We thank Reviewer #3 for their time in reading our manuscript and providing a thoughtful critique. The reviewer's comments and our responses are below.

General Comments:

This manuscript presents a new multivariate extension of the standard Gaspari-Cohn localization function which is compared with 3 other multivariate functions, plus their univariate versions. These techniques are extremely relevant to problem of cross-domain localization in strongly coupled data assimilation and this work is an encouraging step towards developing appropriate methods for such systems. The localization techniques are illustrated using the bivariate Lorenz 96 system.

Whilst it would be nice to have an example illustrating how the new multivariate GC method translates to a more realistic system I appreciate that it is always important to test new ideas like this in a relatively simple system, where results can be more easily interpreted. I hope that the authors have the opportunity to extend this work to a more complex coupled model system in the future.

It is always good to see new work addressing issues related to the application of coupled DA. The article is timely, highly relevant and clearly written; it will make a nice contribution to the coupled DA literature. I suggest it is published after minor revisions.

We are actively working on extending this work to a coupled atmosphere-ocean model.

Specific comments:

1. It is a shame that results were only shown for case where the fast (Y) component is fully observed, and further that the performance of each method was only measured/ reported in terms of the RMSE of the X (slow, unobserved) component. I would like to see some results from experiments where only the X (slow) component is observed, and also where both the X and Y components are observed, both fully and partially. I appreciate that this would potentially increase the number of figures/length of the manuscript, but it may not be necessary to explicitly show all the results. A brief discussion of the results in order to confirm that the general conclusions still hold under different observing scenarios would give the reader greater confidence in the performance of the new GC method.

   We have included experiments observing only the "long" X process and the full coupled system. When we observed only the long process, all localization functions led to very similar performance (Fig. 1). Note that since weak coupling is stable in this configuration we have included results from weakly coupled runs. In the paper we showed only the performance for univariate, multivariate, and weakly coupled Gaspari-Cohn and stated that all other localization functions performed similarly.

   Observing both processes, at least in our configuration, was quite unstable and often led to filter divergence. About 80% of the trials with weakly coupled localization functions led to catastrophic filter divergence. Trials with univariate and multivariate localization diverged less often, but still diverged about 20% of the time. However, in this configuration the precise shape of the localization function appears to have little impact. We did see differences between univariate, weakly coupled, and multivariate localization functions. Figure 2 shows results from only the trials (out of 50 total) which did not diverge. Weakly coupled localization appears to lead to the best performance, when the filter does not diverge. There is some variation in the results across the different localization functions. In particular, multivariate Askey appears to lead to better performance than weakly coupled Askey, but this may be attributable to the issues with stability. The complicated story with the weakly coupled schemes indicates that, in this configuration, filter performance is highly sensitive to the treatment of cross-domain background error covariances. The small ensemble size combined with small true forecast error cross-correlations can lead to spuriously large estimates of background error cross-covariances. Meanwhile, we have nearly complete observations of both processes, so within-component updates are likely quite good. 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.

