Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-62-RC1, 2020

Nonlinear Processes in Geophysics
Discussions

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

Anonymous Referee #1

Received and published: 25 January 2020

This paper reports the results of a comprehensive simulation study comparing different methods for modeling spatial, temporal, and inter-variable correlations in statistically postprocessed ensemble forecasts. With a number of new multivariate methods having been developed in recent years, a study like this is of great interest as it allows readers to get an overview of the strengths and limitations of the different approaches.

General comments:

1. With the goal of this paper being a comparison between different methods for multivariate ensemble postprocessing, I feel that a more detailed discussion of the key features (e.g. optimal sampling of the predictive distribution vs. random sampling, as-

sumption of stationarity of the copula structure vs. flow dependent copula structure, etc.) of the different methods should be given (possibly in the form of a table). This could serve as a motivation for the different simulation settings, which try to mimic situtations where some of the assumptions are met while others are not.

2. Related to 1., I feel that the role of ensemble size (which has a big impact on the representation of the multivariate distribution) should be discussed a bit more. This seems relevant as for some methods it is easy to generate an ensemble of any size while for others it is not. I'm not suggesting that additional experiments should be performed, but a brief discussion of the findings in Wilks (2015) could be useful in a context where strengths and limitations of different multivariate postprocessing approaches are compared.

3. Why has so much focus been given to simulation settings that are based on a time-independent model for the simulated forecasts and observations? I would argue that this (time-independence) is not a feature commonly encountered in applications, but with 3 out of 4 settings being time-independent, multivariate methods that assume a stationary copula structure could be perceived as being more versatile than they really are. Agreed that setting 1 is a natural starting point for such a comparison and that setting 3 is interesting because of the entirely different nature of the marginal distributions (skewness, possibility of heavy tails, mixed discrete-continuous distributions), but what do we learn from setting 2 that we cannot learn from 1 and 3? The main difference to 1 seens to be the poorer performance of GCA, but an explanation for this is not given, and so the insights gained from this setting are limited.

Specific comments:

131-132: Please check if that statement is correct. In my recollection the selection of past observations in Clark et al. (2004) was not random, but was based on the valid date of the forecast

293-297: I find setting 4 the most interesting, but I find the particular definitions of the

model parameters unnecessarily complicated. Specifically, I don't get a good intuition of what kind of time-varying correlations this model implies. Couldn't one simply define:

$\Sigma_{i,j} = \sigma\rho^{|i-j\}}$

as in the other settings, but now make ρ time-varying, e.g. via

$\rho(t) = \rho_0(1 - a/2) + \rho_0(a/2)\sin(\frac{2\pi t}{n}), a \in (0, 1)$

This model would be autoregressive with lag-1 correlations oscillating between ρ_0 and $\rho_0(1 - a)$, and thus have a more intuitive interpretation.

260-264: I find this notation a bit confusing since previously the subscript/superscript 'O' was used for observations and here the subscript '0' (which is hard to discern from 'O' in the NPG font) is used to denote the fraction of zero values. The notation is also inconsistent in that in setting 3 'x' and 'y' are used to denote forecasts and observations, in contrast to the subscript/superscript 'O' for observations in the other settings.

323 '... are identical to those of ECC-Q ...': Is this really true for ECC-S? The way it is described here, ECC-S seems to imply some level of randomization (albeit less than ECC-R), so the sampling is not the same as for ECC-Q.

Fig. 2: I don't think that this figure is really necessary. Why is the (univariate) performace of ECC-Q compared to the raw ensemble relevant to the comparison of multivariate postprocessing approaches?

354 '... SSh never performs substantially worse ...': Why would we expect otherwise? The only drawback of SSh in the present context is the underlying assumption of time-invariance of the correlation structure, which is not a drawback in a time-invariant simulation setting. If not discussed before, this paragraph could be an opportunity to discuss this issue of time-invariant vs time-varying correlations.

360: I wonder if ECC-S gives the better results for the wrong reason here. Maybe I am misunderstanding its key idea, but to me this is essentially a compromise between

ECC-Q and ECC-R. Is it possible that the small amount of randomization in ECC-S weakens the correlations, which is beneficial for $\rho > \rho_0$ and detrimental for $\rho < \rho_0$? An argument against this hypothesis is that the performance of ECC-S is not significantly different from ECC-Q when $\rho = \rho_0$. I just cannot think of any good reason why ECC-S would be better than ECC-Q. If the authors have any explanation for these results I would encourage them to include those in the discussion of the results.

Fig. 6: The caption should state that this is scenario B from setting 3

458 'changes over iterations': I would change this terminology and speak of 'time' instead of 'iterations', and be more specific about what changes/varies over time. Likewise in 464-465 I would clarify what you mean by 'structural change'

Language and typos:

43: ... studies allow one to specifically tailor ...

135: sufficiently many

137: 'not directly straightforward' sounds weird, I'd just say 'not straightforward'

249: ... allows one to generate ...

375: 'can potentially be explained' sounds weird, maybe better 'may be explained'

388: I think you want to say 'In terms of ...'

443: Change to 'the other way round' and 'In accordance with'

445: Change to 'in contrast to'

References:

D.S. Wilks (2015): Multivariate ensemble Model Output Statistics using empirical copulas. QJRMS 141(688), 945-952.

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

2019-62, 2020.