

COVARIANCE CORRECTION FOR ESTIMATING GROUNDWATER LEVEL USING DETERMINISTIC ENSEMBLE KALMAN FILTER

J. Behmanesh* and M.M. Bateni

Department of Water Engineering, Urmia University, Urmia, Iran

Received: 11 March 2014 / Accepted: 15 December 2014 / Published online: 15 January 2015

ABSTRACT

The main problem in developing a groundwater model is to determine model parameters, particularly hydrogeologic coefficients, in a precise way. In this research, Deterministic Ensemble Kalman Filter (*DEnKF*) is described as a modern sequential method for data assimilation and a localization scheme within the framework of *DEnKF* is applied. Najafabad aquifer (in Iran) with area of 1150 km², is modeled in the time window of Oct. 2000 to Sept. 2007 to obtain water table level data when its values of hydrogeologic coefficients calibrated and verified. *DEnKF* assimilated 45 observations of true run into the model with 2, 5, and 10 times of calibrated values of hydraulic conductivity and specific yield. This filter has been run both with and without use of localization. Results show easily-implemented localized *DEnKF* is favorably robust in groundwater flow modeling.

Keywords: Data assimilation, localization, Groundwater flow model, Iran.

1. INTRODUCTION

Groundwater Resources Management and Data Assimilation

Appropriate management of groundwater resources, especially during droughts, leads to better and more effective use of them [1]. An important quantity in this issue is water table depth in pheratic aquifers or piezometric level in confined aquifers [21]. In the past decades, these types of data have been obtained by modeling the aquifer system, but uncertainty in input variables and model parameters has caused the results to be inaccurate [5]. On the other hand, responsible

Authour Correspondence, email: j.behmanesh@urmia.ac.ir

[ICID: 1131844](#)

organizations in water resources management usually have observation wells in aquifers which their water levels are measured regularly. So, using both types of data –observed and model results- to reduce the uncertainty would be a wise approach.

Data assimilation methods are made to combine the observed data of any kind– here, water level data from observation wells- with predictions of the implemented model- here groundwater flow model. In fact, these methods try to find the best combination of observed and model predictions based on the uncertainty involved in each method as the solution [15].

Kalman Filter (*KF*) is the most common sequential data assimilation method that assimilates a state-space expression of a prediction model with noisy observations to give an estimation of the system state with the least square error [11]. *KF* is developed based on the state and observation equations that are respectively stochastic representation of model and observations. System state is a set of variables within the system that contain all past information affecting the future behavior of the system [20].

Ensemble Kalman Filter (*EnKF*) is a large-scale filter based on the representation of probability density of state estimate by finite numbers of system states (N) that are generated randomly [7]. These states, each one called a realization, form the ensemble used in the algorithm of filter. In fact, the only difference between this filter and classical *KF* is that *EnKF* uses an ensemble of predictions for the calculation of covariance of prediction error. This makes it a Monte Carlo implementation of the classical *KF* [4].

2. BACKGROUND

In order to gain observational information in modeling procedure more efficiently, Evensen (1994) introduced the implementational principles of *EnKF* [3]. It was the first alternative for classical *KF* in large scale systems like aquifers that comes with feasible computational difficulty.

Sakov and Oke (2008) presented a simple and efficient linear approximation of *EnKF* which is called Deterministic Ensemble Kalman Filter (*DEnKF*), because of its deterministic characteristic and its similarity to the Kalman filter algorithm [18].

For Hydrogeologic application of *EnKF*, Sun et al. (2009) compared four types of *EnKF* to estimate hydraulic conductivity of an unreal example and found that *DEnKF* outperforms and is more robust than others, even in small ensemble sizes [19].

In this study, groundwater flow is modeled in Najafabad aquifer (Isfahan province, central Iran). The aquifer is divided into five zones (based on general hydrogeologic characteristics)

and identified with different hydrodynamic coefficients. Then, the model results are saved in some points at each time-step. A part of this information, as the observational data, was assimilated into the model with inexact parameters within the *DEnKF* algorithm to evaluate filter robustness. The results of this data assimilation, with different length-scale localization parameters and without localization, are analyzed.

