Addressing model bias and uncertainty in three dimensional groundwater transport forecasts for a physical aquifer experiment

J. B. Kollat,[1] P. M. Reed,[1] and D. M. Rizzo[2]

Received 14 June 2008; accepted 18 July 2008; published 3 September 2008.

[1] This work contributes a combination of laboratory-based aquifer tracer experimentation and bias-aware Ensemble Kalman Filtering (EnKF) to demonstrate that systematic modeling errors (or bias) in source loading dynamics and the spatial distribution of hydraulic conductivity pose severe challenges for groundwater transport forecasting under uncertainty. The impacts of model bias were evaluated using an ammonium chloride tracer experiment conducted in a three dimensional laboratory tank aquifer with 105 near real-time sampling locations. This study contributes a bias-aware EnKF framework that (i) dramatically enhances the accuracy of concentration breakthrough forecasts in the presence of systematic, spatio-temporally correlated modeling errors, (ii) clarifies in space and time where transport gradients are maximally impacted by model bias, and (iii) expands the size and scope of flow-and-transport problems that can be considered in the future. **Citation:** Kollat, J. B., P. M. Reed, and D. M. Rizzo (2008), Addressing model bias and uncertainty in three dimensional groundwater transport forecasts for a physical aquifer experiment, *Geophys. Res. Lett.*, *35*, L17402, doi:10.1029/2008GL035021.

1. Introduction

[2] *Eigbe et al.* [1998] provide an excellent review of groundwater applications of the linear Kalman Filter (KF) as well as the extended KF for nonlinear systems. The review highlights that very few studies exist for three-dimensional groundwater flow-and-transport applications due to these problems' high-dimensional, nonlinear state spaces (i.e., heads and concentrations) and their consequent computational barriers. KF encompasses a general class of time-controlled state estimators that account for both process noise, and measurement noise simultaneously [*Grewel and Andrews*, 1993]. The KF proceeds iteratively in two steps: (i) a time update (or forecast) that projects the state and error covariance of the system ahead in time, and (ii) a measurement update that corrects the estimate and its associated error covariance by assimilating noisy measurement data. The use of extended Kalman filters for nonlinear groundwater flow-and-transport applications has been limited by a curse of dimensionality largely associated with the groundwater head and contaminant covariance matrices as well as the computational burden and nonlinearity posed by the transport problem.

[3] These challenges have presented themselves in several scientific domains and motivated [*Evensen*, 2003] to formulate the EnKF framework for nonlinear oceanographic applications. The EnKF is well suited to high dimensional state spaces where traditional Kalman filtering performs poorly in terms of its accuracy and computational complexity. This study contributes a combination of numerical and physical three-dimensional aquifer modeling to test a computationally efficient formulation for a bias-aware EnKF. The overall goal of this work is to minimize the computational demands associated with the bias-aware EnKF to provide high fidelity groundwater contaminant forecasts that account for measurement uncertainties and dynamic, spatiotemporally correlated model errors. This study provides the first demonstration of the effectiveness of bias-aware filtering for fully three-dimensional groundwater transport. In the broader geophysical research context, this work uses physical tank aquifer modeling to highlight that typical sources of model bias in subsurface chemical transport forecasts yield severe, spatiotemporally complex impacts on predictions. The experimental tank aquifer test case shows that unknown historical dynamics (e.g., concentration source terms, recharge, or discharge) coupled with hydrogeologic uncertainties will severely limit our predictions of groundwater or chemical fluxes across material or process interfaces. This poses a significant challenge to future water cycle research and motivates the need for advances in groundwater transport forecasting [see *National Research Council*, 2004, 2008].

2. Physical Aquifer Experiment

[4] The University of Vermont (UVM) (see Figure 1) scaled physical aquifer transport experiment was constructed using layered porous media within a 254-cm wide $\times$ 356-cm long $\times$ 243-cm high tank constructed using a steel frame and panels. Constant head inlet and outlet reservoirs (7.6-cm wide) are located at opposite ends of the tank to control the tank's hydraulic gradient. Five layers of sand and silt media were carefully packed into the tank in 3-cm lifts. A suspended catwalk was used during the placement of media to ensure an even media distribution and packing.

