"A space-time ensemble Kalman filter for state and parameter estimation of groundwater transport models"

Jessica Briseño[1], Graciela S. Herrera[2]

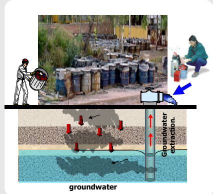
[1] National University of México, Engineering School, México (jevabrir@hotmail.com), [2]National University of México, Geophysics Institute, México, (ghz@geofisica.unam.mx)

Universidad Nacional Autónoma de México

EGU 2010
Vienna, Austria,
2-7 May 2010

1. Introduction

Groundwater monitoring networks are essential in order to estimate groundwater levels and quality parameters and its evolution. Herrera (1998) proposed a method for the optimal design of groundwater quality monitoring networks that involves space and time in a combined form. The method was applied later by Herrera et al (2001) and by Herrera and Pinder (2005). To get the estimates of the contaminant concentration being analyzed, this method uses a space-time ensemble Kalman filter, based on a stochastic flow and transport model.
When the method is applied, it is important that the characteristics of the stochastic model be congruent with field data, but, in general, it is laborious to manually achieve a good match between them.



2. Objective and methodology

The objective is to extend the space-time ensemble Kalman filter proposed by Herrera, to estimate the hydraulic conductivity (K), together with hydraulic head and contaminant concentration, and its application in a synthetic example.

The method has three steps:

1. Given the mean and the semivariogram of the natural logarithm of hydraulic conductivity (ln K), random realizations of this parameter are obtained through two alternatives: Gaussian simulation (SGSim) and Latin Hypercube Sampling method (LHS).
2. The stochastic model is used to produce hydraulic head (h) and contaminant (C) realizations, for each one of the conductivity realizations. With these realization the mean of ln K, h and C are obtained, for h and C, the mean is calculated in space and time, and also the cross covariance matrix h-ln K-C in space and time. The covariance matrix is obtained averaging products of the ln K, h and C realizations on the estimation points and times, and the positions and times with data of the analyzed variables.
3. Finally the ln K, h and C estimate are obtained using the space-time ensemble Kalman filter. The realization mean for each one of the variables is used as the prior space-time estimate for the Kalman filter, and the space-time cross-covariance matrix of h-ln K-C as the prior estimate-error covariance-matrix.

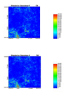
3. Synthetic case study

The synthetic example has a modeling area of 700 x 700 square meters; a triangular mesh model with 702 nodes and 1306 elements is used. A pumping well located in the central part of the study area is considered. For the contaminant transport model, a contaminant source area is present in the western part of the study area.

3.1 Geostatistical analysis for log conductivity and random realization

The semivariogram of ln K is an exponential model.

To obtain the random realizations we used two methods:



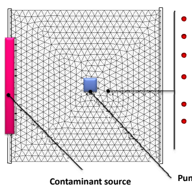
a) Gaussian simulation (SGSim). To obtain model convergence, 3000 realizations of ln K were required using SGSim.

b) Latin Hypercube Sampling method (LHS). To obtain model convergence, 1000 realizations of ln K were required using LHS.

3.2 Groundwater and stochastic flow models

The deterministic flow and transport model was developed using the Princeton Transport Code (PTC, 1993) simulator. The computational grid of the stochastic and deterministic model are the same.

Figure 1. Deterministic flow and transport model

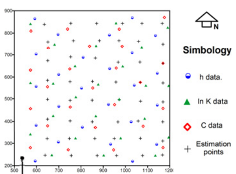


- Specific storage Ss=0.001
- Vertical recharge is not considered.
- Hydraulic conductivity K=1.63 m/day.
- Boundary conditions with specified head.
- 1 layer
- Triangular mesh 702 nodes and 1306 elements

The contaminant source area spreads the contaminant for 12 years (4380 days). The concentration and hydraulic head data are available every 730 days, making a total of 6 periods with data.

The estimation points for ln K, h and C are located on a submesh of the model mesh (same location for h, ln K and C), composed by 56 nodes spread throughout the study area, with an approximately separation of 90 meters between nodes. The nodes of the estimation mesh is where estimations of ln K, h and C will be obtained using the Kalman filter.

Figure 2. Stochastic flow and transport model



- Parameter were obtained from the deterministic flow model, exception K.
- K is a random field generated by :
  * 3000 realizations from SGSim and
  * 1000 realizations from LHS
- 56 nodes were the estimations of ln K, h and C will be obtained.
- ln K possible data points (53 ln K values)
- h possible data points (53 h values)
- C possible data points (56 C values)

3.3 Parameter estimation using the Kalman filter

The Kalman filter requires a prior ln K, h and C estimates (mean ln K, h and C) and a cross covariance matrix ln K-h-C of the estimate errors. In this study, the average of the 3000 ln K (SGSim) and the average of the 1000 ln K (LHS) realizations and the cross covariance matrix ln K-h-C, obtained from the stochastic model realizations, were used to this end.