3. MATERIALS AND METHODS

Groundwater Flow Equation

Groundwater flow equation in a phreatic, isotropic aquifer which is used in this research can be written as follows

$$\frac{\partial}{\partial x} k \frac{h}{x} + \frac{\partial}{\partial y} k \frac{h}{y} + \frac{\partial}{\partial z} k \frac{h}{z} + N - q - N S_s \frac{h}{t} \quad (1)$$

where k is the hydraulic conductivity value (Lt^{-1}), h is the potentiometric head (L), N is recharge (volumetric flux per unit volume, t^{-1}), q is discharge (volumetric flux per unit volume, t^{-1}), S_s is specific yield, and t is time (t) [10].

KALMAN FILTER AND ITS APPLICATION

System and Observational Equations

The groundwater flow model, derived by finite differences, considered as a linear state equation which is defined as follows

$$X_k = AX_{k-1} + Bu_k + W_k \quad (2)$$

where X (system state vector) includes groundwater levels at every mesh points in the aquifer, A is the transition matrix, B is the coefficients matrix and u is the external variable, while k is time index. System noise (W) is defined as the difference between model and actual state [20]. It is assumed that initial system state vector is a random process with mathematical expectancy equal to X_0 and covariance matrix equal to P_0^a .

Observational equation of the filter is

$$Z_k = MX_k + V_k \quad (3)$$

where Z is defined as the vector containing observed (measured) data-observation well records- and V , which is known as observation noise, is a Gaussian process vector with covariance equal to R . Observation matrix (M) which relates system state to observation, contains elements between zero and one, depending on location of the observation wells [20].

The Kalman filter is most often conceptualized as two distinct phases: "Prediction" and "Analysis". The prediction phase uses the state estimate from the previous time step to produce a predicted estimate of the state and its covariance at the current time step (X_k^p, P_k^p). In the analysis phase, the current prediction is combined with current observation information to refine the state estimate. This improved estimate is termed the analyzed state estimate (X_k^a, P_k^a) [16].

Recursive equations to obtain filtered state and its covariance matrix are summarized as (4)

Time updating (prediction):

$$X_k^p = A X_{k-1}^a + B u_k \tag{4}$$

Measurement updating (analysis):

$$X_k^a = X_k^p + K (Z_k - M X_k^p) \tag{5}$$

where Kalman gain matrix (K) which is the weighting matrix for observations versus model results, is

$$K_k = P_k^p M^T (R_k + M^T P_k^p M)^{-1} \tag{6}$$

ENSEMBLE KALMAN FILTER

Prediction and Analysis in EnKF

Prediction equation of *EnKF* is similar to time update equation in *KF*. The only difference is that each ensemble member i.e. each column of the ensemble matrix is operated on individually.

By putting members together again, the predicted ensemble matrix (X^p) is made [4].

In all types of *EnKF*, it has been tried to derive analysis ensemble matrix -from predicted ensemble matrix- with analysis covariance which is consistent with its value of the *KF*. The point is how to make an ensemble and this results in different versions of *EnKF*.

Deterministic Ensemble Kalman Filter (DEnKF)

We consider scaled ensemble anomalies matrix, $X^a = \frac{1}{\sqrt{N-1}} [x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_N - \bar{x}]$, where x_i is the i th ensemble member for $i=1, \dots, N$ and \bar{x} is the ensemble mean. Error covariance matrix can also be derived by squaring it.

With this notation, Sakov and Oke (2008) derived a new *EnKF* formulated as

$$X^a = X^p - \frac{1}{2} K M X^p \tag{7}$$

where K and M are Kalman gain and Observation matrix, respectively. Ensemble mean separately updates to analysed one using equation [5].

Because of its deterministic characteristic and similarity to the KF , this simple and computationally efficient filter is called $DEnKF$ [18]. Because of the presence of prediction error covariance (P^p) in the filter scheme, it is easy to implement schur localization for it.

Problems Concerning Small Ensembles

Using small ensembles for calculating error covariance matrix can introduce sampling errors as false correlations between distant locations or between points that should be uncorrelated. With false updates corresponding to these false correlations, the ensemble variance has an inevitable reduction which causes significant underestimation of actual variance, as time goes on. This situation is called inbreeding and may results in filter failure [14; 2].