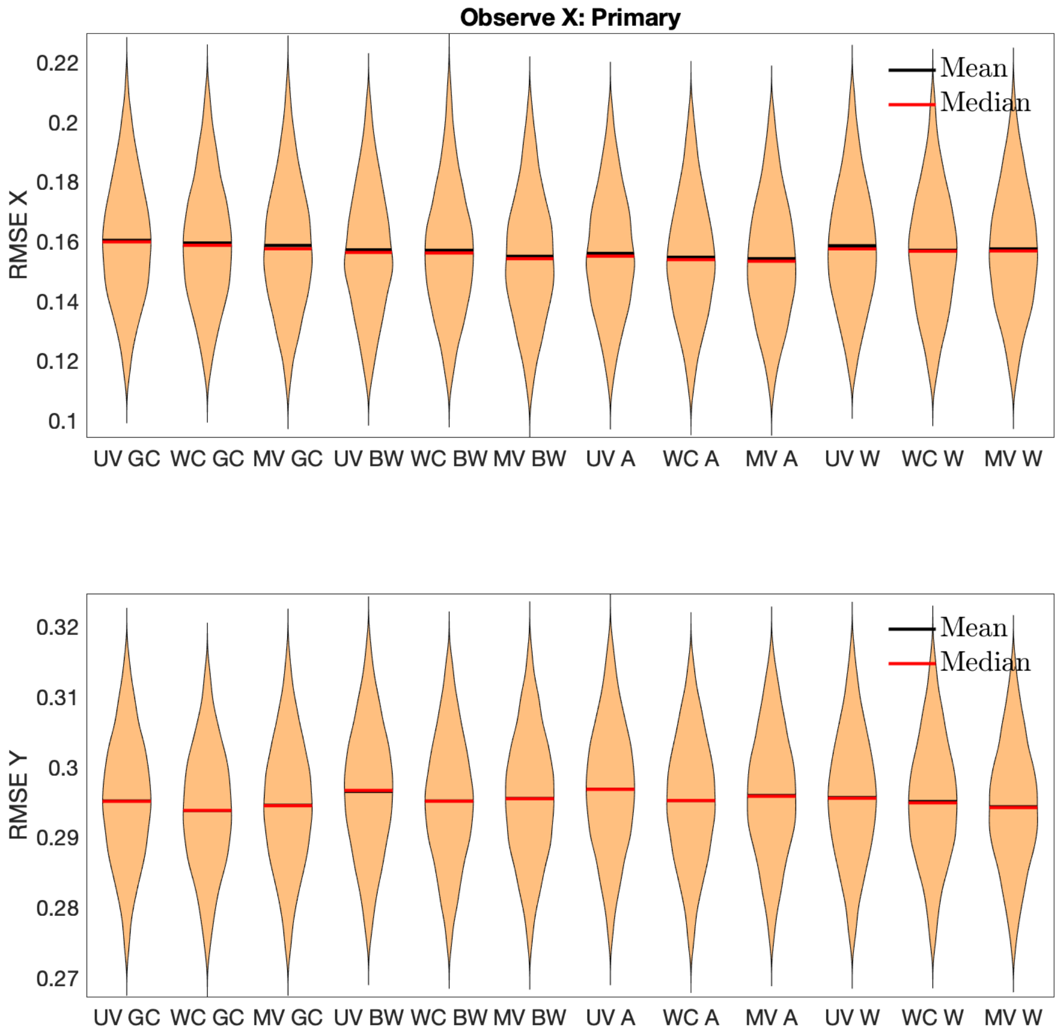

Figure 1: Observe only the X process. UV stands for univariate, WC stands for weakly coupled, MV stands for multivariate.