Six study cases (Table 1) were established to estimate ln K, h and C using different data sets and a prior space-time covariance matrix calculated with SGSim and LHS realizations.

Table 1. Synthetic examples

A priori covariance Matrix h-lnK-C	Study case	Input data
Obtained whit SGSim realizations	1.1	25 data h
	1.2	25 data C
	1.3	25 data h & 25 data C
Obtained whit LHS realizations	2.1	25 data h
	2.2	25 data C
	2.3	25 data h & 25 data C

4. Results

Figure 3, 4 and 5 shows the ln K, h and C prior estimates, the ln K, h and C realizations and the ln K, h and C estimates in the study case 1.3 and 2.3. From a simple analysis of the graphs of this figure, it is difficult to determine the magnitude of the estimates errors obtained with the Kalman filter.

For a more detailed analysis of the results the mean error (ME) and root mean square error (RMSE) were calculated. The estimate error, e_i, is analyzed. This error is calculated as the difference between the data of the ln K, h and C realizations (SGSim or LHS) and the ln K, h and C estimated with the Kalman filter at the points of the estimation mesh with the different case studies.

$$ME = \frac{1}{n}\sum_{i=1}^{n} e_i$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i)^2}$$

The locations where estimations of Ln K, h and C will be obtained are associated with nodes, which we referred to as the estimation mesh shown in Figure 2. This mesh is composed by 48 estimation point distributed throughout the study area (same localization for ln K, h and C).
The results of the ME and MSE and the reduction percentage at different study cases are presented in Table 2 (only for time 6) and the graphs in figure 6 (for all times).

Figure 3. h and C estimation of the study case 1.3

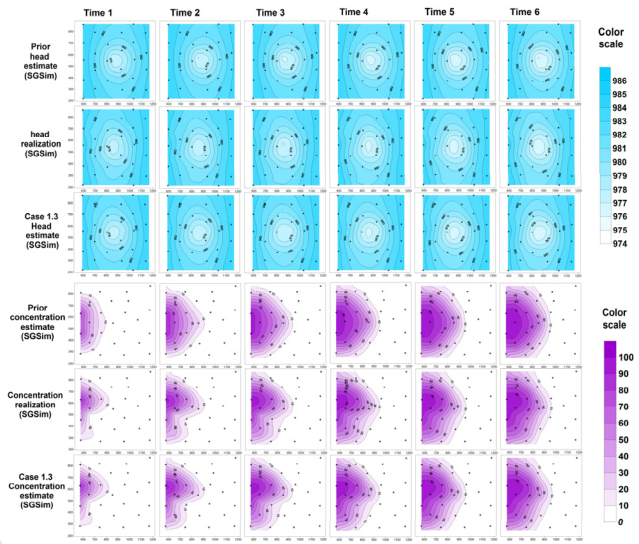


Figure 4. h and C estimation of the study case 2.3

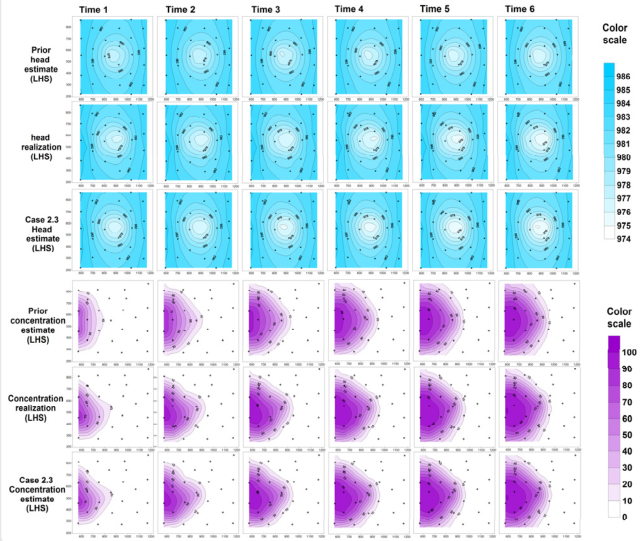


Figure 6. ME and RMSE graph of the Ln K, h and C estimates

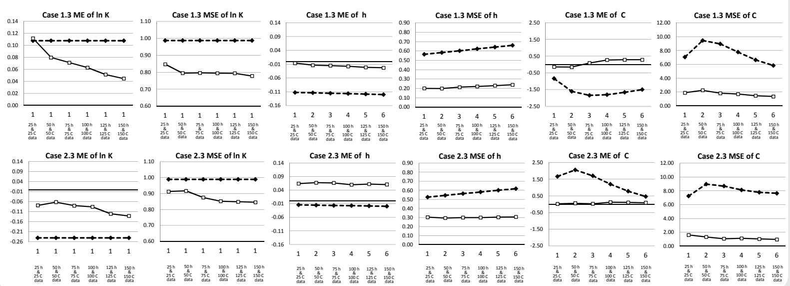
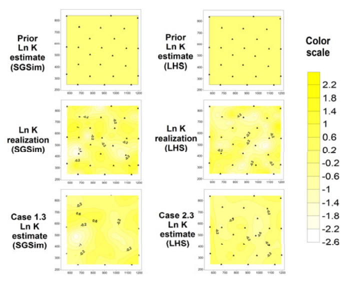


Table 2. Parameter estimation errors.

