ISEIS

Journal of Environmental Informatics 11(1) 11-20 (2008)

Journal of Environmental Informatics

www.iseis.org/jei

Modeling Groundwater Contamination under Uncertainty: A Factorial-Design-Based Stochastic Approach

X. S. Qin[1], G. H. Huang[2,*], and A. Chakma[3]

[1]*Faculty of Engineering, University of Regina, Regina, Saskatchewan S4S 0A2, Canada*
[2]*Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[3]*Department of Chemical Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

Received 12 December 2006; revised 21 June 2007; accepted 7 November 2007; published online 31 March 2008

ABSTRACT. A factorial-design-based stochastic modeling system (FSMS) was developed in this study to systematically investigate impacts of uncertainties associated with hydrocarbon-contaminant transport in subsurface. FSMS integrated a solute transport model, factorial analysis, and Monte Carlo technique into a general framework, and effectively analyze the individual and joint effects of input parameters' uncertainties that are associated with hydrogeological conditions. Four input parameters (i.e. the mean and the variance of permeability as well as the mean and the variance of porosity) were assumed to be of uncertain nature, and the factorial design and Monte Carlo simulation algorithm were incorporated into a groundwater flow and solute transport model developed in this study. Under each factorial experiment, a number of Monte Carlo simulations were implemented. A pilot-scale physical modeling system was used to illustrate the applicability of the proposed methodology. The simulation results reveal that the uncertainties in input parameters pose considerable influences on the predicted output; especially, variations in the mean of porosity will have significant impacts on the modeling output. The results obtained from the systematic uncertainty analysis methods proposed in this study, such as mean, standard deviation, and percentile can provide useful information for further decision-making regarding the petroleum contamination problem.

Keywords: contaminant transport, factorial design, fuzzy modeling, Monte Carlo simulation

1. Introduction

Subsurface contamination by leakage and spill of nonaqueous phase liquids (NAPLs) from petroleum industry has resulted in many environmental concerns in recent years (Chen et al., 1999). Mathematical models are generally employed to simulate fate and transport of the NAPLs for risk assessment and remediation design studies. However, many theoretical and field studies have recognized that the contaminant fate in subsurface is significantly influenced by uncertainties inherent in natural porous media, and thus may affect model predictions (Gelhar, 1993). These uncertainties generally emerge from heterogeneity of hydrogeological environment and scarcity of related data, and they can further be related to aquifer characteristics, and/or physical, chemical and biological properties of the NAPLs being released and transported. Such property parameters can vary significantly from one site to another and also exhibit great spatial variability even within the same site. For example, hydraulic conductivity that spans orders of magnitude at the same site is not uncommon. Even with data available at boreholes, considerable uncertainty usually remains between boreholes. Given the heterogeneity that exists in nature, it is simply not feasible to thoroughly define subsurface conditions at a given site. Attempting to do so would require an infinite number of borings, monitoring wells, sampling and analyses; even if this work were done, the results would still be subject to non-unique interpretations (Adams, 1995). It is therefore essential that the fate and transport of NAPLs are simulated under parameter uncertainties, in order to provide robust information for enhancing performances of remediation actions.

In the past decades, the increasing awareness for uncertainties of porous media led to an improved understanding of contaminant transport behaviors in subsurface. It has been recognized that an optimal best-fit model calibration leading to a single averaged model prediction, even with some sensitivity analysis, is of limited use and furthermore may suggest an unjustified and misleading degree of accuracy (Watts et al., 1996). A deterministic model used for risk assessment and remediation design that underestimates uncertainty inherent in subsurface might cause severe consequences such as damage to human health or property. Conversely, a site remediation system design based on a deterministic model that overestimate uncertainty would waste resources (Maqsood et al., 2003). Over the past years, various techniques were advanced and employed to address the effects of parameter uncertainty

* Corresponding author. Tel.: +1 306 5854095; fax: +1 306 3372305.
E-mail address: huang@iseis.org (G. H. Huang).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.200800106

through transport modeling, including first-order second-moment analysis, response surface method, reliability-analysis-based approaches, and fuzzy-set-based approaches (Hammonds et al., 1994; Chang et al., 1995). For examples, Kaluarachchi et al. (2000) developed a Lagrangian stochastic methodology for analyzing field-scale biodegradation of hydrocarbons in aquifers, using oxygen and nitrate as soluble electron acceptors. The subsurface heterogeneity was represented using a spatially correlated random hydraulic conductivity field with a log-normal distribution. Hu and Huang (2002) developed a nonlocal, first-order, Eulerian stochastic theory for studying reactive chemical transport in a heterogeneous, fractured porous medium under uncertainty. Li et al. (2003) proposed an integrated approach based on a modified fuzzy vertex method and applied it to the simulation of petroleum contamination in the subsurface. Maqsood (2004) developed an integrated fuzzy approach for quantifying relationships among uncertain hydrogeological parameters through techniques of fuzzy multi-attribute decision analysis and factorial design.

The stochastic method has been used a more realistic approach for evaluating uncertainties in groundwater flow and transport simulation systems in recent years (Gelhar, 1993; Mylopoulos et al., 1999; Zhu and Sykes, 2000). It may be generally classified as analytical or numerical. The first includes random parameters as coefficients in partial differential equations that describe the system to be modeled. Then the stochastic differential equations are solved, generally using spectral analysis to analyze perturbed forms of the equations. The second involves solving the governing equation using numerical methods based on the probability distributions of the input parameters, and among which the most popular approach so far is the Monte Carlo simulation (Lahkim and Garcia, 1999). Such a method consists of iterative individual sampling to produce multiple simulation realizations, and then analysis of all of the realizations to present the final output results (Chang et al. 1995; Lahkim and Garcia, 1999). The output realization is usually presented in the form of a probability distribution or a cumulative frequency distribution. Many applications with such approaches were reported in the field of subsurface flow and transport modeling during the past years (Zhu and Sykes, 2000; Clement et al., 2000; Maqsood et al., 2003).

