

## Assimilating water table elevation data into a catchment hydrology modeling framework to estimate hydraulic conductivity

Ryan T. Bailey<sup>1</sup> and Domenico A. Baù

Department of Civil and Environmental Engineering, Colorado State University

**Abstract.** Groundwater models, often used to enhance understanding of hydrologic and chemical processes in local or regional aquifers, are often hindered by inadequate representation of the parameters which characterize these processes. Furthermore, attempts to estimate these parameters are usually limited to studies employing simplified aquifer representations. In this study we present preliminary results of using a data assimilation algorithm, the Ensemble Smoother, to provide enhanced estimates of aquifer hydraulic conductivity within a fully-coupled, surface-subsurface flow modeling framework through assimilation of water table elevation measurements. Based on the Kalman Filter methodology, the algorithm uses residuals between forecasted model results and assimilated measurement data, together with the covariance of model results, to correct model results throughout the model domain. Parameter estimation is achieved by incorporating spatially-variable hydraulic conductivity values into the algorithm, thereby allowing the correlation between water table values and hydraulic conductivity to correct the hydraulic conductivity fields. The applicability of the Ensemble Smoother scheme is demonstrated via a synthetic three-dimensional catchment system incorporating variably-saturated subsurface flow, overland flow, and channel flow. Results indicate that assimilating water table measurements provides an improved estimate of the hydraulic conductivity fields.

### 1. Introduction

Deterministic, numerical hydrologic models, due to mathematical approximation of physical processes and an insufficient understanding of system parameters, are not capable of fully simulating the system they are designed to represent (Van Geer et al., 1991). As such, hydrologic model predictions are often fraught with uncertainty. In an attempt to (1) reduce uncertainty in both model response and parameter variables, and hence bring these variables into accordance with the actual system, and (2) quantify the uncertainty of the system, data assimilation (DA) techniques have been successfully used in hydrologic modeling (Liu and Gupta, 2007). Among the available DA schemes, the Kalman Filter (KF) (Kalman, 1960), designed for systems of linear dynamics, has been used extensively in physically-based modeling studies to assimilate real-world measurement data into model results to provide optimal estimates of system variables.

Following a Bayesian framework, the KF is a statistical routine in which prior information (i.e., numerical model results and knowledge of the system parameters) is merged with information from the actual system (i.e., measurement data) to produce a corrected, posterior system estimate. Limitations of the KF scheme, namely the restriction

---

<sup>1</sup> Department of Civil and Environmental Engineering  
Colorado State University  
Fort Collins, CO 80523-1372  
Tel: (970) 491-5387  
e-mail: rtbailey@engr.colostate.edu

to linear dynamic systems and the requirement to represent model and measurement error statistics with fully-defined probability-density functions (PDF), has led to ensemble methods that use an ensemble of Monte Carlo model realizations to represent uncertainty. Using the results of the realizations, the ensemble mean and standard deviation of model variables at each location throughout the model domain are used to approximate the PDF of the model error statistics, and assimilated measurements are perturbed with a Gaussian error to approximate the PDF of the measurement error statistics. The Ensemble Kalman Filter (EnKF) (Evensen, 1994) was first used in meteorological and oceanographic modeling (Evensen and van Leeuwen, 1996; Houtekamer and Mitchell, 1998), and has since been applied to land-surface and groundwater modeling studies (e.g. Reichle et al., 2002; Das and Mohanty, 2006; Camporese et al., 2009). Typical applications of these studies are the estimation of system response variables such as soil moisture, hydraulic head, and stream flow. Other studies have focused on using system response measurements to condition model parameters, such as hydraulic conductivity (Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008). Chen and Zhang (2006) used 2D and 3D saturated groundwater flow simulations, and Hendricks Franssen and Kinzelbach (2008) used a 2D saturated groundwater flow simulation.

The Ensemble Smoother (ES) (van Leeuwen and Evensen, 1996) is another ensemble scheme which has been used in hydrologic modeling studies (Dunne and Entakhabi, 2005), although to a much lesser extent than the EnKF. In contrast to the basic filtering methods, which uses all available measurements to provide an updated model state estimate at only the current time, the smoother incorporates all previous measurements and model states to provide an updated model state at all previous times. This allows all previous model states to be updated each time measurements are collected and assimilated.

