https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

New and existing covariance estimation techniques for ensemble data assimilation

Shay Gilpin[1,2], Matthias Morzfeld[3], and Kevin K. Lin[1]

[1]Department of Mathematics, University of Arizona
[2]Department of Mechanical & Aerospace Engineering, Princeton University
[3]Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego

Correspondence: Shay Gilpin (sg0095@princeton.edu)

Abstract. Covariance estimation from a small number of samples is a challenging, but necessary task in ensemble data assimilation (DA) and a fundamental problem in statistics and machine learning. Many covariance estimation methods have been developed in geophysics and in statistics, with little exchange of ideas between the two fields. Covariance estimation in statistics and in ensemble DA rely on fundamentally different assumptions, and we study the efficiency and applicability of both approaches to covariance estimation in systematic numerical experiments. Our numerical experiments are designed to feature a correlation structure in which correlation decays with distance globally, but the rate of decay is location-dependent. In such cases, our findings suggest that methods that make even minimal assumptions on the decay of spatial correlations are more accurate than those that do not do so at all. We also describe how to use a generalized Gaspari-Cohn correlation function to design new covariance estimation methods that capture intricate and spatially varying correlation structures and thereby further reduce covariance estimation errors.

1 Introduction

The estimation of covariance matrices from a small number of samples is a fundamental problem in applied mathematics, statistics/machine learning (Ledoit and Wolf, 2004; Friedman et al., 2008; Pourahmadi, 2013), and in geoscience, where covariance estimation is a cornerstone of numerical weather prediction (NWP) and ensemble data assimilation (DA) (e.g., Kalnay, 2003; Evensen, 2009a; Asch et al., 2016). Covariance estimation also naturally arises in inverse problems when iterative ensemble methods are used for optimization (e.g., Iglesias et al., 2013; Tong and Morzfeld, 2023; Gu and Oliver, 2007; Chen and Oliver, 2012, 2013; Emerick and Reynolds, 2012, 2013; Bocquet and Sakov, 2014; Evensen, 2018).

This paper focuses on covariance estimation as it applies to ensemble DA in geophysical applications, where the number of samples (i.e., the ensemble size) is orders of magnitude smaller than the dimension of the estimation problem. Under these conditions, the sample covariance matrix is poorly conditioned, rank-deficient, and contaminated by spurious correlations (Wainwright, 2019; Houtekamer and Mitchell, 2001; Morzfeld and Hodyss, 2023; Hodyss and Morzfeld, 2023). Consequently, the sample covariance is not directly used for covariance estimation in ensemble DA, and alternative estimation techniques have been designed to improve the performance of the DA algorithms.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

In geoscience, especially in NWP, covariance localization has been extremely successful for covariance estimation (Houtekamer and Mitchell, 2001; Evensen, 2009a; Morzfeld and Hodyss, 2023). Covariance localization enforces the assumption of a spatial attenuation of correlation on the covariance estimates via Hadamard (element-wise) products of sample covariances with tapering functions that are often compactly supported, e.g., the Gaspari-Cohn (GC) correlation function (Gaspari and Cohn, 1999, their Eq. 4.10). One can also linearly combine the localized sample covariance with a static "background" covariance, often referred to as "hybrid" estimators (e.g., Hamill and Snyder, 2000). Such "hybrid" covariance estimators have become the de facto standard in NWP (Clayton et al., 2012; Bannister, 2017).

The problem of estimating a covariance matrix from a small number of samples has been studied extensively in statistics (e.g., Pourahmadi, 2013; Wainwright, 2019; Friedman et al., 2008; Ledoit and Wolf, 2004). In these contexts, there is often no natural notion of distance, and statistical estimators such as shrinkage estimators and graphical models are designed to rely the empirical correlation structure regardless of geometry. The generality of these spatially-agnostic estimators make them potentially more broadly applicable than covariance localization or hybrid estimators.

Motivated by the need for flexible covariance estimation schemes in geophysical DA, this paper has four goals. First, we introduce a new localization scheme that is a natural extension of GC localization using the generalized Gaspari-Cohn (GenGC) function of Gilpin et al. (2023). The GenGC function enables the localization to capture intricate and spatially varying correlation structure that its predecessor cannot. We describe two ways to define GenGC localization for ensemble DA, built upon features observed in the state dynamics. This work is the first to implement the GenGC correlation function for localization and hybrid covariance estimators to illustrate its potential in ensemble DA. Second, we compile a diverse list of covariance estimation techniques from statistical and geophysical disciplines to highlight the different assumptions inherent in modern covariance estimation. Third, we illustrate the performance of these methods on the static estimation of a known, synthetic covariance matrix. We find that the performance of the various estimators varies drastically depending on the underlying correlation structure. This discrepancy motivates our fourth goal: to compare the performance of different covariance estimators in a cycling DA framework with the ensemble Kalman filter. We deploy covariance estimators in a series of nonlinear test problems of moderate complexity, noting that some of the estimation techniques are being evaluated in a cycling DA framework for the first time.

The main conclusions from the numerical experiments are as follows. Localization with the standard GC correlation function leads to a massive error reduction (compared to the sample covariance) if correlations in the system decay with distance, even if this localization misses some intricate, smaller-scale correlation features. GenGC localization, which aims to better capture these small-scale features, yields a small but distinct improvement over GC localization (for both localization and hybrid estimators), which aligns well with existing theoretical expectations (Morzfeld and Hodyss, 2023; Ménétrier et al., 2015). The improved performance of GenGC localization and its corresponding hybrid estimators, however, does come at the added cost of requiring additional information during implementation (e.g., specifying grids of hyperparameters). Covariance estimation methods that are spatially-agnostic are not competitive with GC and GenGC localization. The latter is perhaps not surprising: while methods that are spatially-agnostic are designed with minimal assumptions, exploiting the underlying

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

problem structure yields better overall filter performance. Moreover, we observe that some estimation methods which produce covariance estimates that are not positive semi-definite can lead to filter divergence in the ensemble Kalman filter.

The rest of this paper is organized as follows. Section 2 provides background materials and is divided into three parts. Section 2.1 gives an overview of ensemble-based data assimilation (DA) and the (stochastic) ensemble Kalman filter (EnKF); Section 2.2 reviews distance-based localization and hybrid estimators, which are common covariance estimation techniques in geophysical DA; Section 2.3 reviews shrinkage, thresholding, and noise informed covariance estimation (NICE), where shrinkage and thresholding are commonly discussed in the statistics literature. Section 3 describes two variants of the Lorenz 96 model (Lorenz, 1996; Lorenz and Emanuel, 1998) that we use in numerical experiments. Section 4 introduces localization with the GenGC correlation function, detailing its construction for the two Lorenz 96 models we consider. Numerical experiments are presented in Section 5. Conclusions are given in Section 6.

