

## ***Interactive comment on “Verification of ECMWF System4 for seasonal hydrological forecasting in a northern climate” by Rachel Bazile et al.***

**Rachel Bazile et al.**

rachel.bazile@gadz.org

Received and published: 6 September 2017

### **General comments**

**In this research article, the authors present a detailed study on the predictive skill of hydrological ensemble forecasts in 10 watersheds in Québec, Canada. Different methods are employed which are subject to different degrees of complexity. Among these methods, a simple application of historical streamflow data is seen as benchmark for more complex approaches. The second method, the ESP approach is based on historical meteorological data and accounts for initial conditions in each forecast. The initialization using known system states is also relevant in the third approach, a dynamical seasonal forecast method, in which meteorological forcing is obtained from bias-corrected climate model**

[Printer-friendly version](#)

[Discussion paper](#)



forecasts (ECMWF's System4). Given a lead time of 1 month, the dynamical approach provides improved skill in terms of Continuous Ranked Probability Score (CRPS), while for longer lead times the predictive skill is similar to the corresponding ensemble forecasts using ESP. For the period of snowmelt in spring, the CRPS is lowest (best) in the case of ESP and the dynamical approach. In some watersheds, however, the first method which provides forecasts using historical streamflow data performs best. This comparison highlights the fact that the predictability is low in some watersheds. The study is interesting, the results are promising and the paper fits very well into both the special issue on "Sub-seasonal to seasonal hydrological forecasting" in particular and HESS in general. The methodology is comprehensively presented and the results are discussed in a balanced way. Related work and relevant references are mentioned and acknowledged. Especially the assessment of added value provided by each increased level of complexity (using streamflow data only – > ESP – > dynamical forecasts) is very useful. Another important point is that the paper presents a specific case study in which operational forecasts have already been issued and new methods are going to be implemented. This might be relevant for other forecasting centers. However, in my opinion, the paper needs a few minor revisions and technical corrections. It's my impression that the section on reliability seems to be detached to a certain degree given that the findings from this analysis are not really considered in the summary. Moreover, the paper would benefit from some additional explanations that might improve comprehensibility. Please find my suggestions below.

Response :

First of all, we would like to thank you for your detailed review and constructive comments. All the technical corrections as well as the specific comments number 1, 4, 6, 13 and 14 have already been integrated in a revised version of the manuscript. Moreover, answers and clarifications for the other specific comments are detailed

[Printer-friendly version](#)

[Discussion paper](#)



below.

### Specific comments :

1. **Page 1, line 9: the abbreviation “corr-DSP” is not explained in this context and might be omitted here**

Response : The abbreviation will be omitted in the new version of the manuscript.

2. **Page 1, line 9: Would it make sense to point out that “Simulated streamflow computed using observed meteorological data is used as benchmark.”?**

Response: Yes, it would make sense. The sentence "Simulated streamflows are used as observations" will be replaced by "Simulated streamflow computed using observed meteorological data is used as benchmark" in the revised version of the manuscript

3. **Page 2, lines 12-28: In this section, historical streamflow prediction (HSP) and extended ensemble streamflow prediction (ESP) are presented. In my opinion, some additional explanations might be helpful in this context. You could explain that using HSP is in general possible without using a hydrological model, even though, in particular, you involve the output of a model in your specific case study. ESP, in contrast, does require a hydrological model in order to improve forecasts through explicitly incorporating initial states in the forecasts. The relevance of using hydrological models, as already pointed out, might be helpful in the process of understanding the different methods you apply.**

Response : Thank you for bringing this to our attention. The relevance of hydrological models will be further detailed in this section in the revised version of the manuscript. The differences between HSP and ESP in that regard will also be explained more clearly.

[Printer-friendly version](#)

[Discussion paper](#)



4. **Page 3, line 2: I am not sure whether “questioning” is the appropriate verb in this context. As far as I know it would make sense if you have reason to doubt the usefulness. Instead, using “assessing” might be a better option.**

Response: We agree with the suggested modification. "Questioning" was replaced by "assessing" in the revised version of the manuscript.