[5] A network of 21 wells constructed from 5.08-cm polyvinyl chloride (PVC) pipe were distributed throughout the tank. Each well was screened (10-cm) at four depths along its vertical axis. Five sensors were placed along each well (one within each media layer) for a total of 105 sensors. Sensors at each location included a Time Domain Reflec-

[1]Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.
[2]Department of Civil and Environmental Engineering, University of Vermont, Burlington, Vermont, USA.

0094-8276/08/2008GL035021

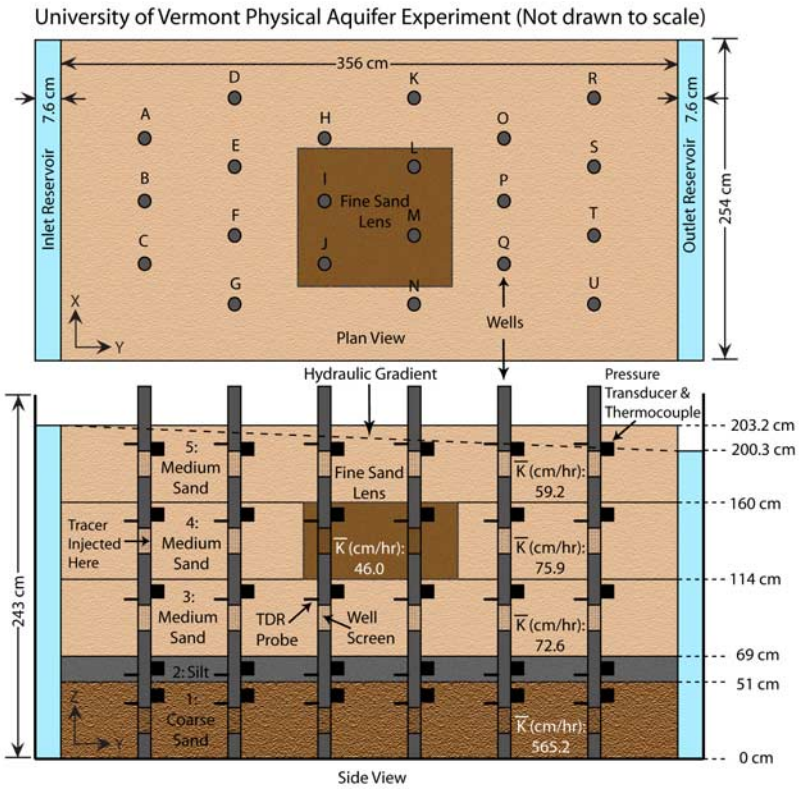

Figure 1. Diagram of the UVM physical tank experiment's dimensions, media layering and average calibrated K values, sensor and well locations, and constant head reservoirs.

tometer (TDR) for measuring concentration data, a thermocouple, and a pressure transducer.

[6] An ammonium chloride tracer was injected through sampling port B4 (see Figure 1) at an average concentration of 1 g-L^{-1} and a rate of 1.5 L-hr^{-1} over a 15-day period. Concentration data were collected at the 63 sampling ports in layers 3 through 5 over a 19-day period at 17.5 minute intervals.

[7] The temporal variability and systematic increasing trend in the source term's loading dynamics (see Figure 2a, port B4) provides a severe test of the bias aware EnKF formulation proposed in this paper. The complex loading history for port B4 would not be known in typical groundwater transport applications and would be modeled (as we have done) with a simplified source loading history within the MT3DMS transport model. Additionally, despite significant experimental effort spent in characterizing the hydraulic conductivity fields in the tank using slug and pump tests, the fine sand lens in the center of the flow field served as a second major source of bias. The fine lens has substantial, fully three-dimensional effects on the spatiotemporal tracer breakthroughs that are also not captured in the MT3DMS simulation.

