Stochastic Inverse Modeling in a Mass Transport Problem

Amir H Hosseini and Clayton V Deutsch

The central idea in this paper is to develop an inverse modeling approach for characterization of uncertainty in residual NAPL dissolution rate and first-order biodegradation rate by tailoring the estimation of these parameters to distributions of uncertainty in source size and hydraulic conductivity field. Such modeling can be used as a screening tool for estimating plume length, total mass, and time of remediation in field applications. The modeling technique is based on sequential self-calibration approach, distance-function approach, and a gradient-based optimization. It is observed that tying the estimation of transport parameters to joint realizations of transmissivity field and source geometry can effectively characterize the uncertainty in these parameters under field conditions and reduce the uncertainty in the state variables. It is also observed that ranking and screening the realizations based on their objective function value can effectively reduce the uncertainty in the source sizes.

Introduction

Estimating the length of time required for natural processes to remove a particular contaminant from a groundwater system is a mass balance problem termed time of remediation problem. There is significant uncertainty associated with source properties and with contribution and efficiency of concentration reducing mechanisms. The important source properties are the source geometry and the dissolution rate of contaminant species into groundwater. The concentration reducing mechanisms are advection, dispersion, dilution, sorption, volatilization, and aerobic and anaerobic biodegradation.

Many biodegradation models that simulate complex kinetics have been developed. It is evident that many of the required kinetic parameters for these models can not be measured or estimated by routine natural attenuation protocols. Simpler approaches with limited number of parameters are often preferred as they can be supported by the available data (Essaid et al. 2003, Rifai and Rittaler 2005). Application of first-order reaction models is common in natural attenuation studies, particularly at screening level. Based on the concentrations measured at monitoring locations, the field-scale first-order rates are estimated by trial and error calibration (Borden et al. 1986), by inverse modeling techniques (Medina and Carrera 1996) or by field approaches such as mass-flux (Borden et al. 1997) and concentration-distance relationships (Buscheck and Alcantar 1995). The parameter estimates by trial and error calibration are modeler dependent and a measure of uncertainty is not often available. Among the inverse modeling techniques, none of them quantifies the confidence in the estimated first-order rates under uncertainty of source properties and hydraulic conductivity distribution. In the case of field estimation techniques, the estimated first-order rates are affected by heterogeneity.

Understanding of the NAPL source dissolution rate is another important factor when investigating different aspects of a TOR problem. A number of experimental (Imhoff et al. 1994), pore- and field-scale numerical (Dillard et al. 2001, Christ et al. 2006) and inverse modeling studies (Sciortino et al. 2000) have been reported to estimate NAPL dissolution rate. None of the above inverse modeling techniques deals with characterization of source properties when the reaction rates are uncertain.

Simultaneous characterization of uncertainty in rate-limited dissolution and field-scale biodegradation is important. Essaid et al. (2003) implemented an inverse modeling in an optimal sense to estimate NAPL dissolution rate and individual first-order biodegradation rates for BTEX compounds as well as other parameters such as recharge rate, hydraulic conductivity, and transverse dispersivity. They only achieved convergence when they estimated a single dissolution rate for all BTEX compounds through simultaneous use of oxygen during aerobic biodegradation (cross-over effect). In other words, they failed to estimate individual dissolution rate and first-order biodegradation rate constants for each BTEX compound due to high correlation between these parameters that results in parameter non-uniqueness. The issue of parameter non-uniqueness is discussed in Hosseini and Deutsch (2009).

For groundwater management purposes in the field-scale, the worth of monitoring data can be used to estimate these parameters through inverse modeling. These estimates, however, will be affected by uncertainty in model structure (source size) and other flow and transport parameters, such as

distribution of hydraulic conductivity and dispersivities. In this paper, the non-linear confidence intervals of first-order biodegradation rate constant and dissolution rate constant are estimated under uncertain source geometry and aquifer transmissivity through a simple inverse modeling approach. Tailoring the estimation of dissolution rate and first-order biodegradation rate to distributions of uncertainty in the source geometry and transmissivity field through Monte Carlo type inverse modeling helps to (1) characterize the inherent uncertainty in the values of these parameters, (2) reduce the uncertainty in the state variables and size and shape of the plume, and (3) possibly reduce the uncertainty in the source sizes by ranking the conditional realizations based on the values of the objective function.

Methodology

A decoupled approach has been adopted in this work: first, sequential-self calibration approach is implemented to generate multiple realizations of transmissivity field conditioned to transmissivity and head data, these realizations are then combined with realizations of source size to create multiple realizations of source/transmissivity, dissolution rate and first-order biodegradation rate constants are then estimated for each joint realization to characterize the uncertainty in these parameters and reduce the uncertainty in the state variables. The forward steady-state flow and transport problems are represented by:

$$\frac{\partial}{\partial x_i}\left(k_i \frac{\partial h}{\partial x_i}\right) = q_{sr} \qquad [1]$$

and,

$$\frac{\partial (C)}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C}{\partial x_j}\right) - \frac{\partial}{\partial x_i}(v_i C) + \max\left[0, k_{dis}\left(C^{eq} - C\right)\right] - \lambda C_s \qquad [2]$$

where, k_i, h, q_{sr}, C, D_{ij}, v_i, k_{dis}, C^{eq}, and λ represent hydraulic conductivity, hydraulic head, dispersion coefficient, seepage velocity, NAPL dissolution rate constant, equilibrium concentration, and first-order biodegradation rate constant. The equilibrium concentration is expressed by:

$$C^{\mathrm{eq}} = \mathrm{f}_s . C_s^{\mathrm{sol}} \qquad [3]$$

where, C_s^{sol} is the solubility limit for pure substrate in water, and f_s is mole fraction of the species s in the mixture of organic and intert/non-biodegradable materials and can be calculated by (Parker et al. 1991):

$$\mathrm{f}_s = \frac{S_s^{NAPL}/\omega_s}{S_s^{NAPL}/\omega_s + T_t^{NAPL}/\omega_t} \qquad [4]$$

where, S_s^{NAPL} is the mass of substrate s per unit mass of dry soil, and T_t^{NAPL} represents the equivalent mass of all inert and non-biodegradable materials t per unit mass of dry soil, ω_s is the molecular weight of substrate s, and ω_t is the equivalent molecular weight of mixture of all non-biodegradable and inert (insoluble) materials. The fraction T_t^{NAPL}/ω_t can be calculated by:

$$\frac{T_t^{NAPL}}{\omega_t} = \sum_{li=1}^{NI}\frac{I_{li}^{NAPL}}{\omega_{li}} + \sum_{lt=1}^{NT}\frac{TR_{lt}^{NAPL}}{\omega_{lt}} \qquad [5]$$

where, I_{li}^{NAPL}, TR_{lt}^{NAPL}, ω_{li} and ω_{lt} represent each inert and tracer (non-biodegradable) material and their associated molecular weights, respectively. The NAPL mass of substrate s per unit mass of dry soil (S_s^{NAPL}) decreases as the dissolution occurs. This process can be represented by (Waddill and Widowson 1998):

$$\frac{\mathrm{d}S_s^{NAPL}}{\mathrm{d}t} = -\frac{\theta}{\rho_b} R_s^{NAPL} \qquad [6]$$

where, ρ_b is the bulk density of the porous medium and R_s^{NAPL} represents the mass transfer rate given by:

$$R_s^{NAPL} = \max\left[0, k_{dis}\left(C_s^{\mathrm{eq}} - C_s\right)\right] \qquad [7]$$

Thus, due to dissolution of NAPL into groundwater, soil NAPL concentration decreases and aqueous concentration increases. In Equation [2], the only term on the left hand side accounts for change in concentration with time (a transient problem); the first term on the right hand side represents the hydromechanical dispersion; the second term represents advection; the third term represents dissolution;

and the forth term represents first-order biodegradation. This equation should be solved numerically to find the distribution of dissolved species concentration in space and time. A simple steady-state groundwater flow simulator, **flsim2d** has been developed for groundwater flow; and a Lagrangian-Eulerian approach, the method of characteristics, has been programmed into the code **snasim** to solve Equations [1] and [2].

The uncertainty in the hydraulic conductivity field is characterized by sequential-self calibration approach (SSC). The details of this approach can be found in (Gomez-Hernandez et al. 1997). The uncertainty in the source geometry based on the existing well arrangement can be modeled by the distance function approach. The uncertainty in areal limits is addressed by:

$$DF^{*}_{ID}(\mathbf{u}_0) = \sum_{i=1}^{N} \alpha^{\frac{DF(\mathbf{u}_i)}{|DF(\mathbf{u}_i)|}} \cdot \lambda_{ID}(\mathbf{u}_i) \cdot \left[DF(\mathbf{u}_i) + \beta \cdot \frac{DF(\mathbf{u}_i)}{|DF(\mathbf{u}_i)|} \right] \quad [8]$$

where, α and β are scaling and separation factors that define a unique uncertainty band and must be calibrated against a large number of synthetic realizations. The objective function for mass transport inverse problem is defined by:

$$F'_{C,j} = \sum_{i=1}^{n_c} q_j w_i \left(C_i^m - C_i\right)^2 \quad \text{with} \quad w_i = \frac{1}{cv_i^2 C_i^2} \quad [9]$$

where, C_i , C_i^m , q_i and cv_i are simulated concentration, measured concentration, source size quantile and coefficient of variation associated with each observation. In order to solve the inverse problem and estimate the values of k_{dis} and λ, one needs to calculate sensitivity coefficients that can be calculated by sensitivity equations:

$$\frac{\partial(S_\alpha)}{\partial t} + \frac{\partial}{\partial x_i}(v_i S_\alpha) - \frac{\partial}{\partial x_i}\left(D_{ij} \frac{\partial S_\alpha}{\partial x_j} \right) = \frac{\partial R_{s/n}}{\partial \alpha} \quad [10]$$

where, α represents either k_{dis} or λ. The minimization of objective function (Equation [9]) and optimization of the transport parameters can be implemented through modified Gauss-Newton approach (Cooley and Naff 1990):

$$\left(\mathbf{C}^T \mathbf{X}_r^T \boldsymbol{\omega} \mathbf{X}_r \mathbf{C} + \mathbf{I} m_r\right) \mathbf{C}^{-1} \mathbf{d}_\mathbf{r} = \mathbf{C}^T \mathbf{X}_r^T \boldsymbol{\omega} \left(\mathbf{y} - \mathbf{y}(\mathbf{b}_\mathbf{r})\right) \quad [11]$$

where, **C** is the diagonal scaling matrix, $\mathbf{X}_r$ is the matrix of sensitivities, $\boldsymbol{\omega}$ is the matrix of weights, **y** is the vector of observed concentrations, $\mathbf{y}(\mathbf{b}_\mathbf{r})$ is the vector of simulated concentrations, and m_r is the Marquardt parameter. In each iteration of the Modified Gauss-Newton approach, the vector of estimated parameters is updated by addition of an updating vector $\mathbf{d}_\mathbf{r}$ multiplied by a damping parameter ρ_r:

$$\mathbf{b}_{\mathbf{r+1}} = \rho_r \mathbf{d}_\mathbf{r} + \mathbf{b}_\mathbf{r} \quad [12]$$

where, $\mathbf{b}_\mathbf{r}$ and $\mathbf{b}_{\mathbf{r+1}}$ are the vectors of the estimated parameters in two consecutive iterations. The damping parameter ρ_r preserves the direction of $\mathbf{d}_\mathbf{r}$ and ensures that the changes in the parameters remain less than the maximum allowed change specified by the user and has a damping effect on likely oscillations that may occur due to opposite directions in consecutive updating vectors ($\mathbf{d}_\mathbf{r}$ and $\mathbf{d}_{\mathbf{r-1}}$). In the modified Gauss-Newton method, the updating vector $\mathbf{d}_\mathbf{r}$ is calculated by Equation [11].

Synthetic Example – Error free observations

A synthetic example is presented to investigate the performance of the Monte Carlo type decoupled inverse modeling in characterization of uncertainty in the dissolution rate and first-order biodegradation rate and to study the effects of error in observed data in the modeling outcomes. A synthetic hydraulic conductivity dataset, two different head observation datasets with two different levels of measurement error, and four concentration datasets are sampled from the reference study sites. Applying the SSC approach, the sampled hydraulic conductivity and head data are used to create multiple realizations of hydraulic conductivity field conditioned to both hydraulic conductivity and head measurements. The distance function approach is used to create multiple realizations of areal extent of the source zone. Inverse modeling is then implemented to estimate the values of dissolution rate and first-order biodegradation rate constants for the sets of joint realizations of source geometry and hydraulic conductivity fields. The performance of the methodology is investigated through studying the distributions of the estimated parameters and source zone sizes and comparing the variations of the state

variables (e.g. plume length and mass loaded into the aquifer) through time with those of the reference study sites. The simulated state variables are also compared to the results of a set of Monte Carlo simulations performed using k_{dis} and λ distributions that represent the range of variability that may be observed under realistic field conditions. The effects of head and concentration measurement errors on the estimation of k_{dis} and λ and the predictions of the state variables are also investigated.

