We thank Reviewer #2 for their time in reading our manuscript and providing a thoughtful critique. The reviewer's comments and our responses are below.

The paper by Stanley et al. develops multi-scale extensions to the traditional families of the parameterized localization functions such as the Gaspari-Cohn 5th order polynomial. A key contribution of the paper is to note that each of these polynomial expressions can be represented as a product of the square-root kernels that can be cross-multiplied to achieve a positive-definite cross-scale localization. Authors, correctly, draw relevance of these new techniques to the problem of cross scale localization encountered in coupled data assimilation. Authors correctly choose the right type of the test problem that is useful-enough to test the mathematics of the developed extensions yet is not too complex to obscure the interpretation of the results. I agree with authors that there is no need to over-interpret the results of this simple experiment with regard to its relevance in a more complex ocean-atmosphere problems.

I found this article very relevant, well-written, with adequate experimental plan, and appropriate interpretation of the experimental results. I congratulate the authors on a nice contribution to the literature and suggest this paper for publication after minor revision.

Summary of suggested changes:

1. This contribution parallels the work of Mark Buehner on multi-scale localization. I suggest that authors draw this parallel by referencing some of his work such as [https://doi.org/10.3402/tellusa.v67.28027]. By including these references, authors can then discuss how their work can also be related to the problem of multi-scale localization (e.g. such as assimilation in convective-resolving models).

   We added a paragraph to Section 2.1 discussing the similarity between our work and the work of Buehner and Shlyaeva.

2. I made some minor suggestion to how authors might consider changing or extending other references in the introduction section (see pdf attached).

   See responses to minor suggestions below.

3. I tried to read the paper from the perspective of someone who might want to implement some of the localization formulas discussed by the authors. I made some suggestions on clarifications. I highly appreciate that authors published the source code for their work. I suggest that authors mention that in the main body of the paper.

   We now mention that code is published in main body of paper.

Minor suggestions from attached pdf:

1. L32: "Penny et al., 2017" this is just a workshop report i dont think it has any results to support your statement. Maybe reference to paper by Sluka et al. 2016 or Frolov et al. 2016 will be more appropriate.

   Changed to suggested references.

2. L34: "Cross-domain correaltions . . . (Penny et al., 2017)" Is this the same as flow-dependent? If yes, then some reference to works by Polly Smith will be approbate. You can also reference the following paper https://doi.org/10.1175/MWR-D-20-0352.1.

   Changed reference to Smith et al. 2017 and Frolov et al. 2021

3. Sec 2.2: I realize that you came to this idea independently. However, similar ideas are also common in the field of multi-scale localization. See papers by Mark Buehner. It would be good to mention this parallel.

   Buehner and Shlyaeva (2015) indeed construct localization matrices that have similar form to the localization matrices we construct, i.e.

$$\mathbf{L} = \begin{bmatrix} \mathbf{L_{XX}} & \mathbf{L_{XY}} \\ \mathbf{L_{YX}} & \mathbf{L_{YY}} \end{bmatrix}. \tag{1}$$

   However their method differs from ours in that they do not construct analytic formulas. Rather, the construct the off-diagonal blocks of the L matrix through $\mathbf{L_{XY}} = (\mathbf{L_{XX}})^{1/2} (\mathbf{L_{YY}})^{T/2}$. This formulation can be related to a discrete approximation to kernel convolution (a perspective that we are developing), but this was not explored by Buehner and Shlyaeva (2015). Constructing the off-diagonal blocks through matrix square roots is appropriate for scale-dependent localization where $\mathbf{L_{XX}}$ and $\mathbf{L_{YY}}$ are of the same

size. It is not immediately obvious how to extend this to multiple spatial domains, as we have in strongly coupled data assimilation. We have mentioned the parallel to multi-scale localization and cited Buehner and Shlyaeva (2015) in Section 2.1 where we introduce multivariate localization.

4. Sec 2.3: You can point out the similarity and differences of your results from Frolov et.al. 2016. Similar to Frolov et.al. 2016, the localization scale for cross-scale localization is the average of the scales in the coupled.

   Added reference to Frolov et al. (2016) and commented on the cross-localization radius they used.

5. Fig 1: please add a comment to the figure why Lx,x seem to drop to zero faster for Wendland than for any other function. Otherwise it looks like you miss-specified the localization radius for Wendland.

   Wendland uses the same $R_{XX} = 45$ value as the other functions. You are correct in observing that it drops to near 0 more quickly than the others, however it does not reach 0 until $d = 45$. We added a sentence to the caption stating the localization radii for all functions.

6. L145: what does κ^{-3} stands for? I am not sure i know how to compute that.

   Added sentence clarifying that $\kappa^2 = \frac{\max\{R_{XX}, R_{YY}\}}{\min\{R_{XX}, R_{YY}\}}$ is a ratio of length scales.

7. Eq. 9: seems like something went wrong with typesetting here.

   Fixed the issue with the '.' appearing in awkwardly at the end of the equation.

8. Eq. 13: I am really confronted by the complexity of this formula and what would it take to implement it in practice. Not sure I follow how to specify ν or γ. Maybe this is something you can explicitly mention after Eq. 13.

   The parameters ν and $\{\gamma_{ij}\}$ are related to the shape of the localization functions (Eq. 10 & 11), and are necessary to guarantee positive semidefiniteness in a given spatial dimension. In general it makes sense to take ν as small as possible while still guaranteeing positive semidefiniteness in the space (for us the space is $\mathbb{R}^3$) and then allow γ_{ij} to vary so that each component localization function can have a different shape. Once ν and $\{\gamma_{ij}\}$ are chosen, Eq. (13) gives the formula for the maximum cross-localization weight factor (i.e. height of the cross-localization function at distance $d = 0$) to guarantee positive semidefiniteness. We added a sentence to section 2.5 describing the parameters ν and γ_{ij}.