3. Bias-Aware Ensemble Kalman Filter

[8] Using similar notation to *Evensen* [2003], the bias-aware version of the EnKF uses N ensemble members to update n modeled states based on m observations for the kth time step. The ensemble of model (ψ) and bias (η) states are defined as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_\psi \\ \mathbf{A}_\eta \end{bmatrix} = \left[\begin{pmatrix} \psi_1 \\ \eta_1 \end{pmatrix}, \begin{pmatrix} \psi_2 \\ \eta_2 \end{pmatrix}, \ldots, \begin{pmatrix} \psi_N \\ \eta_N \end{pmatrix} \right] \in \Re^{2n \times N}. \quad (1)$$

[9] The mean of $\mathbf{A}$ (denoted as $\bar{\mathbf{A}}$) is obtained through multiplication with the matrix containing $1/N$ as its elements. The ensemble perturbations of $\mathbf{A}$ are found from $\mathbf{A}' = \mathbf{A} - \bar{\mathbf{A}}$, and the covariance of $\mathbf{A}$ is found from $\mathbf{P} = \mathbf{A}'(\mathbf{A}')^T/(N-1)$. During a forecast step, (f) the bias portion of the state ensemble is updated using equation (2)

$$\mathbf{A}^f_{\eta,k} = \Lambda_k \mathbf{A}^{a,f}_{\eta,k-1} + \delta_\eta \mathbf{Q}_{\eta,k}, \quad (2)$$

where Λ_k describes the time correlation of the model bias, $\mathbf{Q}_\eta$ is an ensemble of independent sources of zero-mean, spatially correlated noise on the bias states, and δ_η is a scaling factor on this noise. The prior bias state $\mathbf{A}_{\eta,k-1}$ may either be based on a forecast (f) or update (a). Once the bias state is forecast, the ensemble of tracer concentration states is forecast using equation (3)

$$\mathbf{A}^f_{\psi,k} = \mathbf{A}^{a,f}_{\psi,k-1} - \mathbf{A}^f_{\eta,k} + \delta_\psi \mathbf{Q}_{\psi,k}, \quad (3)$$

where $\mathbf{Q}_\psi$ is an ensemble of independent sources of zero-mean, spatially correlated noise on the model states, and δ_ψ is a scaling factor on this noise. Notice that equation (3)

