SAND2007-4273C

On the use of ensemble Kalman filters to predict stream discharge at Barton Springs, Edwards Aquifer, Texas.

K.A. Klise and S.A. McKenna
Geohydrology Department, Sandia National Laboratories, Albuquerque, New Mexico
kaklise@sandia.gov, samcken@sandia.gov

Abstract
Recent drought and population growth in central Texas has increased attention on the sustainability of Barton Springs, located in Austin within the Edwards Aquifer. Degradation of the aquifer is harmful to both residents of the growing city of Austin, who use the aquifer as their primary drinking water source, and for the Barton Springs Salamander, whose only known habitat is the area immediately surrounding the spring. The urban and environmental needs for sustainable water in this region necessitate accurate means to predict stream discharge at Barton Springs. This paper explores the use of an ensemble Kalman filter (EnKF) to make real-time predictions of discharge at Barton Springs.

Analysis using EnKF is based on the MODFLOW model developed by Scanlon et al. (2001). The EnKF model updates an ensemble of state variables (hydraulic conductivity) as control variables (stream discharge) are acquired sequentially. Updates to the state variable are applied in one of two ways: once per time step or through an iterative approach. EnKF results are then compared to results obtained by Scanlon et al. (2001) to assess the ability of EnKF to predict stream discharge.

Introduction
The Barton Springs Salamander was listed as an endangered species in 1997. Declines in the salamander population are accredited to the “degradation of the quality and quantity of water that feeds Barton Springs due to urban expansion over the Barton Springs watershed" (United States Fish and Wildlife Service, 1997). Barton Springs is fed by the Edwards Aquifer, which is used as the primary drinking water source for the city of Austin. The Edwards Aquifer groundwater supply is challenged by recent drought and population growth in Central Texas. Managing discharge at Barton Springs is important to ensure continued habitat for the Barton Springs Salamander and the water supply for the city of Austin.

A comprehensive MODFLOW model of the Barton Springs area was developed for the Lower Colorado River Authority in 2001 (Scanlon et al., 2001). Using measured groundwater pumping, rainfall, and recharge, the model predicts discharge at Barton Springs using hydraulic conductivity assigned to 9 regional zones. Over a 10 year period, the MODFLOW model predicts stream discharge with an average error of 9.5 cfs based on monthly data. The MODLFOW model uses a semi-calibrated hydraulic conductivity to best fit measured stream discharge between 1989 and 1999. This paper explores the possibility of modifying hydraulic conductivity on a monthly basis to provide a better fit to stream discharge using Kalman filter techniques. Though hydraulic conductivity is generally treated as a constant, this parameter may change in time due to changes in the groundwater level, especially in the karstic conduits observed in the Edwards Aquifer (Horvorka et al., 1998).

While several Kalman filter techniques exist, the ensemble Kalman filter (EnKF) is used in this research due to the highly non-linear relationship between the individual hydraulic conductivity values and the spring discharge. Instead of using a single realization of hydraulic conductivity, EnKF continuously updates an ensemble of state variables (hydraulic conductivity) as control variables (stream discharge) are acquired in real time. The EnKF is applied in one of two ways: once per time-step and through an iterative approach. The iterative approach ensures that updated state variables approximate control data within a specified tolerance before advancing to the next time step.

The goal of this study is to improve prediction of discharge at Barton Springs for purpose of preserving the environment for the Barton Springs Salamander and water resources used by the city of Austin. The following paper outlines the implementation of the EnKF used in conjunction with the Scanlon et al. (2001) MODFLOW model. Stream discharge predictions using the MODFLOW model are then compared to predictions made using the EnKF approach. In general, EnKF predicts stream discharge with an average error as low as 6.1 cfs, as compared to 9.5 cfs using the MODFLOW model.