5. **Page 3, line 14: Please add appropriate references of the DEMETER project and also explain the project’s acronym.**

Response : The DEMETER acronym stands for 'Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction'. The corresponding reference is:

Palmer, T., Doblus-Reyes, F., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., Feddersen, H., Cantelaube, P., Terres, J., Davey, M., et al.: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), Bulletin of the American Meteorological Society, 85, 853–872, 2004. The definition of the acronym as well as the above mentioned reference will be added in the revised version of the manuscript.

6. **Page 5, Table 1: Please add mean temperature and mean streamflow if easily available. As mean precipitation is indicated, averages of temperature and streamflow might gain insight into the climate characteristics.**

Response: Mean temperature and mean streamflow are indeed easily available and will be added in Table 1 of the revised version of the manuscript.

7. **Page 5, line 13-14: Does this mean that short-term forecasts are extended by the ensembles generated using ESP? Please consider rephrasing.**

Response : Yes, in the operational forecasting system mentioned in the manuscript, the short-term forecasts are extended by ESP. This will be clearly mentioned in the manuscript.

8. **Page 6, lines 27-28: Do the Nash-Sutcliffe values are computed using daily**

[Printer-friendly version](#)

[Discussion paper](#)





**time series?**

Response : Yes. This precision will be added in the revised version of the manuscript.

**9. Page 7, line 18: Forecasts are also computed using one day time steps?**

Response : Yes. Meteorological ensemble forecasts are really computed for 6-hour time steps. However, for this study, forecasts were only available at daily time steps from 0Z to 0Z. Hydrological forecasts are computed at daily time steps. However, hydrological observations are only available at daily time step between 05Z and 05Z. A monthly aggregation of the different variables was chosen for verification purpose in order to limit the impact of the lag between meteorological and hydrological forecasts. Those precisions will be added in the new version of the manuscript.

**10. Page 8, line 6: Please indicate why the number of forecasts amounts to 420. 35 years x 3 months x 4 seasons?**

Response : The total number of forecasts available for verification is 420 because one forecast is emitted the 1st of each month between 1981 and 2014. Consequently, we have 35 years x 12 months to assess the performance of the forecasting system. However, both meteorological and streamflow observations are not available after the 31th of December 2015, lead-time 2 to 7 counts 419 to 413 forecast-observation pairs for the verification. Those precisions will be included in the revised version of the manuscript.

**11. Page 8, line 23: Is the term "confidence interval" really correct in this context? As far as I understand, say we consider a reliable forecast of a specific event, a probability of 95% should at best also refer to 95 out of 100 occasions in the observed dataset. Please also define the terms "nominal" and "effective".**

Response : We propose to add the following sentence in the revised manuscript

to define the terms 'nominal' and 'effective' as well as to clarify the use of the term 'confidence interval': *The reliability diagram diagnostic tool compares the observed coverage frequency (effective,  $1 - \hat{\alpha}$ ) with the corresponding theoretical confidence levels (nominal,  $1 - \alpha$ ) of predictive confidence intervals calculated from ensemble forecasts. Of course, if forecasts are reliable, these values  $1 - \hat{\alpha}$  and  $1 - \alpha$  should be equal for any confidence level.*

**12. Page 10, lines 13-14: This phrase is hard to understand. Please consider rephrasing.**

Response : We suggest replacing the sentence 'A leave-one-year-out procedure is used, which consists in excluding the forecast to correct from the bias evaluation process.' by '*A leave-one-year-out procedure is used to calculate the bias and correct the forecast. This consists in calculating the bias based on available past forecasts issued on the same month excluding the month under correction.*'

**13. Page 12, line 13: “evolution” might be more appropriate than “maturation” in this context.**

Response: We agree. The revised version of the manuscript will be corrected according to your suggestion.

**14. Page 13, line 1: corr-DSP forecasts**

Response: This will be corrected in the revised version of the manuscript.

**15. Page 14, line 1: Here, you state that Fig. 6 presents a reliability diagram while in the figure’s caption it is labeled as PIT diagram. This is a little bit misleading and might cause confusions even if the type of information is similar to a certain degree. Please confirm or specify the type of figure more detailed.**

Response : We apologize for this typo. The title of Fig. 6 has already been corrected in the revised version of the manuscript.