Study case		Parameter error	ME (meters)	%	MSE (meters)	%
Covariance SGSim	Base	Prior Ln K error	0.1077	100%	0.9872	100%
	1.1	Ln K (h data)	0.0445	58.7%	0.8755	11.3%
	1.2	Ln K (C data)	0.1077	0.0%	0.8601	12.9%
	1.3	Ln K (h and C data)	0.0445	58.7%	0.7779	21.2%
	Base	Prior h error	-0.1196	100%	0.6605	100%
	1.1	h (h data)	-0.0077	93.6%	0.2549	61.4%
	1.2	h (C data)	0.0634	153.0%	0.4663	29.4%
	1.3	h (h and C data)	-0.0228	80.9%	0.238	64.0%
	Base	Prior C error	-1.5082	100%	5.8291	100%
	1.1	C (h data)	1.1725	177.7%	5.5066	5.5%
	1.2	C (C data)	0.1343	108.9%	1.312	77.5%
	1.3	C (h and C data)	0.2862	119.0%	1.3602	76.7%
Covariance LHS	Base	Prior Ln K error	-0.2404	100%	0.9886	100%
	2.1	Ln K (h data)	-0.1281	46.7%	0.9041	8.5%
	2.2	Ln K (C data)	-0.2071	13.9%	0.9122	7.7%
	2.3	Ln K (h and C data)	-0.1312	45.4%	0.8453	14.5%
	Base	Prior h error	-0.0193	100%	0.6189	100%
	2.1	h (h data)	0.0698	461.7%	0.3475	43.9%
	2.2	h (C data)	0.0208	207.8%	0.4665	24.6%
	2.3	h (h and C data)	0.0606	414.0%	0.3059	50.6%
	Base	Prior C error	0.463	100%	7.6435	100%
	2.1	C (h data)	-1.3369	388.7%	3.7126	51.4%
	2.2	C (C data)	0.2432	47.5%	1.3024	83.0%
	2.3	C (h and C data)	0.0844	81.8%	0.9579	87.5%

Figure 5. Ln K estimation



5. Conclusion

- The results analysis was done through the mean error, root mean square error, initial and final estimation maps of h, ln K and C at each time, and the initial and final variance maps of h, ln K and C.
- To obtain stochastic model convergence, in a PC Pentium 4-800 MHz with 4 gigabytes of RAM, 16 hours were required to run 1000 simulations of ln K using LHS, and took three times longer (48 hours) to run 3000 simulations of ln K using SGSim.
- The results show that for both alternatives, the Kalman filter estimates for h, ln K and C using h and C data, have errors which magnitudes decrease as data is added.
- With both methods the error is comparative, but it is important to note that the percentage reduction in MSE with respect to priori MSE for Ln K and h estimate, in cases where the estimation is performed with h, C and h and C data, the percentage reduction was greater using SGSim to LHS. However, the percentage reduction for C estimates using LHS than SGSim was greater.

References

BRISEÑO, J. and HERRERA, G, (2007). Space-time Design of Piezometrics Groundwater Monitoring Networks, Conference on Water Pollution in natural Porous media at different scales. Assessment of fate, impact and indicators. WAPO2. Barcelona, España. Abril, 2007.
DEUTSCH, C. V., and JOURNEL A.G., GSLIB (1998). Geostatistical Software Library and User's Guide, Oxford Univ. Press, New York.
HERRERA, G. S.(1998), Cost Effective Groundwater Quality Sampling Network Design. Ph. D. thesis, University of Vermont, Burlington, Vermont, 172 pp.
HERRERA G., GUARNACCIA J., PINDER G. and SIMUTA R.(2001),"Diseño de redes de monitoreo de la calidad del agua subterránea eficientes", Proceedings of the 2001 International Symposium on Environmental Hydraulics, Arizona, U.S.A.
HERRERA G. S. and PINDER G.F. (2005), Space-time optimization of groundwater quality sampling networks Water Resour. Res., Vol. 41, No. 12, W12407, 10.1029/2004WR003626.
PTC. BABU, D., G. F. PINDER, A. NIEMI, D. P. AHLFELD, and S. A. STOTHOOFF, (1993). Chemical transport by three dimensional groundwater flows (Princeton Transport Code), Rep. 84-WR-3, Princeton University, Department of Civil Engineering. Princeton, N.J.
SIMUTA R. and HERRERA G. (20xx). Convergence analysis for Latin-Hypercube Lattice-Sample Selection Strategies for 3D correlated random hydraulic-conductivity fields. To be submitted to Geofísica Internacional.
ZHANK Y. and PINDER G., (2003) Latin hypercube lattice sample selection strategy for correlated random hydraulic conductivity fields, Water Resources Research, Vol. 39, No. 8, 1226, doi:10.1029/2002WR001822.