Kalman Filters

The Kalman filter technique used to predict system behavior was first introduced by Kalman (1960). Kalman filter techniques are "predictor-corrector" methods and use measured data to correct model predictions by tracking the error covariance matrix. Kalman filtering, however, is only applicable to linear systems. To address this limitation, linearization techniques are added to an extended Kalman filter algorithm to approximate error statistics in the Kalman filter framework (Welch and Bishop, 2003). The linearization techniques, however, are not successful in highly non-linear systems.

The EnKF approach was developed by Evensen (1994) to predict behavior in highly nonlinear systems. A thorough review of EnKF applications has been compiled by Evensen (2003). The EnKF uses an ensemble of model states (state variables) and model estimates (control variables) to predict error statistics forward in time. State variables are then corrected for the next time step based on that error statistic, called the Kalman Gain (KG), and the difference between predictions and physical observations. The KG is defined as follows:

$$KG = \frac{COVA(State, Control)}{VAR(Control)} \tag{1}$$

where *COVA* is the covariance function, *VAR* is the variance function, *State* is the state variable, and *Control* is the control variable. The KG is used to update the state matrix using observational data with the following equation:

$$State_u = State_p + KG * (Control - Observation) \qquad (2)$$

where subscript u denotes the updated state matrix, p denotes the prior state matrix, and *Observation* is the observational data at a specified time. The value of the KG serves as a weighting scheme to place more or less emphasis on the prior state values versus the observation data in updating the state variables.

Initial representation of state variables is sampled from a probability distribution function, forming a state matrix. The state matrix is updated using the EnKF in attempt to improve the model estimate as compared to physical observations. Generally, one update per time-step is carried out. Alternatively, an iterative procedure can be applied to ensure that updated state variable approximates control data within a specified tolerance.

MODFLOW Model

The groundwater flow model produced by Scanlon et al. (2001) is a two-dimensional MODFLOW model used to evaluate spring discharge and groundwater availability in response to groundwater pumping and rainfall conditions during the 10-year period from 1989 through 1999. The model region is bound by the Mount Bonnell fault to the west, a groundwater divide along Onion Creek to the south, a "Bad water" line to the east approximately located at Interstate-35, and the Colorado River and Town Lake to the north (Figure 1). The region consists of a recharge zone to the west and a confined aquifer to the east. Five creeks contribute to recharge within the study area (Onion, Bear, Slaughter, Williamson, and Barton). Barton Springs is the main outflow for the aquifer and is located near the confluence of Barton Creek and the Colorado River.

The geology of the region consists of Lower Cretaceous carbonates. Uplift and erosion to the west have exposed Edwards aquifer in the western region of the study area. This region is characterized by creeks that recharge the aquifer. To the east, lower permeability units remain and the aquifer is confined. The aquifers rapid response to recharge and drought events has been attributed to a high degree of karstification.

The MODFLOW model consists of 120 by 120 grid cells covering an area of approximately 250 square miles. Hydraulic conductivity is divided into 9 zones based on the distribution of head gradients within the region. A combination of trial and error and automated inverse methods was used to produce the final "semi-calibrated" model that best fits the measured stream discharge and groundwater levels (Figure 2).

The MODFLOW model also takes into account the topography of the aquifer, storativity, specific yield, recharge, and groundwater pumping. Topography of the aquifer is based on ground surface elevation in the unconfined region to the west and

