Response to reviewers' comments on "The effect of input data complexity on the uncertainty in simulated streamflow in a humid, vegetated watershed"

Linh Hoang, Rajith Mukundan, Karen E. B. Moore[2], Emmet M. Owens and Tammo S. Steenhuis

Point by point response (indicated by ➔) to reviewers' comments

Please note that the page and line numbers referred to here are from the revised manuscript.

RESPONSE TO REVIEWER 1

W. Vervoort (Referee)

willem.vervoort@sydney.edu.au

Comment

This paper is the revised version of the earlier manuscript. Overall, the authors have done a great job at addressing the reviewer's concerns, and there are only some minor concerns remaining. In general, the paper should be accepted, once these minor issues are resolved.

One issue is that by addressing the reviewer's comments, some level of repetitiveness has crept into the discussion. The authors could remove this by doing some rearrangement of the text. For example, the issue of scale (as indicated in my comments on the pdf) should probably be raised earlier and this would allow reducing the number of mentions of studies that support the findings. I think not all of these are needed and this would make the discussion slightly better to read.

There are a few other comments and some editing that might need to be done, but overall the manuscript is in good shape.

Response

➔ Thank you very much for the reviewer's compliments on our responses to previous comments.

We copied all the comments in the attached pdf excluding the comments on grammar and English as below and responded to them one by one.

Comment

So, in essence there was no real calibration, as you did not specifically optimise using a search algorithm. So while the result is the "best from the MC runs, there could still be a better solution out there. I think this is worth a mention somewhere in the discussion. In addition, was each DEM variation run with the same MC sets?

Response

➔ We respectfully disagree with the reviewer that there was no real calibration. SWAT-CUP, the software that carries out auto calibration and uncertainty analysis for the SWAT model, uses the GLUE approach that is based on Monte Carlo runs as one of its methodologies for auto-calibration. It is true that our results show the best results from Monte Carlo runs, and there could be still better solutions out there. However, we believe that the better solutions, if they exist, are not significantly better, because the number of Monte Carlo parameter sets are high enough to cover the parameter spaces. Moreover, in the context of hydrological models, there is no optimum parameter set but a large set of parameter vectors with reasonably good performance (Beven and Freer, 2001; Bárdossy and Singh, 2008) due to the non-linearity of the parameters and their inter-dependence. The equifinality problem is very common in hydrological modeling that leads to high uncertainties in the model predictions.

In this study, we did not aim at comparing our setups using their optimal parameter sets, but compared them with consideration of parameter uncertainty. Using the GLUE method, we could use the same number of Monte Carlo runs, and same values of parameter sets for all setups to ensure a fair comparison.

We confirm for the reviewer that we used the same Monte Carlo parameter sets for all four DEM setups.

Comment

P13 L4-5: "*Although the effect of DEMs on streamflow prediction is minor, the setups using coarser resolution DEM10m and DEM30m are slightly better and preferred for application.*" Are you going to explain why they are better? It has to do with the smoothing, so worth discussing.

Response

➔ Yes, we added an extensive explanation about why coarser resolution DEMs are better than the finer ones. Please see the discussion in **4.1. What is the most suitable DEM resolution to use in SWAT-HS?**

Comment

P13 L18-19: "The 'good' parameter sets in four setups should give comparable predictions of streamflow," Not necessarily as it might affect the different flow components

Response

➔ Yes, we agree that the '*good*' parameter sets can give different results of flow components. However, what we mean here is that the '*good*' parameter sets should give comparable predictions of overall streamflow, because 'good' parameter sets are chosen based on their good fit to observations.

We revised the sentence as:

"The '*good*' parameter sets in four setups should give comparable predictions of overall streamflow, percentage of watershed area that is saturated, and the time that each wetness class was saturated, which results in similar probability of saturation." **(P14, L1-4)**

Comment

P14 L19-21: "*For coarser resolution DEM10m and DEM30m, saturated areas connect well with each other and with the areas concentrated near streams*". Which presumably is a result of the more connected wetness classes as a result of the smoother DEM.

Response

➔ Yes, we agree with the reviewer. Our discussion for this effect is shown in the discussion **4.1. What is the most suitable DEM resolution to use in SWAT-HS?**

Comment

P14 L25-26: *"Among the four DEMs, DEM10m provides the most realistic representation of saturated areas and reasonable fit to observations."* This is of course slightly qualitative and assumes that what you call "saturated area" in the "observations" is in fact spatially homogeneous, as represented by the model. Maybe the scattered saturated areas in the finer DEMs actually represents the true nature of the "saturated areas"?

