Hydrol. Earth Syst. Sci. Discuss., 11, C2197–C2209, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C2197/2014/

Hydrology and Earth System Sciences
Discussions

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

A. H. Alzraiee et. al

ayman.alzraiee@gmail.com

Received and published: 5 July 2014

[12]article latexsym color

RESPONSE TO COMMENTS FROM REVIEWER # 1

Note: We use text with black color for reviewer comments (RC), blue color for author

comments (AC), and red color for modified text in the manuscript.

Response to General comments

[RC#1]: 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 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.

[AC#1]: [We would like to thank Reviewer #1 for his/her insightful comments about our paper. We incorporated all of these comments and reported our responses in this letter as well as in two response letters for two other reviewers. With regard to the addition of the equivalent of Table 4 for the DF, please see response to comment RC#10.

[RC#2]: For the same reason, Figure 6 and 7 should be presented also for the DF.

HESSD

11, C2197–C2209, 2014

[AC#2]:[As recommended in Comment RC#3, and as indicated by Reviewer #2, the estimation of geostatistical parameters within the inversion requires further, more thorough analyses, which we plan on submitting for peer review in a separate paper. We agree with Reviewer #1 that geostatistical parameter estimation should be investigated in a second separate paper. Accordingly, the analyses included in Sections 2.5, 4.3, 4.4 of our original submission, which contained Figures 6 and 7, have been removed from the revised version of the manuscript.]

[RC#3]: 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.

[AC#3] [We agree with the reviewer. The estimation of geostatistical parameters is worth of further investigation in a separate paper. We believe that it should be more useful to separate the estimation of geostatistical parameters. After careful thoughts, we have thus decided to remove the sections related to the application of the CF method for estimating geostatistical model parameters. This allows also for creating the space necessary to expand the comparison between the CF and DF perfomances as well as include further analyses required by other two reviewers. Parts removed include all sentences in the abstract and conclusions related to geostatistical parameters

HESSD

11, C2197–C2209, 2014

estimation, section 2.5, 4.3, and 4.4, and Figures 6 and 7. New sections and figures are added to focus on the comparison between DF and CF. Further details about these new sections and figures are reported in the following as we respond to the reviewer specific comments.]

[RC#4] 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.

[AC#4] [We agree with the reviewer. We have added a new section (Section 2.3.4) that describes the localization methodology. In addition, we added a new schematic figure (Figure.2 in the revised manuscript) to illustrate the localizations algorithm. Please refer to Comment RC#8 in the response to specific comments for more details.]

Response to specific comments

[RC#5] The presentation of the geostatistical parameter estimation problem is missing in 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.

[AC#5] [We agree with the reviewer. We believe that a full exposition and investigation of the geostatistical parameter estimation requires more tests and extensive literature review. Naturally, this requires further analyses and tests that cannot be completed

HESSD

11, C2197–C2209, 2014

within the time allowed for the revision. Per the reviewer suggestion (RC#3) we have decided to restrict this work to the comparison between the CF and DF approaches. The geostatistical parameter estimation problem will be covered extensively in a future peer review submission.]

[RC#6] 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, which 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.

[AC#6][We agree with the reviewer that there is no forward-in-time forecast after temporal integration of the flow PDE. The text was modified to clarify this point. Please see modified text in the second paragraph in Section 2.3, which directly indicates that no forward-in-time forecast is made. The modified texts read as:

"In the classical implementation of KF, the data assimilation of state follows a two-stage forecast-update process. In the forecast stage, a forward in time prediction of the current state, along with its error covariance is first made. The forecast state is then updated as field measurements become available. In this work, the inversion problem is reduced to a time-independent inversion problem, which means that the forecast stage does not include any forward in-time prediction. That is to say, the forecast stage is limited to the solution of the equivalent steady state groundwater problems expressed by Eq. (7)and (8)".

HESSD

11, C2197–C2209, 2014

[RC#7] 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, m0 = h(Y) and a nonlinear form of the EnKF should be adopted.