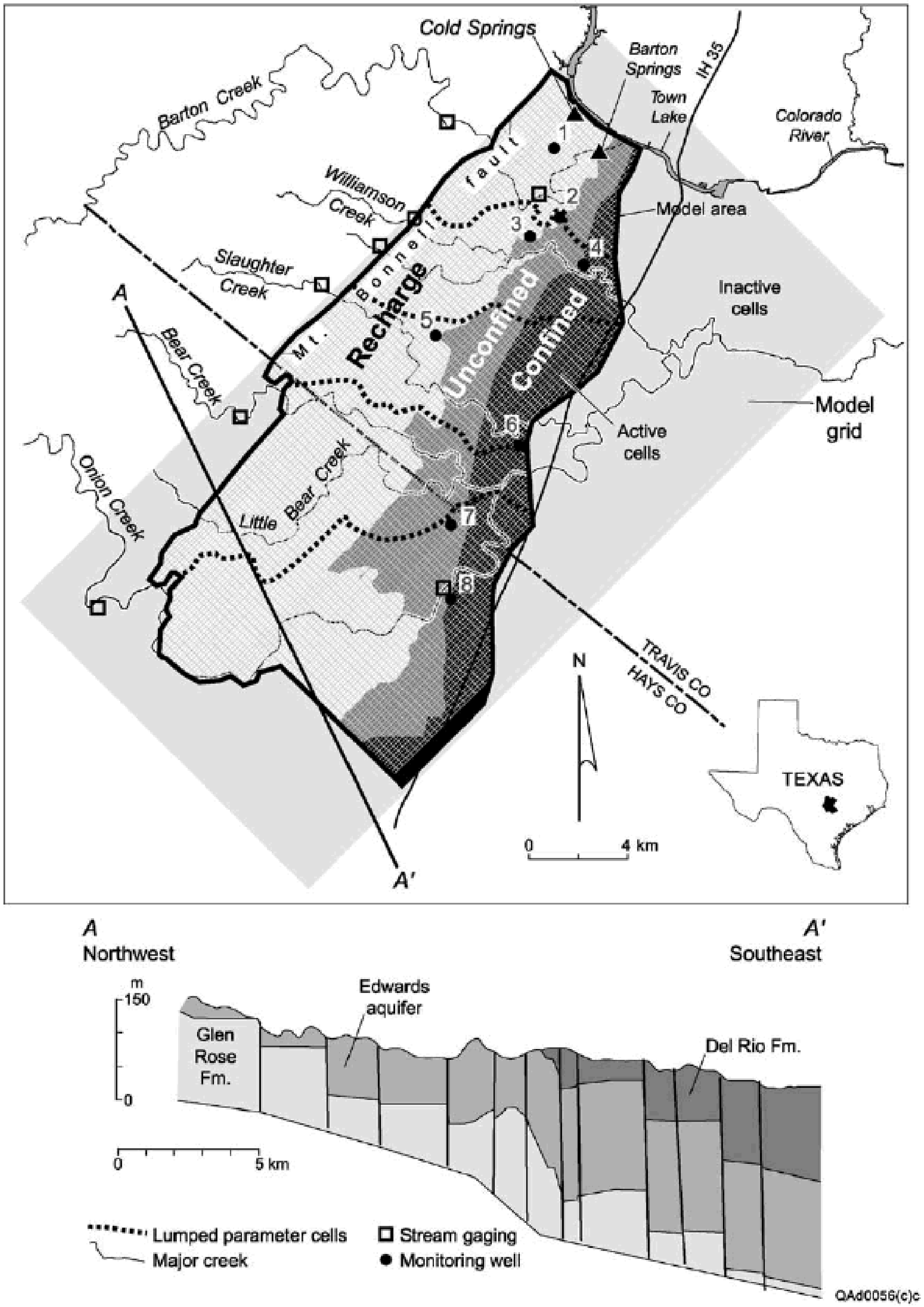
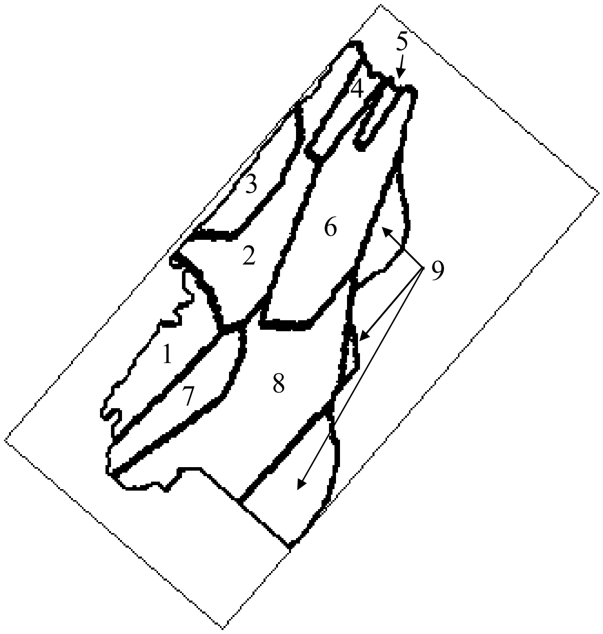