In this study, the ability of the ES algorithm to accurately estimate system parameters using system response measurements is explored using a synthetic coupled surface-subsurface flow simulation. Specifically, water table elevations are assimilated into model results to provide a corrected estimate of the water table field as well as condition the hydraulic conductivity field using measurements from one or more collection times.

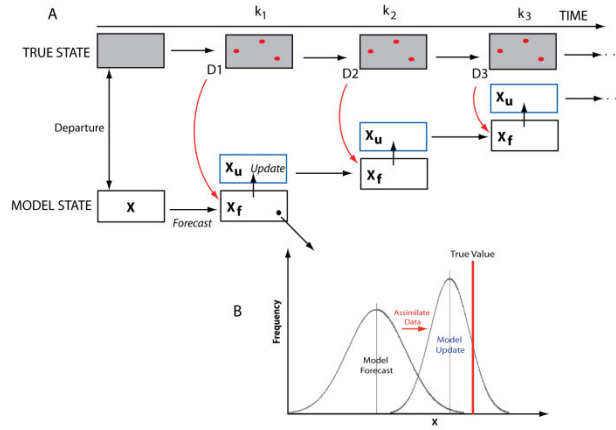
## 2. Theory of the Kalman Filter Data Assimilation Algorithm

The basic KF algorithm follows the sequential *forecast-update* cycle, with update of the system occurring whenever measurements are available (Figure 1). In the *forecast* step, the model state  $X$  is run forward in time based on model formulation, parameters  $P$ , forcing terms  $q$ , boundary conditions  $b$ , model error  $w$  described by a Gaussian (PDF), and solution to the mathematical model  $\Phi$ , generating the prior system information  $X_{k+1}^f$ , where the  $f$  superscript represents *forecast*:

$$X_{k+1}^f = \Phi(X_k; P; q; b) + w_k \quad (1)$$

In the update step, measurement data  $y_{k+1}$  (represented by red circles in Figure 1) are collected from the actual system at time  $k+1$ , perturbed with a Gaussian error  $v$  to create the measurement vector  $D_{k+1}$ , and assimilated into the model forecast results to generate a posterior, corrected state estimate,  $X_{k+1}^u$ , to provide a state estimate that approaches the true state from which the measurements were collected:

$$X_{k+1}^u = X_{k+1}^f + K_{k+1}(D_{k+1} - HX_{k+1}^f) \quad (2)$$



**Figure 1.** DA framework showing the sequential update of the model state using measurements. Measurements are represented by red circles. The updated system state  $X_u$  is brought closer to the true state as more and more measurements are assimilated.

The matrix  $H$  maps model results to model locations at which measurements are taken, so that the matrix  $HX_f^{k+1}$  holds the model results at measurement locations. As such, the difference  $D_{k+1} - HX_f^{k+1}$  provides the residual (i.e. the error) between the model-predicted results and the measurements. The model-forecasted state  $X_f^{k+1}$  is then corrected by this residual, which receives a weight via the Kalman Gain matrix,  $K$ , defined by:

$$K = C^f H^T (H C^f H^T + R)^{-1} \quad (3)$$

where  $C^f$  is the error covariance matrix associated with the model forecast  $X_f^{k+1}$  and  $R$  is the measurement error covariance matrix associated with the perturbed measurements  $D$ . The formulation of  $K$  (1) allows spreading of measurement information throughout the model domain according to spatial correlation of model results, and (2) acts as a weighting term that scales correction terms according to model ( $C^f$ ) and measurement ( $R$ ) error. If the model error is large and the measurement error is low, then the elements of  $K$  approach a value of 1 and the full residual is used to correct the forecast state, i.e. the measurement information is used in its entirety to correct the model results. In contrast, high measurement error relative to model error allows only a portion of the residual to correct the forecast state.

In the ensemble schemes, the system state is represented by an ensemble of Monte Carlo simulations, each of which is forecasted with uncertain parameters according to Equation (1). When measurements are collected, they are perturbed and assimilated into the model results using Equation (2), with  $C^f$  assembled using the deviation of each nodal value from the average nodal value at each model node (Evensen, 2007).

