Hydrol. Earth Syst. Sci. Discuss., 10, C8082–C8091, 2014
www.hydrol-earth-syst-sci-discuss.net/10/C8082/2014/
© Author(s) 2014. This work is distributed under
the Creative Commons Attribute 3.0 License.

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration" *by* S. Gharari et al.

S. Gharari et al.

shervangharari@yahoo.com

Received and published: 3 March 2014

The article presents an approach to constrain the identification of hydrological model parameters, using physical reasoning and expert knowledge on the catchment behaviour. Using models of increasing complexity, the authors provide a method to constrain internal model variables and/or model parameter values. The approach is illustrated on one catchment in Luxemburg.
The article is clear and well-written. The approach proposed by the authors makes sense and may be a way to overcome the problem of model overparameterization generally faced by complex models. Therefore, it is potentially useful

for modellers and should be interest to a broad audience.
However, I found that the article fails to demonstrate the actual value of the proposed approach. Indeed, the approach is tested on a single catchment, with a single calibration/ validation test and on a very short series. Therefore, nothing shows whether the results are general and whether they could be transposed to other conditions. Hence my main recommendation to the authors is to provide a more substantial evaluation of their approach, so one can be sure that their results were not obtained by chance. This means introducing more basins, using longer time series and making more systematic calibration/validation tests. Several aspects of the testing methodology could also be improved.
I also found that the article is somewhat redundant with the other article simultaneously submitted by the authors to HESS presenting the approach. It would probably make sense to merge the two articles or to publish them as companion papers. Below I give a number of suggestions that could be considered by the authors when revising their manuscript. I advise publication after major revision.
We would like to thank the second anonymous reviewer for his/her constructive comments. These comments made it clear for us that we should explain our finding sharper and clearer in the discussion paper.
In general we would like to emphasize few points regarding this discussion paper:
1- We did not look at parameter values or parameter uncertainty. This simply mean the hydrological behaviour might not be related to parameter values if the hydrological model is constraint "properly".
2- This paper is just a simple test on three models with different complexity and number of constraints. We expanded this modelling framework to a large scale basin in China where we showed the power of the Flex-Topo model framework for prediction in validation periods and nested sub-catchments (Gao et al., 2013, currently in HESSD). In addition a further aspect of Flex-Topo, i.e. increased model consistency, analysed with a long-term time series in a small experimental catchment in Fracne is discussed in an additional paper currently in review in WRR (Hrachowitz et al.).

3- We did not want to move the technical note into the main paper as it might distract the reader from the more important message we would like convey. In our point of view the main contribution is formulating and constructing the model structure and constraints. The search algorithm is just a tool to solve this hydrological problem. The efficiency of this search method can be improved. However in our point of view hydrology is not about this search method, it is about how we formulate and use our hydrological knowledge into the modelling practice. For this reason we try to keep the paper separate.
We address the comments mentioned by the reviewer one by one as follow:

Detailed comments

1. Introduction: Maybe "complexity" as it is understood in the paper should be shortly defined in the introduction, to clarify its meaning here.

We agree with reviewer. As far as we know "complexity" itself and measures to quantify it are not well understood. In our study we cannot say for sure which model has higher complexity, "poorly constrained FLEX^A" or "heavily constrained FLEX^C". We tried to make our perception of complexity clear in the revised version by adding a short sentence in the beginning of the introduction.

2. P. 14803, L. 2-4: Lumped conceptual models may not be one endpoint of the modelling spectrum. Black-box models like artificial neural networks are probably further at that endpoint of the spectrum.

We agree with the reviewer. There are more models than stated in the discussion paper, such as artificial neural network and empirical models such as CIA method. In our study we are focusing on models which somehow represent different hydrological processes in a catchment. In our point of view it is rather difficult or even impossible to transfer our knowledge of hydrological processes to a black-box model. We shall make it clear in the revised version.

3. Section 2: The approach should be tested on a larger number of catchments, showing contrasted characteristics (especially various proportions of wetland, hillslope and plateau and various climate conditions). This would help demonstrating whether the approach can be also successful outside the test basin used

here. Does it successfully adapt to other physical and hydroclimatic conditions?

We agree with reviewer, our study is just the first step of exploring a bigger picture. In principle we should be able to apply this modelling framework to different catchments. However for every catchment the dominant runoff processes, topographical HRUs and all the available information should be re-assessed. In simpler words, if we move to another catchment, gauged or ungauged, we should think of ways to gather reasonable information and how to make the best use of it. This will be the topic of our future studies.

We also should mention that we have already tested this methodology for a large scale basin in a cold and arid region of China. The model was able to capture the observed hydrograph at internal nested catchments better than the calibrated lumped or distributed model (Gao et al., 2013). In addition a further study which is currently in review in WRR demonstrated the value of a Flex-Topo approach to increase model consistency for a long term data series in a small experimental catchment in France (Hrachowitz et al.2014, in review).

