Interactive comment on "Is inversion based high resolution characterization of spatially heterogeneous river bed hydraulic conductivity needed and possible?" *by* W. Kurtz et al.

W. Kurtz et al.

w.kurtz@fz-juelich.de

Dear referee,
we thank you for your thoughtful response to our manuscript and for the positive feedback on our work. Here are our replies to your comments:

The main concern for me about this paper is the following. When parameters and state variables are updated simultaneously in each assimilation step with the updated state variable values at t-1 as the initial state variable values at t, the updated

parameters and state variables are only consistent with each other for linear model. The simulation model used here, which predicts the head values from t-1 to t, is a non-linear model. That means updated parameters and updated state variables are NOT consistent, especially when heterogeneity is strong, which is reflected in the fact that the updated state variables at t is different from these predicted state variables at t based on the updated parameters at t. See Thulin et al. (2007) and Gu and Oliver (2006) for references. Therefore, only use RMSEh as the performance assessment criterion is not appropriate in that it may not fully represent the estimation performance of hydraulic conductivity or leakage coefficient in this case. RMSEL or RMSELnL is also necessary to be used directly when hydraulic conductivity estimation is the main task through EnKF. In Particular, the paper points out that when less measurements are available (with only 10 measurement points), the estimation of leakage coefficients can be comparative to the case when more measurements are available (with 100 measurement points), and eventually results in that high resolution representation of river bed hydraulic conductivity is still beneficial. That is an interesting point, but it may be more appropriately justified by RMSELnL instead of RMSEh . Considering that the author shows log10L values on Fig. 10 and 11, I don't see why not use RMSELnL to justify the conclusion.
We agree that the consistency between states and parameters is always an issue when EnKF is used to update model parameters. In Hendricks Franssen & Kinzelbach (2008, WRR) the Restart-EnKF was compared with the traditional EnKF and only minor differences were found for these two approaches. However, for more non-linear systems (e.g., unsaturated conditions, higher degree of heterogeneity) this issue might play a bigger role. In our case, the updating frequency is set to 10 time steps which allows the model states to synchronize to the updated model parameters between the updating cycles. We think that Figures 10, 11 and 14 already give good insights in the way EnKF updates the different parameter ensembles. But we also agree that using RMSE(L) for a more quantitative assessment of updated leakage coefficients would be beneficial

especially for comparing simulations with different amounts of observation data. Therefore, we intend to provide a table with RMSE(L) values for the final parameter ensembles for the different scenarios.

What's more, I think the author should list at least some representative RMSE values (if listing all of them for different reference fields and scenarios is tedious) when making conclusions based on them, such as Line 5 on Page 5848, Line 2-5 and Line 21-22 on Page 5849. It would give more quantitative sense for the readers in that way.
We intend to provide a table that gives information on the average RMSE(h) for the second half of the simulation period for the different references and scenarios.

1. Line 5 on page 5833: It might be better to make it clear that "classical approaches" is actually "classical zonation approaches", because sometimes "classical approach" reminds readers of non-sequential inverse method.
We agree that the term "classical approaches" is a bit misleading in this context. We will replace this term with "zonation approaches".

2. Line 8-10 on P5838: How EnKF can improve the prediction stated by the author is confusing for me. Based on my understanding, the prediction is improved by adding an optimal weighted "innovation term" which is the difference of predicted and observation data. It may not be that "measurement errors and the uncertainty of model predictions are optimally weighted", but use "measurement errors and the uncertainty of model predictions" to weight the innovation term.
We agree that our description is not 100% clear. We will change this part according to your recommendation.

3. Line 7 on P5839: It may be better to explicitly characterize the normal distribution of perturbations, $\epsilon_i \sim N(0, R)$.
We will adapt the description of the perturbation vector accordingly.

4. Line 19 on 5839: Damping factor has similar effect with adding an extra model error in addition to the observation error. It is usually used as a strategy to avoid the ensemble inbreeding. The selection of its value affects the estimated results. Is there any basis to choose this value as 0.1 in this case?
The value for the damping factor of 0.1 is based on findings from Hendricks-Franssen & Kinzelbach (2008, WRR). A similar (optimal) value for the damping factor was found by Kurtz et al. (2012, WRR). We will include these references in the explanation of the damping factor.

5. Line 6 on Page 5847: The analysis about Z5 is not so accurate because we can see the net flux can be overestimated for reference I, VII and others. A more complete and accurate result analysis may be needed here.
We agree that the description of net exchange is a bit limited. What one can see is that the over/underestimation of net exchange for Z5 is usually less severe than for Z3 and Z2. We will make this point more concrete in the presentation of the results for net exchange.