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 non-negative definite localization matrices, which are suitable for use in 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 analogs of all four functions in a simple experiment with the bivariate Lorenz '96 system. In our experiment the multivariate Gaspari-Cohn function leads to better performance than any of the other localization functions.

1 Introduction

An essential part of any data assimilation (DA) method is the estimation of the background error covariance matrix $\mathbf{B}$. The background error covariance statistics stored in $\mathbf{B}$ dictate how information from observations propagates through the domain during the assimilation step (Bannister, 2008). A poorly designed $\mathbf{B}$ 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 $\mathbf{B}$ matrix is estimated through an ensemble average. Using an ensemble to estimate $\mathbf{B}$ allows the estimates of the background error statistics to change with the model state, which is desirable in many geophysical systems. However, this estimate of $\mathbf{B}$ 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 an ensemble estimate of **B** through a Schur (or element-wise) product. The Schur product theorem (Horn and Johnson, 2012, Theorem 7.5.3) guarantees that if the localization matrix is non-negative definite, then the localized estimate of **B** is also non-negative definite. Non-negative definiteness of estimates of **B** is generally desirable in all DA schemes and is essential in variational schemes.

30   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 (Penny et al., 2017; Zhang et al., 2020). However, coupled DA systems face unique challenges as they involve simultaneous analysis of multiple domains spanning different spatiotemporal scales. Cross-domain correlations in particular are found to be spatially

35   inhomogeneous (Penny et al., 2017). Schemes that include cross-domain correlations in the **B** matrix are broadly classified as strongly coupled, which is distinguished from weakly coupled schemes where **B** does not include any nonzero cross-domain correlations. The inclusion of cross-domain correlations in **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,

40   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 localiza-

45   tion functions for each component and are non-negative 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 non-negative definite (Frolov et al., 2016; Smith et al., 2018; Shen et al., 2018). Frolov et al. (2016) report that their proposed localization matrix is experimentally non-negative definite for some length scales. Smith et al. (2018) use a similar method and find cases in which their localization matrix is not non-negative definite.

50   In this work, we build on these methods and investigate distance-based, multivariate, non-negative definite 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 non-negative definite for all length scales and hence appropriate for 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).

55   We investigate the behavior of these functions in a simple bivariate model proposed by Lorenz (1995). 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, artificially decreasing the magnitude of the cross-domain correlation hinders the assimilation of observations. We compare all of the multivariate functions to their univariate analogs and observe that the new multivariate extension of GC outperforms all competitors.

60   This paper is organized as follows. In Sect. 2 we present two new multivariate localization functions and two multivariate localization function from the literature. In Sect. 3 we describe the set up for the experiment whose results are presented in Sect. 4.

2   Multivariate localization functions

2.1   Multivariate localization background

65   Consider the background error covariance matrix $\mathbf{B}$ of a strongly coupled DA scheme with interacting model components X and Y. The $\mathbf{B}$ matrix may be written in terms of background error covariances for components X and Y ($\mathbf{B_{XX}}$ and $\mathbf{B_{YY}}$) and cross-domain covariances between X and Y ($\mathbf{B_{XY}}$ and $\mathbf{B_{YX}}$). Strongly coupled DA is characterized by the inclusion of nonzero cross-domain covariances in $\mathbf{B_{XY}}$ and $\mathbf{B_{YX}}$. Similarly to $\mathbf{B}$, 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

70
$$\mathbf{L} \circ \mathbf{B} = \begin{bmatrix} \mathbf{L_{XX}} & \mathbf{L_{XY}} \\ \mathbf{L_{YX}} & \mathbf{L_{YY}} \end{bmatrix} \circ \begin{bmatrix} \mathbf{B_{XX}} & \mathbf{B_{XY}} \\ \mathbf{B_{YX}} & \mathbf{B_{YY}} \end{bmatrix}, \tag{1}$$

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 $\mathbf{B}_{ij}$ is the sample covariance $\mathrm{Cov}(X_i, X_j)$ where $X_i = X(\mathbf{s_i})$ denotes the 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 $\mathcal{L}(d) = 0$.

75   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 $\mathbf{L_{XY}}$ and $\mathbf{L_{YX}}$ such that, $\mathcal{L}_{XX}, \mathcal{L}_{YY}$, and $\mathcal{L}_{XY}$ together are a multivariate, non-negative definite function in the sense that they produce a non-negative definite $\mathbf{L}$ matrix (Genton and

80   Kleiber, 2015). In this study we compare four different multivariate localization functions, including one that extends GC.

    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 $\mathcal{L}_{XY}(d) = 0$. The cross-localization radius depends on the localization radii R_{XX}, R_{YY} associated with components X, Y. We define the *cross-localization weight factor*, $\beta \geq 0$, as the value of the cross-localization function at

85   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 non-negative definiteness (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 non-negative definite, while values of β less than $\beta_{\max}$ are allowable and may be useful in attenuating

90   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.

95   ## 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 and 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 $\mathcal{L}_{XY}(d) =$

100 $[k_X * k_Y](\mathbf{d})$ is a compatible cross-localization function in the sense that, when taken together, $\mathcal{L}_{XX}, \mathcal{L}_{YY}$, and $\mathcal{L}_{XY}$ form a multivariate, non-negative definite function.

As a proof, we define two processes Z_j, $j = X, Y$, 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

105 processes, $\mathcal{L}_{ij}(d) = \mathrm{Cov}(Z_i(\mathbf{s}), Z_j(\mathbf{t}))$, $i, j = X, Y$, with $\mathbf{s}, \mathbf{t} \in \mathbb{R}^n$ and $d = \|\mathbf{s} - \mathbf{t}\|$. Thus $\mathcal{L}_{XX}, \mathcal{L}_{YY}, \mathcal{L}_{XY}$ form a multivariate covariance function, and hence a multivariate, non-negative definite function.

