number of simplex vertices (MP=NP+1)               *
C*    PRMT : Parameter value at each vertex                     *
C*    YS0  : Measured data                                      *
C*    EE   : Objective function value                           *
C*    FTOL : Error tolerance standard                           *
C*    RTOL : Relative error tolerance                           *
C*    ITMAX : Maximum number of iterations                      *
C*****************************************************************
      PARAMETER(NHZ=17,NP=2,MP=3,NMAX=6)
      COMMON/SIMP/PRMT(3),YS,EE
      COMMON/EXPA/PR,PCEN,ALPHA,DELTA,NDIM,S,P,YPR,IHI
      COMMON/NORM/ZN
      DIMENSION P(NP,MP),Y(MP),PR(NMAX),PRR(NMAX),PCEN(NMAX),
     * YS(NHZ,2),P0(NP,MP)
      DIMENSION S(2),YS0(NHZ,2),ZN(200),SIGMA(100)
      DATA FTOL,ITMAX,ALPHA,BETA/1.0E-10,100,1.0,0.5/
      OPEN(UNIT=5,FILE='START.DAT')
C*****This file contains starting points for each vertex.
      OPEN(UNIT=6,FILE='SIGMA.DAT')
C*****This file contains standard deviation in observations.
      OPEN(UNIT=7,FILE='OPTIM.DAT')
C*****This file contains the optimal estimates.
      OPEN(UNIT=8,FILE='OBS.DAT')
C*****This file contains measured data.
      read(5,*) DELTA,((p0(ii,jj),II=1,NP),JJ=1,MP)
      read(8,*) ((ys0(iz,j),IZ=1,NHZ),J=1,2)
      WRITE(*,*)'TYPE IN THE NUMBER OF DIFFERENT STD'
      READ(*,*)NS1
      WRITE(*,*)'TYPE IN THE SAMPLE SIZE'
      READ(*,*)NS2
      READ(6,*)(SIGMA(I),I=1,NS1)
      DO 111 KK=1,NS1
      WRITE(7,*)'SIGMA=',SIGMA(KK)
      DO 111 KJ=1,NS2
      WRITE(*,*)'SAMPLE=',KJ,'SIGMA=',SIGMA(KK)
      CALL GRN(NHZ)
```

```
C*****Corrupt measured data with noise
      DO 223 J=1,2
      DO 224 IZ=1,NHZ
      IZJ=IZ+(J-1)*NHZ
      YS(IZ,J)=YS0(IZ,J)+SIGMA(KK)*ZN(IZJ)
 224  CONTINUE
 223  CONTINUE
      DO 323 I=1,MP
      DO 323 J=1,NP
      P(J,I)=P0(J,I)
 323  CONTINUE
      NDIM=NP
C*****Run transport model on starting points
      DO 122 I=1,MP
      DO 121 J=1,NP
      PRMT(J)=P(J,I)
 121  CONTINUE
      CALL BAEHR
      Y(I)=EE
 122  CONTINUE
      MPTS=NDIM+1
      ITER=0
 1    ILO=1
C      write(*,*)'iter=',iter
C      write(7,*)'iteration=',iter
C      WRITE(7,'(/1X,A)') 'Vertices of final 2-D simplex and'
C      WRITE(7,'(1X,A)') 'objective values at the vertices:'
C      WRITE(7,'(/3X,A,T11,A,T23,A,T35,A/)') 'I',
C     *'CA','HAW','OBJECTIVE'
C      DO 313 I=1,MP
C      WRITE(*,'(1X,I3,4F15.5)') I,(P(J,I),J=1,NP),Y(I)
C 313     CONTINUE

C*****Find the vertex with the worst objective function value
      iF(Y(1).GT.Y(2))THEN
         IHI=1
         INHI=2
      ELSE
         IHI=2
         INHI=1
      ENDIF
      DO 11 I=1,MPTS
      IF(Y(I).LT.Y(ILO)) ILO=I
      IF(Y(I).GT.Y(IHI))THEN
         INHI=IHI
         IHI=I
      ELSE IF(Y(I).GT.Y(INHI))THEN
       IF(I.NE.IHI) INHI=I
      ENDIF
 11   CONTINUE
C*****Calculate the relative error
      RTOL=2.*ABS(Y(IHI)-Y(ILO))/(ABS(Y(IHI))+ABS(Y(ILO)))
      IF(RTOL.LT.FTOL)GOTO 999
      IF(ITER.EQ.ITMAX) goto 999
      ITER=ITER+1
      DO 12 J=1,NDIM
      PCEN(J)=0.
 12   CONTINUE
C*****Calculate the coordinates of the simplex mass center
      DO 14 I=1,MPTS
      IF(I.NE.IHI)THEN
        DO 13 J=1,NDIM
        PCEN(J)=PCEN(J)+P(J,I)
```

