Responses to Reviewer 2 comments on Manuscript HESS-2019-207

Title: Hydrologic-Land Surface Modelling of a Complex System under Precipitation Uncertainty: A Case Study of the Saskatchewan River Basin, Canada

Authors: Fuad Yassin et al

Manuscript No: hess-2019-207

The review comments are in regular bold typeface, while all responses are in italics and indented paragraphs.

Response to Reviewer 2

General Comments:

Yassin and his colleagues reported their research work conducted by the MESH model. When I first read it, I was quickly saturated with too many trivial details, which are probably very important, but I couldn't remember any of them. When I read the paper again, I found several new things which were hidden among the ocean of numbers: 1) the authors improved the MESH model, by involving irrigation and flow diversion modules; 2) the new MESH model without calibration seems work well not only in streamflow simulation, but also to reproduce ET and TWS. Surprisingly, I did not find any reflection of these innovations in the title, also not any highlight in the main text. More weird. I did not see the comparisons between the original MESH model and the new MESH model.

> *We thank the reviewer for reviewing our manuscript and providing his/her valuable comments. In particular, we appreciate that the reviewer acknowledges there are certain novelties and innovations in this work. We agree that the presentation of the materials and the highlights was sub-optimal in the original version, and there were probably too many details presented that were not warranted. Therefore, we intend to significantly shorten the manuscript in a possible revised version and focus mainly on the innovations. Further, we understand that showing comparison results between the improved model and the original model is important to show the robustness and superiority of the improved model over the original. In response to the reviewer's comments (also pointed out by Reviewer 1), we will include the comparison results between the improved model (including water management) and the original MESH (no water management) in the appendix and keep our main focus on the comprehensive three-stage evaluation strategy for the improved MESH model in the main text.*

Regarding the title, we are considering revising the title, perhaps to something like "Hydrologic-Land Surface Modelling of Complex, Heavily Managed Watershed Systems: Addressing Human Interventions and Precipitation Error". We welcome the reviewer's opinions and suggestions on the title.

Generally, the authors did good research, and a lot of work. But the paper reads like an experimental report, rather than a research article. What can we learn from this paper? It is not necessary to show all the simulated data from the model. What messages did the authors want to deliver to readers? Did the authors want to report their finding something like: "CaPA is the best choice to conduct hydrological research in Saskatchewan River Basin"? If this is the main take-home message, I don't think this paper deserves to be accepted by HESS. Therefore, I suggest that a substantial major revision is needed before further consideration.

We appreciate the value of the reviewer's comment on our work and we agree that the length of the manuscript might affect the efficiency of delivering our main contribution to the readers. Here, we highlight the significance of our work the advances it offers in the the diagnosis of an improved Canadian H-LSM (i.e. MESH with the inclusion of irrigation and flow diversion modules) in modelling the highly complex river system in western Canada (i.e. SaskRB) with the consideration of errors propagation from the precipitation inputs. These advances were shown by presenting a three-stage evaluation strategy for an improved H-LSM. The first-stage evaluation was to assess the error characteristics of several precipitation candidates through the direct and in-direct evaluation methods before calibration. Such evaluation is rarely done in previous studies, especially for process-based H-LSMs that model large-scale heavily-regulated basins. The second-stage evaluation was to conduct a multi-objective multi-station optimization approach using as many streamflow stations as possible for improved model performance and the third-stage evaluation was to further evaluate the model performance by validating the spatial model outputs with additional information from the GRACE data and two evapotranspiration data. Calibrating an H-LSM with multiple stations across a large-scale river basin and validating its spatial outputs are not commonly done in previous studies mainly because H-LSM parameter estimation through calibration is still in its infancy stage. In response to the reviewer's comments, we will vigorously shorten our manuscript (as shown in responding comment 1 in Other Comments Section) and we will highlight the significance of our work in the end of the Introduction Section [P5L17-19].

In addition, inspired by the Reviewer 1's comment 5, we think that the current presentation of our objectives might not fully reflect the main goal of our study and hence reduce the creditability of delivering the main messages to the readers. We will revise the presentation of our study objectives [P4L1-17] to better reflect our work, which is shown as follows:

The aim of this paper is to present a three-stage evaluation strategy for ~~conduct a detailed analysis and evaluation of~~ a physically-based H-LSM ~~for~~ over a highly-managed, large-scale basin, using state-of-the-art calibration strategies and multiple data sources to enable quantification of modelling uncertainty. Such analysis is essential to benchmark model performance, to examine water security vulnerabilities under future conditions, to serve as a test-bed (experimental basin) for the improvement testing of different model process, and to evaluate new datasets. Additionally, such analysis helps to inform H-LSM applications for hydrologic operational forecasts and the management of large-scale basin water resources.

