Reviewer #1:

Comment No. 1

R. *Thank you for the interesting paper on post-processing ensemble forecasts of reference evapotranspiration with lead times up to a week, on which I believe you are right to point out that not much research has been published yet. I do not have any major concerns with this paper. I do have the following suggestions for additions and improving clarity of the paper:*
Introduction:
Please reformulate or leave out 'emerging', 'novel', and 'new' throughout the manuscript when referring to the probabilistic post-processing methods applied in this study, because the methods referred to here, date from over 10 years back. I would suggest to extend the literature review on post-processing methods for meteorological ensemble forecasts with, for example, quantile mapping and extended logistic regression (e.g. Whan and Schmeits 2018; Messner et al. 2014; Verkade et al. 2013). And then provide the reasoning for selecting NGR, BMA, and AKD for this research. (Page 2, line 61).

A. We sincerely thank you for your time and the very useful comments, which considerably helped us to improve the manuscript.
We agree that we used inaccurate terms to describe the methods. This has been amended in the new manuscript. Only in one occasion we refer to "the use of new ETo forecasting strategies" to emphasize that the strategies are for the first time applied to ETo, as the second reviewer has pointed out. In a few cases we now also use the term "state of art" to qualify the methods, which is commonly managed in literature (for example Gneiting, 2014).

References
Gneiting, T.: Calibration of medium-range weather forecasts, European Centre for Medium-Range Weather Forecasts, Technical Memorandum No. 71, 2014.

Comment No. 2

R. Methodology:
Please provide the equations and detailed definition of variables and parameters therein of each of the performance metrics used (Section 2.4). Please add analyses of CRPS(S) results for comparison with other recent hydrometeorological ensemble forecast studies (e.g. van Osnabrugge et al. 2019).

A. Thanks for pointing this out. In the revised manuscript, we provide the equations and detailed definition of variables and parameters therein of each of the performance metrics used. Based on this and a similar request of the other reviewer we added analyses of the CRPS, and compare the results with other recent hydrometeorological ensemble forecast studies (e.g. van Osnabrugge et al. 2019). The tables 2, 3, A1 and A2 (this two last tables are now in the Annex section) now include the CRPS metrics. In addition, considering the other reviewer's requests we added analyses of the rank histograms. The histograms have been added as subpanels in Figs. 3 and 7, as also suggested for the other reviewer.

R. Please keep consistently clear throughout the paper when you are referring to ensemble and probabilistic forecasts. (Because of the presented deterministic performance metrics, at some points in the manuscript impression may arise that deterministic forecasts are concerned). Please see annotated pdf for some examples.
A. Thank you for pointing this out. It has been revised throughout the manuscript.

Comment No. 3

R. Results:
I think that there are too many results presented in the main text. Consider that graphs and tables partly present the same information. Remove redundancies and consider moving part of the results to an Annex.
A. Thank you for pointing out this. The redundancies have been removed in the revised manuscript. We removed one table (originally Table 2) that contained redundancies with other tables/figures and moved Tables 3 and 4 (now A1 and A2) to an Annex.

R. For clarity, I would recommend that every time when reporting or discussing forecast BSS the reference forecasts used to calculate BSS should be mentioned.

A. It has been clarified in the revised manuscript by including the equations and the detailed definition of variables and parameters in Section 2.4. We in all cases compute the BSS associated to the tercile events of the ETo forecasts, such that the sample climatology is equal to $0.3\overline{3}$ and $BS_{\mathrm{clim}} = 0.2\overline{2}$.

R. In general, when discussing forecast performance, please clearly state which forecasts you are referring to as a benchmark (e.g. climatology, persistence, raw ECMWF, or BC ECMWF) Figure 4 - Consider simply presenting BSS with BC-ECMWF as reference forecasts.