A feasible method to attenuate negative ramifications of small ensemble is multiplying the covariance matrix by a semi-positive definite correlation matrix element-wise. Every zero element of the correlation matrix causes the corresponding element of the product matrix to be zero. This can be as effective as increasing the ensemble size, when the ensemble is small [12; 13]. Using such a method often causes a dramatic rise of zero elements of the covariance matrix, while it needs less computational effort than local analysis. Thus, the problem is to find such a matrix that its elements in far enough distance away from each other are zero¹.

The most widely used correlation function is compactly supported 5th order piecewise rational function. The function is [6]

$$\begin{aligned}
 & \frac{r^5}{4c^5} < \frac{r^4}{2c^4} < \frac{5r^3}{8c^3} > \frac{5r^2}{3c^2} < 1, & 0 \leq r \leq c \\
 (r,c) \text{ N } & \frac{r^5}{12c^5} > \frac{r^4}{2c^4} < \frac{5r^3}{8c^3} < \frac{5r^2}{3c^2} > \frac{5r}{c} < 4 > \frac{2c}{3r}, & c \leq r \leq 2c \\
 & 0, & 2c \leq r
 \end{aligned} \tag{8}$$

where r is the distance (positive) between points and c is half of the length-scale parameter, based on prior information or by try and error [4]. If the length-scale is larger than the optimum value it cannot set spurious correlations to zero for relatively far distances (in error covariance matrix). When the selected length-scale is smaller, the filter is not effective enough to adjust

¹ semi-positive definiteness of both matrices to be multiplied guarantees semi-positive definiteness of the product matrix [8].

the model due to artificially setting some real dependencies to zero artificially. The function is used to determine the elements of the correlation matrix.

Houtekamer and Mitchell (2001) multiplied the correlation matrix to both replications of P^p in the Kalman gain matrix equation [9]. In this study, the same method of Houtekamer and Mitchell (2001) is used.

Application Example

Characteristics of the Studied Region

The studied region is Najafabad sub-basin (Fig. 1). It is located between $50^{\circ} 57''$ to $51^{\circ} 44' 26''$ east longitude and $32^{\circ} 20' 13''$ to $32^{\circ} 49' 21''$ north latitude. The area is approximately 1720.23 km^2 and its aquifer is about 1142.67 km^2 . Mean annual evapotranspiration is about 1500 mm while mean annual precipitation is only 158 mm, which nearly all falls in winter months.

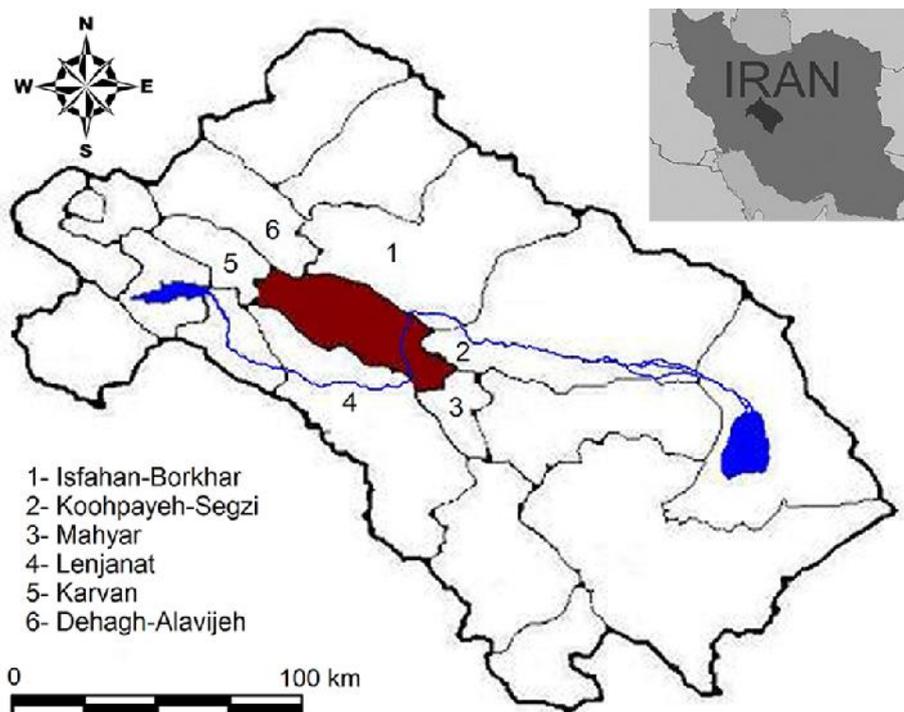


Fig. 1. Zayandehrud River (blue) passes through Najafabad sub-basin (brown) and neighboring sub-basins in Zayandehrud River basin