Figure 1. Location of study area (figure from Scanlon et al. 2003) showing Barton Springs in relation to stream gauging stations, monitoring wells, and creeks. Location of recharge, confined and unconfined zones are shown along with a geologic cross-section of the region.



Zone	K (ft/day)	log10(K)
1	3	0.477
2	3.5	0.544
3	11	1.041
4	100	2
5	1236	3.092
6	320	2.505
7	4.5	0.653
8	96	1.982
9	39	1.591

Figure 2. Map and table of zoned semi-calibrated hydraulic conductivity from Scanlon et al. (2001).

the top of the low permeability clay to the east. The initial head for the transient MODFLOW model are based on results from a steady state model using average recharge for a 20-year period (1979 through 1998) and groundwater pumping records for 1989. Specific yield and specific storage are based on data from Slade et al. (1985).

Both local and diffuse recharge is considered in the model. Local recharge occurs in each of the 5 main creeks that cross the recharge zone. Local recharge is added uniformly along each creek based on analysis of flow loss. Diffuse (interstream) recharge is set at 15 percent of the total stream recharge and is assigned to all active cells in the model. Groundwater pumping as recorded by Barton Springs Edwards Aquifer Conservation District (BSEACD) is also added to the model. Over 100 pumping locations are within the model domain, primarily in the southeast. Domestic pumping that remains unreported was added across the model domain according to countywide estimates.

The MODFLOW model predicts spring discharge at Barton Springs in monthly time increments. Comparison of measured and modeled spring discharge is shown in Figure 3. Measured spring discharge at Barton Springs between January 1992 and July 1992 is unreliable due to flood conditions and these observations and predictions are removed from further analysis. The average difference between measured and modeled stream discharge is 9.5 cfs.

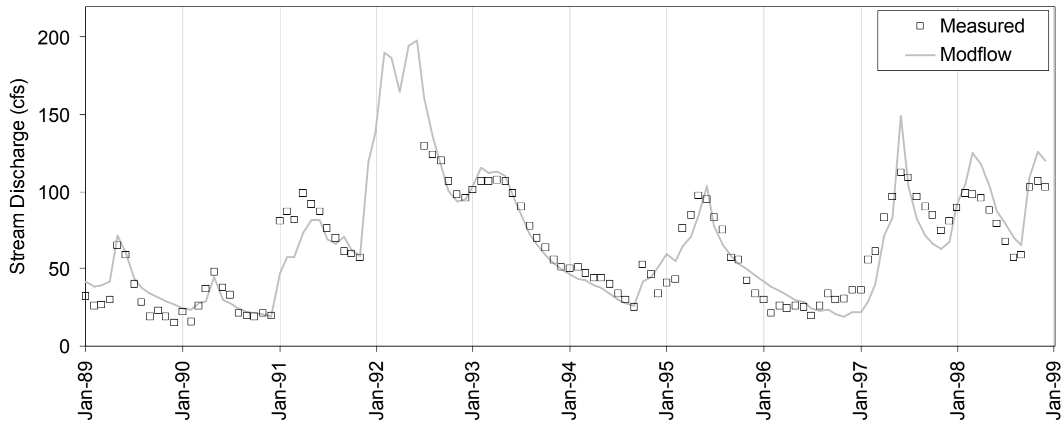


Figure 3. Stream discharge prediction from the Scanlon MODFLOW model as compared to measured discharge.

