Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-331-AC1, 2020

Hydrology and Earth System Sciences Discussions
Open Access

HESSD

Interactive comment

Interactive comment on "On the potential of variational calibration for a fully distributed hydrological model: application on a Mediterranean catchment" *by* Maxime Jay-Allemand et al.

Maxime Jay-Allemand et al.

maxime.jay-allemand@irstea.fr

Received and published: 15 January 2020

Answer RC1, Received and published: 28 Nov 2019

We would like to thank the Reviewer for the careful reading. We gratefully consider all Reviewer's comments and suggestions.

1. The paper lacks clear scientific questions. Three points are mentioned by the authors at the end of the introduction: - upgrading the GRD model, - calibrating GRD with a variational approach, and finally, - upgrading the variational approach. These

three objectives are more similar to a model development process that is described in a project report than to science objectives tackled to fill a gap in knowledge in a research article. They are very specific to the chosen model and very general interms of objectives (upgrading and calibrating are broad terms). They do not introduce valuable objectives for a scientific paper. Moreover, the proposed scientific objectives are not supported by a proper experiment protocol: the new GRD model is not compared to any benchmark (i.e. the former GRD model or another model), the calibration with the variational approach is not compared to a calibration with another approach, only with overly simple homogeneous parameter sets, and the improved variational algorithm is not compared with the classical variational approach.

We agree with the reviewer that the scientific questions have not been explicitly formulated. In the original version of the AIGA model (which is the operational model currently utilized by SCHAPI) the mentioned above "overly simple homogeneous parameter set" is actually used. The main scientific question is: could we benefit from considering the spatially distributed set of coefficients instead of the uniform (homogeneous) set of coefficients and to what extent? In particular, does it help to improve the discharge prediction over the catchment area including 'ungauged' locations? To answer this question the predictive performance of the calibrated model involving spatially distributed coefficients against the one involving uniform coefficients has been performed. We specify more clearly this question in the introductory part. The GRD model comes from the GR family and shares the production and the transfer operators. The lumped GR models (GR4J) have been extensively tested in [Perrin C, 2000] and have shown good performance in modeling the hydrological processes. The distributed versions [De Lavenne A 2019 , Lobligeois F. 2014] take into account the spatially distributed rainfall observations at a refined scale. In this context the GRD model (upgraded AIGA model) designed especially for flood prediction has been developed. We acknowledge that no comparison between the upgraded AIGA model (employing the 'cell-to-cell' routing scheme) and the original version (employing the 'cell-to-outlet' routing scheme) is presented in this paper, but this is done to avoid overloading the paper with technical details. Notice that the cell-to-cell routing scheme is a requirement for our distributed model since it is designed to estimate the discharges at any location inside the watershed (involving ungauged sites), not only at the outlet.

2. Concerning suggested comparison ”the calibration with the variational approach” ”to a calibration with another approach”.

The purpose of such comparison is not clear. Any MAP estimator (KF, EnKF, BLUE, etc.) would produce the same result (at least in theory). We assume that a Bayesian estimator (particle filter, MCMC, etc.), is not really feasible in high-dimensions (of course, it would be possible to go into different dimension reduction techniques, but this is not the direction we decided to follow). The useful comparison would probably be to some robust estimation methods, but this robust estimator is yet to be implemented.

3. Concerning ”the improved variational algorithm is not compared with the classical variational approach”.

Our statement about ’upgrading variational approach’ mistakenly comes from our recent reports and is not relevant in the context of this paper which describes earlier results. Here we only use constrained minimization and scaling, but these features simply represent a ’good practice’ approach in solving the calibration problems. So, this statement is relaxed in the introduction.

4. Concerning “These three objectives are more similar to a model development process that is described in a project report”.

It depends. For example, in computational journals (JCP, CMAME, IJNMF, Comp. Geosciences, etc) this is a usual agenda in the majority of cases.

