ISEIS

Journal of Environmental Informatics 24(2) 80-89 (2014)

Journal of Environmental Informatics

www.iseis.org/jei

State and Parameter Estimation in Three-Dimensional Subsurface Contaminant Transport Modeling Using Kalman Filter Coupled with Monte Carlo Sampling

G. A. Assumaning* and S. Y. Chang

Department of Civil and Environmental Engineering, North Carolina A&T State University, Greensboro, NC 27411, USA

Received 22 June 2012; revised 11 March 2013; accepted 8 September 2013; published online 3 November 2014

ABSTRACT. Accurate contaminant prediction is very important for risk assessment and site remediation. Several predictive tools in the form of mathematical models have been used to model the movement and behavior of contaminants in groundwater. These deterministic models do not account for the heterogeneity and uncertainties in the subsurface environment. Discretization of the models spatially and temporally introduces approximation and truncation errors. In this research, to improve the accuracy of subsurface contaminant prediction in time and space and to assess the impact of first-order decay rate parameter estimation, Kalman filter coupled with Monte Carlo sampling was used in a specified three-dimensional domain space. The filter is perturbed with random Gaussian noise to reflect real life case of contaminant movement. Set of sparse observation points selected at specific locations are used to guide the filter at every time step to improve the accuracy of the prediction. The first-order decay rate parameter is estimated using Monte Carlo sampling method. The algorithms to generate the simulation results were run on Matlab 7.1. The efficacy of the Kalman filter coupled with Monte Carlo sampling, Kalman filter without Parameter Estimation and the numerical method were tested using Mean Absolute Error (MAE) and Maximum Absolute Error (Emax) equations. The results show that the Kalman filter coupled with Monte Carlo sampling performs better than both the Kalman filter without Parameter Estimation and the numerical method. Also, the Kalman filter coupled with Monte Carlo sampling is capable of reducing the error in the numerical solution by approximately 75% when Mean Absolute Error equation is used to estimate the prediction error. The error reduction is due to the adaptive nature of the Kalman filter to the observation data used in the simulation.

Keywords: Kalman filter, parameter estimation, data assimilation, Monte Carlo sampling, groundwater, modeling

1. Introduction

It is important to understand the mechanism, transport and fate of the contaminants in the subsurface medium to improve water quality management and risk evaluation. Accurate prediction of contaminant concentration and model parameters has been a major concern for hydrogeologist and environmental engineers. However, conventional methods such as numerical methods have been used to make predictions. These numerical model results are known to be less accurate due to the errors incorporated in them in the discretization process (Chang and Latif, 2010). The hydrologic parameters are assumed to be constant in the simulation process. Stability and convergence criteria are also setbacks in the numerical method.

The Kalman filter (KF) is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean of the squared error (Welch et al., 2004). It is computationally more efficient than computing the estimate directly from an entire past observed data at each time step (Haykin, 2001). It is widely used in areas such as signal processing, navigation and military technology development.

Several studies have been done in the area of subsurface contaminant transport modeling. Kipp (1987) proposed the use of the Heat-and Solute-Transport Program (HST3D) to simulate ground-water flow and associated heat and solute transport in three dimensional. A numerical algorithm was used for in general quasi 3-D case by Yakirevich et al. (1998). Panday et al. (1993) presented an efficient finite-element model which overcomes many of the challenges encountered in solving complex problems. Guyonneta and Nevilleb (2004) developed a 3-D solute migration from a plane-source. Zhang et al. (20 07) mathematically analyzed the finite element solution of quasi-three-dimensional (quasi-3-D) groundwater flow. A Domenico solution was extended to a prism source by considering off-domain contaminant transport (Promma, 2010). Harrouni et al. (1997) established Dual Reciprocity Boundary Element Method (DRBEM) as a numerical tool coupled with Extended Kalman filtering for aquifer parameter estimation. Chang and Jin (2005) proposed the use of Kalman filter with regional noise to improve the accuracy of a contaminant transport models. Chang and Latif (2010) used Extended Kalman

* Corresponding author. Tel.: +1 336 2852452; fax: +1 336 3347126.
E-mail address: godwinappiah30@yahoo.com (G. A. Assumaning).

ISSN: 1726-2135 print/1684-8799 online
 doi: 10.3808/jei.201400280

filter as data assimilation tool in subsurface porous environment. Hendricks Franssen and Kinzelbach (2009) applied Ensemble Kalman filtering (EnKF) to off-line calibration of transient groundwater flow models with many nodes. Li et al. (2012) applied ensemble Kalman filter to jointly calibrate porosity and conductivity by assimilating concentration data. To handle the large computational cost of EnKF on the flow and transport model, Li et al. (2012) proposed to couple ensemble KF with upscaling to reduce the intensive computational cost. An alternative to reduce computational cost, Xu et al., (2012) proposed a parallel version of ensemble Kalman filter for the aquifer characterization. Chang and Assumaning (2010) also worked on 2-dimensional subsurface radioactive contaminant modeling using Particle and Kalman filter schemes. A 3-D subsurface transport model was used by Cheng (2000) to generate the analytical, numerical and Kalman filter results spatially and temporally under continuous contaminant input conditions. Extended Kalman Filter (EKF) was used to recalculate the polynomial chaos expansions for the uncertain states and the uncertain parameters (Blanchard et al., 2007). Wagner (1992) used parameter estimation/source characterization inverse model combined with groundwater flow and contaminant transport. Aguirre and Haghighi (2003) further developed a finite element solution of the stochastic differential equation for transient contaminant transport. Zhou et al. (2011) proposed the use of normal-score ensemble KF (NS-EnKF) compared to the standard EnKF in log-conductivity characterization and transport predictions. Li et al. (2011) developed a 3-D transport upscaling in highly heterogeneous media.

In this research, the concentrations of a groundwater contaminant are simulated spatially and temporally in porous subsurface environment using prediction techniques such as numerical (finite-difference) method, Kalman filter without Parameter Estimation and Kalman filter coupled with Monte Carlo sampling. The first-order decay rate parameter is initially estimated using Monte Carlo sampling method. The contaminant concentration is then predicted at each time step with the estimated first-order decay rate parameter and compared to the other prediction techniques. A Simulated True value is generated as reference data set to test the accuracy of the results generated from all the techniques. A computer simulation is performed using a 3-D subsurface contaminant transport model in partial differential form with a first-order decay rate parameter. The objective of this work is to improve the accuracy and effectiveness of the 3-D subsurface contaminant transport modeling using the Kalman filters with sparse observation data points and also to assess the impact of estimating the first-order decay rate parameter using Monte Carlo sampling method.