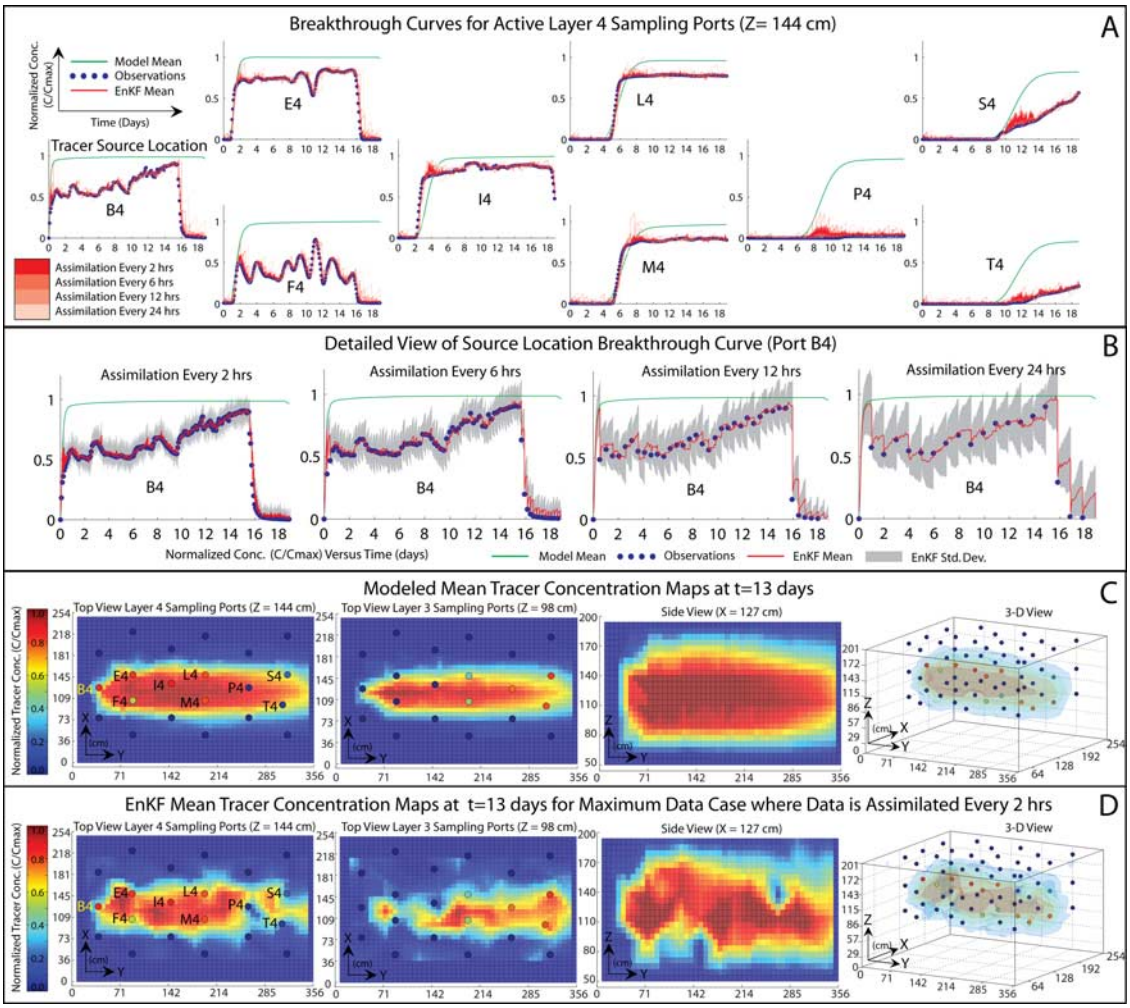

Figure 2. (a) Tracer breakthrough curves associated with the sampling ports in layer 4 plotted as normalized concentration versus time for the model, observed, and filtered concentration time-series. (b) BT curves associated with the tracer source location (port B4) for the various assimilation cases (2-hrs, 6-hrs, 12-hrs, and 24-hrs) where the gray shading represents one standard deviation above and below the ensemble mean. (c) Normalized tracer concentration (C/C_0) maps (two top views of the layer 4 and layer 3 sampling ports, a side view along the center of the tank, and a 3-dimensional view) provided by the model at t = 13-days. In addition, the coloring of the sampling ports indicates observed C/C_0 concentrations at this time step. Also shown are the numerically modeled mean tracer concentration at t = 13 days. (d) Mean filtered maps for assimilation every 2-hrs at t = 13 days (formatted similarly to Figure 2c for comparison).

fully incorporates the bias state forecast into the model state forecast (referred to as full feedback [*Drecourt et al.*, 2006]). When observations are available, the ensemble of bias states can be updated based on those observations. First, a subset of updated bias states is calculated as the difference between the observations and the model state forecasts at the observation locations. However, since m, the number of observations is typically much fewer than n, the number of model states, a Gaussian model is fit to the covariance structure of the bias obtained at the m observation locations. A contribution of this work is to spatially condition the mean bias field and its covariance on observed modeling errors using Sequential Gaussian Simulation (SGS) [*Deutsch and Journel*, 1998]. The bias present at the observation locations is used to conditionally simulate an ensemble of bias states at all n model state locations. Equation (4) is used to update the ensemble of bias states, $\mathbf{A}_\eta$,

$$\mathbf{A}^a_{\eta,k} = \Lambda_k \mathbf{A}^f_{\eta,k-1} + (1 - \Lambda_k)\mathbf{W}_k, \tag{4}$$

where Λ_k again describes the time correlation of the model bias, and $\mathbf{W}$ is the simulated ensemble of bias states based on the newly available observations. Our Bayesian conditioning of the mean and covariance of the bias field substantially improves the spatial representation of systematic model errors and takes advantage of observations to permit more advanced representations of the spatial anisotropy of model error fields. It is interesting to note that equation (4) provides a mechanism for analyzing three-dimensional time varying maps of how transport gradients impact prediction errors. These maps can provide a unique

mechanism for evaluating changes or improvements in the forward model's process formulations. Once the bias ensemble has been reinitialized, the model states are updated using the analysis approach of *Evensen* [2003]. Under the ensemble formulation, random, zero-mean perturbations based on the assumed observation error are used to obtain an ensemble of observations $\mathbf{D}$ with perturbations $\boldsymbol{\Upsilon}$. The covariance of the perturbations is then described by $R = \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^T/(N-1)$. The Kalman Gain matrix, $\mathbf{K}$, which can be thought of as a blending factor to minimize the *a posteriori* error covariance of the filter [*Grewel and Andrews*, 1993], is then calculated using equation (5)