Figure 1 shows the reference study site with the sampling locations, the suspected source zone area, the reference hydraulic conductivity field and the associated head response. The reference hydraulic conductivity field shown in Figure 1-b has a Gaussian distribution in natural logarithmic units with a mean of -10.1 $\log_e$m/s, standard deviation of 1.2$\log_e$m/s, and a spatial correlation defined by a spherical variogram with a nugget effect equal to 0.1 and a range of 32.0 m. The modeling domain is 250m by 160m, which is descritized by 2.0m × 2.0 m squared shape grid cells. The flow boundary conditions involve fixed head boundary conditions at the north and the south of the site equal to 4.0m and 2.0m, respectively. At the east and west of the site, no-flow boundary conditions are assigned. As shown in Figure 1-a, there are a total of 40 observation wells where piezometric heads (steady-state) and concentrations are sampled. There are 11 of these wells (shown by blue circles), with hydraulic conductivity measurements. The solid black wells indicate the observation wells where residual NAPL has been observed. Figure 2-a shows the calibrated uncertainty band for the given well arrangement and suspected source area. The associated optimal values of scaling factor α and separation factor □ are 3.56 and 15.86, respectively. To investigate the performance of the methodology when the actual source size deviates from the average source size that is characterized by the distance function approach, three source sizes corresponding to lower quartile, median and upper quartile of the calibrated uncertainty band are considered as reference cases for source geometry. Figure 2-b shows the CDF of the source sizes associated with the calibrated uncertainty band in Figure 2-a and the selected quartiles. According to Figure 2-b, the reference source sizes (p_{25}, p_{50} and p_{75} quartiles) have areas equal to 643 m^2, 938 m^2, and 1395 m^2. Figure 3 shows the simulated plumes for the smaller source size. For simplicity, it has been assumed that the distribution of residual NAPL (soil concentration) within the areal limits of the source zone is uniform. Variability within areal limits can easily be incorporated. The uniform soil concentration of NAPL is set to 10gr/Kg. The initial mass fraction of the substrate (e.g. Benzene) in NAPL is equal to 0.01. The substrate solubility, substrate and inert molecular weights are set equal to 0.00178gr/cm3, 78.1 and 101gr/mole, respectively. Dry soil density, total porosity and effective porosity are set equal to 1.6gr/cm3, 0.35 and 0.3, respectively. The longitudinal and transverse dispersivities, dissolution rate and first-order biodegradation rates are equal to 1.5m, 0.3m, 0.0015day^{-1} and 0.006day^{-1}, respectively. Zero dispersive flux boundary conditions are assigned at all boundaries. Two synthetic observed datasets for piezometric heads are created by sampling from the reference piezometric head distribution and subsequent addition of Gaussian noise. The first set of head observations is considered to be error-free. The second head dataset is considered to be noisy by addition of Gaussian noise with a standard deviations of σ_{nH} =0.20m.

Applying the SSC technique, two sets of 300 realizations of hydraulic conductivity field conditioned to both hydraulic conductivity and head measurements are constructed form two levels of head measurement errors (σ_{nH} =0.0m and σ_{nH} =0.20m), and combined with realizations of source geometry to create two sets of 300 joint realizations that are used in subsequent estimation of k_{dis} and λ. In terms of the synthetic concentration datasets, an error-free concentration dataset is sampled from the simulated plumes (a total of 520 samples at 40 observation wells over a period of two years – from 5 to 7 years from the start of the simulations). To investigate the effects of errors in measured concentrations on the modeling outcomes, Gaussian noise is added to the synthetic concentration dataset sampled from the first reference case with smaller source size. The added Gaussian noise has a coefficient of variation equal to cv_{nc} =0.35. To study the importance of tailoring the estimation of first-order biodegradation rate constant and dissolution rate to realizations of source geometry and hydraulic conductivity, the results of the decoupled inverse modeling including the simulated state variables should be compared to the available field-scale parameter estimation techniques. Due to the fact that the proposed methodology is aimed to be an advanced screening tool for characterization of uncertainty in the field-scale parameters, its outcomes should be compared to the outcomes of similar screening tools commonly applied to the field. For this purpose, a set of Monte Carlo simulations (MCS) are performed with (1) realizations of

hydraulic conductivity conditioned to conductivity and head data by SSC, (2) realizations of source extend characterized by the DF algorithm in Figure 6-2-a, (3) the values of first-order biodegradation rate drawn from a distribution reported by Bauer et al. (2006), and (4) the values drawn from a distribution of dissolution rate constant representing the uncertainty in a realistic field condition. Bauer et al. (2006) showed that the field-scale method of normalization to a recalcitrant co-contaminant (Wiedemeier et al. 1996) that corrects for the effects of uncertainty in the value of longitudinal dispersivity gives the closest estimate to the true value of the first-order biodegradation rate constant. For an aquifer with a log-normal hydraulic conductivity distribution with a mean of -9.54 log e m/s and a standard deviation of 1.3 logem/s, Bauer et al. (2006) showed that the method of normalization to a recalcitrant co-contaminant overestimates the true first-order rate (on average) by a factor of two, while the standard deviation of the normalized rates is equal to 2. Similar results were found by Bauer et al. (2007) for the improved method of Stenback et al. (2003) with off-centerline measurements. Figure 4-b shows a distribution of first-order biodegradation rate similar to the distribution observed by Bauer et al. (2006) based on normalization to a recalcitrant co-contaminant. To investigate the effect of variability in the dissolution rate, a uniform distribution with an order of magnitude variability (which seems to be a lower bound to variability in this parameter based on the observations in Dillard et al. 2001 and Christ et al. 2006) and a mean equal to 0.0066 day^{-1} (computed by Essaid et al. (2003) for Bemidji site) is considered (Figure 4-a). Figure 5 shows the variations of the mass loaded into the aquifer and the plume length in time for the reference case as well as the mean and quartiles of the state variables based on the results of the Monte Carlo simulations with 100 joint realizations of hydraulic conductivity field and source geometry and the values of k_{dis} and drawn from the distributions in Figures 4-a and 4-b. Figure 6 shows the probability map for the concentrations to exceed a threshold value of 0.005 mg/L (water quality standard for benzene). Figures 5 and 6 show that the MCS may result in large uncertainties in the dimensions of the simulated plume as well as the mass loaded into the aquifer. The distribution of the parameters shown in Figure 4 and the MCS results shown in Figures 5 and 6 will be compared to the results of inverse modeling.

