Multivariate localization functions for strongly coupled data assimilation in the bivariate Lorenz ’96 system

Zofia Stanley, Ian Grooms, and William Kleiber

Department of Applied Mathematics, University of Colorado, Boulder, Colorado

Correspondence: Zofia Stanley (zofia.stanley@colorado.edu)

Abstract. Localization is widely used in data assimilation schemes to mitigate the impact of sampling errors on ensemble-derived background error covariance matrices. Strongly coupled data assimilation allows observations in one component of a coupled model to directly impact another component through inclusion of cross-domain terms in the background error covariance matrix. When different components have disparate dominant spatial scales, localization between model domains must properly account for the multiple length scales at play. In this work we develop two new multivariate localization functions, one of which is a multivariate extension of the fifth-order piecewise rational Gaspari-Cohn localization function; the within-component localization functions are standard Gaspari-Cohn with different localization radii while the cross-localization function is newly constructed. The functions produce positive semidefinite localization matrices, which are suitable for use in both Kalman Filters and variational data assimilation schemes. We compare the performance of our two new multivariate localization functions to two other multivariate localization functions and to the univariate and weakly coupled analogs of all four functions in a simple experiment with the bivariate Lorenz ’96 system. In our ~~experiment~~ experiments the multivariate Gaspari-Cohn function leads to better performance than any of the other multivariate localization functions.

1 Introduction

An essential part of any data assimilation (DA) method is the estimation of the background error covariance matrix $\mathbf{P}^{\mathrm{b}}$. The background error covariance statistics stored in $\mathbf{P}^{\mathrm{b}}$ provide a structure function that determines how observed quantities affect the model state variables, which is of particular importance when the state space is not fully observed (Bannister, 2008). A poorly designed $\mathbf{P}^{\mathrm{b}}$ matrix may lead to an analysis estimate, after the assimilation of observations, that is worse than the prior state estimate (Morss and Emanuel, 2002). In ensemble DA schemes the $\mathbf{P}^{\mathrm{b}}$ matrix is estimated through an ensemble average. Using an ensemble to estimate $\mathbf{P}^{\mathrm{b}}$ allows the estimates of the background error statistics to change with the model state, which is desirable in many geophysical systems (Smith et al., 2017; Frolov et al., 2021). However, this estimate of $\mathbf{P}^{\mathrm{b}}$ will always include noise due to sampling errors because the ensemble size is finite. In practice, ensemble size is limited by computational resources and hence sampling errors can be substantial. The standard practice to mitigate the impact of these errors is localization. A number of different localization methods exist in the DA literature (e.g Gaspari and Cohn, 1999; Houtekamer and Mitchell, 2001; Bishop and Hodyss, 2007; Anderson, 2012; Ménétrier et al., 2015). In this study we concentrate on distance-based localization. Distanced-based localization uses physical distance as a proxy for correlation strength and sets correlations to zero when the distance between the variables in question is sufficiently large. Localization is typically incorporated into the data assimilation in one of two ways - either through the $\mathbf{P}^{\mathrm{b}}$ matrix or through the observation error covariance $\mathbf{R}$ (Greybush et al., 2011). We are focusing on Schur (or elementwise) product localization applied directly to the $\mathbf{P}^{\mathrm{b}}$ matrix. The Schur product theorem (Horn and Johnson, 2012, Theorem 7.5.3) guarantees that if the localization matrix is positive semidefinite, then the localized estimate of $\mathbf{P}^{\mathrm{b}}$ is also positive semidefinite. Positive semidefiniteness of estimates of $\mathbf{P}^{\mathrm{b}}$ is essential for the convergence of variational schemes and interpretability of schemes like the Kalman Filter which are intended at minimizing the statistical variance of the estimation error.

The localization functions presented in this work are suitable for use in coupled DA, where two or more interacting large-scale model components are assimilated in one unified framework. Coupled DA is widely recognized as a burgeoning and vital field of study. In Earth system modeling in particular, coupled DA shows improvements over single domain analyses (Sluka et al., 2016; Penny et al., 2019). However, coupled DA systems face unique challenges as they involve simultaneous analysis of multiple domains spanning different spatiotemporal scales. Cross-domain error correlations in particular are found to be spatially inhomogeneous ~~(Smith et al., 2017)~~(Smith et al., 2017; Frolov et al., 2021). Schemes that include cross-domain error correlations in the $\mathbf{P}^{\mathrm{b}}$ matrix are broadly classified as strongly coupled, which is distinguished from weakly coupled schemes where $\mathbf{P}^{\mathrm{b}}$ does not include any nonzero cross-domain error correlations. The inclusion of cross-domain correlations in $\mathbf{P}^{\mathrm{b}}$ offers advantages, particularly when one model domain is more densely observed than another (Smith et al., 2020). Strongly coupled DA requires careful treatment of cross-domain correlations with special attention to the different correlation length scales of the different model components. Previous studies, discussed below, show that appropriate localization schemes are vital to the success of strongly coupled DA.

As in single domain DA, there is a broad suite of localization schemes proposed for use in strongly coupled DA. Lu et al. (2015) artificially deflate cross-domain correlations with a tunable parameter. Yoshida and Kalnay (2018) use an offline method, called correlation-cutoff, to determine which observations to assimilate into which model variables and the associated localization weights. The distance-based multivariate localization functions developed in Roh et al. (2015) allow for different localization functions for each component and are positive ~~semidefinite~~definite, but require a single localization scale across all components. Other distance-based localization schemes allow for different localization length scales for each component, but are not necessarily positive semidefinite (Frolov et al., 2016; Smith et al., 2018; Shen et al., 2018). Frolov et al. (2016) report that their proposed localization matrix is experimentally positive semidefinite for some length scales. Smith et al. (2018) use a similar method and find cases in which their localization matrix is not positive semidefinite.

In this work, we build on these methods and investigate distance-based, multivariate, positive semidefinite localization functions and their use in strongly coupled DA schemes. We introduce a new multivariate extension of the popular fifth-order piecewise rational localization function of Gaspari and Cohn (1999) (hereafter GC). This function is positive semidefinite for all length scales and hence appropriate for Ensemble-Variational (EnVar) schemes. We compare this to another newly developed multivariate localization function that extends Bolin and Wallin (2016), and to two other functions from the literature (Daley et al., 2015). We investigate the behavior of these functions in a simple bivariate model proposed by Lorenz (1996). In particular, we look at the impact of variable localization on the cross-domain localization function. We show that these functions are compatible with variable localization schemes of Lu et al. (2015); Yoshida and Kalnay (2018). We find that, in ~~our set up~~some setups, artificially decreasing the magnitude of the cross-domain correlation hinders the assimilation of observations, while in other setups the best performance come when there are no cross-domain updates. We compare all of the multivariate functions to their univariate and weakly coupled analogs and observe that the new multivariate extension of GC outperforms all multivariate competitors.

This paper is organized as follows. In Sect. 2 we present two new multivariate localization functions and two multivariate localization functions from the literature. In Sect. 3 we describe ~~the set up for the experiment whose results are presented in Sect. 3.1~~experiments with the bivariate Lorenz 96 model. We conclude in section 4.

2 Multivariate localization functions

2.1 Multivariate localization background

Consider the background error covariance matrix $\mathbf{P}^{\mathrm{b}}$ of a strongly coupled DA scheme with interacting model components X and Y. The $\mathbf{P}^{\mathrm{b}}$ matrix may be written in terms of within-component background error covariances for components X and Y ($\mathbf{P}^{\mathrm{b}}_{\mathbf{XX}}$ and $\mathbf{P}^{\mathrm{b}}_{\mathbf{YY}}$) and cross-domain covariances between X and Y ($\mathbf{P}^{\mathrm{b}}_{\mathbf{XY}}$ and $\mathbf{P}^{\mathrm{b}}_{\mathbf{YX}}$). Here $\mathbf{P}^{\mathrm{b}}_{\mathbf{XY}}$ controls the effect of system X on Y and vice versa for $\mathbf{P}^{\mathrm{b}}_{\mathbf{YX}}$. Strongly coupled DA is characterized by the inclusion of nonzero cross-domain covariances in $\mathbf{P}^{\mathrm{b}}_{\mathbf{XY}}$ and $\mathbf{P}^{\mathrm{b}}_{\mathbf{YX}}$. Similar to $\mathbf{P}^{\mathrm{b}}$, the localization matrix $\mathbf{L}$ may be decomposed into a 2×2 block matrix so that the localized estimate of the background covariance matrix is given by

$$\mathbf{L}\circ\mathbf{P}^{\mathrm{b}} = \begin{bmatrix} \mathbf{L}_{\mathbf{XX}} & \mathbf{L}_{\mathbf{XY}} \\ \mathbf{L}_{\mathbf{YX}} & \mathbf{L}_{\mathbf{YY}} \end{bmatrix} \circ \begin{bmatrix} \mathbf{P}^{\mathrm{b}}_{\mathbf{XX}} & \mathbf{P}^{\mathrm{b}}_{\mathbf{XY}} \\ \mathbf{P}^{\mathrm{b}}_{\mathbf{YX}} & \mathbf{P}^{\mathrm{b}}_{\mathbf{YY}} \end{bmatrix}, \tag{1}$$

where $\circ$ denotes a Schur product. In distance-based localization, the elements in the $\mathbf{L}$ matrix are evaluated through a localization function $\mathcal{L}$ with a specified localization radius R, beyond which $\mathcal{L}$ is zero. For example, if ${\mathbf{P}^{\mathrm{b}}}_{ij}$ is the sample covariance $\mathrm{Cov}(\eta_i,\eta_j)$ where $\eta_i=\eta(\mathbf{s_i})$ denotes the background error in process X at spatial location $\mathbf{s_i}\in\mathbb{R}^n$, then the associated localization weight is $\mathbf{L}_{ij}=\mathcal{L}(d_{ij})$, where $d_{ij}=\|\mathbf{s_i}-\mathbf{s_j}\|$. Furthermore, if $d>R$ then $\mathcal{L}(d)=0$.

Often different model components will have different optimal localization radii and hence one may wish to use a different localization function for each model component (Ying et al., 2018). Let $\mathcal{L}_{XX}$ and the $\mathcal{L}_{YY}$ be the localization functions associated with model components X and Y respectively. A fundamental difficulty in localization for strongly coupled DA is how to propose a *cross-localization function* $\mathcal{L}_{XY}$ to populate both $\mathbf{L}_{\mathbf{XY}}$ and $\mathbf{L}_{\mathbf{YX}}$ such that whenever a block localization matrix $\mathbf{L}$ is formed through evaluation of $\{\mathcal{L}_{XX},\mathcal{L}_{YY},\mathcal{L}_{XY}\}$ then $\mathbf{L}$ is positive semidefinite. We call this collection of component functions a *multivariate positive semidefinite function* if it always produces a positive semidefinite $\mathbf{L}$ matrix (Genton and Kleiber, 2015). We refer to multivariate positive semidefinite functions as *multivariate localization functions* when they are used to localize background error covariance matrices. In this study we compare four different multivariate localization functions, including one that extends GC.

Similar block localization matrices are used in scale-dependent localization, where X and Y are not components, but rather a decomposition of spectral wavebands from a single process. Scale-dependent localization aims to use a different localization radius for each waveband, which leads to the same question of how to construct the between-scale localization blocks. Buehner and Shlyaeva (2015) constructed $\mathbf{L_{XX}}$ and $\mathbf{L_{YY}}$ through evaluation of localization functions with different radii. They then constructed the cross-localization matrix through $\mathbf{L_{XY}} = (\mathbf{L_{XX}})^{1/2}(\mathbf{L_{YY}})^{\mathrm{T}/2}$, with $\mathbf{L_{YX}}$ defined analogously. This is appropriate for scale-dependent localization where X and Y are defined on the same grid and hence $\mathbf{L_{XX}}$ and $\mathbf{L_{YY}}$ are of the same dimension. It is still an open question how to extend this construction to the strongly coupled application where different components are defined on different grids. The multivariate localization functions we construct below could also be used in scale-dependent localization.

In our comparison of multivariate localization functions, we investigate the impact of the shape parameters *cross-localization radius*, and *cross-localization weight factor*. The *cross-localization radius*, R_{XY}, is the smallest distance such that for all $d > R_{XY}$ we have $\mathcal{L}_{XY}(d) = 0$. For all of the functions in this study, the cross-localization radius is related to the within-component localization radii R_{XX} and R_{YY}. We define the *cross-localization weight factor*, $\beta \geq 0$, as the value of the cross-localization function at distance $d = 0$, i.e. $\beta := \mathcal{L}_{XY}(0)$. The cross-localization weight factor β is restricted to be less than or equal to 1 in order to ensure positive semidefiniteness (Genton and Kleiber, 2015) and smaller values of β lead to smaller analysis updates when updating the X model component using observations of Y, and vice versa. Each function we consider has a different maximum allowable cross-localization weight, which we denote $\beta_{\max}$. Values of β greater than $\beta_{\max}$ lead to functions that are not necessarily positive semidefinite, while values of β less than $\beta_{\max}$ are allowable and may be useful in attenuating undesirable correlations between blocks of variables (Lu et al., 2015). ~~However, we find that in our experimental setup the best performance comes when $\beta = \beta_{\max}$.~~

Note that while this example shows model space localization for a coupled model with two model components, the localization functions we develop and investigate may also be used in observation space localization, and can be extended to an arbitrary number of model components.

2.2 Kernel convolution

Localization functions created through kernel convolution, such as GC, may be extended to multivariate functions in the following straightforward manner. Suppose $\mathcal{L}_{XX}(d) = [k_X * k_X](\mathbf{d})$ and $\mathcal{L}_{YY}(d) = [k_Y * k_Y](\mathbf{d})$ where $\mathbf{d} \in \mathbb{R}^n$, $d = \|\mathbf{d}\|$, $(*)$ denotes convolution over $\mathbb{R}^n$, and k_X, k_Y are square integrable functions. For ease of notation let the kernels k_X and k_Y be normalized such that $\mathcal{L}_{XX}(0) = \mathcal{L}_{YY}(0) = 1$, which is appropriate for localization functions. Then the function $\mathcal{L}_{XY}(d) = [k_X * k_Y](\mathbf{d})$ is a compatible cross-localization function in the sense that, when taken together $\{\mathcal{L}_{XX}, \mathcal{L}_{YY}, \mathcal{L}_{XY}\}$ is a multivariate, positive semidefinite function.

