XVIII International Conference on Water Resources
CMWR 2010
J. Carrera (Ed)
CIMNE, Barcelona 2010

INVERSE MODELLING OF HYDRAULIC CONDUCTIVITY DISTRIBUTION BY ASSIMILATION OF RETURN FLOW DATA

Ryan T. Bailey*, Domenico A. Baù*

* Department of Civil and Environmental Engineering
Colorado State University, 1372 Campus Delivery
Fort Collins, CO, 80523-1372, United States
e-mails: rtbailey@engr.colostate.edu, domenico.bau@colostate.edu

Key words: Data assimilation, model uncertainty, groundwater

Summary. This paper demonstrates the use of an Ensemble Smoother, based on the Kalman Filter methodology, to estimate hydraulic conductivity distribution through assimilation of groundwater return flow measurements into groundwater model simulation results.

1 INTRODUCTION AND THEORY

Deterministic, numerical models, such as groundwater flow models are not capable of fully simulating the response of the system they are designed to represent, due to approximation of physical processes and inadequate knowledge of system parameters[1]. In an attempt to address these inadequacies, data assimilation (DA) techniques have frequently been used to reduce uncertainty attached to both state and parameter estimation[2]. The Kalman Filter[3] (KF), designed for linear dynamics, has been used extensively in physically-based modeling studies to assimilate real-world measurement data into model results and provide optimal estimates of state and parameter variables.

Following a standard Bayesian framework, the KF is a statistical routine in which prior information (i.e., numerical model results) is merged with information from the actual system (i.e., measurement data) to produce a corrected, posterior system estimate. The algorithm follows the sequential *forecast-update* cycle, with update of the system occurring whenever measurements are available. In the *forecast* step, the model state $\boldsymbol{X}$ is run forward in time based on model formulation, parameters $\boldsymbol{P}$, forcing terms $\boldsymbol{q}$, boundary conditions $\boldsymbol{b}$, model error $\boldsymbol{w}$ described by a Gaussian probability density function (PDF), and solution to the mathematical model $\boldsymbol{\Phi}$, generating the prior system information $\boldsymbol{X}^f_{k+1}$, where the *f* superscript represents *forecast*:

$$\boldsymbol{X}^f{}_{k+1} = \boldsymbol{\Phi}(\boldsymbol{X}_k; \boldsymbol{P}; \boldsymbol{q}; \boldsymbol{b}) + \boldsymbol{w}_k \tag{1}$$

In the update step, measurement data $\boldsymbol{y}_{k+1}$ are collected from the actual system at time $k+1$, perturbed with a Gaussian error $\boldsymbol{v}$ to create the measurement vector $\boldsymbol{D}_{k+1}$, and assimilated into the model forecast results to generate a posterior state estimate, $\boldsymbol{X}^u_{k+1}$:

$$\boldsymbol{X}^u{}_{k+1} = \boldsymbol{X}^f{}_{k+1} + \boldsymbol{K}_{k+1}(\boldsymbol{D}_{k+1} - \boldsymbol{H}\boldsymbol{X}^f{}_{k+1}) \tag{2}$$

The matrix $\boldsymbol{H}$ maps model results at measurement locations to actual measurement values, creating a residual of the variable in question. The Kalman Gain matrix, $\boldsymbol{K}$, is given by

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

where $\boldsymbol{C}^f$ is the error covariance matrix associated with the model forecast $\boldsymbol{X}_f^{k+1}$ and $\boldsymbol{R}$ is the measurement error covariance matrix associated with the perturbed measurements $\boldsymbol{D}$. The formulation of $\boldsymbol{K}$ i) allows spreading of measurement information throughout the model domain according to spatial correlation of model results, and ii) acts as a weighting term that scales correction terms according to model ($\boldsymbol{C}^f$) and measurement ($\boldsymbol{R}$) error.

Limitations of the KF scheme, namely the restriction to linear dynamic models and the requirement to represent error statistics with fully-defined probability-density functions (PDF), has led to ensemble schemes, i.e. the Ensemble Kalman Filter[4] (EnKF) and the Ensemble Smoother[5] (ES), which use an ensemble of Monte Carlo model realizations to approximate the PDF of the model and measurement error statistics. Whereas the EnKF provides an updated model state given all previous measurement data, the ES scheme incorporates all previous model states and measurement data into the update routine, allowing previous model states to be corrected with the acquisition of new data. These methods have been used extensively in hydrologic modeling to quantify and decrease uncertainty of model results[6,7,8] as well as estimating uncertain system parameters[9,10,11]. In the latter, a typical objective is to estimate hydraulic conductivity distribution through assimilation of hydraulic head measurements.