Figure 7 shows the histograms of the estimated k_{dis} and λ for the 100 joint realizations that are calibrated to concentration measurements from the smaller reference case. It has been assumed that the head and concentration measurements are error-free. Figure 8 shows the variations of the mass loaded into the aquifer and the plume length through time for three sets of 100 realizations corresponding to the smaller reference case. Figure 9 shows the probability maps for concentrations to exceed a threshold value of 0.005 mg/L for the smaller reference case. As expected, the dissolution rate constant is slightly under-estimated for the case with the smaller reference source zone. The proposed approach significantly reduces the uncertainty in the state variables (comparing to the results of the Monte Carlo simulations). Although the ensemble of realizations on average over/under-estimates the reference values, for both state variables, the reference curve falls within the 90% non-linear confidence interval. Comparing Figure 9 to Figure 6, it is also evident that the estimation of dissolution rate and first-order biodegradation rate for joint realizations of hydraulic conductivity and source geometry using concentration data can significantly reduce the uncertainty in the dimensions of the plume. The observed over-estimation and under-estimation of the state variables is partially due to unresolved uncertainties in the source size which can not be fully handled by adjusting the values of k_{dis} and λ by the model. Thus, a ranking-based screening can be applied (Similar to the work of Poeter and McKenna 1995) to choose from the set of realizations based on the values of the modified objective function and to decrease the uncertainty in the source zone sizes previously characterized by the distance-function approach. To investigate the effectiveness of ranking on the reduction of uncertainty and to have enough realizations to explore the space of uncertainty, 300 joint realizations of hydraulic conductivity (conditioned to head data with σ_{nH} =0.0 m) and source geometry are constructed and the concentrations sampled from the three reference cases (with cv_{nc} =0.0) are used to estimate the values of dissolution rate constant and first-order biodegradation rate. Figure 11 shows the CDF of the source sizes for 100 realizations (out of 300 realizations) having smallest values of modified objective function defined by Equation [9].

Comparing Figure 11 to Figure 2-b, one can observe that ranking and screening the realizations can effectively reduce the uncertainty in the source zone sizes for each reference case. To further investigate the effect of ranking, one may also look at Figure 10 where the cross-plot of dissolution rates

and source size quantiles is shown. The color-scale represents the rank of each realization based on the value of the modified objective function (black shows lower values of the modified objective function, higher ranks and therefore accepted realizations). For this example, Figure 10 shows that (1) there is a negative correlation between the size of the source the estimated dissolution rate; and (2) ranking of the realizations can effectively identify the joint realizations that have not properly converged in optimization (solid circles) and the joint realizations that have source sizes that deviate from the reference source size (dashed circles). Figures 11 and 12 show the cross-plots between the dissolution rate and first-order biodegradation rate constant and the cross-plots between the source size quantile and first-order biodegradation rate (after ranking). According to Figure 13, there exists a positive correlation between first-order biodegradation rate and dissolution rate. Looking at Figure 14, one observes that there is little correlation between the values of source quantile and first-order biodegradation rate. The observations in Figures 10, 12 and 13 justify the importance of simultaneous characterization of uncertainty in source areal extent, source dissolution rate and first-order biodegradation rate. Figures 14 to 16 show the estimated parameters, the associated state variables and the probability maps after ranking and screening based on the values of the modified objective function. According to Figures 10 to 16, one can conclude that (1) reduction in the uncertainty of the source zone sizes appears to be achievable by ranking and screening the realizations based on the values of the modified objective function (Equation [9]); (2) for this purpose, an appropriate number of realizations should be selected; and (3) by reducing the uncertainty in the source zone sizes, there will be reductions in the associated uncertainty in the estimated parameters values and the state variables.

Synthetic Example – Erroneous observations

In practice, it is quite rare to consider the observation data as error-free. Due to the fact that the proposed methodology is a decoupled approach, one may generate hydraulic conductivity realizations honoring a particular level of error in head observations (measure of fit *s* close to one); and then estimate the rate constants for the joint realizations, while calibrating to concentrations with a particular level of error in the data values. In the subsequent analysis, Gaussian noise (with a relatively large standard deviation/coefficient of variation) has been added to the head observations and concentration measurements associated with the smaller reference case. It is assumed that (1) a good knowledge of magnitude of error exists in the observations, (2) the error in observations is Gaussian noise with a mean equal to zero (no systematic bias is introduced), and (3) errors at different locations and for heads and concentrations are independent of each other. As it can be observed in the following example, uncertainty in the estimated parameters and the predicted state variables increases with an increase in the measurement errors. Comparing Figures 17, 19 and 18 to Figures7, 8 and 9, one can observe that introducing measurement errors to heads and concentrations results in (1) considerable increase in the standard deviation of the estimated parameters, (2) larger deviation of the average biodegradation rate constant from the reference value, (3) introducing more uncertainty and bias in the estimation of the length (and width) of the plume and (4) increasing the uncertainty in the estimation of mass loaded into the aquifer.

Conclusions

This paper presented a modeling approach and a synthetic example to investigate the performance of the decoupled inverse modeling approach in characterizing the uncertainty in the dissolution rate and first-order biodegradation rate and reducing the uncertainty in the associated state variables. A reference case with smaller source size was considered. First, a set of Monte Carlo simulations were implemented whose results were subsequently compared to the results of the inverse modeling methodology. Comparing the results of the Monte Carlo simulations to the results of inverse modeling for the smaller reference case, it was observed that tailoring the estimation of first-order biodegradation rate and dissolution rate to distributions of uncertainty in the source geometry and hydraulic conductivity field results in characterization of uncertainty in these parameters and significant reduction of uncertainty in the state variables, being mass loaded into the aquifer and the dimensions of the plume. Although the reference values always fell within 90% confidence interval, a bias was observed in the prediction of the reference state variables by the ensemble of simulated realizations. This bias was deemed to be partially due to

