Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-393-RC3, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "The role and value of distributed precipitation data for hydrological models" *by* Ralf Loritz et al.

Wouter Knoben (Referee)

wouter.knoben@usask.ca

Received and published: 11 September 2020

Summary

The authors develop and test a hydrological model that is able to change its spatial complexity in time. In its most simple state, the model represents the Colpach catchment in Luxembourgh as a single representative hillslope. In its most complex state, the model would be able to use 42 hillslope elements to simulate the catchment's response to extremely spatially variable rainfall inputs. The model adds hillslope elements based on the spatial complexity of incoming precipitation and removes hillslope elements based on the change of runoff over time. Both processes use a threshold to decide when upscaling or downscaling the model is needed or possible. The authors

Printer-friendly version

Discussion paper

show that the adaptive model reaches the same KGE scores as a fully distributed model that uses 42 hillslope elements all the time, while the adaptive model needs 16 representative hillslopes at most. This is shown for two short-duration event that occurred during summer.

I have read this paper with much interest and found it generally easy to read and understand. As models grow more complex, computation times go up and studies such as this could open up great opportunities to reduce computation costs by avoiding redundancy in model calculations. However, I have some questions about the tests and metric the authors use to show that the adaptive model is as good as the fully distributed one. These are outlined below. I've provided additional requests for clarification in the line-by-line comments in the hopes that these are helpful.

Kind regards,

Wouter Knoben

Comments

My main concern is the choice of using dQ/dt to reduce the number of model elements. Using the change in discharge over time to measure similarity of states can only work if there is a unique relationship between model state and dQ/dt. Given the equifinality in the fluxes-discharge relation that's typically visible in hydrological models (see e.g. Khatami et al., 2020), I think the section that introduces this concept (P16, l17) is not quite clear about why this dQ/dt assumption can be used together with CATFLOW.

Reading further, the authors address this concern to some extent in section 4.4 (P23, l18). This section however seems to show that CATFLOW does not exhibit such a unique relationship and the model reduces the number of model elements before the groundwater states reach similarity. This does apparently not affect the quality of the simulations much, because the KGE scores in Table 1 seem to indicate the adaptive model is as good as the fully distributed model for the two testing events.

Fig. 4 shows that both testing events are selected in the middle of summer, when presumably the catchment is in quite a dry state (catchment state is not mentioned when selection of the two events is discussed on P18, l20 to P19, l6). The fact that both events are selected during the dry summer could mean that the model can reset itself to mostly empty between the events and as such the long term (seasonal) impacts of not keeping the groundwater states separate can not be investigated with the current two testing events.

Equally, the events concern high flows so the impact of differences in slow groundwater states probably do not register in the dQ/dt values during the falling limb of the hydrograph (and thus the adaptive model simplifies itself).

There is the compounding issue that the KGE scores used to calculate the performance of model c are only calculated during the high flow event and that metrics such as KGE are typically not very sensitive to errors in low flows. This means that the parts of the simulation time series where the differences in groundwater states could be seen are both not included in calculation of the KGE score of model c and if they were, the KGE metric might not be able to pick up on any differences.

Summarizing the above, I'm not sure that the dQ/dt criterion is entirely appropriate to determine when the adaptive model can reduce its complexity, and I'm equally unsure if the current two testing events would be able to show if the dQ/dt criterion is or is not appropriate. The straightforward solution would be to run model c for the year, add these results to Table 1 and briefly investigate for example the relative contributions of different fluxes to the overall water balance and the model's response to a few precipitation events during winter. Given that the adaptive model should be faster than the fully distributed one, this should not be a large computational burden and it will provide a much more complete impression of the capabilities of the adaptive model.

Line-by-line

P5, l5. This question seems quite strongly related to the contrasting results in the literature that the authors discuss in the first and second paragraph of the introduction, where they conclude that the impact of using a distributed model and/or distributed forcing data is conditional on the catchment under investigation. This research question seems a bit generic in that light, given that only a single model and catchment are being investigated in this work. As is, question 1 seems more like a formality to me (it must be answered with "yes" before Q2 can be investigated) and the main focus of the manuscript seems to be on Q2. Perhaps the manuscript can gain a bit in focus if only the current research question 2 is specified, and the work done to answer the current Q1 is presented as a prerequisite to address the current question 2. For example, "We test this hypothesis by first showing that the model CATFLOW applied to the 19.4 km2 Colpach catchment using a gridded radar-based quantitative rainfall estimate improves in performance when it is distributed in space and driven by distributed rainfall. We then address the following research question: "Can adaptive clustering be used to distribute a bottom-up model in space that it is capable to represent relevant spatial differences in the system state and precipitation forcing at the least sufficient resolution to avoid being highly redundant as a fully distributed model?"

