Response letter of hess-2019-324-RC1

Please find the responses to the comments.

Comments made by the reviewer were highly insightful. They allowed me to greatly improve the quality of the manuscript. I described the response to the comments.

Each comment made by the reviewers is written in *italic* font. I numbered each comment as (n.m) in which n is the reviewer number and m is the comment number. In the revised manuscript, changes are highlighted in yellow.

I trust that the revisions and responses are sufficient for my manuscript to be published in *Hydrology and Earth System Sciences*

Responses to the comments of Reviewer #1

The author of this paper used a synthetic case and indicated that topography-driven lateral surface flows induced by heavy rainfalls do matter for data assimilation of hydrological observations into hyper resolution land models. Although this paper reads well and the author provided a long discussion on results, these results are only based on a few deterministic measures, the author needs to clarify more detail and use additional matrices to evaluate his results. All the figures and tables are appropriate.

This manuscript can be considered for publication after carefully addressing all of my concerns.

(1.1) Minor Lines 62-63: "...by the data assimilation of microwave brightness tempera-ture observations..." should be "...by assimilating microwave brightness temperature observations..."

→ I have modified it following the reviewer's instructions.

"Sawada et al. (2015) successfully improved the simulation of root-zone soil moisture by assimilating microwave brightness temperature observations which include the information of vegetation water content."

(1.2) Major: Line 203-206: please use some mathematical relationship to elaborate more what is van Genuchten relationship and how it has been used as H operator to convert pressure head to soil moisture. Parflow does not estimate the soil moisture directly?

→ I have clarified the van Genuchten relationship in the revised version of the paper.

$$S_W(h) = \frac{S_{sat} - S_{res}}{(1+(\alpha h)^n)^{(1-\frac{1}{n})}} + S_{res} \quad (4)$$

$$k_r(h) = \frac{\left(1 - \frac{(\alpha h)^{n-1}}{(1+(\alpha h)^n)^{\left(1-\frac{1}{n}\right)}}\right)^2}{(1+(\alpha h)^n)^{\frac{\left(1-\frac{1}{n}\right)}{2}}} \quad (5)$$

where α [L-1] and n [-] are soil parameters, S_{sat} is the relative saturated water content and S_{res} is the relative residual saturation.

Yes, Parflow does estimate the soil moisture directly so that I did not need to formulate the complicated H for data assimilation. What I wanted to say here is that I directly adjusted pressure head by assimilating (synthetic) volumetric soil moisture observation. The assimilated observation variables are not consistent to the adjusted state variables. Therefore, in the calculation of background covariance (equations (10) and (11)), the van Genuchten relationship can be recognized as H although

I did not need the van Genuchten relationship in data assimilation since volumetric soil moisture has already been calculated by Parflow. This point was indeed unclear in the original version of the paper and I have clarified this issue in the revised version of the paper.

> "I assimilated volumetric soil moisture observations so that $\boldsymbol{y}^f$ and $\boldsymbol{y}^o$ are simulated and observed volumetric soil moisture, respectively. The van Genuchten relationship converts the adjusted state variables $\boldsymbol{x}^f$ to the observable variables $\boldsymbol{y}^f$ and can be recognized as an observation operator $\mathcal{H}$. However, since volumetric soil moisture $\boldsymbol{y}^f$ has already been calculated by Parflow, I did not need the van Genuchten relationship in data assimilation."