Response

➔ We respectfully disagree with the reviewer that the scattered saturated areas in the finer DEMs probably represent the true nature of the "saturated areas". As shown in Figure 8 in the revised manuscript, the saturated areas predicted in DEM1m setup are distributed broadly throughout the watershed area. This can be explained based on Figure 12 showing the relationship between topographic index (TI) and elevation. It is clearly observed in Figure 12 that the distribution of TI values in DEM1m spread more broadly than DEM 10m at all elevations, thus, wetness classes derived from the categorization of TI values have the same pattern. This results in the occurrence of saturated areas in a wide range of elevations when DEM1m is used. However, our observations did not indicate the occurrence of saturated areas in high elevation areas. Therefore, we believe that the finer DEMs (particularly DEM1m) do not represent the true nature of the saturated areas.

Comment

P18 L10-11: *"…we can see that the distribution of TI values in DEM 1m spread out wider than in DEM10m at all elevations."* So are you suggesting that for these finer DEM values you are actually capturing some fine scale topography, which generate locally higher wetness classes. I guess you are arguing that this level of detail in not needed for the streamflow simulation at the catchment level. So in some way you are arguing there is a relationship between scale of the modelled watershed and required accuracy of the DEM. In other words, if you would study a much smaller or much larger watershed, would you answer on the DEM accuracy be different?

Response

➔ Yes, we believe that the required accuracy of DEM depends on the scale of the watershed and also the characteristics of the watershed (more particularly, the dominant hydrological process in the watershed). We added a discussion on this issue in the revised manuscript **(P19 L11 – P20 L16)**

Comment

P18 L13-15: "*Realistically, the highest TI value grids should be located in downslope, near-stream, low elevation areas while the lowest TI value grids should be in upslope, high elevation areas.*". Only when you observe this at a certain scale! You are really discussing fractals here (in a way :-))

Response

➔ We deleted this sentence in the revised manuscript.

Comment

P18 L15-17: "*…Therefore, in this case study, the coarser DEMs (DEM10m and 30m) give a better and more realistic representation of the landscape than the finer DEMs (DEM1m and 3m)…*" Not sure if I would call it "better and more realistic". I think it is "more suitable to the scale of the watershed and the simulation"

Response

➔ We revised this sentence as:

"Because of that, in this case study, the coarser DEMs (DEM10m and 30m) give a more suitable representation of the landscape than the finer DEMs (DEM1m and 3m)." **(P18 L20-22)**

Comment

P19 L8-10: "*Our results show that DEM10m is the best choice among four DEMs tested because of its slightly better performance for streamflow and more importantly, its good fit to observations of saturated areas*." At the scale of the studied watershed!

Response

➔ We revised this sentence to clarify that our results are applicable for the scale of our studied watershed as:

"For the Town Brook watershed, DEM10m is the best choice among four DEMs tested because of its slightly better performance for streamflow and more importantly, its good fit to observations of saturated areas." **(P19 L11-13)**

Comment

P19 L12-14: "*Therefore, DEM10m is the preferred choice to scale-up the application of SWAT-HS to larger watersheds in the New York City water supply system for future applications.*" You might want to test this! It again depends on the scale of the watershed!

Response

➔ We upscaled the application of SWAT-HS to a larger scale watershed (the Cannonsville watershed in which the Town Brook watershed is the headwater area) to simulate streamflow and phosphorus (the manuscript is currently under review). We got very good streamflow prediction for the Cannonsville watershed. However, since the observations of saturated areas are not available at this large-scale watershed, we could not confirm the suitability of using DEM10m to simulate saturated areas for this watershed.

Comment

P20 L4-5: "*Therefore, the sensitivity of DEM resolution may depend on the scale and characteristics of the watershed*." Yes indeed, I think you could mention this much earlier and that would remove some of the repetition in the above section

Response

➔ We tried to bring this discussion earlier (after the first paragraph of section 4.1 as your suggestion). However, we feel that this move interrupted the flow of discussion that we want to tell the readers. We would like to discuss the reason why DEM resolution affects significantly on spatial prediction before discussing about the possible impact of scale on the sensitivity of DEM resolution. We thank you but are sorry that we could not follow your suggestion.

Comment

P22 L15-17: *"Our results show that regardless of the level of detail of input data, we obtained numerous sets of parameter values that give equally good performance for streamflow and saturated area predictions."* I know you used "a lot" of parameter sets, but could this result be at all related to the fact that you don't strictly "optimise" in the calibration? In other words, there is no search algorithm that specifically searches for the "best solution". We can debate whether a "best solution" exists, but in your case, we cannot be sure there is not possibly a single "best solution".

Response

➔ To reiterate our response to your previous comment, in the context of hydrological models, there is no optimum parameter set but a large set of parameter vectors with reasonably good performance (Beven and Freer, 2001; Bárdossy and Singh, 2008) due to the non-linearity of the parameters and their inter-dependence. The equifinality problem is very common in hydrological modeling that leads to high uncertainties in the model predictions. In this study, we did not aim at comparing our setups using their optimal parameter sets, but comparing them with consideration of parameter uncertainty.