```fortran
13      CONTINUE
        ENDIF
14      CONTINUE
C*****Calculate the reflection point of the worst vertex and run transport
C*****model on the new simplex
        DO 15 J=1,NDIM
        PCEN(J)=PCEN(J)/NDIM
        PR(J)=PCEN(J)+ALPHA*(PCEN(J)-P(J,IHI))
        if(pr(j).lE.0.0) pr(j)=1.0E-5
        PRMT(J)=PR(J)
15      CONTINUE
        CALL BAEHR
        YPR=EE
C*****If the reflection point gets the best objective function value,
C*****search along this direction until a local minmum is reached.
        IF(YPR.LE.Y(ILO))THEN
          CALL EXPANSION
          DO 18 J=1,NDIM
          P(J,IHI)=S(J)
18        CONTINUE
          Y(IHI)=YPR
C*****If the reflection point results in worse objective function value,
C*****contract the simplex.
        ELSE IF(YPR.GE.Y(INHI))THEN
          IF(YPR.LT.Y(IHI))THEN
            DO 19 J=1,NDIM
            P(J,IHI)=PR(J)
19          CONTINUE
            Y(IHI)=YPR
          ENDIF
          DO 21 J=1,NDIM
          PRR(J)=BETA*P(J,IHI)+(1.-BETA)*PCEN(J)
          if(prr(j).lE.0.0) prr(j)=1.0E-5
          PRMT(J)=PRR(J)
21        CONTINUE
          CALL BAEHR
          YPRR=EE
          IF(YPRR.LT.Y(IHI))THEN
            DO 22 J=1,NDIM
            P(J,IHI)=PRR(J)
22          CONTINUE
            Y(IHI)=YPRR
          ELSE
            DO 24 I=1,MPTS
            IF(I.NE.ILO)THEN
              DO 23 J=1,NDIM
              PR(J)=0.5*(P(J,I)+P(J,ILO))
              If(pr(j).lE.0.0) pr(j)=1.0E-5
              PRMT(J)=PR(J)
              P(J,I)=PR(J)
23            CONTINUE
              CALL BAEHR
              Y(I)=EE
            ENDIF
24          CONTINUE
          ENDIF
        ELSE
          DO 25 J=1,NDIM
          P(J,IHI)=PR(J)
25        CONTINUE
          Y(IHI)=YPR
        ENDIF
        GO TO 1
```

```
 999    CONTINUE
C       WRITE(7,'(/1X,A,I3)') 'Iterations: ',ITER
C       WRITE(7,'(/1X,A)') 'Vertices of final 2-D simplex and'
C       WRITE(7,'(1X,A)') 'objective values at the vertices:'
C       WRITE(7,'(/3X,A,T11,A,T23,A,T35,A/)') 'I',
C       *               'CA','HSW','OBJECTIVE'
C       DO 113 I=1,MP
C       WRITE(7,'(1X,I3,4F15.5)') I,(P(J,I),J=1,NP),Y(I)
c       write(*,*)'ee=',y(i)
C 113   CONTINUE
        WRITE(7,*)P(1,1),P(2,1)
  111   CONTINUE
        END
C
        SUBROUTINE EXPANSION
        COMMON/SIMP/PRMT(3),YS,EE
        COMMON/EXPA/PR,PCEN,ALPHA,DELTA,NDIM,S,PP,YPR,IHI
        DIMENSION PR(6),PCEN(6),YS(17,2),X1(2,100),X2(2,100),X3(2,100)
        DIMENSION FIB(1000),D(2),XP(2,1000),XQ(2,1000),R(1000),PP(2,3)
       *, S(2),P(2),Q(2),XP1(2)
        FX1=YPR
C*****Calculate the unidirectional search step.
        DO 160 J=1,NDIM
        X1(J,1)=PR(J)
  160   CONTINUE
C*****This loop adjusts DELTA until X2 is better than X1.
  100   DO 170 J=1,NDIM
        X2(J,1)=PCEN(J)+(ALPHA+DELTA)*(PCEN(J)-PP(J,IHI))
  170   PRMT(J)=X2(J,1)
        CALL BAEHR
        FX2=EE
        IF(FX2.LT.FX1) GOTO 150
        DELTA=DELTA/2
        IF(DELTA.LT.1E-4) THEN
          DO 177 J=1,NDIM
  177     S(J)=X1(J,1)
          GOTO 1000
        ENDIF
        GOTO 100
C*****FIB represents Fibonacci number which is used to decide search step.
  150   FIB(1)=1
        FIB(2)=1
        DO 400 J=1,NDIM
        D(J)=X2(J,1)-X1(J,1)
        X3(J,1)=X2(J,1)+D(J)*FIB(1)
  400   PRMT(J)=X3(J,1)
        CALL BAEHR
        FX3=EE
        K=1
        IF(FX3.GT.FX2) GOTO 425
  410   K=K+1
        IF(K.GE.3) FIB(K)=FIB(K-1)+FIB(K-2)
        DO 420 J=1,NDIM
        X1(J,K)=X2(J,K-1)
        X2(J,K)=X3(J,K-1)
        X3(J,K)=X2(J,K)+(X2(J,K)-X1(J,K))*FIB(K)/FIB(K-1)
  420   PRMT(J)=X3(J,K)
        FX2=FX3
        CALL BAEHR
        FX3=EE
C       write(*,*)'fx2=',fx2,'fx3=',fx3
        IF(FX3.LT.FX2) GOTO 410
  425   NK=K
```

183

