EnKF paper, Anonymous Referee #2

Overall evaluation: Minor revision

Major comments: This paper presents an application of distributed modeling data assimilation (DA) using ensemble Kalman filter (EnKF), and investigates the effect of updating frequency as well as the density of streamflow gauge on the performance of the EnKF. The authors carried out synthetic as well as real-world experiments for the basin with multiple stream gauges available, and nicely summarized research findings and discussions. This paper discusses an important topic on developing automateddata assimilation procedures for distributed models and testing with real-world data in order to improve operational streamflow forecasts. The study presented in this paper should be interesting to a broader readership, including operational forecasters, academics, etc. The only major concern I have is the validation method that the authors used. The validation of the DA results has been mostly based on the outlet flow simulation results (Figs. 5, 6, 8, 9, 10) where flow observations are assimilated in four out of five assimilation cases, as presented in Fig. 4. The validation results should be more meaningful if the validation is done at the stream gauge location where the discharge data is not assimilated. The authors may consider presenting and discussing interior flow results generated from both base model simulation and the assimilation procedure. In addition, the reviewer recommends improving clarity by adding more discussions or explanations, or rephrasing, wherever necessary or indicated by the reviewer; please, see specific comments for this. Overall, the reviewer thinks this paper merits publication in this journal.

Answer:
We thank Anonymous Referee #2 very much for her/his review and comments. Below we provide answers to the points, which were raised. We agree that validation of the simulated discharges at interior locations would make the results more convincing and we will include it in the revised manuscript as follows.

In order to validate the results of the real world experiments, the Nash-Sutcliffe model efficiency (NS) (Nash and Sutcliffe, 1970) is calculated at the stream gauge locations, where the discharge data are not assimilated. The median NS of the 64 ensemble updated discharge simulations (Cases A and B) as well as the open loop simulations (without assimilation) are shown in Table I for the updating frequency of 24h. The validation results indicate that the assimilated simulations give performance of the same order of magnitude as the open loop simulations. Small improvements can be still observed when the assimilated observations are upstream of the validation stations (Case B). However, when the assimilated observation is downstream of the validation gauges (Case A), the interior simulations can slightly deteriorate, but this is observed only at one gauge (number 5) out of five gauges. Nevertheless, this seems to be rather acceptable when we consider the large distance between the assimilation gauge (number 1) and validation gauge (number 5) and the fact that the two most upstream gauges are located at two parallel river branches (see Fig. 1a). Overall we can conclude that the validation at the stream gauges without data assimilation makes the presented results more convincing.

Table I: Validation at the stream gauge locations without data assimilation. The median Nash-Sutcliffe model efficiency of the 64 ensemble members for the open loop simulations and the updated simulations being assimilated using Cases A and B.

Gauge number	Open loop	Case A	Case B
1	0.95	– (assimilated)	0.96
2	0.95	0.96	0.97
3	0.96	0.97	0.98
4	0.96	0.96	0.97
5	0.93	0.91	– (assimilated)
6	0.87	0.89	– (assimilated)

Specific comments:

1. l11, p3962: pdf's -> pdfs (?)
Answer: We will correct this in the revised manuscript.

2. l16, p3962: (DA), by →(DA) by
Answer: We will correct this in the revised manuscript.

3. l3, p3964: Lee et al. (2012) -> I do not think this paper deals with soil moisture assimilation; please check again.
Answer: The reviewer is right, we will correct this appropriately in the revised manuscript. Indeed, Lee et al. (2012) analysed the sensitivity of variational data assimilation methods for multiple spatio-temporal adjustment scales, namely assessing 1) different spatial distributions of model states and 2) temporal resolution of biases in precipitation and potential evaporation.

4. l16-17, p3964: Unfortunately -> this word doesn't seem proper in the science paper; please, consider using a different word.
Answer: We will use "However" instead of "Unfortunately".