P5, l14. Assuming that "> 1 m" refers to soil depth, should it be "< 1 m"?

P7, l9. Which numerical scheme is used by CATFLOW?

P7, l20. If possible without using too much space, it might be helpful to the reader to briefly summarize the main findings of Loritz et al. (2017).

P7, l21. What are the outcomes of this quality control?

P7, l28. I'm not quite sure I understand why these distances are given as a range if only a single station is concerned. Does this indicate minimum and maximum distance of the catchment bounds to each radar station?

P9, l13. I find this sentence a bit hard to follow. Is the part from "apart from . . ." onwards necessary here? This is already discussed in the introduction.

P10, l21. Why is the model tested during two events instead of over the full year? How were these events selected?

P11, l14. The conclusion that a distributed model is needed to account for runoff driven by convective precipitation would be stronger if the authors can (briefly) list which processes are represented at too coarse a scale in the reference model for it to properly deal with convective precipitation.

P11, l14. I believe this sentence would be more complete if it also explicitly mentioned that distributed instead of catchment-averaged precipitation data is needed to properly simulate the result of convective precipitation events.

P11, l27. It would be helpful for the reader to repeat that the only difference between reference model and model a is the choice of precipitation data.

P12, l3. Are these variables similar or identical to those used in the reference model?

P12, l4. To clarify, does this mean that model b is run in a gridded fashion with the catchment divided into 42 grids (matching the precipitation grid)? If not, it would be good to clarify this in the text and mention the number of model elements that the precip field similarity approach gives. Line 18 on this page could benefit from a similar clarification.

P12, l23. Are there some observations that could help support the choice for 1 m/s?

P14, l29. It might be good to extend this line of reasoning to soil types and vegetation cover, as these are commonly used as model inputs/parameters.

P15, l7. This sentence is quite general (referring to humid environments) and could use a reference. However, if the authors chose 1 mm hr-1 based on their expertise and knowledge about this catchment, then I think it's more accurate (and in no way worse) to phrase this decision along those lines, e.g.: "We chose this threshold as a reasonable value upon which we expect differences in hydrologic behavior, based on our collective understanding of the Colpach catchment."

P17, l10. I think it's import to repeat the similarity condition of dQ/dt here, because for a model that has no unique relation between model state and dQ/dt values this method cannot be applied without accounting for this difference.

P20, l6. The authors use KGE values in this section and Table 1. I'm not sure to what extent the aggregated value is a useful metric for events that last only a handful of time step. It would be good to at least disaggregate the KGE into its correlation, variability and bias components (e.g. quantify what can be qualitatively estimated from Figure 7) to see if the total KGE scores of the individual models are generated by (roughly) the same types of errors in the simulations.

P21, l25. "acceptable" is somewhat subjective because no standard of acceptability has been defined. It might be cleaner to simply report the correlation component of the KGE to quantify to what extent the hydrograph shape is simulated.

P21, l26. This trial of a direct runoff component seems somewhat ad-hoc to me. I don't think this adds anything to the manuscript and that it will take more space than is available to properly justify this change. I suggest to remove these sentences.

P30, l4-24. These sentences seem as if they would be better placed in the introduction or methodology sections.

Editorial

P1, l13. "capable to dynamically adjust" > "capable of dynamically adjusting"

P5, l3. "by addressing" > "to address"

P7, l6. "dominated" > "dominant"

P9, l13. "Moreover, are the model ..." > "Moreover, the model deficits ... (2017) are ..."

P12, l29. "model b, however, is" > "model b but is"

P12, l30. “3.2.1 to 3.2.3” > “3.3.1 to 3.3.3”

P13, l10. “inside’ > “insight”

P14, l18. “similar” > “similarly”

P15, 5. “structural” > “structurally”

P15, l15. “models” > “model elements”

P21, l24. Figure 7 should be moved and renumbered as Figure 5 if it is mentioned here

P21, l25. “acceptable” > “acceptably”

P27, l21. “perspective, not” > “perspective not”

P29, l22, “there” > “their”

References

Khatami, S., Peel, M. C., Peterson, T. J., Western, A. W. (2019). Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. Water Resources Research, 55, 8922–8941. https://doi.org/10.1029/2018WR023750

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-393, 2020.