2. Methodology

2.1 Model Description

The subsurface environment is made up of a complex, three-dimensional, heterogeneous and hydrogeologic setting. This variability influences groundwater flow and transport, and such a reality can be described accurately only through careful hydrogeologic practice in the field (Konikow, 2011). In this research, a 3-D subsurface advection-dispersion model for a non-conservative solute in a uniform, isotropic, saturated groundwater flow field along the x direction is used. The general advection-dispersion equation obtained when absorption and decay mechanisms are added is given in Equation 1. The deterministic mathematical model was used by Domenico and Schwartz (1990). The model is used to simulate a solute transport in the subsurface in a quasi 3-D form (layers). The dispersion mechanism is in the x-, y- and z- directions. The model considers the concentration of solute sorbed to the porous medium. The presence of first-order decay rate parameter makes the solute reactive (non-conservative):

$$\frac{\partial C}{\partial t} = D_x\left(\frac{\partial^2 C}{\partial x^2}\right) + D_y\left(\frac{\partial^2 C}{\partial y^2}\right) + D_z\left(\frac{\partial^2 C}{\partial z^2}\right) - V\left(\frac{\partial C}{\partial x}\right) - \frac{\rho_b}{\eta}\left(\frac{\partial S}{\partial t}\right) - kC \quad (1)$$

where C is the concentration of contaminant, (mg/L); V is the linear velocity, (m/day); ρ_b is bulk density of the porous medium, (mg/L); D_x, is the dispersion coefficients in the x-direction, (m^2/day); D_y is the dispersion coefficients in the y-direction, (m^2/day); D_z is the dispersion coefficients in the z-direction, (m^2/day); S is the contaminant concentration in the sorbed phase, (mg/L); η is the porosity, x, y and z are the Cartesian coordinates; k is the first-order decay rate parameter, (1/day) and t is time in days. The amount sorbed into the porous medium is given as:

$$S = K_d C \quad (2)$$

$$R = 1 + \rho_b\left(\frac{K_d}{\eta}\right) \quad (3)$$

where K_d is the linear partition coefficient and R is the retardation factor. By assuming a linear partitioning between the C and S, and incorporating retardation factor R into the model, the deterministic model for subsurface contaminant transport is represented in a partial differential equation (PDE) form in Equation 4 (Cheng, 2000):

$$\frac{\partial C}{\partial t} = \frac{D_x}{R}\frac{\partial^2 C}{\partial x^2} + \frac{D_y}{R}\frac{\partial^2 C}{\partial y^2} + \frac{D_z}{R}\frac{\partial^2 C}{\partial z^2} - \frac{V}{R}\frac{\partial C}{\partial x} - \frac{kC}{R} \quad (4)$$

The initial and boundary conditions of the subsurface transport model with instantaneous point source are given in Equations 5 and 6, respectively:

$$C(x,y,z,t)_{t=0} = C_0(x_0,y_0,z_0) \quad (5)$$

$$C(x,y,z,t)_{t=\Omega} = 0 \quad (6)$$

Ω is chosen as the boundary. The initial conditions are simply the values of the state specified everywhere inside the boundary at the start of the simulation.

2.2. Data Assimilation with Sparse Observation Data Points

Practically, sparse observation data can be used in the data assimilation process since it is quite expensive and laborious to take field measurement at every location in the domain space. The idea of few observation points is also to test the accuracy of the data assimilation filters. It is for these reasons that, few observation points were chosen and used to run the data assimilation filters. The domain space used in this research was represented as ($10 \times 10 \times 3$) in a 3-D form with the number of grid points on x-direction, y-direction, and z-direction being 10, 10 and 3, respectively. Therefore, the total grid points in quasi 3-D form is 300, representing a full observation data set. The 300 data points indicate that there are 100 data points on each of the three layers. Six observation data points were selected from each layer to test the efficacy of the filtering methods. In practice, full data points are expensive and labor-intensive to gather hence the use of six observation data points. The six data points also represent 6% of the full observation data points. Hence a total of 18 observation points were chosen to run the simulation. The six observation nodes on the top layer (x-y plane) in the domain space were taken at locations (2, 2), (5, 2), (8, 2), (2, 8), (5, 8) and (8, 8). These sparse observation points were introduced into the filtering process at every time step. Chang and Boateng (2011) adopted the concept of sparse observation data points for their two-dimensional subsurface contaminant modeling.

2.3. Numerical Method

In this work, the 3-D subsurface contaminant transport model is numerically solved using a Finite-differencing scheme called Forward-Time and Central-Space (FTCS). FTCS is a numerical analysis method based on central difference in space and the forward Euler method in time with a first-order convergence in time. The space and time steps chosen satisfy the numerical stability and convergence criteria of the Peclet number. Jin (1996) and Chang and Assumaning (2011) proposed the use of FTCS in their contaminant modeling approach using the Kalman filter and Particle filter. Zou and Parr (1995) used finite-difference method (FDM) to predict two-dimensional aquifer transportation. The FTCS provides the deterministic solution of the model in Equation 4. Typically, finite difference approximations requires the use of Taylor's expansions to the equations i.e. flow and transport and approximating the derivatives in the equation.

The state transition matrix (STM) generated from the model discretization clearly determines the contaminant concentration of the next time step if the prior state is known. The initial concentration, boundary conditions and the STM determine the results of the numerical scheme. The algorithm for the numerical scheme was coded in Matlab 7.1 to estimate the concentration of the contaminant. A pulse input and constant hydraulic parameters were used to generate the numerical solution. The next concentration vector is obtained by multiplying the prior concentration vector with the State Transition Matrix.