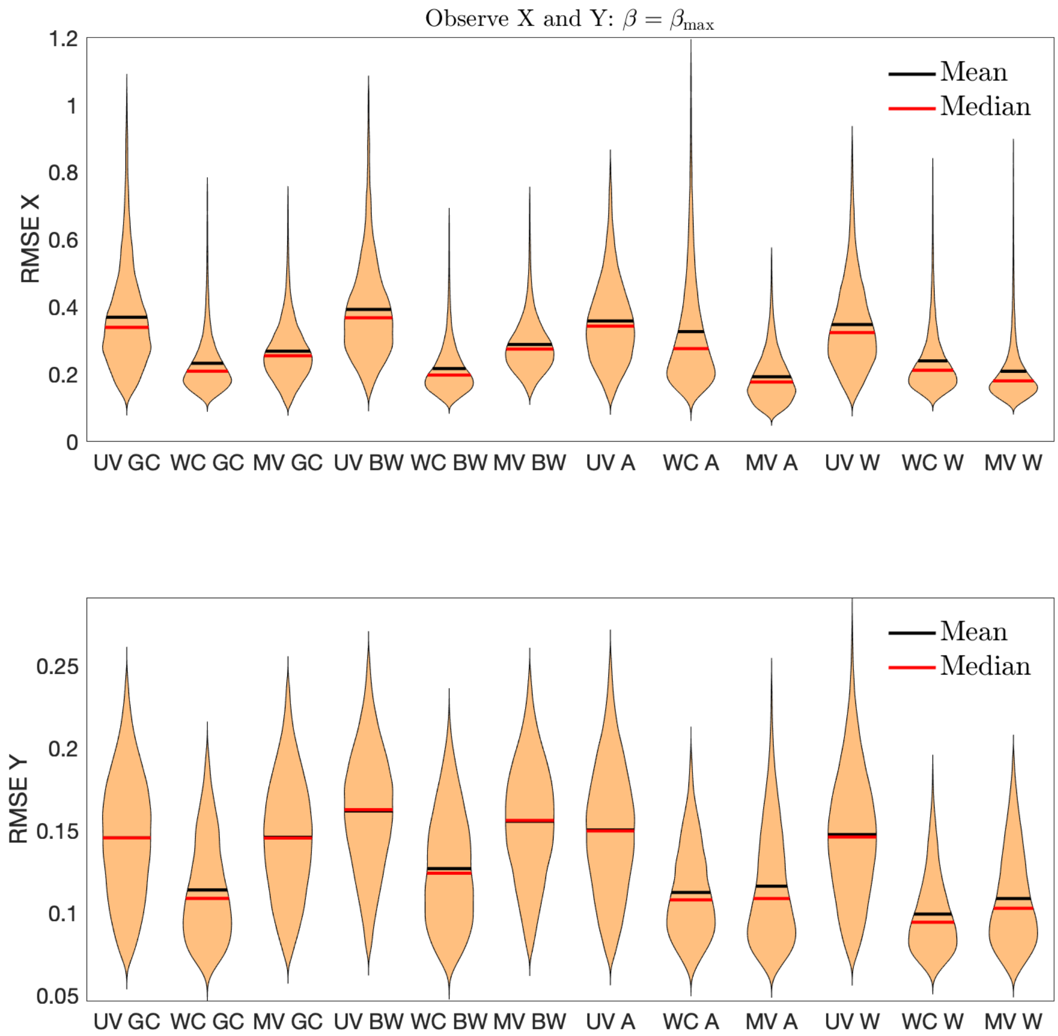

Figure 2: Observe both processes.

2. I am not entirely clear on how the univariate localization functions were implemented. Lines 181-182 state: ”We compare the four multivariate localization functions in Sect. 2 to a simple approach to localization in coupled DA, which is to use the same localization function for all model components. We call this approach univariate localization.” I think this means that each block of the localization matrix L uses the same localization function and radius for all blocks, rather than a different radius for the X and Y blocks and a different function and radius for the cross X,Y block, is this correct? I think what is confusing is that you are calling it univariate localization but you are actually localizing the cross XY blocks of the matrix B. Perhaps this needs to be stated more explicitly somewhere. In systems with very different error correlation scales this type of univariate localization function could be not really be expected to perform well.

   Your interpretation is correct. We have changed the introduction to univariate localization (first paragraph of section 3) so that it is now explicitly stated that the same function is used to localize all matrix blocks.

Minor comments:

1. The references are a bit strange – there are multiple web links for a lot of the papers; the https://doi.org/xxx link will be sufficient in most cases.

   References have been fixed and web links have been removed.

2. Further minor comments and technical corrections are marked in the attached pdf.

   See list below for changes made based on suggestions in pdf.

Minor comments from attached pdf:

1. L27: “non-negative definite” positive semi-definite is the more common terminology

   Changed to positive semidefinite.

2. L35: add reference to Smith et al 2017, https://doi.org/10.1175/MWR-D-16-0284.1

   Added reference.

3. L53: “EnVar” Ensemble-Variational - You need to introduce this type of abbreviation the first time you use it.

   Changed.

4. L107: “$R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$” is this choice required for the matrix L to be positive semi-definite?