The previous studies indicated that the key of the Monte Carlo simulation to obtaining accurate output realization is the modeler's ability to accurately describe each uncertain input parameter's distribution and statistics. This is, however, difficult to achieve when working with real world hydrogeological parameters (Freeze et al., 1990), and there exists some uncertainties in the determination of input parameters' statistics. The inherent uncertainties associated with input parameters and improper selection of variance of the parameters will considerably affect the model prediction, thus causing difficulties to corresponding decision-making. However, there were few real-world applications, with such uncertainties/sensitivities on the prediction of contaminant transport being systematically addressed. In addition, conventional sensitivity analysis in groundwater flow and transport simulation only considers changing one factor at a time, and the joint effects of factors cannot be examined. It is thus desired that an integrated approach be advanced to handle such a complexity.

Therefore, as an extension to the previous studies, this paper attempts to systematically study impacts of uncertainties associated with contaminant transport in subsurface through incorporating numerical modeling, factorial design analysis, and Monte Carlo simulation techniques. This will lead to a factorial-design-based stochastic modeling system (FSMS). A solute-transport physical modeling system will be used for numerical model calibration and verification. The study will be useful for gaining insight into the level of confidence in model prediction, and it can also lead to identification of key sources of uncertainty which will merit further research, as well as the sources of uncertainty that are less important to modeling output.

2. Methodology

2.1. Modeling of Water Flow and Solute Transport in Subsurface

The mass conservation equation for isothermal water flow in an incompressible porous medium under variably saturated condition can be described by the modified form of the Richards equation as follows (Huyakorn et al., 1984):

$$\frac{\partial q_i}{\partial x_i} = -\left(S_w S_s + \phi \frac{dS_w}{d\psi} \right) \frac{\partial \psi}{\partial t} + q \tag{1}$$

where K_{ij} is the saturated hydraulic conductivity tensor [LT^{-1}]; k_{rw} is the relative permeability; ψ is the pressure head [L]; x_i, x_j (i, j = 1, 2, 3) are the spatial coordinates [L]; t is time [T]; u_j is the unit vector pointing in the vertical direction upward; S_w is the water saturation; S_s is the specific storage [L^{-1}]; φ is porosity; q is the source/sink volumetric rate per unit volume of the porous medium [T^{-1}].

The Darcy velocity (q_i) is (Huyakorn et al., 1984):

$$q_i = -K_{ij} k_{rw} \left(\frac{\partial \psi}{\partial x_j} + u_j \right) \tag{2}$$

Assume a typical subsurface transport media has five distinct regions: (1) voids filled with air, (2) mobile water located inside the larger inter-aggregate pores or fractures, (3) immobile water located mainly in the intra-aggregate pores or in the porous media surrounding fractures, (4) a dynamic soil region, in equilibrium with the mobile phase, and (5) a stagnant soil region where mass transfer is diffusion limited. Then, a general solute transport model can be expressed as (van Genuchten and Wierenga, 1976):

$$\begin{aligned} &\frac{\partial C_{wm}}{\partial t}\left(\theta_m + f\rho k_d\right) + \frac{\partial C_{wim}}{\partial t}\left[\theta_{im} + (1-f)\rho k_d\right] \\ &= \frac{\partial}{\partial x_i}\left(\theta_m D_{ij} \frac{\partial C_{wm}}{\partial x_j} \right) - q_i \frac{\partial C_{wm}}{\partial x_i} - q_s (C_{ws} - C_{wm}) - \lambda_{wm} + H_w \end{aligned} \tag{3}$$

where θ_m and θ_{im} are fractions of the soil filled with mobile and immobile water respectively; C_{wm} and C_{wim} are the concentrations [ML^{-3}] of species w in mobile and immobile water respectively; v_i is Darcy velocity [LT^{-1}]; k_d is the linear adsorption coefficient [L^3M^{-1}], respectively; f is fraction of sorption sites which are in direct contact with mobile fluid; ρ is soil bulk density [ML^{-3}]; v_s is volumetric flow rate of fluid injection (or withdrawal) per unit volume of porous medium [T^{-1}]; C_{ws} is concentration of species w in the injected fluid [ML^{-3}]; D_{ij} is hydrodynamic dispersion tensor [L^{-2}]. The following relation should be met (van Genuchten and Wierenga, 1976):

$$\frac{\partial C_{wim}}{\partial t}\left[\theta_{im} + (1-f)\rho k_d\right] = \chi(C_{wm} - C_{wim}) \tag{4}$$

where χ is mass transfer coefficient for diffusive mass exchange between mobile and immobile phases [T^{-1}].

The detailed descriptions on constitutive relations, boundary conditions, and initial conditions can be referred to Huyakorn et al. (1984). The above governing equations can be approximated in three-dimensional space using the Galerkin finite element method (Katyal, 1997). A computationally efficient solution can be obtained by discretizing the solution domain into horizontal slices. These slices are solved in sequence individually to reduce the matrix size. The Picard iterative method is implemented to solve the species transport equations in series for respective aqueous phase concentrations (Chen et al., 1999).

2.2. Factorial Design Analysis

Prediction of contaminant fate and transport in subsurface is complicated with a variety of uncertainties and complexities. These uncertainties are normally derived from many factors related to both hydrogeological and physicochemical conditions. The degrees of influences of these factors (either in single or in multiple forms) on prediction outputs generally have large variations due to nonlinear characteristics of modeling systems. It is thus imperative that individual or joint effects of various inputs be investigated, in order to provide quantitative information for facilitating identification of factors that may pose significant uncertainty impacts.

However, the conventional sensitivity analysis in groundwater flow and transport simulation only considers changing one factor at a time; the joint effects of factors cannot be examined (Maqsood, 2004). The factorial design approach in the field of engineering experimentation can be effectively applied to a sensitivity analysis that considers many factors at the same time (Box et al., 1978). Such a method allows the determination of the coefficients of a linear model with interactions as follows (Box et al., 1978):

$$y = b_0 + \sum_i b_i x_i + \sum_{i \neq j} \sum_{j \neq i} b_{ij} x_i x_j + \ldots \tag{5}$$

where y is the response of the simulation model by changing input parameters; b_0 is the average effect; b_i is the main effect of parameter x_i; and b_{ij} is a second-order interaction effect between x_i and x_j.