### 3. Parameter Estimation Using the Ensemble Smoother

To incorporate parameter estimation into the algorithm, the state matrix  $X$  is augmented to include uncertain model parameter values, e.g. hydraulic conductivity in groundwater flow modeling studies, allowing the spatial correlation between parameter and state variables to correct both the state and parameter values. In this study, water table

elevations (WTE) and hydraulic conductivity (K) values are updated using WTE measurements. In the ES format,  $\mathbf{X}_k^f$  is comprised of both WTE and K variables, from time  $l$  to  $k$ :

$$\mathbf{X}_k^f = [\mathbf{X}_{(WTE)1}, \dots, \mathbf{X}_{(WTE)k}; \mathbf{X}_{(K)}] \quad [(n * k) + e] \times nmc \quad (4)$$

where  $n$  is the number of model nodes,  $e$  is the number of parameters that characterize the system, and  $nmc$  is the number of Monte Carlo simulations. K values are only added once to the state matrix since they are assumed to be time-independent. WTE measurement data is used to assemble the matrix  $\mathbf{D}$ , which holds all measurement data from time  $l$  to  $k$ :

$$\mathbf{D} = [\mathbf{D}_{(WTE)1}, \dots, \mathbf{D}_{(WTE)k}] \quad [(m * k) + e] \times nmc \quad (5)$$

where  $m$  is the number of measurements collected at a given time. In a surface-subsurface modeling framework, an ensemble of random K fields are generated using a sequential Gaussian algorithm, with geostatistical parameters mean ( $\mu$ ), variance ( $\sigma^2$ ), and correlation length ( $\lambda$ ). Boundary conditions, initial conditions, and forcing terms are also defined. An ensemble of simulations, with each realization using a different K field, is then run forward in time, representing the forecast step depicted by Equation (1). In the present proof-of-concept study, an additional K field and associated flow simulation, from which measurements can be taken, provide a “true” state against which the updated fields can be compared.

The update step consists of populating  $\mathbf{X}$  with the ensemble of WTE and K values, taking measurements from the “true” state, and running the ES update routine to provide updated WTE and K ensembles. Measurement coefficient of variation is applied to measurements to incorporate measurement error. The performance of the routine is analyzed by comparing the updated model state to the “true” state via the following two performance parameters (Hendricks Franssen and Kinzelbach, 2008):

$$AE(\mathbf{X}) = \frac{1}{nmc * n} \sum_{j=1}^{nmc} \sum_{i=1}^n |X_{i,j} - X_{i,true}| \quad (6)$$

$$AESD(\mathbf{X}) = \frac{1}{nmc * n} \sum_{j=1}^{nmc} \sum_{i=1}^n |X_{i,j} - \bar{X}_i| \quad (7)$$

The absolute error term (AE) compares the model values to the “true” value at each location in the model domain, and the average ensemble standard deviation (AESD) compares the model values to the ensemble mean at each location, providing a measure of the spread of the values. Hence, AE determines if the updated state approaches the “true” state, and AESD determines the confidence in this estimate.

#### **4. Model Forecast: Hydrologic Flow Simulations**

The model used to provide a forecast of the surface-subsurface flow scenario is the CATHY (CATchment HYdrology) model (Bixio et al., 2000; Camporese et al., 2009; Camporese et al., 2010), which couples three-dimensional variably saturated subsurface flow with one-dimensional diffusion wave overland and channel flow. Equations defining these processes can be found in Camporese et al. (2009). The model computational domain consists of (1) digital elevation model (DEM) cells for the ground surface, with which the one-dimensional surface flow equation is solved; and (2) a three-dimensional tetrahedral mesh for the underlying aquifer, which is created by triangulating the DEM cells and replicating them in the vertical direction downward to the prescribed aquifer base. The surface flow equation is solved using the Muskingum-Cunge diffusion wave approximation (Orlandini and Rosso, 1996), and the subsurface flow equation is solved using the finite element method (Paniconi and Putti, 1994). Unsaturated flow conditions are modeled using the van Gneuchten parameters for the water retention curve.

