Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-24-RC2, 2019

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Does the weighting of climate simulations result in a more reasonable quantification of hydrological impacts?" *by* Hui-Min Wang et al.

Anonymous Referee #2

Received and published: 20 March 2019

Summary and General Comments

The manuscript by Wang et al investigates the impact of multiple ensemble weighting techniques on the simulations of hydrological impacts for two different river basins. The authors compare the results from a hydrological impact model driven by weighted and unweighted GCM projections. In addition, the authors compare the results from bias-correcting the GCM output before weighting or not. They conclude that weighting the bias corrected GCM output has not a large effect while differences are larger when using raw output, improving the representation of the mean hydrograph and reducing the annual streamflow bias. The authors conclude that the equal weighting method is

a conservative approach and still viable given the small effect weighting has on a bias corrected ensemble.

Overall the paper is well and comprehensively written and the analysis extensive. The fact that weighting the bias corrected ensemble has a very small effect is not surprising. Given that bias correcting and weighting for performance have the same goal, bringing the ensemble closer to observed values, I do not understand why you would do both? Some of the risks and disadvantages for weighting, which are all true, also apply for bias correction (e.g. Sippel et al. 2016, Maraun et al. 2017). Both tools need to be applied carefully and have their pitfalls. For instance, it has been shown that out-of-sample testing is crucial for any kind of weighting or sub-sampling (e.g. Herger et al. 2018, Abramowitz et al. 2019), which is still missing so far in this study. In that sense I am not convinced that the authors come to the correct conclusion, even though their arguments are generally not wrong (see below). Weighting a GCM ensemble will conserve dependencies between different variables in a physically consistent way, and in cases where this is important, it might be preferred over bias correction. However, all the risks the authors mention apply, and the study shows nicely that there are still many open questions on how to use these methods properly. I would recommend to rephrase some of the discussion and conclusions more carefully and also account for the assumptions and risks in bias correction.

Specific comments

P1, L23-25: This conclusion is a bit far fetched and ignores the independence issue nicely described on page 2, around line 20. P4, L8: not relevant. P6, L20: The climatological mean of what? Temperature, precipitation, streamflow? All of them together or only individual? That makes a large difference and it has been shown that only using one at a time for PI is risky (Lorenz et al. 2018). P9, L19-23: Yes, but the same assumption applies for bias-correction. P10, L2-5: The testing is all done in sample. Out-of-sample testing is needed. P10, L9-11: At least for PI any metric could be considered, the fact that only climatology was used is because the authors chose

to do it this way, but is not a property of the method. P11-P12, L31-4: While these arguments are true, bias-correction has similar problems. Also, it looks to me that the equally weighted ensemble has the same issue? P12, L27-28: I do not think the results fully support this statement. We might not have found a clearly better way than model democracy, but equal weighting is as at least as arbitrary as weighting. P13, L1-2: Equally weighting is also arbitrary, given that is assumes all models are equally likely and independent, which they are not. P13, L5-10: Again, the same applies to bias-correction. P13, L11: Again, because you chose to only include one metric does not make it a property of the method. At least some of the weighting methods can account for multiple metrics to be included and people argue to do so (e.g. Knutti et al. 2017).

References: Abramowitz, G., et al: ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing, Earth Syst. Dynam., 10, 91-105, https://doi.org/10.5194/esd-10-91-2019, 2019.

Herger, N., et al: Selecting a climate model subset to optimise key ensemble properties, Earth Syst. Dynam., 9, 135–151, https://doi.org/10.5194/esd-9-135-2018, 2018.

Knutti, R., et al.: A climate model projection weighting scheme accounting for performance and interdependence, Geophys. Res. Lett., 44, 1909–1918, https://doi.org/10.1002/2016GL072012, 2017.

Lorenz, R., et al.: Prospects and caveats of weighting climate models for summer maximum temperature projections over North America, J. Geophys. Res.-Atmos., 123, 4509–4526, https://doi.org/10.1029/2017JD027992, 2018. Maraun D. et al: Towards process-informed bias correction of climate change simulations, Nature Climate Change, 7, 764–773, https://www.nature.com/articles/nclimate3418, 2017.

Sippel et al.: A novel bias correction methodology for climate impact Simulations, Earth Syst. Dynam., 7, doi:10.5194/esd-7-71-2016, 71–88, 2016.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-

24, 2019.