Reply letter to Anonymous Referee #3

1)Many details are missing and therefore the numerical experiments are not reproducible. In particular, there is no mention whatsoever of the model parameters (e.g., soil properties) and how they were perturbed to generate the initial ensemble of realizations. Precipitation and potential evapotranspiration were also perturbed but no description on how this was done is currently available in the paper. This is especially relevant for the Ahlergaarde catchment, while perhaps some information on the Karup catchment might be available in previous publications by the same research group.

Details on model parameters as well as on perturbation of parameters and forcings for the Karup model can be found in Zhang et al. (2015) and therefore we found it unnecessary to repeat here. The alike information for the Ahlergaarde model has now been added to the manuscript and a new table (Table 2 in the revised manuscript) is inserted listing the model parameters and associated uncertainty.

2)The "deterministic" model has been calibrated against observation data for both catchments, yet for the Karup catchment this is not showed neither in figures nor tables, while for the Ahlergaarde catchment only a comparison between observed and simulated discharge at the outlet is included, which is a bit limited, compared to the capabilities of the model. I think it is important to show and briefly discuss the model calibration performance in both cases, maybe also in terms of water table, to build some confidence that the subsequent analyses are realistic.

In the revised manuscript, we now briefly discuss the model calibration in both cases. For the Karup model a new figure (Fig.3) is included showing observed and simulated water table and discharge for two selected stations. The performance of the calibrated Ahlergaarde model can be seen in Table 4,5 and Fig.10, where ('NoDA') represents the results of the calibrated model.

3)One of the recurrent results of this study is that assimilation of groundwater head does not improve (or even worsen) soil moisture and vice versa, whereas Camporese et al. (Vadose Zone Journal, 2009), in a similar study, showed that EnKF-assimilation of surface soil moisture can improve the saturated zone and assimilation of groundwater head can improve surface soil moisture. This is probably due to the fact that the model used by Camporese et al. (VZJ, 2009) seamlessly solves the entire subsurface domain with the full 3D Richards equation, while here there might be some "disconnection" between the unsaturated domain and the saturated one. This may lead to the risk of physically inconsistent updates and hence the need for variable localization. I believe that a more insightful discussion is required about the relationship between the DA results and how the saturated zone and the unsaturated one are coupled in MIKE SHE, especially for the readers that, like me, are not entirely familiar with how process coupling is done in such model.

In MIKE SHE the saturated and unsaturated zones are explicitly coupled. This is done to optimize modelling time steps used in the unsaturated zone (minutes to hours) and saturated zone (hours to days), respectively. The flux between the unsaturated and saturated zones is calculated by an iterative procedure that conserves mass for the entire column. This means that assimilation of SM may have an effect on GW and vise versa through this explicit coupling. Although solving the 3D Richards' equation for the entire subsurface represents a seamless solution to flow in the unsaturated and saturated zones, MIKE SHE nevertheless considers interaction between the two zones.

Specific comments

Page 1, lines 18-19: improvements obtained in discharge and ET with data assimilation were only marginal, as acknowledged by the Authors later on in the manuscript. I suggest this statement be relaxed.

We changed the sentence 'a much improved performance' to 'an overall improved performance'.

Page 2, lines 4-6: this is not a general definition of hydrological data assimilation. Please rephrase.

We changed the sentence from 'Hydrological data assimilation aims to efficiently combine both the knowledge represented by integrated hydrological modelling and the information gained from observations.' to 'Hydrological data assimilation aims to utilize the information embedded in available hydrological observations for improving the performance of hydrological models'.

Page 4, lines 22-23: irrigation is not included, yet it is said that the catchment is located in one of the most

irrigated areas of Denmark. I think a much stronger justification is warranted than simply “for computational efficiency”.

We changed the sentence from 'For computational efficiency' to 'For computational efficiency, and due to the fact that the exact irrigation information in terms of both location and amount is not known,…'.

Page 7, line 8: the true state in reality is never known, therefore it is not sure it can be always represented. I suggest this sentence be reformulated.
We changed the sentence from 'the true state and the uncertainty are represented by a limited ensemble of realizations.' to 'the forecast state and its associated uncertainty are represented by a limited ensemble of realizations'.

Page 9, line 15: "appropriate model error" is not sufficient, please give more details.
More details are given in the revised manuscript. The new table 2 is included and more detailed information is given on forcing and parameter perturbation (P14 L6-15).