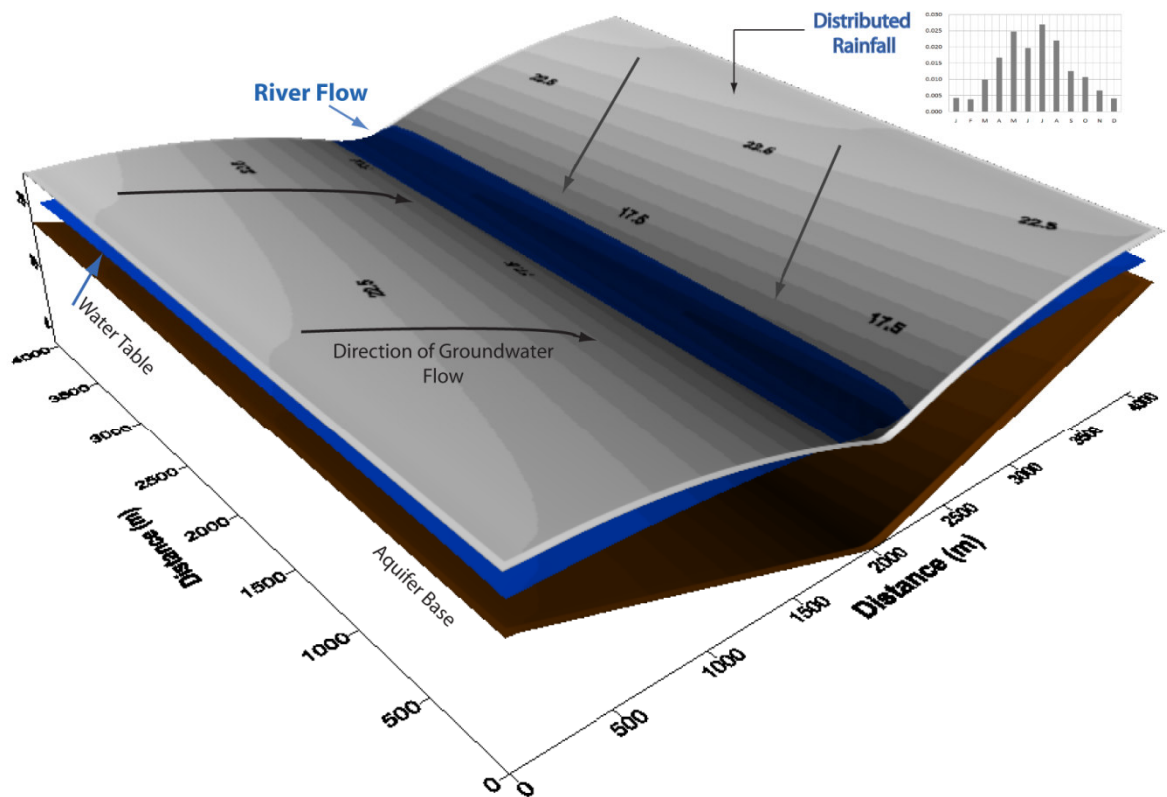
The flow problem used in this study consists of a 4 km by 4 km v-shaped catchment that receives a monthly time series of precipitation and distributes it between overland flow, channel flow, and variably saturated groundwater flow over a one-year period, with channel flow occurring along the central depression. Aquifer thickness varies between 7.5 and 15.5 m, discretized by 10 layers of varying thickness, with the thickest portion of the aquifer placed under the central depression. The aquifer was assigned a porosity of 0.35 and a specific storage of  $0.01 \text{ m}^{-1}$ . Eighty-one DEM cells were defined in both the x and y directions, resulting in 6561 surface cells, 6724 surface mesh nodes, 73964 total mesh nodes, and 393660 tetrahedral finite elements. Figure 2 shows results of a generic simulation with a homogeneous aquifer and a steady precipitation rate, run to steady groundwater and channel flow conditions, showing the aquifer base, the water table, the ground surface elevation, and the flowing channel.

An ensemble of 85 log-normal hydraulic conductivity fields was generated using a sequential Gaussian algorithm, SKSIM (Baù and Mayer, 2008), with mean of  $-4.30$  ( $\log \text{ m sec}^{-1}$ ), variance of  $0.434$  ( $\log \text{ m sec}^{-1}$ )<sup>2</sup>, and correlation length of 1000 m. This resulted in K values ranging from  $0.055 \text{ m day}^{-1}$  to  $216 \text{ m day}^{-1}$ . Using these K fields, an ensemble of 85 CATHY simulations were run using the steady-state conditions presented in Figure 2. Each simulation was run for 181 days to eliminate the bias of the initial conditions, and then run for 365 additional days using a monthly recharge series, with the water table elevation at every location in the aquifer outputted at monthly simulation times. An additional K field and CATHY simulation was run to provide a “true” state against which ES-updated states could be compared. These “true” reference states, along with forecast results, will be presented in the next section.

