Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-691-RC2, 2018

Hydrology and Earth System Sciences
Discussions

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

F. Anctil (Referee)

francois.anctil@gci.ulaval.ca

Received and published: 5 January 2018

In this paper, submitted by Jehn et al., a breakdown approach is proposed in order to simplify a complex model into a structure with "improved model performance, less uncertainty and higher model efficiency" (line 17, page 1). The method is validated on a 3-year time series from a single gauging station in Germany.

General comment:

The main argument in favour of experimenting with the proposed incremental model breakdown is that it may lead to a better model than the more common stepwise bottom-up approaches, arguing that "there is a chance that they have missed an even better model performance by including further modifications" (line 28, page 2). Yet no

comparison with a stepwise model building is presented, providing no evidence that a breakdown approach is superior.

Major comments:

There is possibly some confusion on the size of the watershed, which drains only about 3 km2 according to line 14, page 3. It is more likely that the size be 2977 km2 and not 2.977 km2, in order to accommodate 108 meteorological stations and an altitudinal range from 150 to 950 m a.s.l. A map of the watershed would have allowed to clarify this issue. It is recommended to add one.

Lumped hydrological models often need shorter time series for calibration than distributed ones. But in the context of a research on the selection of structural components, I am surprised that only 6 years of data was selected for calibration and only 3 more for validation (line 3, page 9). This needs to be justified. Longer series offer the advantage of stabilizing the results in regards to climatological variability. Were there no data available after 1988? At least, the authors need to inform on the climatology of the calibration and validation datasets in regard to the general, say, 30-year climatology. For instance, models usually work much better in wet years than in dry years. Was it the reason for selecting observations from the 80's? The authors should also avoid vague statements like "climatic conditions during the calibration (1980-1985) and validation period (1986-1988) were rather similar" (line 31, page 3). Chances are that they are not so similar at least in terms of low flows, otherwise how can one interpret the raise in validation logNS values in Table 2, in comparison to their calibration counterpart.

The issue of shifting the simulated discharge one day into the future to improve overall performance (line 8, page 9), thus simulating Q(t+1) instead of Q(t), typically falls from some failure in the routing components of the model and sounds more like fudging than modelling. What is the operational consequence of that trick? The argument that rainfalls occur in the "later time of a day" is weak and needs to be substantiated. This

information should be included in Figures 1 and 4.

GLUE is a convenient tool to assess the level of the parameter uncertainty of a model and to identify a number of equifinal (behavioural) parameter sets. Its use here as a calibration tool needs to be better justified (line 13, page 9), for example in comparison to more operational calibration schemes. Here, models variants are essentially compared in Table 2 on the basis of their number of behavioural runs that surpass three thresholds advocated by Moraisi et al. (2007), while parameter uncertainty is not explored. In practice, this has two limitations. 1) No performance information is provided for models 2, 3, 5, 8, and 10, for which the suppression of a structural component turned out detrimental. The issue is that we are provided no information on how much detrimental this operation is, which is quite important to the manuscript since model 15 is essentially built around them. 2) A small gain in performance may lead to a large increase in the number of behavioural runs.

Information in Table 2 is not that informative because it reflects only the behavioural runs. For instance, we are told that model 13 should be dismissed even if its metrics are better than model 15, because of a much lower number of runs to compute metrics (line 4, page 16). It would be easier to address that by giving all the information (not just the mean and the standard deviation) for example in the form of a box plot. From an operational point of view, hydrologists are looking for the best possible model, and variant 13 may fit their needs better than variant 15.

Minor comments:

Are the authors aware of any other hydrological studies on the same site that could offer some basis of comparison?

Figure 2 is not much useful.

Figure 3 would be more intelligible if it would be split in two: a figure for model 1 and another one for model 15.

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