(1.3) Lines 210-217: why did you use this approach to identify the closeness of the two PDFs, this seems a very old technique. It would have been much better before using each method you had explained the reason and necessity of using that approach. As this is the synthetic case and you are generating the pressure head and soil moisture observation accordingly, I am not sure how this study can be done on a real-case problem, which is very important, as its result would be more convincing. The author used only a few deterministic measures (e.g., RMSE) to assess the performance of the DA for all the assimilation scenarios in this study. Speaking of uncertainty quantification, both probabilistic and deterministic measures should be used to evaluate the effectiveness and usefulness of the EnKF model. These metrics although show how the simulated quantities could accurately match the observations, it does not provide any insight on the reliability of the predicted values. Therefore, I recommend using the following paper, in which the authors provided a comprehensive description of different probabilistic performance measures, such as Reliability and 95% exceedance ratio (ER95). These measures have been extensively used in many studies to evaluate the quality of the posterior distribution. Abbaszadeh, P., Moradkhani, H., & Daescu, D. N. (2019). The Quest for Model Uncertainty Quantification: A Hybrid Ensemble and Variational Data Assimilation Framework. Water Resources Research,55, 2407–2431. https://doi.org/10.1029/2018WR023629.

→ Please note that I did not use the KLD as an evaluation metrics although the reviewer provided this comment in the context of evaluation metrics. Although the KLD is old, it is widely used in the context of machine learning. The KLD has also been used to measure the Gaussianity in the context of data assimilation so that I used it to evaluate how the ensemble simulation follows the Gaussian distribution. This point was indeed unclear in the original version of the paper. I have clarified this point in the revised version of the paper. This modification includes the response to the other reviewer's comment.

> "To evaluate the non-Gaussianity of the background error sampled by an ensemble, I used the Kullback-Leibler divergence (KLD) (Kullback and Leibler 1951):
>
> $D_{KL}(p,q) = \sum_i p(i) log \frac{p(i)}{q(i)}$ (13)

where $D_{KL}(p, q)$ is the KLD between two probabilistic distribution functions (PDFs), p and q. If two PDFs are equal for all i, $D_{KL}(p, q) = 0$. A large value for $D_{KL}(p, q)$ indicates that the two PDFs, p and q, substantially differ from each other. Therefore, the KLD can be used as an index to evaluate the closeness of two PDFs. In this study, I compared the PDF of the ensemble simulation (p in equation (13)) with the Gaussian PDF which has the mean and variance of the ensembles (q in equation (13)). A large value for $D_{KL}(p, q)$ indicates the state variables simulated by ensembles do not follow the Gaussian PDF. It should be noted that the KLD is not symmetric ($D_{KL}(p, q) \neq D_{KL}(q, p)$). The KLD has been used to quantitatively evaluate the Gaussianity of the sampled background error in the studies on data assimilation (e.g., Kondo and Miyoshi 2019; Duc and Saito 2018)."

The reviewer suggested clarifying the reason of the choice of methodology. Generally, I chose ParFlow and EnKF because they are widely accepted in the community. I would like to clarify how surface lateral flows matter in the widely accepted methodology. For Parflow, this point has already been clarified in the original version of the paper. I additionally emphasized this point in the revised version of the paper (the response to the other reviewer's comment is also included below):

"ParFlow is an open source platform which realizes fully integrated surface-groundwater flow modeling (Kollet and Maxwell 2006; Maxwell et al. 2015). This model can be efficiently parallelized in high performance computers and has been widely used as a core hydrological module in hyperresolution land models (e.g., Maxwell and Kollet 2008; Maxwell and Condon 2016; Fang et al. 2017; Kurtz et al. 2016; Maxwell et al. 2011; Williams and Maxwell 2011; Shrestha et al. 2014). Since I used this widely adopted solver as is and added nothing new to the model physics, I described the method of ParFlow to simulate integrated surface-subsurface water flows briefly and omitted the details of numerical methods. The complete description of ParFlow can be found in Kollet and Maxwell (2006), Maxwell et al. (2015) and references therein."

The EnKF is also widely accepted as the data assimilation algorithm for hyper-resolution land models. This point was unclear in the original version of the paper. I have clarified this point in the revised version of the paper.

