Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/nhess-2021-150-RC1, 2021

Comment on nhess-2021-150

Anonymous Referee #1

Referee comment on "Adaptation and application of the large LAERTES-EU regional climate model ensemble for modeling hydrological extremes: a pilot study for the Rhine basin" by Florian Ehmele et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-150-RC1, 2021

This is an interesting paper that describes the use of a large ensemble of regionally downscaled multi-GCM forcings to drive a hydrological model for impact assessments. The issue of long return period extremes is highly relevant. The paper is very well written, clearly structured and to the point. However, there are some unfortunate shortcuts regarding the model validation which needs to be handled differently.

Main comments:
Both the bias correction and the HBV set ups are validated on the calibration period. While I can accept this for the bias adjustment because it is not anywhere applied outside of the calibration period, it is a big issue for the justification of the hydrological model. HBV is currently calibrated and validated on the same period (1961-2006) based on precipitation and temperatur forcing from gridded observational data sets. When validated on that same period, the results are very good, as seen from the very high NSE values. However, we still know nothing about the model's performance on data it has never seen before, and the main results are based on the downscaled model data. I urge the authors to at least perform a split sample validation where calibration and validation periods are independent, or even a cross-validation. This is standard practice in hydrological model validation.

Bias correction is only performed for precipitation, and no information about potential bias in temperature and how it might affect results is provided. Because temperature, and its translation into evapotranspiration, is an important input to the water balance of the model, it should not be neglected. I would like to at least see a justification for why temperature is not bias corrected (being that the bias is low). In some cases it can be neglected for certain extremes where the pre-conditioning of the river is of minor importance, but also that needs some additional analysis and commenting in the text.

The concluding main result of the paper is presented in figure 7. Although the result is

compelling and seemingly clear, the details may occlude the actual results. First, the lenght of each timeseries has a large effect on the GEV fits and their robustness, as argued in the introduction. Please add the record lenght, i.e. the number of years, in the legend for each data set. Second, it would help the reader a lot to also see the confidence intervals. With so many lines, it might get too busy, but I think adding e.g. the confidence interval for the "Q obs - Weibul" and "LAERTES-EU BC" would be very informative. The confidence intervals would convey two results, one is the fair comparison of the observations and the model that would show the observations results essentially useless beyond 50 years (depending on the lenght of the timeseries), and the other is the added value of the multi-realization simulations which add statistical robustness for the longer return periods.

Minor comments:
L69: Please clarify what you mean with "isolate the effects".

L85-94: Please describe more details about the LEARTES-EU multi-model. It is currently not clear what the driving GCMs are; especially that they area mixture of assimilated reanalysis, decadal initialized forecasts and free GCM simulations. Please repeat more from Ehmele et al. (2020) which provides a good summary, enough for the reader to understand from this paper alone.

L176, 185: I would avoid describing differences between data sets as "bias", but rather use the word "difference" unless you include a well established ground truth observational reference.

Figure 3: Please consider using a log-log scale, which would better show differences between the data sets for the (0,100) mm/day range.

L307: "different forcing and/or assimilation schemes". I refer back to my earlier comment that the LAERTES-EU sources needs to be better described.

L310: "consistend data for precipitation and temperature". This is not really true after bias correction. The depencency between the variables can be severely impacted. You have also not described the potential temperature bias and how it migh affect rain/snow distribution and timing over the year. It might not be useful to retain the dependence it is errouneous?

Figure S7-11: Please change "Observed - Weibul" to "Q obs. - Weibul" as in the main text figure.