5. This lack of clear scientific questions comes with (in my opinion) a deficient introduction. The abstract is written in a very unusual way.

This remarks is consistent with reviewer 2. We agree that these parts should be improved. The introduction is rewritten to emphasis the objectives, the originality and the results of this study.

6. I highly doubt that In hydrology, the variational estimation method as described above (i.e. including the adjoint model) has not been reported so far. (page 3, line 12 ; see also the conclusions To the best of our knowledge, this is the first time when the variational estimation involving the adjoin sensitivities has been applied in the field of hydrology.). For instance, in this journal, HESS, Castaings et al. (2009) seem to have developed a similar method. In addition, in their paper Castaings et al. cite some other works (see Early applications of the adjoint state method to hydrological systems have been carried out in groundwater hydrology (Chavent, 1974; Carrera and Neuman, 1986; Sun and Yeh, 1990). Nguyen et al. (2016) might also be relevant. I encourage the authors to make a proper literature review on this specific aspect, which is necessary for the HESS audience to identify the added value of this specific study.

We agree with the reviewer and acknowledge that this is a significant flaw from our side. The reason for this 'amnesia' is, probably, that in these papers the hydrological problem is represented by the partial differential equations with a nonlinear source term: a standard formulation typical for classical problems in atmospheric science and oceanography. In our case, we introduced the delay-based 'cell-to-cell' conceptual routing scheme, which represents the difference rather than differential equations. Of course, saying "this is the first time when the variational estimation involving the adjoint sensitivities has been applied in the field of hydrology" is incorrect. There are some other papers besides the mentioned, but still, only a very few. Let us also note, that the adjoint of an operational geophysical model is always an exclusive piece of work, despite availability of the modern Automatic Differentiation tools. The bibliography is completed by the references suggested and some others. We highlight the differences with our work (model, objectives, methodology).

7. The assessment of the performance is not developed. Only NSE values are calculated, while the authors specifically want to address flash floods. No criterion about

peak-over-threshold, timing, intensity, is used. Since the study is already limited to a single watershed, limiting the analysis to a single criterion is out of the standards of nowadays hydrological studies.

We agree these remarks. First, the NSE criteria is not enough for assessing the performance since this criterion was used in calibration. We suggest to also use the KGE criteria. We agree that it makes sense to assess the performances over flood events with others criteria (timing). Our model runs continuously and such analysis would require splitting the discharge series and analyzing a selection of events. This work can be done and added to this paper.

8. The presentation of scores is poor. In figures 3, 4 and 5, the stations are ranked by their performance. It is a pity that we cannot identify anymore the stations. What a hydrologist would like would be to analyze whether there is a difference between experiments for a specific station, whether there are links between performances over upstream/downstream stations, etc. This kind of analyses is impossible to perform from the presented graphs. In addition, the fact that scores are mixed between the two periods (P1 and P2), if I understand well, is even more confusing.

The presentation chosen (results ranked by the value of NSE) was used to keep the analysis simple. While this representation could be suitable for analysis of a large set of catchments, we agree that for one catchment this may not be sufficient (only 4 stations and 2 periods). Thus, we provide a deeper analysis of the model performance by stations and by periods.

9. First, it is clear from figures 6 and 7 that the parameter values are highly different when calibrated over P1 or P2. For many grid points, Cp and Ct can reach the lower bound for a period C4 and the upper bound for the other. The authors blame the change of precipitation between the periods or the chosen model. In my opinion, it simply indicates equifinality. It comes from the fact that not enough information is provided to the algorithm to calibrate 540 x 3 parameter values.

We have exactly the same opinion. Equifinality manifests itself in the fact that for different 'test' signals we obtain quite different estimates of parameters. In the discussion part, we mention that equifinality issues may be reduced by introducing additional constraints to the minimization problem (spatial correlations, better background estimation, removing data outliers, etc.). This is done in more recent work. However one can notice the stability of the routing parameter "v" and some similarities of the others parameters between both periods (higher values at the North and South parts of the watershed).