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^f \mathbf{H}^T + \mathbf{R}_k)^{-1}, \tag{5}$$

where $\mathbf{P}$ is the forecast covariance of the state ensemble, $\mathbf{H}$ is a measurement operator that maps the m observation locations onto the n model states, and $\mathbf{R}$ is the covariance of the observation noise ensemble. $\mathbf{K}$ is a $2n \times m$ matrix where the first n rows represent the model state (ψ) portion of $\mathbf{A}$ and the last n rows, the bias state (η) portion of $\mathbf{A}$. Finally, the state ensemble is updated using equation (6)

$$\mathbf{A}_{\psi,k}^a = \mathbf{A}_{\psi,k}^f + \mathbf{K}(\mathbf{D}_k - \mathbf{H}\mathbf{A}_{\psi,k}^f), \tag{6}$$

where $\mathbf{A}_k^f$ is the state forecast and $\mathbf{A}_k^a$, the updated or analyzed state. The updated covariance of $\mathbf{A}$ is then found from the forecast covariance using equation (7)

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^f. \tag{7}$$

4. Computational Experiment

[10] The tracer experiment was modeled in MODFLOW [*Harbaugh et al.*, 2000] and MT3DMS [*Zheng and Wang*, 1999] using a grid resolution of x = 35 cells (δ_x = 7.837-cm), y = 50 cells (δ_y = 7.874-cm), and z = 33 cells (δ_z ranging from 7.3-cm to 5.08-cm) for a total of 57,750 model cells. No flow boundary conditions were established at the bottom and sides of the tank and constant head boundary conditions of 203.2-cm at the inlet and 200.6-cm at the outlet reservoirs were established.

[11] A separate, non-point source ammonium chloride tracer test (1 g-L^{-1}) was used to calibrate MODFLOW and MT3DMS for the tank. Tracer breakthrough curves for the sampling ports in all media layers were fit simultaneously using PEST 10.2 [*Doherty*, 2006] by adjusting hydraulic conductivity (K), porosity, and dispersion to minimize the RMSE between observed and modeled transport. The calibrated K values were used as conditioning points for simulating multiple realizations of the K-field using SGS. This provides a distribution of hydraulic conductivity fields, in turn providing an ensemble of transport model realizations to be used within the ensemble framework of the filter.

[12] Since the media in the tank was layered during construction, the K-field of each media layer was conditionally simulated individually to eliminate SGS smoothing effects. The flow realizations were used in MT3DMS to model 100 realizations of the tracer transport over time steps of 8.1 minutes and two stress periods (3330 steps of tracer injection and 209 steps with no tracer) for a total simulation period of 20 days. Longitudinal dispersivities ranged from 6.44 cm in the course sand layer to 0.44 cm in the silt layer and the dispersivity ratios (relative to longitudinal) used in the transverse and vertical directions were 0.17. The diffusion coefficient was specified as 1×10^{-4}, and a Courant number of 0.9 was used to minimize numerical dispersion. An initial tracer concentration of zero was assumed throughout the domain.