Printer-friendly version

Discussion paper



16. **Page 14, line 11: Here, you state that Fig. 7 presents a PIT histogram while in the figure's caption it is labeled as rank histogram. Is this in line with your explanations in Sect. 4?**

Response : You are right, Figure 7 should be entitled 'PIT diagram' instead of 'Rank histogram'. This will be corrected in the revised version of the manuscript. As mentioned in section 4, PIT histograms and rank histograms are equivalent in terms of interpretation. However, Fig. 7 really represents a PIT histogram.

17. **Page 14, lines 11-12: Further explanations might improve comprehensibility (e.g., by stating that an equal distribution indicate accurate ensemble forecasts).**

Response : Further explanations about the interpretation (flat, bias and over/under-dispersive cases) of the PIT histogram will be added in the revised version of the manuscript to help the interpretation of Figure 7.

18. **Page 14, Figure 6: In my opinion, labeling each row of the diagram by stating the watershed's numbers might be more intuitive (see, e.g., Figure 2). This is also relevant in the case of Figures 7, 8, 9, and 10.**

Response : Yes. The watershed's numbers will be added directly on the rows of the figures 6, 7, 8, 9 and 10 instead of featuring only in the figures' labels.

19. **Page 15, line 16: The bias correction is applied for each month. Single events at time scales smaller than one month might be subject to biases different to the monthly values.**

Response : We completely agree. This is certainly a limit of the chosen bias correction method (which has the advantage of being simple, but it is possible that in this case a more sophisticated bias correction method would be worth the additional complexity). This issue will be discussed in the revised version of the manuscript.

20. **Page 16, line 11: By the way, the term dispersion is often used throughout**

[Printer-friendly version](#)

[Discussion paper](#)



**the manuscript if the variability is overestimated (or underestimated). Variability might be more appropriate as mentioned in line 6 on page 15.**

Response : The same issue has been addressed by referee #2. Throughout the manuscript, *Dispersion* refers directly to the spread of a single ensemble forecast, namely the variability of the members. This definition of the term *dispersion* will be added in the revised manuscript. We use the term *variability* to characterize the variation in a data set, such as the CRPS of different forecasts.

**21. Page 17, line 11: Do you mean corr-DSP when discussing the results of ensemble meteorological forecasts?**

Response : No. The denomination corr-DSP refers to streamflow forecasts only. The general denomination 'ensemble meteorological forecasts' refers to uncorrected temperature and precipitation forecasts. Bias corrected ensemble meteorological forecasts are the forecasts used to produce corr-DSP (by passing them to the hydrological model).

**22. Page 17, line 12: Is it possible to prove if the skill is significant or not significant from your analyses? The term significant should be proved by providing statistical measures.**

Response : We agree. It would be feasible to compute approximate confidence intervals for the CRPS using a bootstrap procedure. These intervals could then be used to add some more formal indications about the significance of the verification results. This will be done for the revised version of the manuscript. The remaining of the manuscript will also be verified to make sure that there are no other instances.

**23. Page 18, line 5-6: Please explain in brief why corr-DSP is less reliable. Is this finding relevant for winter or all seasons? Maybe you can refer to the reliability diagram?**

Response : As shown in Figure 6 (reliability diagrams), the reliability for corr-

[Printer-friendly version](#)

[Discussion paper](#)





DSP varies in with lead-time, season and watershed. The causes of this lack of reliability are more visible in Figure 7. In some cases, such as 1a) or 2d), biases are still present. In other cases, under-dispersive behaviors are observed such as in cases 1b) or 2c). This inadequate forecast uncertainty representation behavior could be caused by the bias correction which may have reduced the dispersion of the precipitation forecasts. Those explanations will be included in the revised version of the manuscript.

**Technical corrections :**

**Page 1, line 6: Please add the CRPS (abbreviation) here as it is mentioned later without explanation (cf. line 14)**

**Page 2, line 32: I would suggest using the singular form of precipitation**

**Page 8, line 20: distributions (plural)**

**Page 10, Figure 2: Please add the dimension of the precipitation bias in the color bar.**

**Page 15, line 15: remain**

**Page 17, line 11: “is predictable” instead of “are predictable”**

**Page 18, line 6: originates**

Response: We agree with all the suggested technical corrections and we thank you for pointing them out. They have already been included in the revised version of the manuscript that we are preparing.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2017-387>, 2017.

[Printer-friendly version](#)[Discussion paper](#)