For multivariate localization functions defined in this manner, the cross-localization radius is the average of the two localization radii for model components X and Y, $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$. Note that if $\mathcal{L}_{XX}, \mathcal{L}_{YY}, \mathcal{L}_{XY}$ is a multivariate, non-negative definite function, then so is $\mathcal{L}_{XX}, \mathcal{L}_{YY}, \beta\mathcal{L}_{XY}$ for $\beta < 1$ (Roh et al., 2015). Thus, we may choose to attenuate

110 the cross-domain covariances through a cross-localization weight factor β. To aid in comparisons to other cross-localization functions, we re-define kernel-based cross-localization functions as,

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

where $\beta_{\max} = [k_X * k_Y](\mathbf{0})$, and $\beta \leq \beta_{\max}$.

For most kernels, closed form analytic expressions for the convolutions above are not available. In the following two sections

115 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. The kernel is normalized so that $\mathcal{L}(0) = 1$, $z_+ = \max\{z, 0\}$, and c is the radius of the kernel. The resulting con-

120 volution is zero at distances greater than $2c$, where the two kernels are entirely non-overlapping. Thus c is referred to as the
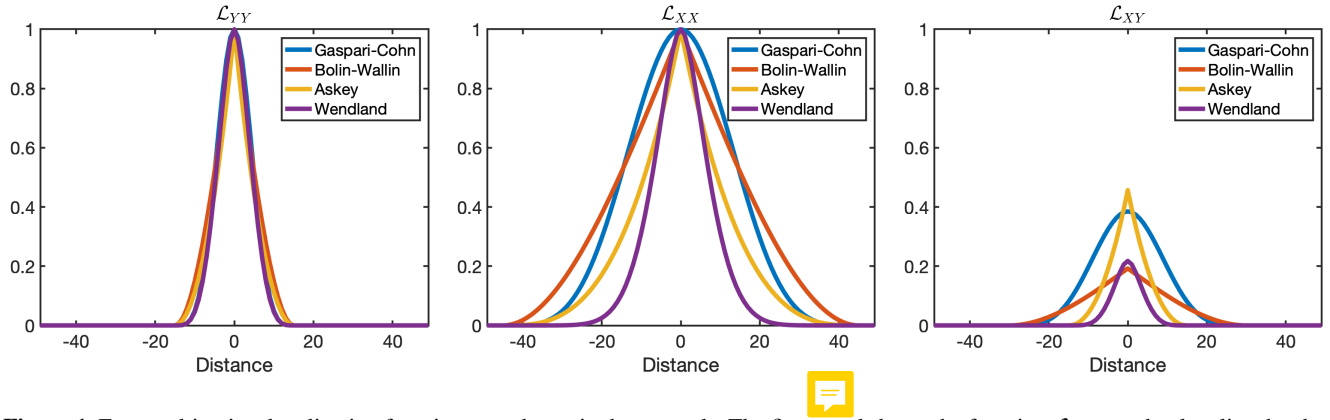
Figure 1. Four multivariate localization functions are shown in three panels. The first panel shows the function $\mathcal{L}_{YY}$ used to localize the short process, Y. The second panel shows the function $\mathcal{L}_{XX}$ used to localize the long process, X. The third panel shows the cross-localization function $\mathcal{L}_{XY}$. In each panel, the color of the line shows the different multivariate functions: Gaspari-Cohn (blue), Bolin-Wallin (red), Askey (yellow), and Wendland (purple). In the case of univariate localization, the functions in the left panel are used to localize all processes.

localization half-width, and is related to the localization radius by $c = R/2$. 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 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)

125 with half-widths $c_j = R_{jj}/2, j = X, Y$. The derivation of and resulting formula for $\mathcal{L}_{XY}^{(GC)}(d) = [k_X * k_Y](\mathbf{d})$ are quite lengthy and are thus included in Appendix A. The maximum cross-localization weight factor for the multivariate GC function is $\beta_{\max} = \frac{5}{2}\kappa^{-3} - \frac{3}{2}\kappa^{-5}$, where $\kappa^2 = \frac{\max\{R_{XX}, R_{YY}\}}{\min\{R_{XX}, R_{YY}\}}$. The support for the cross-localization function is $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$. 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.

130 **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 $r \leq c_j$ and 0 otherwise. The resulting univariate localization function for

135 within-process localization is

$$\mathcal{L}_{jj}^{(BW)}(d) = \frac{3}{2\pi c_j^3} \cdot V_3\left(c_j, \frac{d}{2}\right) \text{ if } 0 \leq d < R_{jj} = 2c_j, \quad j = X, Y. \tag{6}$$

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

Table 1. Selected univariate Wendland functions

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

$$V_3(r,x) = \begin{cases} \frac{\pi}{3}(r-x)^2(2r+x) & |x| < r \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The label (BW) references Bolin and Wallin, who perform the necessary convolutions in a work aimed at a different application
140 of covariance functions (Bolin and Wallin, 2016). While Bolin and Wallin never develop 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).