large variabilities in the size of the source zone which cannot be fully resolved by adjusting the values of dissolution rate and first-order biodegradation rate. In this work, it was observed that ranking and screening the conditional realizations based on the value of modified objective function (Chapter 5) can effectively reduce the uncertainties in the size of the source zone and uncertainty and bias in the prediction of the state variables. The value of the modified objective function is deemed to be independent of overall level of concentrations in the modeling domain due to the fact that it is a dimension-less number and a normalization that takes place by defining the weights inverse proportional to the value of simulated concentrations. On the other hand, the number of realizations that are kept after ranking and screening may be considered 'problem dependent' for a large part. Despite this apparent subjectivity, it was observed through a sensitivity analysis that ranking the realizations and keeping any number of realizations can still be useful as it can give a general idea about the size of the source, while reducing its uncertainty. The importance of simultaneous characterization of uncertainty in the parameters was investigated through cross-plots of the parameters, where it was observed that a positive correlation exists between the values of dissolution rate and first-order biodegradation rate, a negative correlation exists between the values of dissolution rate and source size quantile and little correlation exists between first-order biodegradation rate and source size quantile. To investigate the effects of errors in observation data on the modeling outcomes, relatively large levels of observation errors (Gaussian noise with a mean of zero and pre-specified standard deviation/coefficient of variation) were added to head and concentration observations and the transport parameters were estimated. According to the results, existence of Gaussian noise in the data resulted in an increase in the uncertainty and bias of the estimated parameters and the predicted state variables. Comparing these results to the results of Monte Carlo simulations indicated that even if the observed data are subject to a relatively large level of Gaussian noise, the uncertainty in the predicted state variables are still smaller than the results of Monte Carlo simulations.

References

Bauer, S., C. Beyer, and O. Kolditz (2006), Assessing measurement uncertainty of first order degradation rates in heterogeneous aquifers, *Water Resources Research*, 42, W01420, doi: 10.1029/2004WR003878

Borden, R.C., P.B., Bedient, M.D., Lee, C.H., Ward, J.T. Wilson (1986), Transport of dissolved hydrocarbons influenced by oxygen-limited biodegradation: 2. Field application, *Water Resources Research*, 22(13), 1983-1990

Buscheck, T.E., and C. M. Alcantar (1995), Regression techniques and analytical solutions to demonstrate intrinsic bioremediation, in Intrinsic Bioremediation, edited by R.E. Hinchee, T.J. Wilson, and D. Downey pp. 109-116, Battelle Press, Columbus, Ohio

Christ, J.A., C.A. Ramsburg, K.D. Pennell, and L.M. Abriola (2006) Estimating mass transfer from dense nonaqueous phase liquid source zones using upscaled mass transfer coefficients: An evaluation using multiphase numerical simulations, *Water Resources Research*, 42(11), W11420

Cooley R.L., and R.L. Naff (1990), *Regression modeling of ground-water flow*, Techniques of Water-Resources Investigations of the US Geological Survey, 232 pp.

Dillard, L.A. H.I. Essaid, M.J. Blunt (2001), A functional relation for field-scale nonaqueous phase liquid dissolution developed using a pore network model, *Journal of Contaminant Hydrology*, 48(1-2), 89-119

Essaid, H.I., I.M. Cozzarelli, R.P. Eganhouse, W.N., Herkelrath, BA. Bekins, G.N. Delin (2003), Inverse modeling of BTEX dissolution and biodegradation at the Bemidji, MN crude-oil spill site, *Journal of Contaminant Hydrology*, 67(1-4), 269-299

Gomez-Hernandez, J.J., Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data – 1.Theory (1997), *Journal of Hydrology*, 203(1-4): 162-174

Hosseini, A.H., C.V. Deutsch (2009), Stability of the inverse problems: a case study, *Center for Computational Geostatistics, Report 11*, paper 129

Imhoff, P.Y., P.J. Jaffe, and G.F. Pinder (1994), An experimental study of complete dissolution of a non-aqueous phase liquid in saturated porous media, *Water Resources Res.*, 30(2), 307-320

Medina, A., and J. Carrera (1996), Coupled estimation of flow and solute transport parameters, *Water Resources Research*, 32(10): 3063-3076

Poeter, E.P., S. McKenna (1995), Reducing uncertainty associated with ground water flow and transport predictions, *Ground Water*, 33(6), 899-904

Rifai, H. S., and T. Rittaler (2005), Modeling natural attenuation of benzene with analytical and numerical models, *Biodegradation*, 16, 291-304

Sciortino, A., T.C. Harmon and W.G. Yeh (2000), Inverse modeling for locating dense nonaqueous pools in groundwater under steady flow conditions, *Water Resources Research*, 36(7): 1723-1735

Stenback, G.A., S.K. Ong, S.W. Rogers, and B.H. Kjartonson (2004), Impact of transverse and longitudinal dispersion on first-order degradation rate constant estimation, *Journal of Contaminant Hydrology*, 73, 3-14, doi: 10.1016/j.jconhyd.2003.11.004

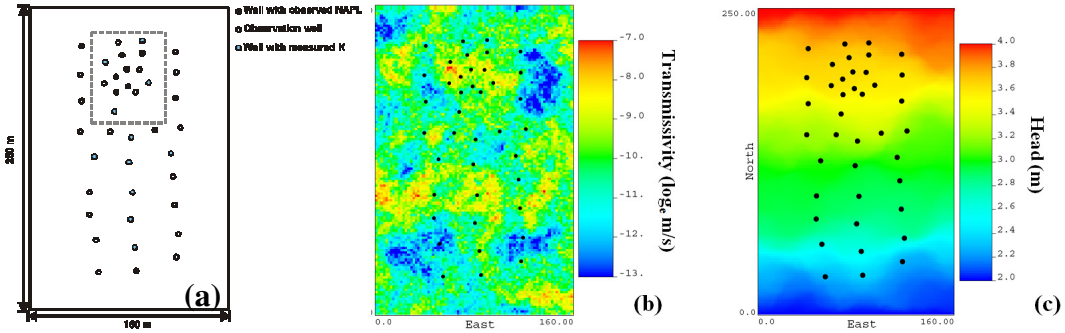

Figure 1: (a) The reference study site with monitoring locations and suspected source zone area (dashed box), (b) the reference hydraulic conductivity field, and (d) the reference hydraulic head distribution.

