Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-62-RC3, 2020

Nonlinear Processes in Geophysics
Discussions

Interactive comment on "Simulation-based comparison of multivariate ensemble post-processing methods" *by* Sebastian Lerch et al.

Zied Ben Bouallegue (Referee)

zied.benbouallegue@ecmwf.int

Received and published: 7 February 2020

This paper present results of an intercomparison study. Different multivariate ensemble post-processing methods are compared using toy-model simulations. The focus is on empirical copula methods that are generally applied as a second post-processing step, after univariate post-processing step, in order to provide coherent multivariate structure to ensemble calibrated forecasts.

As the reviewer is the developer of one of the methods compared here, he prefers to make himself know. There exists no conflict of interest (stricto sensu) but potential cognitive biases from the reviewer side when scrutinizing the results.

The paper is clear, well structured, and well written. However, the choice regarding the selection of illustrations is not sufficiently motivated to my opinion. This choice is important because it drives the discussion and the main conclusion of the study. A 3-point argument is developed below to explain this criticism.

You claim in the conclusion (L462/463) that the 4 simulation settings aim to mimic different situations and challenges occurring in practical situations. However, the link with practical situations is sometimes weak or missing. In particular, it would be interesting to link misspecification definitions with practical examples. What sigma > 1 and sigma < 1 mean, and similarly what does rho > rho0 and rho < rho0 mean and "look like" in practice? In which situations should one expect to encounter these types of misspecification? Which type of misspecification situations are the most common in practice? Are combinations of misspecified elements more common than others (sigma <0 and rho< rho0 for example)?

Once the link to the applications is clarified, illustrations could be chosen consistently. Result material is abundant, so one selection criterion could be to focus on the main misspecification encountered in practical situations. For example, you use sigma equals 1 to illustrate results in Setting 2. Does that often occur in practice? One could rather illustrate Setting 2 with sigma<1. Similarly, for Setting 3, scenario B is used for illustration purposes. It corresponds to the case where the ensemble forecasts have a heavier right tails and slightly higher point mass at zero than the observations. Does that often occur in practice? No justification for the use of this scenario as reference is provided. Scenario A (the ensemble comes from a distribution with smaller spread) or Scenario C (the observation distribution has a much heavier right tail) seem to be more likely to be faced in practice.

Your conclusion points to the robustness of the SSh method and so you encourage post-processing practitioners to consider this method as their first choice. Based on the results presented in the manuscript and the ones in the supplemental material document, one could draw the opposite recommendation. First choice methods are

available for different types of misspecification and so one could encourage to apply SSh only when the misspecification is unknown or difficult to identify. Specific recommendations would read:

1.If the ensemble is underdispersive and the ensemble correlation is too weak, d-ECC is the best option, both in terms of ES and VS.

2.If the ensemble is underdispersive and the ensemble correlation is too strong, ECC-S is the best option in terms of VS and one of the best options in terms of ES.

3.If the ensemble is overdispersive and the ensemble correlation is too strong, d-ECC is the best option, both in terms of ES and VS.

4. If the ensemble is overdispersive and the ensemble correlation is too weak, GCA is the best option in terms of VS, but is less performant in terms of ES. GCA proves to work well even when the overdispersiveness error characteristic is relaxed.

5. Otherwise consider using SSh, in particular when the correlation structures in the forecasts and observations are very dissimilar.

This is valid for all types of distributions (so all types of weather variables). Are these conclusions still valid in case of time varying misspecification? Verification results are missing to conclude here. Therefore, the authors are encouraged to investigate various sigmas in Setting 4 in order to collect evidences, and could draw conclusions accordingly.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2019-62, 2020.