5. l4-6, p3965: Hence it is . . . forecast. -> this sentence does not read smoothly; please, consider improving readability.
Answer: We will rephrase this sentence as follows: "Hence it is interesting to test the optimal updating frequency at which the hydrological data are assimilated into the forecasting model to obtain the most accurate forecast."

6. l7,p3965: DA framework -> it appears in various places in the text, the authors are using DA "framework", DA "scheme", DA "machinery" but they all denote the same thing, I think; the reviewer suggests using the same word, e.g., DA procedure, throughout the paper to be consistent or not to create any confusion.
Answer: As the reviewer suggests, we will use the consistent terminology "DA procedure" throughout the revised manuscript.

7. l6,p3969: NS's -> NSs
Answer: We will correct this in the revised manuscript.

8. l8,p3969: root mean square error (rmse) -> Root Mean Square Error (RMSE)
Answer: We will change this in the revised manuscript.

9. Eqs. (8) and (9): What is different between H in Eq. (8) and H in Eq. (9)? Please, describe it in the text.
Answer: We thank the reviewer for this remark. Actually, there is no difference between H in Eq. (8) and H in Eq. (9). Both represent the same, a function (operator), which translates the model states into the simulated discharges, in our case the kinematic wave model (definition in l.4, p3970). However, we agree that the notation should be unified in both equations and in the text, otherwise it could be confusing the reader. Therefore, we will unify the notation in the revised manuscript.

10. l16,p3972: synthetic observation Qobs,k -> How is this different from Qobs in Eq. (14)? Is Qobs in Eq. (14) synthetic observation or actual streamflow data? If they are different, then Qobs in Eq. (14) should be described separately below Eq. (14).
Answer: We thank the reviewer for this correct remark. Qobs at l16,p3972 and Qobs in Eq. (14) are different. We always assimilated perturbed observed discharges, but we calculated the RMSE from the unperturbed observations using measured discharges for the real world case and Qtrue for the synthetic case. We will distinguish this more clearly in the revised manuscript.

11. Qfor in Eq.(14) -> It is not clear if Qfor is an ensemble mean or ensemble member? So, it is also not clear if RMSE presented in Figs. 5, 8, and 9 are based on an ensemble mean or all ensemble members.
Answer: This is an important remark. Indeed, Qfor stands for each individual ensemble member. Therefore, we will add one more summation notation over the ensemble size in Eq.(14). Then it will be more clear and mathematically correct.

12. l20, p3973: a variance of (0.1Qobs,k)2 -> Where does 0.1 come from? Is it based on a sensitivity run, or based on a data analysis, or based on the literature? I think Clark et al. (2008), which is cited in this paper, used 0.1 as well. The authors may describe a little bit detail on this or may simply add a reference.
Answer: The choice of the variance (0.1Qobs,k)2 was based on the literature, namely Weerts and El Serafy (2006) and Clark et al. (2008). We will add these references to the revised manuscript.

13. Subheading of the subsection 3.1.1 -> the current subheading doesn't seem good because of RMSE used in the title. Please consider renaming it, e.g., discharge forecast, as similar to the title of the subsection 3.1.2, or model performance on discharge forecast.
Answer: We will change the name of subsection 3.1.1 to "Model performance regarding discharge forecast".

14. l11-13, p3974: "the benchmark case A . . . to the catchment outlet" seems inconsistent with what is written in lines 17 to 20 in the same page "Additionally, . . . to the outlet."
Answer: We agree that these two sentences are inconsistent. Therefore, we will reformulate l11-13, p3974 and also implement comment # 23 as follows:
"...the benchmark case A performs worst of all 5 cases, although the differences between case A and B are marginal, given the small difference in their RMSE values as well as the high NS value in the case of base model simulation. Moreover, ..."

15. l23-25,p3974: "Slightly . . . frequency." This finding seems counter-intuitive. Please, consider adding a description in the text on why this is happened.
Answer: We agree that this part on l20-24,p3974 needs some reformulation. We propose the following wording: "For the updating frequency of 6 h, there is no pronounced improvement in RMSE, some of the forecasts even deteriorate in comparison to the 12-h updating frequency. This could be expected, because within the 6 h between updating moments, hardly any rainfall is transformed into discharge, even at the most upstream gauges, as is shown in Table 1."

