Hess-2019-242-RC1
Reply to Hessel Winsemius' comments

Emery et al. propose a new data assimilation scheme (AEnKF), to be used for assimilation of wide-swath altimeter information from the upcoming Surface Water and Ocean Topography (SWOT) mission. Given the large-scale nature and long-time scale revisit times, using an asynchronous scheme, that to the best degree possible utilizes time and space varying availability of information seems like a logical and useful choice, compared to more synchronous approaches such as classical EnKF. I consider this to be a useful contribution to HESS and very much in scope, and useful as a preparation for using SWOT in large-scale hydraulic simulations and forecasts.

The authors would like to thank referee Hessel Winsemius for his detailed and very helpful comments, which helped us to better articulate our paper. Please find below our replies and associated modifications to the manuscript in order to address these concerns. To easily address all comments, we numbered each comment: they are referenced by a "M" for the major comments or by a "m" for the minor comments.

We kept the reviewer's comments and question in bold while our replies are in italic. When the associated modifications in the manuscript remain small, we inserted the modified paragraph in our reply in plain text: the black text corresponds to the original unmodified text, the crossed text corresponds to deleted text and the blue text corresponds to new text.

*The abstract, the introduction (the latest paragraphs) and the discussions sections have been heavily modified to take into account several remarks from the three reviewers. Therefore, we attached to our reply separate files with the new version of these three sections. In these rewritten sections, we used a color code to differentiate which reviewer made the comment and suggested a modification: comments from reviewer 1 (Hessel Winsemius), 2 (Claire Michailovsky) and 3 (Paul Bates) are in purple, orange and green respectively. Each modification is also referenced by a code in bracket indicating the reviewer (R#***) and the type and index of the comment (M/m#***) such as: "[R#3-M#1]".*

I do have a number of comments that lead to my verdict that this paper requires major revisions.

<u>MAJOR COMMENTS</u>

M#1. My largest comments are a) the choice to update Manning's n (rather than a state); and related to this, b) the choice to only evaluate the assimilation performance on the basis of water levels (or depth). In most applications, the user will require a good estimate of the river flow (besides water levels), because river flow (a volume in time) controls availability of water for some process that is to be predicted by the model used, not just the water level. A hydrology-hydraulic model-cascade could for instance be used to provide inputs to water allocation predictions for the forthcoming weeks/months (requiring an amount of flow over a given time span), or an upstream boundary condition for a flood simulation of a downstream river stretch. For all such applications, an accurate volume per unit of time is required, not just a stage (except for a flood simulation in a steep area, where floodplain storage is negligible to event accumulated flow, but these are

generally small streams, definitely not comparable with the size of the Amazon and its tributaries). I consider this a large (and unnecessary) weakness in the approach, with the additional risk that the Manning roughness will change to physically highly illogical values (which it in fact does in this study!), because e.g. either the water balance of the underlying hydrological simulations of ISBA are biased, or the channel dimensions are poorly defined. To become useful for typical applications, data assimilation should be as much as possible aimed at correcting the amount of water in channel sections, so that predictions after state updating can be made useful and reliable. This is now not proven. I find it a pity that the authors decided to apply this AEnKF on parameters with (As far as I can find in the text) the sole reason being that other authors already used it for state estimation experiments. This makes the study purely theoretical, as I don't really see how the experiments would ever be applied in a real-world case. The authors should at least show river flow as an additional benchmark variable and show how the 3 experiments affect the accuracy of river flow and discuss this result. My logical feeling is that discharge will be quite heavily impacted especially in experiment 3 where a bias is introduced.

We agree with the reviewer on the fact that river flow/discharge is one of the main variable of interest for river modeling applications and, following that, state estimation are well adapted for hydrology data assimilation applications. However, we would like to emphasize that we already dedicated another study focused on the correction of the state and the discharge in the context of remotely-sensed hydrology products, see Emery et al. (2018). Moreover, as SWOT is a scientific mission with a nominal lifespan of 3 years, an important application of SWOT data would be to calibrate hydrological models in order to get better simulations over past and future periods. Therefore, the present parameter estimation study works as a complementary study focusing on correcting crucial parameters for hydrological applications that are still not well-known. To set this framework clearer, we added those remarks at the end of the introduction. Still, it is possible to apply the current framework (e.g. twin experiments with SWOT-like observations) to correct either water depths or river discharges (state estimation). We added this point to the perspectives and future works in the manuscript's conclusions.

