Influence of Parameter Uncertainty on Petroleum Contaminants Distribution in Porous Media

J. B. Li[1], G. H. Huang[1], G. M. Zeng[2], A.Chakma[1], and Z. Chen[1]
[1] Faculty of Engineering, University of Regina, Regina, Sask S4S 0A2, Canada
Tel: 1-306-5854095, Fax: 1-306-5854855, E-mail: huangg@uregina.ca
[2] Department of Environmental Engineering, Hunan University, Changsha, Hunan 410082, China
Tel: 86-8823701, Fax: 86-8823413, E-mail: zgming@mail.hunu.edu.cn

Abstract: A methodology based on factorial design and Monte Carlo methods is developed and implemented for incorporating uncertainties within a multiphase subsurface flow and transport simulation system. Due to uncertainties in intrinsic permeability and longitudinal dispersivity, the predicted output is also uncertain based on the well-developed multiphase compositional simulator. The simulation results reveal that the uncertainties in input parameters pose considerable influences on the predicted output, and the mean and variance of permeability will have significant impacts on the modeling output. The proposed method offers an effective tool for evaluating uncertainty in multiphase flow simulation system.

1. Introduction

Many theoretical and field studies have suggested that mathematical modeling of petroleum contaminant fate in subsurface systems is significantly influenced by presence of uncertainties in the porous media system [1]. As a result, it should accommodate this kind of uncertainties in the simulation process to evaluate the extent of contamination and to conduct successful risk assessment and remediation system design.

Monte Carlo simulation is the most popular approach to account for uncertainties so far [2]. However, the key to obtaining accurate output realization through this method, is the modeler's ability to accurately describe each uncertain input parameter's distribution and statistics. Unfortunately, this is very difficult to achieve when working with real world hydrogeological parameters [3], and there exists some uncertainties in the determination of input parameters' statistics (i.e., mean and variance). The inherent uncertainties in input parameters and erroneous selection of mean and variance of the parameters will considerably affect the model prediction, thus causing difficulties to corresponding decision-making. However, few previous work have considered such uncertainties/sensitivities on the prediction of contaminant transport in subsurface, in a systematic way. 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. Fortunately, the factorial design approach in the field of engineering experiment, can be effectively applied to the sensitivity analysis considering many factors [4]. This paper attempts to systematically study the influence of input parameters' uncertainties on modeling output by using multiphase compositional modeling approach. Two uncertain parameters including intrinsic permeability (K_{xx}) and longitudinal dispersivity (α_L), are considered to be uncertain, and the determination of the parameters' statistics are also assumed to be of uncertain nature. The factorial experiments are implemented to study the sensitivity of input parameters' mean and variance on the predicted benzene concentration in groundwater. Thus the problem of multiphase flow and transport with uncertain information is addressed, and the systematic uncertainty analysis in this paper can provide 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 not important to modeling output.

2. Methodology

2.1 Modeling NAPLs Flow and Transport

A complete description of flow and transport in the subsurface must include flow of the fluid phases (water, gas, NAPLs, etc.), mass transfer of components between these phases, and transport of components in each of the phases, therefore, the multiphase compositional simulators are recognized as effective tools in investigating such complicated processes. The multiphase compositional modeling approach is used in this paper, and the basic mass conservation equation for components in the subsurface can be written as [5, 6]:

$$\frac{\partial}{\partial t}(\phi \tilde{C}_k \rho_k) + \overset{\rho}{\nabla} \cdot [\sum_{l=1}^{n_p} \rho_k (C_{kl} \overset{\rho}{u}_l - \phi S_l \overset{\beta}{D}_{kl} \cdot \overset{\varpi}{\nabla} C_{kl})] = R_k \quad (1)$$

where k is component index, l is phase index, ϕ is soil porosity, $\tilde{C}_k$ is overall concentration of component k (volume fraction), ρ_k is density of component k [ML^{-3}], n_p is number of phases, C_{kl} is concentration of component k in phase l (volume fraction), $\overset{\varpi}{u}_l$ is Darcy velocity of phase l [LT^{-1}], S_l is the saturation of phase l, R_k is total source/sink term for component k (volume of component k per unit volume of porous media per unit time), and $\overset{\beta}{D}_{kl}$ is the dispersion tensor. The overall concentration $\tilde{C}_k$ is the volume of component k summed over all phases.