The idea of a factorial design is to arrange the simulations in such a way that the variations in simulation responses, obtained under different settings (i.e., mean and variance of input parameters), can be traced back to the variations of the factors. By proper arrangements of the factor settings, it will be possible to determine not only the main effect of each factor but also the joint effects among the factors on the variations of the modeling responses. If the simulations are implemented at minimum and maximum values (two levels) for each of the k factors, the design is then called 2^k factorial design which needs 2^k sets of experimental tests. Through the factorial design, the main effect for each factor can be determined, which is the difference between two averages (Box et al., 1978):

$$b_i(main\ effect) = \bar{y}_+ - \bar{y}_- \tag{6}$$

where $\bar{y}_+$ is the average response for the factor when it takes a maximum value, and $\bar{y}_-$ is the average response for the factor when it takes a minimum value. The measure of the interaction effect of $A \times B$ is defined as a half of the difference of the effects from factor A, when factor B is at its maximum and minimum levels (Box et al., 1978):

$$AB = \frac{1}{2}[A_{B+} - A_{B-}] \tag{7}$$

Rapid calculation of the effects is possible using Yates's algorithm, which is described in detail in Box et al. (1978). If there is an interaction effect AB, this means that the influence of changing factor A will depend on the setting of factor B.

2.3. Monte Carlo Simulation

Monte Carlo simulation will be used to tackle uncertainties that can be described by probability distribution functions (PDFs). Monte Carlo techniques utilize repeated executions of numerical models to simulate stochastic processes of groundwater flow and contaminant transport. Each execution of the model produces a sample output. The output samples can then be examined statistically and distributions can be determined. The primary components of a Monte Carlo simulation include (1) probability distribution functions, (2) random number generator, (3) sampling rule, (4) scoring, (5) error estimation, (6) variance reduction techniques, and (7) parallelization and vectorization. Monte Carlo techniques have a number of advantages, such as (1) the ability to handle uncertainty and variability associated with model coefficient, (2) it can potentially be applied in deterministic modeling structure, and (3) there is a great deal of flexibility with respect to the types of probability distributions that can be used to characterize model inputs.

2.4. FSMS Development

Based on above descriptions, the subsurface modeling

system can quantitatively investigate the groundwater flow and contaminant transport in porous media; the factorial analysis is effective in examining individual or joint effects of various model inputs; the stochastic simulation techniques are capable of addressing system uncertainties. Integration of these tasks leads to a factorial-design-based stochastic modeling system (FSMS). Such a system combines advantages of different methodologies, and thus has features of tackling uncertainties in more thorough and efficient ways. Figure 1 presents a schematic of FSMS, with details being summarized as follows:

(a) For k uncertain parameters, design 2^k sets of factorial experimental tests;

(b) Generate a random number between zero and one for uncertain parameters;

(c) Transform the random number to its corresponding random variate for a lognormal statistical distribution;

(d) Substitute the random variates into the solute transport model;

(e) Calculate output contaminant concentrations within the transport model;

(f) Store the resulting output of the concentrations for further statistical analysis;

(g) Repeat steps (b) to (f) a number of times;

(h) Stop when done, and exit to step (g);

(j) Analyze the outputs and perform uncertainty analysis.

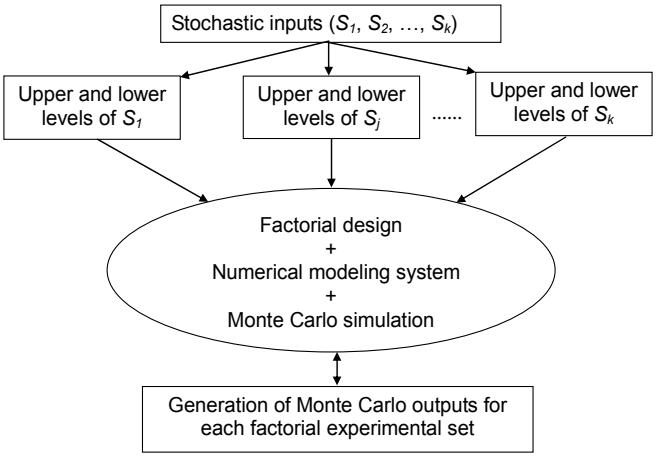

Figure 1. General framework of FFSMS.

3. Case Study

A pilot-scale experimental system was developed for simulating the groundwater flow and solute transport under uncertainties (Huang, 2004). It has features of multi-dimensionality and spatial heterogeneity. It will be examined to demonstrate the practicability and applicability of the proposed FSMS.

3.1. Pilot-Scale Physical Modeling System

In physically simulate the subsurface and support calibration and verification for numerical models, a three-dimensional pilot-scale system was developed in this study. It was of cuboid shape with an interior dimension of L × W × H = 3.6 × 1.2 × 1.4 m^3. The reactor was composed of four sections, each of which contained a supporting part, a loading manhole, and two observation windows. Water-table level gauges were installed in the first and fourth sections to monitor water depth inside the reactor. An observation window was built on the front side of each section while another on the top. The side windows were used to observe the subsurface conditions, and the top ones were built to observe the soil surface. The four sections were connected to each other with flanges, each of which had 44 bolts. Gaskets made of anti-organic solvent and anti-high temperature rubber and silicone pastern were placed between the flanges to prevent the leakage of liquid or gaseous substances. The pilot-scale system was set up to physically simulate the on-site conditions. A thermostatic room, in which the pilot-scale system and the accessorial equipment were assembled, was built to realize various temperatures by an air conditioner. Water and drainage containers were each connected to the upstream inlet and downstream outlet, respectively. A contaminant container was used to facilitate the leakage of petroleum into the system (Huang, 2004).