```fortran
      IF(NK.GT.2) GOTO 460
      DO 464 J=1,NDIM
      S(J)=X2(J,NK)
      YPR=FX2
464   CONTINUE
      GOTO 1000
460   DO 470 J=1,NDIM
      P(J)=X1(J,NK)
470   Q(J)=X3(J,NK)
      DO 480 IK=NK,2,-1
      R(IK)=FIB(IK-1)/FIB(IK)
      DO 482 J=1,NDIM
      XQ(J,IK)=P(J)+R(IK)*(Q(J)-P(J))
482   PRMT(J)=XQ(J,IK)
      CALL BAEHR
      FXQ=EE
      DO 484 J=1,NDIM
      XP(J,IK)=Q(J)-R(IK)*(Q(J)-P(J))
484   PRMT(J)=XP(J,IK)
      CALL BAEHR
      FXP=EE
      IF(FXQ.LT.FXP)THEN
         DO 490 J=1,NDIM
         Q(J)=XP(J,IK)
         XP(J,IK-1)=XQ(J,IK)
490      XQ(J,IK-1)=P(J)+R(IK)*(Q(J)-P(J))
      ELSE
         DO 492 J=1,NDIM
         P(J)=XQ(J,IK)
         XQ(J,IK-1)=XP(J,IK)
492      XP(J,IK-1)=Q(J)-R(IK)*(Q(J)-P(J))
      ENDIF
480   CONTINUE
      DO 494 J=1,NDIM
      IF(FXQ.LT.FXP)THEN
         S(J)=XQ(J,IK)
         YPR=FXQ
      ELSE
         S(J)=XP(J,IK)
         YPR=FXP
      ENDIF
494   CONTINUE
1000  RETURN
      END

      SUBROUTINE GRN(N)
      COMMON/NORM/ZN
      DIMENSION ZN(200)
      DO 555 J=1,N
      CALL RANDOM(U1)
      CALL RANDOM(U2)
      ZN(J)=((-2*ALOG(U1))**0.5)*COS(2*3.14159*U2)
      ZN(J+N)=((-2*ALOG(U1))**0.5)*SIN(2*3.14159*U2)
555   CONTINUE
      RETURN
      END
```

VITA

Jian Yue

Candidate for the Degree of

Doctor of Philosophy

Thesis: PARAMETER ESTIMATION AND UNCERTAINTY FOR
            VOLATILE ORGANIC TRANSPORT

Major Field: Biosystem Engineering

Biographical:

   Personal Data: Born in Zhaotong City, Yunnan, China, on
      April 15, 1964, the daughter of Hongyuan Yue and
      Yingzheng Jing.

   Education: Graduated from Zhaotong First High School,
      Zhaotong City, Yunnan in July 1980. Received
      Bachelor of Science degree in Civil Engineering
      from Chengdu Science and Technology University,
      Chengdu, China in July 1984. Received Master of
      Science degree in Civil Engineering from Beijing
      Water Resource Conservancy and Research Institute
      in June 1987. Completed the requirements for the
      Doctor of Philosophy degree with a major in
      Biosystem Engineering at Oklahoma State University
      in December 1994.

   Experience: Employed by Kunming Dam Designing Firm as an
      engineer from June 1987 to January 1990. Employed
      by Oklahoma State University, Department of
      Biosystem Engineering as an research associate,
      January 1990 to July 1994. Employed by Oklahoma
      Department of Environmental Quality as
      environmental engineer, August 1994 to present.