The dispersion tensor $\overset{\beta}{D}_{kl}$ can be expressed as [7].

$$\overset{\beta}{D}_{klij} = \frac{D_{m,kl}}{\tau}\delta_{ij} + \frac{\alpha_{TI}}{\phi S_l}|\overset{\varpi}{u}_l|\delta_{ij} + \frac{(\alpha_{LI} - \alpha_{TI})}{\phi S_l}\frac{u_{li}u_{lj}}{|\overset{\varpi}{u}_l|} \quad (2)$$

where τ is tortuosity (defined with a value greater than 1), $D_{m,kl}$ is molecular diffusion coefficient of component k in

phase l [L^2T^{-1}], δ_{ij} is Kronecker delta function, α_{Ll} and α_{Tl} are longitudinal and transverse dispersivities of phase l, respectively [L], u_{li} and u_{lj} are Darcy velocities of phase l in directions i and j, respectively [LT^{-1}], and $|\vec{u}_l|$ is magnitude of the vector flux for phase l [LT^{-1}].

The phase flux is calculated from the multiphase form of Darcy's law [5]:

$$\vec{u}_l = -\frac{k_{rl}\vec{K}}{\mu_l}\cdot(\vec{\nabla}P_l - \rho_l g\vec{\nabla}z) \quad (3)$$

where k_{rl} is relative permeability of porous medium to phase l, K is intrinsic permeability tensor [L^2], μ_l is viscosity of phase l [$ML^{-2}T^{-1}$], ρ_l is density of phase l [ML^{-3}], g is acceleration of gravity [LT^{-2}], z is vertical distance which is defined as positive downward [L], P_l is pressure of phase l [$ML^{-1}T^{-2}$].

By substituting Darcy's law for the phase velocity terms in the mass balance equation, Eq. (1), and summing over all components (n_{cv}), as well as noting that $\sum_{k=1}^{n_{cv}} C_{kl} = 1$, a pressure equation can be developed which is written explicitly in terms of water phase pressure as [5, 6]:

$$\phi C_t \frac{\partial P_w}{\partial t} + \vec{\nabla}\cdot\vec{K}\cdot\lambda_{rTc}\vec{\nabla}P_w = -\vec{\nabla}\cdot\sum_{l=1}^{n_p}\vec{K}\cdot\lambda_{rlc}\vec{\nabla}z + \vec{\nabla}\cdot\sum_{l=1}^{n_p}\vec{K}\cdot\lambda_{rlc}\vec{\nabla}P_{clw} + \sum_{k=1}^{n_{cv}}Q_k \quad (4)$$

where C_t is total compressibility, P_w is water phase pressure, λ_{rlc} is relative mobility, λ_{rTc} is total relative mobility, P_{clw} is capillary pressure difference between phase l and the water phase, and Q_k is injection/production rate for component k per bulk volume.

The term λ_{rlc} and λ_{rTc} can be defined as:

$$\lambda_{rlc} = \frac{k_{rl}}{\mu_l}\sum_{k=1}^{n_{cv}}\rho_k C_{kl} \quad (5)$$

$$\lambda_{rTc} = \sum_{l=1}^{n_p}\lambda_{rlc} \quad (6)$$

Compositional multiphase models require multiple constitutive relations to close the system of equations, and the typical constitutive relations include pressure-saturation-permeability (p-S-k) relations that are described by Delshad et al. [6] and Lenhard and Parker [8, 9]. The above governing equations can then be solved numerically with the block-centered finite difference approach [6].

