Response to comment/review - *egusphere-2022-434*:

Guidance on how to improve vertical covariance localization based on a 1000-member ensemble

by Tobias Necker, David Hinger, Philipp Johannes Griewank, Takemasa Miyoshi, and Martin Weissmann

We would like to thank the editor for his helpful comments and for answering questions. During the revision process, we performed additional computations and studied different approaches that allow for achieving positive semi-definiteness of a localization matrix. The revised version of the manuscript includes these outcomes, which we summarize in a newly added section. We also expanded the theory and conclusion sections, which now discuss the positive semi-definite character of covariances and localization.
Below we respond to the editor's comments and further explain how we addressed them in the revised manuscript. The original queries are bold, and our responses are normal text. All changes in the manuscript are clearly marked in the pdf latex diff document.

EC2: ['Reply on AC3'], Olivier Talagrand, 19 Sep 2022

To the authors,

I thank you for the revised version of your paper. You now discuss the symmetric positive semi-definite (SPSD) character of the localization matrix C. You also mention in the response in the Interactive Discussion that this matrix is not SPSD (with eigenvalues ranging from -0.5 to 50, which suggests that the negative part of the matrix is in a sense 'small'). You also mention that the matrix C' obtained by setting the negative eigenvalues to 0 changes the localized values by up to 15%.

These results are interesting, and should in my mind be included (together with other possible results) in your paper. More generally, I think that the diagnostics you have performed on the basis of the matrix C (starting with subsection 3.1) should be performed again with the corrected matrix C' (or actually with any other appropriate SPSD matrix you might build from your original matrix C). I do not think the required cost would be prohibitive, and that would greatly increase the interest and significance of your paper. Actually, the diagnostics you have performed with your matrix C could be removed from the paper if they are replaced with new diagnostics obtained with an SPSD localization matrix. It may be that the new diagnostics will turn out to be similar to the previous ones. If so, that would actually be very instructive.

Answer: We addressed your comments and substantially revised the manuscript with respect to the SPSD aspect. We now compare different approaches that allow for achieving SPSD of the EOL localization matrix. This analysis includes the application of the nearest correlation matrix (NCM) algorithm (Higham 2002), which is a useful tool and, to our knowledge, was never applied in a data assimilation context.
Overall, we found that guaranteeing SPSD of the constructed localization matrix results in minor changes in the raw EOL estimate (see Fig. 1&2 in the Appendix). The conclusions of the error reduction are unaffected when using an SPSD localization. For this reason, we mainly present "raw" EOL results that are optimal in terms of sampling error reduction without any additional constraints.

The revised version of the manuscript includes additional results, which we summarize in a newly added section. The new section analyzes changes due to enforcing SPSD by discussing Figure 1 (see Appendix/Supplement). We also adapted the theory and conclusion sections, which now discuss the SPSD characteristic of covariances and localization in more depth.

I make three additional remarks

- It would be appropriate to mention explicitly that the matrix Ploc defined by the Schur product (3) is SPSD for any SPSD matrix P if, and only if, the localization matrix C is itself SPSD (I cannot mention an explicit reference for that, although I know there exist some, but I can provide you with a proof if necessary).

Answer: Done. We now reference the Schur product theorem and adapted the theory section accordingly.

- You write As some DA algorithms require SPSD covariance matrices (l. 237 of your Authors'.Tracked.changes file) and covariances or localization matrices often need to be symmetric positive semi-definite (l. 465). Well, covariance matrices are SPSD by definition. As for DA algorithms, all algorithms that are based, explicitly or implicitly, on a Bayesian approach (and that is the case of Kalman Filter and of all its variants, particularly ensemble variants, in which finite samples describe probability distributions) require proper covariance matrices (if that condition is not verified, the corresponding computations are meaningless and can lead to absurd results, as has actually been observed by many authors).

Answer: Indeed, this sentence was not well phrased, as all covariance matrices should be symmetric positive semi-definite. We adopted the conclusion based on our latest results. Overall, we agree that proper mathematical covariances are a crucial ingredient for data assimilation. However, we struggled to find literature that describes how the negative definiteness of covariance matrices or localization affects the estimation process. To avoid speculation, we decided not to discuss the effect of definiteness on the estimation process.

- Localization is necessary in the various forms of Ensemble Kalman Filter because of the relatively small size of the ensembles that are used. It might be useful, if this is possible, to give some quantitative information on how the need for localization decreases when the ensemble size increases.

Answer: Thank you for this comment. We have already thought about studying the effect of ensemble size on localization and consider to do so in the future. However, such an analysis is beyond the scope of the present study as it comprises a new question.

I am going to ask you to make a major revision of your paper along the lines above. I will submit your new revised version to further review.

References: *Nicholas J. Higham, Computing the nearest correlation matrix—a problem from finance, IMA Journal of Numerical Analysis, Volume 22, Issue 3, July 2002, Pages 329–343, https://doi.org/10.1093/imanum/22.3.329*

Appendix/Supplement

Figure 1: Examples of EOL-based localization matrices for a single vertical column: (a) matrix C constructed based on the \textit{SINGLE} case, (b) resulting nearest correlation matrix C following the NCM algorithm, and (c) changes due to enforcing positive definiteness. (This figure is now included in the manuscript as Fig. 10.)

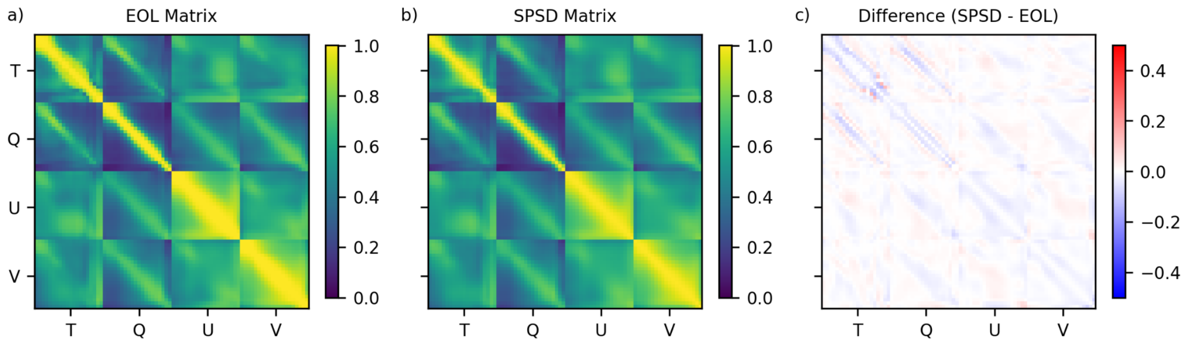

Figure 2: Similar data/analysis as in Figure 1 but snapshots of the matrices showing temperature and reference level 500hPa.

a) EOL b) SPSD EOL c) Differences