As a proof, we define two processes Z_j, where j can represent either X or Y, as the convolution of the kernel k_j with a white noise field $\mathcal{W}$:

$$Z_j(\mathbf{s}) = \int_{\mathbb{R}^n} k_j(\mathbf{s} - \mathbf{t}) \mathrm{d}\mathcal{W}_{\mathbf{t}}. \tag{2}$$

It is straightforward to show that the localization functions we have defined are exactly the covariance functions for these two processes, $\mathcal{L}_{ij}(d) = \mathrm{Cov}(Z_i(\mathbf{s}), Z_j(\mathbf{t}))$, with $i,j = X,Y$, locations $\mathbf{s},\mathbf{t} \in \mathbb{R}^n$, and distance $d = \|\mathbf{s}-\mathbf{t}\|$. Thus $\{\mathcal{L}_{XX}, \mathcal{L}_{YY}, \mathcal{L}_{XY}\}$ is a multivariate covariance function, and hence a multivariate, positive semidefinite function (Genton and Kleiber, 2015).

For localization functions created through kernel convolution the localization radii are related to the kernel radii. Suppose the kernels k_X and k_Y have radii c_X and c_Y, i.e. $k_j(d) = 0$ for all $d > c_j$. The convolution of two kernels is zero at distances greater than the sum of the kernel radii. Thus the implied within-component localization radii are $R_{jj} = 2c_j$, for processes $j = X,Y$. Further, the implied cross-localization radius is the sum of the two kernel radii $R_{XY} = c_X + c_Y$. Equivalently, the cross-localization radius is the average of the two within-component localization radii, $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$, which is how we will write it going forward. Interestingly, this is exactly the cross-localization radius used in Frolov et al. (2016) and Smith et al. (2018).

Unlike *within-component* localization functions, *cross-localization* functions created through kernel convolution will always have $\mathcal{L}_{XY}(0) < 1$ whenever $k_X \not\equiv k_Y$. The maximum allowable cross-localization weight factor ($\beta := \mathcal{L}_{XY}(0)$) is exactly the value produced through the convolution, i.e. $\beta_{\max} = [k_X * k_Y](\mathbf{0})$. Smaller cross-localization weight factors also lead to positive semidefinite functions since if $\{\mathcal{L}_{XX}, \mathcal{L}_{YY}, \mathcal{L}_{XY}\}$ is a multivariate, positive semidefinite function, then so is $\{\mathcal{L}_{XX}, \mathcal{L}_{YY}, \beta\mathcal{L}_{XY}\}$ for $\beta < 1$ (Roh et al., 2015). To aid in comparisons to other cross-localization functions, we re-define kernel-based cross-localization functions as,

$$\mathcal{L}_{XY}(d) = \frac{\beta}{\beta_{\max}} [k_X * k_Y](\mathbf{d}) \tag{3}$$

with $\beta \le \beta_{\max}$. In this way $\mathcal{L}_{XY}(0) = \frac{\beta}{\beta_{\max}} [k_X * k_Y](\mathbf{0}) = \beta$ which is consistent with our definition of the cross-localization weight factor in the previous section. We will experiment with the impact of varying β, but must always ensure $\beta \le \beta_{\max}$ to maintain positive semidefiniteness.

For most kernels, closed form analytic expressions for the convolutions above are not available. In the following two sections we present two cases ~~below~~ where a closed form is available. The kernels used in these two cases are the tent function (GC) and the indicator function (Bolin-Wallin).

2.3 Multivariate Gaspari-Cohn

The standard univariate GC localization function is constructed through convolution over $\mathbb{R}^3$ of the kernel, $k(r) \propto \left(1 - \frac{r}{c}\right)_+$ with itself. Here we define $r = \|\mathbf{r}\|$ with $\mathbf{r} \in \mathbb{R}^3$ and $z_+ = \max\{z, 0\}$. The kernel has radius c and is normalized so that $\mathcal{L}(0) = [k * k](\mathbf{0}) = 1$. As discussed in the previous section, the localization radius, R, is related to the kernel radius through $R = 2c$. We develop a multivariate extension of this function through convolutions with two kernels,

$$k_j(r) \propto \left(1 - \frac{r}{c_j}\right)_+, \quad j = X,Y. \tag{4}$$

The resulting within-component localization functions $\mathcal{L}_{jj}^{(GC)}(d) = [k_j * k_j](\mathbf{d})$ are exactly equal to GC, Eq. (4.10) in Gaspari and Cohn (1999). The formula for the cross-localization function $\mathcal{L}_{XY}^{(GC)}(d) = [k_X * k_Y](\mathbf{d})$ is quite lengthy and is thus included in Appendix A.

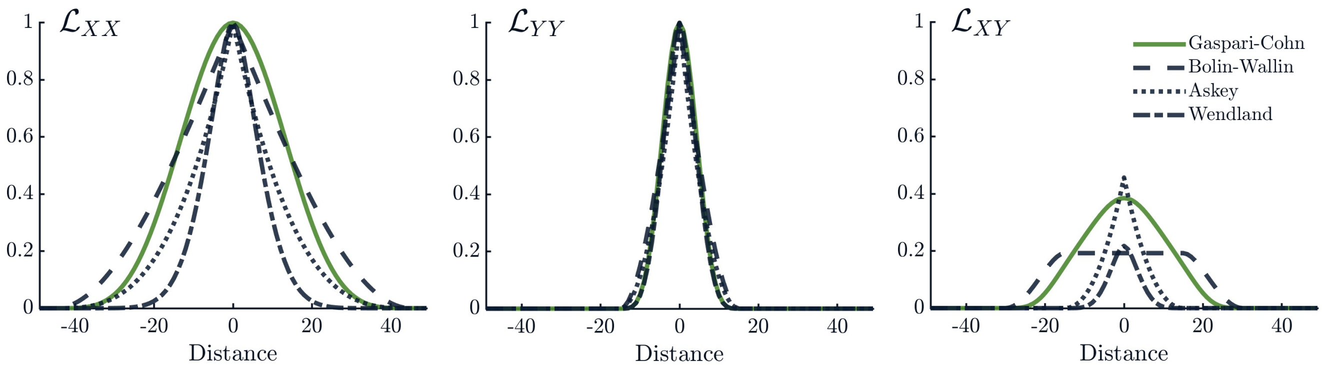


Figure 1. Four multivariate localization functions are shown in three panels. The first panel shows the function ~~$\mathcal{L}_{YY}$~~ $\mathcal{L}_{XX}$ used to localize the ~~short~~ long process, ~~Y~~ X. The second panel shows the function ~~$\mathcal{L}_{XX}$~~ $\mathcal{L}_{YY}$ used to localize the ~~long~~ short process, ~~X~~ Y. The third panel shows the cross-localization function $\mathcal{L}_{XY}$. In each panel, the color of the line shows the different multivariate functions: Gaspari-Cohn (~~blue~~green, solid), Bolin-Wallin (~~red~~dark, dashed), Askey (~~yellow~~dark, dotted), and Wendland (~~purple~~dark, dash-dot). In the case of univariate localization, the functions in the ~~left~~ middle panel are used to localize all processes. The within component localization radii are $R_{YY} = 15$ and $R_{XX} = 45$ for all functions. The cross-localization radii are $R_{XY} = 30$ for Gaspari-Cohn and Bolin-Wallin and $R_{XY} = 15$ for Askey and Wendland.

Recalling from the previous section that the maximum cross-localization weight factor is $\beta_{\max} = [k_X * k_Y](\mathbf{0})$, we find that for multivariate GC $\beta_{\max} = \frac{5}{2}\kappa^{-3} - \frac{3}{2}\kappa^{-5}$, where for convenience we define $\kappa^2 = \frac{\max\{R_{XX}, R_{YY}\}}{\min\{R_{XX}, R_{YY}\}}$ as a ratio of the within-component localization radii. As with all localization functions created through kernel convolution, the cross-localization radius is the average of the within-component localization radii, $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$. An example multivariate GC function with $R_{XX} = 45$, and $R_{YY} = 15$, and $\beta = \beta_{\max}$ is shown in Fig. 1. The multivariate GC localization function for three or more model components is discussed in Appendix A3.

2.4 Multivariate Bolin-Wallin

We derive our second multivariate localization function through convolution of normalized indicator functions over a sphere in $\mathbb{R}^3$. As in the previous section, the kernels are supported on spheres of radii c_X and c_Y,

$$k_j(r) = \frac{\sqrt{3}}{\sqrt{4\pi c_j^3}} I_{c_j}(r), \quad j = X, Y, \tag{5}$$

where $I_{c_j}(r)$ is an indicator function which is 1 if $r \le c_j$ and 0 otherwise. The resulting within-component localization function with localization radius $R_{jj} = 2c_j$ is

$$\mathcal{L}_{jj}^{(BW)}(d) = \left(\frac{1}{2R_{jj}^3}\right)(R_{jj} - d)^2 (2R_{jj} + d) \qquad \text{if } d \le R_{jj} \tag{6}$$

This is commonly referred to as the spherical covariance function. The label (BW) references Bolin and Wallin, who performed the convolutions necessary to create the associated cross-localization function in a work aimed at a different application of covariance functions (Bolin and Wallin, 2016). While Bolin and Wallin never developed multivariate covariance (or in our case localization) functions, the algebra is the same. We present only the localization functions that result from the convolution over $\mathbb{R}^3$, though similar functions for $\mathbb{R}^2$ and $\mathbb{R}^n$ are available in Bolin and Wallin (2016). Note that there is a typo in Bolin and Wallin (2016), which has been corrected below.

Let $c_X > c_Y$ be kernel radii, then the resulting cross-localization function is,

$$\mathcal{L}_{XY}^{(BW)}(d) = \frac{\beta}{\beta_{\max}} \cdot \frac{3}{4\pi\left(c_X c_Y\right)^{3/2}} \cdot \begin{cases} \frac{4\pi c_Y^3}{3} & \text{if } d < c_X - c_Y \\ V_3\left(c_X, \frac{d^2+c_X^2-c_Y^2}{2d}\right) + V_3\left(c_Y, \frac{d^2+c_Y^2-c_X^2}{2d}\right) & \text{if } c_X - c_Y \leq d < c_X + c_Y. \end{cases} \tag{7}$$

Here $V_3(r,x)$ denotes the volume of the spherical cap with triangular height x of a sphere with radius r, which is given by

$$V_3(r,x) = \begin{cases} \frac{\pi}{3}(r-x)^2(2r+x) & |x| < r \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

As with multivariate GC, it is convenient to define a ratio of within-component localization radii by $\kappa^2 = \frac{\max\{R_{XX}, R_{YY}\}}{\min\{R_{XX}, R_{YY}\}}$. Then we can write the maximum cross-localization weight factor as $\beta_{\max} = \kappa^{-3}$. The cross-localization radius for BW is $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$ because it is created through kernel convolution. An example multivariate BW function with $R_{XX} = 45$, $R_{YY} = 15$, and $\beta = \beta_{\max}$ is shown in Fig. 1.

2.5 Wendland-Gneiting functions

We compare the two functions of the preceding sections to the Wendland-Gneiting family of multivariate, compactly-supported, positive semidefinite functions. This family is not generated through kernel convolution, but rather through Montée and Descente operators (Gneiting, 2002). The simplest univariate function in this family is the the Askey function, which is given by

$$\mathcal{L}(d) = \left(1 - \frac{d}{R}\right)_+^{\nu} \tag{9}$$

with shape parameter ν and localization radius R. Other functions in this family are called Wendland functions. Several examples of univariate Wendland functions are displayed in Table 1.

Porcu et al. (2013) developed a multivariate version of the Askey function, where the exponent in ~~equation~~ Eq. (9) can be different for each process while the localization radius R is constant across all processes. Roh et al. (2015) found that this multivariate localization function outperforms common univariate localization methods when assimilating observations into the bivariate Lorenz 96 model. Daley et al. (2015) extended the work of Porcu et al. (2013) and constructed a multivariate version of general Wendland-Gneiting functions that allows for different localization radii for different processes. The multivariate

Original Wendland Functions
$\widetilde{\psi}_{3,1}(d) = (1-d)_+^4(4d+1)$
$\widetilde{\psi}_{4,2}(d) = \frac{1}{3}(1-d)_+^6\left(35d^2+18d+3\right)$
$\widetilde{\psi}_{5,3}(d) = (1-d)_+^8\left(32d^3+25d^2+8d+1\right)$
$\widetilde{\psi}_{6,4}(d) = \frac{1}{5}(1-d)_+^{10}\left(429d^4+450d^3+210d^2+50d+5\right)$

Table 1. Selected univariate Wendland functions

Askey function from Daley et al. (2015) has the form,

$$\mathcal{L}_{ij}^{(A)}(d) = \beta_{ij}\left(1-\frac{d}{R_{ij}}\right)_+^{\nu+\gamma_{ij}+1}, \quad \beta_{ij} = \begin{cases} 1 & i=j \\ \beta & i\neq j \end{cases}, \quad i,j = X,Y \tag{10}$$

The general form for multivariate Wendland functions is,

$$\mathcal{L}_{ij}^{(W)}(d) = \beta_{ij}\widetilde{\psi}_{\nu+\gamma_{ij}+1,k}\left(\frac{d}{R_{ij}}\right), \quad \beta_{ij} = \begin{cases} 1 & i=j \\ \beta & i\neq j \end{cases}, \quad i,j = X,Y \tag{11}$$

where $\widetilde{\psi}$ is defined as,

$$\widetilde{\psi}_{\nu+\gamma+1,k}(w) = \frac{1}{\mathrm{B}(2k+1,\nu+\gamma+1)}\int_w^1 \left(u^2-w^2\right)^k(1-u)^{\nu+\gamma}\mathrm{d}u \tag{12}$$

with B the beta function, $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\mathrm{d}t$. The parameters ν and $\{\gamma_{ij}\}$ are related to the shape of the localization functions, and are necessary to guarantee positive ~~definiteness~~ semidefiniteness in a given dimension. The parameter k determines the differentiability of the Wendland functions at lag zero (Gneiting, 2002). Note that the Askey function in ~~equation~~ Eq. (10) is a special case of the Wendland function (11) which corresponds to the case $k=0$. Daley et al. (2015) gave sufficient conditions on the parameters ν, k, R_{ij}, γ_{ij}, and β to guarantee that ~~equation~~ Eq. (11), and hence (10), is positive semidefinite. In particular for two processes ~~X, Y, equation~~ X and Y, Eq. (11) is positive semidefinite on $\mathbb{R}^n$ when $\nu \geq \frac{1}{2}(n+1)+k$, $\min\{R_{XX}, R_{YY}\} \geq R_{XY}$, $\gamma_{XY} \geq \frac{R_{XY}}{2}\left(\frac{\gamma_{XX}}{R_{XX}}+\frac{\gamma_{YY}}{R_{YY}}\right)$, and

$$\beta \leq \beta_{\max} := \sqrt{\left(\frac{R_{XY}^2}{R_{XX}R_{YY}}\right)^{\nu+2k+1}\frac{\mathrm{B}(\nu+2k+1,\gamma_{XY}+1)^2}{\mathrm{B}(\nu+2k+1,\gamma_{XX}+1)\mathrm{B}(\nu+2k+1,\gamma_{YY}+1)}}. \tag{13}$$

Going forward we consider the multivariate Askey function (10) and the multivariate Wendland function with $k=1$ in (11). Note that with both of these functions the cross-localization radius depends only on the smallest localization radius. In fact, we choose $R_{XY} = \min\{R_{XX}, R_{YY}\}$, although smaller values for R_{XY} also produce positive semidefinite functions. Thus for given R_{XX} and R_{YY}, the cross-localization radius for Askey and Wendland functions is always smaller than the cross-localization radius for GC and BW. With the choice $R_{XY} = \min\{R_{XX}, R_{YY}\}$, we see that $\beta_{\max}$ depends on the ratio

$\frac{\max\{R_{XX},R_{YY}\}}{\min\{R_{XX},R_{YY}\}}$, as in GC and BW. Examples of multivariate Askey and Wendland functions with $R_{XX} = 45$, $R_{YY} = 15$, $R_{XY} = 15$, and $\beta = \beta_{\max}$ are shown in Fig. 1. Important parameters for the four multivariate localization functions presented in this section are summarized in Table 2.