The choice of the Manning coefficient as control variable is directly linked to the results of a sensitivity analysis (SA) of the ISBA-CTRIP river outputs to its routing parameters (Emery et al 2016). This SA showed that, among all CTRIP parameters and their tested ranges, simulated water depths are essentially sensitive to the Manning coefficients. Then, as SWOT water elevations product is closest to ISBA-CTRIP's water depths, it was chosen to build a framework based on the assimilation of SWOT-like water depths to correct the model most sensitive input parameters.

Besides, the reviewer also made very relevant points regarding the consequence of forcing/LSM bias on the roughness coefficient value. Indeed, in a real-case framework, if these types of bias are not considered, the assimilation scheme will try to change Manning coefficient values to compensate water elevations variations due to these biases. We realized that this point was not discussed in the initial manuscript, so we dedicated a new paragraph in the Discussions section where we acknowledge this limitation of our current approach and suggest solutions to handle it in future developments.

Moreover, we would like to clarify that it is not the direct value of the Manning coefficient that is corrected but a multiplying factor applied to the Manning distribution (see our reply to your M#2). Therefore, the high values for the control variable displayed in Figures 7 and 9 correspond to these multiplying factors which remain within a range of 0.5 and 1.5, while the corresponding Manning coefficients are within physical value between 0.02 and 0.07.

Finally, as suggested by the reviewer, the model and data assimilation performances on water depth and discharge were easily calculated. Given the idealized framework of the OSSE, the statistics were exceptionally good for both water depth and discharge and did not bring essential new information to the manuscript. These tables were then added to the manuscript in a dedicated appendix.

- *Modified introduction >> see attached file: Hess-2019-242-corrected-introduction.pdf*
- *Modified manuscript in the study's perspectives (p.19, l.15-22):*

These experiments offer several perspectives. They mainly consist in going towards more realistic data assimilation experiments that take into account more sources of uncertainties between the model and the observations, ~~(~~such as correlated observation errors or uncertainties in the forcing and LSM surface and sub-surface runoff~~)~~. To test the performances' limitations regarding the DEM/bathymetry bias issue, one can use simulated water surface elevations referenced to a geoid instead of water depths from the model or even assimilate water depths from another model where the bathymetry is different. As most applications generally require a good estimate of the river flow and river water volume, another lead of investigation could maintain the SWOT-based OSSE framework but to correct the simulated water storage and/or discharge, either as a single state estimation framework or as a dual state-parameter estimation framework (similarly to dual discharge-bathymetry inference methods developed by Oubanas et al., 2018 and Brisset et al. 2018 for some hydraulic models). Moreover, along with observations of water surface elevations, SWOT will also provide two-dimensional maps of river widths and surface slopes. One can also study the possibility of assimilating such product to ~~correct the corresponding parameters in ISBA-CTRIP such as the model river width or maybe~~ constrain other parameters such as the bankful depth that controls the model flooding scheme.

- *Modified discussions >> see attached file: Hess-2019-242-corrected-discussions.pdf*

M#2. Second point: I don't fully understand the zonal approach to updating Manning's n. To me it would make more sense to use an upstream-downstream relation in manning coefficients (e.g. update manning coefficient at location, as well as upstream and downstream) which could easily introduce a logical covariance between n values (rather than assuming everything to be independent). In fact, the full zonal approach with areas that may have very little relationship to each other suggests, that there is a 100% covariance across the zone. Why was this selected in this way (or would things become overcomplicated if done in a different way)? Consider discussing this in the last sections of the paper.

The zonal approach is applied to the control variables only. As previously mentioned in the reply to M#1, the control variable are not directly the Manning coefficients but rather a set of multiplying factors applied to the Manning coefficient distribution. It is those factors that are set constant over the 9 hydro-geomorphologic areas while the Manning coefficient distribution remains spatially-distributed at the grid-cell scale and built on an upstream-downstream relationship (specified in Eq. 1 of the manuscript). The interest of using such zonal approach to correct the Manning coefficient distribution was indeed to maintain this upstream-downstream relationship between the grid-cell (which was not the case when each Manning coefficient was individually updated in Pedinotti et al. 2014). We realized that this aspect might not be clear enough in the manuscript as reviewer #2 also *raised similar questions. Therefore, we modified the manuscript adequately in Section 3.2.2 to better present the definition of the control variable and recall it in Sections 4 and 6.*