2 Background on Ensemble Data Assimilation and Tested Covariance Estimation Techniques

2.1 Ensemble data assimilation

Data assimilation (DA) is a class of statistical estimation techniques that combine numerical models with sparse and noisy observations. DA is a critical component of numerical weather prediction (NWP), and DA methods can be broadly separated into two groups: Monte Carlo-based, ensemble Kalman methods (EnKF) (e.g., Evensen, 1994; Burgers et al., 1998; Tippett et al., 2003) and optimization-based, variational methods (e.g., Talagrand and Courtier, 1987; Courtier and Talagrand, 1987). The most successful methods are "hybrids" that combine the Monte Carlo approach with optimization (e.g., Hamill and Snyder, 2000; Lorenc, 2003a; Zhang et al., 2009; Buehner et al., 2013; Kuhl et al., 2013; Poterjoy and Zhang, 2015). Collectively, we refer to EnKFs and hybrid ensemble-variational methods as "ensemble DA."

In ensemble DA, uncertainties are represented by an ensemble of states and an associated covariance matrix. For an ensemble of n_e members $\mathbf{x}_i \in \mathbb{R}^{n_x}$, $i = 1, 2, \ldots, n_e$, the $n_x \times n_x$ sample covariance matrix is

$$\mathbf{S} = \frac{1}{n_e - 1}\sum_{i=1}^{n_e}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}}, \quad \bar{\mathbf{x}} = \frac{1}{n_e}\sum_{i=1}^{n_e}\mathbf{x}_i\,, \tag{1}$$

where $\overline{\mathbf{x}}$ is the ensemble mean and superscript T denotes the transpose. The sample covariance arises naturally in DA algorithms, for example, in ensemble Kalman filters (EnKFs), which update an ensemble of states using the Kalman formalism (Kalman, 1960).

For a standard EnKF, an ensemble of forecasted states, generated by a dynamical model, are updated with observation information as follows. Let $\mathbf{x}_i^f \in \mathbb{R}^{n_x}$, $i = 1, 2, \ldots, n_e$, be an ensemble of states (the forecast ensemble) at time t^n and let $\mathbf{P}^f$ be the corresponding forecast covariance matrix, estimated based on a modification of the sample covariance matrix $\mathbf{S}$ in Eq. (1) (see Sections 2.2 and 2.3 for a review of different covariance estimation techniques). Assimilating the observation

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

$\mathbf{y} \in \mathbb{R}^{n_y}$ at time t^n (i.e., one step of the EnKF) amounts to computing the analysis ensemble $\mathbf{x}_i^a$,

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}\left(\mathbf{y} - (\mathbf{H}\mathbf{x}_i^f + \boldsymbol{\eta}_i)\right), \quad i = 1, 2, \ldots, n_e, \tag{2a}$$

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{2b}$$

where $\mathbf{R} \in \mathbb{R}^{n_y \times n_y}$ is the observation error covariance matrix, $\mathbf{H} \in \mathbb{R}^{n_y \times n_x}$ is a (linear or linearized) observation operator, and $\boldsymbol{\eta}_i$ are independent samples from a Gaussian distribution with mean zero and covariance $\mathbf{R}$. The $n_x \times n_y$ matrix $\mathbf{K}$ in Eq. (2b) is the Kalman gain. This variant of EnKF is called the "stochastic EnKF" (van Leeuwen, 2020, Sec. 4). Other variants include ensemble adjustment Kalman filters (EAKF; Anderson, 2001) and ensemble transform filters (e.g., Tippett et al., 2003; Ott et al., 2004). Following the analysis update, the analysis ensemble at time t^n is propagated to the next observation time t^{n+1} via the dynamical model,

$$\mathbf{x}_i^f(t^{n+1}) = g\left(\mathbf{x}_i^a(t^n)\right), \tag{3}$$

where $g(\cdot)$ represents the time evolution of the underlying dynamical system from time t^n to t^{n+1}. Equation (3) may include a stochastic forcing term to account for "model error" (Daley, 1991, Ch. 13.3). Equations (2) and (3) define one data assimilation cycle.

We use the stochastic EnKF with a linear observation operator as a representative ensemble DA method to test and compare various covariance estimation techniques. Working directly with $\mathbf{P}^f$ allows a straightforward application of the various covariance estimation techniques during the DA cycles. Nonlinear observation operators, directly estimating the Kalman gain in ensemble square-root filters (e.g., Anderson, 2001; Whitaker and Hamill, 2002; Tippett et al., 2003; Ott et al., 2004), or covariance estimation in observation space (e.g., Greybush et al., 2011) are common techniques in NWP, but these are not amenable to some of the covariance estimation methods we consider without modifications, which is beyond the scope of this paper.

2.2 Covariance estimation in ensemble data assimilation

Typically, the number of unknown state variables, n_x, is much larger than the ensemble size, n_e. In global NWP, for instance, n_x is on the order of 10^9 and the ensemble size n_e is on the order of 10^2. The sample covariance in Eq. (1) is unbiased and consistent, i.e., the large ensemble limit yields the true covariance and the estimator variance decreases with ensemble size. The sample covariance, however, is plagued by large errors when the ensemble size is small compared to the state dimension, $n_e \ll n_x$ (e.g., Wainwright, 2019). All ensemble DA methods used in practice modify the sample covariance to improve its accuracy when the ensemble size is small. Below, we review two common covariance estimation techniques in ensemble data assimilation: localization and hybrid covariance estimators.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

2.2.1 Covariance localization

Covariance localization can reduce errors in the sample covariance by imposing the additional constraint that covariances and correlations attenuate with distance. Such a "distance-based" localization is often implemented via the Hadamard (element-wise) product of the sample covariance $\mathbf{S}$ with a tapering matrix $\mathbf{T} \in \mathbb{R}^{n_x \times n_x}$,

$$\mathbf{P}_{\text{loc}} = \mathbf{T} \circ \mathbf{S}\,, \tag{4}$$

which imposes spatial attenuation of covariances. Since $\mathbf{S}$ is always positive semi-definite (PSD), the Schur Product Theorem (Schur, 1911; Horn and Johnson, 1985, Thm. 7.5.3) guarantees that $\mathbf{P}^f_{\text{loc}}$ is PSD provided $\mathbf{T}$ is also PSD. This is the case whenever $\mathbf{T}$ is itself a covariance matrix for *some* probability distribution. Accordingly, tapering functions used to build $\mathbf{T}$ are usually taken to be correlation functions to ensure that the resulting covariance estimates are PSD.