In this paper, the ability of the ES to accurately estimate system parameters using system response measurements is explored using a synthetic 2D transient groundwater flow simulation. Specifically, groundwater return flow volumes (RFV) to a stream are used to condition the hydraulic conductivity (K) field using measurements from one or more simulation times. Sensitivity analyses are carried out to gain insights into the influence of measurement error, the number of stream gage locations, the number of assimilation times, and the correlation length of the K fields. For real stream-aquifer systems, fluxes to the stream from groundwater could be calculated as long as a water balance for a given reach of the stream is conducted.

2 PARAMETER ESTIMATION USING THE ENSEMBLE SMOOTHER

To achieve parameter estimation, the state matrix $\boldsymbol{X}$ is augmented to include uncertain model parameter values, allowing the spatial correlation between parameter and state variables to correct both the state and parameter values. In this work, return flow volumes (RFV) and hydraulic conductivity (K) values are updated using RFV measurements. In the ES format, $\boldsymbol{X}^f_k$ is comprised of both RFV and K variables, from time *1* to *k*:

$$\boldsymbol{X}_k^f = \left[\boldsymbol{X}_{(RFV)1}, \dots, \boldsymbol{X}_{(RFV)k}; \boldsymbol{X}_{(K)}\right] \qquad [(n * k) + e] \text{ x } nmc \tag{4}$$

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. Initially, only RFV measurements are used to condition the ensemble of K fields, although K measurement data can also be added to $\boldsymbol{D}$ for further conditioning:

$$\boldsymbol{D} = \left[\boldsymbol{D}_{(RFV)1}, \ldots, \boldsymbol{D}_{(RFV)k}; \boldsymbol{D}_{(K)}\right] \qquad [(m * k) + e] \text{ x } nmc \tag{5}$$

where *m* is the number of measurements collected at a given time. In a groundwater modeling framework, the forecast step consists of running the simulations. An ensemble of groundwater flow simulations is initialized with an ensemble of K fields and initial hydraulic head fields. The K fields are generated using a sequential Gaussian algorithm, called SKSIM[12] with geostatistical parameters mean (μ), variance (σ^2), and correlation length (λ). Boundary conditions and forcing terms are applied throughout the simulation. An additional K field and associated flow simulation, from which measurements can be taken, provide a “true” state against which the updated K fields can be compared.

The update step consists of populating $\boldsymbol{X}$ with the ensemble of RFV and K values, taking measurements from the “true” state, and running the ES update routine to provide an updated model state. 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[11]:

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

$$AES(\boldsymbol{X}) = \frac{1}{nmc * n} \sum_{j=1}^{nmc} \sum_{i=1}^{n} \left| X_{i,j} - \bar{X}_i \right| \tag{7}$$

The absolute error term (AE) compares the model values to the “true” value at each location in the model domain, and the average ensemble spread (AES) compares the model values to the ensemble mean at each location, providing a measure of the spread of the values.

3 GROUNDWATER FLOW SIMULATIONS AND PARAMETER ESTIMATION

3.1 Forecast

The 2D transient groundwater flow problem consists of an areal aquifer 2000 m west-east by 4000 m north-south (Figure 1), solved using the finite-element code SAT2D[13]. An initial ensemble of 100 log-normal K fields was generated with SKSIM[12] using an exponential correlation model and with mean of -4.30 (log m sec^{-1}), variance of 0.434 (log m $sec^{-1})^2$, and correlation length of 300 m. Three other K ensembles, using correlation lengths of 500 m, 1000 m, and 1500 m, were also created to study the influence of correlation length. The triangular-element mesh consists of 3321 nodes and 6400 elements, with each block of 2 triangular elements assigned a different K value.

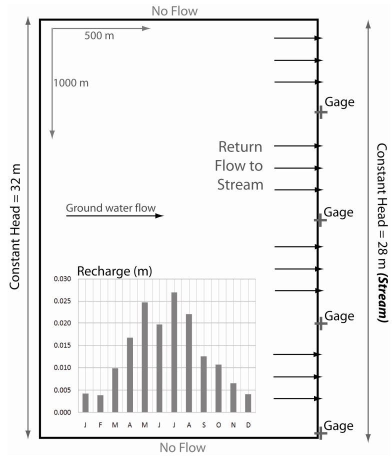

Figure 1: Plan view of 2D conceptual model, showing representation of return flows and recharge series.