Ensemble Kalman Filter Model

Using the MODFLOW model described above, the EnKF is implemented by updating the state variable after each time-step. To apply the EnKF with the MODFLOW model, state variable, state matrix, control variable and observational data must be defined for use in (1) and (2). For the Barton Springs model, the state variable is hydraulic conductivity assigned to 9 zones. The state matrix is an ensemble of realization of the state variable. 50 realizations are used in this study. To build the state matrix, state variables for each of the 9 hydraulic conductivity zones are selected from a lognormal distribution (with a mean as specified in Figure 2 and standard deviation of 1). These are drawn independently from each of the nine specified distributions. The control variable is the monthly model prediction of discharge at Barton Springs while the observational data is monthly measured discharge.

Using the ensemble of hydraulic conductivity models and associated predicted discharge, a KG is calculated for each zone in the model. For each time-step in the model, EnKF updates the state matrix using this error statistic. For example, if there is positive correlation between the hydraulic conductivity in one zone and the predicted stream discharge, KG is positive. In that case, if the model predicted stream discharge is low compared to the observational data, the hydraulic conductivity will increase proportional to the KG.

For this study, EnKF updates are based on 2 methods: a single update per time step, and an iterative process. Iterative updates are carried out until the difference between the predicted and measured stream discharge is less than 5 cfs. Once this error tolerance is met, the model is advanced to the next time-step. Starting head for MODFLOW at successive time steps is the ensemble average of heads resulting from the previous time step.

To ensure proper convergence in the MODFLOW model, several checks were included while running the EnKF. First, upper and lower limits for hydraulic conductivity are assigned. Hydraulic conductivity is constrained between 0.5 ft/day and 1000 ft/day. The zone that contains the spring has an upper limit of 2000 ft/day. Ensuring that hydraulic conductivity stays within a reasonable range greatly enhances convergence. Changes to hydraulic conductivity are also limited to 0.5 orders of magnitude for each update. This prevents hydraulic conductivity from quickly migrating to the outer limits and allows the model to take successive time steps (or iterations) into consideration when the filter calculates a large change in the state variable. If the updated state variable is outside the outer limits, the variable is set to half the distance between the breached threshold and its previous value. This helps maintain variability throughout the ensemble. For each zone, the ensemble of hydraulic conductivity must have a standard deviation greater than 0.05. If the variability is too low, then hydraulic conductivity cannot sufficiently change in the EnKF updating routine. In the case where the MODFLOW model still does not converge, the state variables in that model are perturbed using a random variable and the model is re-run. The state variable is not updated between January and July 1992 due to unreliable stream discharge measurements associated with a flood that winter.

Results

The EnKF is run both with a single update and with an iterative process. The single update approach changes the state variable once per time-step in attempt to better predict the control variable forward in time. The ensemble average of predicted stream discharge using this method as compared to the MODFLOW model is detailed in Figure 4. The resulting change in hydraulic conductivity over the 10 year period shows that some zones remains moderately stable (zone 1, 5, 7) while other zones fluctuate over 2 orders of magnitude. As compared to the MODFLOW model, using the EnKF with a single update decreases the average difference between measured and modeled stream discharge from 9.5 cfs to 6.6 cfs, a 31% improvement.

Combining the EnKF with an iterative approach adds convergence criteria to the model before the model advances to the next time-step. Prediction of stream discharge for each time-step is made with the first iteration, similar to the single update method. After a prediction has been made, the observed stream discharge at that time-step is used to continue updating the hydraulic conductivity until the difference between measured and modeled stream discharge less than 5 cfs. Once convergence is met, the EnKF predicts stream discharge at the next time-step. The prediction of stream discharge using the iterative method is detailed in Figure 4. Using an iterative approach decreases the average difference to 6.1 cfs, a 36% improvement relative to the results obtained with the MODFLOW model.