"In this paper, the ensemble Kalman filter (EnKF) was applied to assimilate soil moisture observations into ParFlow. The EnKF has widely been applied to hyper-resolution land models (e.g., Camporese et al. (2009); Camporese et al. (2010); Ridler et al. (2014); Zhang et al. (2015); Kurtz et al. (2016); Zhang et al. (2018)). I examine if surface lateral flows matter for data assimilation of soil moisture observations into hyperresolution land models using this widely adopted data assimilation method. "

The reviewer suggested clarifying how to convince what I found here by real world applications. I believe that I could perform the similar experiment using in-situ soil moisture observations in the intensively observed river basins or using high resolution satellite observation. I have clarified this point at the end of the revised paper:

> "Future work will focus on the real-world applications using intense in-situ soil moisture observation networks and/or high-resolution satellite soil moisture observations."

The reviewer also suggested using the probabilistic measures to evaluate the performance. Figure R1 shows the spatial distributions of 95% exceedance ratio (ER95) in the HIGH_K-DOWN_O experiment. In the NoDA experiment, ER95 is 0% everywhere. Since I assumed the large uncertainty in rainfall and saturated hydraulic conductivity and it is not mitigated in the NoDA experiment, the forecast is completely underconfident. Data assimilation made this too large ensemble spread smaller. However, the spatial averaged ER95 is 31% and much larger than 5% so that the ensemble forecast in the DA experiment is overconfident. This is probably because the number of rainfall events and/or the frequency of rainfall events is small (see Figure S1 and Figures of Abbaszadeh et al. (2019)). In hydrological models, rainfall events are the primary factor to increase the ensemble spread so that it is difficult to maintain the appropriate ensemble spread with the small number of rainfall events. Interestingly, the regions of good ER95 corresponds to the regions where RMSE is greatly reduced (please compare Figure R1b and Figure 2d) so that RMSE can be used as a good proxy of the probabilistic measure. Same conclusions can be obtained in the other synthetic experiments.

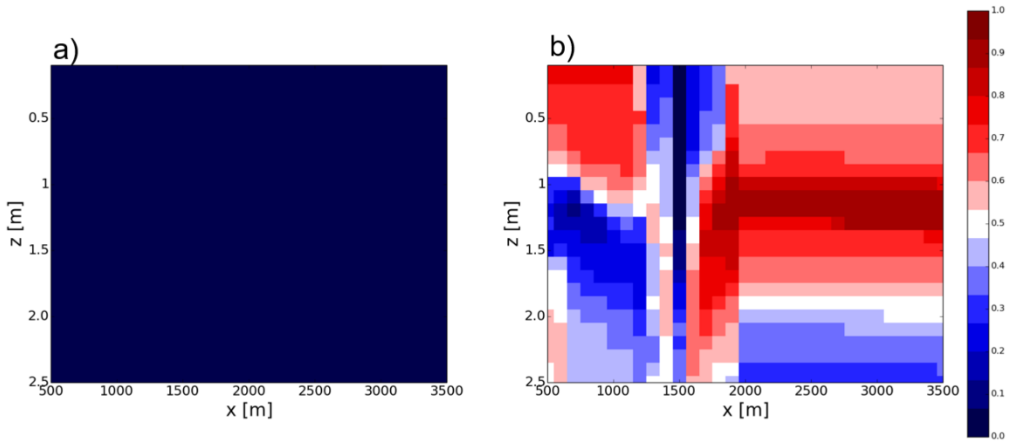

Figure R1. ER95 for (a) the NoDA experiment and (b) the DA experiment in the HIGH_K-DOWN_O setting.

I would like to propose not to include Figure R1 and the detailed discussion of the evaluation by the probabilistic measure although I briefly mentioned the importance of the probabilistic measure in the revised version of the paper. First, as the other reviewer revealed, this theoretical paper has already been very complicated for readers outside the community of theoretical and hydrologic data assimilation. To get many potential readers, I believe that I should not further add the results and figures if it is not absolutely necessary. Please note that the uncertainty quantification is the quite advanced topic. To my best knowledge, the probabilistic measures that the reviewer raised have been used mainly in the data assimilation of lumped and conceptual hydrologic models (e.g., Abbaszadeh et al. (2019)), which is the most matured research field in the hydrologic data assimilation. Currently, the studies on the data assimilation of hyper-resolution land models have not used these evaluation metrices. I believe that the take-home-message of this study can be described and validated without this probabilistic measure.