2.4. Process and Observation Models

Two main data sets are required to run the data assimilation filters. The two governing equations needed to generate these data sets are process and observation equations. These two equations are dynamic and stochastic in nature. Stochastic approaches have resulted in many significant advances in characterizing subsurface heterogeneity and dealing with uncertainty (Gelhar, 1993). The process or system model in this research is the numerical model with an additional white (Gaussian) noise. The process equation is a discrete-time controlled process (Cheng, 2000). The generic form of the process equation is given as:

$$X_{t+1} = AX_t + w_t \quad t = 0, 1, 2, 3, \ldots \tag{7}$$

where X_{t+1} is the vector of contaminant concentration at all nodes at time, $t + 1$; X_t is the vector of contaminant concentration at all nodes at time, t, A is the STM that runs till the last time step, w_t is the vector of system noise at time, t. The system noise vector w_t is assumed to be normally distributed with covariance of Q_t and a zero mean. A standard deviation of 10% as part of the error component was added to the numerical scheme to generate the dynamic system states.

Typically, field data is used as observation. However in this research, the observation/measurement data set is generated from the Simulated True value with additional white Gaussian noise. The noise component in the model reflects the heterogeneous and stochastic nature of the subsurface environment. The equation governing the observation data is given as:

$$Z_t = HX_t^T + O_t \tag{8}$$

where Z_t is the vector of the observed values for all nodes at time step t, X_t^T is the Simulated True value of the state for all nodes at time step t, O_t is vector of the observation error and H is the measurement sensitivity matrix. The observation error vector O_t is assumed to be normally distributed with covariance of R_t and zero mean. A standard deviation of 5% was chosen and added to the Simulated True value to generate the observation data set. The parameters chosen are based on typical hydrologic values and values from literature.

2.5. Kalman Filter without Parameter Estimation

The Kalman filter is considered to be a very powerful tool since it supports estimation of past, present and even future states even when the precise nature of the modeled system is unknown (Welch et al., 2004). The solution is recursive since updated estimate of the state (concentration) is computed from

the previous estimate and new input data (Haykin, 2001). The Kalman filter approach only stores the previous (prior) estimate. Cheng (2000), Chang and Jin (2005) and Chang and Assumaning (2010) used the Kalman filter to predict the concentration of contaminant in 2- and 3-dimensional subsurface environment. Kalman filter is used in this study to estimate the state (concentration) of the contaminant in a groundwater. The data assimilation is carried out primarily to reduce the variance estimates of the states. The two state-space models presented in Equations 7 and 8 represent the process and observation equations, respectively control the dynamics of the Kalman filter estimation. The Kalman filter estimation equation depicting the stochastic condition is given as:

$$X_t(+) = X_t(-) + K_t\left[Z_t - HX_t(-)\right] \tag{9}$$

where $X_t(+)$ is the vector of estimated states after the Kalman filter adjustment, $X_t(-)$ is the vector of estimated states before the Kalman filter adjustment and K_t is the Kalman gain matrix. The K_t is derived by minimizing the trace of the posterior error covariance matrix, $P_t(+)$. K_t determines how much the estimated value using Kalman filter can gain from the observation. The K_t is determined by:

$$K_t = P_t(-)H^T(HP_t(-)H^T + R_t)^{-1} \tag{10}$$

$(\ )^T$ and $(\ )^{-1}$ denote the transpose and inverse of matrix respectively in Equation 10. The posterior and prior error covariance matrices given in Equations 11 and 12, respectively are advanced recursively for all time steps by:

$$P_t(+) = (I - K_tH)P_t(-) \tag{11}$$

$$P_{t+1}(-) = AP_t(+)A^T + Q_t \tag{12}$$

The initial value of the posterior error covariance matrix is given as;

$$P_0 = E\left[(X_0 - E\left[X_0\right])(X_0 - E\left[X_0\right])^T\right] \tag{13}$$

where X_0 is the initial estimated state at time step zero.

2.6. Kalman Filter Coupled with Monte Carlo Sampling

Since the subsurface environment is heterogeneous in nature, the hydrologic parameters are characterized by uncertainties. These uncertainties in parameters are mostly due to changes in chemical, physical and biological processes of the contaminant. For this reason, the first-order decay rate parameter, k is estimated using Monte Carlo sampling method. Walker (2006) estimated the model parameters using Extended Kalman filter. Moradkhani et al. (2005) also used Ensemble Kalman filter for dual state-parameter estimation of hydrological models. The algorithm used in the parameter estimation is shown in Figure 1.

The initial first-order decay rate parameter, k to be sampled is taken from the k used in the numerical scheme. A Monte Carlo sampling technique is initiated with k value of 0.35 1/day, standard deviation of 5% and normally distributed random error. The idea of Monte Carlo technique is to randomly generate inputs from a probability distribution over the domain. The standard deviation and normally distributed random error are incorporated at every time step. A total of n samples of k are generated to begin the estimation process. The k values in practice cannot be negative, so a condition is set in the algorithm to store positive k values and set negative values to zero. Out of these samples, n number of new STM and states vectors is generated. The error differences between the states and the observations at every time step are evaluated. The k that produces the minimum error is selected to be the actual k value for the numerical scheme and it is also sampled to begin the next time step. Similarly, the state vector with minimum Mean Absolute Error (MAE) values is selected for the next time step. In this research, fifty number of k samples were generated at each time step.

The Kalman filter algorithm is then applied to generate the states after the k is estimated for all the time steps. However, a modification is made to the process equation as shown in Equation 7 to incorporate the new k estimated. The parameter estimation is also performed as a means to calibrate the numerical model and to determine the actual decay rate parameter value. Automated parameter-estimation techniques improve the efficiency of model calibration (Konikow, 2011). The algorithm used to estimate the states after the parameter estimation of the first-order decay rate parameter k is shown in Figure 2. The estimated k is introduced into the process model at each time step to begin the evolution of the states. The new k alters the STM at every time step. The posterior and prior error covariance matrices are subsequently changed due to the change in the State Transition Matrix. The posterior and prior error covariance matrices are required to be positive

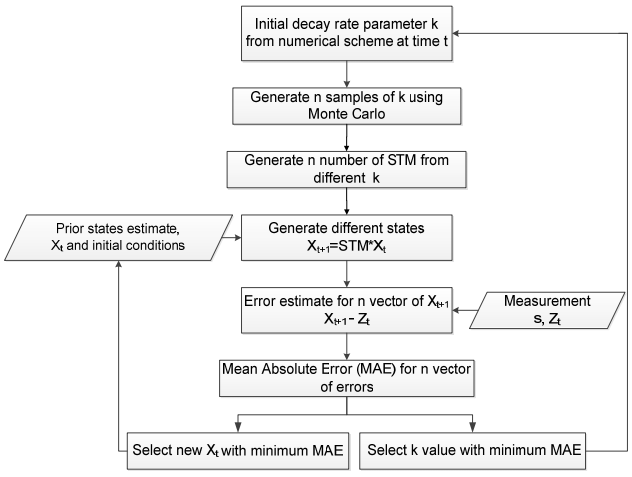

Figure 1. Flowchart of first-order decay rate parameter estimation using Monte Carlo sampling method.