In atmospheric applications, it is common to construct $\mathbf{T}$ using the piecewise rational, compactly-supported approximation to a Gaussian of Gaspari and Cohn (1999), their Eq. (4.10), often referred to as the Gaspari-Cohn (GC) function. Houtekamer and Mitchell (2001) were the first to implement the GC function for localization. Ever since, GC localization is considered by many to be standard practice in ensemble DA (Hamill et al., 2009), where it is used primarily for horizontal localization (Anderson, 2012), with some limited applications to vertical localization (e.g., Necker et al., 2023). We discuss the GC function, as well as a new generalization of GC localization in Section 4.

There is also a vast body of literature on "optimal localization" methods, as discussed in Morzfeld and Hodyss (2023) and references therein. Optimal localization computes a tapering matrix $\mathbf{T}_{\text{opt}}$ from the true correlations ρ_{ij}, $i,j = 1,2,\ldots,n_x$, where the $i,j-$th entry is defined as

$$(\mathbf{T}_{\text{opt}})_{ij} = \frac{\rho_{ij}^2(n_e-1)}{1+\rho_{ij}^2 n_e}, \tag{5}$$

(Morzfeld and Hodyss, 2023, their Eq. 12). We note that the optimal localization in Eq. (5) does not rely on a spatial attenuation of correlation, however, the "true" correlations ρ_{ij} are unknown in practice. Nonetheless, we will use optimal localization as a benchmark in the numerical examples in Section 5.1 where the "true" correlations are known.

Localization, as described above, does not make sense if the problem at hand is not characterized by a spatial attenuation of covariance or if the problem does not have a natural notion of distance. In these situations, one can instead taper the covariance matrix using functions of sample correlations, rather than spatial distances, and is often referred to as "correlation-based localization." Correlation-based localization is popular in reservoir engineering and history matching (e.g., Luo et al., 2018b, a; Luo and Bhakta, 2020), and has been developed for ensemble Kalman smoothers (Vossepoel et al., 2025). Most correlation-based localization schemes rely on the fact that small correlations are noisy while large correlations are typically more "trustworthy." The sampling error correction (SEC) of Anderson (2012), for example, corrects sample correlations based on a look-up table of correction factors that depend on the sample correlation and on the ensemble size. The same idea –

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

small correlations are noisy and should be damped – also underpins methods defined by "raising correlations to a power" (e.g., Bishop and Hodyss, 2007, 2009a, b), which is an idea we return to in Section 2.3.3.

2.2.2 Hybrid covariance estimation

Hybrid estimators linearly combine the localized sample covariance matrix in Eq. (4) with a prescribed (static) "background" covariance matrix $\mathbf{B}$,

$$\mathbf{P}_{\text{hyb}} = \beta_1 \mathbf{B} + \beta_2 \mathbf{P}^f_{\text{loc}}, \quad \beta_1, \beta_2 \geq 0, \tag{6}$$

(Hamill and Snyder, 2000; Lorenc, 2003b; Buehner, 2005; Bishop and Satterfield, 2013; Satterfield et al., 2018). The coefficients β_1 and β_2 generally sum to one, but are not limited to this constraint (Clayton et al., 2012; Ménétrier and Auligné, 2015; Caron and Buehner, 2022). The advantage of hybrid estimators is that they combine flow-dependent information from the ensemble through the localized sample covariance with additional information from the background covariance matrix $\mathbf{B}$. Hybrid methods are widely applied in geophysical DA with much success, and were shown to provide more accurate covariance estimates than localization alone (Ménétrier et al., 2015).

There are several methods for constructing background covariance matrices $\mathbf{B}$. For instance, $\mathbf{B}$ can be estimated from climatological information, via the National Meteorological Center (NMC) method, by wavelet-based and spectral-based approaches, or by an "ensemble of data assimilations" (EDA), and others (e.g., Parrish and Derber, 1992; Fisher, 2003; Buehner, 2005; Bannister, 2008; Isaksen et al., 2010; Ménétrier and Auligné, 2015). For the numerical experiments in Section 5, we construct an "average of the day" (AoD) background covariance matrix, motivated by the numerical experiments of Ménétrier and Auligné (2015). The AoD matrix, $\mathbf{B}_{\text{AoD}}$, is the homogeneous and isotropic average (i.e., averaged across all distances in space) of the sample covariance matrix "of the day," by which it remains fixed for a "day," or for a fixed number of integration time steps/DA cycles. The AoD matrix is updated at the beginning of the next "day." This matrix serves as a representative background covariance matrix that incorporates information from the cycling DA system, which is then combined with the flow-dependent information from the localized sample covariance matrix.

2.3 Covariance estimation beyond localization

Covariance estimation from a limited number of samples from a (comparably) high-dimensional space is also an important problem in statistics (e.g., Pourahmadi, 2013; Wainwright, 2019). In these statistics applications, estimation techniques have been primarily developed for *static* covariance estimation, i.e. estimating a covariance matrix that does not evolve in time, with little overlap with the estimation techniques developed in DA. These covariance estimation techniques replace assumptions about a spatial attenuation of correlation (localization) or similarity to a target matrix (hybrid estimators) with a different set of assumptions that are often spatially-agnostic. We review three different covariance estimation techniques (shrinkage, thresholding, and noise informed covariance estimation), which we implement in the experiments in Section 5.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

2.3.1 Shrinkage

Shrinkage estimators linearly combine a fixed covariance matrix $\mathbf{B}$ with the sample covariance $\mathbf{S}$ in Eq. (1),

$$\mathbf{P}_{\text{shr}} = \alpha_1 \mathbf{B} + \alpha_2 \mathbf{S}, \quad \alpha_1 + \alpha_2 \leq 1 \text{ and } \alpha_1, \alpha_2 \geq 0. \tag{7}$$

This method is often referred to as "shrinkage" because the eigenvalues of $\mathbf{S}$ "shrink" towards the eigenvalues of the target matrix $\mathbf{B}$ (Pourahmadi, 2013). The covariance matrix $\mathbf{B}$ is typically the identity matrix, and the coefficients α_1 and α_2 are determined by optimizing an objective function (Pourahmadi, 2013, Sec. 4.1). Ledoit and Wolf (2004) find α_1 and α_2 by minimizing mean squared error in the Frobenius norm (Pourahmadi, 2013, p. 100) and by ensuring a consistent covariance estimate. This method, which we refer to as "Ledoit-Wolf," is commonly used for static covariance estimation problems, but showed little promise when used within an EnKF (Nino-Ruiz et al., 2021). We note that shrinkage estimators and hybrid estimators are very similar: both linearly interpolate between a fixed covariance matrix ($\mathbf{B}$) and an ensemble covariance matrix. Hybrid estimators in NWP, however, use a localized sample covariance estimate and carefully construct the target covariance matrix. For static covariance estimation problems, simpler choices for the target matrix $\mathbf{B}$, e.g., the identity matrix, are typically used, but more sophisticated techniques are used to determine the interpolation coefficients α_1 and α_2.