[13] An ensemble size of 100 realizations was used for the EnKF. To enhance the computational efficiency of the EnKF, a filtering subdomain was defined at a courser resolution of 588 cells bounding the simulated regions with non-zero tracer concentration within the overall numerical model grid containing 57,750 cells. The filtering subdomain was defined to have at least one grid cell between every cell containing observation points in all three dimensions to support spatial evaluation of the filter. The time correlation factors, Λ for the bias forecast and reinitialization were set to 99-percent. The spatially correlated noise fields, $\mathbf{Q}_\eta$ and $\mathbf{Q}_\psi$ were generated for the filtering subdomain using a Gaussian covariance model with sill = 1.0 and range = 5 cells. These fields were then scaled using noise scaling factors (δ_b and δ_m) of 15-percent. The noise applied to the observations was normally distributed random noise with mean 0 and standard deviation 0.05 to represent approximately 5-percent error. The filter was then run for the full 19-day experiment with varying assimilation intervals. Filter forecasts were conducted at intervals of 2-hrs while observation assimilation was conducted at intervals of 2-hrs, 6-hrs, 12-hrs, and 24-hrs for four separate test cases. The 2-hr forecast, 2-hr assimilation is defined as the maximum data case, meaning that the filter was used to forecast the tracer, and then was immediately updated in the same time step with the observation data. For the 24-hr case, 12 forecasts occurred before the filter was updated using the observations at the 24th hour.

5. Results and Discussion

[14] Figure 2 shows that the bias-aware filtering is able to effectively capture complex tracer loading behavior, reduce errors due to the spatial distribution of hydraulic conductivities (i.e., the fine sand lens), and resolve advective gradients not captured in the original MODFLOW/MT3DMS simulation. It should be noted that complex contaminant source histories are rarely known and represent a very large source of model bias when simplified source area dynamics are assumed.

[15] Figure 2a shows the tracer breakthrough (BT) curves associated with the sampling ports in layer 4 plotted as normalized concentration versus time. The green line indicates the modeled BT curve, the blue dots represent observed concentration, and the ensemble means from the bias-aware filter assimilation cases are portrayed using a red color scale. Dark red lines represent the maximum data case (assimilation every 2-hrs) and the light pink lines represents the minimum data case (assimilation every 24-hrs). Without bias awareness, the model would require reinitialization after each assimilation step, resulting in costly Monte Carlo simulations of the groundwater flow and transport models.

[16] Figure 2a demonstrates that by modeling the bias, the filter can effectively correct for large systematic model errors. If this were not done, the filter would quickly track back to the modeled BT curves whenever an observation was not available. Figure 2b shows the BT curve associated with the tracer source location for each of the assimilation rates (2-hrs, 6-hrs, 12-hrs, and 24-hrs). The gray shading in these plots represents one standard deviation above and below the ensemble mean. Figure 2b shows that the complexity of source loading history is a significant source of error in the model. In addition, Figure 2b shows a clear increase in uncertainty as the time between data assimilation is increased from 2-hrs to 24-hrs. However, it is interesting to note that even for the minimal data case (assimilation every 24-hrs), the observed tracer concentration is always captured in the range of the bias aware filter's projected uncertainty ranges without the need to reinitialize and rerun the full Monte Carlo ensemble.

[17] The mean tracer concentration maps of the numerical model are shown in Figure 2c for t = 13-days. Four different views of the tracer plume are shown in Figure 2c: (i) an X-Y slice through the layer 4 sampling ports, (ii) an X-Y slice through the layer 3 sampling ports, (iii) a Y-Z slice (side view) through the center of the tank, and (iv) a 3-dimensional view of the tracer plume. The coloring of the maps is based on the normalized tracer concentration (C/C_o) where red indicates maximum concentration and blue indicates concentration below the detection limit. The coloring of the sampling ports indicates observed normalized tracer concentrations at this time step. Figure 2c shows that the numerical model fails to capture the true behavior of the plume because the sampling ports in layer 4 indicate low tracer concentration at the leading edge of the plume and the sampling ports in layer 3 indicate low tracer concentration at the source region of the plume. The impact of the fine-sand lens in the center of the tank results in significant systematic errors in the spatial and temporal contaminant BT behavior for the tracer. However, Figure 2d provides the mean filtered maps of the maximum data case (forecasts following by assimilation every 2 hours) at t = 13 days. The maps in Figure 2d indicate that bias-aware assimilation of observations is far more effective for capturing the experimental observations for the system. Not only are the observation locations appropriately corrected by the filter, but the spatially adjacent locations are improved as well. The effect of the fine sand lens is shown in these maps as a build-up of tracer developing at the source edge of the lens, and subsequent movement down and through the lens (as would be expected when hydraulic conductivity decreases at a boundary).