The three-stage evaluation strategy consists of three specific objectives, which is shown as follows~~ of this paper are as follows~~:

- *To identify a suitable precipitation dataset for the H-LSM modeling based on: (1) precipitation error characteristics against ground-based observation, and (2) performance measure criteria based on streamflow simulation when used to drive default parametrized H-LSM.*
- ~~*To identify the most accurate precipitation dataset by evaluating error characteristics of multiple gridded precipitation datasets against ground-based observations.*~~
- ~~*To evaluate the quality of gridded precipitation datasets in terms of how well they reproduce observations of multiple streamflow gauges when used to drive an H-LSM.*~~
- *To conduct a multi-objective multi-station optimization approach, ~~To improve the H-LSM parameterization using a state-of-the-art computationally-efficient calibration approach,~~ and evaluate the effectiveness of parameter transferability through validation in time and space, using independent multiple streamflow gauges not used in calibration.*
- *To test the model performance using multiple sources of observational information on model storage and output fluxes, to ensure that the optimal parameters obtained are as realistic as possible (giving the "right answers for the right reasons") without error compensation across multiple outputs.*

Subsequently, we will also revise the abstract [P1L14-26] to reflect the changes that address both reviewers' comments:

> *A three-stage evaluation strategy ~~analysis~~ of the MESH model performance was carried out ~~in two steps~~. First, the reliability of multiple precipitation products was evaluated against climate station observations and based on their performance in simulating streamflow across the basin when forcing the MESH model with a default parameterization. Second, a state-of-the-art multi-criteria multi-station optimization ~~calibration~~ approach was*

applied, using multiple streamflow gauge stations across the river basin ~~various observational information including streamflow, storage and fluxes~~ for calibration ~~and validation~~. Third, various observational information including storage and fluxes were used for further model validation. ~~The first analysis shows that the quality of precipitation products had a direct and immediate impact on simulation performance for the basin headwaters but effects were dampened when going downstream. In particular, the Canadian Precipitation Analysis (CaPA) performed the best among the precipitation products in capturing timings and minimizing the magnitude of error against observation, despite a general underestimation of precipitation amount. The subsequent analyses show that the MESH model was able to capture observed responses of multiple fluxes and storage across the basin using a global multi-station calibration method. Despite poorer performance in some basins, the global parameterization generally achieved better model performance than a default model parameterization. Validation using storage anomaly and evapotranspiration generally showed strong correlation with observations, but revealed potential deficiencies in the simulation of storage anomaly over open water areas.~~ The first-stage evaluation revealed the different error characteristics of precipitation datasets that are directly propagate to H-LSM modeling, and allowed identify the better precipitation dataset candidates for better H-LSM modelling. The comprehensive analysis in the subsequent stages demonstrated the capability of MESH (H-LSM) to model highly regulated and complex basin as well as the possibility to improve the model simulation through global multi-station parametrization than a default model parametrization, while revealing potential deficiencies in simulation of water storage anomaly over open water areas.

Other Comments:

(1) **The paper is too long to read (40 pages), and quite easy to drain readers' energy and patience. It needs substantial shortening and condensing.**

We agree that the manuscript is lengthy and we will vigorously shorten our manuscript to improve the readability. We will focus on shortening and revising the Results and Discussion Section.

(2) **Figure 1 is not clear. Please make sure all the words in the figure can be read.**

Thanks for the reviewer's comment. We will revise Figure 1 to make sure all the words are visible to the readers.

(3) **Many confusing points. For example, Page 16 line 31: "Such cases, could imply that the errors from the precipitation products were outweighed by other errors." If other errors outweigh precipitation uncertainty, is it convincing to use precipitation as input of the MESH to evaluate the quality of precipitation data?**

We appreciate the value of the reviewer's comment on the validity of evaluating the quality of precipitation data through a process-based H-LSM (as also commented by Reviewer 1). We wanted to highlight the fact that the rationale behind this statement is based on two assumptions we made (as stated in the manuscript P16L24-29) and the belief that we are more confident in a process-based H-LSM in which we have explicitly represented the known hydrological processes and water management practices in the model using our best knowledge and understanding on those processes (e.g. snow processes, vegetation and soil processes, irrigation). We think that the basis of the two assumptions is reasonable, and therefore, we could obtain new information and insight about the quality of the precipitation products through a process-based H-LSM. However, we are also very aware of the deficiencies of the MESH model through our work (as discussed in Sections 6.3 and 6.4) and we acknowledge that caution is needed when interpreting the results of precipitation products, particularly in any sub-basins where model deficiencies are found.

We will clarify the validity of evaluating the quality of precipitation data through a process-based H-LSM by extending the discussion in Section 6.2 [P1623-34; P17L1-4].