EGUsphere, referee comment RC2
https://doi.org/10.5194/egusphere-2022-506-RC2, 2022


# Comment on egusphere-2022-506

Anonymous Referee #2

---

Referee comment on "Adjoint-based spatially distributed calibration of a grid GR-based parsimonious hydrological model over 312 French catchments with SMASH platform" by François Colleoni et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-506-RC2, 2022

---

The article presents the application of the SMASH distributed hydrological model on a large set of French catchments. The authors evaluate the performance of two model versions, the original and a modified one, and conclude that the modified version is more efficient. The two models outperform a lumped model, which is also applied to the same catchments.

I have several major concerns about this article. I found that the way the modifications were introduced in the model is overall not justified. There are also several results difficult to understand, typically on the interception store.

I suggest major revision before the article could be reconsidered for publication.

**Major comments**

- Section 2.1: The authors introduced three modifications taken from model versions developed by other authors and tested them all together without explaining why they are all necessary individually. Therefore it is very difficult to understand what brings some improvement in model performance. Were all these modifications actually necessary?
- L134: The authors choose to introduce two routing stores with the same mathematical formulation. In the work they cite by Pushpalatha et al., there is one power-law store and one exponential store. Could the authors explain why they made a different choice here?
- L199 and 219: Nothing is said about the GR5H model tested here nor about the calibration algorithm used. It is difficult to know which version was used and how it was

implemented. Furthermore, the purpose of including this model in the article is not clear. It is stated twice (lines 255 and 262) that GR5H is used as a reference but not for benchmarking purposes (actually I do not understand the difference between the two in the context of this article). If the objective is to show that SMASH is better than a model "taken from the shelf", I wonder why the choice was made to take GR5H. It is very similar to the SMASH unit brick (model structure), as mentioned in the article, but not really the same. Therefore it is difficult to conclude anything from the comparison proposed here. Does the better performance of SMASH comes from the fact that it is distributed or from the fact that there are differences in the structure of the unit brick? I found it would be more useful to test the original and modified versions of the SMASH model structure in a lumped mode to answer the previous question.

- Table 2: I do not understand why the upper bound for the interception store is so high. It is physically a nonsense to have an interception store of 100 mm. An interception store capacity is typically less than 10 mm. If the capacity is that high in the calibration process, it means that this store does not only play the role of an interception store.
- Table 2: I also did not understand how the calibration process ensures that the "fast" (r) and "slow" (l) tanks actually play this role, i.e. that ctr is lower than ctl. If there is no explicit constraint in the optimisation process, there must be catchments where this is not the case, depending on the proportion of the base flow. It may end up in the fast routing store simulating slower recession than the slow routing store. Furthermore, the fact that both stores have the same formulation must generate equivalent parameter sets for some basins that would only need one reservoir, and thus generate indetermination in the response surface.
- Section 4.4: The comments on the greater variability (greater standard deviation) of some parameters are in my view biased by the fact that the parameters vary over very different orders of magnitude. It might have been more appropriate to compare the coefficients of variation to assess which parameters are indeed the most variable relative to their mean value. Some of the comments (L350-352, L357-359, L360-361) remain rather general, and hypotheses that are difficult to verify as they stand. I therefore question the added-value of these comments.
- Tables 4 and 5: I was quite surprised by the average values taken by the capacity of the interception reservoir here (15 to 25 mm sometimes). One could expect a capacity ten times lower. Clearly, the role taken by this reservoir goes beyond a simple interception function, with a possible interaction on the water balance function (exchange in particular) or a smoothing role on rainfall inputs going beyond the interception process. In Ficchi's work cited by the authors, this reservoir had capacities of a few millimetres on average. How can these differences be interpreted? Furthermore, the stability of the average values of the model parameters does not mean that the values are stable when comparing periods basin by basin. Have scatterplots been drawn to check that the apparent stability of the averages is confirmed when looking at the basins individually? Last, for version D, have spatial analyses of the stability of parameter fields been carried out? There may be equivalent fields in terms of mean or variability, but with very different spatial configurations.

**Other comments**

- Asbtract: It could probably be reduced. Some introductory sentences sound very general and some details about the results do not seem essential.

- General: most of the table and figure captions are not detailed enough to fully understand what is shown. This should be improved.
- L5 and 7: VAD is defined twice.
- L20: Unclear what "Uniform basins" means here.
- L46: please explain what semi-lumped means
- L70: "over a large catchment sample"
- L73: "the initial study by"
- L82-84: It is unclear why the sensitivity analysis is performed.
- L91: "a uniform calibration"
- L108: Here again there is no justification why this cell-to-cell routing was chosen. There are many options possible. Furthermore, the authors never discuss the possible interactions between the in-cell routing and the cell-to-cell routing. There are probably many cases where one compensates for the other. The overall stability of all these parameters between periods may give some answers.
- L109: "described in Fig. 1"
- L158 and Eq. A20: The NSE is defined by 1 minus the ratio of quadratic error and variance. The authors should stick to this definition to avoid confusion.
- L180: "continental France" (?)
- L199: The GR5H model is mentioned here without having been presented before.
- L204: QM/PM min/max are not defined. Is it the min/max over the twelve months?
- L208: I understand from this sentence that there was no criterion applied to remove basins subject to artificial influences such as dams. Is it actually the case? How could the model simulate artificial behaviours?
- Figure 3: Add basin boundaries and the main river network for better readability. It is only mentioned in the legend of this figure that there are two sub-samples (upstream and downstream). This should be mentioned in the text.
- L226-229: Is the initialization period of one year sufficient for basins where groundwater is dominant? Often it is not the case for such basins Furthermore it is not clear whether this initialization period was also used in validation.
- L237: "in calibration, i.e. considered as ungauged" (?)
- L254-255, L261-262: Again, I do not understand these sentences (see comment above on the use of GR5H in the article)
- L285-288: I do not understand why this assumption is made without being verified. It should be possible to check whether there is a link, in the sample of basins, between the hydroclimatic differences observed between the two sub-periods and the evolution of model performance.
- L300: "each flood event"
- L302: "Sutcliffe"
- L318-321: How significant is this improvement in S6 over S3? Give some quantitative evidence of this result.
- Tables 4 and 5: Put cr and ml in the same order as Table 2.
- Table 4: Do the ctr and ctl values were compared? The mean values seem very close on some catchments, which may indicate that a single store would be sufficient and that the extra complexity by using two stores is not justified and may even create identification issues. Was this investigated?
- Lines 395-397: Which fixed value would this be in this case? I found this conclusion strange. It should be further investigated in light of the comments above on the odd values of the interception store capacity.
- Lines 437: "French".
- Appendix A6: The formula for NSE and KGE used to illustrate performance is "1 -" the formulas given here.