Thank you to our editor, Olivier Talagrand, for continuing to handle the review of our manuscript. We are also grateful for further thoughtful input from two reviewers. Our responses (in **bold**) to each review are below.

1. Eq. (1) (and elsewhere). The matrix $\mathbf{L}$ and $\mathbf{P}^b$ being symmetric, you might mention that the submatrices $\mathbf{L_{XY}}$ and $\mathbf{L_{YX}}$ on the one hand, and $\mathbf{P_{XY}}^b$ and $\mathbf{P_{YX}}^b$ on the other, are necessarily transpose of each other.

   We added several sentences to section 2.1 mentioning the relationship between $\mathbf{P_{XY}}^b$ and $\mathbf{P_{YX}}^b$ as well as $\mathbf{L_{XY}}$ and $\mathbf{L_{YX}}$.

2. Table 3 shows RMS errors in the assimilated fields. Is that necessary (the corresponding results are shown on Fig. 3, and the same information is not included in Tables B1 and B2) ?

   We have removed the median RMS errors from table 3.

3. Over how long periods of assimilation (or at which assimilation time) have the results shown in Figures 3 and 4 been obtained (the EnKF being a purely sequential algorithm, which may take some time to reach a stationary regime, that may be of some importance) ?

   We run each DA scheme for 3,000 analysis cycles, discarding the first 1,000 cycles and reporting statistics from the remaining 2,000 cycles. (See final paragraph of section 3.2.) Figures 3 and 4 show results from the final 2,000 cycles.

4. L. 275. The significance of the parameter σ_λ^2 is not clear.

   This parameter is not directly relevant to the study at hand. We included it in the manuscript for the sake of reproducibility. However, we have attempted to clarify the role of this parameter in the adaptive inflation scheme.

5. The final paragraph, which you have apparently introduced in response to one of my previous comments, is actually inappropriate Although we tested the localization functions in an EnKF our results should translate to EnVar schemes as well, because 3D-Var and the analysis step of the Kalman Filter are equivalent in the case of a linear observation operator (Daley, 1993). The positive semidefiniteness of the localization matrix is essential to ensure convergence of the numerical optimization methods used to implement EnVar (Bannister, 2008). The localization functions we have presented may be used in variational schemes without the need to numerical verify that the localization matrix is positive semidefinite each time a new set of localization radii is tested. Firstly, variational schemes (at least in their present form) do not build covariance matrices from discrete ensembles, and therefore do not need localization. Secondly the equivalence between variational and Kalman Filter algorithms requires linearity, not only of the observation operators, but also of the dynamical model. It therefore does not apply in the present case of the Lorenz model. Although that paragraph was introduced following one of my comments, I think it would be preferable that you remove it.

   We have removed this paragraph.

We thank S. G. Penny for his time in reading our manuscript and providing a thoughtful critique. His comments and our responses (**in bold**) are below.

I appreciate that the authors have expanded their experiments to include a more thorough assessment of the coupled data assimilation problem, including varying observation errors, observation coverage, and coupling strength. The updated results appear to provide a stronger case for the use of the multivariate Gaspari-Cohn localization.

1. "The note of S. Rasp is in reference to subgrid-scale parameterization with this model, so it is not directly relevant. "
   Certainly this was his focus application, but that does not negate his point of the highly linear relationships between the slow/fast variables in this model. I disagree that this dynamic is not directly relevant.

   We have now acknowledged the linear coupling in the bivariate L96 model in the conclusions.

2. "However, important couplings between atmosphere and ocean can be linear, e.g. their exchange of sensible heat, which is approximately linearly proportional to the temperature difference. "
   Perhaps this dynamic in the Lorenz model, and the potential application mentioned here, can be highlighted in the manuscript. This would give the reader a better understanding of the potential applicability of the results.

   We added two sentences to the conclusions which discuss linear coupling between atmosphere and ocean model components.

3. "the cross-assimilation decreases"
   I'm not sure what this means.

   We meant to say that the magnitude of the of the cross-domain analysis increment decreases as the strength of the dynamical coupling decreases.

4. "The coupling strength is h = 2 in the figure in the paper. The biggest change we saw is that the magnitude of the analysis errors in the unobserved X process increased with decreasing h. This is not surprising"
   But the errors also decreased in the observed variables as you reduced the coupling strength. Perhaps some attention should be given to the complete coupled state estimate rather than only the unobserved variables. It does appear that the overall RMSE of the coupled XY state may reduce with stronger coupling, but it might still be worth calculating and reporting.

   It is certainly interesting that varying the coupling strength impacts the magnitude of the analysis errors in both the observed and unobserved variables. However, the aim of this manuscript is to understand the impact of different multivariate localization functions, not different coupling strengths. The relative performance of the localization functions does not change across the range of coupling strengths we considered. Understanding the impact of coupling strength on the magnitude of analysis errors is an interesting area for further research and is outside of the scope of this manuscript.