Najafabad aquifer (Fig. 2) is a phreatic aquifer, which mainly consists of sediments related to the fourth geologic period. Important sources of recharge for the aquifer are Zayandehrud River, deep percolation from irrigation practices and precipitation, in order of

importance. Water withdrawals from the aquifer are only for agricultural uses². Groundwater withdrawal takes place by 2168 deep wells, 8452 semi-deep wells, 73 qanats and one spring.

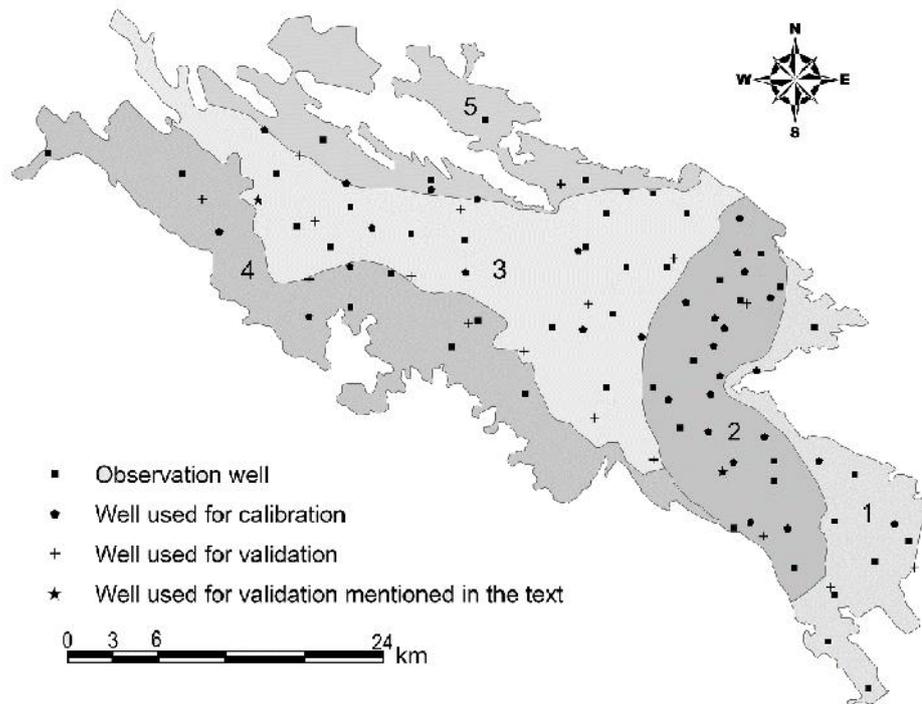


Fig. 2. Najafabad aquifer divided into five zones for calibration and its particular wells

Modeling Details

The model is a Mesh-centered finite difference with fixed cell size of 500×500 m, monthly stress periods, and time span from October 2000 to September 2007.

The aquifer was divided into five zones for modeling, based on geological characteristics and previous modeling studies [17]. Constant hydrogeologic coefficients for each zone were considered (Fig. 2).

The hydrogeological coefficients of these five zones were derived through calibration with the index of sum of squared differences between monthly data of 32 observation wells (out of 49 observation wells) from October 2000 to March 2005. The results are summarized in Table 1. Increasing or decreasing of hydrogeologic parameters within a wide neighborhood of the calibrated values showed that the model is much more sensitive to values of specific yield in comparison with hydraulic conductivity values.

The aquifer was modeled with these calibrated parameters from April 2005 to September 2007. Modeling results were compared with remaining 17 observation wells' data. The results of the

² Obtained from Zayandehrud River Basin Report by Isfahan Regional Water Authority, Isfahan, (2008).

model showed a good performance. For example, the modeled and observed values of two observation wells in northwest and southeast of the aquifer, which have different recharge and discharge rates, are shown in Fig. 3 for a 30-month period. As we want to evaluate the filter performance in groundwater problem exclusively, the calibrated model results assumed as true data, therefore the calibration is not effective in evaluation of the filter performance explicitly.

