Hydrology and
Earth System
Sciences
Discussions

Open Access

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.

Anonymous Referee #1

Received and published: 15 April 2015

General comments

In this study the Ensemble Transform Kalman Filter is used to update the states and estimate the parameters of a hydrological model. The specific aim is to evaluate the influence of localization method, number of observations and number of ensemble members on the simulation results. The scope of the study is well suited for the journal, and the manuscript is mostly well written. I have listed some issues below, mostly technical ones, which if correctly identified need to be corrected before eventual publication.

C1052

Specific comments

Abstract:

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.

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.

Page 2268, Line 26:

Should it say "on one part of the hydrological cycle"?

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".

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.

Page 2272, Line 21:

Change to "mica sand".

Page 2273, Line 6:

Change to "saturated hydraulic conductivity".

Page 2274, Eq. 2:

I think a transpose might be missing in the equation. According to Harlim and Hunt the equation is C = (Yˆb)ˆTRˆ(-1). Please check.

Page 2274, Eq. 3:

The matrix C has dimensions n x k and the matrix Yˆb has dimensions k x n. Their matrix product CYˆb 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.

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.

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.

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.

Page 2279, Line 1-5:

The sentences on these lines sound awkward. Please rephrase.

Page 2279, Lines 19-22:

Why were the parameters transformed?

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.

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.

Page 2283, Line 1 and 7:

Change to Fig. 3.

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?

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.

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).

Page 2285, Line 23:

Which localization method was used? Please specify.

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.

Page 2286, Line 23:

Change to Fig. 6.

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.

Page 2287, Line 20:

Change to Fig. 8.

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?

Conclusions:

The findings of this study are well depicted in the conclusions section.