Let $c_X > c_Y$ be localization half-widths, then the resulting cross-localization function is,

$$\mathcal{L}_{XY}^{(BW)}(d) = \frac{\beta}{\beta_{\max}} \cdot \frac{3}{4\pi \left(c_X c_Y\right)^{3/2}} \cdot \begin{cases} \frac{4\pi c_Y^3}{3} & \text{if } d < c_X - c_Y \\ V_3\left(c_X, \frac{d^2 + c_X^2 - c_Y^2}{2d}\right) + V_3\left(c_Y, \frac{d^2 + c_Y^2 - c_X^2}{2d}\right) & \\ & \text{if } c_X - c_Y \le d < c_X + c_Y. \end{cases} \tag{8}$$

145 where $\beta_{\max} = \kappa^{-3}$, $\kappa^2 = \frac{\max\{R_{XX}, R_{YY}\}}{\min\{R_{XX}, R_{YY}\}}$. As with the multivariate GC, the cross-localization radius for BW is $R_{XY} = \frac{1}{2}(R_{XX} + R_{YY})$. Note that there is a typo in Bolin and Wallin (2016), which has been corrected here. 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,
150 non-negative definite 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

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

Other functions in this family are called Wendland functions. Several examples of univariate Wendland functions are displayed
155 in Table 1.

Porcu et al. (2013) develop a multivariate version of the Askey function, where the exponent in equation (9) is allowed to vary by process and the localization radius R is constant across all processes. Roh et al. (2015) find that this multivariate

localization function outperforms common univariate localization methods when assimilating observations into the bivariate Lorenz 96 model. Daley et al. (2015) extend the work of Porcu et al. (2013) and construct a multivariate version of general

160 Wendland-Gneiting functions that allows for different localization radii for different processes. The multivariate 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,

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

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

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

with B the beta function. Note that the Askey function in equation (10) is a special case of the Wendland function (11) which corresponds to the case $k = 0$. Daley et al. (2015) give sufficient conditions on the parameters ν, k, R_{ij}, γ_{ij}, and β to guarantee that equation (11), and hence (10), is non-negative definite. In particular for two processes X, Y, equation (11) is non-negative

170 definite 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

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

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 non-negative definite functions.

175 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.

3 Experimental design

We compare the four multivariate localization functions in Sect. 2 to a simple approach to localization in coupled DA, which is to use the same localization function for all model components. We call this approach *univariate localization*. A second simple

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 functions.

approach is to use different localization functions for each process and then zero out all cross-correlations between processes. This approach leads to a "weakly" coupled scheme, which is not the focus of this work. Additionally, in our setup we observe
185 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.

3.1 Bivariate Lorenz model

The bivariate Lorenz 96 model is a simple model of two coupled variables with distinct temporal and spatial scales. This model
190 is a conceptual model of atmospheric processes, where the "short" process can be thought of as rapidly-varying small-scale convective fluctuations and 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 "long" process with 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,

$$195 \quad \frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k - (ha/b)\sum_{j=1}^{J} Y_{j,k} + F \tag{14}$$

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

The Y process is divided into K sectors, with each sector corresponding to one "long" variable X_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.
200 We follow Lorenz (1995) and let $K = 36, J = 10$, so there are 10 times as many "long" variables as "short" 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 (1995) 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}. The solutions are output with a time interval of 0.005 nondimensional units,
205 or 36 minutes, as in Lorenz (1995). This time scale is 10 times shorter than the time scale typically used in the univariate Lorenz 96 model (0.05 nondimensional units, or around 6 hours). The factor of 10 is consistent with the understanding that the
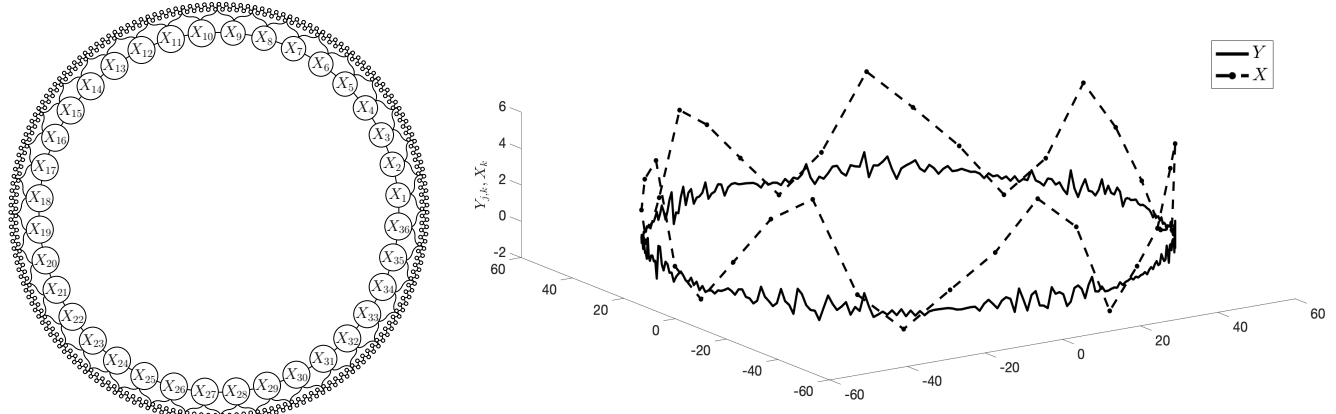
Figure 2. Left: Schematic illustrating the location of the different variables in the bivariate Lorenz 96 model, inspired by Wilks (2005). The set up has $K = 36$ sectors, with $J = 10$ "short" process variables per sector. The "long" process is shown on the inner circle and each X variable is labeled. 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 dashed line and has amplitude about 10 times larger than the "short" process Y (solid black line).

"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 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.