   For Gaspari-Cohn, Bolin-Wallin, and other functions created through kernel convolution, there R_{XY} is completely determined by R_{XX} and R_{YY} and is equal to $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$. We have changed our discussion in Sec. 2.2 to clarify this point.

5. Fig 1: Does Wendland use the same R_x value? Maybe add values of R_x, R_Y and R_{XY} in the caption

   Wendland does use 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 localiztaion radii for all functions.

6. Eq. 10: what are ν, $\gamma_{i,j}$? How are they chosen/ defined?

   The parameters ν and $\{\gamma_{ij}\}$ are related to the shape of the localization functions, and are necessary to guarantee positive definiteness in a given dimension. A sentence has been added to describing these parameters.

7. L166: “B the beta function” This needs to be defined or an appropriate reference added. Also, this isn’t the greatest choice of notation given that the background error covariance matrix is denoted B

   We defined the beta function and changed the notation for the background error covariance matrix from B to $\mathbf{P^b}$.

8. L174: "we choose $R_{XY} = \min\{R_{XX}, R_{YY}\}$" has this value been experimented with?

   To maintain positive semidefiniteness we require $R_{XY} \leq \min\{R_{XX}, R_{YY}\}$. From our understanding of the true forecast error correlations we estimate that for optimal performance R_{XY} should be at least as large as R_{YY} (which is the smaller of the two within-component localization radii). For this reason, we have not experimented with smaller values for R_{XY}.

9. L181: grammar?

   This section has been reworded.

10. L184: "Additionally in our setup we observe only one of the two processes and we find that when the assimilation is not allowed to update the unobserved process the result is prone to catastrophic divergence." This is interesting ... it would be good to understand why this occurs. Does it depend on the temporal/ spatial frequency of the obs?

    We have yet to find a stable configuration where we observe only one process and do not allow any updates to the unobserved process. It is possible that we could increase the coupling strength enough that the unobserved process would synchronize with the observed process, but short of that it is difficult to imagine a stable configuration when one process is entirely detached from observations.

11. L200: "so there are 10 times as many long variables as short variables" 10 times more short variables than long variables

    Changed.

12. L204: how does .005 translate to 36 minutes?

    We rewrote this sentence to clarify that this is a result from Lorenz (1996). The reasoning used in Lorenz (1996) is based on relative error growth rates. A full discussion of the reasoning from Lorenz (1996) is superfluous to the discussion here and is hence not included.

13. L213: "We choose to place the variable X_k in the middle of the sector" is this choice important to the performance of the localisation?

    Placing X_k at the beginning of the sector, as in Roh et al. (2015), means that half of the nearby Y variables are nearly uncorrelated with X_k. The analysis errors are larger when X_k is placed at the beginning of the sector rather than the middle of the sector. However, the relative performance of all the localization functions is the same in both cases. Figure 3 shows the distribution of analysis errors when we observe the "short" process with X_k placed at the beginning of the sector.

14. L217: This paragraph is clunky - needs revising

    This paragraph has been rewritten following this suggestion and a suggestion from S. Penny.

15. L220: observation operator?

    The changed paragraph now says "linear observations".

16. L221: delete "ensemble of"

    Deleted in rewrite of paragraph.

17. L234: How frequent in time are the obs?

    In the original experiment in the first submission, we observe every 0.005 model time units.

18. L260: "Figure 3 shows RMSE for process X" what about Y? How is RMSE computed? vs a 'truth'?

    Yes that is correct. RMSE is computed vs. a 'truth', which is also used to generate the 'observations'. We added a sentence clarifying this.

19. Eq. A15 ?

    Could the reviewer clarify what about this expression is confusing?