[AC#7][Yes, this is possible. The forecast matrix would then be X = [Y, M0'], where M0' includes the temporal moments at observation locations only. While this would certainly reduce the size of the forecast matrix, it would not significantly reduce the CPU time as the heavy computation required by the EnKF occurs when we invert the matrix (HPH'+R) in Equation (13), which has the same size when M0 includes temporal moments at all nodes or when we M0' includes moments at observation wells only. Specifically, if we invert M0 as in formulation A, the size of (HPH'+R) is 36 by 36 when M0 includes all nodes or when M0 include moments at observation locations only. In both cases, the matrix is computationally easy to invert.]

[RC#8] 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).

[AC#8][To improve the description of the localized fusion algorithm, we have added a new subsection (subsection 2.3.4 in the revised manuscript) to illustrate the localized inversion. Also, we added a new figure (Figure 3) to better illustrate the algorithm. The new section reads as :

Localization of Decentralized Fusion: The inversion of the matrix $\mathbf{C}$ in Eq. (21) constitutes the most intensive part of the GMF. In HT, it is typically required to estimate

HESSD

11, C2197–C2209, 2014

hydrogeological parameters at high resolution, which often renders the GMF approach computationally very intensive. To circumvent this obstacle, we propose the following novel localized fusion algorithm.

In essence, instead of computing Eq. (21) for all the cells in the domain at once, the fused estimate at any given cell is computed by considering only a circular block of cells within a specified radius around the cell to be fused (Fig. 3). The localized DF algorithm visits each cell within the domain sequentially or in parallel and fuse these circular blocks. The resulting fused estimate for the cell at the center is returned, and the algorithm moves to next cell. Indicating as n' ($< n$) the number of grid cells within a specified distance from the cell of interest, the resulting size for the "local" matrices in Eq. (21) is: $n' \times n' N_p$ for $\mathbf{B}$, and $n' N_p \times n' N_p$ for $\mathbf{C}$.

The implicit assumption behind this method is that neighboring cells will have the majority of influence on the estimation. The GMF localization is meant to improve the computational efficiency in two ways. First, the inversion of matrices $\mathbf{C}$ of smaller size is less CPU intensive; second, the fusion algorithm can be directly parallelized on multi-core processors.

[RC#9] None of the performance metrics is based on the collection of the observations, 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.

[AC9#][We have added two figures, one for CF and another for DF, illustrating the simulated head using estimated parameters and heads simulated using true parameters. The statistics of errors are computed and discussed in the results. Figure 5 and Figure 7 show the true heads versus heads simulated using estimated parameters for CF and DF respectively. The performance statistics are shown panels (a) in both figures.

HESSD

11, C2197–C2209, 2014

These figures are also discussed in paragraph 8 in section 4.1, which now reads

"Plots in Figure 5 compare the simulated heads using the estimated Y and Z fields using CF methods with heads obtained by simulating true parameter fields. Figure 5a shows a scatter plot of simulated heads versus reference heads resulting from the five pumping tests and for heads observed at 36 observation wells. The performance statistics L_1, L_2, and r are 0.09, 0.015, and 0.998, respectively, indicating fairly good performance of the inversion method. Figure 5b Figure 5f show one sample of hydraulic head hydrographs resulting from the five pumping test at observation well 15 (See Figure 2), which is located approximately in the middle of the simulated domain. The figures show a general agreement between observed and simulated head hydrographs."

and in paragraph 6 in section 4.1, which now reads

"Figure 8 compares hydraulic heads obtained by simulating estimated Y and Z fields using DF method with observed hydraulic heads. Comparing the performance criteria of DF method, shown in Figure 8a, with performance criteria of CF method, shown in Figure 5a, shows that the performance of CF method is better than that of DF method. The hydraulic heads at observation well No. 15 are plotted in Figures 8b to 8f for the five pumping tests. A general agreement can be observed between simulated heads and true ones."

[RC#10] 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 then 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.

HESSD

11, C2197–C2209, 2014

[**AC#10**] [To address the reviewer comment, a new table (Table 4 in the revised manuscript) has been added to summarize the performance of the decentralized fusion. Thus, the results and discussion section were modified to present the new table. The reason this table was not added to the original manuscript is that the comparative performance is independent from inversion method. that is to say, if Formulation A is better than Formulation B in CF method, then the same can be seen using DF method. The modified text reads as]"The performance criteria for DF method using different formulations are summarized in Table 4. Comparing performance criteria for DF method shown in Table 4 with performance criteria for CF in Table 3, reveals that the performance of different formulations is independent from the fusion method used. For example, formulation A outperforms formulations B and C in estimating Y field for both CF and DF methods, and formulation E outperforms formulation D in estimating Z for both CF and DF."