10. In other words, the optimization problem is ill posed.

Equifinality means non-uniqueness, i.e. given the input, the same or nearly the same observed model output can be generated with different parameter sets. Regularization is used to chose one out of many possible solutions. On the contrary to what the Reviewer said, the calibration problem given by eq.(16) is well-posed. This does not help too much in practice, since the solution depends on the background pâĹŮ . The latter is obtained by considering the uniform coefficients (just 3 in total to calibrate: CÌĎ p , CÌĎ t , vÌĎ) and using the Monte-Carlo to build the joint posterior distribution, from which the mean values are taken as the background values. Even for these three coefficients we have different values of CÌĎ p , CÌĎ t , vÌĎ for different periods P 1 and P 2 , see Tables 2,3. Thus, the problem is not due to over-parametrization.

11. It indicates that calibrating these three parameters over each grid cell is not possible with only discharge time series and the variational algorithm. While the presentation of negative results is interesting

What exactly is the criterion of successful calibration to assess whether the results are negative or positive? For the model which conceptually imitates the natural system, there exist no 'true set' of parameters. If the parameters are quantities of interest (QoI) by themselves, than the results are rather negative. If the QoI is the predicted state (discharge), it is rather positive, as have been shown using cross-validation. It is, of course, a comfortable feeling if for different inputs one could get consistent sets of parameters, i.e. the estimates are stable, but is this so crucial for forecasting purpose ?

12. If different precipitation patterns are the key factor for explaining these results, then the reader has no element to assess that: no mean or extreme precipitation values or even maps are provided.

True, this will be provided. Particular events are studied (hydrograms and hydrographs are added). Moreover we would like to assess if the calibration (and the corresponding optimal solution) could be controlled by one particular event. For that, we compute a sliding NSE criteria over the calibration period and eventually detect if one particular event strongly influence the NSE.

12. The discussion of the results comes with several rude and coarse judgments, not supported by evidence. For instance, the authors blame a structural deficiency of the chosen model (page 16, line 1), saying it is not surprising since the model is conceptual. Then, they state that the hydrological modeling at the cell scale is very primitive. These two elements may explain why the parameter values are so different according to the authors. First, these assertions are very surprising, as the authors did develop the model they used, if I understand well. Second, being simple or conceptual is not necessarily a deficiency. On the opposite, it is often considered as being an advantage, as such models are easier to run or to understand. Third, if the authors identified a structural deficiency, then it has to be shown and analyzed, and solutions for improvement must be discussed.

“Structural deficiency” only means that the model (of a given structure) cannot be used to fit observations by manipulating its coefficients. This can be revealed by considering the ’forecast minus-observations’ residuals. Despite a supposed over-parametrization (3 parameters for each cell), the residuals still contain some regular structures which cannot be explained solely by the data noise. While the presence of a structural deficiency is revealed, its correct interpretation and subsequent model improvement are not easy. It could be that the reservoir models currently used are not perfectly suitable for the catchment under consideration, for example. It could be something else. Any discussion on this issue would be rather speculative at this stage. Some rewording of this part is done to avoid misunderstanding.

14. It is true that some processes are not modeled, but what shows that this is the reason of the poor results?

This is relevant to the discussion point raised above. What means 'the poor result' in our circumstances? Stability of estimates? Is it so vital to achieve this stability? In our view the results are promising: we get better discharge predictions at the gauged locations and, in average, in ungauged locations.

14. In addition, the authors also suspect the routing scheme (line 3). I am surprised by that, as figure 7 and page 14, lines 29 to 30 indicate that the routing velocity is the best determined parameter

This may be true, we agree with the reviewer.

All minor remarks are taken into account. Overall we agree that the presentation have to be improved and some additional materials have to be added in terms of results and their interpretation.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-331, 2019.