Average aquifer saturated thickness and specific yield were 30 m and 0.20, respectively, for all simulations, and constant-head boundaries of 32 m and 28 m were placed on the west and east ends of the aquifer. Head fields produced by steady-state simulations were used as initial conditions for the transient problem, which consisted of a 365-day simulation using a time step of 1 day. The constant-head boundary on the east side of the aquifer was treated as a stream, with the flows leaving the model domain along this boundary treated as return flows to the stream. Return flow volumes were calculated by summing flows between designated stream gage locations between two moments in time (Figure 1). An additional K field and flow simulation were created to provide "true" fields from which measurements were collected and against which update ensemble could be compared. AE and AES for the ensemble of K fields are 0.482 and 0.346, respectively.

3.2 Update using RFV Measurements

Conditioning of K fields using RFV measurements was undertaken for various measurement times, stream gage locations, measurement error, and K correlation length. The number of assimilation times ranged from 1 (measurements taken only at 365 days) to 52 (weekly measurements); the number of gage locations ranged from 1 (gage located at south end of stream) to 20 (gages located every 200 m); measurement coefficient of variation ranged from 0.00 to 3.00; and correlation lengths used were 300 m, 500 m, 1000 m, and 5000 m. Assimilating RFV measurements once a year produced AE and AES values of the K ensemble of 0.384 and 0.286, respectively, an improvement of 20.7% and 17.7%, respectively, from the forecast values of 0.482 and 0.346. Increasing the number of assimilation times only slightly improves the AE and AES terms (Figure 2A). Using 1 stream gage and assimilating measurements bi-weekly produced AE and AES values of the K ensemble of 0.424 and 0.328, and improvement of 12.4% and 5.5% from the forecast values. These values are greatly improved when 4 gages are used (Figure 2B), with a reduction of 25.0% and 25.1% from the forecast values. Minor improvement is made by using 20 gages instead of 4 (Figure 2B).

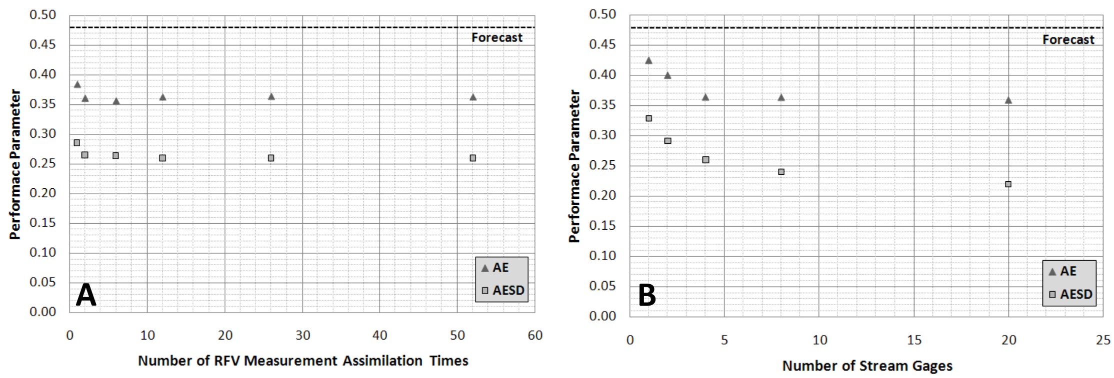

Figure 2: Effect of (A) the number of assimilation times and (B) the number of stream gage locations on the update K ensemble.

The correlation length used in creating the initial K ensemble dramatically influences the K update, with AE and AES improvement of only 7.9% and 14.0% when a length of 300 m is used, opposed to an improvement of 29.1% and 28.9% when a length of 1500 m is used. Comparisons of the K "true" state with the K update ensemble mean and update ensemble standard deviation for a correlation length of 1500 m (Figure 4) provides a much stronger conditioning of K than for the scenario using a length of 300 m (Figure 3).