Response To Minor Comments

[RC#11] 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 *mk* used in equations (2-8) are the moments of the IRF. Please correct the text accordingly.

[AC#11] [We agree with the reviewer. We modified the manuscript, whenever it is applicable, to clearly indicate that we assimilate the temporal moments of the Impulse Response Function of the drawdown instead of the temporal moments of the drawdown data.]

[RC12] Pages 4171-4172: please define explicitly the relation between f (used in equa-

HESSD

11, C2197–C2209, 2014

tion (5)), the pumping rates Qi and the well location xw (used in equation (6)). Also define xw after its first occurrence (equation (6)) and not later.

[AC12][Changes suggested are made in equation 5, and in the second paragraph in Section 2.2. The modified texts read as "For a unit impulse extraction $Q(x;t) = \delta(x_{\mathrm{w}})$ at location x_{w},..."]

[RC13] Page 4173: in equation (9) the normalization term should be $p(m,l)$ and not *p(m/l).*

[AC13] [We corrected equation 9 as suggested by reviewer #1 and reviewer #2. The corrected equation reads as follows: $p(\phi|m,I) = \frac{p(m|\phi,I)p(\phi,I)}{p(m,I)}$]

[RC14] Page 4177: to better understand equation (14), you should add $\sum_{i=1}^{N_p} \mathbf{w}_i \hat{\mathbf{Y}}_u^i$

[AC14][We modified the equation as suggested by the reviewer.The modified equation reads as follows $\tilde{Y} = \mathbf{W^T} \cdot \hat{Y}_{\mathrm{u}}^{1:N_{\mathrm{p}}} = \sum_{i=1}^{N_p} \mathbf{w}_i \hat{Y}_u^i$]

[RC15] 5. Page 4178: in equation (15), the ensemble matrix Y should be replaced with the true field Ytrue.

[AC15][We would like to point out that Y is already refer to the true Y field.]

HESSD

11, C2197–C2209, 2014

[RC16] 6. Page 4182, line19: discard the second closed parenthesis.

[AC16] We made the changes required.

[RC17] 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).

[AC17][Yes, after 10 days the system is practically at steady state. In preliminary tests we verified that steady state heads occur at all observation wells. The newly added figures (Figure (5) and Figure (8)) show the drawdown at observation well 15 reaches steady state level in about four days. It is worth noting here that when the pumping is not continuous or temporally variable, the IRF function cannot be obtained using equations (3) and (4). Fitting a parametric function for the IRF should be applied in this case.]

[RC18] 8. Page 4183: lines 10-12 are a repetition of lines 19-21 of page 4174.

[AC18][We agree with the reviewer. We remove the repetition in Section 3.1]

[RC19] 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?

[AC19][We agree with the reviewer. We remove the repetition in Section 3.1]

HESSD

11, C2197–C2209, 2014

[RC20] 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?

[AC20][We corrected equation 25 to calculate the correlation between two images (2D fields). Essentially, this equation is the classical Pearson's correlation coefficient, but slightly modified to find the correlation of two images instead of two vectors. Certainly, those 2D images can be reshaped from 2D to 1D vector. In both cases, the correlation values are the same.

[RC21] 11. Figure 2: only 25 pumping wells are depicted in figure 2 (not 36).

[AC21][Figure 2 (Figure 3 in the revised manuscript) is corrected. We added pumping wells numbers and observation well numbers.]

[RC22] 12. Figure 3: the title of panel (d) should be 'Estimated Z'.

[AC22][The Figure is corrected (See Figure 4).]

[RC23] 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.

[AC23][This section is removed as suggested by the reviewer.]

HESSD

11, C2197–C2209, 2014

[RC24] 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?

[AC24][This section is removed as suggested by the reviewer.]

HESSD

11, C2197–C2209, 2014