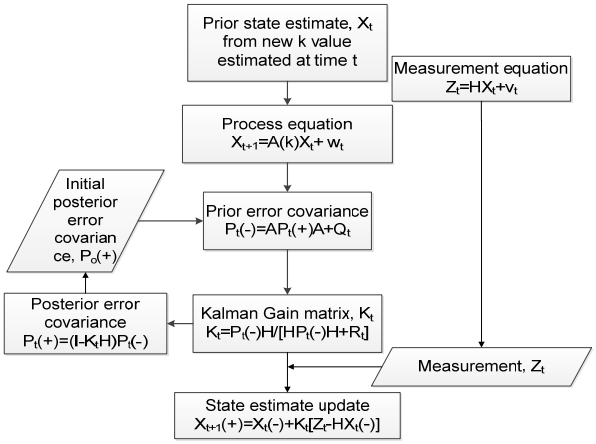

Figure 2. Flowchart of Kalman filter algorithm with estimated first-order decay rate parameter.

semi definite matrices to avoid degeneracy and numerical instability.

The sparse observation data are introduced at each time step to guide the estimation made. Also, priori knowledge of process error covariance, Q_t and measurement error covariance, R_t are required to begin the filtering process. However, the Q_t and R_t are updated with time. The optimal Kalman gain, K_t which acts a weight factor is a function of the measurement sensitivity matrix, the posterior and prior error covariance matrices and the measurement error covariance. The difference $[Z_t - HX_t(-)]$ is called the measurement innovation. The innovation reflects the discrepancy between the estimated state and the measurement. The Kalman filter basically applies minimum mean-square error to estimate the states (concentrations) with time and involves matrix manipulation.

2.7. Testing Numerical and Data Assimilation Filters Results

The accuracy and effectiveness of the data assimilation filters and the numerical solution were determined by comparing their results to the Simulated True value at each time step. Two error estimation equations were used to analyze the prediction results. These methods are Mean Absolute Error (MAE) and Maximum Absolute Error (E_{max}). The MAE and E_{max} equations are defined in Equations 14 and 15, respectively:

$$MAE(t) = \sum \frac{\left|C^E(x,y,z,t) - C(x,y,z,t)\right|}{N} \quad (14)$$

$$E_{\max}(t) = \sum \frac{\left|C^E(x,y,z,t) - C(x,y,z,t)\right|}{C_{\max}(x,y,z,t)} \quad (15)$$

where N is the number of sampling nodes, C^E is the estimated concentration of the contaminant at time, t, C is the Simulated True value at time, t and C_{max} is the maximum Simulated True value at time, t.

3. Results and Discussion

3.1. Model Parameters

The hydrogeologic properties like porosity, advection and dispersion are needed to perform the simulations using the data assimilation filters and the numerical scheme. The model parameters used were assumed based on the previous research done by Cheng (2000) and typical parameter values used in literature. The parameter values used in this 3-D contaminant transport modeling is provided in Table 1. An instantaneous contaminant was injected into the grid point at coordinates (5, 5, 1).

Table 1. Model Parameters and Corresponding Values

Model Parameter	Value
Initial contaminant concentration, C_o	10,000 mg/L
Total number of nodes or grid points, N	300 (10×10×3)
Sparse observation points, M	18
First-order decay rate parameter, k	0.35 1/day
Linear velocity for numerical method, V_n	0.15 m/day
Linear velocity for Simulated True value, V_s	0.22 m/day
Retardation factor, R	1.125
Dispersion in x-direction, D_x	0.35 m^2/day
Dispersion in y-direction, D_y	0.35 m^2/day
Dispersion in z-direction, D_z	0.3 m^2/day
Grid interval in x,y and z direction (d_x,d_y,d_z)	2 m
Simulation time	12 days
Time interval, dt	0.3 day
Total time steps	40

3.2. Numerical Solution (FTCS)

From Figure 3 where numerical solution at time step 30 for all three layers is shown, the concentration of the contaminant decreases with corresponding increase in time. The concentration reduction is as a result of the decay, dispersion and flow of the contaminant. The minimum concentration for layer 1, 2 and 3 are 22, 10 and 2.6 mg/L, respectively. The contours of the contaminant plume shows that the plume from numerical method moves slower than the Simulated True value mainly due to the relatively smaller linear velocity used in the estimation. The numerical solution is characterized with error due to the constant parameters in the simulation model. These errors tend to propagate with time.

3.3. Kalman Filter without Parameter Estimation Results

The Kalman filter without Parameter Estimation and Simulated True value contours of the contaminant plume developed for time step 30 are shown in Figure 4. The contours are represented in quasi 3-D form to depict the concentration and plume at each of the three layers. The closeness of estimated results to the Simulated True value is visibly shown. However, the Kalman filter without Parameter Estimation results are randomized due to the random Gaussian noise introduced into

the simulation at each time step.

Similarly, the concentration of the subsurface contaminant tends to decrease with increase in time as depicted by the numerical method. However, the Kalman filter without Parameter Estimation results is relatively closer to the Simulated True value than the numerical solution. The closeness is primarily due to the data assimilation nature of the Kalman filter. From Figure 4, the concentration contours spread out with time and reduces from the top layer to the bottom layer. The spread of the contaminant is more along the x direction. The presence of dispersion coefficients and linear velocity parameters in the contaminant model contributed to the spreading of the contaminant with time. The Kalman filter without Parameter Estimation with sparse observation data (18 data points) introduced into the assimilation process to estimate the concentration of the contaminant spatially and temporally yielded a relatively better solution than the numerical solution. The maximum concentration shown in Figure 4 for layers 1, 2 and 3 are 88, 40 and 10.4 mg/L, respectively.