The monitoring wells are for facilitating access to the groundwater so that a "representative" view of the subsurface hydrogeology can be obtained, either through the collection of water samples or the measurement of physical and hydraulic parameters. Locations of the monitoring wells are presented in Figure 2. Totally there are 25 wells allocated in four sections of the pilot system. Soil in the system was stratified into four layers, with the third and forth layers being saturated with water. Each layer is 30 cm deep. Among the wells, 13 of them (with PVC pipes) were installed to reach the third soil layer; the other 12 wells could reach the forth layer. Figure 4 shows the detailed soil types and well depths. Small holes were uniformly made around the bottom sections of the pipes. Screens were used to wrap the pipes to prevent from soil clogging. Soil particles were prevented from moving into the wells while the groundwater could infiltrate into them. The wells were sealed by rubber caps at the tops. For each well, a hose was installed that passed through the caps and reached its bottom. The outside of the hose was clamped by a clip so that air and groundwater in the well were isolated from the atmosphere (Huang, 2004).

To simulate hydrocarbon leakage, 12 liters of benzene solution were injected into the bottom of the second soil layer at an upper stream location during 1.5-day period, as shown in Figure 3. At the same time, tap water from a water container was pumped into the system as groundwater inflow with a rate of 20 liter/day (through a peristaltic pump). Water level in the upstream gauge was 55 cm high and that in the downstream one was 45 cm high. After the leakage period, such flow conditions were maintained for 40 days to simulate the process of contaminant flow and transport in the subsurface. Water samples from different locations were collected every other day. Benzene concentrations in these samples were analyzed (Huang, 2004). A peristaltic pump was used to

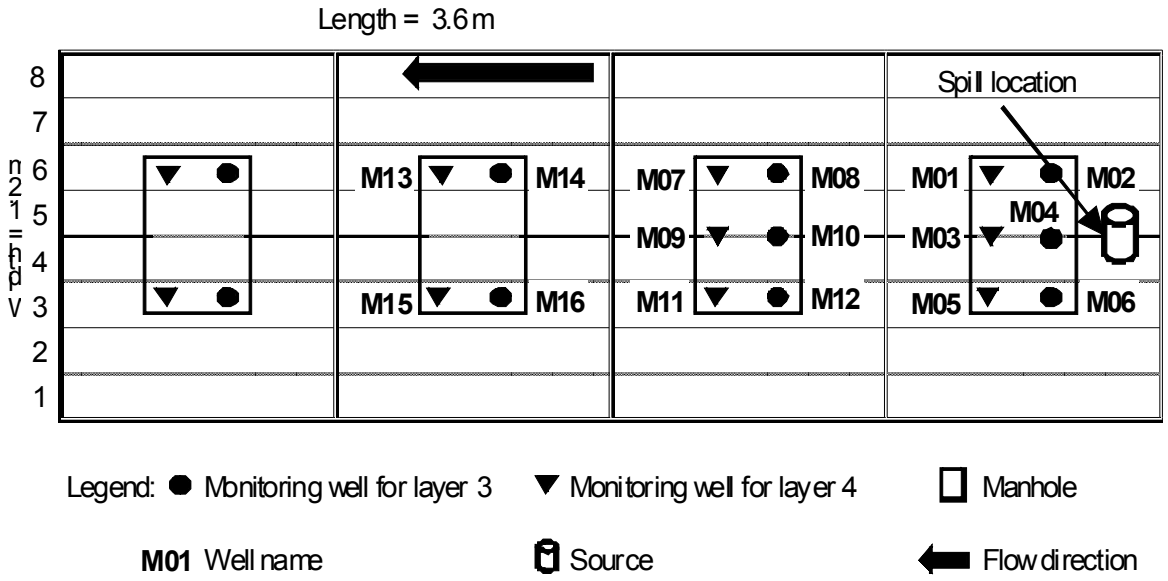

Figure 2. Well locations.

obtain groundwater samples from pilot-scale reactor through pre-installed monitoring wells. For monitoring wells, groundwater sample was collected into a 20 ml glass bottle which was sealed by a cap. Varian CP-3800 Gas Chromatograph (GC) was used for analyzing contaminant contents in water samples. The GC system was equipped with a flame ionization detector (FID), a photo ionization detector (PID), and a Chrompack WCOT fused silica 0.53 mm × 30 m capillary column. The initial oven temperature was programmed at 65 ℃. From 65 ℃, the oven temperature was increased to 135 ℃ at a rate of 10 ℃ per minute. The oven temperature was held for 5 minutes and temperatures of the injector port and the detector were maintained at 250 and 200 ℃, respectively. Helium, at an initial flow of 5.5 mL per minute, served as the carrying gas. Benzene concentrations within water samples in the pilot-scale experiment were measured. For each set of measurement, four standard benzene samples were then used to produce the standard curve. The integrated peak areas on the chromatograph charts resulted in the actual concentration of benzene based on the standard curves.

3.2. Numerical Model Development and Verification

The study system is defined as a three-dimensional (3-D) domain. The simulation domain is an area of 3.6 × 1.2 m^2 and a depth of 1.2 m. Vertically, the simulation domain has two layers, corresponding to the third and fourth simulation layers of the physical modeling system; each layer is located at the middle of the grid block that facilitates the application of a block-centered finite difference scheme. In horizontal plain, each layer is discretized into 24 × 8 grids. Each grid has dimensions of 0.15, 0.15, and 0.30 m in x, y, and z directions, respectively. The total number of grids in this 3-D computational system are 384 (24 × 8 × 2). The concerned two layers (i.e. layers 1 and 2) are located in the saturated zone.

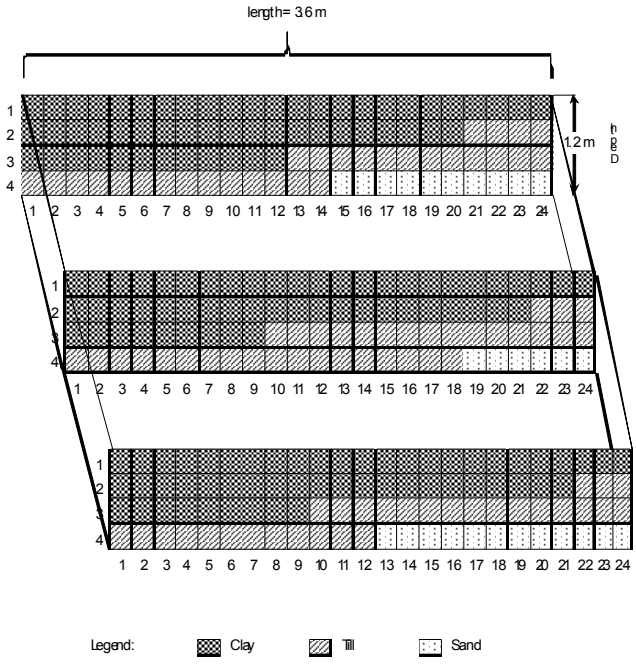

Figure 3. Soil types and well depths.