- *Modified manuscript in section 3.2.2 (p.7, l.17-21)*

Following the conclusions from the ISBA-CTRIP sensitivity analysis to its routing parameters in Emery et al. (2016), we determined that assimilating water-depth-like observations would be efficient to correct the distribution of the river Manning coefficients. These coefficients are spatially-distributed at the grid-cell scale. However, from Pedinotti et al. (2014), equifinality issues were raised by correcting the distribution at this scale while also affecting its upstream-to-downstream spatial distribution. Thus, we chose to correct it by applying multiplying factors defined at a coarser scale, namely at the scale of the 9 hydro-geomorphological areas defined in Section 2.3 and illustrated in Figure 1b. Within a same area, the Manning coefficients of all grid-cells are identically updated by being multiplied by the same correcting factor. Thus, data assimilation will focus on directly adjusting these multiplying factors. Therefore, the control vector is composed of the nx multiplying factors $N_{mult;i}$, i = 1: : :n_x, applied to correct the distribution of the river Manning coefficient:

$$x_k = \left[N_{mult,1}, \dots, N_{mult,n_x}\right]^T, (8)$$

giving nx = 9.

- *Modified manuscript in section 4 (p.11, l.2-5)*

In the incoming experiments, the true control variables xt are:
…. (17)
and their ~~background~~a priori values at~~for~~ the first assimilation cycle xb are:
… (18)

- *Modified manuscript in section 6 (p.14, l.20-21)*

We present now the results from the data assimilation experiments presented in Table 2 and in Section 4.4.2. Recall that these experiments aim at correcting a set of 9 multiplying factors applied to the manning coefficient distribution and constant over 9 hydro-geomorphological zones ~~the spatially-varying Manning coefficient in the nine zones~~ covering the Amazon basin.

M#3. The English writing and sentence constructions are not everywhere up to standards. Please make sure the manuscript is reviewed by a (near-)native English person.

Reviewer #2 had a similar remark. Therefore, while preparing the replies to the reviewers, we submitted the manuscript to an independent English-speaking proofreader to improve the English.

MINOR COMMENTS

m#1. Introduction: there are many "however"s in the text. Some or many of these can be removed.

Thank you for noticing this. We read through the introduction and modified it to use different linking words.

m#2. P. 3, l. 32 "at a coarser scale", please just describe the scale.

Here, the "coarser resolution" relates to the zonal distribution, in opposition to the grid-cell finer resolution. As the zonal distribution is not introduced yet in the manuscript, it is true that it might be confusing. Therefore, we modified the sentence in the introduction to be more explicit.

- *Modified manuscript in Introduction (p.3, l.32-33)*

For the current study, it was decided to update the Manning coefficient distribution, not a the grid-cell resolution but at a coarser zonal resolution, by applying multiplying correcting factors constant over each zones, ~~at a coarser regional scale~~ identical to the one used in Emery et al. (2016).

m#3. p. 4. l. 13: "gravitational drainage", do you mean groundwater outflow?

Not exactly. When we use the term "gravitational drainage", we are describing the LSM where the water flows toward the deep soil and feeds CTRIP's groundwater reservoir (denoted G). The term "groundwater outflow" is used for CTRIP and represents the flow from the groundwater reservoir G into CTRIP's river/surface reservoir S. It is now specified in the manuscript.

- *Modified manuscript in section 2.1 (p.4, l.12-13)*

In particular, ISBA gives a diagnostic of the surface runoff (QISBA,sur) and the gravitational drainage (QISBA,sub, , i.e. water percolating to the deep layers of the soil) later used as forcing inputs for the RRM denoted CTRIP.

m#4. p. 4. l. 20. Replace "empties" by "spills"

Thank you this suggestion. The modification was made into the manuscript.

m#5. p. 4, l. 29, Replace "fixes" by "results in"

Thank you this suggestion. The modification was made into the manuscript.

m#6. p. 5 l. 2 (p.5): ". . .values between 0. 75 and 1 for smaller and mountainous tributaries. . ." I guess you mean 0.075 and 0.1 s m-1/3. The values you mention are ridiculously high!

Yes, you are absolutely right. This is a typing error, thank you for noticing it. The manuscript has been corrected.

m#7. p. 5. Eq. 1. Why is SOmax not simply 1 as it is only a way to scale values?

Actually, there is another typing error here. SO is the stream order taking values ranging from 1 at source cells (grid cells without any upstream grid cells, according to the river network) to a maximal SO associated to the outlet grid cell (depending on the depth of the river network). It is Nmin and Nmax that takes values of 0.04 and 0.06, following the configuration from Decharme et al (2012). These definitions were corrected in the manuscript.