The iterative approach uses, on average 1.6 iterations per time-step. The highest number of iterations occurs where the stream discharge changes greatly from one month to the next. The iterative approach takes 13 iterations to converge when stream discharge changes from 20 to 81 cfs in one month (December 1990 to January 1991). Figure 5 shows the changes in predicted stream discharge (for each realization and for the ensemble average) associated with those iterations. Allowing the hydraulic conductivity to change according to the large variation in stream discharge experienced in January 1991 allows the EnKF model to better predict successive time steps. Using a single update EnKF model, it takes 4 months (time steps) for the model to achieve a comparable fit to the measured discharge. The MODFLOW model does not adequately predict the same jump in stream discharge for 6 months. Using yearly averages, a full comparison of predicted stream discharge using the MODFLOW and EnKF models is detailed in Table 1.

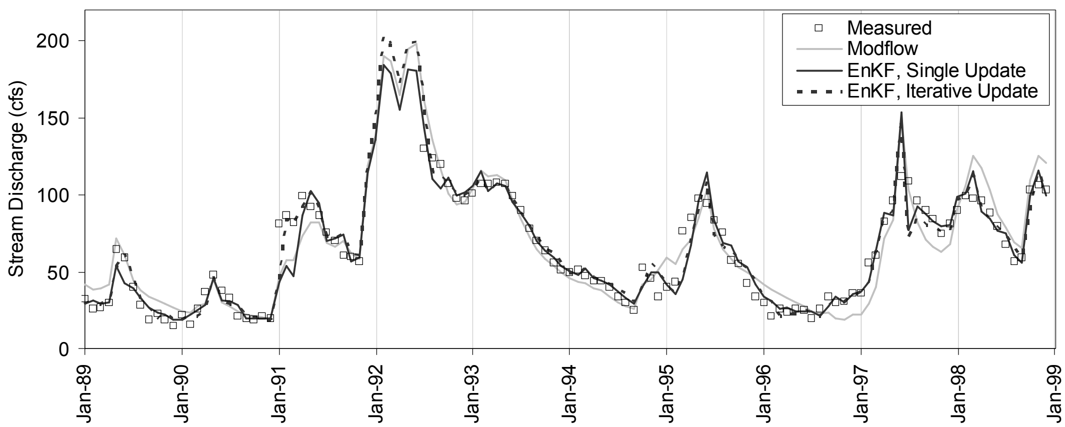


Figure 4. Comparison of measured stream discharge with discharge predicted by the MODFLOW model and EnKF model (ensemble average).