Function name	Maximum cross-localization weight factor, $\kappa^2 = \frac{\max\{R_{XX},R_{YY}\}}{\min\{R_{XX},R_{YY}\}}$	Cross-localization radius
Gaspari-Cohn	$\frac{5}{2}\kappa^{-3} - \frac{3}{2}\kappa^{-5}$	$\frac{1}{2}(R_{XX} + R_{YY})$
Bolin-Wallin	κ^{-3}	$\frac{1}{2}(R_{XX} + R_{YY})$
Askey	$\kappa^{-(\nu+1)}\sqrt{\frac{B(\nu+1,\gamma_{XY}+1)^2}{B(\nu+1,\gamma_{XX}+1)B(\nu+1,\gamma_{YY}+1)}}$	$\min\{R_{XX},R_{YY}\}$
Wendland	$\kappa^{-(\nu+2k+1)}\sqrt{\frac{B(\nu+2k+1,\gamma_{XY}+1)^2}{B(\nu+2k+1,\gamma_{XX}+1)B(\nu+2k+1,\gamma_{YY}+1)}}$	$\min\{R_{XX},R_{YY}\}$

Table 2. Summary of ~~localization~~ important shape parameters for four cross-localization functions.

3 ~~Experimental design~~Experiments

~~We compare the~~ In this section we investigate the performance of a data assimilation scheme using each of the four multivariate localization functions presented in Sect. 2~~to~~. We choose a setup which isolates the impact of the cross-localization functions and relate the filter performance to important cross-localization shape parameters. As a baseline for comparison, we also test two simple approaches to localization for coupled DA. The first method uses a single localization function and radius to localize all within- and cross-component blocks of the background error covariance matrix, i.e. $\mathcal{L}_{XX} \equiv \mathcal{L}_{YY} \equiv \mathcal{L}_{XY}$. We call this approach *univariate localization*. ~~The second simple approach is to use~~ In systems with very different optimal localization radii this type of univariate localization is likely to perform poorly, however it does provide a useful comparison point. The second approach uses different localization functions for each process and then ~~zero~~ zeroes out all cross-correlations between processes~~. This approach~~, i.e. $\mathcal{L}_{XX} \neq \mathcal{L}_{YY}$, and $\mathcal{L}_{XY} \equiv 0$. We call this approach *weakly coupled localization* as it leads to a ~~"weakly " coupled scheme, which is not the focus of this work. 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. Hence going forward we focus on the comparison between univariate and multivariate localization. In this section we outline the details of the model and assimilation scheme.~~ weakly coupled data assimilation scheme. All of the experiments are run with the bivariate Lorenz 96 model, which is described below (Lorenz, 1996).

3.1 Bivariate Lorenz model

The bivariate Lorenz 96 model is a ~~simple model of two coupled variables~~ conceptual model of atmospheric processes and is comprised of two coupled processes with distinct temporal and spatial scales. ~~This model is a conceptual model of atmospheric processes, where the~~ The "short" process can be thought of as rapidly-varying small-scale convective fluctuations ~~and~~ while the "long" process can be thought of as smooth large-scale waves. The model is periodic in the spatial domain, as a process on a fixed latitude band would be. ~~In the bivariate Lorenz 96 model, X is the~~

The "long" process~~with~~, X, has K distinct variables, X_k for $k = 1, \ldots, K$. The "short" process, Y, ~~has JK distinct variables, $Y_{j,k}$ for $j = 1, \ldots J, k = 1, \ldots, K$. The governing equations are,~~

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}\left(X_{k-2} - X_{k+1}\right) - X_k - (ha/b)\sum_{j=1}^{J} Y_{j,k} + F$$

$$\frac{\mathrm{d}Y_{j,k}}{\mathrm{d}t} = -abY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - aY_{j,k} + (ha/b)X_k.$$

~~The Y process~~ is divided into K sectors, with each sector corresponding to one "long" variable X_k. There are J "short" process variables in each sector, for a total of JK distinct Y variables, $Y_{j,k}$ for $j = 1, \ldots, J, k = 1, \ldots, K$. The Y variables, arranged in order, are $Y_{1,1}, Y_{2,1}, \ldots, Y_{J,1}, Y_{1,2}, Y_{2,2}, \ldots, Y_{J,K}$. Succinctly, $Y_{j-J,k} = Y_{j,k-1}$ and $Y_{j+J,k} = Y_{j,k+1}$, with periodicity conditions $Y_{j,k+K} = Y_{j,k-K} = Y_{j,k}$ for all j,k. The X process is also periodic with $X_{k+K} = X_{k-K} = X_k$ for all k.

We represent the variables on a circle where the arc length between neighboring Y variables is 1. Equivalently, the radius of the circle is $r = \frac{JK}{2\pi}$. Variable $Y_{j,k}$ is located at $(r\cos(\theta_{j,k}), r\sin(\theta_{j,k}))$ where $\theta_{j,k} = \frac{2\pi}{JK}(J(k-1)+j)$. We choose to place the variable X_k, located at $(r\cos(\phi_k), r\sin(\phi_k))$, in the middle of the sector whose variables are coupled to it, e.g. if $J = 10$ then X_k is halfway in between $Y_{5,k}$ and $Y_{6,k}$ and $\phi_k = \frac{2\pi}{10K}(10(k-1)+5.5)$. The placement of these variables is illustrated in Fig. 2. The chord distance between any two variables is $2r\sin\left(\frac{\Delta\theta}{2}\right)$, where $\Delta\theta$ is the angle increment, e.g. the angle increment between Y_{j_1,k_1} and Y_{j_2,k_2} is $\Delta\theta = |\theta_{j_1,k_1} - \theta_{j_2,k_2}|$.

The governing equations are,

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}\left(X_{k-2} - X_{k+1}\right) - X_k - \left(\frac{ha}{b}\right)\sum_{j=1}^{J} Y_{j,k} + F \tag{14}$$

$$\frac{\mathrm{d}Y_{j,k}}{\mathrm{d}t} = -abY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - aY_{j,k} + \left(\frac{ha}{b}\right)X_k. \tag{15}$$

We follow Lorenz (1996) and let $K = 36, J = 10$, so there are 36 sectors and 10 times more "short" variables than "long" variables. We let $a = 10$ and $b = 10$, indicating that convective scales fluctuate 10 times faster than the larger scales, while their amplitude is around $1/10$ as large. For the forcing we choose $F = 10$, which Lorenz (1996) found to be sufficient to make both scales behave chaotically. All simulations are performed using an adaptive fourth-order Runge-Kutta method with relative error tolerance 10^{-3} and absolute error tolerance 10^{-6} (Dormand and Prince, 1980; Shampine and Reichelt, 1997). The solutions are output ~~with~~ each assimilation cycle. Unless otherwise specified, the assimilation cycles are separated by a time interval of 0.005 model time units (MTU), which Lorenz (1996) found to be similar to ~~or~~ 36 minutes in more realistic settings. This time scale is 10 times shorter than the time scale typically used in the univariate Lorenz 96 model. The factor of 10 is consistent with the understanding that the "short" process evolves 10 times faster than the "long" process, where the "long" process is akin to the univariate Lorenz 96 model. In choosing the coupling strength, we follow Roh et al. (2015) and set $h = 2$, which is twice as strong as the coupling used by Lorenz. ~~Increasing~~ Varying the coupling strength ~~leads to larger covariances between the forecast errors in processes X and Y, thereby making the effect of cross-localization more pronounced and easier to study.~~

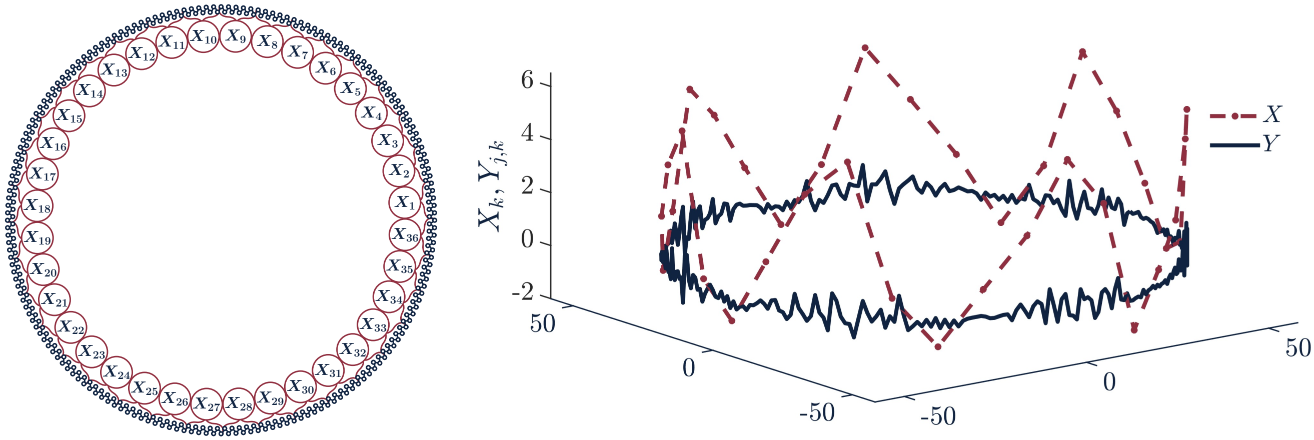


Figure 2. Left: Schematic illustrating the location of the different variables in the bivariate Lorenz 96 model, inspired by Wilks (2005). The setup has $K = 36$ sectors, with $J = 10$ “short” process variables per sector. The “long” process is shown on the inner circle and each X variable is labeled. The “short” process is shown, unlabeled, on the outer circle. Brackets show the sectors. Right: A single snapshot of the bivariate Lorenz 96 model with variables placed on a circle. The “long” process X ~~is shown with a~~ (red, dashed~~line and~~) has fluctuations with amplitude about 10 times larger than the fluctuation of the “short” process Y (dark, solid~~black line~~).

h across values $\{\frac{1}{2}, 1, 4\}$ changes the magnitude of the analysis errors, but does not change the relative performance of different localization functions in our experiments.

~~The variables are periodic, so we represent them on a circle where the arc length between neighboring Y variables is 1. There are 360 Y variables, so the radius of the circle is $r = 180/\pi$. Variable Y_j, k is located at $(r\cos(\theta_{j,k}), r\sin(\theta_{j,k}))$ where $\theta_{j,k} = \frac{\pi(10(k-1)+j)}{180}$. We choose to place the variable X_k in the middle of the sector whose variables are coupled to it, i. e. X_k is halfway in between $Y_{5,k}$ and $Y_{6,k}$. Variable X_k is located at $(r\cos(\phi_k), r\sin(\phi_k))$ where $\phi_k = \frac{\pi(10(k-1)+5.5)}{180}$. The placement of these variables is illustrated in Fig. 2. The chord distance between any two variables is $2r\sin\left(\frac{\Delta\theta}{2}\right)$, where $\Delta\theta$ is the angle increment, e.g. the angle increment between Y_{j_1,k_1} and Y_{j_2,k_2} is $\Delta\theta = |\theta_{j_1,k_1} - \theta_{j_2,k_2}|$.~~

3.2 ~~The assimilation~~ Assimilation scheme

~~We develop localization functions for data assimilation schemes that rely on Schur product modification of the background error covariance matrix B.~~ In our experiments we use the stochastic Ensemble Kalman Filter (EnKF) (Evensen, 1994; Houtekamer and Mitchell, 1998; Burgers et al., 1998). ~~However, because 3D-Var and the analysis step of the Kalman Filter are equivalent in the case of a linear observation operator (Daley, 1993), our results translate to EnVar schemes as well. The positive semi-definiteness of the localization matrix is essential to ensure convergence of the numerical optimization methods used to implement EnVar (Bannister, 2008). The~~ The EnKF update formula for a single ensemble member is

$$\mathbf{x}^{\mathrm{a}} = \underline{+\sqrt{\lambda}}\mathbf{x}^{\mathrm{b}}\underline{-} + \mathbf{K}\left(\mathbf{y} + \boldsymbol{\eta} - \mathbf{H}\mathbf{x}^{\mathrm{b}}\right) \tag{16}$$

where $\mathbf{x}^{\mathrm{a}}$ is the analysis vector, $\mathbf{x}^{\mathrm{b}}$ is the background state vector, ~~$\overline{\mathbf{x}^{\mathrm{b}}}$ is the background ensemble mean, λ is the inflation factor,~~ $\mathbf{y}$ is the observation, each element of $\boldsymbol{\eta}$ is a random draw from the probability distribution of observation errors, and $\mathbf{H}$ is the linear observation operator. The Kalman gain matrix $\mathbf{K}$ is

$$\mathbf{K} = \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \tag{17}$$

where $\mathbf{P}^{\mathrm{b}}$ is the background error covariance matrix and $\mathbf{R}$ is the observation error covariance matrix. The background covariance matrix is approximated by a sample covariance matrix from an ensemble, $\mathbf{x}_i$ for $i = 1, \ldots, N_e$ where N_e is the ensemble size. ~~The localization matrix $\mathbf{L}$ is incorporated into the estimate of the background covariance matrix through a Schur product as in equation (1).~~ In this experiment we use the adaptive inflation scheme of El Gharamti (2018) and ~~apply the inflation to the prior estimate~~inflate each prior ensemble member through,

$$\mathbf{x}^{\mathrm{b}}_{\lambda} = \overline{\mathbf{x}^{\mathrm{b}}} + \boldsymbol{\Lambda}^{1/2}\left(\mathbf{x}^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}}\right), \tag{18}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with each element on the diagonal containing the inflation factor for one variable and $\overline{\mathbf{x}^{\mathrm{b}}}$ is the background ensemble mean. We then use $\mathbf{x}^{\mathrm{b}}_{\lambda}$ in place of $\mathbf{x}^{\mathrm{b}}$ in Eq. (16) and in the estimation of $\mathbf{P}^{\mathrm{b}}$. Note that estimating $\mathbf{P}^{\mathrm{b}}$ with the inflated ensemble is equivalent to estimating it with the original ensemble and then multiplying by $\boldsymbol{\Lambda}^{1/2}$ on the left and right, $\boldsymbol{\Lambda}^{1/2}\mathbf{P}^{\mathrm{b}}\boldsymbol{\Lambda}^{1/2}$. We initialize the inflation ~~factor at $\lambda = 1.1$ everywhere, and the inflation variance at~~ factors with $\boldsymbol{\Lambda} = 1.1\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. El Gharamti (2018) defined a prior variance associated with each inflation factor. We initialize the prior inflation variance with $\sigma^2_{\lambda} = 0.09$ ~~.~~ and update the variance following the enhanced adaptive inflation algorithm. The localization matrix $\mathbf{L}$ is incorporated into the estimate of the background covariance matrix through a Schur product as in Eq. (1).