Table 1. Calibrated values for hydrogeological coefficients

Zone number	Hydraulic conductivity (m/day)	Specific yield (%)
1	0.88	0.25
2	0.73	0.44
3	0.61	0.49
4	1	0.31
5	1	0.31

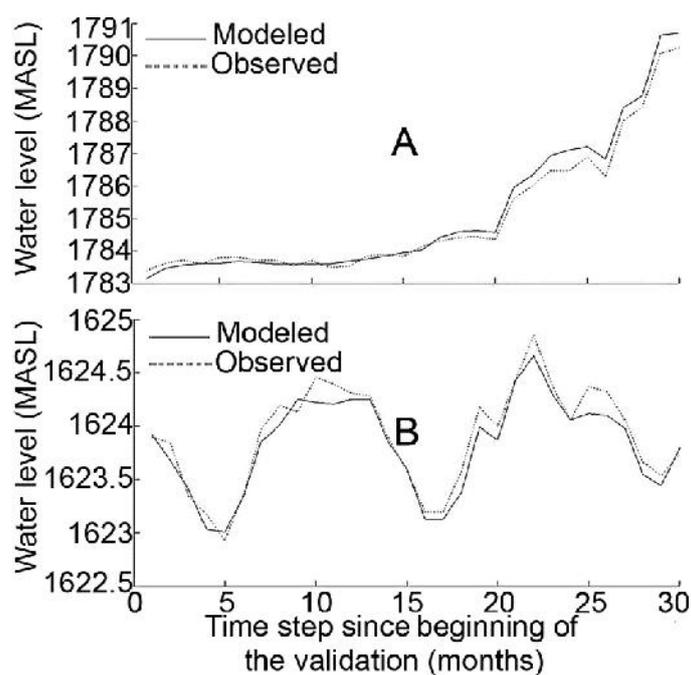


Fig. 3. Two examples of model validation results; A: Azizabad well (west well in Fig. 2) and B: Cham-va-Ardal well (east well in Fig. 2)

Filtering Details

Simulation results for all grid nodes within the aquifer for 84 months are recorded. The results are treated as actual data in assessment of the filter performance.

Records of 45 selected cells in the simulation are treated as observation data to be used in the filtering process. Observational error is assumed as a Gaussian stochastic process with zero mathematical expectancy (white noise) and variance of 0.01 m^2 and is artificially added to the observed values. This variance is recommended by the local water authority experts and based on the precision of measuring instruments for the observation wells. It is tried to select these 45 cells so that they contain all types of discharge and recharge regimes existing in the region. Positions of all observation wells and points selected as the observation wells in the filtering procedure are shown in Fig. 2.

Hydraulic conductivity and specific yield values of the aquifer for all five zones are modified by 2, 5 and 10, respectively, in six different simulations, while a set of observations from the simulation with the calibrated values of the coefficients are available. The results of the simulation with inaccurate parameters are considered as state variables and assimilated within the *DEnKF* with 100 ensemble members with and without localization. Localization is implemented with three length-scale parameters³ (5×500 , 7×500 and 10×500 m). The results of each run are compared with the simulated results for all cells in the flow domain.

Initial sampling

Adding an appropriate random matrix to the initial condition of the system can lead to an applicable initial state for the ensemble filter. The random matrix has N columns (equal to the members of the desired ensemble) and n rows (equal to the dimension of the system). It has zero mean for each column and the covariances with other elements of the same column are based on the derived variogram.

To obtain such a matrix, the state field defined by the variogram is transferred to Fourier space. Sampling is done randomly and finally the random sample from Fourier space is returned to the real space by reverse transformation (For more detailed procedure see [4]).

Filter Performance Evaluation

To evaluate the performance of the filter, a feasible criteria would be the square root of mean squared differences (*RMSE*) between the filter results and true model. However, because of the absence of true results, the variance of the analysis ensemble is considered in majority of literature.

