

Hydrol. Earth Syst. Sci. Discuss., author comment AC2
<https://doi.org/10.5194/hess-2021-147-AC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Pierre Nicolle et al.

Author comment on "Technical note: RAT – a robustness assessment test for calibrated and uncalibrated hydrological models" by Pierre Nicolle et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-147-AC2>, 2021

Thank you very much for your review. We provide general answers below and specific answers in the attached file.

We agree that the RAT is an imperfect evaluation test, and in this technical note we took great care to underline its limits, especially in conclusion. We also agree that because the RAT is data-driven, it does not allow to separate model deficiencies from data deficiencies on a case by case basis. This is why we consider that this test is a relative rather than an absolute test, allowing the comparison of several models or model variants: if there are errors in the streamflow or the precipitation data, they will affect all models or model variants similarly. Our view is that data quality issues should be addressed before modelling (even if we recognize that some issues can well be overseen... and appear only during the modelling phase)

We agree with you that "we do not want to reject a model that is not wrong", but like with any test, this risk (type I error) will exist. This is the risk of judging guilty an innocent person, which, in the case of a model is not so worrying: it is in the very nature of hydrological models to be guilty or imperfect. A test such as RAT may yield a false positive for a few catchments, but it will nonetheless offer a possibility for comparing modelling alternatives.

Also, on this issue of data errors, we would like to stress that this is not specific to the RAT: this problem affects all robustness tests, all the derivatives of the split sample test.

Concerning your suggestion to test the model with different input datasets in order to try to distinguish between metrological and model causes to a positive reaction to RAT, we think that global datasets will be unavoidably much more imprecise than Meteo-France's ground-based interpolated data, which will increase considerably the amplitude of bias (we have a long experience of using Meteo France's SAFRAN product as catchment forcing, and we consider it to be the best option in France). The alternative you suggest (introduce some noise in the meteorological input and assess its impact) is interesting, but the data quality issues that modellers are the most concerned with are not random errors, they are the trends or the sudden changes linked with modifications of the observation network (we could still add noises with a trend but it would be a

rather obvious result that these are detectable).

Concerning your question on the hydrological year, you are right, a perfect model does not need a specific season to compute bias. But in the case of snow-affected catchments, it seems always more careful to avoid separating the snow accumulation and the snow melt seasons. However we do not agree with you on the following point: *"The model output bias and climatic variables may not be dependent at daily scale but show dependency at coarser temporal scale, or vice versa."* We consider that because models are so imperfect, it is more reasonable to work on bias at coarse time steps first, and then only to move to fine time steps. But we agree that the absence of bias for any given year, may well hide a positive bias during half of the year and a negative year during the other half. This is why we underline in conclusion that *"the RAT should not be seen as the only basis for model choice."*