We use $N_e = 20$ ensemble members, except where otherwise noted. The small ensemble size is chosen to accentuate the spurious correlations and hence the need for effective localization functions. 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. Each experiment is repeated 50 times with independent reference states~~and observation errors. The observation operator $\mathbf{H}$ is such that all of the Y variables are observed, and none of the X variables are observed. In this way we can isolate the effect of the localization on the performance of the filter for the X variable. The observation error variance for the Y process is $\sigma^2_Y = 0.005$, which is about 5% of the climatological variance of the Y process.~~ , which serve as the “truth” in our experiments. We generate “observations” by adding independent Gaussian noise to the reference state.

3.3 ~~Univariate vs. multivariate setup~~

3.3 Experimental design

~~We compare univariate and multivariate versions of four localization functions: Gaspari-Cohn, Bolin-Wallin, Askey, and Wendland. Across all functions, the univariate localization function is equivalent to the~~ The experiments described in this section compare the performance of each of the four multivariate localization functions presented in Sect. 2. Performance is measured through the root-mean-square distance between the analysis mean and the true state, which we refer to as the root-mean-square error (RMSE). To standardize the comparison of the different shapes, we use the same within-component localization radii for all multivariate functions. We also investigate the performance of univariate and weakly coupled localization functions. The univariate localization functions are chosen to be equal to the within-component localization function for ~~the~~ Y ~~process, $\mathcal{L} = \mathcal{L}_{YY}$.All univariate functions use a localization radius of $R = 15$. All multivariate functions use localization radius $R_{YY} = 15$ for the Y process and $R_{XX} = 45$ for the X process. For multivariate Gaspari-Cohn and Bolin-Wallin this implies a~~ , i.e. $\mathcal{L} \equiv \mathcal{L}_{YY}$. The within-component weakly coupled localization functions are equal to the within-component multivariate localization functions. However, the weakly coupled cross-localization ~~radius equal to $R_{XY} = 30$. For multivariate Askey and Wendland the cross-localization radius is $R_{XY} = 15$. These localization radii are chosen in accordance with sensitivity experiments, described~~ functions are identically zero. The free localization function parameters are chosen to balance performance of the univariate, multivariate, and weakly coupled forms of each localization function. We estimate these parameters through a process which we describe in Appendix ~~??~~B1.

~~For all functions we use the maximum allowable~~ We test the performance of multivariate localization functions using three different observation operators. First we observe all “short” variables and none of the “long” variables. In this setup we isolate the impact the of the cross-localization ~~weight factor, $\beta = \beta_{\max}$. We find that, because we observe only the Y process and hence the only updates to X are through observations of Y, smaller~~ function, which allows us to make conjectures about important cross-localization ~~weight factors lead to degraded performance. Details of the sensitivity experiments involving β are provided in Appendix ??. We hypothesize that an important factor in the performance of~~ shape parameters. Next we flip the setup and observe all “long” variables and none of the “short” variables. We compare and contrast our findings with those from the previous case. Finally, we observe both processes and observe behavior reminiscent of both of the previous cases. The experimental setups are grouped by observation operator below. The source code for all experiments is publicly available (see Code Availability).

3.3.1 Observe only the “short” process

To isolate the impact of the ~~multivariate localization functions is the size of $\beta_{\max}$ and that functions with a larger $\beta_{\max}$ will allow for more information to propagate across model domains which will lead to better performance in our setup. With our chosen parameters, the multivariate Askey function has the largest~~ cross-localization functions we fully observe the “short” process and do not observe the “long” process at all. In this configuration, analysis increments of the “long” process can be fully attributed to the cross-domain assimilation of observations of the “short” process. The treatment of cross-domain background error covariances plays a crucial role in the analysis of the “long” process, so we expect that changes in the cross-localization ~~weight factor at $\beta_{\max} \approx 0.46$, followed by Gaspari-Cohn ($\beta_{\max} \approx 0.38$), Wendland ($\beta_{\max} \approx 0.22$), and Bolin-Wallin ($\beta_{\max} \approx 0.19$). A visual representation of this ordering is shown in the third panel of Fig. 1. In this figure we see that the shape of each cross-localization~~ function ~~varies not only in its height at zero, but also in its radius and smoothness near zero. While Askey peaks higher than GC, GC is generally smoother near zero and has a larger cross-localization radius. All of these differences in shape impact directly how much information propagates across model domains, so we hypothesize that GC~~

~~allows for larger cross-domain updates than any of the other multivariate localization functions presented here. The parameter choices for each function in the experiment and Fig. 1 are assembled~~ will lead to changes in the “long” process analyses. All observations are assimilated every 0.005 MTU. We use an observation error variance of $\sigma_Y^2 = 0.005$ both in the generation of synthetic observations from the reference state and in the assimilation scheme. The observation error variance is chosen to be about 5% of the climatological variance of the Y process. We also run the experiment with $\sigma_Y^2 = 0.02$, or about 20% of climatological variance, and find that the analyses are degraded, but the relative performance of the different localization functions is the same.

The localization parameters we use in this experiment are given in Table 3~~and described in further detail in Appendix ??. The source code for all experiments is publicly available (see Code Availability~~. For all functions we use the maximum allowable cross-localization weight factor, $\beta = \beta_{\max}$. In estimating the optimal cross-localization weight factor we find that since the only updates to X are through observations of Y, smaller cross-localization weight factors lead to degraded performance (Appendix B1).

Function name	Univariate parameters	Multivariate parameters	Univariate median RMSE in X	Multivariate median RMSE in X
Gaspari-Cohn	$R = 15$	$R_{YY} = 15$, $R_{XX} = 45$, $\boldsymbol{R_{XY} = 30}$, $\boldsymbol{\beta \approx 0.38}$	0.036	~~0.028~~ **0.028**
Bolin-Wallin	$R = 15$	$R_{YY} = 15$, $R_{XX} = 45$, $\boldsymbol{R_{XY} = 30}$, $\boldsymbol{\beta \approx 0.19}$	0.035	0.033
Askey	$R = 15$, $\nu = 1$	$R_{YY} = 15$, $R_{XX} = 45$, $\boldsymbol{R_{XY} = 15}$, $\boldsymbol{\beta \approx 0.46}$, $\nu = 1$, $\gamma_{YY} = 0$, $\gamma_{XX} = 1$, $\gamma_{XY} = \frac{1}{6}$	0.034	0.035
Wendland	$R = 15$, $\nu = 2$, $k = 1$	$R_{YY} = 15$, $R_{XX} = 45$, $\boldsymbol{R_{XY} = 15}$, $\boldsymbol{\beta \approx 0.22}$, $\nu = 2$, $\gamma_{YY} = 0$, $\gamma_{XX} = 5$, $\gamma_{XY} = \frac{5}{6}$, $k = 1$	0.036	0.047

Table 3. ~~Parameter choices~~ Localization parameters for the experiment ~~comparing univariate and~~ observing only the “short” process. Note that weakly coupled parameters are not shown in this table because they are exactly equal to the multivariate ~~localization~~parameters except with $\beta = 0$. The right columns show the median analysis RMSE of the X process for each function. Median RMSE for the weakly coupled case is not shown because weak coupling leads to catastrophic filter divergence.

4 ~~Univariate vs. multivariate results~~

3.0.1 Observe only the “long” process

Next we investigate the impact of the different localization functions when we fully observe the “long” process and do not observe the “short” process at all. The “long” process fluctuates about 10 times more slowly than the “short” process, so we use an assimilation cycle length that is 10 times longer than the one in the previous configuration. All observations are assimilated every 0.05 MTU. We use an observation error variance of $\sigma_X^2 = 0.28$ both in the generation of synthetic observations from the reference state and in the assimilation scheme. The observation error variance is chosen to be about 5% of the climatological variance of the X process. We also run the experiment with $\sigma_X^2 = 1.1$, or about 20% of climatological variance, and find that the X analyses are degraded, but the relative performance of the different localization functions is the same. The localization parameters we use in this experiment are given in Table B1. We find that the analysis errors are similar with all values of β. For consistency with the previous experiment we use the maximum allowable cross-localization weight factor, $\beta = \beta_{\max}$.

3.0.2 Observe both processes

Finally, we observe both processes and note the impact of the different localization functions. In this configuration we observe 75% of the variables in each process, with the observation locations chosen randomly for each trial. All observations are assimilated every 0.005 MTU, in line with the analysis cycle length for the observation of the "short" process only. We use observation error variances of $\sigma_Y^2 = 0.01$ and $\sigma_X^2 = 0.57$ in the generation of observations and in the assimilation scheme. The observation error variance is chosen to be about 10% of the climatological variance of each process. We also run the experiment with $\sigma_Y^2 = 0.02$ and $\sigma_X^2 = 1.1$, or about 20% of climatological variance, and find that the performance is similar. The localization parameters we use in this experiment are given in Table B2. We find that the analysis errors grow with increasing β. Nonetheless, to distinguish between multivariate localization, which allows for cross-domain information transfer, and weakly coupled localization, which does not, we use $\beta = \beta_{\max}$ for all multivariate functions.

3.1 Results

~~Figure 3 shows the root mean square error (RMSE) for process~~

3.1.1 Observe only the "short" process

Figure 3 shows the distribution of RMSEs for the configuration described in Sect. 3.3.1. With weakly coupled localization functions no information is shared in the update step between the observed Y process and the unobserved X ~~for the localization functions defined in Table 3. The RMSE compares analysis states to the "truth", which is the state we perturb to generate the "observations ". Each panel compares performance with univariate and multivariate versions of a function.The performance of all~~ process. This leads to no updates of the X variables and eventually to catastrophic filter divergence. In principle the system might be able to synchronize the unobserved ("long") process through dynamical couplings with the observed ("short") process, but in our setup this does not happen. Hence weakly coupled localization functions are not included in the figure. The RMSE distributions for the observed Y process are similar for all functions except multivariate Wendland. For the unobserved X process, the analysis errors are comparable across all of the univariate localization functions~~is very similar~~. This is consistent with the fact that all of the univariate localization functions have similar shapes as seen in the ~~first~~ second panel of Fig. 1. The ~~univariate functions, while they do not allow for longer-range cross-domain updates, do allow for the largest cross-domain updates at small distances, with $\mathcal{L}(0) = 1$ and hence provide a consistent benchmark against which to test the multivariate functions. The multivariate~~ multivariate localization functions, on the other hand, show great diversity of performance. ~~GC shows improved performance when using the multivariate localization function . By contrast, the~~ The Wendland function leads to significantly worse performance with the multivariate function when compared to the univariate functions. BW and Askey functions show no statistically significant difference between the multivariate and univariate ~~versions. The Wendland function shows significantly worse performance with the multivariate function when compared to the univariate version~~functions. Out of all of the localization functions we consider, the best performance is achieved with multivariate GC.

~~As noted in~~ To understand the improved performance with multivariate GC, we consider two different shape parameters. Recall from Sect. ~~3.3 the shape of the GC~~ 3.3.1 that smaller cross-localization ~~function appears to allow for the most cross-domain localization because it is nearly as tall as Askey, fairly smooth near $d = 0$, and it decays to zeromore slowly than any other function, with the possible exception of BW. We~~ weights led to worse performance when holding all other localization parameters fixed. Extending this finding, we hypothesize that functions with a larger $\beta_{\max}$ will allow for more information to propagate across model domains, thereby improving performance in this setup. With the chosen localization parameters, the multivariate Askey function has the largest cross-localization weight factor with $\beta_{\max} \approx 0.46$, followed by GC with $\beta_{\max} \approx 0.38$. A visual representation of the cross-localization weight factor is shown as the height of the cross-localization function at zero in the third panel of Fig. 1. The shape of each cross-localization function varies not only in its height at zero, but also in its radius and smoothness near zero. For example, while the Askey cross-localization function peaks higher than GC, GC is generally smoother near zero and has a larger cross-localization radius. All of these differences in shape impact how much information propagates across model domains. Based on its height and width, we hypothesize that GC allows for sufficient cross-domain information propagation at both short and long distances and this is why multivariate GC outperforms all other functions in this experiment.