2.3.2 Thresholding

Thresholding methods set small covariances to zero (Bickel and Levina, 2008a, b; Rothman et al., 2009; Pourahmadi, 2013; Wainwright, 2019). These methods are applied element-wise to the sample covariance matrix, and therefore have no underlying structure requirements, in contrast with localization. For example, soft thresholding is defined as

$$(\mathbf{P}_{\text{st}})_{ij} = \begin{cases} \mathbf{S}_{ij} - \lambda\, \text{sign}(\mathbf{S}_{ij}) & \text{if } |\mathbf{S}_{ij}| > \lambda \\ 0 & \text{if } |\mathbf{S}_{ij}| \leq \lambda, \end{cases} \qquad i \neq j. \tag{8}$$

(In Eq. (8), $\text{sign}(x) = +1$ for $x > 0$ and -1 for $x < 0$.) The resulting covariance estimate depends continuously on both the soft threshold λ and elements $\mathbf{S}_{ij}$. One can interpret soft thresholding as a form of tapering in which a spatial distance is replaced by the "distance" between the covariances and the threshold. There are several different thresholding schemes, including hard thresholding (Pourahmadi, 2013), smoothly clipped absolute deviation (SCAD; Fan and Li, 2001), and adaptive thresholding schemes (Cai and Liu, 2011). One may also apply thresholding to the sample correlation matrix; in this case thresholding would be another example of a correlation-based localization technique.

From a theoretical perspective, thresholding is attractive because it can be interpreted as optimal covariance estimation with sparsity-promoting regularization (e.g., Pourahmadi, 2013; Wainwright, 2019). Thresholding has been successfully applied in high-dimensional statistical problems (e.g., Rothman et al., 2009), but applying thresholding to the forecast error covariance matrices has yet to be tested in the context of DA, where only some theoretical results are available in (Al-Ghattas et al., 2025).

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

2.3.3 Noise-informed covariance estimation (NICE)

Rather than working with the sample covariance matrix, one can design covariance estimation techniques that work with the sample correlation matrix to improve overall covariance estimates. The variance-correlation factorization of the sample covariance matrix is

$$\mathbf{S} = \mathbf{V}^{1/2}\mathbf{C}\mathbf{V}^{1/2}, \tag{9}$$

where $\mathbf{V} \in \mathbb{R}^{n_x \times n_x}$ is a diagonal matrix whose diagonal entries are variances, and $\mathbf{C} \in \mathbb{R}^{n_x \times n_x}$ is the sample correlation matrix, which itself is PSD. Noise-informed covariance estimation (NICE; Vishny et al., 2024), for example, modifies the sample correlation matrix under the premise that small to medium correlations are most effected by sampling error and should be damped. This damping is performed by raising correlations to a power, which dampens small and medium correlations while only minimally impacting larger correlations. This approach underpins the "ensemble correlation raised to power" methods of Bishop and Hodyss (2009a, b), where these implementations require manually tuning of such "power." NICE is an adaptive variant of raising correlations to a power that alleviates the need for manual tuning.

The damping of the spatial correlations in NICE can be written as the following Hadamard product,

$$\mathbf{C}_{\text{NICE}} = \mathbf{C} \circ \mathbf{L}(\alpha), \tag{10}$$

where the "damping" matrix $\mathbf{L}(\alpha)$ linearly interpolates between two applications of raising sampling correlations to a positive, even integer a,

$$\mathbf{L}_{ij}(\alpha) = \alpha|\mathbf{C}_{ij}|^{a} + (1-\alpha)|\mathbf{C}_{ij}|^{a-2}, \quad i,j = 1,2,\ldots,n_x, \tag{11}$$

The covariance estimate is then recovered by rescaling by the sample variances.

$$\mathbf{P}_{\text{NICE}} = \mathbf{V}^{1/2}\mathbf{C}_{\text{NICE}}\mathbf{V}^{1/2}. \tag{12}$$

By restricting the power a to be a positive, even integer, NICE guarantees a PSD covariance estimate. Additionally, the pair of powers and interpolation factor α are adaptively chosen based on the estimated "noise level" of the sample correlations and a discrepancy principle.

3 Variants of the Lorenz 96 models

We describe two variants of the Lorenz 96 dynamical system (Lorenz, 1996; Lorenz and Emanuel, 1998) that we use in the numerical experiments of Section 5. We show that both test problems exhibit a spatial attenuation of correlation, so that the overall assumptions underpinning covariance localization are satisfied, however, the strength of the attenuation varies across

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

the domain. It is therefore not obvious if covariance estimation techniques with strong assumptions (e.g., covariance localization or hybrid estimators) can outperform techniques that make fewer assumptions on the (spatial) covariance structure (e.g., thesholding and NICE). Moreover, we describe a new covariance estimation technique that can incorporate spatial variations in the covariance structure and explore in numerical experiments if these new ideas can lead to improved covariance estimators and thereby improved ensemble DA.

3.1 Modified storm-track Lorenz 96

We first consider a modified version of the "storm-track" Lorenz 96 model (Bishop et al., 2017):

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = (x_{j+1} - x_{j-2})x_{j-1} - [1 + 2.5\cos(j\pi/80)^4]x_j + \hat{F}_j(t), \quad j = 1, 2, \ldots, n_x = 80, \tag{13}$$

where $x_j(t)$ is periodic in j, i.e., $x_0(t) = x_{n_x}(t)$ and $x_{-1}(t) = x_{n_x-1}(t)$, and $\hat{F}_j(t)$ is a stochastic forcing term. We discretize Eq. (13) with the standard explicit fourth-order Runge-Kutta (RK4) scheme with time step $\Delta t = 0.05$. The forcing $\hat{F}_j(t)$ is an independent random variable at every grid point that is updated at each integration timestep, where for F_{old} at the previous time step, the new forcing F_{new} is defined as

$$F_{\text{new}} = rF_{\text{old}} + (1-r)G, \tag{14}$$

where $r = e^{-1/3}$ and G is a random draw from a gamma distribution with a shape value of eight and scale value of one. With these choices, the mean value of the forcing is eight, which is also the time-independent forcing in the original Lorenz 96 model (Lorenz, 1996). Note that we increase the magnitude of the damping to generate more pronounced asymmetric features on either side of the central "storm-track" (compare to Bishop et al. (2017), who use $[0.5 + 2\cos(j\pi/80)^4]x_j$ in Eq. 13).

