Hydrol. Earth Syst. Sci. Discuss., 12, C1797–C1805, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C1797/2015/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance" *by* J. Rasmussen et al.

J. Rasmussen et al.

jr@geo.ku.dk

Received and published: 27 May 2015

Response to Anonymous referee #1

The authors would like to extend our gratitude to the referee for the detailed comments and suggestions. We have here tried to answer the referee's comments point-by-point.

Abstract:

- Comment: As far as I understood the study, no real observations were used for

assimilation. In the abstract, please point out that the study uses synthetic data for assimilation.

Response: We agree that the synthetic test used in the study should be emphasized in the abstract.

Modifications: The following has been added to the abstract: "The study is conducted using synthetic data only".

- Comment: Is the term "local analysis localization" commonly used in the data assimilation context? If not, I would suggest renaming this localization method throughout the paper, for example referring to it as "distance based localization" or similar. I think that would better describe the method.

Response: We agree; "local analysis" is indeed more rarely used and less descriptive than "distance based".

Modifications: "local analysis localization" has been changed to "distance-based localization" throughout the paper.

- Comment: Page 2268, Line 26: Should it say "on one part of the hydrological cycle"?

Modifications: Changed.

- Comment: Page 2270, Line 22-24:The first objective seems a bit too specific to the site and experiment. I would assume that the optimal ensemble size varies depending on location, data availability, model, type of data (synthetic observations or real-world observations) and other factors. Perhaps remove the last part of the sentence: "with the objective of assessing an optimal ensemble size".

Response: We agree that the formulation used is too general. The optimal ensemble size does indeed vary strongly depending on a variety of factors.

Modifications: "with the objective of assessing an optimal ensemble size" removed.

- Comment: Page 2271, Methods: I miss a description of the model input data and for which time period the simulations were performed in the methods section.

Response: We agree that information on the model input and simulation span should be presented in the paper.

Modifications: The following has been added to section 2.2.2: “The model simulations span five years, from 1968 to 1972 (both included). The first two years is the spin up, where the model is allowed to stabilize and the ensemble of states is allowed to develop a spread without the assimilation of observations. The following three years, observations are assimilated using the filter. However, only the last two years, 1971 – 1972 are used for evaluating the filter performance.

Applied precipitation in the model is based on measured daily precipitation from nine gauges located in the catchment. The measured data is extrapolated to the model domain using Thiessen polygons, thus applying the measured precipitation to the model grid points located closest to the measuring location. Spatially uniform daily values of potential ET are specified. “

- Comment: Page 2272, Line 21: Change to “mica sand”.

Modifications: Changed.

- Comment: Page 2273, Line 6: Change to “saturated hydraulic conductivity”.

Modifications: Changed.

- Comment: Page 2274, Eq. 2: I think a transpose might be missing in the equation. According to Harlim and Hunt the equation is C = (YËĘb)ËĘTRËĘ(-1). Please check.

Response: That is correct.

Modifications: Changed

- Comment: Page 2274, Eq. 3: The matrix C has dimensions n x k and the matrix

YËĘb has dimensions k x n. Their matrix product CYËĘb should have dimensions n x n which means that the identity matrix I also must have dimensions n x n. I also suspect that the factor (k-1) in Eq. 3 actually should be (n-1) according to the Harlim and Hunt paper. Please check which is correct.

Response: Correct; the error arose due to the difference in notation between this paper and Harlim and Hunt (2005).

Modifications: The notation has been corrected and changed to match that of Harlim and Hunt (2005).

- Comment: Page 2274-2275, Eq. 4 to 6:

It seems as if the dimensions of the matrices are stated wrongly. I think most of the k’s should be n’s instead. Please check and correct if that is the case. I would suggest using the exact same notation as given in the Harlim and Hunt paper. Please excuse me if I have read and interpreted the equations wrongly.

Response: please see above.

- Comment: Page 2275, Line 7: There seems not to be any study by Anderson from 2001. The Bishop and Hodyss from 2007 is not included in the reference list, but one from 2009 instead. Please check all references throughout the paper carefully.

Response: Indeed, the papers that should have been referred to are Anderson (2007) and Bishop and Hodyss (2009).

Modifications: The references have been corrected, and all the other references of the paper have been checked as well.

- Comment: Page 2275, Line 6 to Page 2275, Line 9: I think the adaptive localization method should be described in much more detail since it is one of the main objectives of the study. I was unable to understand how the localization was performed from the current description.

Response: We realize that the localization algorithm used is conceptually difficult. However, we believe that the current description already is rather detailed and takes up much space. We have reconsidered the formulation and made a few clarifications. Further, we reference three studies which contain additional information that compliments the description in the manuscript.

Modifications: None.

- Comment: Page 2279, Line 1-5: The sentences on these lines sound awkward. Please rephrase.

Modifications: The sentences have been changed to the following: The choice of parameters to estimate was based on a sensitivity analysis which was performed using the AUTOCAL software (Madsen, 2003; MIKE by DHI, 2014). The included parameters were those with scaled sensitivities (Table 1) of 1% or more of the sensitivity of the most sensitive parameter.

- Comment: Page 2279, Lines 19-22: Why were the parameters transformed?

Response: The parameters were transformed to account for the expected parameter uncertainty, which may span several decades.

Modifications: Changed to "Four of the five estimated parameters, the hydraulic conductivities of meltwater sand and quaternary sand, as well as the stream bed leakage coefficient and the drain time constant were transformed logarithmically in the filter as the expected parameter uncertainty is expected to span several decades. . .".

- Comment: Page 2279, Line 26 to Page 2280, Line 4: First, please make clear in the abstract, introduction, figure describing the field site and conclusions that the study does not use real-world observations for assimilation but synthetic ones. Second, I found it difficult to understand this paragraph. Weren't observations simply generated by running the model once using perturbed parameters and states. In Table 2, were the "Initial values" used for the base model? Please specify.