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 scheme

We develop localization functions for EnVar schemes where non-negative definiteness of the localization matrix is essential to ensure convergence of the numerical optimization. Since the minimizer of the 3D-EnVar objective function is the same as the EnKF analysis mean in the case of linear observation (Lorenc, 1986), in this experiment we make use of the EnKF rather than implement an ensemble of 3D-EnVar assimilation scheme (Evensen, 1994; Houtekamer and Mitchell, 1998; Burgers et al., 1998). The EnKF update formula for a single ensemble member is

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

where $\mathbf{x}^{\mathrm{a}}$ is the analysis vector, $\mathbf{x}^{\mathrm{b}}$ is the background state vector, $\mathbf{y}$ is the observation, each element of $\boldsymbol{\eta}$ is a random draw

225 from the probability distribution of observation errors, and $\mathbf{H}$ is the linear observation operator. The Kalman gain matrix $\mathbf{K}$ is

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

where $\mathbf{B}$ 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

230 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. We initialize the inflation factor at $\lambda = 1.1$ everywhere, and the inflation variance at $\sigma_\lambda^2 = 0.09$.

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 time steps, discarding the first 1,000 time steps and reporting statistics from the remaining 2,000 time steps. Each experiment is repeated 50

235 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 $\sigma_Y^2 = 0.005$, which is about 5% of the climatological variance of the Y process.

3.3 Univariate vs. multivariate setup

240 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 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 cross-localization radius equal to $R_{XY} = 30$. For multivariate Askey and Wendland the cross-localization radius is

245 $R_{XY} = 15$. These localization radii are chosen in accordance with sensitivity experiments, described in Appendix B1.

For all functions we use the maximum allowable 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 cross-localization weight factors lead to degraded performance. Details of the sensitivity experiments involving β are provided in Appendix B2. We hypothesize that an important factor in the performance of the multivariate localization functions is the size of $\beta_{\max}$ and

250 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 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

255 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 that 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 in Table 3 and described in further detail in Appendix B3.

Function name	Univariate parameters	Multivariate parameters	Univariate median RMSE	Multivariate median RMSE
Gaspari-Cohn	$R = 15$	$R_{YY} = 15, R_{XX} = 45, \boldsymbol{R_{XY} = 30}, \boldsymbol{\beta \approx 0.38}$	0.036	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 for the experiment comparing univariate and multivariate localization

4 Univariate vs. multivariate results

260 Figure 3 shows the RMSE for process X for the localization functions defined in Table 3. Each panel compares performance with univariate and multivariate versions of a function. The performance of 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 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

265 test the multivariate functions. The 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 BW and Askey functions show virtually no difference between the multivariate and univariate versions. The Wendland function shows significantly worse performance with the multivariate function when compared to the univariate version.

As noted in Sect. 3.3 the shape of the GC cross-localization function appears to allow for the most cross-domain localization

270 because it is nearly as tall as Askey, fairly smooth near $d = 0$, and it decays to zero more slowly than any other function, with the possible exception of BW. 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

275 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.
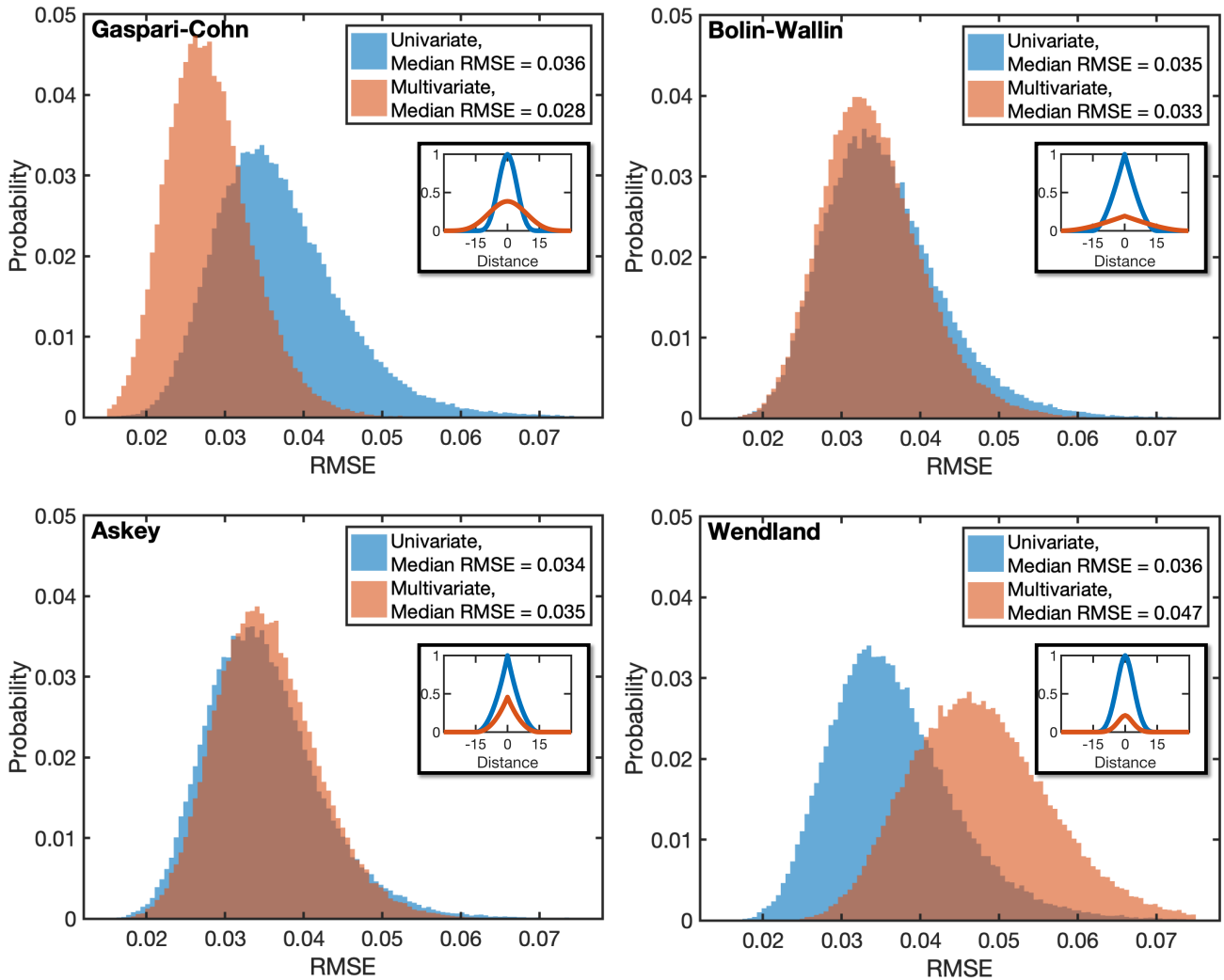
Figure 3. Histograms of RMSE for the X process with different localization functions. In each plot the blue histogram shows the distribution of RMSE when univariate localization is used. The red histogram shows the distribution when multivariate localization is used. Insets in each panel show the cross-localization 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.