A. This issue has been addressed in the revised manuscript following your suggestions. The changes in the Methods section also helped to address this issue. We would prefer to keep figure 4 in the current format by using the same reference throughout the document to avoid any confusion, and also for the purposes of comparison with other studies. On the other hand, we think that the differences between BSS are easier to interpret, because the Brier score of climatology is constant in the experiments.

Comment No. 4

R. Discussion:
It would be interesting if you could discuss some of the earlier published results of post-processing ensemble forecasts of temperature, wind speed, and radiation, and how using these post-processed products, instead of the raw forecasts, to construct ETo forecasts would compare to the post-processed ensemble forecasts of ETo of this research.

A. As requested, we added a new section to discuss these issues. We discuss some of the earlier published results of post-processing ensemble forecasts of temperature, wind speed, and radiation. We sustain that the post-processing of meteorological forecasts for producing the ETo ensembles may require accounting for the multivariate dependence among those forcing, which is often difficult (e.g. Wilks, 2015). Kang et al (2010) found that post-processing of the streamflow forecasts provided more accurate predictions than post-processing the forcing alone, while Vekade et al (2013) showed that the improvements in precipitation and temperature through the post-processing hardly benefited the streamflow forecasts. Lewis et al., 2014 showed that the performance of the ETo forecasts can largely surpass that of the individual input variables. Please find detailed discussions in the revised manuscript.

References

Kang, T.H., Kim, Y.O. and Hong, I.P., 2010. Comparison of pre - and post - processors for ensemble streamflow prediction. Atmospheric Science Letters, 11(2), pp.153-159.

Verkade, J.S., Brown, J.D., Reggiani, P. and Weerts, A.H., 2013. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. Journal of Hydrology, 501, pp.73-91.

Wilks, D.S., 2015. Multivariate ensemble Model Output Statistics using empirical copulas. Quarterly Journal of the Royal Meteorological Society, 141(688), pp.945-952.

Lewis, C.S., Geli, H.M. and Neale, C.M., 2014. Comparison of the NLDAS weather forcing model to agrometeorological measurements in the western United States. Journal of hydrology, 510, pp.385-392.

Comment No. 5

R. Conclusion (and Abstract):

The relevance of differences in computational efficiency (Page 13, line 404) depends on what the computational time is, what the intended application is, and what will be the hardware on which these expected applications will run. None of these considerations are currently written here, which is too limited for a discussion on computational time (also not discussed earlier as criterion earlier in the paper, just mentioned). Please expand or consider leaving out the issue of computational efficiency.

A. Yes, we agree. Thanks for this helpful comment. In the revised manuscript, we have removed the related statements.

R. Please find detailed comments and editorials in the annotated pdf.

A. We want to thank the reviewer again for the helpful comments. We followed all the suggestions except the one suggesting adding a map with the locations of the three forecast grids used. Instead we added the following comment:

"The forecasts were interpolated to the same 0.5^o × 0.5^o grid using the TIGGE data portal". We pondered adding the grid in Fig. 1 but it involves a too dense set of points, which affect the quality of the figure.

All the following references have been added in the revised manuscript:

Bremnes, J. B.: Probabilistic Wind Power Forecasts Using Local Quantile Regression, Wind Energy, 7, 47–54, 2004.

Davò, F., Alessandrini, S., Sperati, S., Delle Monache, L., Airoldi, D. and Vespucci, M. T.: Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting, Solar Energy, 134, pp.327-338, 2016

Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B. and Searight, K.: Probabilistic weather prediction with an analog ensemble, Monthly Weather Review, 141(10), pp.3498-3516, 2013.

Hagedorn, R., Hamill, T. M. and Whitaker, J. S.: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: Two-meter temperatures. Monthly Weather Review, 136, 2608–2619, 2008.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather and Forecasting, 15(5), pp.559-570, 2000.

Kang, T. H., Kim, Y. O. and Hong, I. P.: Comparison of pre - and post - processors for ensemble streamflow prediction, Atmospheric Science Letters, 11(2), pp.153-159, 2010.

