Reply on RC2
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-AC1, 2021

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

We have written this paper with the conviction that, because it is in the very nature of hydrological models to be imperfect, a test such as RAT will never offer a definitive answer and should mainly be used for comparing modelling alternatives. But we may have been imprecise sometimes, as you mentioned.

As far as the critic of key point 3 is concerned, you are right, there is a logical contradiction with key point 4. Indeed, the RAT is useful to identify models that are “unsafe” to use. Or even (because we believe that no model can be considered as perfectly safe), RAT can be used to compare models and identify “the less unsafe” one. We should perhaps reword this sentence as follows “the RAT method can be used to determine whether a hydrological model cannot be safely used for climate change impact studies”.

We believe it is very important to underline the limit of any method and this is what we tried to do in this note too... the difficulty being that we did not want to be only negative either. With the expression “climate proof” we did not mean to give an absolute statement (one may still get wet with weatherproof clothing under particularly extreme weather!), but we understand your point and will replace this expression.

Concerning your observations on section 4.2, we agree that it is very descriptive. We will include in the revised version a discussion on what can be done when different types of model failure happen (even if, as you mention, we may ultimately conclude that “it depends on the study”).

Concerning the suggestion of illustrating the possible impact of a change in flow rating curves, this is more complicated. A synthetic example would not be very interesting. We have in our dataset two stations that have been moved to a nearby location on the same river:

- 
  - Le Serein à Chablis (H2342020): the gaging station was moved in January
2001 (significant dependency to P and P/E);
- La Sauldre à Salsbris (K6402520): the gaging station was moved in February 2006 (significant dependency T)

As can be seen in the bias plots of the supplementary material, no abrupt change is visible at these dates... however, it is true that a climate dependency is detected. We are unable to conclude whether this is caused by the modification of the streamgaging section (no one can make statistics with a sample of two stations...), but it remains interesting to note that the two stations where this change occurred showed some dependency.

A last precision on our use of the term “post-processing”: if we identify a linear dependency between model bias and temperature for example, one could be tempted to fit a linear correction model in order to unbiase the results (this is what we call “post-processing”). And when writing that we do not recommend it, we wanted to stress that we should not focus on the symptoms but rather try to identify the causes of the bias.