The benzene spill event initially caused a contaminated area in groundwater. The zero-flow boundary conditions were enforced at the top and bottom of the modeling domain, as well as at the sides parallel to *x*-axis. Constant hydraulic heads were employed at the left and right boundaries, allowing con-

tinuous water flow in the aquifer. Model calibration and verification are undertaken using data obtained from the pilot-scale experiments. The input parameters related to groundwater flow and benzene transport are presented in Table 1. Reasonable accuracy levels are obtained in the models for the contaminant flow and transport. Figure 4 presents the verification results for layer 1 on the 28th day as well as the temporal variations of benzene concentrations in well 10. The results demonstrated that a reasonable level of prediction accuracy is obtained. The absolute errors between the simulated and observed concentrations range from 0.08 to 0.65 mg/L with a mean of 0.34 mg/L. The root-mean-square error (RMSE) is 0.45 mg/L, and the correlation coefficient is 0.97. The RMSE is calculated using the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{16}$$

where y_i is the observed value, $\overline{y}_i$ is the predicted value, and n is the number of samples.

Table 1. Modeling Parameters

Parameter	Value
Benzene volume	12 L
Diffusive coefficient of benzene	0.00009 m^2/d
Relative viscosity of benzene	0.65
Benzene density	0.873 g/cm^3
Benzene solubility	1760 mg/L
Benzene surface tension	27.90 dyne/cm
Permeability of sand/ till/ clay	1000 MD
Porosity of sand/ till/ clay	0.35
Longitudinal dispersivity of sand/ till/ clay	0.015 m
Transverse dispersivity of sand/ till/ clay	0.015 m
Hydraulic gradient	0.03 m/m
Simulation period	40 day

It is indicated that differences between the predicted and observed concentrations are generally acceptable. A few exceptions exist due to potential sampling errors, complexities of biological processes, and subsurface stratification. In general, the numerical models developed in this study could reasonably simulate the fate and transport of the contaminant. The verified models can then be used for investigating the effects of parameter uncertainties on prediction of benzene concentrations.

3.3. Uncertainty Characterization

For many real-world sites, hydrogeological parameters could vary significantly from one site to another and exhibit high spatial variability even within the same site. In this study, soil structures are complicated with a variety of uncertainties due to system complexities, leading to imprecise/vague modeling parameters. As indicated in Table 1, two key parameters are assumed to be uncertain; they include the logarithmic value of intrinsic permeability (InK_{xx}) and the porosity (n). In this study, they are considered as stochastic parameters with normal distributions, with probability density functions (PDFs) being referred to literatures (Li et al., 2003). The mean level (M_k) and variance (σ_k^2) of InK_{xx} are [−9.5, −6.5] and [0.50, 1.0], respectively; the mean level (M_p) and variance (σ_p^2) for porosity are [0.2, 0.6] and [0.1, 0.4], respectively. The above-mentioned four factors including M_k, σ_k^2, M_p and σ_k^2 will be selected to examine impact of uncertainties on modeling outputs.

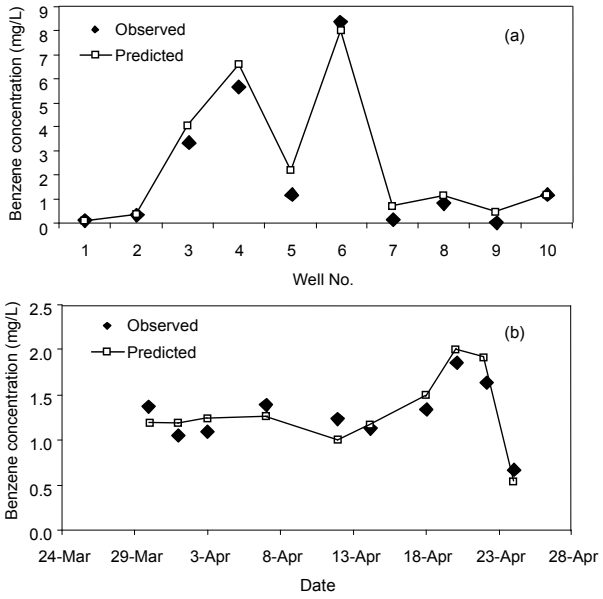

Figure 4. Verifications of modeling results: (a) spatial verification at layer 3 on the 28th day; (b) temporal verification for results of well 10 at node (16, 5).

Since these stochastic parameters have two levels obtained from upper and lower bounds of its interval. Consequently, a total of 96 (16 × 6) experiments can be designed and implemented based on the 2^4 factorial designs. Table 2 lists the combination of factor settings for each experiment. Under each experiment, 100 sets of values of porosity are created by the normal-distribution random-number generation algorithm that has been incorporated into the numerical simulator in this study. The node (22, 4) of well 6 and the node (16, 5) of well 10 in layer 1 as shown in Figure 2 are selected to analyze the predicted benzene concentrations under uncertainties. This is due to the fact that well 6 is close to the hydrocarbon source and well 10 is located at the downstream of the initial source zone.

3.4. Result Analysis

Table 3 presents the predicted benzene concentrations and the parameter uncertainty effects under 16 sets of experiments. It can be seen that the 16 experimental sets, based on

the simulated benzene concentrations, may be grouped into four major supersets: sets 1, 5, 9 and 13 may be grouped into superset 1; sets 2, 6, 10, and 14 into superset 2; sets 3, 7, 11, and 15 into superset 3; and sets 4, 8, 12, and 16 into superset 4. It appears that the differences in the means of the benzene concentrations are significant for supersets 1 and 3, as well as supersets 2 and 4, with each superset pair having different levels of M_p. This implies that the mean of porosity will have significant impacts on modeling predictions. However, the standard deviations associated with the means of the benzene concentrations are relatively insignificant for supersets 1 and 2, as well as supersets 3 and 4, with each superset having different levels of σ_p^2. This indicates that the impacts of σ_p^2 on the means of benzene concentrations are much lower than those of M_p. This fact can also be verified through observing the differences in the standard deviations of the benzene concentrations for supersets 1 and 3 (i.e. 0.0970 and 0.0792 for well 6 at the 28th day, respectively), and supersets 2 and 4 (i.e. 0.1714 and 0.1406 for well 10 at the 28th day, respectively). Nevertheless, the variations in the standard deviations of benzene concentrations are negligible for supersets 1 and 3, where lower levels of σ_p^2 exist; whereas, the impacts are considerable for supersets 2 and 4 that have higher levels of σ_p^2.