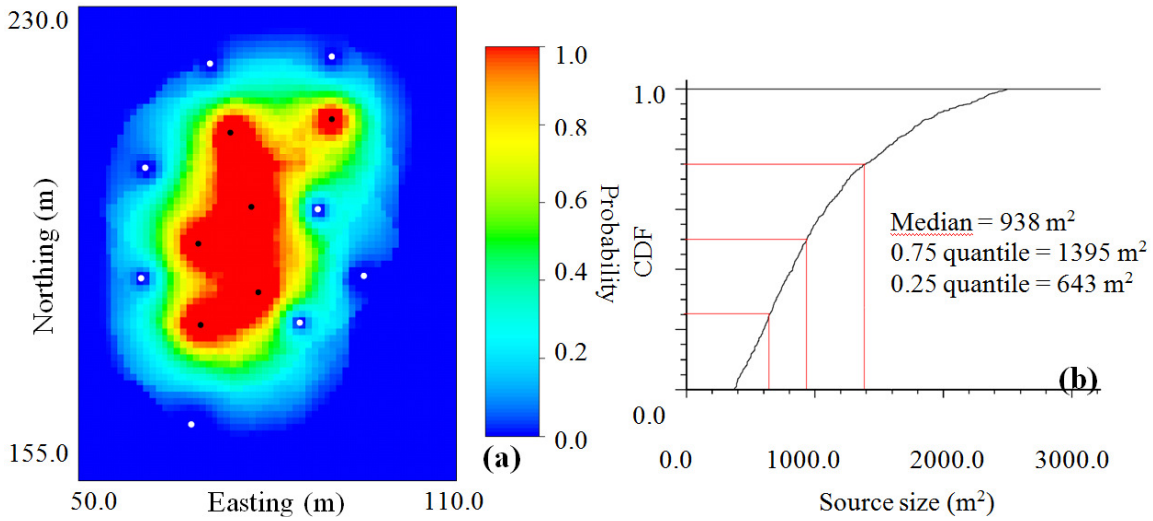

Figure 2: (a) The calibrated band of uncertainty for the contaminant source zone, and (b) the CDF of the source sizes.

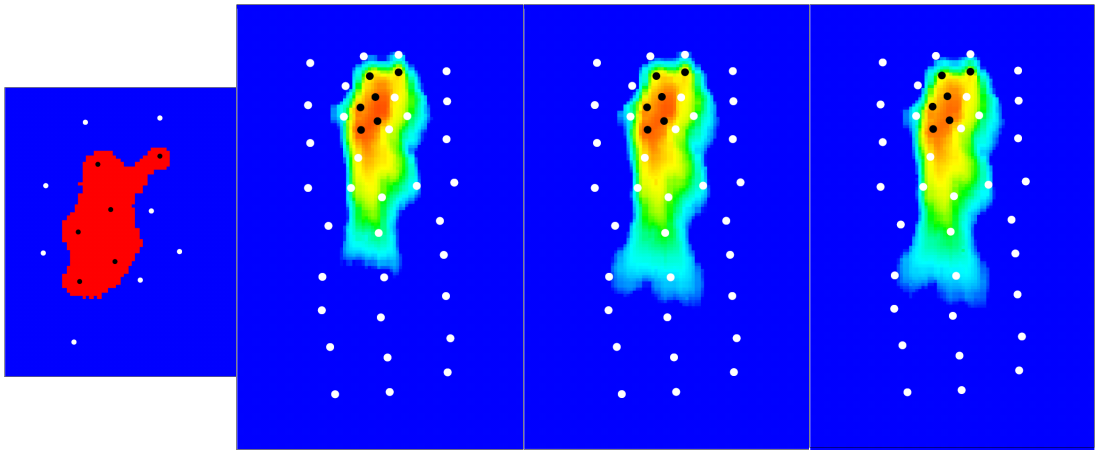

Figure 3: (a) The smaller source zone size corresponding to p_{25} of the calibrated uncertainty band, (b) the simulated plume after 550 days, (c) the simulated plume after 1281 days, and (d) simulated plume after 2562 days.

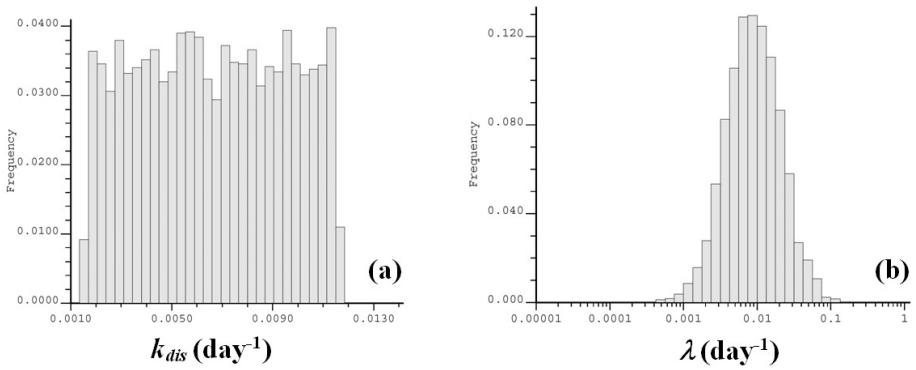

Figure 4: The distribution of uncertainty in (a) the dissolution rate constant and (b) the first-order biodegradation rate constant, used in the subsequent MCS.

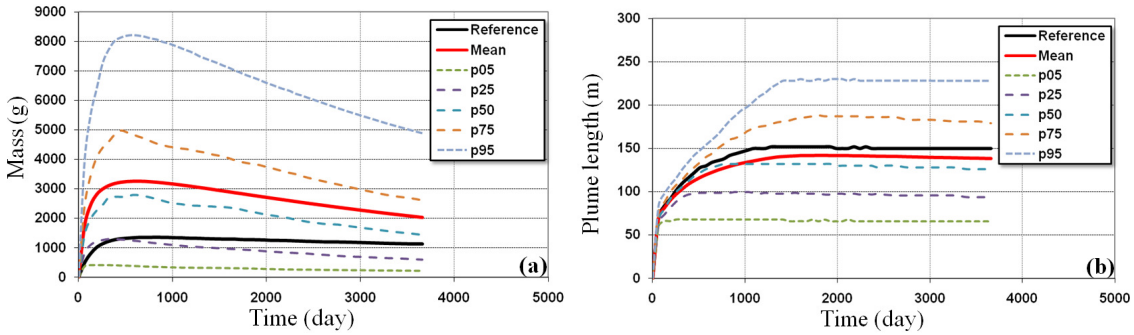

Figure 5: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (p_{50} source size) (a) total mass loaded into the aquifer and (b) plume length based on the results of the MCS. The reference curve is associated with the median source size (Figure 5-24)