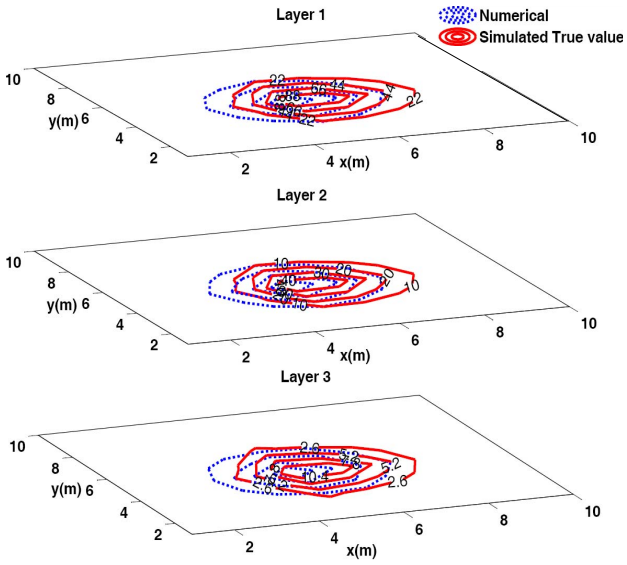

Figure 3. Numerical solution contours at time step 30 for 3 layers.

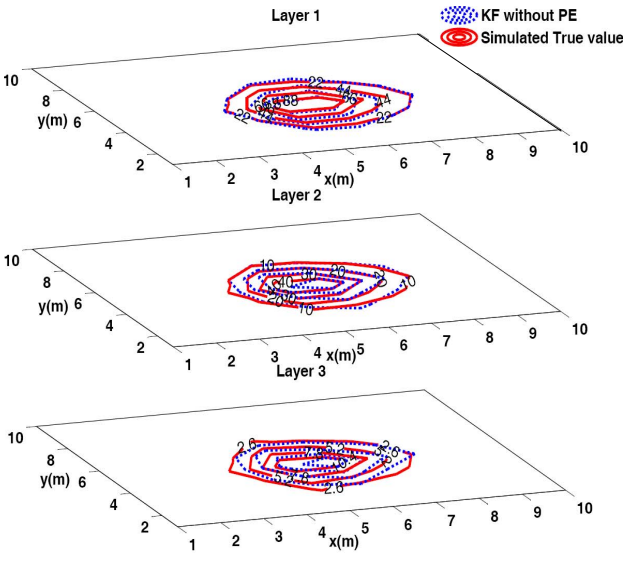

Figure 4. Kalman filter without Parameter Estimation contours at time step 30 for 3 layers.

3.4. Kalman filter coupled with Monte Carlo Sampling Results

Parameter estimation is essential in heterogeneous field like the subsurface environment. This helps the prediction model to mimic real life situation of contaminant transport and it improves the prediction accuracy. However, the computational cost and challenges can be higher due to the volume of data used in the data assimilation process. In practice, the hydrologic parameters are not constant due to the change in physical, biological and chemically processes and other uncertainties in the subsurface environment. Thiemann et al. (2001) used Bayesian recursive approach for parameter estimation and hydrologic prediction. As part of the research, a first-order decay rate parameter is estimated using Monte Carlo sampling method. The estimated value is then introduced into the Kalman filter to predict the contaminant concentration at every time step.

The profile of the first-order decay rate parameter is shown in Figure 5. The estimation of the parameter converges and fairly stabilized at time step 3 with a value of about 0.25 1/day. This indicates that the actual first-order decay rate parameter is around 0.25 1/day. The first-order decay rate parameter value estimated at the beginning of the simulation is 1.2 1/day. Although higher initially, the estimated values were found to be closer to the actual value with time. A small degree of randomness is shown in the profile from time step 3 to 39 due

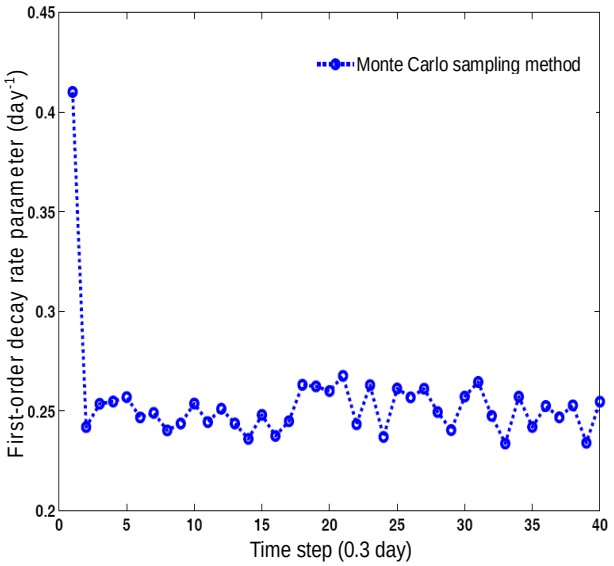

Figure 5. Profile of first-order decay rate parameter estimation using Monte Carlo sampling method.

to the sampling technique used.

The improvement in concentration prediction when first-order decay rate parameter estimated is incorporated into the model can be seen in Figure 6. The figure exhibits characteristics similar to that of Kalman filter without parameter estimation in terms of contour shape and concentration values. The distinction between the two is visible in the error estimation profiles. The quasi 3-D contour shape of the contaminant shows the contaminant plume in layers along the z-direction. Every layer is a cross-sectional view of the contaminant shape in 2-D form along the x and y directions. The closeness of the predicted concentration to the Simulated True value can also be realized in the figure. Similarly, sparse observation data were introduced into the assimilation process to estimate the concentration of the contaminant. The predicted concentration profile shows erratic contaminant plume mainly due to the stochastic nature of the Kalman filter and the parameter estimation processes.