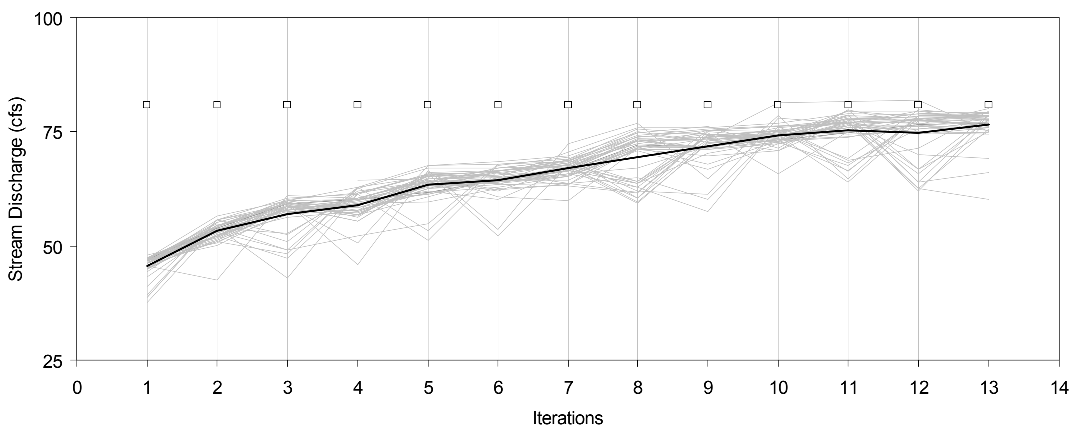


Figure 5. Change in stream discharge using the EnKF iterative method for January 1991. The initial low prediction is due to the large change in stream discharge between December 1990 (20 cfs) and January 1991 (81 cfs). Black squares show the measured discharge (81 cfs). Gray lines are the stream discharge results for individual realizations; the black line is the ensemble average.

Table 1. Average difference (modeled vs. measured) in stream discharge (cfs) for each year of the model and for the 10-year average.

Year	MODFLOW Scanlon et al. (2001)	EnKF single update	EnKF iterative update
1989	9.4	5.0	4.6
1990	3.8	3.3	4.0
1991	13.9	14.8	7.6
1992	9.6	9.2	9.2
1993	4.3	2.6	2.8
1994	5.7	4.3	5.3
1995	9.7	9.5	9.6
1996	9	3.4	3.4
1997	17.1	9.8	10.9
1998	12.7	6.1	5.6
10-year average	9.5	6.6	6.1

Conclusions

Both the single update and iterative EnKF model improve the ability of the Scanlon et al. (2001) MODFLOW model to predict stream discharge at Barton Springs over the 10 years between 1989 and 1999. While prediction on a monthly basis is not always improved, on average the error is decreased by 31% using the single update method and 36% using the iterative approach. The EnKF allows hydraulic conductivity, the state variable, to fluctuate to best predict the discharge at each month. While some fluctuation in conductivity is reasonable, especially considering the karstic nature of the aquifer, the EnKF model shows some changes in hydraulic conductivity over 2 orders of magnitude. Large changes in hydraulic conductivity are not necessarily real changes to the aquifer, but merely a means to better describe the aquifer in terms of stream discharge. The MODFLOW model developed by Scanlon et al. (2001) also compares measured and modeled water levels at 8 observation wells throughout the model domain. Water level may also be used as a control variable in the EnKF approach and may further improve the ability to predict stream discharge using the MODFLOW model.

Acknowledgment

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

Evensen, G., 1994, Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research, 99:10143-62.

Evensen, G., 2003, The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, Ocean Dynamics, 53, 343-367.

Hovorka, S. D., Mace, R. E., and Collins, E. W., 1998, Permeability structure of the Edwards aquifer, South Texas—implications for aquifer management: The

University of Texas at Austin, Bureau of Economic Geology Report of Investigations No. 250, 55 p.

Kalman, R.E, 1960, A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82.

Scanlon, B. R., R.E. Mace, B. Smith, S. Hovorka, A.R. Dutton, R. Reedy, 2001, Groundwater Availability of the Barton Springs Segment of the Edwards Aquifer, Texas: Numerical Simulations Through 2050, Lower Colorado River Authority (UTA99-0), 99p.

Scanlon, B. R., R. E. Mace, M. E. Barrett, and B. Smith, 2002, Can we simulate regional groundwater flow in a karst system using equivalent porous media models? Case study, Barton Springs Edwards Aquifer, USA. Journal of Hydrology, 276:137-158.

Slade, R.M., L. Ruiz, and others, 1985, Simulation of the flow system of Barton Springs and associated Edwards aquifer in the Austin area, Texas: U.S. Geological Survey, Water Resources Investigations Report 85-4299, 49p.

United States Fish and Wildlife Service, 1997, Endangered and Threatened Wildlife and Plants; Final Rule To List the Barton Springs Salamander as Endangered, Federal Register, Vol. 62, No. 83.

Welch, G., and G. Bishop, 2003, An introduction to the Kalman filter. Technical report, Department of Computer Science, University of North Carolina, 2003. Technical Report No. TR 95-041.