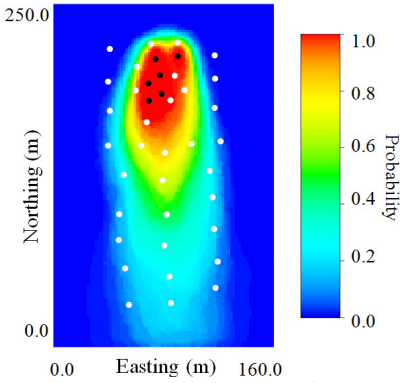

Figure 6: The probability of concentrations exceeding 0.005 mg/L based on the results of the Monte Carlo simulations

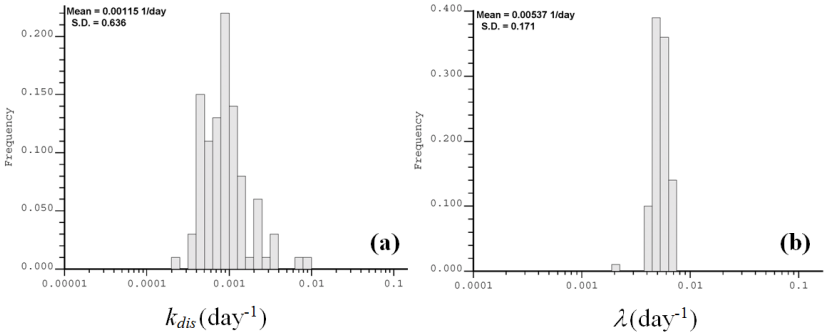

Figure 7: The histograms of (a) k_{dis} and (b) λ for the case with σ_{nH} = 0.0 m and cv_n = 0.0 and the smaller reference source size.

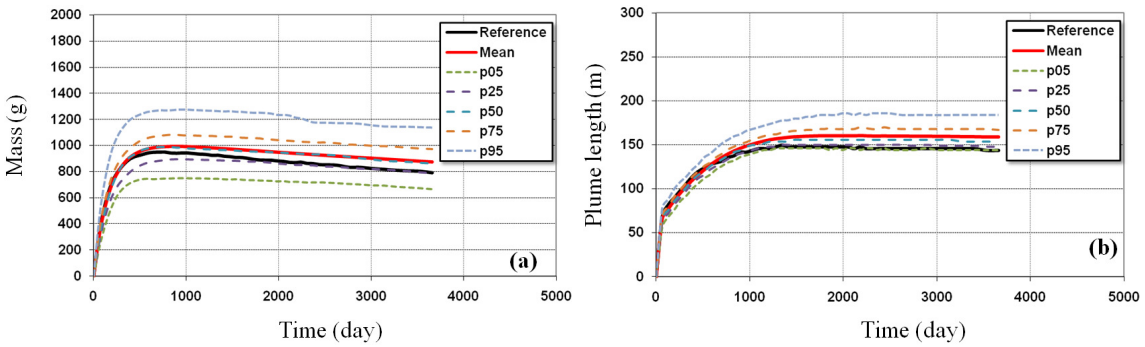

Figure 8: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the smaller reference source size.

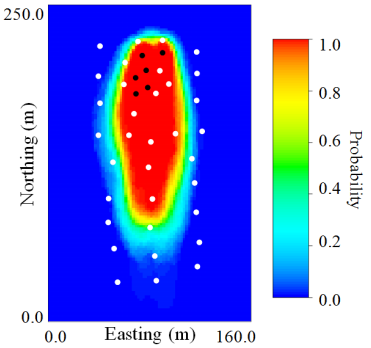

Figure 9: The probability of concentrations exceeding 0.005 mg/L after conditioning to concentrations for the smaller reference case

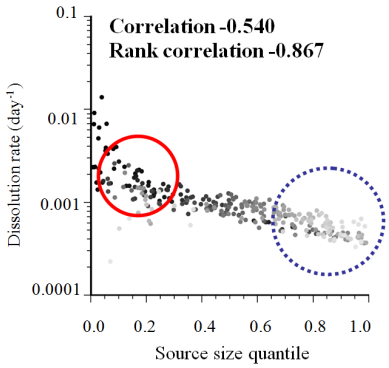

Figure 10: The cross-plot between the source size quantiles and the estimated dissolution rate. The solid circles show the realizations that are likely not converged and the dashed circles show the realizations that their source sizes significantly deviate from the reference source size. The color scale shows the rank of realizations based on their modified objective function value (black represents lower values of the objective function)

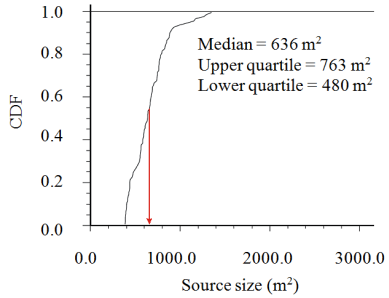

Figure 11: The CDF of the source sizes of the 100 accepted realizations after ranking based on the modified objective function value. The red arrows show the reference source size for each case.

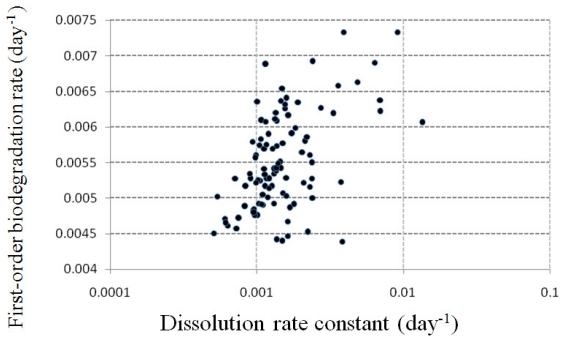

Figure 12: The cross-plot between the dissolution rate constant and firs-order biodegradation rate constant for the reference case with smaller source size (p_{25}) with a correlation coefficient equal to 0.438.