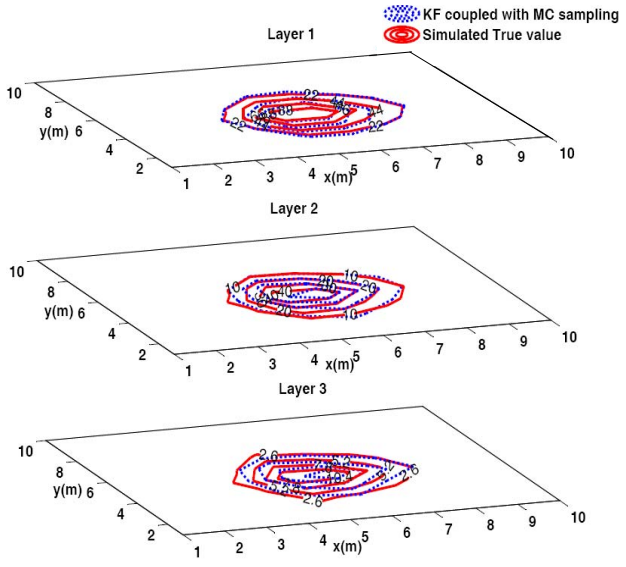

Figure 6. Kalman filter coupled with Monte Carlo sampling contours at time step 30 for 3 layers.

3.5. Comparison of all Prediction Techniques Results

To further analyze the accuracy of each prediction results at different location and time concurrently, the contour of all techniques results for layer 1 at time step 40 were plotted as shown in Figure 7. The contour plot is aimed at facilitating comparative analysis of the prediction results. From Figure 7, numerical solution is farther away from the Simulated True value. This indicates that the error in the numerical solution is relatively the largest compared to other prediction techniques. The closest prediction technique to the Simulated True value is Kalman filter coupled with Monte Carlo sampling. The deviation by this technique from the Simulated True value is relatively smaller. This suggests that the Kalman filter coupled with Monte Carlo sampling is relatively better than the rest of the techniques used in this research.

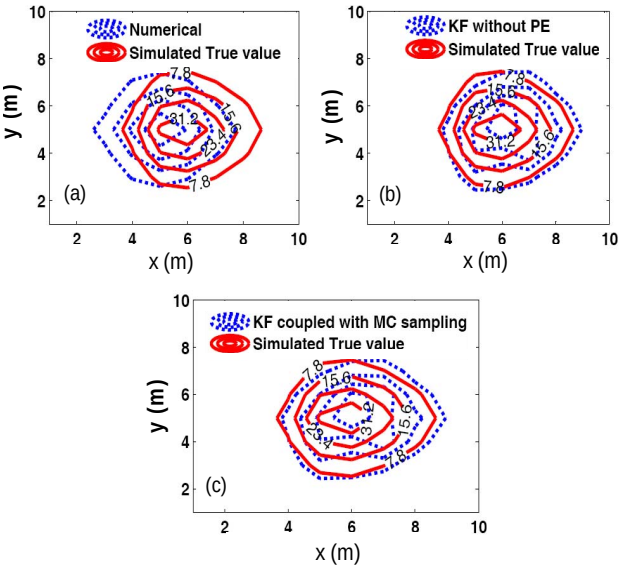

Figure 7. Contaminant concentration contours for all the prediction techniques at time step 40 for 1 layer.

3.6. Accuracy of the Numerical Method and Data Assimilation Filters

The measure of accuracy of the results of each prediction technique is assessed using Mean Absolute Error (MAE) and Maximum Absolute Error (E_{max}). The purpose of the error/residual analysis is to show the deviation from the Simulated True value with time.

The Mean Absolute Error (MAE) is used in this work to find the absolute difference between the predicted and the Simulated True value at every time step. Figure 8 shows the profile of the MAE profiles for all prediction techniques. The data assimilation filters were found to be unstable in nature due to fewer observation data used and the random Gaussian noise introduced into the data assimilation process. The deviation by the numerical solution from the data assimilation filters is very visible and indicates the degree of error in the results. The profile also shows the erratic nature of the filters from time step 1 to 20. The randomized nature of the filters is as a result of the stochastic Markov chain process used in simulating the contaminant transport. At the end of the simulation, errors in the numerical solution and the filters were approximately 1.8 and 0.4 mg/L, respectively. This indicates that the filters are capable of reducing the error in the numerical method by 77% at the end of the prediction. Also, the Kalman filter coupled with Monte Carlo sampling has the least error estimate.

In other to establish which prediction technique works better, Maximum Absolute Error (E_{max}) was calculated at each time step for all the techniques. Figure 9 shows the plot of the E_{max} for all the prediction techniques. The error in the numerical solution is highly increasing due to the accumulation of error with time. Among the data assimilation filters, the Kalman filter without Parameter Estimation shows the result which is highly randomized. The profiles further confirm the accuracy

of the Kalman filter coupled with Monte Carlo sampling.

4. Conclusions

Several numerical and analytical models have been used as predictive tool in solute transport. In this research, data assimilation filters namely Kalman filter without Parameter Estimation and Kalman filter coupled with Monte Carlo sampling were used as tools for predicting contaminant concentration in subsurface porous environment. Monte Carlo sampling method was also adopted as a first-order decay rate parameter estimation technique due to the uncertainties associated with hydrologic parameters and aquifer properties. Fifty samples were drawn from a known probability distribution for the parameter estimation process.

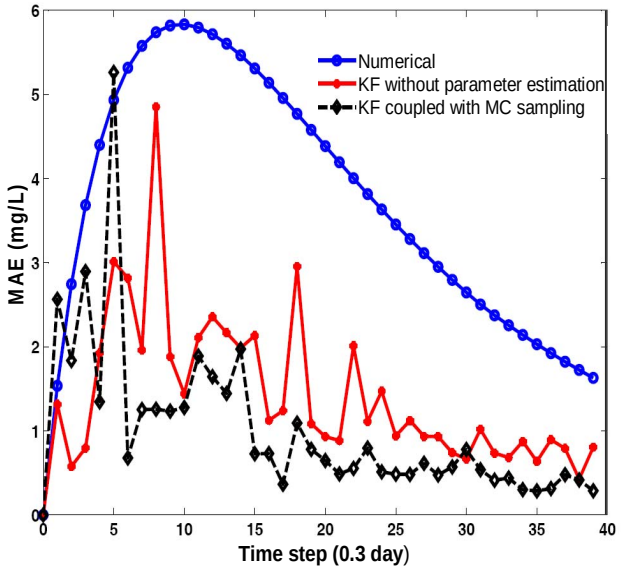

Figure 8. Mean Absolute Error (MAE) profiles for all prediction techniques.