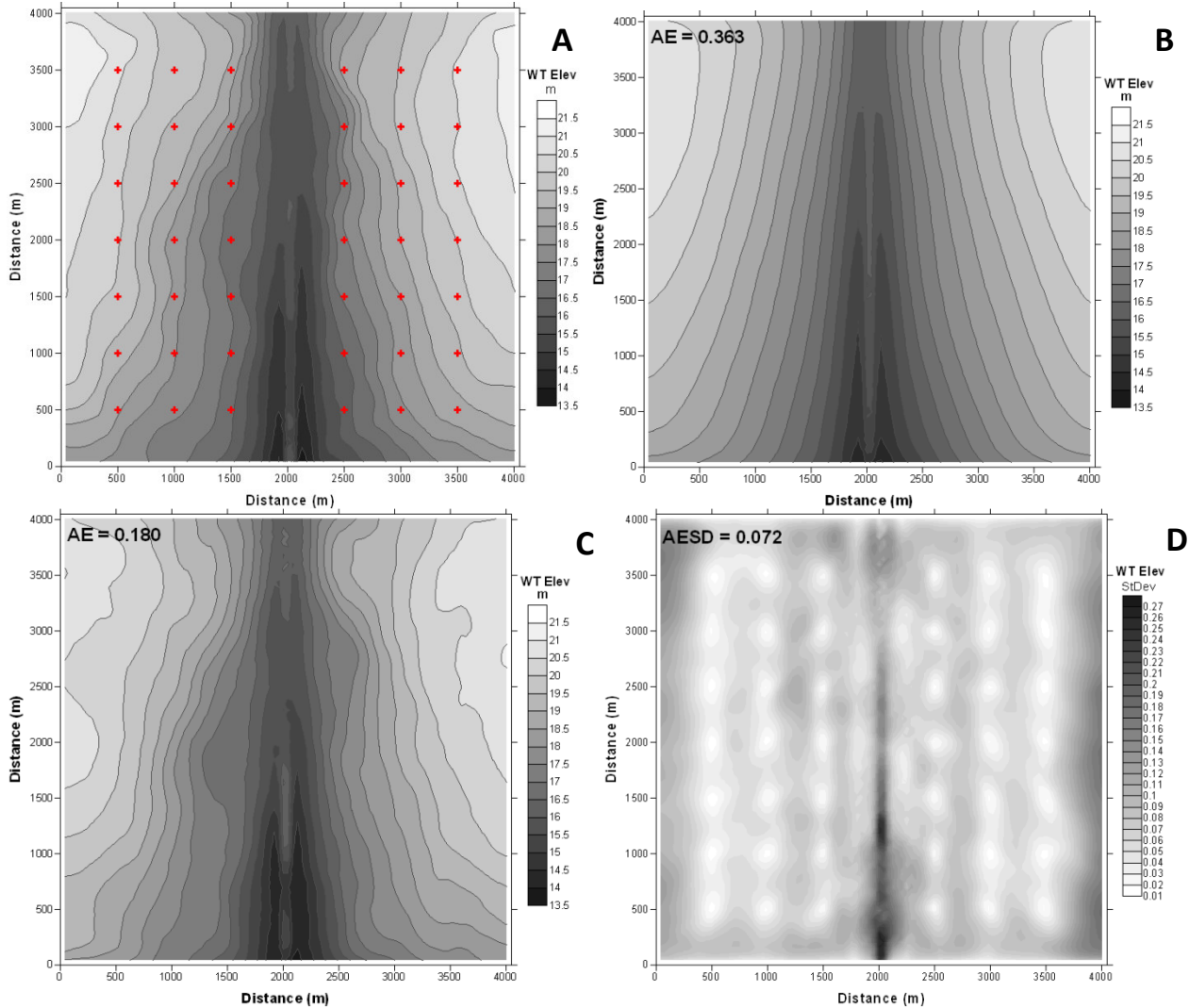
#### **5. Model Update: Water Table and Hydraulic Conductivity Estimation**

