Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-369-RC2, 2019

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Comparison of probabilistic post-processing approaches for improving NWP-based daily and weekly reference evapotranspiration forecasts" *by* Hanoi Medina and Di Tian

Anonymous Referee #2

Received and published: 24 September 2019

The manuscript "Comparison of probabilistic post-processing approaches for improving NWP-based daily and weekly reference evapotranspiration forecasts" discusses the application of different standard post-processing approaches to (multi-model) ensemble forecasts of reference evapotranspiration (ETo). Though the methods used are quite standard in the field of statistical post-processing, its application to ETo is novel. The authors performed an exhaustive comparison of different post-processing approaches, non-homogenous regression (NGR), Bayesian model averaging (BMA), and affine kernel dressing (AKD) based on three meteorological models and 9 regions.

In general the paper is well written. However, the results section is a bit lengthy (large number of figures and tables). I think it can be condensed without loosing any information. Further, the manuscript would highly benefit from showing some standard verification metrics like the CRPS (Matheson and Winkler 1976; Gneiting and Raftery 2007) and histograms of probability integral transforms (PIT; Diebold et al. 1998, Gneiting et al. 2007).

Specific comments

- As mentioned by reviewer 1 the post-processing methods NGR, BMA, and AKD are not novel and have been used for over ten years in the field of probabilistic (hydro-)meteorological forecasting. Hence, I highly recommend not to use words like "novel" so many times, even though to my knowledge these methods haven't been applied to ETo so far.

- Reviewer 1 mentioned that the continuous ranked probability score (CRPS) should be included for the verification of the probabilistic forecasts. I strongly agree on this, as the CRPS is a widely accepted verification score that considers errors in location and spread simultaneously.

- Further, PIT histograms of the different forecasts should be shown. Like the spread skill plots they measure reliability. However, PIT histograms may help to detect additional issues like potentially wrongly specified parametric distributions in NGR. Subpanels with PIT histograms for selected forecasts could be added to Figures 3 and 7.

- The figures captions are not detailed enough. I recommend to move some of the figure descriptions from the main text to the figure captions.

- Tables 3 and 6 show only positive biases even after statistical post-processing. However, for a well calibrated forecast, I would expect that the bias is zero in expectation with some random fluctuations. Accordingly, I would expect to see some negative biases for the post-processed forecasts in tables 3 and 3. Why is this not the case?

- L42-45: Please rephrase in a way that it becomes clear that the methods developed by Gneiting et al. (2005) and Raftery et al. (2005) correct also errors in the variance.

- L66-68: Are the presented post-processing approaches really suitable for extreme values? Being some kind of extended regression models I expect the post-processing methods to perform well in most of the forecast cases, but not for extremes, for which appropriate training data are typically very rare. Further, for extremes the skill of the post-processing approaches highly depends on the shape of the parametric distributions used. As the post-processing approaches presented in this manuscript are all based on normal distributions, there is no possibility to specifically fit the tails of the forecast distributions, which are key when it comes to extremes. Accordingly, I recommend to either remove the sentence on lines 65-66 or explain your reasoning in more detail.

- L76: “AF . . . rely on the mean of retrospective reforecasts, thus neglecting information about their dispersion” This is not generally true for analog forecasts. Using a similarity measure that considers also the second moment of the forecast distribution may allow to consider dispersion as well.

- L93: “. . .while better fits the user’s actual needs” This part of the sentence is a bit out of sync and difficult to understand. What do you mean?

- L116-118: Exchangeability is not relevant here, because only ensemble statistics are considered.

- L128: I assume that z denotes a vector? Does u(z) then map the z to a scalar?

- L346-L347: PIT histograms would help to corroborate this

Technical corrections

- Equations: colon prior to equations can be omitted in most of the cases

- L43: Gneiting instead of Gneitting

- L60: . . .ETo forecasts is considerably. . .

- L71: . . .post-processing of ETo. . .

- L155: . . .generated at a specific. . .

- L156-157: . . .independent from each other.

- L231: . . .the coverage rations in Table 2 provides. . .

- Tables 2,4,5,6: “best” values in bold font like in table 3

- L270: . . .the weights of the NCEP model. . .

- L330: see L43

- Figure 2: What is the unit of the lead time?

- Figure 2 / caption: Please mention here that results for daily forecast s are shown here. This comment applies also to the other figures: Please indicate whether they show results for daily or weekly forecasts.

- Figure 4a: The uppermost line probably shows Brier scores (BS) and not Brier skill scores (BSS). Please modify the caption accordingly.

- Figure 6: Which quantiles are shown by the box-plots?

References:

- Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. International Economic Review 39, 863–883.

- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378.

- Matheson, J. E. and R. L. Winkler (1976). Scoring rules for continuous probability distributions. Management Science 22, 1087–1096.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-369, 2019.