~~Askey has the largest cross-localization weight factor of any of the multivariate functions we consider, and yet shows no improvement over the univariate version. Both the Askey and Wendland functions have smaller cross-localization radii and fall off very rapidly compared to GC and BW. Thus multivariate Askey allows for the largest cross-domain updates at short distances, but not at longer distances. By contrast, Wendland and BW both have small cross-localization weight factors, so that even at short distances the ability to propagate information across model domains is limited. Thus, BW allows for minimal cross-domain updates at short distances, but this falls off slowly at longer distances. Multivariate Wendland, which allows for minimal cross-domain updates at short distances and falls off very quickly, shows the worst performance.~~

3.1.2 Observe only the "long" process

When we observe only the "long" process (as described in Sect. 3.0.1), we find that all localization functions lead to very similar performance. In this case the shape of the localization function is not important. Rather, the dynamics of the bivariate Lorenz model are driving the behavior. In this configuration, the true background error cross-correlations are very small (less than 0.1 even at short distances). The Y variables are restored towards $\left(\frac{h}{b}\right)X$ in their sector (Eq. 15). Thus even when the assimilation does not update the Y variables, we expect to recover the mean of the Y process. Based on climatology we find that the conditional mean of $Y_{j,k}$ given $X_k = x$ is $E[Y_{j,k}|x] \approx 0.0559x$. The median root-mean-square difference between Y and its conditional mean is 0.294. Our results show that the median RMSE in the Y process ranges from 0.294 to 0.297. Thus, the filter does not improve upon a simple linear conditional mean prediction, which is perhaps unsurprising given the small magnitude of error cross-correlations. Figure 4 shows the distribution of RMSE for univariate, weakly coupled, and multivariate GC. The distributions for other functions are nearly identical and hence not shown.

3.1.3 Observe both processes

When we observe both processes the precise shape of the localization function appears to have little impact. We do see differences between univariate, weakly coupled, and multivariate localization functions. Figure 4 shows RMSE distributions for the three different versions of GC, which are broadly representative of the behavior seen in other functions as well. This configuration is quite unstable. About 80% of the trials with weakly coupled localization functions lead to catastrophic filter divergence. Trials with univariate and multivariate localization diverge less often, but still diverge about 20% of the time. Figure 4 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.

4 Conclusions

In this work we developed a multivariate extension of the oft-used GC localization function, where the within-component localization functions are standard GC with different localization radii, while the cross-localization function is newly constructed to ensure that the resulting localization matrix is positive semidefinite. A positive semidefinite localization matrix guarantees, through the Schur product theorem, that the localized estimate of the background error covariance matrix is positive semidefinite (Horn and Johnson, 2012, Theorem 7.5.3). We compared multivariate GC to three other multivariate localization functions (including one other newly presented multivariate function), and their univariate and weakly coupled counterparts. We found that ~~, in a toy model,~~ the performance of different localization functions is highly dependent on the observation operator. When we observed only the “long” process, all localization functions performed similarly. In an experiment where we observed both processes, weakly coupled localization led to the smallest analysis error. When we observed only the “short” process, multivariate GC led to better performance than any of the other localization functions we considered. We hypothesized that the shape of the GC cross-localization function allows for larger cross-domain assimilation than the other functions. There is still an outstanding question of how multivariate GC will perform in other, perhaps more realistic, systems.

~~In this work we investigated the importance of the~~ We found that choosing an appropriate cross-localization weight factor, β, ~~which~~ is crucial to the performance of the multivariate localization functions. This parameter determines the amount of information which is allowed to propagate between co-located variables in different model components. We found that this parameter should be as large as possible ~~. This is likely unique to our setup as other studies~~ when observing only the "short" process. By contrast, the parameter should be small or even zero when both processes are well observed. This is consistent with other studies which have shown the value in deflating cross-domain updates between non-interacting processes (Lu et al., 2015; Yoshida and Kalnay, 2018). ~~This can easily be incorporated in this framework by taking β to be small or even zero.~~

A natural application of this work is localization in a coupled atmosphere-ocean model. Multivariate GC allows for within-component covariances to be localized with GC exactly as they would be in an uncoupled setting, using the optimal localization length scale for each component (Ying et al., 2018). ~~In this work we discussed the importance of the~~ The cross-localization length scale for GC is the average of the two within-component localization radii, which is the same as the cross-localization radius ~~in determining performance. However, this work did not address the question of~~ proposed in Frolov et al. (2016). We hypothesize that the cross-localization radius is important in determining filter performance. However, the functions considered here did not allow for a thorough investigation of the optimal cross-localization radius~~selection~~, which is an important area for future research.

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.

Appendix A: Derivation of multivariate Gaspari-Cohn

A1 Multivariate Gaspari-Cohn cross-localization function

Let c_X, c_Y be the kernel radii associated with model components X and Y. Without loss of generality, we take $c_X > c_Y$. The formula depends on the relative sizes of c_X and c_Y, with two different formulas for the cases (i) $c_X < 2c_Y$ and (ii) $c_X \geq 2c_Y$. In both cases, the notation is significantly simplified when we let $c_X = \kappa^2 c_Y$. The first case we consider is $c_Y < c_X < 2c_Y$. In this case, the GC cross-localization function is,

$$
\mathcal{L}_{XY}^{(GC)}(d) = \frac{\beta}{\beta_{\max}} \begin{cases}
-\frac{1}{6}\left(\frac{|d|}{\kappa c_Y}\right)^5 + \frac{1}{2}\left(\frac{d}{\kappa c_Y}\right)^4 \frac{1}{\kappa} - \frac{5}{3}\left(\frac{d}{\kappa c_Y}\right)^2 \frac{1}{\kappa^3} + \frac{5}{2}\frac{1}{\kappa^3} - \frac{3}{2}\frac{1}{\kappa^5} & 0 < |d| < c_X - c_Y \\[2ex]
\begin{array}{l} -\frac{1}{4}\left(\frac{|d|}{\kappa c_Y}\right)^5 + \frac{1}{4}\left(\frac{d}{\kappa c_Y}\right)^4 \left(\kappa + \frac{1}{\kappa}\right) + \frac{5}{8}\left(\frac{|d|}{\kappa c_Y}\right)^3 \\ \quad -\frac{5}{6}\left(\frac{d}{\kappa c_Y}\right)^2 \left(\kappa^3 + \frac{1}{\kappa^3}\right) + \frac{5}{4}\frac{|d|}{\kappa c_Y}\left(\kappa^4 - \kappa^2 - \frac{1}{\kappa^2} + \frac{1}{\kappa^4}\right) \\ \quad +\frac{1}{6}\frac{\kappa c_Y}{|d|}\left(\kappa^6 + \frac{1}{\kappa^6}\right) - \frac{3}{8}\frac{\kappa c_Y}{|d|}\left(\kappa^4 + \frac{1}{\kappa^4}\right) + \frac{5}{12}\frac{\kappa c_Y}{|d|} \\ \quad -\frac{3}{4}\left(\kappa^5 + \frac{1}{\kappa^5}\right) + \frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right) \end{array} & c_X - c_Y < |d| < c_Y \\[2ex]
\begin{array}{l} -\frac{1}{12}\left(\frac{|d|}{\kappa c_Y}\right)^5 + \frac{1}{4}\left(\frac{d}{\kappa c_Y}\right)^4 \left(\kappa - \frac{1}{\kappa}\right) + \frac{5}{8}\left(\frac{|d|}{\kappa c_Y}\right)^3 \\ \quad -\frac{5}{6}\left(\frac{d}{\kappa c_Y}\right)^2 \left(\kappa^3 - \frac{1}{\kappa^3}\right) + \frac{5}{4}\frac{|d|}{\kappa c_Y}\left(\kappa^4 - \kappa^2 - \frac{1}{\kappa^2} - \frac{1}{\kappa^4}\right) \\ \quad +\frac{1}{6}\frac{\kappa c_Y}{|d|}\left(\kappa^6 - \frac{1}{\kappa^6}\right) - \frac{3}{8}\frac{\kappa c_Y}{|d|}\left(\kappa^4 + \frac{1}{\kappa^4}\right) + \frac{5}{12}\frac{\kappa c_Y}{|d|} \\ \quad -\frac{3}{4}\left(\kappa^5 - \frac{1}{\kappa^5}\right) + \frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right) \end{array} & c_Y < |d| < c_X \\[2ex]
\begin{array}{l} \frac{1}{12}\left(\frac{|d|}{\kappa c_Y}\right)^5 - \frac{1}{4}\left(\frac{d}{\kappa c_Y}\right)^4 \left(\kappa + \frac{1}{\kappa}\right) + \frac{5}{8}\left(\frac{|d|}{\kappa c_Y}\right)^3 \\ \quad +\frac{5}{6}\left(\frac{d}{\kappa c_Y}\right)^2 \left(\kappa^3 + \frac{1}{\kappa^3}\right) - \frac{5}{4}\frac{|d|}{\kappa c_Y}\left(\kappa^4 + \kappa^2 + \frac{1}{\kappa^2} + \frac{1}{\kappa^4}\right) \\ \quad -\frac{1}{6}\frac{\kappa c_Y}{|d|}\left(\kappa^6 + \frac{1}{\kappa^6}\right) - \frac{3}{8}\frac{\kappa c_Y}{|d|}\left(\kappa^4 + \frac{1}{\kappa^4}\right) + \frac{5}{12}\frac{\kappa c_Y}{|d|} \\ \quad +\frac{3}{4}\left(\kappa^5 + \frac{1}{\kappa^5}\right) + \frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right) \end{array} & c_X < |d| < c_Y + c_X.
\end{cases} \tag{A1}
$$

where $\beta_{\max} = \frac{5}{2}\kappa^{-3} - \frac{3}{2}\kappa^{-5}$ and $\beta \leq \beta_{\max}$. Note that when we take $c_X \to c_Y$, which implies $\kappa \to 1$, this multivariate function converges to the standard univariate ~~Gaspari-Cohn~~ GC function, as we would expect.

The second case to consider is $c_X > 2c_Y$. Again, let $c_X = \kappa^2 c_Y$. In this case, the cross-localization function is

$$\mathcal{L}_{XY}^{(GC)}(d) = \frac{\beta}{\beta_{\max}} \begin{cases} -\frac{1}{6}\left(\frac{|d|}{\kappa c_Y}\right)^5 + \frac{1}{2}\left(\frac{d}{\kappa c_Y}\right)^4 \frac{1}{\kappa} - \frac{5}{3}\left(\frac{d}{\kappa c_Y}\right)^2 \frac{1}{\kappa^3} + \frac{5}{2}\frac{1}{\kappa^3} - \frac{3}{2}\frac{1}{\kappa^5} & 0 \le |d| < c_Y \\[2ex] -\frac{5}{2}\frac{|d|}{\kappa c_Y}\frac{1}{\kappa^4} - \frac{1}{3}\frac{\kappa c_Y}{|d|}\frac{1}{\kappa^6} + \frac{5}{2}\frac{1}{\kappa^3} & c_Y \le |d| < c_X - c_Y \\[2ex] \begin{aligned} &-\tfrac{1}{12}\left(\tfrac{|d|}{\kappa c_Y}\right)^5 + \tfrac{1}{4}\left(\tfrac{d}{\kappa c_Y}\right)^4\left(\kappa - \tfrac{1}{\kappa}\right) + \tfrac{5}{8}\left(\tfrac{|d|}{\kappa c_Y}\right)^3 \\ &\quad -\tfrac{5}{6}\left(\tfrac{d}{\kappa c_Y}\right)^2\left(\kappa^3 - \tfrac{1}{\kappa^3}\right) + \tfrac{5}{4}\tfrac{|d|}{\kappa c_Y}\left(\kappa^4 - \kappa^2 - \tfrac{1}{\kappa^2} - \tfrac{1}{\kappa^4}\right) \\ &\quad +\tfrac{1}{6}\tfrac{\kappa c_Y}{|d|}\left(\kappa^6 - \tfrac{1}{\kappa^6}\right) - \tfrac{3}{8}\tfrac{\kappa c_Y}{|d|}\left(\kappa^4 + \tfrac{1}{\kappa^4}\right) + \tfrac{5}{12}\tfrac{\kappa c_Y}{|d|} \\ &\quad -\tfrac{3}{4}\left(\kappa^5 - \tfrac{1}{\kappa^5}\right) + \tfrac{5}{4}\left(\kappa^3 + \tfrac{1}{\kappa^3}\right) \end{aligned} & c_X - c_Y \le |d| < c_X \\[2ex] \begin{aligned} &\tfrac{1}{12}\left(\tfrac{|d|}{\kappa c_Y}\right)^5 - \tfrac{1}{4}\left(\tfrac{d}{\kappa c_Y}\right)^4\left(\kappa + \tfrac{1}{\kappa}\right) + \tfrac{5}{8}\left(\tfrac{|d|}{\kappa c_Y}\right)^3 \\ &\quad +\tfrac{5}{6}\left(\tfrac{d}{\kappa c_Y}\right)^2\left(\kappa^3 + \tfrac{1}{\kappa^3}\right) - \tfrac{5}{4}\tfrac{|d|}{\kappa c_Y}\left(\kappa^4 + \kappa^2 + \tfrac{1}{\kappa^2} + \tfrac{1}{\kappa^4}\right) \\ &\quad -\tfrac{1}{6}\tfrac{\kappa c_Y}{|d|}\left(\kappa^6 + \tfrac{1}{\kappa^6}\right) - \tfrac{3}{8}\tfrac{\kappa c_Y}{|d|}\left(\kappa^4 + \tfrac{1}{\kappa^4}\right) + \tfrac{5}{12}\tfrac{\kappa c_Y}{|d|} \\ &\quad +\tfrac{3}{4}\left(\kappa^5 + \tfrac{1}{\kappa^5}\right) + \tfrac{5}{4}\left(\kappa^3 + \tfrac{1}{\kappa^3}\right) \end{aligned} & c_X \le |d| < c_Y + c_X \end{cases} \tag{A2}$$

where, as in the above case, $\beta_{\max} = \frac{5}{2}\kappa^{-3} - \frac{3}{2}\kappa^{-5}$ and $\beta \le \beta_{\max}$. Note that when $c_X = 2c_Y$ (A1) is equal to (A2).