RMSE is defined as

³ Values for length-scale parameter are selected based on an intuitive knowledge came from the variogram obtained from average of monthly records of all 49 observation wells.

$$RMSEN = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i^t - x_i^t)^2} \tag{9}$$

where \hat{x}_i is the i th element of vector of filter results \hat{X} , x_i is the i th element of vector of true model results X and n is the size of both vectors, which is equal to the system dimension.

4. RESULTS

Filter Performance Without Localization

In case of inexact hydraulic conductivity values, $RMSE$ values for the filter without localization are small and reasonable. The $RMSE$ values for these three non-localized filters show severe reduction in the first few steps and then they show smooth curves and no significant drop is observed.

Poor filter performance in the case of inexact specific yields can be justified due to higher sensitivity of the model to specific yield. For runs with inexact specific yield values, after a severe drop of $RMSE$ in the first few time steps, the criteria began to increase. The increase of $RMSE$ is a sign of filter failure. Fig. 4 and Fig.5 show the mentioned points. In these figures, c equals to half of the length scale as defined for equation [8]. In case of no localization, it is 'inf' and in other cases, c is in kilometers unit.

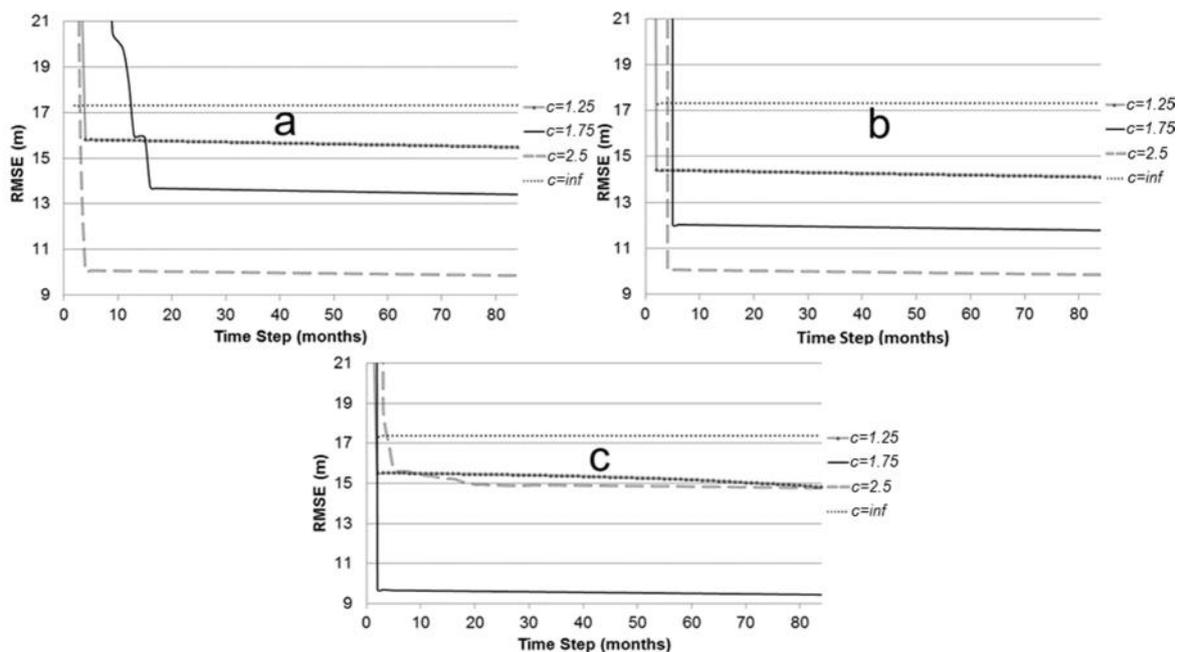


Fig. 4. $RMSE$ values for filter; a: with $2 \times K_{exact}$, b: with $5 \times K_{exact}$ and c: with $10 \times K_{exact}$

