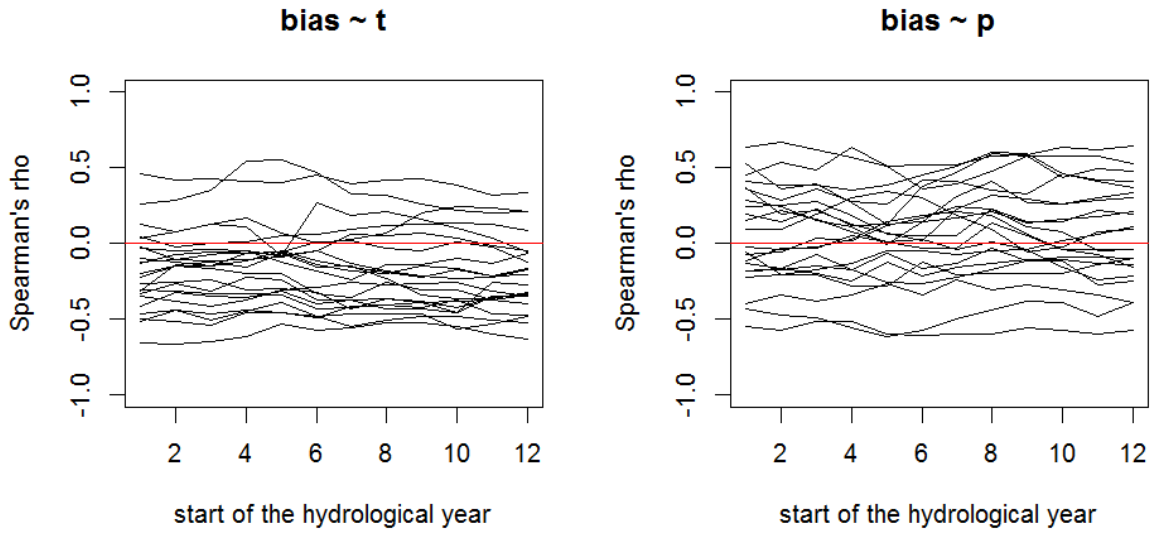
Dear Editor,

Thank you for taking care of our Technical Note, and for your comments. Please find below short answers to your questions.

This technical note proposes an additional method to check whether it is reasonable to use a model under climate change conditions, a method that does not rely on numerous re-calibration. I see the value of the idea presented here but in addition to the reviewers' comments, I would like to raise some additional points.

First of all, the choice of the hydrological year might in my view have a considerable impact on the results and I am surprised that it is not tested in this note (easy to do: check the robustness of the results for other closing dates). Why? The hydrological year is most probably chosen such as to have little carry over of water storage from one year to the next (something that might need to be specified in the paper); for soil moisture and snow, August (as chosen in the paper) is certainly a reasonable choice but what about catchments that show the driest month in September? Will the carry-over effect result in spurious model bias correlations to temperature?

We tested the twelve possible "hydrological years", and found that it only marginally affects the result of the test (see below). We will add this discussion in the revised manuscript.



This brings me to the next point: the authors use a simple model, which has the main advantage that it should be straight-forward to explain what model "problem" could cause a correlation between bias and air temperature (or precipitation). It would be nice to see how we could attempt to interpret the climate dependency of the model. What model parameterizations can in fact influence the annual bias in this model? What does it tell us about the model if the bias increases with temperature or with precipitation? Obviously, such an interpretation is not possible for complex models but I really see this as a missed opportunity to share the authors' expertise in the use of a conceptual model to give guidance on how to further interpret the results. I.e. I would like to see a further

elaboration of Section 4.2 (in addition to the comments raised by reviewer 2 on this section).

We do agree that it is very interesting to use a model (or several models…) on a wider catchment set to identify statistically the deficiencies of the model structure, which may explain their non-robustness. But if we introduce such a section in our note… we will need to describe the dataset, the model(s), interpret the results, propose a typology of "reacting" catchments… this will not be a technical note anymore, but rather a very very long paper. We would prefer to keep this note short and rather simple, and keep the large sample interpretation for a future paper. We do not like "Salami publications", but the fact is that there is already many details to be explained and understood.

This attempt for interpretation might in fact even unravel unexpected reasoning: is it not a good sign if e.g. a simple snow model gives stronger bias in snow rich years as opposed to snow poor years?

This is exact.

Furthermore: I do not understand how the GSST points in Figure 4 are obtained; how is the bias for two simulations over two different periods (validation, calibration) defined.

The black points in Figure 4 correspond the GSST as initially published by Coron et al. (2012): all possible 10-year periods are used to calibrate the model and for each calibration, each 10-year sliding period over the remaining available period, strictly independent of the calibration one, is used to evaluate the model. The red points correspond to the "RAT" approach with a single simulation (and in our case, calibration) and as many validation periods as there are years.

And there is a formulation error in line 223, which reads like: "We then identified the catchments where the RAT procedure detected a lack of dependency of streamflow bias to climate variables." I guess it should be "identified a dependency"

No, our sentence was right, but we agree that it was uselessly complicated. We propose to write : "We then identified the catchments where the RAT procedure detected a dependency of streamflow bias to one or several climate variables".

References

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. Water Resources Research, 48, W05552. https://doi.org/10.1029/2011WR011721