Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-517-RC1, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Uncertainty of gridded precipitation and temperature reference datasets in climate change impact studies" *by* Mostafa Tarek et al.

Anonymous Referee #1

Received and published: 9 December 2020

In sparsely gauged regions, satellite, reanalysis and merged products datasets are valuable for hydro-meteorological studies. However, it is not yet well understood how uncertainty cascades from the choice of a reference dataset down to future climate streamflows in hydroclimatic impact studies. To bridge this gap, the authors assess relative uncertainty contributions from 4 main contributors (element of the hydroclimatic modelling chain; GCMs, precipitation and temperature datasets and hydrological models) for future climate streamflow predictions in across 1145 African catchments. They show that GCMs and precipitation datasets are the main contributors of uncertainty, hinting at the importance of the choice of the modelling chain and reference datasets for hydroclimate impact studies.

This paper is overall well-written and aims to bridge an important gap in hydroclimate impact studies. The methodology used however currently suffers from a few major limitations which should be addressed by the authors (see the first three key comments below). Please find below a few comments which will hopefully help improve the paper for publication.

Key comments

- I wonder about the conservation of the important energy and water balance when combining P and T datasets from different sources - e.g. MSWEP precipitation and ERA5 temperature. Could you please comment on this and potential impacts in the paper?

- On P10 L280-281 you mention the impacts of the reduction in the ensemble size of the precipitation dataset on the variance analysis: "Unsurprisingly, it shows that reducing the size of the precipitation ensemble results in a consistent decrease in the variance attributed to precipitation". This leads me to question what the impacts may be of the ensemble sizes from the contributors you investigate on your conclusions: 10 GCMs, 2 hydrological models, 2 temperature and 9 precipitation datasets. Wouldn't it be more adequate to have the same number of ensemble members coming from these various components of the chain? For example, temperature appears to play a minor role in the analysis, however, only 2 members were used here, which could impact this conclusion somewhat. Please reflect on this in your paper.

- The calibration strategy may be problematic for this study given the climate timescales explored. As mentioned by Arsenault et al. (2018): "In this study, the effect of calibration and validation is investigated on three catchments that did not show signs of non-stationarity, i.e. the mean annual streamflow did not contain a trend over a 25-year period. This allowed randomly sampling from the database to generate calibration and validation sets. This raises the question as to how the method would fare on a catchment that is subject to non-stationarity. Obviously, in this scenario, the independent

test period would need to be in the most recent years and those years could not be randomly selected from the entire time series." Arsenault et al. go on to suggest alternative methods which could be used in such catchments. Could you please reflect on the adequacy of this calibration strategy for your study?

- You mention 51 different streamflow metrics, yet all results are shown for only 6 metrics. Could you please comment on the results from the additional metrics not shows here? These could perhaps go in as supplementary material?

- Your figures are very rich in results. Please guide the readers a bit more by mentioning what is shown in the columns/rows, etc. when introducing each figure (especially for Fig 5, 7 and 8).

Minor comments

- P1 L9-12: Please clarify here if these datasets are deterministic or ensembles.

- P1 L15: Please explain here what CMIP5 GCMs stands for.

- P3 L60: Please clarify here what GHGES stands for, it only comes later on L66.

- P3 L63: Could you please give readers a brief explanation of what the "change factor approach for downscaling" is here?

- P4 L91: Since you are looking at hydroclimate impacts, it might make sense to also refer to the Hydrological Climate Classification by Knoben et al. (2018), more adapted to hydrological studies: Knoben, W. J., Woods, R. A., & Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. Water Resources Research, 54(7), 5088-5109, https://doi.org/10.1029/2018WR022913.

- P4 L107: Please clarify here what NAC2H stands for.

- P4 L112: Please consider replacing "(or better)" with "or higher".

- P5 L122: While it is implicit, you do not actually explicitly mention that you have used

GRDC data in this paper.

- P5 L136: Could you please provide your reasoning for selecting the L5 vector layer instead of the other 3 shown on Fig. 1?

- P5 L137: I would find it helpful if you could briefly go over the steps of the entire hydroclimatic modelling chain from Fig. 2 in the text as well. For example, I only found out from Fig. 2 that two calibrations are performed.

- P5 L146: Please summarise what the "4 groups of components of the uncertainty modeling chain" are here for added clarity for the readers.

- P6 L150-152: Could you please summarise briefly in the text as well which metric(s) and time period were used for the calibration? I read the added information later on in Section 3.1.3, please mention here that more details are given in that later section.

- P6 L155-158: This is a repetition to an extent of P6 L147-149. Please consider merging these two paragraphs for more clarity. It is also unclear to me how the 1150 African stations (L155) became 1145 catchments?

- P6 L162-163: More importantly, have they been shown to perform well in Africa specifically?

- P7 L201-204: Is this single simulation used as a reference against which to verify the other simulations produced as part of the analysis? Please clarify as it confused me a little bit. When you say "Based on the hydrological modeling performance on the 350 gauged catchments", do you mean the calibration performance? Please clarify here.

- P8 L213: I would have liked to read a bit more about the variance analysis, about the methods and the aim of this analysis. E.g. What are the variance components and what do they tell you? Is such an analysis computationally expensive to run?

- P8 L214-216: Could you please provide an overview of the metrics computed, perhaps in a table?

- P8 L232-233: Please clarify that this observation is with regards to calibration for these 350 catchments. It could otherwise be misinterpreted taken out of context.

- P9 L256-256: The mean Summer Q also appears to show a different signal from the other metrics.

- P9 L260-261: This is arguable and quite complicated to see. Perhaps putting the metrics in columns and the contributors in rows might help see these better?

- P9 L268-269: Which variance contributor is Fig. 6 shown for?

- P10 L284-285: It seems to be that most of the drop is seen between ensembles 1 and 2, rather than 2 and 3. Please also add the ensemble numbers 1-4 from table 4 to Fig. 7.

- P10 L290-292: Hydrological model uncertainty appears dominant over precipitation uncertainty for low flows.

- Figure 2: In order to be clearer for the readers, please consider adding clear sub-heading for each box in this diagram.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-517, 2020.