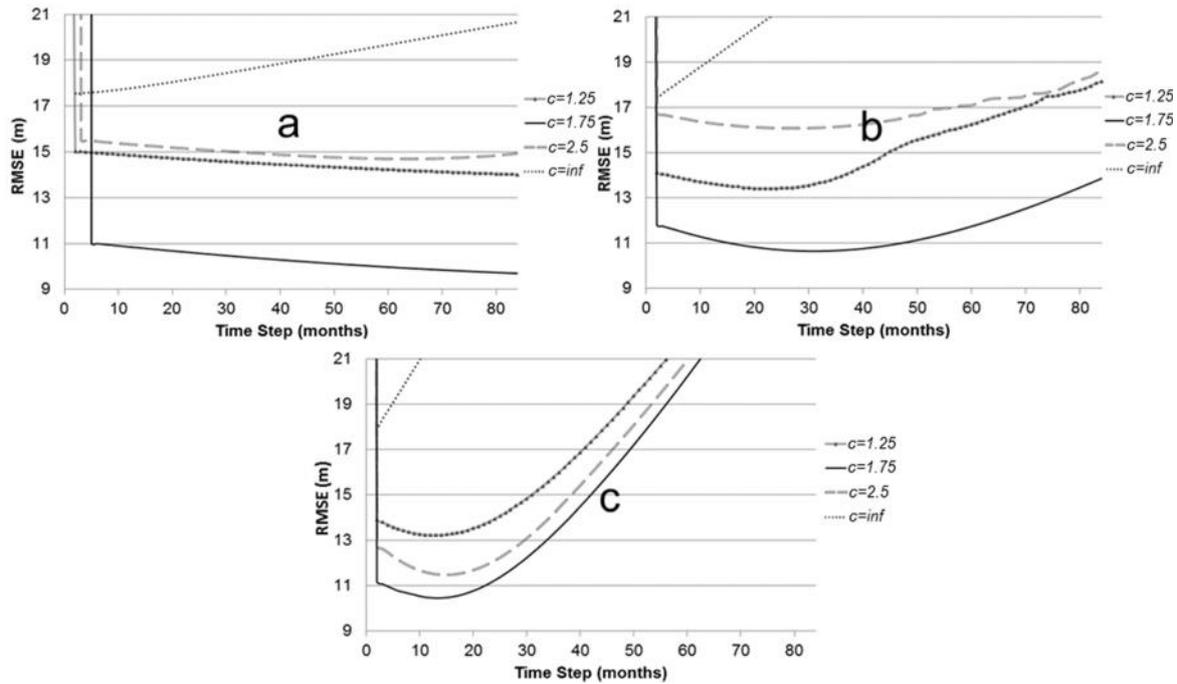


Fig. 5. RMSE values for filter; a: with $2 \times \text{Sexact}$, b: with $5 \times \text{Sexact}$ and c: with $10 \times \text{S exact}$

Filter Performance With Localization

As it is shown in Fig. 4 and Fig. 5, localization with all the three length-scales improved the filter performance in all cases. However, in the case of 2 or 5 times of the true hydraulic conductivity, this improvement was maximized at length scale of 5000 m, and among all other cases, 3500 m length scale was the best. However, in the case of 5 or 10 times of the exact specific yield, localization is not helpful for filter divergence.

After all, localization caused better results (compared to lack of localization) in all cases, e.g. in the case that specific yield is twice the true value, localization with length scale of 2500 and 3500 m prevents the filter from divergence.

5. DISCUSSIONS

In this study, Najafabad aquifer was modeled to obtain water table level data and its values of hydrogeologic coefficients calibrated and verified. Deterministic Ensemble Kalman Filter (*DEnKF*) assimilated 45 observations of true run (the calibrated model) into the model with 2, 5, and 10 times of calibrated values of hydraulic conductivity and specific yield. This filter has been run both with and without use of localization.

The use of *DEnKF* for inexact groundwater flow modeling shows remarkable improvement. The filter improves the results of the inexact model in each time step based on the information of observed data. Usually, this improvement is shown mainly after the first few

time steps which indicates efficiency of the filter. In the case of non-localized filters, less values of *RMSE* for inexact hydraulic conductivity values can be justified under the lesser sensitivity of the model to this parameter.

Using this filter without localization, shows a rising trend in most cases after significant reduction of error in the first few steps. This rising trend is related to spurious correlations in the predicted error covariance matrix (P^p) which are generated during updating process of the filter. The model as a transient one is initialized by the values of the previous time step. Therefore, error in previous steps due to false correlations can be treated as an ancillary source for of the huge *RMSE* values of final steps. Localization eliminates these false correlations and as a result, the filter will experience less error, especially in final time steps.