16. l3, p3975: "In other words to check" Please, rewrite these words.
Answer: We will rephrase it as follows: "In other words, we wanted to check".

17. l22, p3977: "which is contradicting" Please, add an explanation in the text on what caused the contradictory result.
Answer: We will reformulate this part and add an explanation as follows: "It is worth mentioning that case B outperforms case A for lead times from 5 h to 20 h, which is not observed using the long-term statistics (Fig. 8). This is caused by the spatial properties of this major flood peak, during which the importance of the upstream gauges is clearly shown, however, completely averaged out in the long term statistics."

18. l7,p3978: "explicit routing" I do not know what the authors mean "explicit" here.
Answer: We will rephrase the whole sentence as follows: "The advantage of a grid-based hydrological model with grid-based routing over a lumped model with a unit hydrograph-type of channel delay is that"

19. l5-6,p3979: "mainly. . .EnKF scheme." In the reviewer's opinion, one of main reasons for the routing states to be more sensitive than the rest model states is that the EnKF, as formulated in this study, does not explicitly consider the high correlation between soil moisture states in the immediate past and streamflow at the time of forecast. As in Eq. (4), the model state vector is composed of water balance and routing states at the concurrent time step; in this case, it may be

difficult to build a covariance matrix among water balance model states (i.e., SM, UZ, LZ) via assimilating discharge observations. This seems briefly mentioned in lines 24 to 29 at page 3980, but not exactly discussing the issue described above. Please, consider discussing this in the text.
Answer: We thank reviewer for this comment. We will discuss this issue in the revised manuscript and extend l4-8,p3979 as follows (same as our answer to Anonymus Referee #1, comment #1):

"In this study, mainly the pdf's of the two routing model storages were changed, while the other model states (SM, UZ, LZ) were found to be less sensitive to the EnKF scheme. This is because the current formulation of the EnKF (see Eq. 4) does not explicitly consider the strong correlation between soil moisture states in the immediate past and streamflow at the time of forecast. Therefore, it may be difficult to build a covariance matrix among the water balance model states (i.e., SM, UZ, LZ) via assimilating discharge observations. Based on our results we can state that, given a measured discharge downstream, it is difficult to adjust (and justify) the soil moisture upstream (in a spatially distributed coherent manner) using an EnKF. Other filters like the Ensemble Kalman Smoother (EnKS), which calculate the analysis from several previous time steps (Evensen and Leeuwen, 2000), may result in better adjustment of the spatially distributed soil moisture states, which may improve forecasts for even longer lead times.
In this study, however, with a larger number of assimilated discharge gauges, both the forecasted and updated pdf's of SM, UZ and LZ had more narrow peaks around their actual true values."

20. l2-3, p3981: A reference may be necessary to support this sentence.
Answer: This sentence is a general hydrological assumption that soil moisture has higher spatial variability than deeper groundwater.

21. Fig. 3: consider presenting time series of interior flows.
Answer: It would certainly be possible to present time series of interior flows, but because of the already large number of figures in the manuscript, we would prefer including only the Nash-Sutcliffe model efficiency coefficients between the open loop simulations and observation (at the interior locations). See our answer to general comments and Table I (Open loop simulations).

22. Caption of Fig. 5: "EnKF assimilation . . . (right.)" Make this a complete sentence.
Answer: We will rephrase this caption as follows: "EnKF assimilation was carried out every 24h (left), 12h (center), 6h (right)".

23. Fig. 5: As to the RMSE results for cases A and B in the left plot, I am not sure if A and B results are considered significantly or noticeably different, given the small difference in their RMSE values as well as high NS value in the case of base model simulation. If their difference is marginal, please make it clear at the text.
Answer: We agree that the differences between cases A and B in the left plot of Fig. 5 are marginal and will make it clear in the text (see answer to comment # 14).