Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-691-RC1, 2017

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Incremental model breakdown to assess the multi-hypotheses problem" *by* Florian U. Jehn et al.

L.A. Melsen (Referee)

lieke.melsen@wur.nl

Received and published: 27 December 2017

Jehn et al. provide a case-study of incremental model-breakdown; starting off with a (benchmark) model including a high number of processes (and parameters), the model is compared to models where fewer processes are explicitly represented. Finally, based on this information, a simplified model is presented with a higher model performance than the benchmark model.

The manuscript is well written and well-structured, and the figures and tables are to the point. I liked the fluxogram. There are, however, some questions, especially about the rationale, that I think need to be addressed, and the results and discussion sections are limited. This can improve with a more in-depth analysis of the results, for which I

Printer-friendly version

Discussion paper

provide a (first) suggestion.

About the rationale:

In the introduction and the conclusion the 'incremental model-breakdown' is presented as an alternative next to step-wise model building and comparison of pre-defined structures.

1) My intuition would be to conduct a sensitivity analysis, and based on that determine which processes are relevant and which are not. What is the advantage of doing the incremental model-breakdown rather than a sensitivity analysis? (except that the parameter is completely removed from the model rather than fixed at some point).

2) Another alternative, besides the incremental model-breakdown, step-wise building, and pre-defined structures, is to replace formulations of certain processes with alternative formulations, for example the SUMMA framework which you cite (Clark et al, 2015ab). How does the incremental model breakdown compare to this approach?

3) What is the added value of the incremental model-breakdown compared to all the alternatives? p.2,l.28 states that only a minor quantity of the vast space of possible model structures is explored, but isn't this also true for the incremental model-breakdown as presented in the manuscript, since only a single 'complex' model was employed?

Main points:

The model was run with a daily time step for a catchment in the order of 3000 km2. As becomes clear later on (section 2.5), the response time of the catchment is less than a day. How do you expect this influences your results? Obviously, this temporal resolution is not sufficient to capture the dynamics of the catchment. (follow up on that; It is unclear to me why you had to move the time-series; the river-part could easily be implemented as a routing with a time delay rather than a storage-system, which is more common for rainfall-runoff models).

The discussion of equifinality in the manuscript seems incosistent. Generally, the risk

on equifinality is higher with more degrees of freedom (more parameters compared to the information in the available data for calibration). But on page 13, l.3 is written: '[incremental model-breakdown]..have a positive impact on model performance, given the increased number of behavioural runs'. So; more behavioural runs is positive? But also an implication of equifinality? On p.14, l.11 it states '[incremental model-breakdown].. is a good way to improve model performance and reduce equifinality'. Please clarify.

This relates to my next point: is it a fair comparison to take the mean of the behavioural runs? I have not figured it out myself completely yet, but I don't see why a particular model should be 'punished' for having more (or less) behavioral parameter sets, see e.g. p.16, l.3-7. Perhaps consider another metric to compare the models.

p.9, l.20; for every model, a LHS of 300.000 is taken, despite the number of parameters. So, for models with fewer parameters, each parameter is sampled more often. This could explain why the more frugal models (fewer parameters) have more behavioral runs. Do you think this is the case?

Please add a motivation why you chose these three objective functions. None of the objective functions focusses on high flows, but still peaks and high flows are continuously discussed in the results and discussion section (e.g. p.13,l.17), while low flows are not discussed at all.

Can you provide an order of magnitude for the drinking water abstraction? The process is included in the model because water is abstracted for 80,000 inhabitants (p.4,l.2) but turns out to be unimportant, possible because of low population (159 persons per km2, p.15,l.10). In other words: where did you base the min and max parameter boundaries for drinking water extraction on? (Table 3)

In general, please provide references or motivation how and why you defined these boundaries for your parameters (Table 3).

Printer-friendly version

Discussion paper

To continue on that, I would also like to suggest for further analysis; why not showing the distribution of the parameters for the different model formulations? I would be interested to see if any of the parameters is taking over the job of one of the parameters that has been left out. This would result in a shifted parameter distribution. If I may undisclosed refer to my own work; see figure 8 in https://doi.org/10.5194/hess-20-2207-2016 Then, it would be interesting to see which parameters compensate for which processes.

The manuscript lacks a discussion of how the calibration period relates to the validation period. More parameters could fit better in the calibration period but can be flawed in the validation, which is something that should be discussed in relation to model complexity and number of parameters (see Kirchners paper on being right for the right reasons).

Other points:

Please mention the model time step and the temporal resolution of the input data in the 'model input and validation data' section.

Please check the units of Eq.1. V0 is a volume (p.4,l.24) but has the units of a rate. What are the units of V and Q?

Calculating the mean for a NSE is tricky since the NSE is highly non-symmetrical (+1 to minus infinity). Consider using the median.

Figure 3, caption. I think the word 'uncertainty' in 'uncertainty of the behavioural model runs' is not in place here. All you look at is the spread in your behavioural runs, which is certainly different from uncertainty.

The same holds true for p. 17, l. 1, 'less uncertain'

p.2, l.21 comparability -> comparison

p.4, l.6 unnecessary brackets around CMF, 2017

p. 10, table 3; caption; ‘indented’, in the table: ‘intendent’ -> intended

p. 16, l.8-13 repetition of p. 15, l.26

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-691, 2017.

Printer-friendly version

Discussion paper