A Hovmöller diagram and the time-averaged climatological covariance matrix of the modified storm-track model are shown in Figure 1. Both figures are generated from a 6000 time step integration of Eq. (13) with the RK4 scheme and time step $\Delta t = 0.05$. We refer to this run as the "free run" of the modified storm-track model. Figure 1(a) plots the states $x_j(t)$ for the last 50 time steps of the free run, and the climatological covariance in Figure 1(b) is estimated from the last 500 time steps of the free run.

The Hovmöller diagram illustrates how the modified storm-track model generates asymmetric damping around the noisy "storm-track" region in the center of the domain, which in turn results in spatially varying correlation length scales, as illustrated by the time-averaged covariance matrix in Figure 1(b). In Section 5, we use the modified storm-track model for two sets of experiments: we use the climatological covariance matrix (Figure 1b) as a static test-case and we use the dynamical model in a set of cycling DA experiments.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

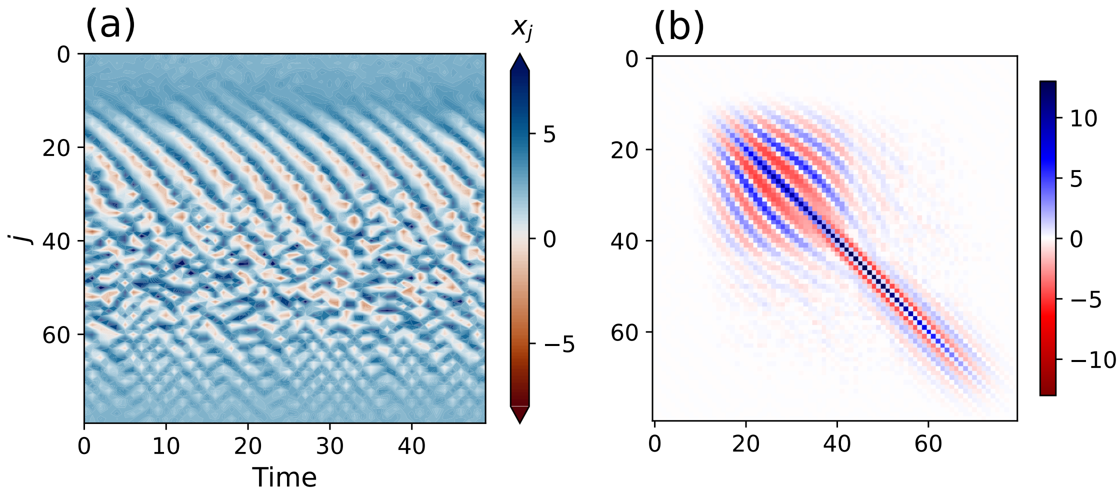

Figure 1. State dynamics and time-averaged climatological covariance matrix of the modified storm-track model Eq. (13). Panel (a): Hovmöller diagram of the state dynamics with time on the horizontal axis, coordinate j on the vertical axis, and magnitude of x_j denoted by the colorbar. Panel (b): climatological covariance matrix estimated from a free run of the modified storm-track model.

3.2 Modified Lorenz 96

The second dynamical system we consider, which we refer to as "modified L96," is a modified version of the Lorenz 96 model introduced in Eqs. (55)-(56) of Vishny et al. (2024):

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F_j \;, \quad j = 1, 2, \ldots, n_x = 400, \quad F_j = 8 + 6\sin(4\pi j/n_x) \;, \tag{15}$$

where both $x_j(t)$ and F_j are periodic in j, the F_j are constant in time, and we consider $n_x = 400$ state variables (Vishny et al. (2024) use only $n_x = 100$). The modified L96 model replaces the constant-in-space forcing of the original Lorenz 96 model with a spatially varying, constant-in-time forcing. These features lead to more structure and variability in the state dynamics and inhomogeneities in the climatological covariance matrix. Figure 2 illustrates the modified L96 model with a Hovmöller diagram of the spatiotemporal evolution of the state and the time-averaged, climatological covariance matrix. The Hovmöller diagram and climatological covariance matrix are constructed from a 6000 time step integration of Eq. (15) using an explicit RK4 scheme with time step $\Delta t = 0.05$, which we refer to as the "free run." The Hovmöller diagram corresponds to the last 50 time steps of this free run, and the climatological covariance matrix is estimated from the last 3500 time steps of the free run.

Regarding the dynamics, we observe that the model is "calmer" in regions where the forcing F_j is small and more "turbulent" in regions where F_j is large (Figure 2a and b). These dynamics lead to regions with lower variances and longer correlations, or "bulges" in climatological covariance matrix, and regions with higher variance and tighter correlation length scales.

In our numerical experiments, we use the modified L96 model to perform state and parameter estimation through a series of cycling DA experiments, where the parameters are the forces F_j in Eq. (15). The state and parameter estimation problem

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

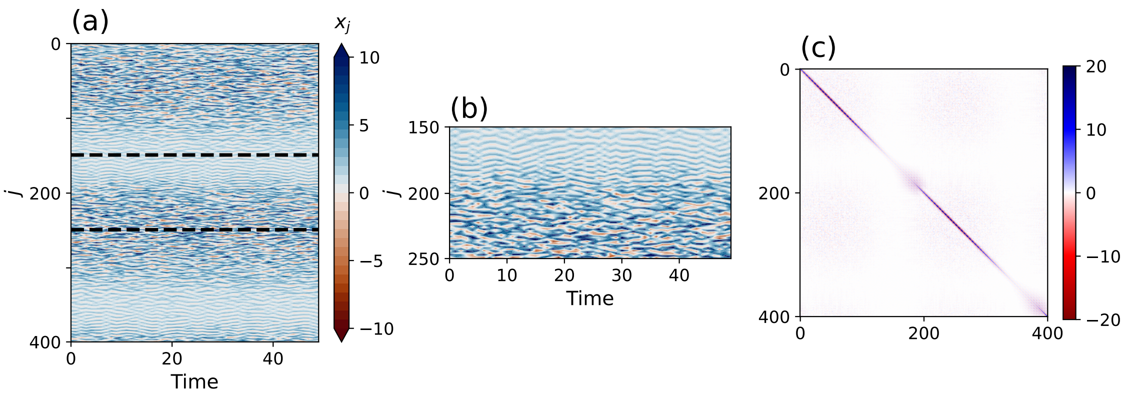

Figure 2. State dynamics and time-averaged climatological covariance matrix for the modified L96 model Eq. (15). Panel (a): Hovmöller diagram of the state dynamics with time on the horizontal axis, coordinate j on the vertical axis, and magnitude of x_j denoted by the colorbar. Panel (b): Zoomed in portion of the Hovmöller diagram (zoom is between the black-dashed lines in (a)). Panel (c): Climatological covariance matrix of the modified L96 model.

