Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-55-AC1, 2019

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Interactive comment on "Order of operation for multi-stage post-processing of ensemble wind forecast trajectories" *by* Nina Schuhen

Nina Schuhen

nina.schuhen@nr.no

Received and published: 10 December 2019

Printer-friendly version

Discussion paper

Response to R1

We thank the reviewer for their thoughtful comments and overall positive assessment of our paper. Please find below a point-by-point response to the comments.

Specific comments

1. 161: For RAFT$_{ens}$, were separate correlation matrices calculated for each member or were data pooled across all members and a single correlation matrix estimated and used for all members?

   The correlation matrices for RAFT$_{ens}$ are unique to the ensemble members. We added a sentence to the revised manuscript to clarify this.

2. 191: The explanation here is somewhat confusing. What exactly is averaged here? Please add some more detailed explanation. Also, while I don't fully understand in which sequence the different lead times are being processed, I am surprised that both the preceding and subsequent lead times are assumed to have been processed at this point. Please add some explanation for this as well.

   We added some text to the description of the algorithm that makes it clear that Step 3 is only carried out after Step 2 has finished for all lead times and that in 3(b)

Printer-friendly version

Discussion paper

we are averaging the values for the adjustment period length of the neighboring lead times.

3. 242: I am not happy with the formulation 'A flat histogram denotes perfect calibration'. A flat histogram is a necessary but not sufficient for calibration, and examples can be constructed where uncalibrated forecasts yield near flat histograms (e.g. Hamill, 2001, or Thorarinsdottir et al., 2016, their Fig. 7).

   This is of course correct. We changed the text to reflect that a flat histogram is not a sufficient condition for calibration.

4. 368: 'From this we can conclude ...': I don't disagree with the conclusion in general but I find this pair of sentences logically confusing. If EMOS + RAFT$_m$ + ECC results in a better score, how does this imply that RAFT$_{ens}$ preserves the multivariate correlation structure? Please clarify.

   We changed this sentence to "The almost identical variogram scores suggest that RAFT$_{ens}$ manages to preserve the multivariate correlation structure throughout its multiple iterations." The variogram score is sensitive to misspecifications in the multivariate correlation structure. Thus the results indicate that applying RAFT$_{ens}$ does not destroy the dependencies between lead times, as it performs similar to RAFT$_m$ + ECC, which follows the ensemble's multivariate structure.

5. 371: I would be careful with the use of the term 'underdispersed' in connection with the average rank histogram. What we see is a U-shape, but the discussion in Thorarinsdottir et al., 2016 suggests that this can be caused by miscalibration other than underdispersion of the ensemble. I'd refer to the univariate histograms first as their interpretation is less ambiguous.

   We changed the text here, so that it is clear that we consulted the band depth histogram to eliminate other causes of miscalibration than underdispersion. The

same holds for our interpretation of the EMOS + RAFT$_m$ + ECC average rank histogram as the correlation between components being too weak.

6. 383: Can the author provide a reference for the claim that 'the CRPS is usually more sensitive to the error in the forecast mean'? This is not obvious to me since the CRPS- unlike RMSE or Euclidean error - assesses the calibration of the full distribution.

   We added text to the revised manuscript that this follows from Figure 4 in Friederichs and Thorarinsdottir (2012) and is also the case for the multivariate version of the CRPS, the energy score (Pinson and Tastu, 2013).

7. Fig. 8b: While I agree that at this point it is not required to always show results for both gEMOS and logEMOS given that it has been found previously that their performance is comparable, I wonder why the author chose to alternate between the two. This could create the impression that figures where hand picked to support a particular conclusion. I'd either consistently use gEMOS or logEMOS, or provide the underlying rationale if there is indeed a good reason to consider gEMOS in one and logEMOS in another context.

   The reviewer is correct in that using the logEMOS plot here might create an impression we did not intend to convey. We have switched Figure 8b for the gEMOS version, which looks almost identical.

8. 346-347: it has be to applied ->it has to be applied

   This has been corrected in the revised manuscript.

9. 351: is is ->it is

   This has been corrected in the revised manuscript.

References

Friederichs, P. and Thorarinsdottir, T. L.: Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction, *Environmetrics*, 23, 579–594, https://doi.org/10.1002/env.2176, 2012.

Pinson, P. and Tastu, J.: Discrimination ability of the energy score, Technical report, Technical University of Denmark (DTU), 2013.