A2 Derivation of multivariate Gaspari-Cohn cross-localization function $\mathcal{L}_{XY}^{(GC)}$

The multivariate GC cross-localization function is created through the convolution of two kernels, $\mathcal{L}_{XY}^{(GC)}(d) = [k_X * k_Y](\mathbf{d})$, with $k_j(\mathbf{r}) = k_j^0(\|\mathbf{r}\|) = (1 - \|\mathbf{r}\|/c_j)_+$, $j = X, Y$, and $\mathbf{r} \in \mathbb{R}^3$. Theorem 3.c.1 from Gaspari and Cohn (1999) provides a framework for evaluating the necessary convolutions. It is shown that for radially symmetric functions $k_j(\mathbf{r}) = k_j^0(||\mathbf{r}||)$ compactly supported on a sphere of radius c_j, $j = X, Y$, with $c_Y \le c_X$ the convolution over $\mathbb{R}^3$ given by

$$P_{XY}^0(\|\mathbf{d}\|) = \int k_X^0(\|\mathbf{r}\|) k_Y^0(\|\mathbf{d} - \mathbf{r}\|)\,\mathrm{d}\mathbf{r}, \tag{A3}$$

can equivalently be written,

$$P_{XY}^0(d) = \frac{2\pi}{d} \int_0^{c_Y} r k_Y^0(r) \int_{|r-d|}^{r+d} s k_X^0(s)\,\mathrm{d}s\,\mathrm{d}r. \tag{A4}$$

Equation (A4) is normalized to produce a localization function with $\mathcal{L}_{XX}(0) = \mathcal{L}_{YY}(0) = 1$. The normalization factor $P_{jj}^0(0)$ is given by

$$P_{jj}^0(0) = 4\pi \int_0^{c_j} \left(r k_j^0(r)\right)^2\,\mathrm{d}r, \quad j = X, Y. \tag{A5}$$

The resulting cross-localization function is a normalized version of (A4),

$$\mathcal{L}_{XY}(d) := \frac{P_{XY}^0(d)}{[P_{XX}^0(0) P_{YY}^0(0)]^{1/2}}, \tag{A6}$$

With this framework, we are now able to compute the cross-localization function using the GC kernels. We first compute the normalization factor $P^0_{jj}(0)$ using GC kernels. Plugging in $k^0_j(r) = (1 - r/c_j)_+$ gives,

$$P^0_{jj}(0) = 4\pi \int_0^{c_j} r^2(1 - r/c_j)^2 \,\mathrm{d}r = \frac{2\pi}{15} c_j^3, \quad j = X, Y. \tag{A7}$$

Thus the denominator in Eq. (A6) is

$$[P^0_{XX}(0) P^0_{YY}(0)]^{1/2} = \frac{2\pi}{15} \sqrt{c_X^3 c_Y^3}. \tag{A8}$$

To compute the numerator in Eq. (A6), which is precisely (A4), we consider eight different cases, four cases for each formula presented above.

The case $c_X > 2c_Y$ and $0 \leq |d| < c_Y$ is shown in detail here. The other cases are derived similarly and are not shown. The inner integral in ~~equation~~ Eq. (A4) is

$$\int_{|r-d|}^{r+d} s k^0_X(s) \,\mathrm{d}s = \int_{|r-d|}^{r+d} s(1 - s/c_X) \,\mathrm{d}s = 2rd - \frac{1}{3c_X} \begin{cases} 2r^3 + 6rd^2 & \text{if } r \leq d \\ 6r^2 d + 2d^3 & \text{if } r \geq d \end{cases} \tag{A9}$$

The outer integral in (A4) is

$$\int_0^{c_Y} r(1 - r/c_Y)(2rd) \,\mathrm{d}r - \frac{1}{3c_X} \int_0^{d} r(1 - r/c_Y)(2r^3 + 6rd^2) \,\mathrm{d}r - \frac{1}{3c_X} \int_d^{c_Y} r(1 - r/c_Y)(6r^2 d + 2d^3) \,\mathrm{d}r \tag{A10}$$

which simplifies to

$$\frac{1}{6} d c_Y^3 - \frac{1}{3c_X} d \left[\frac{1}{30 c_Y} d^5 - \frac{1}{10} d^4 + \frac{1}{3} c_Y^2 d^2 + \frac{3}{10} c_Y^4 \right]. \tag{A11}$$

Substituting (A11) into (A4) we see,

$$P^0_{XY}(d < c_Y) = 2\pi \left(\frac{1}{6} c_Y^3 - \frac{1}{3c_X} \left[\frac{1}{30 c_Y} d^5 - \frac{1}{10} d^4 + \frac{1}{3} c_Y^2 d^2 + \frac{3}{10} c_Y^4 \right] \right). \tag{A12}$$

With the proper normalization, we have the cross-localization function,

$$\mathcal{L}_{XY}(d < c_Y) = \frac{15}{2\pi \sqrt{c_X^3 c_Y^3}} P^0_{XY}(d < c_Y). \tag{A13}$$

Now make the substitution $\kappa^2 = \frac{c_X}{c_Y}$ and this becomes

$$\mathcal{L}_{XY}(d < c_Y) = -\frac{1}{6} \left(\frac{d}{\kappa c_Y} \right)^5 + \frac{1}{2\kappa} \left(\frac{d}{\kappa c_Y} \right)^4 - \frac{5}{3\kappa^3} \left(\frac{d}{\kappa c_Y} \right)^2 + \frac{5}{2} \left(\frac{1}{\kappa^3} \right) - \frac{3}{2} \left(\frac{1}{\kappa^5} \right). \tag{A14}$$

Other cases are calculated similarly, with careful consideration of the bounds of the kernels and integrals.

A3 Multivariate Gaspari-Cohn with three or more length scales

Suppose we have p processes, $X_1,\ldots,X_p$ with p different localization radii $R_{11},\ldots,R_{pp}$. Define the associate kernel radii by $c_j = R_{jj}/2$ and the associated kernels by $k_j(r) \propto (1-r/c_j)_+$. Then the localization function used to taper background error covariances between process X_i and X_j is $\mathcal{L}_{ij}(d) = \alpha_{ij}[k_i * k_j](\mathbf{d})$, with

$$[\alpha_{ij}]_{i,j=1}^{p} \tag{A15}$$

a positive semidefinite matrix with 1's on the diagonal, i.e. $\alpha_{ii} = 1$. When $i=j$, $\mathcal{L}_{ii}$ is precisely the standard univariate GC function. When $i \neq j$, $\mathcal{L}_{ij}$ is given by Eq. (A1) if $\max\{R_{ii},R_{jj}\} < 2\min\{R_{ii},R_{jj}\}$ or Eq. (A2) otherwise. The ratio of length scales κ is defined as $\kappa^2 = \frac{\max\{R_{ii},R_{jj}\}}{\min\{R_{ii},R_{jj}\}}$. We have written (A1) and (A2) with a coefficient $\beta/\beta_{\max}$, which is convenient for the case of two components. Here we replace $\beta/\beta_{\max}$ by α_{ij} to emphasize the importance for three or more length scales is in choosing α_{ij} such that (A15) is positive semidefinite. Wang et al. (2021) discussed how to construct a similar matrix for multiscale localization using matrix square roots. The simplest case is to let $\alpha_{ij} = 1$ for all i,j.

Appendix B: ~~Sensitivity experiments~~Estimation of localization parameters

B1 ~~Localization radius~~

A fair comparison between the univariate, weakly coupled, and multivariate localization functions requires that thoughtful attention be paid to the many parameter choices in the different localization functions. ~~While each localization function has its own parameters and constraints on those parameters, there are two~~ We estimate different localization parameters for each observation operator. This section describes our reasoning behind the selection of the localization parameters ~~which are shared by all functions:~~ for the experiment where we observe only the "short" process. We follow the same estimation procedure for the other two observation operators as well.

Some of the parameters are shared across functions. For example, every univariate function has a localization radius R~~, and~~ . To aid in comparisons between functions, we estimate a single univariate localization radius which is shared by all univariate functions. Indeed, whenever different methods share a parameter we estimate a single value for it. We estimate a separate cross-localization weight factor β ~~. We discuss the localization radius in this section and the cross-localization weight factor in ??.~~ for each function because each function has a different upper bound on this parameter.

We estimate the localization parameters iteratively in the the following way. First, note that Wendland is a family of functions, with parameter k controlling the smoothness. In sensitivity experiments (not shown) we found that increasing k degrades the performance of the filter. Thus, we choose to use $k=1$ for all experiments.

~~We first pick an appropriate localization radius for univariate localization functions~~Next, we pick appropriate localization radii for each process. We use a large (500-member) ensemble with no localization to compute forecast error correlations, hereafter called the "true" forecast error correlations, and shown in Fig. B1. We see that the true forecast error correlations for the "short" process Y degrade to zero in just a few spatial units. The forecast errors for the "long" process X, by contrast, have meaningful correlations out to about 40 spatial units. ~~We observe the entire Y process and none of the X process, thus the dominant behavior for the purposes of constructing a background error covariance matrix is determined by the "short" Y process. Hence we choose the localization radius for the~~ This gives us a baseline for the range of localization radii we should investigate. We compare the performance of all univariate localization functions ~~to be consistent with the non-zero correlation range of the Y process. Sensitivity experiments (not shown) reveal that~~ with the radius: $R \in [5, 10, 15, 20, 30, 45]$. In these sensitivity experiments we use $\nu = 1$ for Askey and $\nu = 2$ for Wendland. These values of ν are as small as possible while still guaranteeing positive semidefiniteness. Figure B2 shows that univariate localization radius $R = 15$ ~~is an appropriate localization radius for univariate localization~~leads to the best performance.

~~For multivariate localization , we keep the same localization radius for the "short" process ,~~

Using this univariate localization radius, we now estimate ν for univariate Askey and Wendland. To maintain positive semidefiniteness we require $\nu \geq 1$ for Askey and $\nu \geq 2$ for Wendland. We compare RMSEs for process X and Y with $\nu \in [1, 1.5, 2, 2.5]$ for Askey and $\nu \in [2, 2.5, 3, 3.5]$ for Wendland. In general we find that smaller value of ν lead to less peaked localization functions and better performance, and choose ν to be as small as possible, i.e. ~~$R_{YY} = R = 15$, and allow the radius for~~ $\nu = 1$ for Askey and $\nu = 2$ for Wendland.

Next we estimate the optimal *multivariate* localization radii. We want to eliminate as much ambiguity as possible in our comparison of univariate and multivariate localization functions. For this reason we choose to set the univariate localization radius equal to one of the within-component localization radii. From Fig. B1 we know that the univariate localization radius $R = 15$ is closer to the ~~X variable to vary. Informed by the true forecast error correlations we choose $R_{XX} = 45$, which is approximately the~~ range of ~~non-negligible correlations for the forecast errors in the X process .~~

~~For the four functions under consideration here,~~ significant true forecast error correlations for process Y than for process X so we set $R_{YY} = R = 15$. Now for the within-X localization radius, we consider the following localization values: $R_{XX} \in [30, 40, 45, 50, 6$ For Askey and Wendland we use $R_{XY} = \min\{R_{XX}, R_{YY}\}$, and $\gamma_{ij} = 0$ for all $i, j = X, Y$. For all functions we use $\beta_{\max}$ as the ~~choice of these two localization radii determines the~~ cross-localization ~~radius R_{XY}. As shownin Table 2,~~ weight factor. The RMSE for both processes is minimized with values of R_{XX} between 40 and 50 (not shown). Informed by the true forecast error correlations, we pick $R_{XX} = 45$. Now we turn to the cross-localization radius~~for~~. For Gaspari-Cohn and Bolin-Wallin ~~is $R_{XY} = 30$, while for~~ R_{XY} is determined by R_{XX} and R_{YY}, with $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$. For Askey and Wendland we ~~have $R_{XY} = 15$. Note that with~~ require $R_{XY} \leq \min\{R_{XX}, R_{YY}\}$ to maintain positive semidefiniteness. From the true forecast error correlations we see that the correlation length scale for X is larger than the cross-correlation length scale, which is in turn larger than the length scale for Y. This intuition tells us that, ideally we would have $R_{YY} < R_{XY} < R_{XX}$. However, because of the requirement for positive semidefinitess in Askey and Wendland ~~we~~ the closest we can come is $R_{YY} = R_{XY} < R_{XX}$. We could choose to use a smaller cross-localization radius, but the true forecast error correlation indicates that this would be a mistake, as there are non-negligible cross-correlations out past 15 units. Thus, we choose $R_{XY} = R_{YY} = \min\{R_{YY}, R_{XX}\}$.

Using all of the previously estimated multivariate localization parameters, we now estimate γ_{ij}, for all processes $i, j = X, Y$ for both Askey and Wendland. For Askey we consider all combinations of $\gamma_{YY} \in [0, 1, 2]$ and $\gamma_{XX} \in [0, 1, 2, 3]$. For Wendland we consider all combinations of $\gamma_{YY} \in [0, 1, 2]$ and $\gamma_{XX} \in [0, 1, 3, 4, 5, 6, 7, 9]$. The guarantee of positive semidefiniteness restricts our search for γ_{XY} to $\gamma_{XY} \geq \frac{R_{XY}}{2}\left(\frac{\gamma_{XX}}{R_{XX}} + \frac{\gamma_{YY}}{R_{YY}}\right)$. For simplicity, we take γ_{XY} to be at the edge of the allowable range, $\gamma_{XY} = \frac{R_{XY}}{2}\left(\frac{\gamma_{XX}}{R_{XX}} + \frac{\gamma_{YY}}{R_{YY}}\right)$. While investigating γ, we use the maximum allowable cross-localization weight factor. For Askey we find that the best performance comes with $\gamma_{XX} = 1$ and $\gamma_{YY} = 0$. For Wendland we see that performance improves as γ_{XX} increases, all the way out to $\gamma_{XX} = 5$. We hypothesize that this is because increasing γ_{XX} allows for an increased cross-localization weight factor. We use $\gamma_{XX} = 5$ and $\gamma_{YY} = 0$ for Wendland.

B1 ~~Impact of cross-localization weight factor~~

~~The~~ The final localization parameter to estimate is the cross-localization weight factor, β~~,~~. This parameter determines how much cross-domain information propagation occurs between the X and Y processes. ~~This parameter is investigated for the Askey localization function with the same support for both processes by Roh et al. (2015).~~ Each multivariate localization function has a different upper bound on β, which depends on a ratio of localization radii, as shown in ~~Table 2 and~~ Fig. B3.

~~When~~ Note that setting $\beta = 0$ ~~no information is shared in the update step between the observed Y process and the unobserved X process. In our setup, this leads to no updates of the X variables and eventually to catastrophic filter divergence. In principle the system might be able to synchronize the unobserved (“long”) process through dynamical couplings with the observed (“short”) process, but in our setup this does not happen. The~~ leads to a weakly coupled scheme, so to distinguish between multivariate and weakly coupled we consider only value of β greater than 0.1. For each multivariate localization function, we vary β between $\beta_{\max}$ and 0.1 while holding all other parameters fixed. In this setup, the best performance generally comes when the cross-correlation is at or near its maximum allowable value, as shown in Fig. B3. Figure B3 shows visually that the ~~Gaspari-Cohn~~ GC cross-correlation is always greater than the ~~Bolin-Wallin~~ BW cross-correlation, which is easily verified analytically since $\kappa^{-3} \leq \frac{5}{2}\kappa^{-3} - \frac{3}{2}\kappa^{-5}$ for all $\kappa \geq 1$ (true by the definition of κ). Similarly we see that the cross-localization weight factor for Askey is greater than cross-localization weight factor for Wendland across the range of parameters considered here.