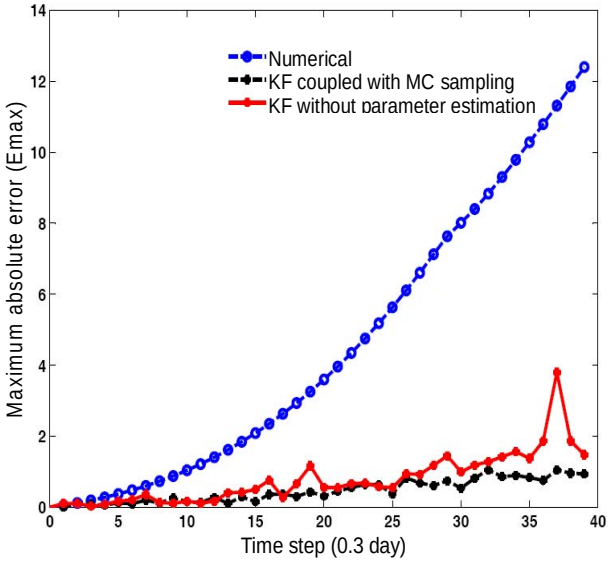

Figure 9. Maximum Absolute Error (Emax) profile for all prediction techniques.

The filters used sparse observation data points (18 points) as opposed to the full data set (30 points) used by Chang and Jin (2005) and Chang and Li (2008) in their research. However, the filters guided by the sparse observation data in the filtering process, provided a better contaminant concentration prediction than the numerical method. The discretization of the 3-D subsurface contaminant transport model by FTCS introduces round-off, truncation and approximation errors into the numerical model.

The filters are initiated with unknown noise statistics and uncertain states. However, they show a strong convergence trends in the error estimation profiles given. They are initially erratic but subsequently stabilize and converge with time indicating the effectiveness of the filtering process. The filters have an advantage of minimizing the error or residual between the observed values and the estimated values. From the MAE profile, the errors in the numerical solution and the filters were approximately 1.8 and 0.4 mg/L, respectively at the end of the simulation. This indicates that the data assimilation filters are capable of reducing the error in the numerical method by 75% at the end of the prediction. All the profiles on MAE and E_{max} confirm the higher accuracy of the Kalman filter coupled with Monte Carlo sampling.

The contaminant transport simulations for all the data assimilation filters and the numerical method were run on a personal computer (PC) with processor speed of 2.99 GHz and RAM of 3.25 GB. The filters with Parameter Estimation take about eight minutes each whilst the numerical method takes three minutes. The filters have high computational time and challenges demands due to the Parameter Estimation, prediction and updating processes of the algorithm. However, it is worth using the Kalman filter coupled with Monte Carlo sampling due to its accuracy in prediction and effectiveness.

Acknowledgments. This work was sponsored by the Department of Energy Samuel Massie Chair of Excellence Program under Grant No. DF-FG01-94EW11425. The views and conclusions contained herein are those of the writers and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the funding agency.

Notations

The following symbols are used in this paper:

A = state transition matrix (STM)

C = concentration of the contaminant

d_x, d_y, d_z = grid interval at the x, y and z-coordinate, respectively

D_x, D_y, D_z = dispersion coefficients in the x, y and z direction, respectively

H = measurement sensitivity matrix

k = first-order decay rate parameter

k_{t+1} = vector of estimated first-order decay rate parame- ter at time, t + 1;

k_t = vector of first-order decay rate parameter at time, t;

K_t = Kalman optimal gain matrix

K_p = linear partition coefficient

η = porosity

N = number of sampling nodes

O_t = observation error vector

ρ_b = bulk density of the porous medium, (mg/L)

$P_t(-)$ = prior optimal estimate error covariance matrix

$P_t(+)$ = posterior optimal estimate error covariance matrix

R = retardation factor

R_t = measurement covariance matrix

S = contaminant concentration in the sorbed phase

t = time in days

V = average linear velocity

w_t = vector of process error

X_t = vector of contaminant concentration at all nodes at time, t

X_{t+1} = vector of contaminant concentration at all nodes at time, t + 1

$X_t(-)$ = vector of estimated value before the Kalman filter adjustment

$X_t(+)$ = vector of estimated value after the Kalman filter adjustment

X_t^T = Simulated True value of the state for all nodes at time step t

x, y, z = Cartesian coordinates

Z_t = state vector for observed values for all nodes at time step t

References

Aguirre, C.G., and Haghighi, K. (2003). Stochastic modeling of transient contaminant transport. *J. Hydrol.,* 276, 224-239. http://dx.doi.org/10.1016/S0022-1694(03)00059-3

Blanchard, E., Sandu, A., and Sandu, C. (2007). Parameter Estimation Method using an Extended Kalman Filter. Proc. of the Joint North America, Asia-Pacific ISTVS Conference and Annual Meeting of Japanese Society for Terramechanics. Fairbanks, Alaska, USA, June 23-26, 2007.

Chang, S.Y., and Jin, A. (2005). Kalman filtering with regional noise to improve accuracy of contaminant transport models. *J. Environ. Eng.,* 131(6), 971-982. http://dx.doi.org/10.1061/(ASCE)0733-9372(2005)131:6(971)

Chang, S.Y., and Assumaning, G. (2011). Subsurface radioactive contaminant modeling using Particle and Kalman filter schemes. *J. Environ. Eng.,* 137(4), 221-229. http://dx.doi.org/10.1061/(ASCE)EE.1943-7870.0000317

Chang, S.Y., and Boateng, L.K. (2012). Adaptive Kalman filtering scheme for the simulation of benzene in subsurface environment. *J. Environ. Eng.,* 138, 542-551. http://dx.doi.org/10.1061/(ASCE)EE.1943-7870.0000500

Chang, S.Y., and Latif, S.M.I. (2010). Extended Kalman filtering to improve the accuracy of a subsurface contaminant transport model. *J. Environ. Eng.,* 136(5), 466-474. http://dx.doi.org/10.1061/(ASCE)EE.1943-7870.0000179

Cheng, X. (2000). Kalman filter scheme for three-dimensional subsurface transport simulation with a continuous input. Master of Science Thesis, Civil and Environmental Engineering; North Carolina Agricultural and Technical State University.