P^p is not presented directly in the equations of most ensemble filters; when its square root or other forms exist, applying the localization is not directly possible. Therefore, localization often leads to computational difficulties. Unlikely, localization is easily possible due to the presence of P^p in the equations of *DEnKF*. On the other hand, as it was mentioned before, localization results in better filter results in all cases. Therefore, localization (with this tapering function) is always recommended.

6. CONCLUSIONS

We applied a new deterministic modification of the traditional *EnKF*, which is called *DEnKF*, with the use of traditional localization to a realistic groundwater flow model in central Iran. Such a localization can be easily implemented within the framework of the *DEnKF*. The filter was run with inexact specific yield and hydraulic conductivity. The model was more sensitive to specific yield than hydraulic conductivity. Our experiments indicated that *DEnKF* can be used as a robust filter in groundwater flow modeling. Furthermore, the localized *DEnKF* is less susceptible to filter divergence than non-localized one. Overall, we concluded that the *DEnKF* combines the feasible performance with simplicity of the implemented localization in its framework and therefore represents an interesting alternative to ensemble-based Kalman filters in groundwater modeling.

REFERENCES

- [1] Alley W.M., Reilly T.E., Franke O.L., Sustainability of ground-water resources, U.S. Geological Survey Circular Denver, Colorado, 1999.
- [2] Ehrendorfer M., Meteorol. Z. 2007, 16 (6), 795-818.

- [3] Evensen G., *J. Geophys. Res.* 1994, 99(C5), 10143-10162.
- [4] Evensen G., *Data assimilation: the ensemble Kalman filter*, Springer, New York, 2009.
- [5] Fu J., Jaime Gómez-Hernández J., *J. Hydrol.* 2009, 364, 328-341.
- [6] Gaspari G., Cohn S.E., *Quart. J. Roy. Meteor. Soc.* 1999, 125 (554), 723-757.
- [7] Heemink A.W., Verlaan M., Segers A.J., *Mon. Wea. Rev.* 2001, 129 (7), 1718-1728.
- [8] Horn R.A., The hadamard product. In: *Proc. Symp. Appl. Math*, 1990, pp. 87-169.
- [9] Houtekamer P.L., Mitchell H.L., *Mon. Wea. Rev.* 2001, 129 (1), 123-137.
- [10] Kresic N., *Hydrogeology and Groundwater Modeling*, CRC press, Taylor & Francis group, Washington D. C., 2006.
- [11] Moradkhani H., Sorooshian S., Gupta H.V., Houser P.R., *Adv. Water Resour.* 2005, 28 (2), 135-147.
- [12] Oke P.R., Sakov P., Corney S.P., *Ocean Dynm.* 2007, 57(1), 32-45.
- [13] Petrie R.E., *Localisation in the ensemble Kalman Filter*. University of Reading, Reading, 2008.
- [14] Petrie R. E., Dance S. L., *Weather*, 2010, 65(3), 65-69.
- [15] Reichle R.H., *Adv. Water Resour.* 2008, 31 (11), 1411-1418.
- [16] Reichle R.H., Crow W.T., Keppenne C.L., *Water Resour. Res.* 2008, 44 (3), W03423.
- [17] Safavi H.R., Alijanian M.A., *J. Irrig. Drain. Eng.* 2011, 137 (6), 383-397.
- [18] Sakov P., Oke P.R., *Tellus A.* 2008, 60 (2), 361-371.
- [19] Sun A.Y., Morris A., Mohanty S., *Adv. Water Resour.* 2009, 32 (2), 280-292.
- [20] Van Geer F.C., *Application of Kalman filtering in the analysis and design of groundwater monitoring networks*, Technical University of Delft, Delft, 1987.
- [21] Zhen L., Zöbisch M.A., *Resource use and agricultural sustainability: risks and consequences of intensive cropping in China*, Kassel Univ. Press, Kassel, 2006.

How to cite this article:

Behmanesh J, Bateni M.M. Covariance correction for estimating groundwater level using Deterministic Ensemble Kalman Filter. *J Fundam Appl Sci.* 2015, 7(1), 1-13.