In the first measurement scenario, 42 WTE measurements at simulation times of 59, 120, 181, 243, 304, and 365 days were collected and assimilated into the model results to provide updated model states at these times. The WTE measurement data was assigned a coefficient of variation of 0.0, representing negligible measurement error. Figure 3 shows the “true” water table field at time = 365 days, with the red cross-marks depicting locations where measurements were collected (Figure 3A), the forecasted average value of the WTE at each location (ensemble mean) at time = 365 days (Figure 3B), and the updated

ensemble mean of the WTE (Figure 3C). The updated water table field closely resembles the “true” water table field, with the AE performance parameter reduced from 0.363 for the forecast state to 0.180 for the update state, a reduction of 50.4%. The standard deviation of the ensemble values at each model location is shown in Figure 3D. Notice that the standard deviation approaches 0.0 near the measurement locations, due to the measurements assigned a coefficient of variation of 0.0.

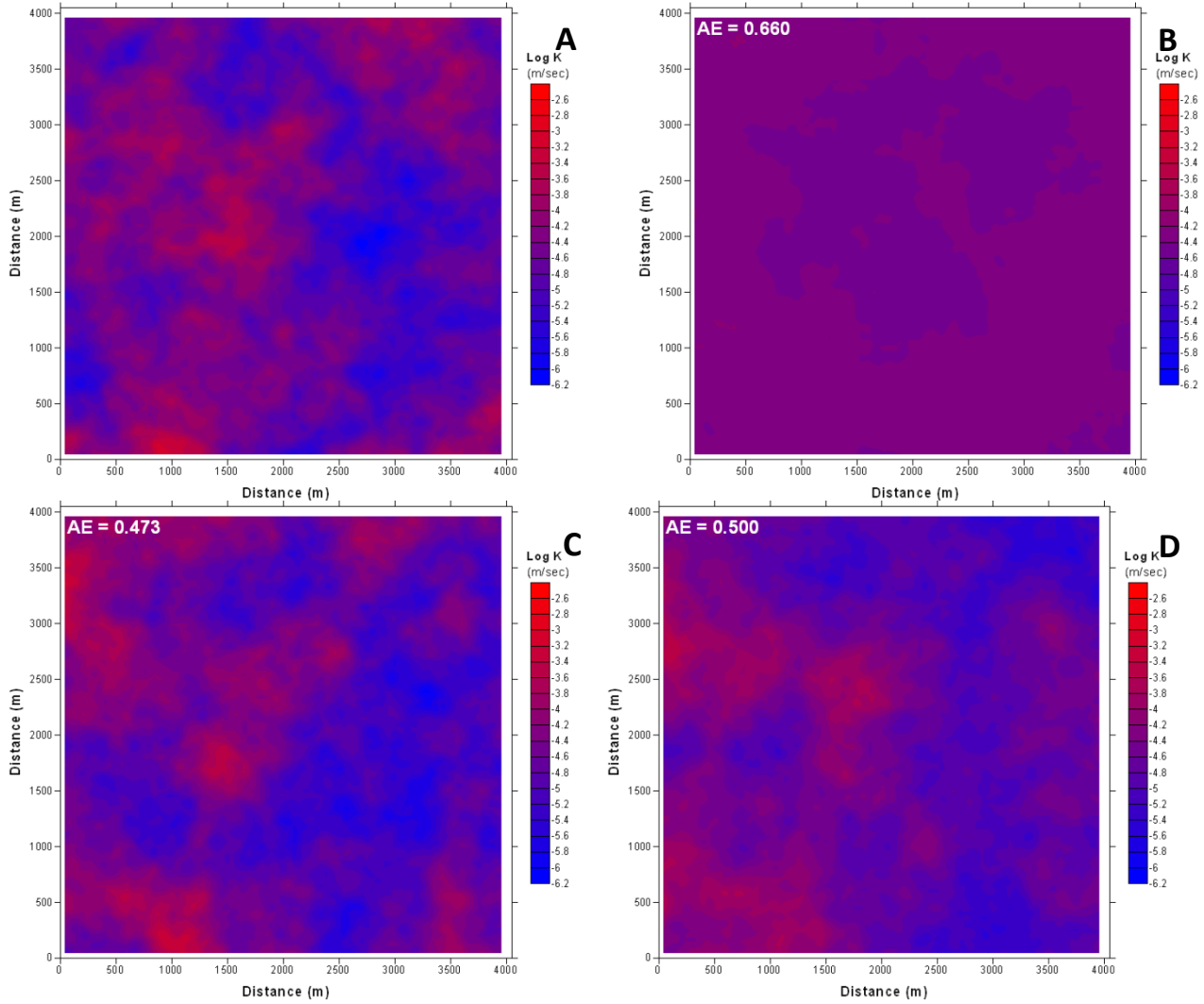


**Figure 2.** Ground surface elevation, aquifer base, and steady-state water table and channel flow delineation, showing direction of groundwater flow in catchment and the rainfall series used in the transient simulation. The catchment is 4 km by 4 km.



