

Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts

by Louise Crochemore, Maria-Helena Ramos, Florian Pappenberger

General comments

This is a revised version of the manuscript that I already reviewed in the first round (I therefore refrain from writing a summary). The authors have carefully revised most reviewer comments and did so in a very transparent way which is easy to follow for a reviewer. Thanks for that. I went through the revised manuscript and mainly checked if the various reviewer comments were addressed or not. Only in a very few minor issues, I made new comments.

Please find my comments below. I used the same comment labels as they were written in the reply. If anything is not listed below, this means that I accepted the authors reply. From my viewpoint, the main issue that needs a bit further discussion is the MAE results of GDM-y and EDM-y for JJA and its discrepancy with respect to the bias-results in Fig. 6. This is mainly due to my first comment not being concise enough. My apologies for that. I'm happy to discuss this in the open discussion and could reply to a posted comment if the authors posted a reply to that specific comment before July 15, as I'm out-of-office after that. Otherwise, I leave it to the editor to decide whether or not the reply is satisfactory.

Overall, I think the manuscript has much improved and is much better readable than in the first version. I suggest that the article is ready for publication after a few minor comments have been taken into account.

Detailed comments

New comments to the revised version:

Page 6, line 8-10: The sentence "A scaling factor higher (lower) than 1 indicates that the mean ensemble forecast underpredicts (overpredicts) the mean observed value. A value of 1 indicates no bias in the forecasts." could be left out since the analysis of the scaling factors has been removed from the manuscript. Instead, a reference for the method could be given (for e.g. Lenderink et al. (2007), or any other paper that gives a more detailed explanation of the approach).

Page 7, line 19: "equally reliable" instead of "reliable"

Page 24, lines 6-7: This sentence about the challenges for the bias-correction methods would seem to fit better into a later paragraph, since the current paragraph is describing the raw forecasts. Consider moving this sentence to a different paragraph.

Reviewer comments that were not addressed entirely satisfactorily:

Reviewer 1, comment 1: I think the rephrased introduction section about novelty and the main goals of the study is fine. However, the statement about "... no previous study has compared bias correction methods and their impact on streamflow forecasting in a systematic way, with a focus on understanding how the main attributes of forecast performance are impacted by bias correction." does not hold. First, it is contradictory to the statement on line page 3, line 7 where you say that such studies are rare – which implies that there are at least some studies -, and second, after a short search on the internet, I found the paper by Hashino et al. (2007). It seems that seasonal forecasts in combination with bias-correction methods has been studied before. This said, I do not think this impedes on the quality of the paper. It is just the statement above that needs to be removed or adjusted.

Reviewer 2, Page 4, line 23: I understand now what the authors mean by interannual potential evapotranspiration, but I still think that the term “interannual” is misleading. It is not the potential evapotranspiration between the years, but a multi-year climatological mean year. I suggest rephrasing the term, for e.g. to something alike “climatological potential evapotranspiration”.

Reviewer 2, Page 9, line 12: In the revised version, you refer to the lines corresponding to the 5% significance test. In your reply, you mention though that the value of 0.1 for the bands is not exactly correct for the 5% significance level. I’m a bit confused by the different information given. I do not doubt that the value of 0.1 is a good choice, but if you show the lines at 0.1, you should not write that they represent the 5% significance level. Please correct the manuscript accordingly.

Reviewer 2, Page 11, line 8-9: Thanks a lot for the explanations. I indeed misinterpreted the figures in the first place.

Reviewer 2, Section 5.3 and figures 8 and 9: I agree with the authors that even in case of strong biases in GDM-y and EDM-y, the MAE might still, in principle, show some skill improvement with respect to uncorrected forecasts. For me, this would be logical in case the strong biases after bias-correction are still smaller than before bias-correction.

Related to the last condition stated above, it appears now to me that I have formulated my comment unclearly. I should have written that I’m particularly surprised to see such a clear gain in MAE skill for summer since the bias after bias-correction is larger than before bias-correction. In Fig. 6 and for the months JJA, the bias skill after bias-correction shows a clear underprediction whereas for the data before bias-correction, the bias-skill is very close to 1 or slightly on the overprediction side for the catchment 4. It is this before bias-correction vs after bias-correction discrepancy as well as the very clear skill gain in figure 7 for JJA that make me doubting the results for GDM-y and EDM-y in figure 7. Or to look at the same thing from a different perspective, I would expect the methods with low bias in Fig. 6 to also show better skill in MAE than methods with high bias in Fig. 6. (as long as no strong non-linearities are to be expected, which I doubt here). How could one for example explain that EDMD-m that has a bias of very close to 1 in Fig. 6 for all the months in JJA only leads to skill gain in MAE in about 60% of the catchments, but EDM-y which has a clear underprediction in Fig. 6 and is clearly worse than EDMD-m still leads to an extremely high gain increase?

Could the authors please comment a bit more on this issue? I’m happy to follow this on the interactive discussion. However, I’m out-of-office from July 15th until August 15th.

Reviewer 2, figure 6: I appreciate that the authors took into account my comment. It seems to me that in the revised manuscript on page 11, line 13 and following, the text explaining the bias values shown in figure 6 has not been changed accordingly. Also, if you are plotting the relative bias in figure 6 now, why is the interval from 1 to 4 so much larger than the interval from 0.25 to 1? Is there still some transformation involved? If there is no reason for the unequal interval size, I suggest to use a linear colorscale for easier interpretation (i.e. equal distance on the colorscale means equal factor difference to the no-bias point).

Reviewer 3, RC: page 3, line 23: I agree with the reviewer that the term interannual PET is confusing, and the suggestion “long-term mean” sounds fine to me. The authors should change the wording since two of three reviewers observed the same issue.

RC: 3) I noticed that in the result section, the authors still refer to points when discussing the PIT diagram (for e.g. on page 9, line 8). For consistency reason, this should be changed according to the changes in section 3.

References

Hashino, T., Bradley, a. a., & Schwartz, S. S. (2007). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences*, 11, 939–950. doi:10.5194/hess-11-939-2007

Lenderink, G., Buishand, a., & van Deursen, W. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, *11*(3), 1145–1159. doi:10.5194/hess-11-1145-2007