Second, the current experiment design was not appropriate to deeply discuss the evaluation by the probabilistic measures. As I discussed in Figure R1, in the synthetic experiment, the number of rainfall events is small, and the timing and magnitude of rainfall were not diversified. Therefore, I could not expect the enough amount of data to evaluate the long-term statistical property of the ensemble simulation as Abbaszadeh et al. (2019) did. This point is important to move on to the more realistic experiment implemented in the future. In the revised version of the paper, I have included this limitation in the discussion section citing Abbaszadeh et al. (2019).

> "The other limitation of this study is that I could not thoroughly evaluate the skill of the ensemble data assimilation to quantify the uncertainty of its prediction. Following Abbazadeh et al. (2019), I calculated the 95% exceedance ratio and found that the ensemble forecast was systematically overconfident (not shown). In the synthetic experiments of this study, the number of rainfall events was small, and the timing and magnitude of rainfall were not diversified. Due to this limited amount of data, it is difficult to deeply discuss the accuracy of the quantified uncertainty by data assimilation. While the skill of lumped hydrological models was often evaluated by the probabilistic performance measures such as the 95% exceedance ratio (e.g., Abbazadeh et al. (2019)), the uncertainty quantification of the simulation of hyper-resolution land models is in its infancy. How surface lateral flows affect the accuracy of the uncertainty quantification by data assimilation should be investigated using more realistic data."

(1.4) Lines 622 and 623: "Although particle filtering in a high dimensional system suffers from the "curse of dimensionality", please highlight that this can be resolved through improvements of

importance sampling in PFs, and therefore it provides the potential for data assimilation application in large-scale systems" for more discussion the readers can be referred to the following papers: P. Van Leeuwen. (2009). Particle Filtering in Geophysical Systems. Mon. Weather Rev., 137 (12), pp. 4089-4114. https://doi.org/10.1175/2009MWR2835.1

→ I fully agree with this reviewer's comment. Recently, some studies provided the methodological advances of PF although their applicability to hydrological models has not been discussed. Note that many of works cited by van Leeuwen et al. used conceptual models such as Lorenz96 and the applicability of these methodological advances to the real-world problems is still debated. I have included this issue in the revised version of the paper.

> "Although particle filtering in a high dimensional system suffers from the "curse of dimensionality" (e.g., Snyder et al. 2008), some studies developed the methodology to improve the efficiency of particle filtering (e.g., van Leeuwen 2009; Poterjoy et al. 2019)."

(1.5) Lines 633-634- How do you convince that this "In addition, in the virtual experiment of this paper, I neglected some of the important land processes such as transpiration, canopy interception, snow, and frozen soil." is a correct pre-assumptions.

→ First, these neglected processes can be modelled as a source term of ParFlow in many hyperresolution land models. Therefore, these processes do not modify the fundamental physical process simulated by ParFlow so that what I found in this study can be robust to the models which include these processes. This point was indeed unclear in the original version of the paper and I have clarified it in the revised version of the paper.

> "These processes affect the source term of equation (1) in hyper-resolution land models (e.g., Shrestha et al. 2014). Since the inclusion of the neglected processes do not change the structure of the original ParFlow, the findings of this study can be robust to the models which include these processes."

In addition, here I focused on the propagation of overland flows, whose timescale is relatively short compared with the neglected processes. Therefore, neglecting these processes may not have a large impact on the conclusion of this paper quantitatively. This point has already been discussed in the original version of the paper.

> "Although they are generally not primary factors in the propagation of overland flows generated by extreme rainfall, which has a shorter timescale than the neglected processes, those processes should be considered in the future."