Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-78-RC3, 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts" *by* Louise Crochemore et al.

Anonymous Referee #2

Received and published: 26 March 2016

Summary

The study analyses the skill of ECMWF's System 4 seasonal forecasting system for precipitation and streamflow forecasts in 16 French catchments, using the hydrological model GR6J for transferring the meteorological forecasts into streamflow forecasts. In particular, the study focusses on the effect of bias-correcting precipitation in different ways. Main conclusions are that linear scaling and EDMD bias-correction with monthly calibration windows perform better than other methods. In general, bias-correction was found to improve the skills, but the result varies for the different skill scores. Often, a trade-off between decreasing sharpeness and increasing reliability was found when

applying bias-correction. When comparing the differences of the bias-correction effect on precipitation and streamflow, it was found that in cases when sharpeness and overall performance increased by bias-correction, it increased more strongly for streamflow than for precipitation. The opposite was found to be true for reliability.

General comments

The study addresses a relevant scientific question and the methods used are sound. A few results need some further clarification and/or discussion and I have listed the open issues in the detailed comments. Additionally to the detailed comments, I would like to add two general comments.

- The study uses mainly modelled streamflow as a reference. Nevertheless, I miss some indication of the hydrological model performance in the 16 basins. This is particularly relevant since also the observed streamflow is used as a reference forecast in one part of the manuscript, and this analysis would critically depend on systematic biases of the hydrological model.
- The manuscript covers a large body of results and is therefore lengthy. I think that it could be streamlined without losing too much information.

Over all, I suggest acceptance of the manuscript after my comments have been taken into account. I'm looking forward to the revised manuscript.

Detailed comments

Page 3, line 1: Some reference needed to support the statement that linear scaling and distribution mapping are widely used methods in seasonal forecasting.

Page 4, line 13: Which parametrization was used to derive potential evapotranspiration?

Page 4, line 23: What is meant by interannual potential evapotranspiration? I would have understood the manuscript in such a way that potential evapotranspiration is derived from raw, i.e. non-bias-corrected, forecasts, but in this case, the term interannual potential evapotranspiration does not make sense. I probably misunderstood something and would like that the authors clarify the manuscript in that respective.

Page 5, lines 3-4: Just a comment, nothing to change: leave-one-year-out might result in the validation years not being really independent, as interannual serial correlation might be quite high. Maybe it would be interesting to test larger block sizes in future studies.

Page 6, lines 21-25: In the case of EDM-m and GDM-m, only 29 data points are used to derive a cumulative distribution function for the reference data. This is a rather low number of data points, potentially leading to estimated cumulative distributions that are non-robust. Maybe, and this is of course rather speculative without analyzing the data, this could be a reason for the worse bias validation of EDM-m and GDM-m in Fig. 6.

Page 6, lines 21-25: I'm not aware of a study that applied gamma distribution fitting for monthly precipitation data. Could you please cite a study to support the method GDM-m? I'm a bit worried that the gamma distribution might not be a good choice for monthly mean precipitation values.

Page 6, lines 25: It is unclearly written how exactly the EDM and GDM correction is applied to daily values. I assume it is done as such that the monthly values are corrected following the quantile mapping procedure. After that, a correction factor is estimated between the corrected and the uncorrected monthly mean value and this correction factor is applied to all daily values. The text on line 25 is though misleading as the actual correction in a quantile-mapping framework is the mapping of the uncorrected values to the cumulative probability space, from which a corrected value is derived following an inverse mapping based on the reference data. As the mapping is calibrated for monthly values, it cannot be used for daily values directly. Please clarify the text.

Page 8, lines 23-24: The first sentence in this paragraph is redundant. Consider removing it.

Page 8, lines 27-28: According to section 3.3, all data was first converted to weekly means, thus a seven-day moving average cannot be derived. Please clarify the contradictions.

Page 9, line 12: Why is the value +0.1 and -0.1 for the deviations from the diagonal chosen?

Page 9, line 19: Unclear use of the word "translate".

Sections 4-6: The presentation of the results could be improved and shortened. When I read the manuscript, I would have liked to have the comparison of the raw and bias-corrected forecasts closer together and I suggest combining the discussion of the raw and the EMDD corrected forecasts. It would be much easier for the reader to follow the discussion if, for e.g., figure 2 and 10 are to be combined into one figure. Similarly for all other seasonal skill score figures in sections 4 and 6.

Page 11, line 8-9: I thought that the reference forecast is the streamflow simulated using the reference precipitation. Thus, any model deficiencies regarding low flows should not affect the skill score as also the reference forecast would suffer from those deficiencies. Also, similarly as for the low flows, the PIT diagram reports difficulties to forecast the high flow. What could be the reason for this issue? The explanations give in the manuscript so far are not fully convincing.