2.2 Factorial Design and Monte Carlo Simulation

Factorial design is a classical experimental design method which allows the determination of the coefficients b_i, b_{ij}, etc. of a linear model with interactions as in the following equation:

$$y = b_0 + \sum_i b_i x_i + \sum_{i\neq j}\sum_{j\neq i} b_{ij}x_i x_j + \mathrm{K} \quad (11)$$

where y is response of the simulation model by changing input parameters x_i, x_j, etc., 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 variation in simulation response obtained with different settings of the factors, can be traced back to the variations of the factors. By proper arrangement of the factor settings, it will be possible to determine not only the main effect of each factor but also the joint effects between factors on the variations of the modeling response. If the simulations are implemented at the minimum and maximum values (two levels) of each of the k factors, the design is called 2^k factorial design which needs 2^k sets of experimental tests. Rapid calculation of effects is through Yates's algorithm which was described in detail in Box et al. [4]. If there is an interaction effect AB, this means that the influence of changing factor A will depend on the setting of factor B.

In this paper, the input parameters K_{xx} and α_L are assumed to be uncertain, and they are lognormally distributed. The mean of K_{xx} and α_L, as well as the variance of $\ln K_{xx}$ and $\ln\alpha_L$, are assumed to have uncertainties in their selection, therefore, the influences of such uncertainties on modeling response should be quantified. The maximum and minimum values of these four factors (means of K_{xx} and α_L, variances of $\ln K_{xx}$ and $\ln\alpha_L$) are determined based on literature review, and the 2^4 factorial experiments will be conducted to examine the sensitivities of these factors on modeling output. Therefore, there will be 16 sets of experimental tests, and under each set of test, 100 Monte Carlo simulation runs will be implemented based on the developed multiphase compositional simulator, and then we can systematically analyze the uncertainties in model prediction.

3. Case Study

An aquifer with the dimensions of 12 m (x-direction) × 10 m (y-direction) × 6 m (z-direction), is considered to study the NAPLs flow and transport behavior in saturated zone

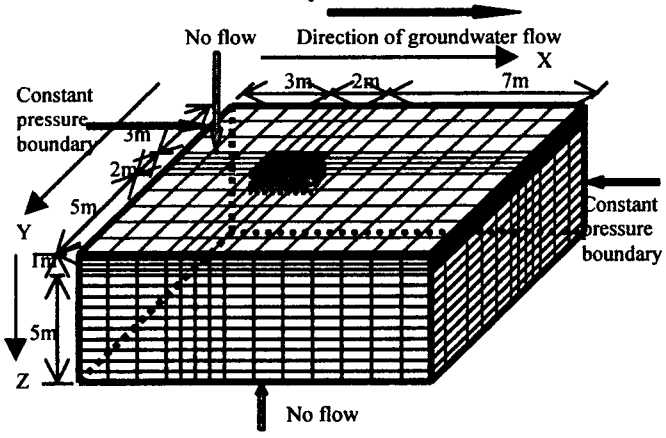

Figure 1 3D simulation domain

under the influence of hydrogeological parameter uncertainty.

The NAPLs is assumed to initially occupy a rectangle between 3 and 5 m in the x-direction, between 3 and 5 m in the y-direction, and between 0 and 1 m in the z-direction which can be shown in figure 1. The zero-flow boundary conditions are enforced at the top and bottom of the model domain, as well as at the sides parallel to x-axis. The constant hydraulic heads are employed at the left and right

boundaries, allowing water flow in the aquifer from left to right under a hydraulic gradient of 0.0167 m/m. The domain is discretized into 2352 grid cells (14 × 12 × 14) using the nodal spacing of 1 m in the x direction, 1 m in the y direction, and 0.5 m in the z direction. Near the location of NAPLs source with initial saturation, the nodal spacing is 0.5 m in the x-direction, 0.5 m in the y direction, and 0.25 m in the z direction (see figure 1).