B1 ~~Technical details of Askey and Wendland functions~~

~~For the multivariate Askey we must choose the parameters $R_{XY}, \gamma_{XX}, \gamma_{YY}, \gamma_{XY}$, and ν. Both R_{XY} and γ_{XY} have bounds restricting the possible range of values to ensure positive semidefiniteness. For simplicity, we take these values to be at the edge of the allowable range, $R_{XY} = \min\{R_{XX}, R_{YY}\}$~~The localization parameters for the other two observation operators are estimated following the same procedure. The localization parameters for the experiment where we observe only the “long” process are given in Table B1. The localization parameters for the experiment where we observe both processes are given in Table B2.

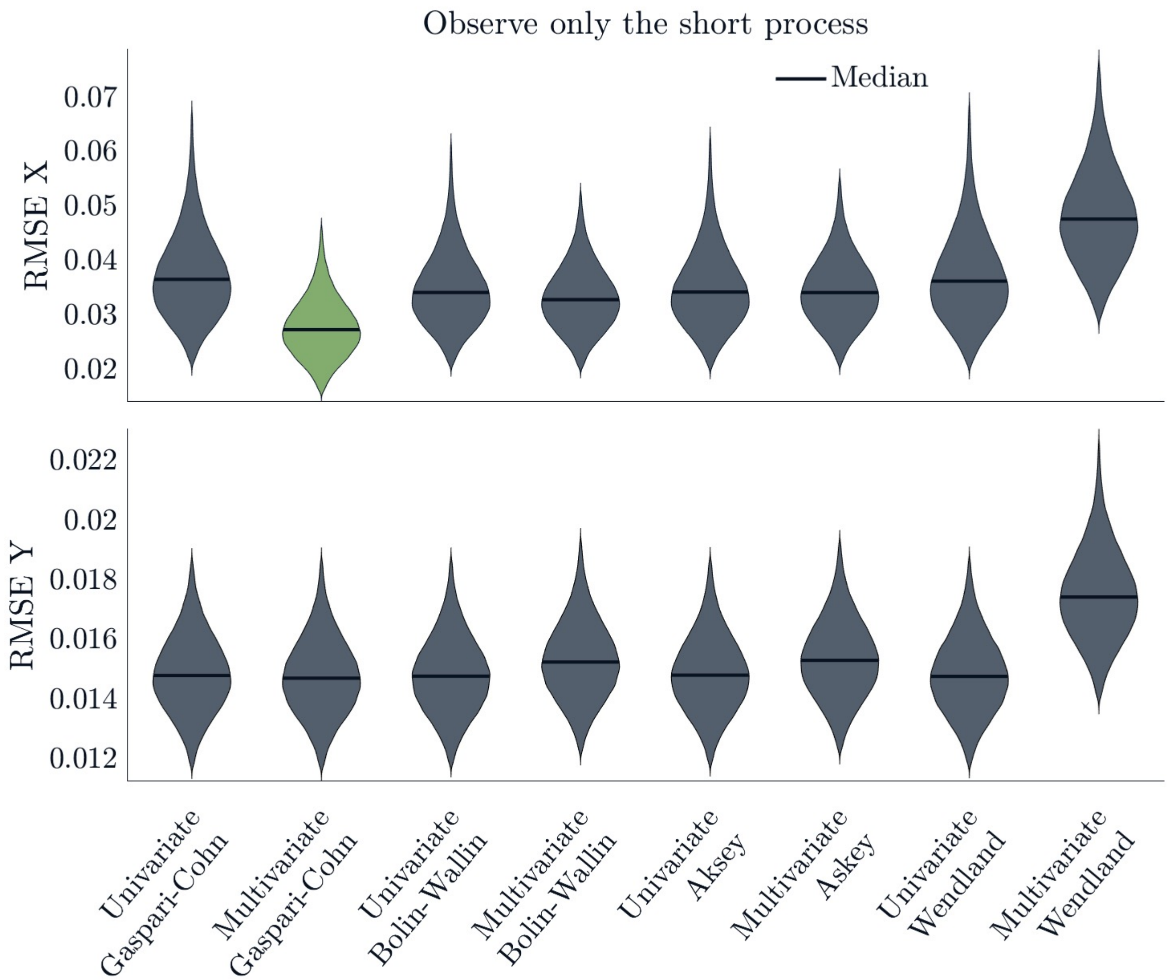
~~For univariate Askey we need only choose the parameter ν in equation (9). For positive semidefiniteness, we require $\nu \geq 1$. Smaller values of ν allow for larger cross-localization weights and longer effective cross-localization radii (Fig. ??), both of which are desirable and improve performance in sensitivity experiments (not shown). We choose to use~~

Function name	Univariate parameters	Multivariate parameters
Gaspari-Cohn	$R = 20$	$R_{YY} = 20$, ~~$\gamma_{XY} = \frac{R_{XY}}{2}\left(\frac{\gamma_{XX}}{R_{XX}} + \frac{\gamma_{YY}}{R_{YY}}\right)$. Note that the values of γ enter into the multivariate Askey fun~~
Bolin-Wallin	$R = 20$	$R_{YY} = 20$, $R_{XX} = 40$, $R_{XY} = 30$, $\beta \approx 0.35$
Askey	$R = 20, \nu = 1$	$R_{YY} = 20$, $R_{XX} = 40$, $R_{XY} = 20$, $\beta \approx 0.41$
		$\nu = 1$, $\gamma_{YY} = 2$, $\gamma_{XX} = 0$~~or $\gamma_{YY} = 0$. Increasing γ_{XX} while keeping $\gamma_{YY} = 0$ serves to decrease~~
Wendland	$R = 20, \nu = 2, k = 1$	$R_{YY} = 20$, $R_{XX} = 40$, $R_{XY} = 20$, $\beta \approx 0.14$
		$\nu = 2$, $\gamma_{YY} = 2$, $\gamma_{XX} = 0$~~decreases the effective localization radius of process Y and hence decrea~~

Table B1. Localization parameters for the experiment where we observe only the "long" process.

Function name	Univariate parameters
Gaspari-Cohn	$R = 15$
Bolin-Wallin	$R = 15$
Askey	$R = 15, \nu = 1$
Wendland	$R = 15, \nu = 2, k = 1$~~here as higher values lead to unreasonably small cross-localization weight factors and hence degraded perform~~

Table B2. Localization parameters for the experiment where we observe both processes.



~~Histograms of RMSE for the X process with different localization functions. In each plot the blue histogram shows the distribution of RMSE when univariate localization is used. The red histogram shows the distribution when multivariate localization is used. Insets in each panel show the cross-localization functions in the case of univariate (blue) and multivariate (red) localization. All four univariate localization functions perform similarly with median RMSE ranging from 0.034 to 0.036, while there is a greater range in performance for the multivariate versions of these functions. Multivariate Gaspari-Cohn shows improvement over its univariate counterparts. Univariate and multivariate Bolin-Wallin and Askey functions appear to perform similarly. For Wendland, the multivariate function performs significantly worse than the univariate function.~~

Figure 3. Violin plots show the distribution of RMSE for the X and Y process with different localization functions (Hoffmann, 2015). All four univariate localization functions perform similarly, while there is a greater range in performance for the multivariate versions of these functions. Multivariate Gaspari-Cohn shows improvement over its univariate counterparts. Univariate and multivariate Bolin-Wallin and Askey functions appear to perform similarly. For Wendland, the multivariate function performs significantly worse than the univariate function.

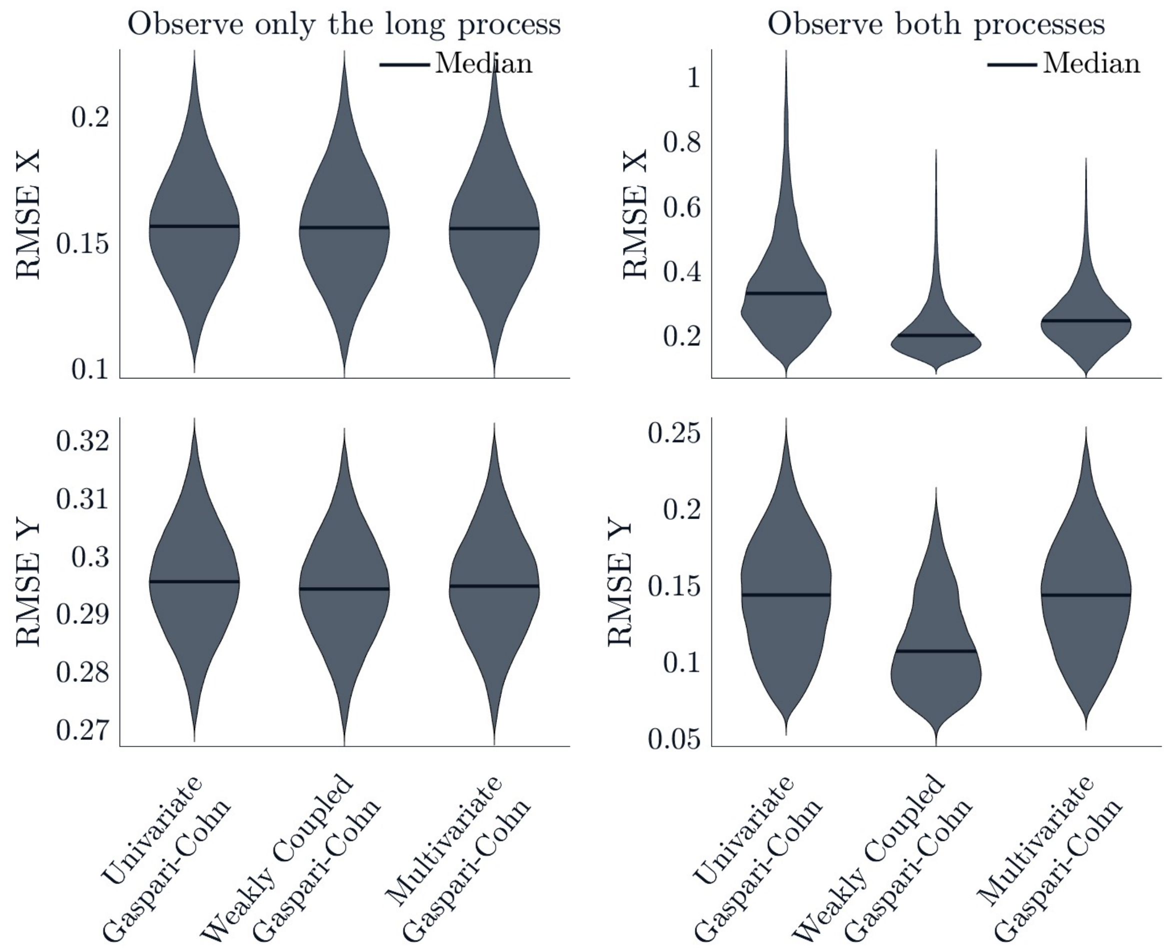


Figure 4. Violin plots show the distribution of RMSE for all versions of the Gaspari-Cohn localization function (Hoffmann, 2015). Left: results from the experiment where we observe only the "long" process. All functions perform similarly. Right: results from the experiment where we observe both processes. The weakly coupled localization functions appear to lead to the best performance, but are highly unstable.

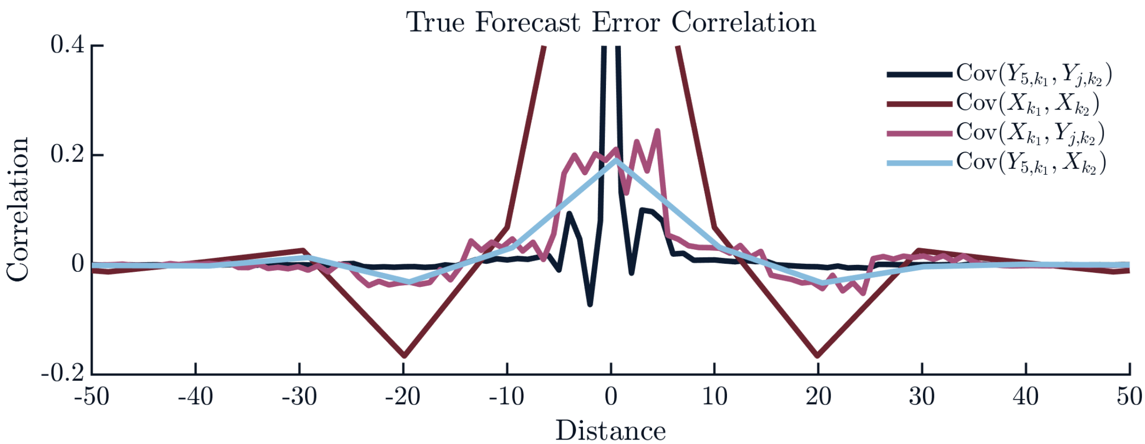


Figure B1. True forecast error correlations for variables in the middle of each sector, X_k and $Y_{5,k}$. Correlations between Y variables (dark blue) decay to zero after about 5 spatial units, while correlations between X variables (dark red) are significant up to 40 spatial units away. Cross-correlations (~~yellow~~ pink and ~~purple~~light blue) are small everywhere, but still significant out to at least 20 spatial units.