Table 2. Combination of Parameters for 2^4 Factorial Design

Set No.	Mean of $\ln K_{xx}$ (M_k)	Variance of $\ln K_{xx}$ (δ_k^2)	Mean of porosity (M_p)	Variance of porosity (δ_p^2)
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	1
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1

Tables 4 and 5 show the magnitude of input effects on the predicted benzene concentrations at wells 6 and 10, where factors A, B, C, and D represent M_k, δ_k^2, M_p, and σ_p^2, respectively. From these tables, it is found that the mean of porosity (M_p) has dominant effects on the predicted benzene concentration, followed by the mean of $\ln K_{xx}$ (M_k). From Table 4, it is found that the variance of $\ln K_{xx}$ (δ_k^2) also shows a large main effect on predicted benzene concentration at well 6. However, such an effect becomes negligible for results at well 10. From Table 5, it is shown that the interactive effect of factors A and C is notably large compared with those of other interactive effects. The variance of porosity seems to have trivial effects on the predicted benzene concentrations at both wells.

Examining the effects on the predicted benzene concentration at well 6 in the 28th day, we can obtain such insights as: (1) the main effects of factors M_k, δ_k^2, and M_p are −7.1389, −6.7441, and −17.3572 ppm, respectively, demonstrating that increasing the values of these four factors will decrease the predicted concentration; (2) the interactive effects of ACD and BCD are −2.5448 and −2.9072 ppm, respectively, and this means that the interactive effect of the mean and the variance of porosity on the modeling output may come from further interaction with the mean or the variance of $\ln K_{xx}$, since mean latter two have much larger main effects; (3) the difference between the interactive effect of AD (−0.5867 ppm) and the main effect of A (−7.1389 ppm) is significant, indicating that the effect of the variance of porosity (σ_p^2) on modeling output may mainly due to itself. It is apparent from analysis of table 4 that the mean of porosity, the mean of $\ln K_{xx}$, and the variance of $\ln K_{xx}$ have the significant impacts on the predictions, while the remaining factors have small impacts which can sometimes be neglected due to their effects are mainly due to interactions with the most sensitive factors.

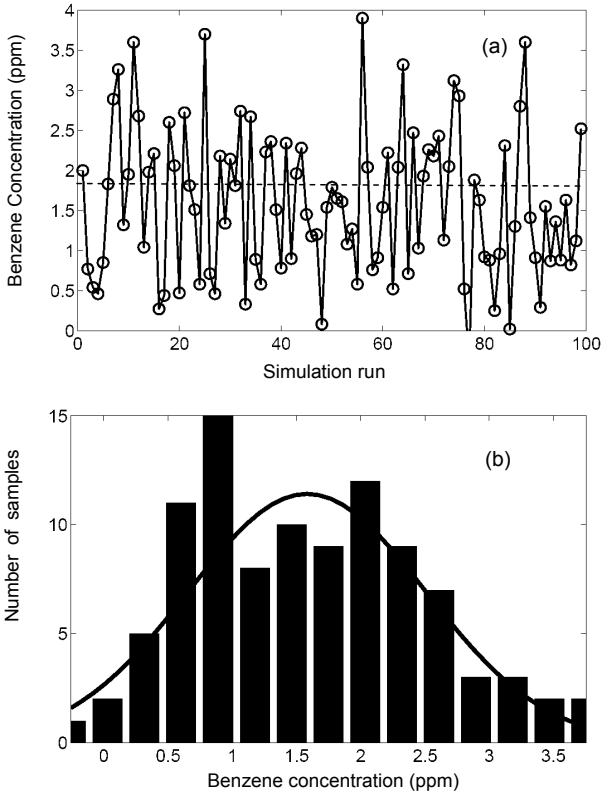

Figure 5. Monte Carlo simulation results at well 6 for experiment set 1.

The influences of uncertainties on the predicted benzene concentration for each factorial experimental set can be examined through analyzing the Monte Carlo simulation outputs. For example, Figure 5 presents the Monte Carlo simulation results of well 10 on the 28th day for experimental set 6. It is indicated that the uncertainties in the mean of porosity will sig-

Table 3. Predicted Benzene Concentrations

Experiment No.	Well 6 at the 28th day		Well 6 at the 40th day		Well 10 at the 28th day		Well 10 at the 40th day	
	Mean (ppm)	S. Dev. (ppm)	Mean (ppm)	S. Dev. (ppm)	Mean (ppm)	S. Dev. (ppm)	Mean (ppm)	S. Dev. (ppm)
1	8.6611	0.0970	7.3792	0.1218	2.0691	0.1115	2.6732	0.1495
2	7.8935	0.1885	7.3173	0.2104	1.3123	0.1714	1.7068	0.1667
3	6.7873	0.0792	6.0229	0.1045	0.8645	0.0381	1.2050	0.0464
4	7.4307	0.2356	6.1047	0.3997	0.8991	0.1406	1.2036	0.1663
5	7.7071	0.0863	6.8638	0.1289	2.2250	0.0703	2.5605	0.0689
6	8.4032	0.1011	7.3736	0.1273	1.9277	0.3384	2.2886	0.4256
7	6.2449	0.0500	5.5742	0.0716	1.0389	0.0223	1.2995	0.0250
8	6.1923	0.1776	5.5231	0.2180	0.9947	0.0912	1.2389	0.0998
9	7.6787	0.0768	6.5142	0.0887	3.3500	0.2150	3.6500	0.1874
10	8.0722	0.0752	6.6730	0.1643	3.6200	0.6697	4.1777	0.6426
11	6.4445	0.0675	5.6093	0.0877	1.3055	0.0611	1.6650	0.0671
12	6.2151	0.1865	5.5373	0.2554	1.0235	0.1225	1.7477	0.2236
13	7.1173	0.0792	6.1127	0.0958	3.2700	0.1735	3.3850	0.1498
14	7.3316	0.0672	6.2710	0.0854	2.7155	0.3606	2.7905	0.2927
15	5.8869	0.0643	5.1900	0.0934	1.5250	0.0576	1.7650	0.0604
16	5.6209	0.1457	5.0268	0.1702	2.0691	0.1115	1.4314	0.1307