Benzene is selected as the constituent of interest in NAPLs phase in this study, and the benzene concentration in the aquifer is simulated by the developed model. The initial NAPLs saturation in the source zone is 0.04, and the mass transfer processes between phases are assumed to be in equilibrium. The mean of K_{xx}, the variance of logarithm of K_{xx}, the mean of α_L, and the variance of logarithm of α_L, are in the range [4000, 12000] milli Darcy, [0.10, 0.40], [0.015, 0.075] m, and [0.10, 0.40], respectively.

4. Results and Analysis

In order to test the sensitivity and uncertainty of the four factors described above on the modeling response, 16 experiments were implemented based on the 2^4 factorial design. Under each experiment, 100 sets of values of K_{xx} and α_L were created by the lognormally-distributed-random-number generation algorithm which was incorporated into the multiphase compositional simulator. The Monte Carlo modeling mechanism was also embedded into the developed multiphase compositional simulator, and 100 sets of modeling output realizations and the corresponding mean and standard deviation for that experiment were then obtained by running the simulator. Node (5, 5) and node (11, 6) were selected to analyze the sensitivity/uncertainty in the predicted benzene concentration. Node (5, 5) is the 5th grid cell in the x-direction and 5th grid cell in the y-direction, and it is located at the initial NAPLs source zone. Node (11,6) is located at the downstream of the initial source zone. Figure 2 shows the mean and standard deviation of the predicted benzene concentration at node (11, 6) in layer 2 from 8 of the 16 experiments.

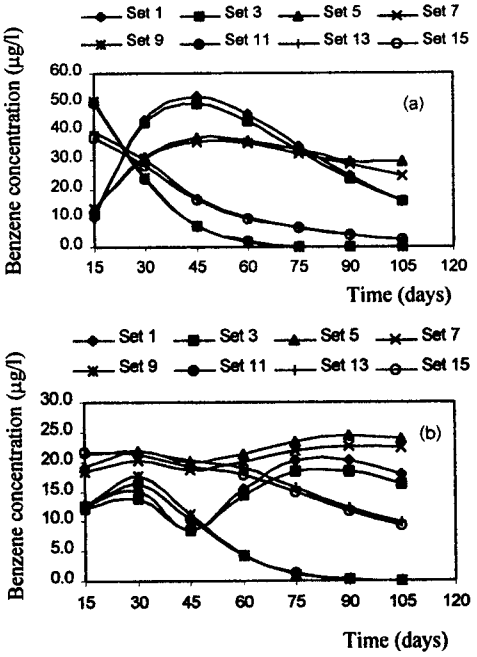

Figure 2 Mean and standard deviation of benzene concentration, (a) mean at node (11,6) and (b) standard deviation at node (11,6)

From figure 2, we can find that: (1) the changes in mean and standard deviation of benzene concentration are small for pair of sets 1 and 3, pair of sets 5 and 7, pair of sets 9 and 11, and pair of sets 13 and 15, and each pair has only different mean of α_L, and this may mean that the mean of α_L will have small influence on the modeling prediction; (2) the impacts on modeling outputs are large for pair of sets 1 and 5, pair of sets 3 and 7, pair of sets 9 and 13, and pair of sets 11 and 15, and each pair has only different variance of $\ln K_{xx}$, and this may mean that change of the variance of $\ln K_{xx}$ will have large influence on the modeling response, and (3) the change in the mean of K_{xx} will have large influence on the model response based on the comparison between group of sets 1, 3, 5, and 7 and group of sets 9, 11, 13, and 15.