5 Conclusions

280

In this work we develop 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 non-negative definite. A non-negative definite localization matrix guarantees, through the Schur product theorem, that the localized estimate of the background covariance matrix is non-negative defi-

285 nite (Horn and Johnson, 2012, Theorem 7.5.3). We compare multivariate GC to three other multivariate localization functions (including one other newly presented multivariate function), and their univariate counterparts. We find that, in a toy model, multivariate GC leads to better performance than any of the other localization functions we consider. There is still an outstanding question of how multivariate GC will perform in other, perhaps more realistic, systems.

In this work we investigate the importance of the cross-localization weight factor, which determines the amount of informa-

290 tion which is allowed to propagate between co-located variables in different model components. We find that this parameter should be as large as possible. This is likely unique to our setup as other studies 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-

295 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 discuss the importance of the cross-localization radius in determining performance. However, this work does not address the question of optimal cross-localization radius selection, which is an important area for future research.

Appendix A: Derivation of multivariate Gaspari-Cohn

A1 Multivariate Gaspari-Cohn cross-localization function

300

Let c_X, c_Y be the localization half-widths associated with model components X and Y. Without loss of generality, we may take $c_X > c_Y$. We find there are two different cases to consider, when $c_X < 2c_Y$ and when $c_X \geq 2c_Y$. In both cases, the notation is significantly simplified when we let $c_X = \kappa^2 c_Y$. The first case we consider is when $c_Y < c_X < 2c_Y$. In this case, the GC

cross-localization function is,

$$
305 \quad \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{aligned}
&-\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 \\
&-\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^4} - \frac{1}{\kappa^2}\right) \\
&+\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|} \\
&-\frac{3}{4}\left(\kappa^5 + \frac{1}{\kappa^5}\right) + \frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right)
\end{aligned} & c_X - c_Y < |d| < c_Y \\[2ex]
\begin{aligned}
&-\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 \\
&-\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^4} - \frac{1}{\kappa^2}\right) \\
&+\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|} \\
&-\frac{3}{4}\left(\kappa^5 - \frac{1}{\kappa^5}\right) + \frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right)
\end{aligned} & c_Y < |d| < c_X \\[2ex]
\begin{aligned}
&\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 \\
&+\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^4} + \frac{1}{\kappa^2}\right) \\
&-\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|} \\
&+\frac{3}{4}\left(\kappa^5 + \frac{1}{\kappa^5}\right) + \frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right)
\end{aligned} & c_X < |d| < c_Y + c_X.
\end{cases} \tag{A1}
$$

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

The second case we 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 \\[4mm] -\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 \\[4mm] \begin{aligned} &-\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 \\ &-\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) \\ &+\frac{5}{12}\frac{\kappa c_Y}{|d|} - \frac{3}{8}\frac{\kappa c_Y}{|d|}\left(\kappa^4 + \frac{1}{\kappa^4}\right) + \frac{1}{6}\frac{\kappa c_Y}{|d|}\left(\kappa^6 - \frac{1}{\kappa^6}\right) \\ &+\frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right) - \frac{3}{4}\left(\kappa^5 - \frac{1}{\kappa^5}\right) \end{aligned} & c_X - c_Y \le |d| < c_X \\[8mm] \begin{aligned} &\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 \\ &+\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) \\ &+\frac{5}{12}\frac{\kappa c_Y}{|d|} - \frac{3}{8}\frac{\kappa c_Y}{|d|}\left(\kappa^4 + \frac{1}{\kappa^4}\right) - \frac{1}{6}\frac{\kappa c_Y}{|d|}\left(\kappa^6 + \frac{1}{\kappa^6}\right) \\ &+\frac{5}{4}\left(\kappa^3 + \frac{1}{\kappa^3}\right) + \frac{3}{4}\left(\kappa^5 + \frac{1}{\kappa^5}\right) \end{aligned} & c_X \le |d| < c_Y + c_X \end{cases} \tag{A2}$$

310 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 $c_X = 2c_Y$, and hence $\kappa = \sqrt{2}$, both (A1) and (A2) are the same.

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 frame-

315 work 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 $\mathbb{R}^3$,

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

can be written as,

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

320 The normalization factor $P_{jj}^0(0)$ is equal to

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

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

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

Now we compute the normalization factor $P^0_{jj}(0)$ using the GC kernels. Plugging in $k^0_j(r) = (1 - r/c_j)_+$ gives,

$$325 \quad P^0_{jj}(0) = 4\pi \int_0^{c_j} r^2 (1 - r/c_j)^2 \, \mathrm{d}r = \frac{2\pi}{15} c_j^3, \quad j = X, Y. \tag{A7}$$

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.