**Figure 3.** (A) “True” water table field at time = 365 days, with the red cross-marks depicting locations from which measurements are taken and assimilated; (B) Forecast ensemble mean at time = 365 days, (C) Update ensemble mean of water table elevation at every node in the model domain, at time = 365 days, and (D) Standard deviation of the updated water table elevation at every node in the model domain, at time = 365 days. Notice that the standard deviation approaches 0.0 near the measurement locations.

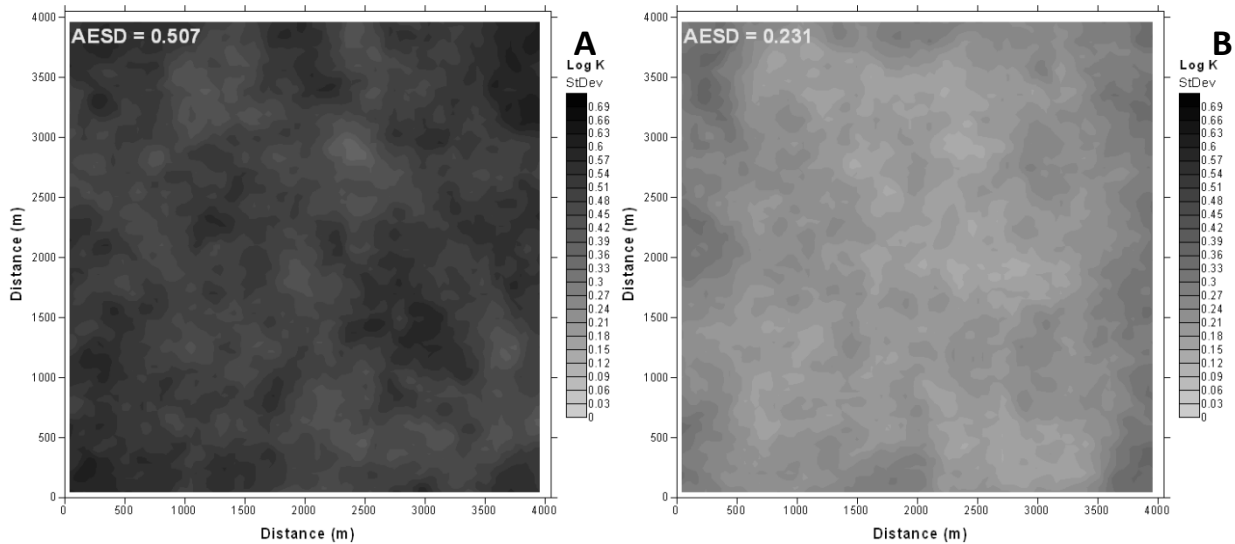
The K field ensemble was also included in the system state, allowing the K values to be conditioned by the WTE measurements. Figure 4 shows the “true” K field (Figure 4A), the forecasted ensemble mean of the K values at each model location (Figure 4B), and the updated ensemble mean of the K values (Figure 4C) when 42 WTE measurements are assimilated. The updated K ensemble seems to match quite reasonably the spatial K distribution of the “true” state. The AE was reduced from 0.660 for the forecast K ensemble to 0.473 for the updated K ensemble, a reduction of 28.3%. Figure 4D shows the updated K ensemble mean when only 21 WTE measurements are assimilated. The overall trend of the K distribution approaches the “true” state in Figure 4A, although not to the same extent as when 42 WTE measurements are assimilated. With 21 WTE measurements assimilated, the AE value for the K ensemble is 0.500, a reduction of 24.2% from the forecast value of 0.660.



**Figure 4.** (A) “True” K field; (B) Forecasted K ensemble mean; (C) Updated K ensemble mean when 42 WTE measurements are assimilated, and (D) Updated K ensemble mean when 21 WTE measurements are assimilated.