Page 10, line 23 and Figure 3: I assume RMSE is averaged not only in space but also in time. Please clarify.
Yes the RMSE is averaged not only in space but also in time. Clarified in the figure text in the revised manuscript.

Page 12, line 18: does it mean that previous experiments were run with an ensemble size of 60? If so, please clarify. Also, please provide justification for using 60 (e.g., previous sensitivity analysis?).
Yes, previous experiments were for an ensemble size of 60, which is now emphasized in the revised manuscript. Justification for using 60 is also provided, which is by considering both assimilation performance and simulation time (P12,L33 – P13,L1).

Page 13, lines 22-26: more details are needed regarding the perturbation of precipitation, evapotranspiration, and parameter values. For instance, how were the variables perturbed? With additional Gaussian noise? And using what variance? Also the model parameters (e.g., soil properties) must be specified.
More details are given in the revised manuscript. In the first paragraph of section 5 (P14,L4-15), detailed information on forcing/parameter perturbation is given together with new Table 2.

Page 13, lines 29-30: from the figure it is clear that the model consistently underestimates low flows and overestimates peak flows. Therefore, I do not think that "the model is good at predicting low flow" and suggest this sentence be reformulated. What about performance of the model against other data (e.g., water table, soil moisture)?
The sentence is reformulated. The model performance against other data can be seen in Table 4 and 5 as well as in Fig.10

Page 14, lines 7-10: this procedure for "adjusting" the observations and make them closer to the model predictions is a strong assumption and represents a significant limitation of this study. By doing this, what would be a real test case becomes basically another synthetic experiment. If the Authors really think the observations are biased (and it seems they justify this hypothesis based only on the fact that the calibrated model is not able to match the data, which is a weak explanation, in my opinion), why don't they use the framework proposed by themselves in Rasmussen et al., "Data assimilation in integrated hydrological modelling in the presence of observation bias", HESS, 2016? Such an analysis would make the paper more robust and interesting.

This study is based on the overall assumption that neither the assimilated observations nor model are significantly biased. Based on this strong assumption, we assumed the bias comes from observation such that the observation bias can be removed prior to assimilation. We also think using the bias-aware filter is a more natural approach to deal with the bias. The scope of current study mainly concerns the results with respect to assimilating different variables. We believe including bias-aware filter is very interesting to be explored in future studies.

Page 14, lines 16-19: this explanation is needed earlier in the manuscript. See previous point concerning Page 4, lines 22-23.
We moved this explanation to an earlier section in the revised manuscript.

Page 14, lines 19-20: more details are needed here. Which criteria were used to remove unreliable observations from the dataset?
The unrealistic observations here refer to the negative values or the sudden spike data, most probably due to technical errors. We simply remove those data empirically.

Page 16, line15: data used in this study are not exactly "real" (see previous comment). Please relax this statement.
We relax this in the revised manuscript. Instead of saying 'real data', we changed to 'data from real observations'.

Page 17, line 4: improvements in discharge and ET are very small, I would say "marginal", instead of "relatively small".
We changed to 'very small' in the revised manuscript.

Page 20, Table 2: please remove the column with cumulative ET and replace it with measures of performance of the various DA scenarios, as done with discharge at the outlet.
Changed as suggested.

Page 22, Table 4: first column (cumulative ET) can be removed.
Changed as suggested.

Page 31, Figure 9: not clear what this figure should represent. Here RMSE is computed with respect to the "deterministic" model, which is clearly not the truth, is surely affected by strong uncertainty, and may well be erroneous (as shown by model results in Figure 6). If the Authors want to quantify the spatial distribution of the system corrections made by DA, I suggest replacing RMSE with, e.g., AAD (average absolute deviation) between the DA runs and the deterministic model. Finally, please add labels to x- and y- coordinate axes.
The RMSE between the assimilated and deterministic model show (not exactly quantify) the corrections made by DA spatially. RMSE is used rather than other measures to keep the consistency as RMSE is used throughout the manuscript to compare between assimilated model results and truth/observations.

Technical corrections

All over the manuscript: correct all citations where the Authors are erroneously in parentheses, e.g., (De Lannoy et al., 2007) applied EnKF . . .
Page 4, line 4: replace "relative" with "relatively".
Page 4, line 27: correct "Metrological". In general, small edits of English are required throughout.
Figure 5: please correct labels in the top three panels.

We fixed all above technical errors.

Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C., and Jensen, K. H.: Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model, Adv. Water Resour., 86, Part B, 400-413, http://dx.doi.org/10.1016/j.advwatres.2015.07.018, 2015.