has the additional challenge of estimating the state-parameter cross covariances from the ensemble, and appropriately applying the various covariance estimation techniques to these cross-covariance matrices. As such, this problem is a suitable testbed to compare covariance estimation techniques such as thresholding and NICE, which make minimal assumptions, with the standard techniques of localization and hybrid estimators, which rely on stronger structural assumptions.

4 Covariance estimation with Gaspari-Cohn and Generalized Gaspari-Cohn (GenGC)

The most frequently used function for distance-based localization in atmospheric DA is Eq. (4.10) of Gaspari and Cohn (1999), often referred to as the "Gaspari-Cohn" (GC) function. The GC function, which we denote by $C_0(||x-y||;a,c)$, is a parametric correlation function defined on two- or three-dimensional Euclidean space (or on the circle/sphere embedded in the corresponding spaces) where $||x-y||$ denotes the distance between x,y in the domain Ω. There are two globally fixed hyperparameters: a influences the shape of the correlation function and c is a cut-off length in which the GC function becomes identically zero on a disk (or sphere) of radius $2c$ within Ω. For $a = 1/2$, the GC function is a compactly-supported approximation to a Gaussian, defined piecewise in Eq. (4.10) of Gaspari and Cohn (1999),

$$C_0(||x-y||;1/2,c) = \begin{cases} -\frac{1}{4}(||x-y||/c)^5 + \frac{1}{2}(||x-y||/c)^4 + \frac{5}{8}(||x-y||/c)^3 - \frac{5}{3}(||x-y||/c)^2 + 1, & 0 \le ||x-y|| \le c, \\ \frac{1}{12}(||x-y||/c)^5 - \frac{1}{2}(||x-y||/c)^4 + \frac{5}{8}(||x-y||/c)^3 & \\ \quad + \frac{5}{3}(||x-y||/c)^2 - 5(||x-y||/c) + 4 - \frac{2}{3}c/||x-y||, & c \le ||x-y|| \le 2c, \\ 0, & 2c \le ||x-y||. \end{cases} \tag{16}$$

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

Since both hyperparameters c and $a = 1/2$ are globally fixed, the correlation function, and corresponding localization, is homogeneous and isotropic.

Implementing GC localization requires specifying (i.e., tuning) the fixed cut-off length parameter c. In practice, the cut-off parameter c can depend on ensemble size, spatial resolution, observation network, and can vary between state variables or change with vertical height (e.g., Anderson, 2007; Buehner and Shlyaeva, 2015; Destouches et al., 2021; Necker et al., 2022, 2023). Generalizations of the GC function allow for multiple hyperparameters (e.g., multiple cut-off values c) and thereby introduce inhomogeneity and anisotropy into the localization (Gaspari et al., 2006; Stanley et al., 2021). Other localization strategies that introduce inhomogeneity and anisotropy blend GC functions of different scales (e.g., Buehner and Shlyaeva, 2015; Wang et al., 2021) or determine localization functions via machine learning methods (e.g., Lei and Anderson, 2014; Wang et al., 2023).

We introduce a new method of covariance localization that is a natural extension of GC localization using the generalized Gaspari-Cohn (GenGC) correlation function (Gilpin et al., 2023). In contrast to the globally-fixed hyperparameters c and $a = 1/2$ of the GC function, GenGC allows for spatially-varying *fields* of hyperparameters c and a, so that the hyperparameters can vary over subregions of the domain Ω specified by the user (see Sec. 2 of Gilpin et al., 2023). More specifically, for disjoint subregions Ω_k of Ω for $k = 1, 2, \ldots, K$, one can specify hyperparameters a_k, c_k for each subregion. The GenGC correlation function is then constructed using these parameters over the different subregions, where for $x \in \Omega_k$ and $y \in \Omega_\ell$, GenGC is given by

$$C_{k\ell}(||x - y||; a_k, a_\ell, c_k, c_\ell). \tag{17}$$

For example, the domain Ω can be split into two subregions Ω_1 and Ω_2 with hyperparameters a_1, c_1 and a_2, c_2 for each subregion (Gilpin et al., 2023, Figure 2), or each subregion can correspond to a grid cell in Ω and the hyperparameters a and c can be defined as functions over these grid cells (Gilpin et al., 2023, Figures 4–5). The GenGC correlation function, like its predecessor, is compactly-supported, but is supported on non-uniform regions determined by the choice of cut-off lengths c_k. Allowing both hyperparameters to vary over different subregions allows users to design intricate localization schemes while maintaining its compact support and, if a is fixed at $1/2$, its Gaussian-like correlation structure. Moreover, rather than blending multiple GC functions to generate spatially varying correlation structure (e.g., Bishop et al., 2017), GenGC captures complex localization structures with a single correlation function that guarantees positive semi-definiteness by definition. The added flexibility of the parameters does result in a lengthy set of coefficients determined by the various relationships between the cut-off lengths c_k and c_ℓ defined in a piecewise fashion (see Appen. B of Gilpin et al., 2023). Python and Fortran implementations of the GenGC function are freely available (see Gilpin, 2025).