4. P. 14817, L. 20-22: I did not understand this sentence: when calibrating the model, the model must still be run to evaluate model efficiency.

We will try clarify this point. We introduced two type of constraints: parameter constraints and process constraints. Parameter constraints can be checked a priori without running the model. Process constraints should be evaluated after the model run, because the reservoirs and fluxes time series are needed to be evaluated. In other words, before the model is run for a parameter set, at least all the parameter sets should be satisfied.

5. P. 14819, L. 1-5: This paragraph is unclear.

We agree that the paragraph is not clear enough. We compare the transpiration from the hillslope and plateaus. The same comparison can be done between wetland and plateau or hillslope. However based on a pervious study (Gharari et al., 2011) the minimum DEM resolution to capture wetland in this catchment should be at least 20 meters while the LANDSAT 7 images have the resolution of 30 meters only.

6. P. 14819: The use of the runoff coefficient is a bit strange, since it is based on flows. It directly provides information about water balance. Hence it is not a typical physical characteristic available on ungauged basins. The use of simple water balance models may introduce strong errors in water balance. Their potential impact on model results is not analysed. Why runoff coefficient was used? What would be the results without this characteristic?

We agree with the reviewer that the water balance is hidden in the runoff ratio but this information is generally available based on the regional Budyko curve. Even in the absence of runoff measurement one can say something about approximate yearly runoff ratio just by analysing typically available precipitation and potential evaporation estimates with in the Budyko framework.

The second question that the reviewer asked is highly interesting, but not in the scope of our current study. As we explained briefly, it would be interesting to look at the effect of each constraints on the final set of behavioural parameter sets, the ones which satisfy all the constrains, an analysis which is provided in an additional paper currently in review (Hrachowitz et al., see replies above). This is from significant importance and will help us understand which constraints contains the bigger portion of information.

7. P. 14820: Equations could be presented in a more general way based on mean and standard deviation, instead of giving the values for the case study.

We agree that giving more general form might be brief but we feel giving the values can be easier to understand rather than reporting mean and standard deviation using a more general equation.

8. P. 14822, L. 14-16: It is unclear why the authors chose to use such a short period for model testing, which potentially increases the dependency of modelling results to the characteristics of the period. Longer time series should be used. If the authors wish to test the robustness of their approach when data availability is decreasing (which is an interesting question), they should do the corresponding tests and show results.

We agree with the reviewer to some extent. We are not comparing the performance

of one model in calibration and validation period. We are comparing the difference between three models. All these three models are subject to the choice of the length of the calibration and validation time series. So the time length of calibration and validation is similar for the three models.

In calibration, usually, the intention is to increase the calibration length period to include every possible events. However our intention is also not really to cover the entire possible events. It is rather interesting for us to evaluate the model before the so called "parameter stabilization" stage. We would like to see in what way the given constraints help to predict events which are not even in calibration period. We showed in an additional paper that by introducing constrains we can capture events which are different from calibration period while an unconstrained models fails to do so (Gao et al., 2013).

9. Section 3.4: The testing methodology could be improved. First, instead of only doing a single calibration/validation, the authors should do cross validation on the two periods. This would give a more complete assessment and provide a way to analyse the consistency of parameter sets identified on two different time periods (potentially with differing characteristics and therefore different impacts of constraints on parameter selection). Second, the authors could also introduce as a benchmark the models calibrated without any constraint (neither on parameters nor on states). The comparison with the author cases would be interesting to discuss.

For "constrained but un-calibrated" parameter sets, calibration and validation or cross calibration-validation are meaningless as the parameter sets are not trained based on the observed hydrograph.

We agree with the reviewer that looking at the parameter can give another layer of analysis, however in our study we did not look at parameter sets and parameter values as this is in our point of view beyond the scope of the study.

About the benchmark model, yes, we already did this test and the complex models, FLEX^C and FLEX^B, start to produce some strange result which seems not to be realistic at all. In our point of view the paper is already too long and bringing in this "calibrated but unconstrained" will open a new front for further discussion. However in our recent work (Gao et al., 2013; Hrachowitz et al. 2014, in review) we did such a test.

10. P. 14822, L. 10-13: Why KGE measures were not used instead of NSE efficiencies, which are known to be biased (see the work of Kling and Gupta)?

We are fully aware of the benefit of KGE over NSE. Both efficiencies are measuring the performance for the modelled hydrograph compared to the measure hydrograph. Our goal is just to show the improvement in performance; in our point of view both KGE and NSE will give the same conclusion. In this regard we prefer to use NSE as a well-known efficiency measure in order not to distract reader by the newly formulated, though indeed less biased, KGE.

11. P. 14822, L. 26: It cannot be said that flow observation was not use, since it is indirectly used in the runoff coefficient.

We agree with the reviewer, the flow observation is used indirectly. However this observation does not need temporal high resolution measurement of observed flow which is often needed for calibration. It can be regional estimation of annual runoff measures and does not need to be necessarily from the catchment of interest. It can even be approximated from precipitation and potential evaporation data only by using the Budyko framework. We shall make it clear in the revised version. We also explain this point answering comment No. 6.