nificantly affect the simulation outputs. For example, the predicted benzene concentration are 0.770 and 0.540 mg/L for simulation runs 2 and 3, respectively; and the benzene concentration are 2.814 and 3.646 mg/L for runs 87 and 88; whereas the mean concentration is 1.9277 mg/L. The Monte Carlo simulation results can also be presented in the form of probability distribution. Figure 5 presents the plotted probability density function for the predicted benzene concentrations, where the maximum and minimum values of the predicted outcomes of experiment set 6 are 0.152 and 3.901 mg/L, respectively, with the 97.5 percentile being 1.928 mg/L. The histogram of the 100 simulation runs shows normal distributions, and the outputs passed the Kolmogorov-Smirnov hypothetical test for normality.

Figures 6 and 7 present the mean and standard deviations of the predicted benzene concentrations at the 40th day for experimental sets 1 and 16, respectively. A set 1 and 16 corresponds to situations where all concerned stochastic factors are at their lower and upper bounds, respectively. It is revealed that high standard deviations exist at locations with high mean benzene concentrations, and there exist significant uncertainties in benzene concentration. Comparing Figure 6 with Figure 7 indicates that the means of benzene concentrations in experimental set 1 are much higher than those in set 16, but the corresponding standard deviations are much lower, implying that the uncertainties associated with intrinsic permeability will have significant impacts on the prediction of benzene concentrations.

Generally, the above result analysis demonstrated that proposed FSMS can efficiently incorporate uncertainties expressed not only as probability density functions but also their combinations with discrete intervals. It could be used for quantitatively analyzing the individual and combined influence of input parameters' uncertainties that are associated different hydrogeological parameters to the modeling outputs. Such studies would provide strong basis for perform successful risk assessment and efficient remediation design for the management of contaminated site.

Table 4. Estimated Effects of Uncertain Factors on Benzene Concentration at Well 6

	28th day		40th day	
	Mean (ppm)	S. Dev. (ppm)	Mean (ppm)	S. Dev. (ppm)
Ave.	8.738	0.1487	7.9272	0.1446
A	-7.1389	-0.4068	-6.9338	-0.3816
B	-6.7441	-0.378	-5.2962	-0.6682
C	-17.3572	0.378	-13.9945	0.5352
D	0.9102	0.9286	-0.5343	0.9646
AB	-0.3296	0.2183	-0.332	0.2824
AC	-0.0328	0.1546	0.7611	-0.0953
AD	-0.5867	-0.3275	0.4853	-0.1747
BC	-1.7098	-0.0439	-1.1796	-0.0818
BD	0.7956	-0.2469	0.19	-0.584
CD	-0.6354	0.63	-0.2445	0.7854
ABC	2.1428	-0.0791	0.4907	0.0622
ABD	-1.4175	0.0926	-0.7725	0.1326
ACD	-2.5448	0.0583	-1.7537	-0.3281
BCD	-2.9072	0.0331	-1.3297	-0.0271
ABCD	1.815	-0.1206	0.7502	0.0153

Table 5. Estimated Effects of Uncertain Factors on Benzene Concentration at Well 10

	28th day		40th day	
	Mean (ppm)	S. Dev. (ppm)	Mean (ppm)	S. Dev. (ppm)
Ave.	1.8361	0.1737	2.1743	0.1814
A	6.7149	0.8122	6.4362	0.6061
B	0.4895	-0.28	-1.2696	-0.3967
C	-11.6017	-1.441	-11.6762	-1.2639
D	-1.9185	1.281	-1.618	1.3935
AB	-1.5931	-0.4012	-2.4674	-0.5775
AC	-4.1279	-0.6422	-3.112	-0.3175
AD	0.2089	0.2822	0.9826	0.2563
BC	0.9159	0.1696	1.0966	0.0217
BD	-0.4501	-0.076	-0.9032	-0.1041
CD	0.7587	-0.6586	0.9922	-0.5505
ABC	1.9185	0.5516	1.7752	0.5541
ABD	-1.2115	-0.4252	-2.1738	-0.6929
ACD	-1.3303	-0.3454	-1.3604	-0.1921
BCD	0.2799	0.0428	-0.0478	-0.1585
ABCD	1.3565	0.5264	1.4596	0.6107

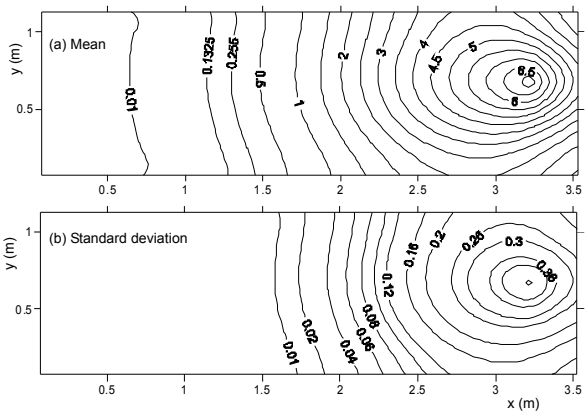

Figure 6. Simulation results of benzene concentration of under experiment set 1.