We leverage prior knowledge about the underlying covariance structure and corresponding state dynamics to design the cut-off length parameters c_k. In the next two sections, we describe the construction of GenGC localization for two cycling DA experiments with the two variants of the Lorenz model described in Section 3.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

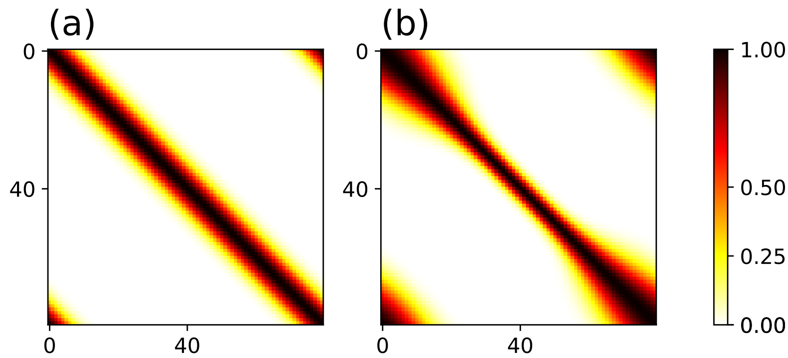

Figure 3. Localization matrices using GC (a) and GenGC (b) for the modified storm-track DA experiments.

4.1 GenGC localization for storm-track Lorenz 96 DA experiments

Motivated by the quiet and noisy regions in Figure 1(a) driven by the spatially-varying damping term in the modified storm-track model, Eq. (13), we define the cut-off length $c(x)$ for GenGC localization as

$$c(x) = c^* \left[1 + 2.5\cos(\pi x)^4\right], \quad x \in [0,1], \tag{18}$$

where the domain $\Omega = [0,1]$ maps to the periodic "grid points" of the modified storm-track model. The additional hyperparameter c^* is tuned to minimize root mean square error over a series of DA experiments. The resulting localization matrix is shown in Figure 3(b), along with a GC localization matrix with a fixed cut-off length c in Figure 3(a). We see that defining the cut-off length function $c(x)$ to correspond to the spatially varying damping effectively captures the correlation structures associated with this damping term, which elongates correlations in the regions of high damping, and shortens correlations in the center where the damping is weaker. This GenGC localization function is constructed for the forecast error covariance matrix of the cycling DA experiments, which, as shown in Figure 9 in Section 5, has different correlation structures than the time-averaged climatological covariance matrix in Figure 1(b). In static covariance estimation experiments, where we estimate the climatological covariance matrix from samples of a mean-zero Gaussian, we construct the GenGC localization scheme to capture the features observed in Figure 1(b); this is described in Section 5.1.

4.2 GC and GenGC localization for state-parameter estimation in a modified L96 model

We consider joint state- and parameter estimation in the modified L96 model: given observations of the state, we estimate the state x_j *and* the forcing parameters F_j for $j = 1, 2, \ldots, n_x$. Thus, the EnKF analysis uses the extended state vector $\mathbf{z}^{\mathrm{T}} =$

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

$(\mathbf{x}^{\mathrm{T}}, \mathbf{F}^{\mathrm{T}})$, and a forecast covariance matrix for $\mathbf{z}$, which we may write in terms of four blocks:

$$\mathbf{P}^f_{\mathrm{zz}} = \begin{pmatrix} \mathbf{P}_{\mathrm{xx}} & \mathbf{P}_{\mathrm{xF}} \\ \mathbf{P}^{\mathrm{T}}_{\mathrm{xF}} & \mathbf{P}_{\mathrm{FF}} \end{pmatrix}. \tag{19}$$

Here, $\mathbf{P}_{\mathrm{xx}}$ and $\mathbf{P}_{\mathrm{FF}}$ are the covariance matrices of the state $\mathbf{x}$ and parameters $\mathbf{F}$, respectively, and $\mathbf{P}_{\mathrm{xF}}$ is the cross-covariance between the state and the parameters. Localization can be applied in the usual way to the state and parameter auto-covariances, $\mathbf{P}_{\mathrm{xx}}$ and $\mathbf{P}_{\mathrm{FF}}$; to localize the state-parameter cross covariances $\mathbf{P}_{\mathrm{xF}}$, we take advantage of the block structure of the forecast covariance $\mathbf{P}^f_{\mathrm{zz}}$. Let $\mathbf{T}_{\mathrm{xx}}$ and $\mathbf{T}_{\mathrm{FF}}$ be tapering matrices for the state and forcing with Cholesky factors $\mathbf{L}_{\mathrm{x}}$ and $\mathbf{L}_{\mathrm{F}}$, respectively. To construct a PSD tapering matrix $\mathbf{T}_z$, we follow Buehner and Shlyaeva (2015) and set the tapering matrix for the cross covariance to be $\mathbf{T}_{\mathrm{xF}} = \mathbf{L}_{\mathrm{x}}\mathbf{L}^{\mathrm{T}}_{\mathrm{F}}$. This results in the state-parameter tapering matrix

$$\mathbf{T}_{\mathrm{z}} = \begin{pmatrix} \mathbf{T}_{\mathrm{xx}} & \mathbf{T}_{\mathrm{xF}} \\ \mathbf{T}^{\mathrm{T}}_{\mathrm{xF}} & \mathbf{T}_{\mathrm{FF}} \end{pmatrix}, \tag{20}$$

which is PSD provided that $\mathbf{T}_{\mathbf{xx}}$ and $\mathbf{T}_{\mathbf{FF}}$ are PSD.

We construct the state tapering matrix $\mathbf{T}_{\mathrm{xx}}$ using GenGC by defining the cut-off length function $c(x)$ following the spatially varying forcing term F_j, where

$$c(x) = c^* \left(1 + \frac{6}{8}\sin(4\pi x)\right), \quad x \in [0,1]. \tag{21}$$

The ratio $6/8$ arises from rescaling the forcing in Eq. (15), and the hyperparameter c^* is tuned. The domain $\Omega = [0,1]$ maps to the periodic “grid points” of the modified L96 model, as in the modified storm-track case. For the parameter tapering matrix $\mathbf{T}_{\mathrm{FF}}$, a fixed cut-off length c is sufficient, and the cross covariance tapering matrices are constructed using the Cholesky factors as described above. The resulting GenGC localization matrix is given in Figure 4, with vertical and horizontal lines separating the four distinct blocks. As noted in Section 3.2, the spatially varying forcing term introduces noisy and quiet regions into the state dynamics, shrinking and increasing the correlation length scales in the respective regions. Taking the cut-off length parameter to vary in the same manner as this forcing term in Eq. (21) effectively captures these features. Using Cholesky factors for cross-covariance localization incorporates the different localization structures in the states and the parameters into the cross-localization blocks while ensuring the overall tapering matrix $\mathbf{T}_z$ is PSD (Buehner and Shlyaeva, 2015).