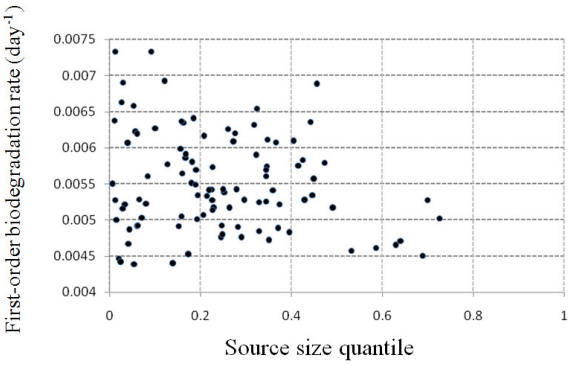

Figure 13: The cross-plots between the source size quantile and first-order biodegradation rate constant for the reference case with smaller source size (p_{25})

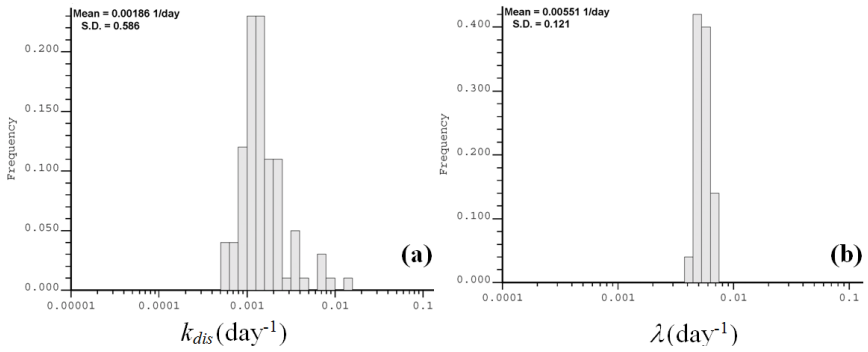

Figure 14: The histograms of (a) k_{dis} and (b) ? for the accepted realizations after ranking, based on the reference case with the smaller source size.

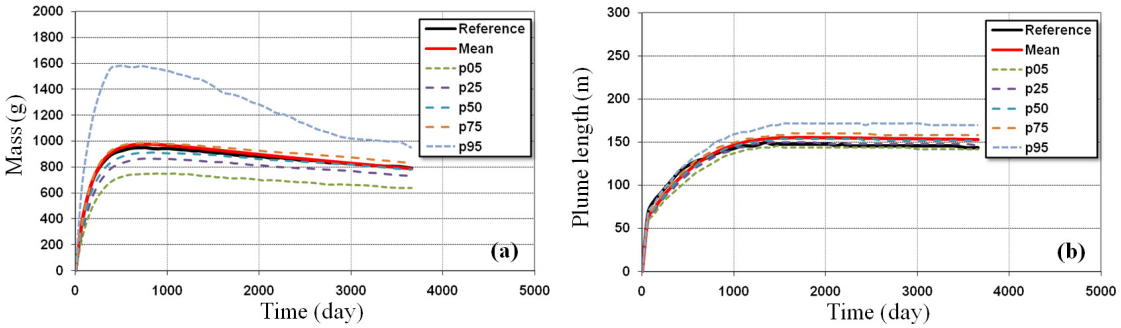

Figure 15: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the smaller source size after ranking.

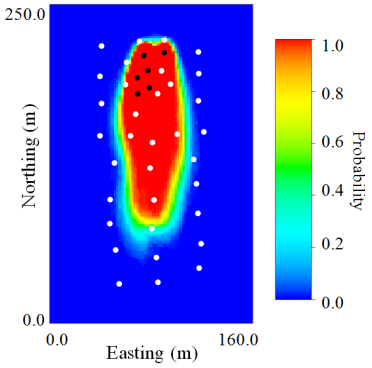

Figure 16: The probability of concentrations exceeding 0.005 mg/L after conditioning to concentrations and ranking for the smaller reference case

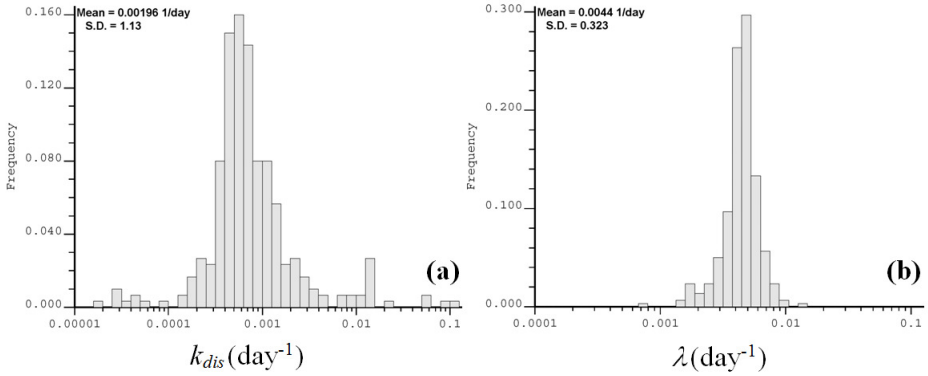

Figure 17: The histograms of (a) k_{dis} and (b) ⍰ for the case with ⍰$_{nH}$ = 0.2 m and cv_n = 0.3 and the smaller reference source size.

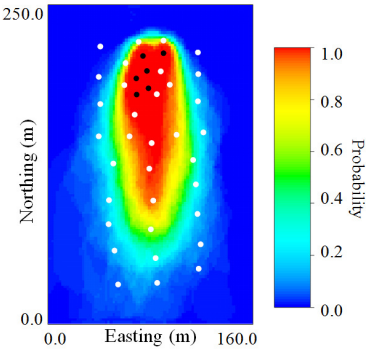

Figure 18: The probability of concentrations exceeding 0.005 mg/L for the case with ⍰$_{nH}$ = 0.2 m and cv_n = 0.3 and the smaller reference source size.

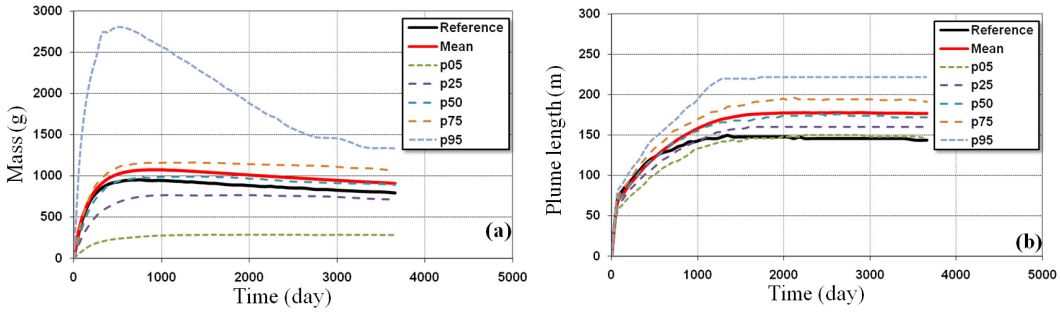

Figure 19: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the case with ⍰$_{nH}$ = 0.2 m and cv_n = 0.3 and the smaller reference source size.