6. Conclusions

[18] The water cycle research community has highlighted that uncertainties in the subsurface [*Loescher et al.*, 2007] and frameworks for improving flux predictions are top research needs [*National Research Council*, 2008]. This work contributes a combination of laboratory tracer experiments and numerical modeling to clearly demonstrate that commonly occurring limitations in our knowledge of historical dynamics (source loading in this study) and subsurface flow properties yield severe sources of model bias, even for a well characterized sand-dominated laboratory system. Using a 19-day ammonium chloride tracer experiment, this study shows that filtering systematic model errors in groundwater transport predictions separately and allowing them to feed back to concentration predictions yields enhanced forecasting accuracy while eliminating the need to reinitialize and rerun forward model ensembles. Consequently, this work considerably expands the size and scope of groundwater flow-and-transport problems where the bias-aware EnKF can be applied. It would be expected that the need for and value of the bias-aware EnKF framework will only increase for more complex field environments. As our ability to predict multi-process flow-and-transport at watershed scales advances [*Kollet and Maxwell*, 2008], the bias-aware EnKF framework presented in this study can transform observation networks and forecasting frameworks to better account for limits in our knowledge of the subsurface.

[19] **Acknowledgments.** The authors of this work were partially supported by the U.S. NSF under CAREER grant CBET-0640443 and a Science to Achieve Results Graduate Research Fellowship (agreement FP-916820) awarded by the U.S. EPA. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not reflect the views of the NSF or the EPA.

References

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York.

Doherty, J. (2006), *PEST: Model-Independent Parameter Estimation*, 5th ed., Watermark Numer. Comput., Oxley, Queensland, Australia.

Drecourt, J.-P., H. Madsen, and D. Rosbjerg (2006), Bias aware Kalman filters: Comparison and improvements, *Adv. Water Resour.*, *29*, 707–718.

Eigbe, U., M. B. Beck, H. S. Wheater, and F. Hirano (1998), Kalman filtering in groundwater flow modelling: Problems and prospects, *Stochastic Hydrol. Hydraul.*, *12*(1), 15–32.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367.

Grewel, M. S., and A. P. Andrews (1993), *Kalman Filtering: Theory and Practice*, Prentice-Hall, Englewood Cliffs, N. J.

Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), MODFLOW-2000: The U. S. Geological Survey modular ground-water model—User guide to modularization concepts and the ground-water flow process, *U. S. Geol. Surv. Tech. Rep. Open File Rep., 00-92*.

Kollet, S. J., and R. M. Maxwell (2008), Demonstrating fractal scaling of baseflow residence time distributions using a fully-coupled groundwater and land surface model, *Geophys. Res. Lett.*, *35*, L07402, doi:10.1029/2008GL033215.

Loescher, H., J. J. O. Wendroth, D. Robinson, G. Poulos, K. McGuire, P. Reed, B. Mohanty, J. Shanley, and W. Krajewski (2007), Enhancing water cycle measurements for future hydrologic research, *Bull. Am. Meteorol. Soc.*, *88*, 669–676.

National Research Council (2004), *Groundwater Fluxes Across Interfaces*, Natl. Acad., Washington, D.C.

National Research Council (2008), *Integrating Multiscale Observations of U.S. Waters*, Natl. Acad., Washington, D. C.

Zheng, C., and P. P. Wang (1999), MT3DMS: A modular, three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide, *Tech. Rep. Contract Rep. SERDP-99-1*, Eng. Res. and Dev. Cent., U.S. Army Corp of Eng., Hanover, N. H.

J. B. Kollat and P. M. Reed, Department of Civil and Environmental Engineering, Pennsylvania State University, 406 Sackett Building, University Park, PA 16802-1408, USA. (juk124@psu.edu; preed@engr.psu.edu)

D. M. Rizzo, Department of Civil and Environmental Engineering, University of Vermont, 219 Votey Hall, Burlington, VT 05405-0156, USA. (drizzo@cems.uvm.edu)