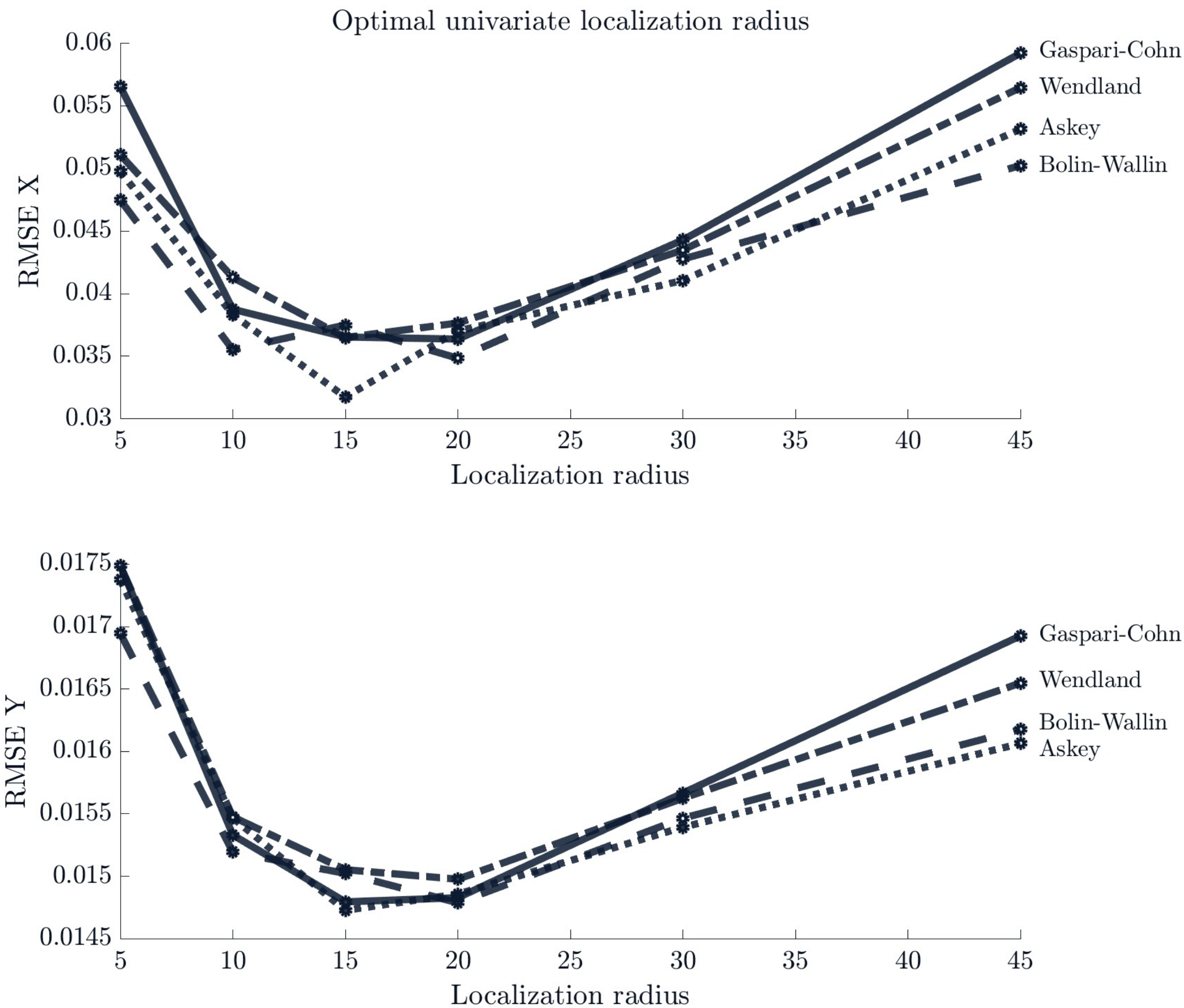


Figure B2. RMSE for different univariate localization radii. Considering all functions, the best performance comes when $R = 15$.

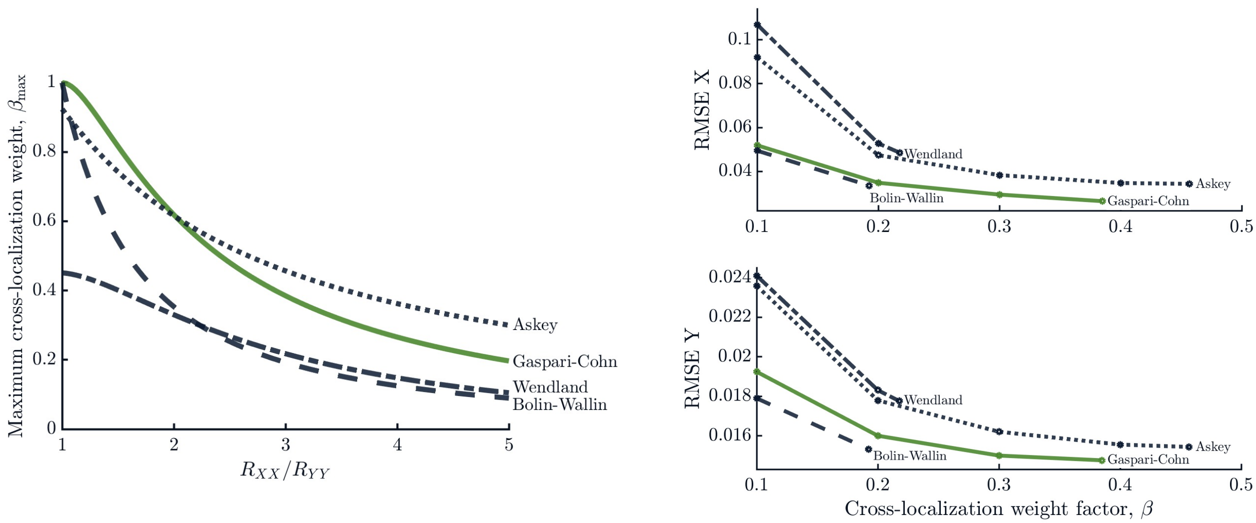


Figure B3. Left: Maximum cross-localization weight factor as a function of R_{XX}/R_{YY}. Right: Average RMSE ~~for the X process~~ is shown on the y-axis for different multivariate functions. The top (bottom) plot shows RMSE for the X (Y) process. For all functions, as the cross-localization weight factor increases, the analysis errors (RMSE~~decreases~~) decrease.

Code availability. Code used for this study is available in the GitHub repository: https://github.com/zcstanley/Multivariate_Localization_Functions.

Author contributions. The authors all contributed to the conceptualization and formal analysis. ZS and IG contributed to software development. ZS performed the experiments and analysis of the results, and prepared the manuscript, which was edited by IG and WK.

Competing interests. The authors declare they have no conflict of interest.

Acknowledgements. ZS is supported by NSF DGE-1650115. IG is supported by NSF DMS-1821074. WK is supported by NSF DMS-1811294 and NSF DMS-1821074. This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University.

References

Anderson, J. L.: Localization and sampling error correction in ensemble Kalman filter data assimilation, Monthly Weather Review, 140, 2359–2371, 2012.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, Quarterly Journal of the Royal Meteorological Society, 134, 1951–1970, https://doi.org/10.1002/qj.339, 2008.

Bishop, C. H. and Hodyss, D.: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation, Quarterly Journal of the Royal Meteorological Society, 133, 2029–2044, 2007.

Bolin, D. and Wallin, J.: Spatially adaptive covariance tapering, Spatial Statistics, 18, 163–178, https://doi.org/10.1016/j.spasta.2016.03.003, 2016.

Buehner, M. and Shlyaeva, A.: Scale-dependent background-error covariance localisation, Tellus A: Dynamic Meteorology and Oceanography, 67, 28 027, https://doi.org/10.3402/tellusa.v67.28027, 2015.

Burgers, G., Leeuwen, P. J. v., and Evensen, G.: Analysis Scheme in the Ensemble Kalman Filter, Monthly Weather Review, 126, 1719–1724, https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2, 1998.

Daley, D. J., Porcu, E., and Bevilacqua, M.: Classes of compactly supported covariance functions for multivariate random fields, Stochastic Environmental Research and Risk Assessment, 29, 1249–1263, https://doi.org/10.1007/s00477-014-0996-y, 2015.

Daley, R.: Atmospheric Data Analysis, Cambridge Atmospheric and Space Science Series, Cambridge University Press, 1993.

Dormand, J. and Prince, P.: A family of embedded Runge-Kutta formulae, Journal of Computational and Applied Mathematics, 6, 19–26, https://doi.org/10.1016/0771-050X(80)90013-3, 1980.

El Gharamti, M.: Enhanced Adaptive Inflation Algorithm for Ensemble Filters, Monthly Weather Review, 146, 623–640, https://doi.org/10.1175/MWR-D-17-0187.1, 2018.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, https://doi.org/10.1029/94JC00572, 1994.

Frolov, S., Bishop, C. H., Holt, T., Cummings, J., and Kuhl, D.: Facilitating Strongly Coupled Ocean–Atmosphere Data Assimilation with an Interface Solver, Monthly Weather Review, 144, 3–20, https://doi.org/10.1175/MWR-D-15-0041.1, 2016.

Frolov, S., Reynolds, C. A., Alexander, M., Flatau, M., Barton, N. P., Hogan, P., and Rowley, C.: Coupled Ocean–Atmosphere Covariances in Global Ensemble Simulations: Impact of an Eddy-Resolving Ocean, Monthly Weather Review, 149, 1193–1209, https://doi.org/10.1175/MWR-D-20-0352.1, 2021.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, https://doi.org/10.1002/qj.49712555417, 1999.

Genton, M. G. and Kleiber, W.: Cross-Covariance Functions for Multivariate Geostatistics, Statistical Science, 30, 147–163, https://doi.org/10.1214/14-STS487, 2015.

Gneiting, T.: Compactly Supported Correlation Functions, Journal of Multivariate Analysis, 83, 493–508, https://doi.org/10.1006/jmva.2001.2056, 2002.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R.: Balance and Ensemble Kalman Filter Localization Techniques, Monthly Weather Review, 139, 511–522, https://doi.org/10.1175/2010MWR3328.1, 2011.

Hoffmann, H.: Violin Plot, https://www.mathworks.com/matlabcentral/fileexchange/45134-violin-plot, 2015.

Horn, R. A. and Johnson, C. R.: Matrix analysis, Cambridge University Press, Cambridge ; New York, 2nd edn., 2012.

Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, Monthly Weather Review, 126, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 129, 15, 2001.

Lorenz, E. N.: Predictability—A problem partly solved, vol. 1, pp. 1–18, Proc. Seminar on Predictability, ECMWF, Reading, Berkshire, United Kingdom, 1996.

Lu, F., Liu, Z., Zhang, S., and Liu, Y.: Strongly Coupled Data Assimilation Using Leading Averaged Coupled Covariance (LACC). Part I: Simple Model Study, Monthly Weather Review, 143, 3823–3837, https://doi.org/10.1175/MWR-D-14-00322.1, 2015.

Morss, R. E. and Emanuel, K. A.: Influence of added observations on analysis and forecast errors: Results from idealized systems, Quarterly Journal of the Royal Meteorological Society, 128, 285–321, https://doi.org/10.1256/00359000260498897, 2002.

Ménétrier, B., Montmerle, T., Michel, Y., and Berre, L.: Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part I: Optimality Criteria and Application to Variance Filtering and Covariance Localization, Monthly Weather Review, 143, 1622–1643, https://doi.org/10.1175/MWR-D-14-00157.1, 2015.

Penny, S. G., Bach, E., Bhargava, K., Chang, C.-C., Da, C., Sun, L., and Yoshida, T.: Strongly Coupled Data Assimilation in Multiscale Media: Experiments Using a Quasi-Geostrophic Coupled Model, Journal of Advances in Modeling Earth Systems, 11, 1803–1829, https://doi.org/10.1029/2019MS001652, 2019.

Porcu, E., Daley, D. J., Buhmann, M., and Bevilacqua, M.: Radial basis functions with compact support for multivariate geostatistics, Stochastic Environmental Research and Risk Assessment, 27, 909–922, https://doi.org/10.1007/s00477-012-0656-z, 2013.

Roh, S., Jun, M., Szunyogh, I., and Genton, M. G.: Multivariate localization methods for ensemble Kalman filtering, Nonlinear Processes in Geophysics, 22, 723–735, https://doi.org/10.5194/npg-22-723-2015, 2015.

Shampine, L. F. and Reichelt, M. W.: The MATLAB ODE Suite, SIAM Journal on Scientific Computing, 18, 1–22, https://doi.org/10.1137/S1064827594276424, 1997.

Shen, Z., Tang, Y., Li, X., Gao, Y., and Li, J.: On the localization in strongly coupled ensemble data assimilation using a two-scale Lorenz model, preprint, Predictability, Data Assimilation/Climate, Atmosphere, Ocean, Hydrology, Cryosphere, Biosphere, https://doi.org/10.5194/npg-2018-50, 2018.

Sluka, T. C., Penny, S. G., Kalnay, E., and Miyoshi, T.: Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation, Geophysical Research Letters, 43, 752–759, https://doi.org/10.1002/2015GL067238, 2016.

Smith, P. J., Lawless, A. S., and Nichols, N. K.: Estimating Forecast Error Covariances for Strongly Coupled Atmosphere–Ocean 4D-Var Data Assimilation, Monthly Weather Review, 145, 4011–4035, https://doi.org/10.1175/MWR-D-16-0284.1, 2017.

Smith, P. J., Lawless, A. S., and Nichols, N. K.: Treating Sample Covariances for Use in Strongly Coupled Atmosphere-Ocean Data Assimilation, Geophysical Research Letters, 45, 445–454, https://doi.org/10.1002/2017GL075534, 2018.

Smith, P. J., Lawless, A. S., and Nichols, N. K.: The role of cross-domain error correlations in strongly coupled 4D-Var atmosphere–ocean data assimilation, Quarterly Journal of the Royal Meteorological Society, 146, 2450–2465, https://doi.org/10.1002/qj.3802, 2020.

Wang, X., Chipilski, H. G., Bishop, C. H., Satterfield, E., Baker, N., and Whitaker, J. S.: A Multiscale Local Gain Form Ensemble Transform Kalman Filter (MLGETKF), Monthly Weather Review, 149, 605–622, https://doi.org/10.1175/MWR-D-20-0290.1, 2021.

Wilks, D. S.: Effects of stochastic parametrizations in the Lorenz '96 system, Quarterly Journal of the Royal Meteorological Society, 131, 389–407, https://doi.org/10.1256/qj.04.03, 2005.

Ying, Y., Zhang, F., and Anderson, J. L.: On the Selection of Localization Radius in Ensemble Filtering for Multiscale Quasigeostrophic Dynamics, Monthly Weather Review, 146, 543–560, https://doi.org/10.1175/MWR-D-17-0336.1, 2018.

Yoshida, T. and Kalnay, E.: Correlation-Cutoff Method for Covariance Localization in Strongly Coupled Data Assimilation, Monthly Weather Review, 146, 2881–2889, https://doi.org/10.1175/MWR-D-17-0365.1, 2018.