Comment

P22 L19-20: "*The number of randomly generated Monte Carlo parameter sets is sufficiently high to give a good coverage of parameter space.*" Do you show this anywhere? Or test this?

Response

➔ The calibration procedure in this manuscript is based on our previous study (Hoang et al., 2017). Our calibration procedure includes 2 stages: snowmelt calibration (5 parameters) and flow calibration (9 parameters among which 1 parameter, ALPHA_BF, is not very sensitive based on our previous study). In each stage, we generated 10,000 parameter sets which were run with SWAT-HS and the results were compared with observations.

We assume that we generate 1 random sample in at least high, middle, and low range for each parameter. In the stage of snow melt calibration, with 5 parameters involved, 3^5 = 243 is the minimum number of MC parameter sets required to cover the parameter space. In the flow calibration stage, with 8 sensitive parameters involved, 6561 combinations are the minimum number required. Therefore, we think that using 10,000 MC parameter sets for each stage of calibration is sufficiently high to give a good coverage of parameter space.

Anonymous Referee #2

Comment

I thank the authors for the careful answers and explanations to my comments regarding the first version of their manuscript. I am satisfied with the answers and the adaptations made in the manuscript. I have mainly three concerns left, which I think would be worth to address in the manuscript:

Response

➔ We also thank the reviewer for your valuable comments which helped to improve our manuscript. We have tried our best to address your three remaining concerns.

Comment

Both reviewers asked for the reason for choosing an NSE value of 0.65 as a threshold for good simulations. I assume that many readers will later on have the same question. From this perspective, I would highly recommend that the authors add a short statement on the reason for choosing NSE 0.65 in their manuscript.

Response

➔ We added two sentences in the revised manuscript to clarify our choice of NSE value of 0.65 as a threshold for good simulations:

"Our choice of NSE threshold at 0.65 is based on the guideline for model performance evaluation by Moriasi et al. (2007) that suggested "*good*" model performance for streamflow as corresponding to monthly NSE higher than 0.65. As NSE values at the monthly time step are usually higher than the daily values, we believe that that our choice of NSE higher than 0.65 as "*good*" model performance at daily time step is a reasonable choice." **(P9 L25- P10 L2)**

Comment

2) I like that the authors provided a more detailed explanation of the two-stage calibration process. However, this raises the question (as also pointed out be reviewer 1) why a two-stage calibration is needed if there is only streamflow data, but no snow data available. I think it would be worth to address this question when explaining model calibration. Furthermore, I would also mention how you selected the values of the 9 flow parameters during the first calibration round. Did you use default values?

Response

➔ We chose to calibrate in two stages because we did not want to include a lot of parameters in one calibration which requires a much higher number of Monte Carlo runs. Moreover, this watershed is located in a region that has 1/3 of the total precipitation falling as snow. Therefore, prediction of snow storage and snowmelt will significantly affect the prediction of the timing and volume of streamflow in winter and early spring. Consequently, we decided to divide the calibration in two stages to reduce the number of Monte Carlo runs and to focus on calibrate snow processes/parameters before adjusting other processes/parameters.

We added 2 sentences in the revised manuscript to clarify the reason that we divided the calibration into 2 stages:

"Since the Town Brook watershed is located in a region that is heavily impacted by snow, the prediction of snow storage and snowmelt will significantly affect the timing and volume of predicted streamflow in winter and early spring. Consequently, we divided the calibration in 2 stages in order to reduce the number of calibrated parameters involved in one calibration and to focus on getting the right results for snow processes before adjusting other processes." **(P9 L1-6)**

Comment

3) As mentioned in the first review round, I recommend to consistently use (HESS guidelines for) units and references. E.g. DEM resolution of 30 m vs. 30m DEM, mm/°C vs. days-1 (in Table 1), group(i) vs. group(2) at P16 L6, some references with doi while others without, etc.

Response

➔ We had a careful look at the HESS guidelines, and had a careful check throughout the manuscript to correct the remaining mistakes on the units, figures, tables, and references.

REFERENCES

Bárdossy, A., and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrology and Earth System Science, 12, 1273-1283, https://doi.org/10.5194/hess-12-1273-2008, 2008.

Beven, K., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11-29, http://doi.org/10.1016/S0022-1694(01)00421-8, 2001.

Hoang, L., Schneiderman, E. M., Moore, K. E. B., Mukundan, R., Owens, E. M., and Steenhuis, T. S.: Predicting saturation-excess runoff distribution with a lumped hillslope model: SWAT-HS, Hydrological Processes, 31, 2226-2243, https://doi.org/10.1002/hyp.11179 2017.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885-900, http://doi.org/10.13031/2013.23153, 2007.