The case $c_X > 2c_Y$ and $0 \le |d| < c_Y$ is shown in detail here. The other cases are derived similarly and are not shown. The

330 inner integral in equation (A4) is

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

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

$$335 \quad \frac{1}{6} d c_Y^3 - \frac{1}{3c_X} d \left[\frac{1}{30 c_Y} d^5 - \frac{1}{10} d^4 + \frac{1}{3} c_Y^2 d^2 + \frac{3}{10} c_Y^4 \right]. \tag{A11}$$

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,

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

340 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 localization

345 half widths by $c_j = R_{jj}/2$ and kernels $k_j(r) \propto (1 - r/c_j)_+$. Then the localization function between process X_i and X_j is
$\mathcal{L}_{ij}(d) = \alpha_{ij}[k_i * k_j](\mathbf{d})$, with

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

a non-negative definite matrix with 1's on the diagonal. 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 $\max\{c_i, c_j\} < 2\min\{c_i, c_j\}$ or Eq. (A2) otherwise. We have written (A1) and (A2) with a

350 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 non-negative definite. The simplest case is to let $\alpha_{ij} = 1$ for all i, j.

Appendix B: Sensitivity experiments

B1   Localization radius

355 A fair comparison between the univariate 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 parameters which are shared by all functions: localization radius R, and cross-localization weight factor β. We discuss the localization radius in this section and the cross-localization weight factor in B2.

360 We first pick an appropriate localization radius for univariate localization functions. 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

365 background error covariance matrix is determined by the "short" Y process. Hence we choose the localization radius for the univariate localization functions to be consistent with the non-zero correlation range of the Y process. Sensitivity experiments (not shown) reveal that $R = 15$ is an appropriate localization radius for univariate localization.

For multivariate localization, we keep the same localization radius for the "short" process, i.e. $R_{YY} = R = 15$, and allow the radius for the X variable to vary. Informed by the true forecast error correlations we choose $R_{XX} = 45$, which is approximately

370 the range of non-negligible correlations for the forecast errors in the X process.

For the four functions under consideration here, the choice of these two localization radii determines the cross-localization radius R_{XY}. As shown in Table 2, the cross-localization radius for Gaspari-Cohn and Bolin-Wallin is $R_{XY} = 30$, while for Askey and Wendland we have $R_{XY} = 15$. Note that with Askey and Wendland 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

375 cross-correlations out past 15 units.

B2   Impact of cross-localization weight factor

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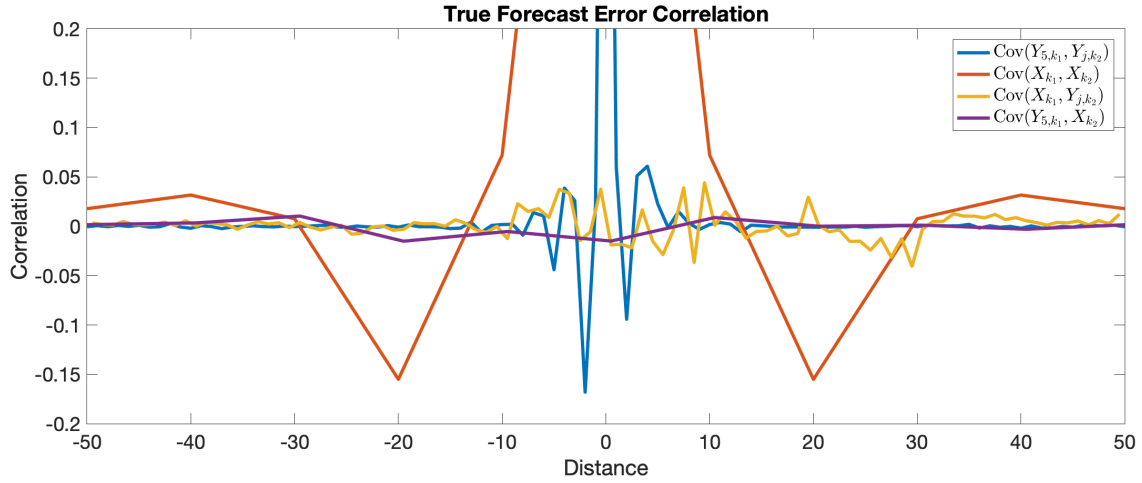
Figure B1. True forecast error correlations for variables in the middle of each sector, X_k and $Y_{5,k}$. Correlations between Y variables (blue) decay to zero after about 5 spatial units, while correlations between X variables (red) are significant up to 40 spatial units away. Cross-correlations (yellow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

by Roh et al. (2015). Each multivariate localization function has a different upper bound on β, which depends on a ratio of

380   localization radii, as shown in Table 2 and Fig. B2.

When $\beta = 0$ 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 best performance generally comes when the cross-correlation is at or

385   near its maximum allowable value, as shown in Fig. B2. Figure B2 shows visually that the Gaspari-Cohn cross-correlation is always greater than the Bolin-Wallin 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.