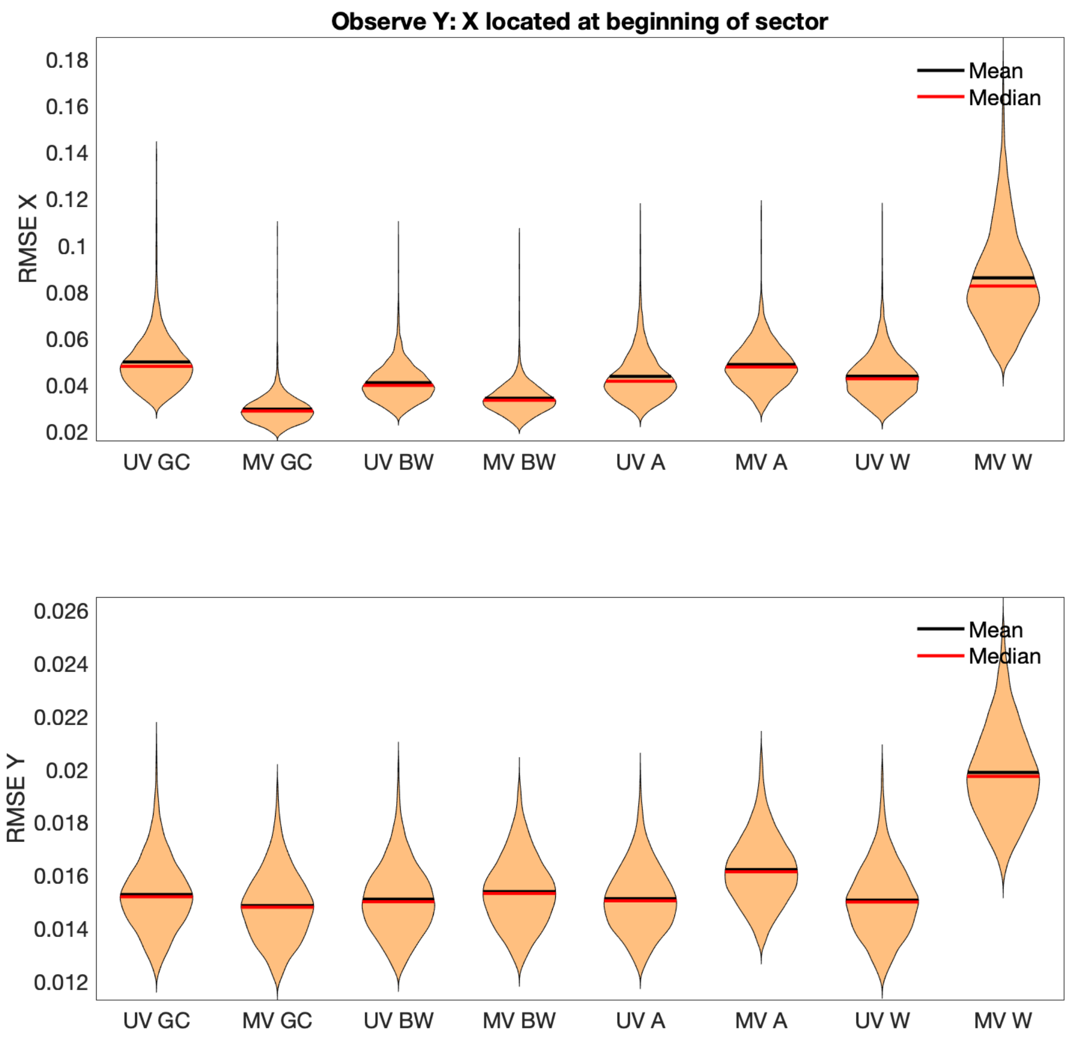

Figure 3: Move the location of X_k to the beginning of the sector, instead of the middle of the sector.