Response: We agree that the formulation used is unclear. The methodology used is the common synthetic test setup, as described by the reviewer.

Modifications: The section has been changed to: "This study uses a twin test approach in which observations are generated by extracting selected state values from a forward run ("True" model), and adding normally distributed noise to emulate typical real world observation errors. For comparison, the results of a model similar to the true model, but with perturbed initial parameter values, will be shown. This shows the states of the model if no state updating or parameter estimation is applied, and will in the following be denoted the base model".

"Initial value" in table 2 has been changed to base value, to avoid confusion.

- Comment: Page 2280, Line 6-9: What impact on the result does the omission of spatially correlated noise have? Please discuss this issue in detail in the paper. Also discuss how covariances can arise between adjacent grid cells even though spatially correlated noise was omitted.

Modifications: The following is added to section 2.5: "The noise added to both forcings and parameters is based on experience with uncertainty in real data and parameters. The magnitude of parameter uncertainty is for many parameters well understood, as sensitivity analysis and calibration has been performed on several hydrological models, including the Karup catchment model (Refsgaard, 1997). Correlation in parameter values is only included where this is widely accepted to exist and easily quantifiable (i.e. horizontal and vertical hydraulic conductivity). The noise added to the forcings represents a significant simplification of the understanding of forcing uncertainty, which is likely to be highly correlated both temporally and spatially. A better description of the correlation in forcing noise would most likely have resulted in better description of the error covariances, which currently is determined based on the difference in model behavior between the ensemble members, and thereby better filter performance in terms of distributing the state updates. However, spatially and temporally correlated

ensembles of forcings are difficult to generate, and outside the scope of this study".

Comment: Page 2283, Line 1 and 7: Change to Fig. 3.

Modifications: Changed.

- Comment: Figure 3 and 4: The mean head RMSE for the period displayed in Fig.4 is roughly 0.9 for the case "Localization dist 10 km", which is much higher than the mean head RMSE of roughly 0.16 for the same case shown in Fig. 3. Why do the results differ that much?

Response: The values in figure 3 are wrong, due to the way that the results were extracted from the result files.

Modifications: The values in figure 3 have been corrected, and now reflect those of figure 4.

- Comment: Figure 4: Please change the y-axis of the lower panel in the plot so that all data is displayed properly. Which cases are shown in the lower panel displaying discharge? Only "No localization"? Please clarify in the figure caption.

Response: We agree that the readability of the figure is poor.

Modifications: The figure has been completely redesigned.

- Comment: Figure 5: Which localization algorithm was used? Please specify this in the caption. Shouldn't the titles ("Hydraulic head observations"/"Discharge observations") be on the left hand side of the plots? Finally, I understand that the columns display different locations of the observations. However, please tell the reader explicitly that the point of the figure is to visualize/compare patterns of localizations weights for different positions of observations. I also think it would be helpful to see the results obtained by the "local analysis localization" method for one scenario (for example the 10 km case).

Response: The figure shows the localization weights as determined by the adaptive algorithm, for two groundwater observations and for two discharge observations, in

both the groundwater domain and the stream flow domain. As such, for both of the observation types, the algorithm will determine localization weights in processes other than the process observed (e.g. groundwater head observations are assigned specific weights when updating the stream flow).

Modifications: An example of the distance-based localization weights is added, and the figure is revised to better show the structure. Furthermore, the following is added to the figure caption: "The two left columns show localization weights for groundwater head observations, the third and fourth columns show localization weights for discharge observations. An example of the localization weights as obtained using the distance based localization method, with a localization distance of 10 km, is included for comparison on the right."

- Comment: Page 2285, Line 23: Which localization method was used? Please specify.

Modifications: The localization type (adaptive) has been specified in the text and in the figures.

- Comment: Page 2286, Line 6-9: Why do the results differ between the cases "0 obs" and "0 obs incl. Local" in Figure 6? If no observations are included in the assimilation, isn't it impossible to compute Eq. 2 to 6? The localization matrix Pobs shouldn't influence the results at all. Please explain the difference between the results of the two cases.

Response: "0 obs" referrers to the number of groundwater head observations, as stated in the method section (although not clearly enough). Discharge observations are still assimilated, and localizing these is paramount, and greatly improves the filter performance.

Modifications: The following is added to the method section 2.6 regarding data availability: "Note that the distribution names only describe the groundwater head observations and that stream discharge observations are always assimilated unless otherwise

is stated in the scenario name (see section 2.7)".

- Comment: Page 2286, Line 23: Change to Fig. 6.

Modifications: Changed.

- Comment: Figure 8: It is hard to see the difference between the lines for the "2 obs" and "2 obs incl Local" cases. Please improve the figure.

Modifications: The markers of both lines have been enlarged, to improve readability.

- Comment: Page 2287, Line 20: Change to Fig. 8.

Modifications: Changed.

- Comment: Page 2288, Line 3-7: Is the inference on these lines correct? The parameters estimated in the InclParNoQ experiment shows a better match with the "true" parameter values (magenta lines in Fig.7) than the InclParInclQ experiment. The later experiment shows a better match with the base model (black lines in Fig.7), but this model was not used to generate the observations if I understand correctly. Which is the target, the parameters of the base model or the "true" model?

Response: The true parameter value in figure 7 is indicated by the black horizontal line, while the base value is indicated by the magenta line.

Modifications: To avoid future misunderstandings, the figure caption regarding the horizontal lines has been changed to: "The horizontal lines show the true parameter value (black line) and the base parameter value (magenta line)".

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 2267, 2015.