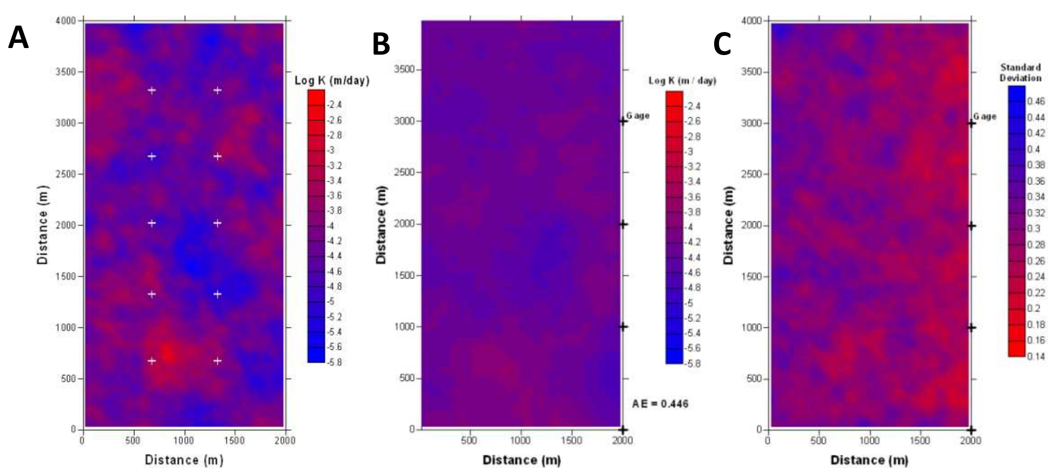

Figure 3: (A) K "true" state, (B) K update ensemble mean, and (C) K update ensemble standard deviation, conditioned by bi-weekly RFV measurements at 4 gaging locations, using a correlation length of 300 m.

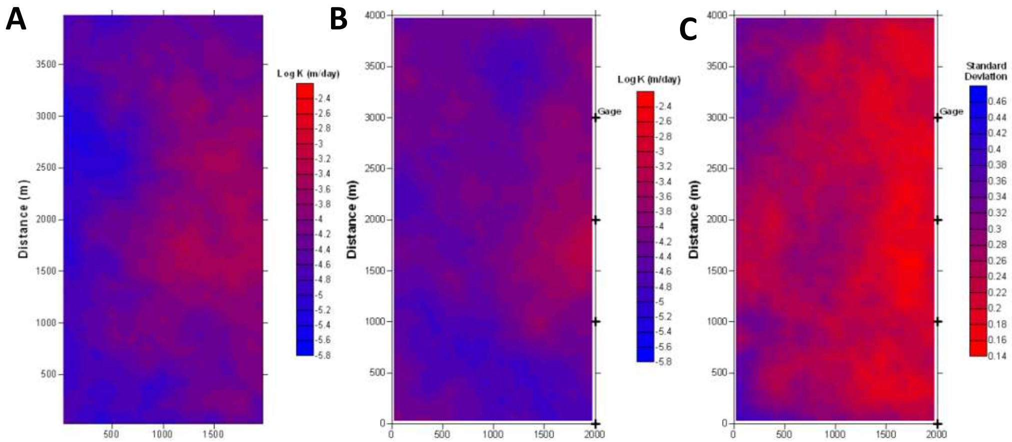

Figure 4: (A) K "true" state, (B) K update ensemble mean, and (C) K update ensemble standard deviation, conditioned by bi-weekly RFV measurements at 4 gaging locations, using a correlation length of 1500 m.

3.3 Update using both RFV and K Measurements

Further update scenarios were run using both RFV and K measurements to jointly condition the K ensemble across the four correlation lengths. Figure 5A shows the values of AE across all correlation lengths for the three scenarios of (a) only 10 K measurements are assimilated, the measurement location shown in Figure 3A, (b) only RFV measurements are assimilated, at 4 gage locations and collected bi-weekly, and (c) both RFV and K measurements are assimilated.

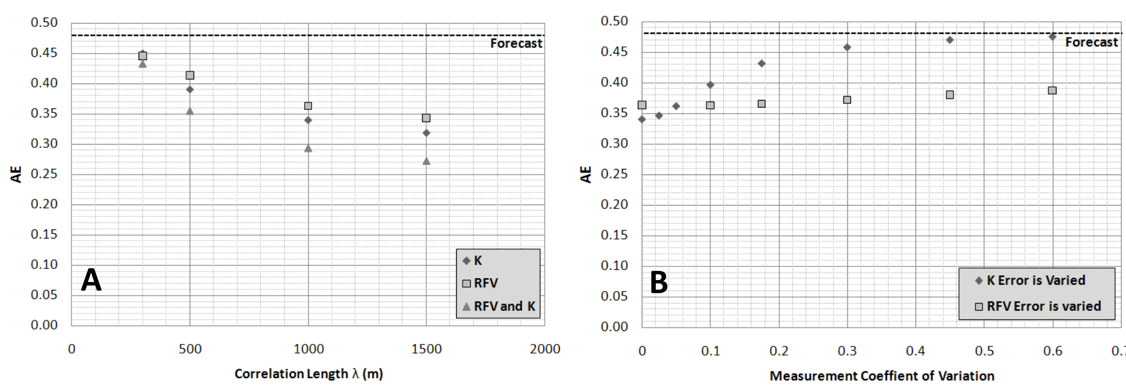

Figure 5: (A) Effect of correlation length on K conditioning for scenarios of both K and RFV measurement assimilation, and (B) Effect of measurement error on K conditioning.