Page 14, lines 19-23: The reasoning is unclear to me, probably due to an unclear explanation how the bias-correction works. If it is done in the way I described in the comment regarding EDM and GDM, I don't think that the reasoning is correct. Everything stated for the monthly correction would also apply for the daily correction. Also on the daily time scale, the rank structure (see comment below) of the forecast is not the same as for the reference data. In both cases (monthly and daily correction), the distribution mapping should be able to correct differing rank structures and remove biases in the monthly mean effectively. In fact, I would have expected the daily correction to perform worse than monthly correction when evaluated on the monthly scale since it is not targeted to the monthly scale but the daily scale. I rather think it has to do with a higher sensitivity of monthly corrections to overfitting as evaluated within the cross-validation framework. Admittedly, distribution mapping can lead to unforeseen effects and it might very well be that I'm wrong. If the authors are convinced that their reasoning is correct, I would like them to describe in the reply a case where the distribution mapping fails in more detail, for e.g. by showing how the reference and forecast distribution look like and how the mapping fails to come up with a correct monthly mean value.

Page 14, lines 19 and 21: Usage of the term "time structure" seems to be misleading. I understand this term in a way that it refers to the temporal sequence of values, i.e. that the day n in the reference corresponds to day n in the forecasts. However, distribution mapping does not have this requirement. It is rather the rank structure as I would call it: Rank n in the reference has to correspond to the rank n in the forecasts. Please correct the terminology or explain in more detail what "time structure" means.

Section 5.2: In my opinion, this section does not give new information which is not already present in figure 6 (time varying bias-correction factors can be inferred from the panel "Before bias correction") and I suggest removing it for the sake of shortening the result section. The only new aspect is that the correction factors for EDMD vary more than for LS, but this comparison is not valid in my point of view as one should not compare a mean correction factor with a correction factor for a quantile level. I'm pretty sure that if you would calculate the correction factor for the mean in the case of EDMD, it would be very similar to the LS factor.

Section 5.3 and figures 8 and 9: I very much like this analysis. I'm not sure though if I really understand the analysis completely. MAE is partly related to the bias analysis in figure 6, i.e. if biases in figure 6 are substantial, then MAE should be even larger since MAE does not allow for a compensation of errors. EDM-y and GDM-y have large biases throughout the year in figure 6, and in some cases and particularly in summer, the bias is even larger than in the uncorrected data. However, in figure 8 the two methods stick out for MAE and IQR in summer lead to skill improvements in all catchments up to a lead time >60 days. To me, this seems to be contradicting. Could you please

explain this particularity?

Page 19, line 1: If I read the figure 10 correctly, there are negative skill score values and therefore, the statement that the skill scores are always larger than zero does not hold.

Page 20, line 3: If I read the figure 12 correctly, there are negative skill score values and therefore, the forecast performs sometimes worse than ESP, which is the opposite of what is stated on this line.

Page 20, lines 12-13: It is not clear to me why this is expected. I would expect that comparison to streamflow climatology is a harder check and therefore the skillfull lead-time should be smaller than in the comparison to the baseline reference run since also the hydrological model bias deteriorates the skill. I surely misunderstand something but I think it would be good to add a bit more explanation in the manuscript.

Page 22, line 4: As for the uncorrected forecast discussion, I do not understand why it is the hydrological model that causes the problems with low-flow overestimation. The reference data is also output of the same hydrological model driven by the reference precipitation data. I would therefore rather think that it is some characteristics in the input data which the bias-correction cannot correct for that causes the problem (for e.g. dry-spell lengths). If the authors still think their statement holds, I would like to have a bit more explanations why this can be the case.

Page 27, lines 13-15: References needed

Section 6.3: Just a comment: I very much like this analysis.

Section 6.4: Although I like the illustrative character of this section, it stands a bit loose within the rest of the manuscript. I suggest to either motivate the section better or, for the sake of brevity, to remove it. In my opinion, the main statements of this sections have already been made, i.e. increased sharpeness after bias-correction compared to ESP.

Figures

Figure 2: "... and all seasons." The figure only shows two seasons, please correct the caption.
Figures 3, 5, 11, 14: The dashed lines should be explained in the figure as well, and not just in the text describing figure 3.
Figure, 6: Although certainly correct, I do not see a reason why to transform the simple relative bias into 1-bias. I understand that this transformation turns the bias into a skill score. However, in my opinion, the interpretation is not following the one for skill scores anyway. The perfect bias-correction would not yield 1 but 0. I suggest plotting the relative bias without transformation. The scale would be much easier interpretable as it directly refers to a percentage over- or underestimation.
Figures 8 and 9: Why are there different color scales for the different seasons?
Figure 15: What are the colours standing for? There is probably also an error in the caption where it reads "shown for all seasons".

Technical comments

Page 19, line 10: precipitation instead of precipitations

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-78, 2016.