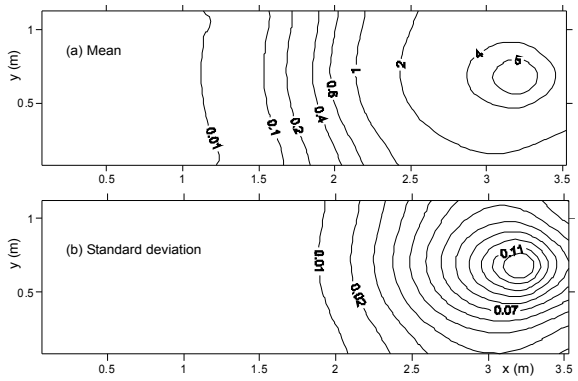

Figure 7. Simulation results of benzene concentration under experiment set 16.

4. Conclusions

A factorial-design-based stochastic modeling system (FSMS) was developed in this study to systematically investtigate impacts of uncertainties associated with hydrocarbon-contaminant transport in subsurface. FSMS integrated a solute transport model, factorial analysis, and Monte Carlo technique into a general framework, and effectively analyze the individual and joint effects of input parameters' uncertainties that are associated with hydrogeological conditions.

A pilot-scale physical modeling system was used to illustrate the applicability of the proposed methodology, and it was found that the uncertainties in the mean of porosity had more significant impacts on the predicted benzene concentrations than the mean of the logarithmic value of intrinsic permeability as well as their variances. The results obtained from the systematic uncertainty analysis methods proposed in this study, such as mean, standard deviation, and percentile can provide useful information for further decision-making regarding the petroleum contamination problem. However, as stochastic methods normally require a large amount of historical data to generate the required PDF information, the applicability of the proposed method would be significantly restricted when the available data is limited. Fuzzy approaches are promising alternatives that could handle such a difficulty and are desired to be incorporated in future studies.

Acknowledgments. This research was supported by the Major State Basic Research Development Program of MOST (2005CB724200 and 2006CB403307) and the Natural Science and Engineering Research Council of Canada.

References

Adams, R.B. (1995). *Uncertainty in the investigation and remediation of contaminated groundwater*, Geotechnical Special Publication (ASCE), 46(1), 1-15.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for Experimenters*, John Wiley & Sons, New York.

Chang, C.M., Kemblowski, M.W. and Kaluarachchi, J. (1995). Stochastic analysis of multiphase flow in porous media: 1, spectral/perturbation approach, *Stoch. Hydrol. Hydraul.*, 9(3), 239-267, doi:10.1007/BF01581722.

Chen, Z., Huang, G.H. and Chakma, A. (1999). Methodology for subsurface modeling, In Numerical Modeling, Risk Assessment, and Site Remediation, Technical Report, *Energy and Environment Program*, University of Regina: Regina, SK, Canada.

Clement, T.P., Johnson, C.D., Sun, Y.W., Klecka, G.M. and Bartlett C. (2000). Natural attenuation of chlorinated ethene compounds: model development and field-scale application at the Dover site, *J. Contam. Hydrol.*, 42(2-4), 113-140, doi:10.1016/S0169-7722 (99)00098-4.

Gelhar, L.W. (1993). *Stochastic subsurface hydrology. Prentice-Hall*, Englewood Cliffs, New Jersey.

Hammonds, J.S., Hoffman, F.O. and Bartell, S.M. (1994). An introductory guide to uncertainty analysis in environmental and health risk assessment, *Technical Rep. No. ES/ER/TM-35/R1, Oak Ridge National Laboratory*, Oak Ridge, Tenn.

Hu, B.X. and Huang, H. (2002). Stochastic analysis of reactive solute transport in heterogeneous, fractured porous media: a dual-per-

meability approach, *Transport in Porous Media*, 48, 1-39, doi: 10.1023/A:1015671227317.

Huang, Y.F. (2004). Development of environmental methodologies for supporting system simulation, optimization and process control in petroleum waste management, *Ph.D. Dissertation*, University of Regina, Regina, Saskatchewan, Canada.

Huyakorn, P.S. and Pinder, G.F. (1983). *Computational methods in subsurface flow*, Academic Press, Inc., New York, pp.473.

Kaluarachchi, J.J., Cvetkovic, V. and Berglund, S. (2000). Stochastic analysis of oxygen and nitrate based biodegradation of hydrocarbons in aquifers, *J. Contam. Hydrol.*, 41, 335-365, doi:10.1016/S0169-7722(99)00072-8.

Lahkim, M.B. and Garcia, L.A. (1999). Stochastic modeling of exposure and risk in a contaminated heterogeneous aquifer, 1: Monte Carlo uncertainty analysis, *Environmental Engineering Science*, 16 (5), 315-328.

Li, J.B., Huang, G.H., Chakma, A., Zeng, G.M. and Liu L. (2003). Integrated fuzzy-stochastic modeling of petroleum contamination in subsurface, *Energy Sources*, 25(6), 547-563.

Maqsood, I., Li, J.B. and Huang, G.H. (2003). Inexact multiphase modeling system for the management of uncertainty in subsurface contamination, *ASCE Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, 7(2), 86-94, doi:10.1061/(ASCE)1090-025X(2003)7:2(86).

Maqsood, I. (2004). Development of Simulation- and Optimization-Based Decision Support Methodologies for Environmental Systems Management, *Ph.D. Dissertation*, University of Regina, Regina, Saskatchewan, Canada.

Mylopoulos, Y.A., Theodosiou, N. and Mylopoulos, N.A. (1999). A stochastic optimization approach in the design of an aquifer remediation under hydrogeologic uncertainty, *Water Resour. Manage.*, 13 (5), 335-351, doi:10.1023/A:1008182906373.

van Genuchten, M.Th. and Wierenga, P.J. (1976). Mass transfer studies in sorbing media 1, Analytical solutions, *Soil Sci. Soc. Amer. J.*, 40, 473-480.

Watts, L., Nanin, S., McGary, R. and Silver, A. (1996). A methodology for estimating predictive uncertainty in groundwater contaminant modeling using the hydrologeochemical transport code, TRAFFIC, *Calibration and Reliability in Groundwater Modelling*, IAHS publ. No. 237, 571-578.

Zhu, J. and Sykes, J.F. (2000). Stochastic simulations of NAPL mass transport in variably saturated heterogeneous porous media, *Transport in Porous Media*, 39 (3), 289-314, doi:10.1023/A:1006669123903.