- *Modified manuscript in section 2.2 (p.5, l.8-10)*

~~with~~ SO is~~being~~ the stream size relative measure at the current cell; SOmax ~~= 0:06~~ is the same measure at the river mouth (which value depends on the depth of the river network) and SOmin = ~~0:04~~1 the measure at source cells (namely cells without any upstream cells according to the river network) ~~(Decharme et al., 2012)~~. The Manning coefficient is then set to be constant in time while its spatial values decrease as the cells approaches the river outlet, taking values between Nmin=0.04 and Nmax=0.06 (Decharme et al 2012). ~~(following the river network).~~

m#8. p. 5, l. 10 replace “as the cells approaches” for “towards”

Thank you this suggestion. The modification was made into the manuscript.

m#9. p. 5, l 11. V(t) is not the surface flow, but the average cross-sectional flow velocity.

You are right. By “surface”, we meant the flow velocity in CTRIP’s surface reservoir S. The confusion is now corrected in the manuscript.

- *Modified manuscript in section 2.2 (p.5, l.11-12)*

All these parameters are eventually essential to estimate the spatially- and time-varying ~~surface~~ average cross-sectional flow velocity in the surface reservoir v(t) following the Manning formula.

m#10. p. 5, eq. 3. S is not defined

Actually, it was introduced p4, l.17, but it is true that it is hidden within the text and does not clearly appear as a variable. Therefore, we added it in p.5

- *Modified manuscript in section 2.2 (p.5, l.14)*

where hS is the river water depth estimated from the river storage S by

m#11. p. 6, l. 3. “forcings are considered perfect”. This is my point above. They never are and the assimilation should work to correct these forcings. In the case of ISBA this concerns errors in the water balance, and in CTRIP errors in the transport of mass and momentum through the channel network.

We agree with the reviewer that the forcings are never perfect. By “perfect” here, we meant that the forcing uncertainties are not included in the generation of the ensemble, but they should definitely be included in future studies. First, in the manuscript, we withdraw the use of the term “perfect” when presenting the forcing (section 2.4) but instead write it is considered “as such” in the data assimilation framework (section 3.4.3). Then, following a similar remark from reviewer #2, we dedicate a paragraph to the “perfect” forcing assumption in the discussions. Finally, the objective to directly correct these forcings with assimilation is here out of the scope of the study but it is pointed out in the discussions.

- *Modified manuscript in section 2.4. (p6, l.2-4)*

~~In the entire study, those forcing are considered perfect.~~

- *Modified manuscript in section 3.4.3. (p10, l.3)*

To generate the background control ensemble, we solely stochastically perturb the variables within the control vector. Note that, by only perturbing the variables in the control vectors to generate the ensemble, we assume that all other features of the forward model, e.g. the atmospheric forcings, the LSM structure and therefore the surface and sub-surface runoff, are perfect. While this is the case for purely OSSE, such features are never perfect in real-case experiment. This assumption is further discussed in the Section 6.

- *Modified discussions >> see attached file: Hess-2019-242-corrected-discussions.pdf*

m#12. p. 6, l. 26. "white noise", is this a reasonable assumption? And if reality is different, how would it affect your results? Discuss this in Section 7.

When the study was developed, there was, at the time, no large-scale SWOT simulator allowing considering correlated SWOT-like errors so the white noise assumption was the most reasonable, in default of having a better error model. Nevertheless, reviewer #2 had a similar comment inviting us to discuss more those observation errors in the Discussions; therefore we added a dedicated paragraph in the Discussions.

- *Modified discussions >> see attached file: Hess-2019-242-corrected-discussions.pdf*

m#13. p. 11, eq 17 and 18. Are these the selected zonal Manning;s n values? These are unrealistically high. Why are these selected in such a strange domain?

No, those are the multiplying factors of the Manning coefficient distribution. This explains why their values are around 1.0. Following the modifications from M#2, this is now explicitly specified in the Section.

m#14. p. 11, 19. Describe briefly what your expected results are (i.e. why these experiments) on both water levels and flows!

Following the reviewer's suggestion, we added a few sentences in the Section 4.2.

- *Modified manuscript in section 4.2. (p11, l.20-24)*

The first experiment, denoted as PE1, is configured from the aforementionned sensitivity test outcome. The parameters defining the experiment (spinup, starting date, ensemble size, control error) will be those giving the best results in the sensitivity tests in Table 2. Also, the reference level between the observed and simulated water depths is the same. In other words, there is no bias in the observation. This first idealized experiment serves as proof-of-concept as the observations nature matches exactly the type of the simulated variables. Consequently, with this experiment, we expect to retrieve the true value of the control variables and hence the correct water depths and discharges.