12. P. 14823, L. 4-5: The number of constraints accounted for in each model could be indicated.

We agree with the reviewer, we will make this point clear in revised version.

13. P. 14823, L. 10-11: Uncertainty was estimated based on the width of the confidence interval. This is not a sufficient measure of the quality of the uncertainty estimates. A confidence interval may be sharp but not reliable. Hence the authors should use a more comprehensive evaluation measure that accounts for the reliability of uncertainty estimates.

We may not fully agree with the reviewer on this point. What we are showing in the discussion paper is hydrograph uncertainty for behavioural parameter sets. The pa-

rameter uncertainty might be higher for FLEX^C compare to FLEX^A and in the discussion paper we do not analyse the parameter sets in detail. The message we would like to convey cautiously is that "actual parameter values and consequently parameter uncertainty become less important if the problem is well constrained".

14. P. 14824, L. 23 - P. 14825, L. 2: The reason for the difference between the two models is unclear. Is it due to the difference in complexity or the difference in the number of constraints? This is not clearly shown.

Simply both, we try to make it clear in revised version.

15. P. 14825, L. 5-6: The interval is narrower, which is probably the direct consequence of a larger number of constraints. However, is it still reliable? Sharper confidence intervals are generally less reliable and therefore of more limited use to qualify the actual model predictive uncertainty.

This is a good point raised by the reviewer. The answer to this question might not be that simple. Reliability of "constrained but un-calibrated" models are directly linked to the reliability of each imposed constraint. A more detailed analysis of this is, in our opinion, out of the scope our work for this paper. As explained briefly in the manuscript in this case study we treated the constraints in a deterministic fashion where behavioural parameter sets are ones which satisfies all the constraints. However the sensitivity of final set of parameter sets to each constraint is rather unknown in this stage. Knowing the effect of every constraints will help us understand the importance of information given by each of them. A preliminary analysis of the effect of different constraints is given by Hrachowitz et al. 2014, which is currently in review in WRR.

16. P. 14825, L. 16-17: Same as previous comment.

The same answer as comment 15.

17. Section 4: The consistency of parameters identified on two test periods should be analysed since it would also be an indication of the robustness of the proposed approach. It would also provide good arguments in favour of the application in the ungauged case in the context of regionalization.

We understand the point raised by the reviewer. We would like to emphasize that in our understanding hydrology is not about parameter values [calibration and parameter uncertainty] however it is rather about what we perceive about parameters' or fluxes' or reservoirs' relations. In simpler words, we are not concerned with the fact that maximum interception of hilllsope is exactly 2.5 mm or 1.5 mm. It is rather important for us that any value it has should be more than wetland maximum interception. If maximum interception for hillslope is 1.5 mm in wetland it can be 0-1.5 but the exact value is not of primary importance.

18. Section 4.3: I found that this section largely repeats what was already said in sections 4.1 and 4.2. Probably a more synthetic presentation and comparison of results could be sought in sections 4.1 to 4.3, to avoid repetitions.

We agree with the reviewer and we are aware of such a repetition. However we wanted to summarize the story for "constrained but un-calibrated " and "constrained but calibrated" and meanwhile explain the figure 5 in details.

19. P. 14827, L. 25-26: To which extent conclusions are weaker in this case? Why is it so?

We simply meant that the difference in performance of different model structure for low is less distinguishable as NS values would be less sensitive to peak matching. We shall remove this part to avoid ambiguity.

20. P. 14829, L. 4-7: Which respective impacts have these two aspects on model results?

Simply the improvement in performance for parameter sets which satisfy all the constraints (un-calibrated but constrained).

21. Table 2: In the equation of the slow reservoir, I did not understand why the Rp component does not appear, whereas it seems to act on the reservoir in Fig. 2.

We will correct this. In table 2, we should add both percolation and capillary rise for the water balance equation of the slow reservoir.

22. Figures 3 and 4: Differences between models are difficult to see. Therefore these figures are not very useful.

Figure 3, clearly shows a strong difference between the three models, and figure 4 doesn't show significant difference. Figure 3 is emphasizing the difference of "un-calibrated by constraint" parameter sets for three models and figure 4 is emphasizing the fact that calibrated models are rather similar however in our point of view figure 4 is important as we see over-estimation of runoff by FLEX^C, this simply shows that constraints or model structure might be biased given the observation. We shall make it clear in the revised version.

Once again we would like to thank the second anonymous reviewer for his/her constructive comments.

References:

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Testing the realism of a topography driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrol. Earth Syst. Sci. Discuss., 10, 12663-12716, doi:10.5194/hessd-10-12663-2013, 2013.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Freer, J., Savenije H.H.G., Gascuel-Odoux, C.: Process Consistency in Models: the Importance of System Signatures, Expert Knowledge and Process Complexity, in review, Water Resources Research, 2014.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 14801, 2013.