Table 1 Estimated effects from the 2^4 factorial design

Factors	Effects on benzene concentration at node (11,6) in layer 2 (time: day)			
	On mean concentration (µg/l)		On standard deviation (µg/l)	
	t=15	t=45	t=15	t=45
Average	25.5	26.8	14.8	12.9
A	6.9	-34.1	0.0	0.0
B	-8.8	-4.1	4.1	6.4
C	-5.0	-3.1	-3.7	-4.2
D	-5.4	-2.1	-1.4	-1.3
AB	-11.3	9.7	-0.7	-2.2
AC	-5.6	-1.2	-1.6	-1.5
AD	-5.4	-2.1	-2.8	-3.5
BC	-4.9	-1.9	-4.0	-3.8
BD	-4.5	-2.0	-2.1	-1.4
CD	-4.7	-2.1	-1.9	-1.3
ABC	-4.5	-2.5	-2.2	-1.1
ABD	-4.5	-2.1	-3.6	-3.4
ACD	-4.7	-2.0	-3.4	-3.5
BCD	-4.3	-2.0	-2.4	-1.4
ABCD	-4.3	-2.0	-3.9	-3.4

In order to get deeper insights of the sensitivity of the factors to modeling response, the main effects and interaction effects should be quantified, and table 1 lists the magnitude of various calculated effects of the 4 factors. In table 1, factors A, B, C, and D represent mean of K_{xx}, variance of $\ln K_{xx}$, mean of α_L, and variance of $\ln \alpha_L$, respectively. From table 1, we can find that the mean of K_{xx} has the largest main effect on the predicted benzene concentration, while variance of $\ln K_{xx}$, mean of α_L, and variance of $\ln \alpha_L$ also have considerable main effects. Examining the effects on mean concentration at node (11, 6) in layer 2 after 45 days, we can obtain such insights as: (1) the main effects of mean K_{xx}, variance of $\ln K_{xx}$, mean α_L, and variance of $\ln \alpha_L$, are -34.1, -4.1, -3.1, and -2.1 µg/l, respectively, and this means that increasing of the values of

these four factors will decrease the mean concentration; (2) the interaction effects of AC and AD are –1.2 μg/l and –2.1 μg/l, respectively, and they cannot be neglected compared with the main effects of C and D, and this means that the effects of mean α_L and variance of $\ln\alpha_L$ on the modeling output may come from interaction with mean K_{xx} because mean K_{xx} has a large main effect; (3) the difference between the interaction effect of AB (9.7 μg/l) and main effect of B (-4.1 μg/l) is large, and this means that the effect of variance of $\ln K_{xx}$ on modeling output may mainly due to itself. It is apparent from analysis of table 3 that mean K_{xx} and variance of $\ln K_{xx}$ have significant impacts on the prediction, while mean α_L and variance of α_L have small impacts which can sometimes be neglected because their effects are mainly due to interaction with mean K_{xx}.

The influence of parameter uncertainty on the predicted benzene concentration for each factorial experiment can be examined from the analysis of Monte Carlo simulation results. Figure 3(a) shows the comparison between 6 realization results and the mean of the 100 realizations for node (5, 5) in layer 2 under experiment 1. The concentration decreases with time and shows large uncertainties, for example, at t=15 days, concentration has a value of 230 μg/l for realization 2, and of 150.1 μg/l for realization 4. Figure 3(b) presents the plotted probability density function for the output at node (5,5) after 30 days, and the maximum and minimum value of the predicted concentration for this experiment set are 241 and 16.7 μg/l, respectively, and the 97.5th percentile is 219.4 μg/l. The histogram of the 100 simulation runs show normal distributions, and output results have passed the Kolmogorov-Smirnov hypothetical test for normality with the calculated P value of 0.15.