B3   Technical details of Askey and Wendland functions

390   For the multivariate Askey we must choose $R_{XY}, \gamma_{XX}, \gamma_{YY}, \gamma_{XY}$, and ν. Both R_{XY} and γ_{XY} have bounds restricting the possible range of values to ensure non-negative definiteness. For simplicity, we take these values to be at the edge of the allowable range, $R_{XY} = \min\{R_{XX}, R_{YY}\}, \gamma_{XY} = \frac{R_{XY}}{2}\left(\frac{\gamma_{XX}}{R_{XX}} + \frac{\gamma_{YY}}{R_{YY}}\right)$. Note that the values of γ enter into the multivariate Askey function (10) as sums with ν. Since ν is constant across all processes, we demand that either $\gamma_{XX} = 0$ or $\gamma_{YY} = 0$. Increasing γ_{XX} while keeping $\gamma_{YY} = 0$ serves to decrease the effective localization radius of process X and increase the cross-

395   localization weight factor (Fig. B3). Meanwhile, increasing γ_{YY} while keeping $\gamma_{XX} = 0$ decreases the effective localization radius of process Y and hence decreases the cross-localization weight factor. Thus it is no surprise that sensitivity experiments
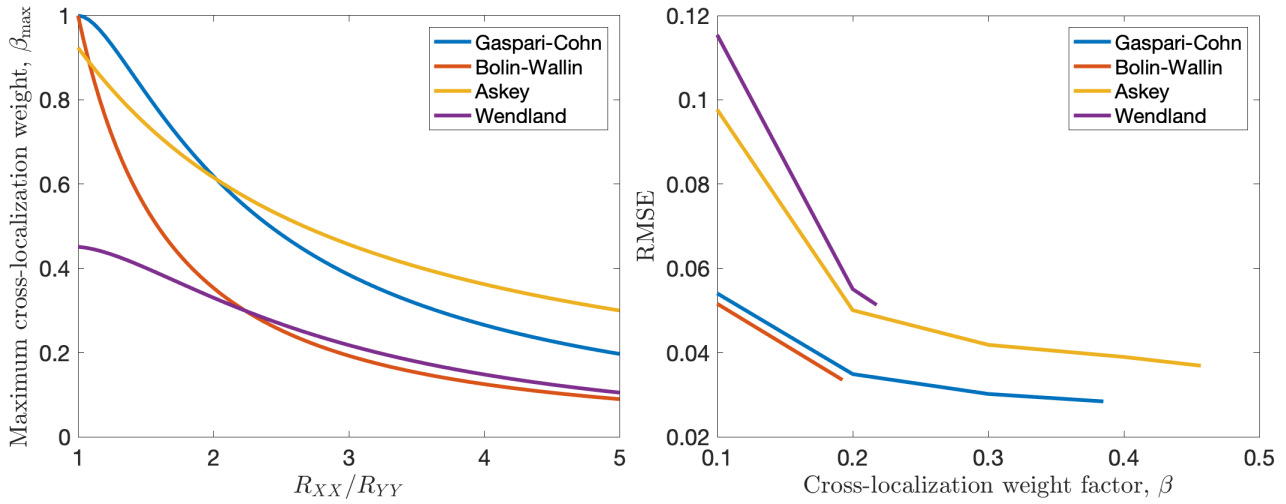
Figure B2. Left: Maximum cross-localization weight factor as a function of R_{XX}/R_{YY}. Right: RMSE for the X process is shown on the y-axis for different multivariate functions. For all functions, as the cross-localization weight factor increases, the RMSE decreases.
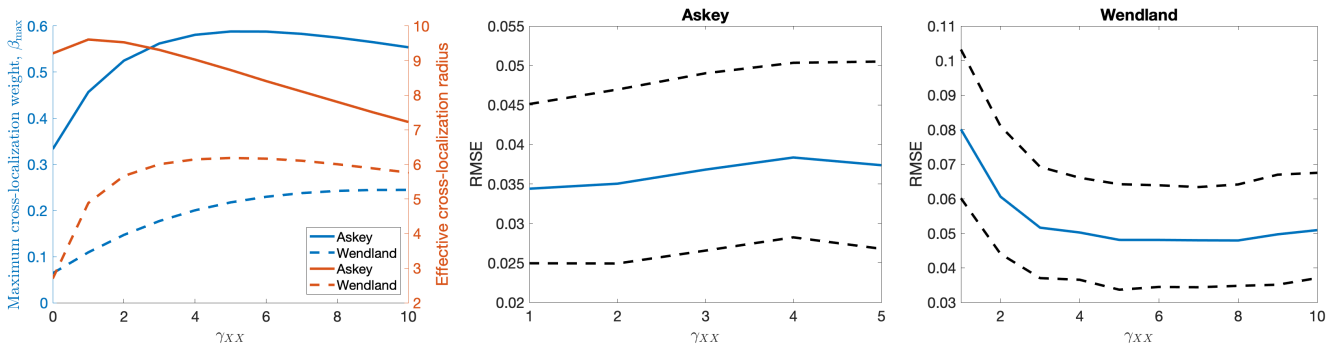
Figure B3. Impact of increasing γ_{XX} in Askey and Wendland functions. Left panel shows that increasing γ_{XX} allows for larger cross-localization weights for Askey (solid blue) and Wendland (dashed blue). The effective cross-localization radius, the distance at which $\mathcal{L}_{XY}(d) < 0.05$, for Askey (solid red) peaks at $\gamma_{XX} = 1$, and increases steadily for Wendland (dashed red). Sensitivity experiments for Askey and Wendland are shown in the right two panels. The optimal value for Askey is $\gamma_{XX} = 1$ and for Wendland is $\gamma_{XX} = 5$.

(not shown) indicate that fixing $\gamma_{YY} = 0$ leads to better results than fixing $\gamma_{XX} = 0$. Sensitivity experiments also show that the best performance comes with $\gamma_{XX} = 1$, shown in Fig. B3.