- *Modified manuscript in section 4.2. (p12, l.8-10)*

First, in experiment PE2, there will still be no bias between the observeo and simulated river bathymetry to see how the assimilation of anomalies performs. Similarly to PE1, we expect this experiment to be able to retrieve the true control and state variables. Finally, the last experiment PE3, which introduces a constant relative bias between CTRIP and SWOT, will be carried out. For this experiment, we anticipate that the assimilation will still be able to retrieve the model state variables. The use of anomalies as observations should limit the impact of the inserted bias however, we do not exclude that it may be slightly echoed on the control variables.

m#15. Section 5.1. Describe what you I hydrological sense expect from the spinup time experiment. You can relate the expected required spinup to the time of concentration of the considered basin.

Thank you for this suggestion. We added the following sentences at the end of section 5.1.

- *Additional remark in section 5.1 (p.13, l.5)*

This period corresponds to the basin concentration time or, in other terms, the required time for the river network to totally empty.

m#16. p. 13. L. 9, you only show spatial average results. Why not spatial patterns? That may reveal the locations where things go right / wrong.

You are right. Initially, we chose to display the results averaged over the basin to keep the figure easy to read and limit the number of figures within the manuscript. However, these plots can definitively be generated. Therefore, we added the corresponding figure in appendix.

- *Added figure >> see attached file*

m#17. p. 15, l. 16-28. I find this paragraph not very clear, it is not clear why the results behave so differently from zone 1,2,3. I was wondering if there is not simply too little water in the system to get correct results in this part of the domain? If you only change manning's n, you can never introduce new water or take water out of the system (see main comment).

You are right. We focused on the sensitivity analysis results to explain our results but it is true that, during the low flow period, there might be just too little water entering the system and the assimilation is unable to perform efficiently probably due to a forecast ensemble that is not spread enough. The section was slightly modified to include this comment and to simplify the interpretation.

- *Modified manuscript in section 6.1 (p.15, l.16-28)*

Subsequently, zones 6, 7 and 8 correspond to right-bank tributaries, namely the Juruá and Purus rivers (zone 6), the Madeira river (zone 7) and the Tapajós and Xingu rivers (zone 8). These right-bank tributary zones are characterized by a strong seasonal cycle (see Figure 8, zones 6-8). Then, by comparing the corresponding plots in Figures 7 and 8, we notice that the period when the analysis control variable spreads from the truth corresponds to the low flow season in these zones. According to the global sensitivity analysis results, water depths in these zones are ~~not~~less sensitive to the Manning coefficient in low flow conditions. Additionally, there is very little water in the zones during this period. Consequently, the background control ensemble is not spread enough for the EnKF to be efficient. Meanwhile, the EnKF still "sees" ~~a positive discrepancy between the model and the observations (i.e.~~ that the observations are higher than the model predictions~~),~~ (as seen with the positive innovation in these zones shown ~~in the time evolution per zone of the innovation~~ in Figure A1).Therefore, in order to increase the simulated water depths, the EnKF corrects the Manning coefficient so that its value gets higher (a higher Manning coefficient means a slower flow velocity and then a higher simulated water depth). ~~However, the water being insensitive to the Manning coefficient during this period, the correction is not transferred to the simulated water depth and the Manning coefficient value keeps being increased during the low flow season.~~ Finally, once the low flow season ends, the analysis Manning coefficient converges back to the truth (see the last assimilation cycles).

m#18. p. 17, l. 20. Around this part you should definitely discuss the state updating versus parameter updating, and the water storage errors that you can never resolve with parameter updating.

Following the reviewer's suggestion, we completed the Discussions paragraph on the forcing and LSM bias.

- *Modified discussions >> see attached file: Hess-2019-242-corrected-discussions.pdf*

m#19. p. 17, l. 31. I am very curious what kind of exceptional hydrological event you mean here

Maybe the term was not the most adapted here. By "exceptional", we meant intense flooding for example. We also changed the term by "extreme".

m#20. Some of the figures have too small fonts, please make the figures readible throughout the text.

Following the reviewer's suggestion and a similar comment from reviewer #2, we re-worked the figures, namely Figures 7 to 10.

- *Modified figures >> see attached file*