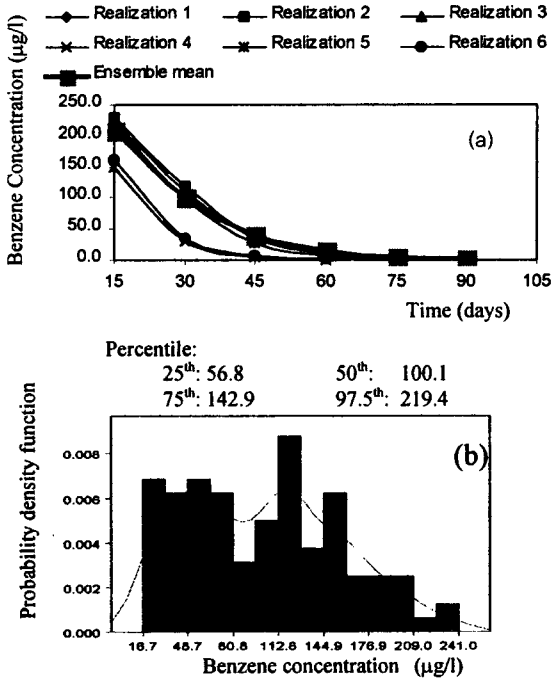

Figure 3 Simulated benzene concentration at node (5, 5) i layer 2 for experimental set 1, (a) compariso between realization results and ensemble mean, (b) probability density function after 30 days

5. Conclusions

A methodology based on factorial design and Monte Carlo simulation was proposed in this study to examine the effects of parameter uncertainties on prediction of NAPLs flow and transport fate in porous media. The longitudinal dispersivity α_L and intrinsic permeability K_{xx} were considered to be uncertain with known probability distributions. The determination of the mean and variance of these two parameters, were also assumed to be uncertain, and a 2^4 factorial design was conducted to evaluate the uncertainty impacts of these four factors on the modeling output. Under each factorial experiment, a series of Monte Carlo simulations were implemented, so as to assess the influence of uncertainties in the two input parameters on the simulation outputs uncertainties. It was found that the mean of K_{xx} and variance of $\ln K_{xx}$ have significant impacts on the prediction results, while the mean of α_L and variance of $\ln\alpha_L$ have minor influences on the modeling results, but significant interaction effects of these factors were observed, and it is suggested that great caution should be paid in determination of mean K_{xx} and variance of $\ln K_{xx}$ for simulation. The Monte Carlo simulation results revealed that the uncertainties in input parameters K_{xx} and α_L will introduce significant uncertainties in the simulation results. The proposed method will act as an effective tool for systematically evaluating uncertainties in the field of groundwater flow and transport simulation.

References

[1] Gelhar L. W., *Stochastic Subsurface Hydrology*, Prentice-Hall, Englewood Cliffs, NJ., 1993.

[2]Lahkim M. B., and Garcia L. A., "Stochastic modeling of exposure and risk in a contaminated heterogeneous aquifer, 1: Monte Carlo uncertainty analysis." *Environ. Eng. Sci.*, Vol. 16, no. 5, pp.315-328, 1999.

[3]Freeze R.A., Massmann J., Smith L., Sperling J., and James B., "Hydrogeological decision analysis, 1, A framework." *Ground Water*, Vol. 28, no. 5, pp.738-766, 1990.

[4] Box G.E.P., Hunter W.G., and Hunter J.S., *Statistics for Experimenters*, John Wiley & Sons, NY., 1978.

[5] Brown C. L., *Simulation of Surfactant Enhanced Remediation of Aquifers Contaminated with Dense Non-aqueous Phase Liquids*, PhD Dissertation, University of Texas at Austin, TX., 1993.

[6] Delshad M., Pope G. A., and Sepehrnoori K., "A compositional simulator for modeling surfactant enhanced aquifer remediation, 1. Formulation." *J. Contam. Hydrol.*, no. 23, pp.303-327, 1996.

[7] Bear J., *Hydraulics of Ground Water*, McGraw-Hill, New York, 1979.

[8] Lenhard R. J., and Parker J. C., "Measurement and prediction of saturation-pressure relationships in three-phase porous media systems." *J. Contam. Hydrol.*, no.1, pp.407-424, 1987a.

[9] Lenhard R. J., and Parker J. C., "A model for hysteretic constitutive relations governing multiphase flow, 2. Permeability-saturation relations." *Water Resour. Res.*, Vol. 23, no.12, pp.2197-2206, 1987b.