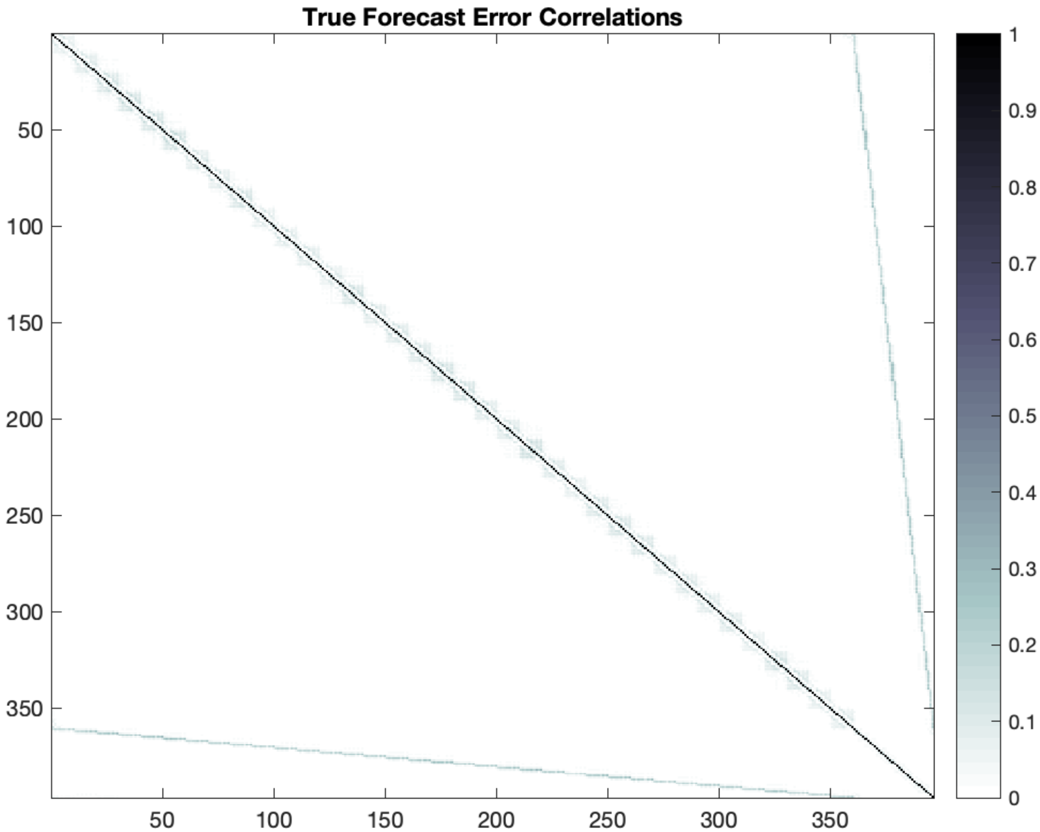

Figure 4: True forecast error correlation matrix. Upper left block shows correlations between the "short" Y variables. Lower right blocks shows correlations between the "long" X variables.

20. L352: how is κ defined for this case?

    κ is defined as $\kappa^2 = \frac{\max\{R_{ii},R_{jj}\}}{\min\{R_{ii},R_{jj}\}}$. We added this clarification to the manuscript.

21. Fig B1: Have you plotted the whole correlation matrix? Or even the individual blocks? It may be easier to see the correlation structures, particularly for the cross-correlations

    We have plotted the whole correlation matrix, see Fig. 4. We find it difficult to estimate the correlation length scales from the full matrix because the "short" and "long" variables are on different grids and these grids do not show up clearly in the matrix. For example, looking just at this matrix one might reasonably conclude that the correlation length scale is longer for the Y process (upper left block) than the X process (lower right block), when in fact the opposite is true.

22. L390: "$\gamma_{XX}, \gamma_{YY}, \gamma_{XY}$, and ν" I don't understand what these parameters represent.

    See response to #6. The parameters ν and $\{\gamma_{ij}\}$ are related to the shape of the localization functions, and are necessary to guarantee positive definiteness in a given dimension. A sentence has been added to section 2.5 describing these parameters.

23. Fig B3: Are these showing the RMSE for X?

    Yes, however the section describing the estimation of the localization parameters has been rewritten and this figure has been removed. We felt that this figure was unclear and causing confusion.