    For univariate Askey we need only choose the parameter ν in equation (9). For non-negative definiteness, we require $\nu \geq 1$.
400 Smaller values of ν allow for larger cross-localization weights and longer effective cross-localization radii (Fig. B4), both of which are desirable and improve performance in sensitivity experiments (not shown). We choose to use $\nu = 1$, which is as small as possible while still ensuring a non-negative definite function.
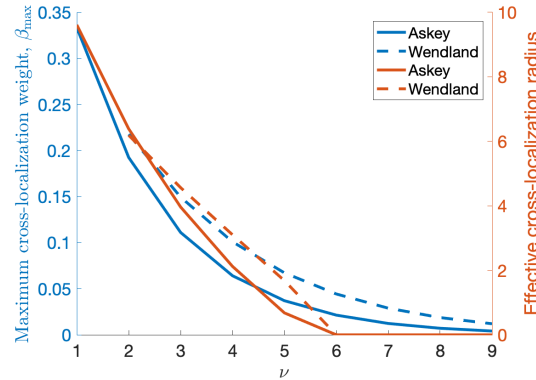
Figure B4. Impact of increasing ν in Askey and Wendland functions. Increasing ν leads to smaller cross-localization weight factors for Askey (solid blue) and Wendland (dashed blue). The effective cross-localization radius, the distance at which $L(d) < 0.05$, also decreases for both Askey (solid red) and Wendland (dashed red).

Now Wendland comes with further parameters to estimate. In particular, we must choose k, which determines the smoothness of the function. We choose to work with $k = 1$ here as higher values lead to unreasonably small cross-localization weight factors and hence degraded performance. For $k = 1$, we require $\nu \geq 2$. As with the Askey function, sensitivity experiments (Fig. B4) indicate that the best performance comes when ν is as small as possible. Hence we pick $\nu = 2$. Again, as with the Askey function we fix $\gamma_{YY} = 0$ and investigate $\gamma_{XX} > 0$. Interestingly, we see that performance improves as γ_{XX} increases, all the way out to $\gamma_{XX} = 5$ (Fig. B3). We hypothesize that this is because increasing γ_{XX} allows for an increased cross-localization weight and effective localization radius (Fig. B3).

References

420   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/https://doi.org/10.1002/qj.339, http://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.339, _eprint:
425   https://onlinelibrary.wiley.com/doi/pdf/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: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 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,
430   http://www.sciencedirect.com/science/article/pii/S2211675316000245, 2016.

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, https://journals.ametsoc.org/view/journals/mwre/126/6/1520-0493_1998_126_1719_asitek_2.0.co_2.xml, publisher: American Meteorological Society Section: Monthly Weather Review, 1998.

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

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, https://journals.ametsoc.org/doi/full/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 statis-
440   tics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, https://doi.org/https://doi.org/10.1029/94JC00572, http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JC00572, _eprint: https://onlinelibrary.wiley.com/doi/pdf/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, http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0041.1, 2016.

445   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, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712555417, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/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, http://projecteuclid.org/euclid.ss/1433341471, 2015.

450   Gneiting, T.: Compactly Supported Correlation Functions, Journal of Multivariate Analysis, 83, 493–508, https://doi.org/10.1006/jmva.2001.2056, https://linkinghub.elsevier.com/retrieve/pii/S0047259X01920561, 2002.

Horn, R. A. and Johnson, C. R.: Matrix analysis, Cambridge University Press, Cambridge ; New York, 2nd ed 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, https://journals.ametsoc.org/view/journals/

455   mwre/126/3/1520-0493_1998_126_0796_dauaek_2.0.co_2.xml, publisher: American Meteorological Society Section: Monthly Weather Review, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, MONTHLY WEATHER REVIEW, 129, 15, 2001.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 112, 1177–1194,
460   https://doi.org/https://doi.org/10.1002/qj.49711247414, http://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711247414, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49711247414, 1986.

Lorenz, E.: Predictability: a problem partly solved, 1995.

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, https://journals.ametsoc.
465   org/view/journals/mwre/143/9/mwr-d-14-00322.1.xml, publisher: American Meteorological Society Section: Monthly Weather Review, 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, http://doi.wiley.com/10.1256/00359000260498897, 2002.

470   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, https://journals.ametsoc.org/view/journals/mwre/143/5/mwr-d-14-00157.1.xml, publisher: American Meteorological Society Section: Monthly Weather Review, 2015.

Penny, S., Akella, S., Alves, O., Bishop, C., Buehner, M., Chevallier, M., Counillon, F., Draper, C., Frolov, S., Yosuke, F., Karspeck, A.,
475   Kumar, A., Laloyaux, P., Mahfouf, J.-F., Martin, M., Peña, M., Tardif, R., Wang, Y., and Wu, X.: Coupled Data Assimilation for Integrated Earth System Analysis and Prediction: Goals, Challenges, and Recommendations, p. 45, 2017.

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, http://link.springer.com/10.1007/s00477-012-0656-z, 2013.

480   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, https://www.nonlin-processes-geophys.net/22/723/2015/, 2015.

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, https://npg.copernicus.org/preprints/npg-2018-50/, 2018.

485   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/https://doi.org/10.1002/2017GL075534, http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL075534, _eprint: https://onlinelibrary.wiley.com/doi/pdf/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/https://doi.org/10.1002/qj.3802,
490   http://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3802, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3802, 2020.

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/https://doi.org/10.1256/qj.04.03, https://rmets.onlinelibrary.wiley.com/doi/abs/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, http://journals.ametsoc.org/doi/full/10.1175/MWR-D-17-0336.1, publisher: American Meteorological Society, 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, http://journals.ametsoc.org/doi/10.1175/MWR-D-17-0365.1, 2018.

Zhang, S., Liu, Z., Zhang, X., Wu, X., Han, G., Zhao, Y., Yu, X., Liu, C., Liu, Y., Wu, S., Lu, F., Li, M., and Deng, X.: Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: a review, Climate Dynamics, 54, 5127–5144, https://doi.org/10.1007/s00382-020-05275-6, https://doi.org/10.1007/s00382-020-05275-6, 2020.