Of the three scenarios, (b) has the smallest influence on conditioning the K ensemble, followed by (a) and then (c). However, varying measurement error for K shows that conditioning ceases as K measurement error increases to 0.70 (Figure 5B). In contrast, K conditioning remains practically unchanged when RFV measurement error increased to 0.70

(Figure 5B). The conditioned K ensemble mean from scenario (c) (Figure 6) has an AE of 0.293, an improvement of 39.5% from the forecast ensemble, and accurately reflects the K distribution from the "true" state, shown in Figure 4A.

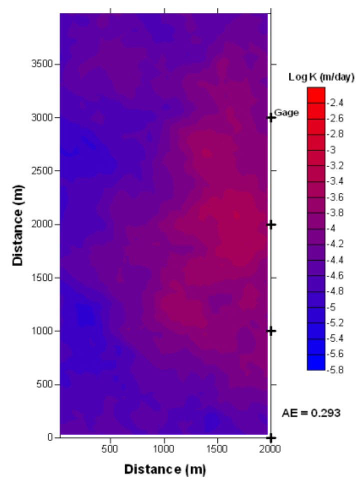

Figure 6: K update ensemble mean using bi-weekly RFV measurements at 4 gaging stations in addition to 10 K measurements. Compare with reference K field in Figure 4A.

4 CONCLUSIONS

From results shown in section 3, the ES update scheme is successful in using RFV measurements to condition the K fields to approach the "true" K field. This conditioning is most sensitive to the correlation length used in generating the K fields (Figures 3,4), followed by the number of stream gages used and the number of measurement assimilation times (Figure 2). The vast improvement in K conditioning with increased correlation length is due to the RFV occurring on only one side of the model domain. In order for spatial correlations to exist between the measurement locations and other aquifer locations, and hence for the RFV measurements to condition the K values throughout the aquifer, the correlation lengths must be significant. Assimilating 10 K measurements in the 800 ha aquifer conditions the K ensemble better than assimilating bi-weekly RFV measurements at 4 gaging locations. If errors are assigned to measurements, however, RFV measurements provide a better conditioning of the K ensemble.

Future studies might include conditioning of K fields using ground water flows to a stream using a model that could simulate more realistic surface water/ground water interactions, such as a catchment hydrology model that couples surface and variably-saturated subsurface flow.

REFERENCES

[1] 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.

[2] 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.

[3] Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35-45.

[4] 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.

[5] van Leeuwen, P.J., and G. Evensen (1996), Data assimilation and inverse methods in terms of probabilistic formulation. *Mon. Weather Rev.*, **124**, 2898-2913.

[6] Reichle, R.H., D.B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather Rev.*, **130**, 103-114.

[7] Camporese, M., C. Paniconi, M. Putti, and P. Salandin (2009), Ensemble Kalman filter data assimilation for a processbased catchment scale model of surface and subsurface flow, *Water Resour. Res.*, **45**, W10421, doi:10.1029/2008WR007031.

[8] Dunne, S. and D. Entekhabi (2005), An ensemble-based reanalysis approach to land data assimilation. *Water Resour. Res.*, **41**, W02013, doi:10.1029/2004WR003449.

[9] Hantush, M.M., and M.A. Marino (1997), Estimation of spatially variable aquifer hydraulic properties using Kalman filtering. *J. Hydraul. Engrg. ASCE*, **123**(11), 1027-1035.

[10] 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.

[11] 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.

[12] Bau, 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.

[13] Gambolati, G., M. Putti, and C. Paniconi (1999), Three-dimensional model of coupled density-dependent flow and miscible salt transport in groundwater, In J. Bear, A.H.D. Cheng, S. Sorek, D. Ouazar and I. Herrera Eds., *Seawater Intrusion in Coastal Aquifers: Concepts, Methods and Practices*, Kluwer Academic Publ., Dordrecht, The Netherlands, **10**, 315-362.