5. "Multivariate Gaspari-Cohn still led to better performance than any of the other functions"
   Yes, it seems the case for the multivariate Gaspari-Cohn has been strengthened by the further experiments.

   We agree!

6. "We have included experiments observing only the "long" X process and the full coupled system."
   Lorenz uses the terminology "large" and "small" scales. He also designed the system to represent different timescales, spanning slower growing instabilities associated with planetary and synoptic scales, and the fast evolving mesoscale motions and convective clouds at smaller scales. I think it would be preferable to use one of these two terminologies rather than adding 'short' and 'long' as new descriptors for such an old and frequently studied model.

   We have changed the terminology to "large" and "small" as suggested.

7. Further, I would suggest reviewing a worthwhile analysis performed by Ginelli and collaborators that might provide additional insights: Carlu, Ginellie, Lucarini, Politi, 2019: Lyapunov analysis of multiscale dynamics: The slow bundle of the two-scale Lorenz '96 model. https://arxiv.org/pdf/1809.05065.pdf

It's a very interesting paper, but they consider a very different parameter regime. Specifically, they add a constant forcing to the small-scale equations, which leads to very interesting but qualitatively different dynamics.

8. "When we observed only the long process, all localization functions led to very similar performance (Fig. 4)." The errors might need to be scaled with some reference here. Using the absolute errors is less informative when working with different scales. Perhaps, for example, you could rescale the errors as a percentage of climatological variability.

   We have remade figures 3, 4, B2, and B3 with rescaled analysis errors.

9. "Observing both processes, at least in our configuration, was quite unstable and often led to filter divergence." This is concerning, and could point to a problem in the DA approach (perhaps because of the use of the stochastic EnKF? Could the presence of multiple scales make the system more sensitive to magnitude of random noise applied to different components?). It would be useful to understand why this is the case. Could there be some relationship to the imbalance of the effects on the observed versus unobserved variables as mentioned above?

   The EnKF can blow up even when implemented correctly. We have added a short discussion of this and three relevant references to section 3.4.3.

10. " filter performance is highly sensitive to the treatment of cross-domain background error covariances." Yes we have seen similar results in more complex models.

    This is reassuring.

11. "Thus, zeroing out the cross terms, as in weakly coupled schemes, may improve state estimates. On the other hand, inclusion of some cross-domain terms appears to be important for stability." It would be interesting to develop a strategy to approach this more rigorously.

    Agreed. This is an interesting area for further research.

12. "The BW method looks slightly improved, and the Askey method slightly degraded. This is true, however the difference is not statistically significant. " The dichotomy (significant vs. non-significant) is problematic since it is based on a somewhat arbitrary threshold for the p-value and can set misguided incentives in the evaluation and interpretation of a study. I suggest a review of the ASA's guidance on the use of the term, and taking care in discarding a result that may have some value. For example, from The American Statistician special issue on the Statistical Inference in the 21st Century:

    - Don't base your conclusions solely on whether an association or effect was found to be "statistically significant"
    - Don't believe that an association or effect exists just because it was statistically significant.
    - Don't believe that an association or effect is absent just because it was not statistically significant.
    - Don't believe that your p-value gives the probability that chance alone produced the observed association or effect or the probability that your test hypothesis is true.
    - Don't conclude anything about scientific or practical importance based on statistical significance (or lack thereof).

    Wasserstein et al., 2019: "Moving to a World Beyond "$p < 0.05$""
    https://www.tandfonline.com/doi/full/10.1080/00031305.2019.1583913

    We have removed the mentions of statistical significance from the manuscript.

13. "Could you clarify your interpretation of within-component localization?... They are both tapering a covariance between variables based on distance" Localization is needed across different variables, and when two different components of a system have different length scales, then it is not clear that a simple Euclidean distance metric is sufficient. Applying the localization in a more abstract space, or with an appropriately tailored distance metric, may be more appropriate.

    Agreed. This is a very interesting area for further research and one that we are interested in pursuing.

14. “A setup where the atmosphere influences the ocean state, but not vice versa would necessarily be associated with a background error covariance matrix which is not symmetric (which is not possible).”
    This might then indicate that this is not the best approach for the coupled data assimilation problem, e.g. if such more strongly one-way interactions are necessary for constraining coupled systems. Could this have any relevance to the failures of the fully observed cases?

    The background error covariance matrix must be symmetric. It is hard to see how to get around this.

Unfortunately we were not able to access any of the comments from Reviewer #2. We would be interested to see them and happy to respond to them if a new document could be uploaded with the comments included.