Domenico, P.A., and Schwartz, F.W. (1900). Physical and Chemical Hydrology. Wiley-Interscience publication, New York.

Gelhar, L.W. (1993). Stochastic Subsurface Hydrology. Prentice-Hall, Englewood Cliffs: 390 pp.

Guyonneta, D. and Nevilleb, C. (2004). Dimensionless analysis of two analytical solutions for 3-D solute transport in groundwater. *J. Contam. Hydrol.,* 75(1-2), 141-153. http://dx.doi.org/10.1016/j.jconhyd.2004.06.004

Harrouni, K.E., Ouazar, D., Wrobel, L.C. and Cheng, A.H.D. (1997). Aquifer parameter estimation by extended Kalman filtering and boundary elements. *Eng. Anal. Bound. Elem.,* 19, 231-237. http://dx.doi.org/10.1016/S0955-7997(97)00008-8

Haykin, S., (2001). Kalman filtering and Neural Networks. John Wiley and Sons, Inc. ISBN: 0-471-36998-5. http://dx.doi.org/10.1002/0471221546

Hendricks Franssen, H.J., and Kinzelbach, W. (2009). Ensemble Kalman filtering versus sequential self-calibration for inverse modeling of dynamic groundwater flow systems. *J. Hydrol.,* 365, 261-274. http://dx.doi.org/10.1016/j.jhydrol.2008.11.033

Jin, A. (1996). An optimal estimation scheme for subsurface contaminant transport model using Kalman-Bucy filter. Master's Thesis, Department of Civil and Environmental Engineering, North Carolina A&T State University, Greensboro, NC

Kipp, K.L., Jr. (1987). HSTSD: HST3D: A Computer Code for Simulation of Heat and Solute Transport in Three-Dimensional Ground-Water Flow Systems.

Konikow, L.F. (2011). Use of Numerical Models to Simulate Groundwater Flow and Transport. U.S Geological Survey, Reston, Virginia.

Li, L., Zhou, H., and Gómez-Hernández, J.J., (2011). Transport upscaling using multi-rate mass transfer in three-dimensional highly heterogeneous porous media. *Adv. Water Resour.,* 34(4), 478-489. http://dx.doi.org/10.1016/j.advwatres.2011.01.001

Li, L., Zhou, H., Hendricks Franssen, H.J., and Gómez-Hernández, J.J., (2012). Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via Ensemble Kalman filter. *J. Hydrol.,* 428-429, 152-169. http://dx.doi.org/10.1016/j.jhydrol.2012.01.037

Li, L., Zhou, H., Hendricks Franssen, H.J., and Gómez-Hernández, J.J., (2012). Modeling transient groundwater flow by coupling Ensemble Kalman filtering and upscaling. *Water Resour. Res.,* 48, 1-19. http://dx.doi.org/10.1029/2010WR010214

Moradkhani, H., Sorooshian,S., Gupta, H.V., and Houser, P.R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.,* 28, 135-147. http://dx.doi.org/10.1016/j.advwatres.2004.09.002

Panday, S., Peter, S. H., Therrien, R., and Nicholas, R.L. (1993). Improved three-dimensional finite-element techniques for field simulation of variably saturated flow and transport. *J. Contam. Hydrol.,* 12(1-2), 3-33. http://dx.doi.org/10.1016/0169-7722(93)90013-I

Promma, K. (2010). Approximate solution to simulate dissolved contaminant transport in groundwater from prism source. *J. Hydrol.,* 389, 381-389. http://dx.doi.org/10.1016/j.jhydrol.2010.06.021

Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models. *Water Resour. Res.,* 37, 2521-2001 http://dx.doi.org/10.1029/2000WR900405

Wagner, B.J. (1992). Simultaneous parameter estimation and conta-

minant source characterization for coupled groundwater flow and contaminant transport modeling. *J. Hydrol.,* 135, 275-303. http://dx.doi.org/10.1016/0022-1694(92)90092-A

Walker, D.M. (2006). Parameter estimation using Kalman filters with constraints. *Int. J. Bifurcation Chaos,* 16(4), 1067-1078. http://dx.doi.org/10.1142/S0218127406015325

Welch, G., and Bishop, G. (2006). An introduction to the Kalman filter. Department of Computer Science, University of North Carolina at Chapel Hill.

Xu, T., Gómez-Hernández, J.J ., Li, L., and Zhou, H., (2012). Parallelized ensemble Kalman filter for hydraulic conductivity characterization. *Comput. Geosci.,* 52, 42-49. http://dx.doi.org/10.1016/j.cageo.2012.10.007

Yakirevich, A., Borisov, V., and Sorek, S. (1998). A quasi three-dimensional model for flow and transport in unsaturated and saturated zones:1. Implementation of the quasi two-dimensional case. *Adv. Water Resour.,* 21(8), 679-689. http://dx.doi.org/10.1016/ S0309-1708(97)00031-6

Zhang, X.W., Takeuchi, K., and Chen, J. (2007). Temporal and spatial discretization on quasi-3-d groundwater finite element modelling to avoid spurious oscillation. *J. Hydrodynamics,* 19(1), 68-77. http://dx.doi.org/10.1016/S1001-6058(07)60030-4

Zhou, H., Gómez-Hernández, J.J., Hendricks Franssen, H.J., and Li, L. (2011). An approach to handling non-Gaussianity of parameters and state variables in Ensemble Kalman filtering. *Adv. Water Resour.,* 34(7), 844-864. http://dx.doi.org/10.1016/j.advwatres.2011.04.014

Zou, S, and Parr, A. (1995). Optimal Estimation of two-dimensional contaminant transport. *Ground water,* 33, 319-325. http://dx.doi.org/10.1111/j.1745-6584.1995.tb00287.x