Kann, A., Haiden, T. and Wittmann, C.: Combining 2-m temperature nowcasting and short-range ensemble forecasting, Nonlinear Processes in Geophysics, 18, 903–910, 2011.

Kann, A., Wittmann, C., Wang, Y. and Ma, X.: Calibrating 2-m temperature of limited-area ensemble forecasts using high-resolution analysis. Monthly Weather Review, 137, 3373–3387, 2009. Hagedorn, R.: Using the ECMWF reforecast data set to calibrate EPS forecasts. ECMWF Newsletter, 117, 8–13, 2008.

Madadgar, S., Moradkhani, H. and Garen, D.: Towards improved post-processing of hydrologic forecast ensembles. Hydrological Processes, 28(1), pp.104-122, 2014.

Messner, J. W., Mayr, G. J., Zeileis, A. and Wilks, D. S.: Heteroscedastic Extended Logistic Regression for Postprocessing of Ensemble Guidance. Mon. Wea. Rev., 142, 448–456, https://doi.org/10.1175/MWR-D-13-00271.1, 2014.

Møller, J. K., Nielsen, H. A., and Madsen, H.: Time-Adaptive Quantile Regression, Computational Statistics & Data Analysis, 52, 1292–1303, 2008.

Osnabrugge, B. V., Uijlenhoet, R. and Weerts, A.: Contribution of potential evaporation forecasts to 10-day streamflow forecast skill for the Rhine River. Hydrology and Earth System Sciences, 23(3), pp.1453-1467, 2019.

Pinson, P., and Madsen, H.: Ensemble-Based Probabilistic Forecasting at Horns Rev, Wind Energy, 12, 137–155, 2009.

Scheuerer, M. and Büermann, L.: Spatially adaptive post - processing of ensemble forecasts for temperature, Journal of the Royal Statistical Society: Series C (Applied Statistics), 63(3), pp.405-422, 2014.

Sloughter, J. M., Gneiting, T. and Raftery, A. E.: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging, Journal of the american statistical association, 105(489), pp.25-35, 2010.

Vanvyve, E., Delle Monache, L., Monaghan, A.J. and Pinto, J.O.: Wind resource estimates with an analog ensemble approach, Renewable Energy, 74, pp.761-773, 2015.

Verkade, J. S., Brown, J. D., Reggiani, P. and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, Journal of Hydrology, 501, pp.73-91, 2013.

Verzijlbergh, R. A., Heijnen, P. W., de Roode, S. R., Los, A. and Jonker, H. J.: Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications, Solar Energy, 118, pp.634-645, 2015

Whan, K. and Schmeits, M: Comparing Area Probability Forecasts of (Extreme) Local Precipitation Using Parametric and Machine Learning Statistical Postprocessing Methods. Mon. Wea. Rev., 146, 3651–3673, https://doi.org/10.1175/MWR-D-17-0290.1., 2018.

Wilks, D. S.: Multivariate ensemble Model Output Statistics using empirical copulas, Quarterly Journal of the Royal Meteorological Society, 141(688), pp.945-952, 2015.

Wilks, D.S.: Comparison of ensemble-MOS methods in the Lorenz'96 setting, Meteorological Applications, 13(3), pp.243-256, 2006.

Wilks, D.S.: Extending logistic regression to provide full probability distribution MOS forecasts, Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 16(3), pp.361-368, 2009.

Wilks, D.S.: Sampling distributions of the Brier score and Brier skill score under serial dependence, Q J Roy Meteor Soc, 136(653): 2109-2118, 2010.

Williams, R. M., Ferro, C. A. T. and Kwasniok, F.: A comparison of ensemble post-processing methods for extreme events. Quarterly Journal of the Royal Meteorological Society, 140(680), pp.1112-1120, 2014.

Zhang, J., Draxl, C., Hopson, T., Delle Monache, L., Vanvyve, E. and Hodge, B. M.: Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods, Applied Energy, 156, pp.528-541, 2015.