Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-176-RC2, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Behind the scenes of streamflow model performance" *by* Laurène J. E. Bouaziz et al.

Anonymous Referee #2

Received and published: 1 June 2020

Bouaziz et al. (2020) evaluate 12 hydrologic models for three medium-sized Belgian catchments, all established and calibrated by eight research groups. Although the spatially aggregated streamflow performance differences among models are negligible, the internal model states and processes (can) differ significantly. This paper is an interesting diagnostic study, with nice figures. I have some minor comments which the authors should consider to address.

First of all, it is nice to see the huge collaborative efforts across many institutes behind this model inter-comparison study. This study with many details shows large differences among 12 hydrologic models and even larger differences against different remotely sensed products. Something what a reader would expect. I encourage the authors to stress more clearly, what is the main "take-home" message of the main

paper. Because the authors did not use an ensemble of model structures from a modular framework (e.g., FLEX, FUSE), which could properly address those differences or individual model deficiencies step by step (by identifying individual hypothesis), in their study they cannot clearly separate and identify, which hidden hydrological processes can help improve model functioning against those reanalysis products. Could you please comment on this?

Further details in chronologic order:

Line 80,140+: evaporation => "evapotranspiration"? Please don't forget about the plants! Hargreaves-Samani formula is for evapotranspiration, not for evaporation only.

Line 85: streamflow => "runoff", because of the unit

Line 86: You should start this sentence that this is a headwater basin of ID1

Line 101: I guess the authors could have used a bit more advanced method for interpolation rain gauge observation instead of the Thiessen polygons, to better account for input error uncertainty, e.g. kriging or its variants. The uncertainty in the meteorological inputs is not mentioned in the manuscript.

Line 114: PET method is based on Priestley Taylor, which is different from section 3.1. How is it compatible with section 3.1 and overall results?

Line 143: how did you spatially average soil moisture?

Line 153: I guess, your entire study domain is just a single GRACE pixel. I am quite skeptical for using it at all, as it's beyond the limits of its usability. The original raw GRACE signal is based on a much larger region (3degrees). You may better wait for the GRACE-FO, which has much finer native resolution...

Section 4.1 I guess all models were applied in spatially lumped manner, i.e. no spatially distributed mode, isn't it? Please write down explicitly in this section.

Line: 179-181 This analysis was done here, or in previous study? Not clear, please specify, and provide link to the transferability results. Curious to see them.

How many behavioral parameter sets per model were used? Is the number same per hydrological model? Here referring to error bars in Fig. 3, Fig. 4 and elsewhere.

Figure 9: Is it possible to rank the models according to their performance? Which one seems to be most relevant and how it compares to e.g. an operational model, if that's available? Please think about putting some implications to the paper.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-176, 2020.