The standard deviation of the K ensemble values at each model location is shown in Figure 5, with the updated ensemble having an AESD value reduced by 54.4% from the forecasted value (0.231 compared to 0.507).



**Figure 5.** Standard deviation of the (A) forecasted hydraulic conductivity ensemble and (B) updated hydraulic conductivity ensemble at every element in the model domain. The AESD value is decreased from 0.507 to 0.231.

## 6. Conclusions

The Ensemble Smoother (ES), a statistical data assimilation routine that merges uncertain, model-produced values with measurement data, was implemented and evaluated in its ability to condition hydraulic conductivity (K) fields using water table elevation (WTE) measurement data. Preliminary results demonstrate that, within the confines of a statistically homogeneous, randomly-generated K field ensemble, the ES scheme provides an updated K ensemble that approaches the state from which the WTE measurement data was collected. These results confirm results from earlier parameter estimation studies using the Ensemble Kalman Filter (EnKF), and extended its application to a more realistic framework by using the hydrologic model CATHY, a fully coupled surface-subsurface flow model. Further research will include an investigation into the influence of the number of WTE measurements, error assigned to assimilated WTE measurement data, the number of assimilation times, the correlation length used in generating the K fields, and using groundwater return flows to the stream as another system response variable to condition the K fields.

### Acknowledgements

The majority of this work has been made possible by a Colorado Agricultural Experiment Station (CAES) grant (Project No. COL00690).

## 7. References

- Baù, D.A., and A.S. Mayer (2008), Optimal design of pump-and-treat systems under uncertain hydraulic conductivity and plume distribution, *Jour. Cont. Hydrol.*, **100**, 30-46.
- Bixio, A.C., S. Orlandini, C. Paniconi, and M. Putti (2000), Physically-based distributed model for coupled surface runoff and subsurface flow simulation at the catchment scale, in *Computational Methods in*

- Water Resources*, vol. 2, *Computational Methods, Surface Water systems and Hydrology*, edited by L. R. Bentley et al., pp. 1115-1122, Balkema, Rotterdam, Netherlands.
- Camporese, M., C. Paniconi, M. Putti, and P. Salandin (2009), Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow, *Water Resour. Res.*, **45**, W10421, doi:10.1029/2008WR007031.
- Camporese, M., C. Paniconi, M. Putti, and S. Orlandini (2010), Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data, *Water Resour. Res.*, **46**, W02512, doi:10.1029/2008WR007536.
- Chen, U., and D. Zhang (2006), Data assimilation for transient flow in geologic formations via Ensemble Kalman Filter. *Adv. Water Resour.*, **29**(8), 1107-1122.
- Dunne, S. and D. Entekhabi (2005), An ensemble-based reanalysis approach to land data assimilation. *Water Resour. Res.*, **41**, W02013, doi:10.1029/2004WR003449.
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10, 143-10,162.
- Evensen, G., and P.J. van Leeuwen (1996), Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Weather Rev.*, **124**, 85-96.
- Evensen, G. Data assimilation. *The Ensemble Kalman Filter*. Springer-Verlag Berlin Heidelberg, 2007.
- Hendricks Franssen, H. J., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resour. Res.*, **44**, W09408, doi:10.1029/2007WR006505.
- Houtekamer, P.L., and H.L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.*, **126**, 796-811.
- Liu, Y., and H.V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, **43**, W07401, doi:10.1029/2006WR005756.
- Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35-45.
- Orlandini, S., and R. Rosso (1996), Diffusion wave modeling of distributed catchment dynamics, *J. Hydraul. Eng. ASCE*, **1**(3), 103-113.
- Paniconi, C., and M. Putti (1994), A comparison of Picard and Newton iteration in the numerical solution of multidimensional variably saturated flow problems, *Water Resour. Res.*, **30**(12), 3357-3374.
- Reichle, R.H., D.B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather Rev.*, **130**, 103-114.
- Van Geer, F.C., C.B.M. Te Stroet, and Y. Zhou (1991). Using Kalman filtering to improve and quantify the uncertainty of numerical groundwater simulations, 1, The role of system noise and its calibration, *Water Resour. Res.*, **27**(8), 1987-1994.
- van Leeuwen, P.J., and G. Evensen (1996), Data assimilation and inverse methods in terms of probabilistic formulation. *Mon. Weather Rev.*, **124**, 2898-2913.