

Interactive comment on “Estimation of heterogeneous aquifer parameters using centralized and decentralized fusion of hydraulic tomography data from multiple pumping tests” by A. H. Alzraiee et. al

Anonymous Referee #1

Received and published: 30 April 2014

This is an interesting paper that contributes in improving the characterization of the hydraulic properties of an aquifer using hydraulic tomography data from multiple pumping tests. The ensemble Kalman filter is employed for the assimilation and inversion of the temporal moments of the impulse response function, avoiding the computational burden of transient Monte Carlo simulations. In particular the paper aims to understand which scheme is more efficient for the inversion of the hydraulic head data collected from multiple pumping tests: the direct inversion of the whole set of data (centralized fusion, CF), or the separate inversion of the data of each pumping test, combined with

C1205

the Generalized Millman Formula to fuse together the results of the single inversions (decentralized fusion, DF). The question and the proposed methodologies are of interest for the readers of HESS. The numerical simulations considered in the paper are effective in showing that the CF scheme consistently outperforms the DF scheme. However, I recommend adding more results to complete the comparison between the two schemes and corroborate the conclusions. For example, the equivalent of Table 4 for the DF is necessary to have a quick look at the results of the comparison. For the same reason, Figure 6 and 7 should be presented also for the DF.

A second objective of the paper is to evaluate the efficiency of the CF in estimating the geostatistical parameters of the random fields Y and Z . Although the identification of geostatistical parameters is a crucial question in real applications, I find that the addition of this part makes the reading of the paper more difficult and the main focus of the work is lost. This is partially due to the fact that this second objective is not presented in the introduction of the paper. Further investigation is necessary to understand when the procedure proposed in section 2.5 reduces the uncertainty on the geostatistical parameters (are more observations necessary? Is the procedure effective with different configuration of the true parameter?). In my opinion, this is the material for another (interesting) paper. The author should either discard sections 2.5 and 4.4 or rewrite these sections focusing on the comparison between CF and DF.

Finally, the author states that the novel localized DF is essential for the inversion of the matrix C . However, only few lines are dedicated to the description of the localized inversion. More details on the construction of the localized inverse matrix of C will increase the significance and novelty of the paper.

I think that the work might be publishable after major revisions. In the following I will try to outline where and how the manuscript can be improved.

MAJOR COMMENTS

1. The presentation of the geostatistical parameter estimation problem is missing in

C1206

the introduction. You fail to provide an overview of existing methodologies that estimate the mean, variance, and correlation length of the spatial distributions of Y and Z . This needs to be done to prove the novelty of the methodology presented in Section 2.5.

2. As stated in section 2.3, the KF and EnKF are typically applied in transient problems where the update step of equation (9) is combined with the forecast step (Kolmogorov equation). However, the methodology presented in sections 2.3.1-2.3.2 employs only the update step of EnKF. In fact, EnKF is not applied to time dependent equations. Equations (7) and (8) constitute the observation operator, that is the link between the parameters and the assimilated data. For these reasons the term 'forecast' adopted in sections 2.3.1 and 2.3.2 is misleading. The Section 2.3 should state clearly that there is no forecast step (in time) in the proposed methodology, and that you indicate with forecast the posterior distribution of the moments given the prior distribution of the parameters Y and Z .
3. I like the idea of proposing different formulations of the state vector for the EnKF. Since you are interested in updating only the parameters, is it possible to adopt as forecast matrix only the matrix Y ? This drastically reduces the dimensions of the matrices involved in the update for both DF and CF. In this case the observation operator is nonlinear, $m_0 = h(Y)$ and a nonlinear form of the EnKF should be adopted.
4. End of page 4179: the description of the novel localized fusion algorithm for the inversion of matrix C is not clear. In my opinion the relation between the components of matrix C and the cells of the domain is not straightforward. A detailed description of this novel algorithm is necessary to understand and reproduce the DF method (maybe add an appendix with the algorithm).
5. None of the performance metrics is based on the collection of the observations,
C1207

which are the only data available in real application. I recommend adding a performance metric based on the error between the observed hydraulic heads and the heads estimated with the posterior distribution of the parameters.

6. For what concern the presentation of the numerical results, why do you restrict the comparison between DF and CF only to the formulations A and E? The comparison on all the formulations will be useful to corroborate the results that the formulations A and E are better than B, C, and D, and that CF always outperform the DF. Why for DF do you only compute the correlation coefficient? The author should present the analogous of Table 4 also for the DF.

MINOR COMMENTS

1. In Sections 2.1 and 2.2 the temporal moments of the impulse response function (IRF) are confused with the moments of hydraulic head: from Li et al. (2005), the moments m_k used in equations (2-8) are the moments of the IRF. Please correct the text accordingly.
2. Pages 4171-4172: please define explicitly the relation between f (used in equation (5)), the pumping rates Q_i and the well location x_w (used in equation (6)). Also define x_w after its first occurrence (equation (6)) and not later.
3. Page 4173: in equation (9) the normalization term should be $p(m, l)$ and not $p(m|l)$.
4. Page 4177: to better understand equation (14), you should add $= \sum_i^{N_p} w_i \tilde{Y}_u^i$.
5. Page 4178: in equation (15), the ensemble matrix Y should be replaced with the true field Y_{true} .
6. Page 4182, line19: discard the second closed parenthesis.

7. Page 4182, line 24: does the system reach the steady state in 10 days? This is a crucial for the computation of the temporal moments with equations (3) and (4).
8. Page 4183: lines 10-12 are a repetition of lines 19-21 of page 4174.
9. Page 4183: lines 20-21 are a repetition of lines 1-2 of the same page. Lines 21-25 should be moved after line 2. Which are the initial conditions for the generation of the measurements?
10. Page 4184: The classical definition of the correlation coefficient is different from the definition in equation (25). Why do you chose a different formula for the correlation coefficient?
11. Figure 2: only 25 pumping wells are depicted in figure 2 (not 36).
12. Figure 3: the title of panel (d) should be 'Estimated Z' '.
13. Figure 6 (and discussion). For the values of α considered, the expected value of K varies only of one order of magnitude, while the expected value of S varies of several order of magnitude. Can this be the cause of the different sensitivity analysis between K and S ? In my opinion, an additive perturbation of the parameter μ is more adequate than a multiplicative perturbation.
14. Figure 7: why the prior distributions of σ are different in panels (b) and (e)? Why the true values on the mean μ are different from the values reported in Table 2?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 4163, 2014.