Review for manuscript HESS-2019-331

"On the potential of variational calibration for a fully distributed model: application on a Mediterranean catchment"

This study describes a new method based on variational data assimilation to calibrate a distributed hydrological model used operationally for flash flood forecasting in France. The method is applied on a French mountainous basin monitored by four stream gauge stations and rainfall radars. The authors first present the basics of the GRD distributed model and the new development that enables the differentiability of the model. Then the variational calibration method, the core of the study, is developed, as well as the methodology to evaluate it. The last two sections present the results and the discussion. Overall, the calibration strategy proposed by the authors is quite new and relevant for high dimensional model calibration, making this paper suitable for publication in HESS. Nevertheless, I have a few major comments I think the authors should address before publication. I would highly encourage the authors to further develop the evaluation of the methodology; the paper would then largely gain in visibility.

Major comments:

1. The methodology is quite well described but lacks for deeper evaluation. The results section is only one page and the discussion is reduced to less than 20 lines. I am sure that the study could benefit a lot from a more comprehensive analysis of this new methodology. For instance, some hydrographs could be included to show how the model behaves at the different stations depending on whether they are used or not in the calibration process. Also, do some stations always show good/bad performances? It lacks some time series to get an idea of interannual and spatial variability, flood severity (Qmax/Qmin).

2. Mathematical notations should be carefully revised as they are not always defined and sometimes not consistent (see following comments). This is very confusing and does not help the reader to fully understand how the method works.

Minor comments

P2L20. Do the authors know if there is any effort in the community to relate parameters of lumped or conceptual models to physical characteristics (e.g. average slope of the basin, concentration time...)? This could have important implication for the extension to ungauged basins.

P4L25. Is there any name or reference for the radar precipitation estimates provided by Météo-France?

P6. It appears to me that q (from transfer function, Eq. (8)) represents the runoff, while Q (from the routing model) represents the discharge in rivers. It is not clear in the text (L7 and L19).

P6L12. “Assuming Pn is the impulse function”: isn’t it Pr?

P6L22. Tau_i represents the runoff (or discharge? See previous comment) delay from node i-1 to node i. In a drainage network, a particular node i may have multiple direct upstream nodes i-1. Does Eq. (9) mean that all nodes i-1 flowing into node i share the same velocity and distance to node i? Rather, I would have defined v_i and d_i as the velocity in node i and the distance between node i and node i+1. Please correct me if I am wrong.

P6L22. N is not defined. Also, I would have removed “i=1,...,N” from the formula. Idem for Eq. (10)

P6L27. Variable q_i in Eq. (10) is not defined. I guess it is q at node i.

P7L15. K is not defined.

P7L20. What does “without the initial shock” mean?

P7L26. Identifiability is not only due to scarcity of observations in space. It may also comes from the model structure (concurrent parameters) and from processes (concurrent variables).

P8Fig1. I would suggest that variables that appear in the figure should be defined in the figure caption, even if they are defined in the text.

P9L1. Ns is not defined. Does it represent the number of nodes within the entire domain? Is it equal to N that appears previously (but not defined as well)?

P9L2. It is not clear from the formulation of Eq. (15) if Qk(t) only depends on P(xk,t),E(xk,t) or P and E over the entire domain. Actually, Qk(t), which represents the discharge at node xk and time t, should depend on P and E over the upstream sub-basin (which could be extended to the entire domain with a zero impact for nodes not inside the upstream sub-basin) and over a past period with a certain depth (which depends on the flow velocity of all the upstream cells, and which could be extended to the entire past period). Given that, I would have removed “k=1,Ns” from the equation and I would have changed t in the right hand side into t’ with t’ within the interval (0,t). Besides, in all the equations of section 2.1, i represents the node (discrete space) and k the time step (discrete time), which is not consistent with Eq. (15).

P9L3. h(x) should be h(x,0). Moreover, h(x,0) only represents the states of production and transfer reservoirs at node x, not over the entire domain. Same for p(x).

P10L1. The authors may explain that the scaling of p is done to get parameters within a [0,1] range.

P10L6. May p_tilde be equal to 0 or 1?

P10L7-11. Could the authors explain the implications of such assumptions?

P11L6. “Study area”

P11L27. “Investigating methodology”

P13L4. Letter “v” is missing in “validation”. Please check this out throughout the section.

P14L9. Please show it in a new figure (e.g. time series of P averaged over the entire domain).

P14L11-12. Is this shown in Fig. 4?

P14L25-26. Could this be shown (see also previous comment)?

P16Table2. Is Table 2 cited in the text?

P16L1-2. Physically based model can also suffer from structural deficiencies.

P16L5. What justifies the given bound values? Would the results be better without these bounds? In my opinion, it is better not to constrain parameter values except in case of numerical constraints. This is especially true for a conceptual model for which parameter values have no interpretable physical meaning.

P17Table3. Is Table 3 cited in the text?

P17L1-2. Typically, this is a result that could be shown in a figure.

P17L5. In addition to the cumulative surface, the distributed parameters could be compared to land cover maps or soil texture maps. If any relationship appears, such data could be used to better constrain the calibration process and then limit equifinality issues.

P17L14-17. Please reformulate.

P23Figure6. Parameter values are very different whether P1 or P2 is considered as the calibration period. What are the implications on the stability of the calibration (similar performances with P1 or P2)?