5 Numerical Experiments

We now compare the covariance estimation techniques in systematic numerical experiments. We first compare covariance estimation methods for a “static” problem, i.e., we estimate a fixed covariance matrix from a given ensemble. In this experiment, we define the underlying “true” covariance matrix and the ensemble is generated from a mean-zero Gaussian with that

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

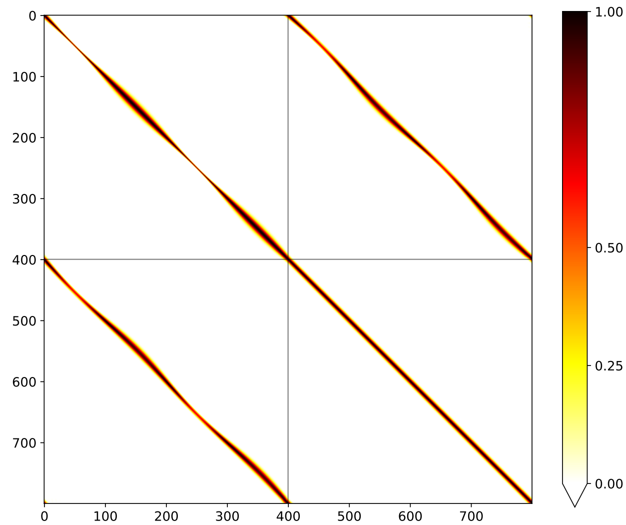

Figure 4. GenGC localization matrix for the modified L96 state-parameter DA experiments. The upper left block corresponds to the state localization matrix $\mathbf{T}_{\mathrm{xx}}$ constructed using the GenGC localization function. The lower right block corresponds to the parameter localization matrix $\mathbf{T}_{\mathrm{FF}}$. The upper right and lower left blocks correspond to the cross-covariance localization matrices constructed using Cholesky factors. These cross-covariance localization blocks contain a few small, negative entries from the Cholesky factors.

Case	Covariance type	Discretization	Ensemble size
$\mathbf{P}_1$	GenGC	$n_x = 1000$	$n_e = 30$
$\mathbf{P}_2$	Modified storm-track	$n_x = 80$	$n_e = 20$

Table 1. Specifications of numerical experiments with two static covariance matrices $\mathbf{P}_1$ and $\mathbf{P}_2$.

covariance matrix. A second set of experiments uses covariance estimation in cycling DA experiments and we compare the covariance estimation methods via their implied DA performance, specifically the analysis root mean squared errors.

5.1 Static covariance estimation

We illustrate the various covariance estimation techniques on two mean-zero multivariate Gaussian distributions with covariance matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, defined over a periodic domain $[0, 2\pi)$, discretized on a uniform grid with n_x grid points (see Table 1). The first covariance matrix $\mathbf{P}_1$ is defined by the generalized Gaspari-Cohn (GenGC) correlation function of Gilpin et al. (2023) with

$$a(x) = \frac{1}{2}\sin(x) - \frac{1}{10}, \tag{22a}$$

$$c(x) = \frac{3}{4}\sin\left(\frac{3x}{2}\right) + 1. \tag{22b}$$

The second covariance matrix $\mathbf{P}_2$ is the time-averaged, climatological covariance of the modified storm-track model (Figure 1(b)).

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

For both cases we draw n_e samples (see Table 1) and use these samples to estimate the "true" covariance matrices. We use GC localization (tuned, fixed length scale) and GenGC localization with $a(x) = 1/2$. For case 1, we use the "true" cut-off length Eq. (22b) for localization and for case 2 we use

$$c(x) = 0.72\left[\sin(x/2) + \sin(x) + \cos(x) + 1\right], \tag{23}$$

to mimic the correlation length scales observed in Figure 1(b). The coefficient of 0.72 was obtained by tuning to minimize errors in the Frobenius norm (for an $n_x \times n_x$ matrix $\mathbf{A}$, the Frobenius norm is defined as $\|\mathbf{A}\|_{\text{Fro}} := \sqrt{\sum_{i,j=1}^{n_x} \mathbf{A}_{ij}^2}$). We further use optimal localization, NICE, Ledoit-Wolf, and thresholding. We do not apply hybrid or shrinkage covariance estimators because there is no natural choice of background matrix $\mathbf{B}$. We repeat that optimal localization (Section 2.2.1) serves as a benchmark, but is not a viable method in practice.

Figures 5 and 6 summarize the results for the estimation of covariances $\mathbf{P}_1$ and $\mathbf{P}_2$. In each figure, panel (a) shows the covariance matrix (either $\mathbf{P}_1$ or $\mathbf{P}_2$) and panels (b)–(h) show covariance estimates for a single ensemble draw. To gain insight into the accuracy of the covariance estimation methods, we also consider 50 independent ensemble draws and associated covariance estimates. We evaluate the accuracy of a covariance estimation method by computing the Frobenius norm of the difference of the covariance matrix and its approximation, normalized by the Frobenius norm of the covariance matrix. The results of 50 experiments are shown in panel (i) of Figures 5 and 6 where we plot errors, relative to the errors of the sample covariance matrix: values less than one (green) correspond to estimates that are more accurate than the sample covariance and values greater than one (purple) are less accurate. Vertical bars, from left to right, correspond to the 20th, 50th, and 80th quantiles, respectively.

Considering first the matrices from a single experiment (panels a-h), we note that the covariance matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ have inhomogeneous structures (panel a), but overall, covariances and correlations decay with distance. The sample covariances are inaccurate due to the small ensemble size (panel b). Optimal localization (panel c) is effective in capturing the inhomogeneous covariance structure and serves as a reference for what a "good" localization should achieve. Localization with GenGC (panel d) also captures the inhomogeneous structure because the tapering matrix is inhomogeneous. Localization with GC (panel e), tuned with a relatively large cut-off, reveals the inhomogeneous covariance structure at the expense of sampling error and noise near the diagonal, where covariances are more local. This noise is efficiently reduced by the more sophisticated localization with GenGC. The spatially-agnostic methods NICE (panel f) and thresholding (panel h) yield qualitatively similar results and are also effective at capturing the inhomogeneous covariance structure. These estimates, however, are more noisy than localization with GC or GenGC. Ledoit-Wolf, which shrinks to the identity matrix, remains rather noisy and does not effectively capture the structures that are better captured by the other estimation techniques. These experiments thus underline the importance of a careful design of the "target" matrix in shrinkage/hybrid estimators.

Now considering the errors over all 50 experiments (panel i of Figures 5 and 6), we find that GC localization results in a significant reduction in errors (compared to the sample covariance), close to what can optimally be achieved. GenGC, a more flexible and general method that explicitly accounts for inhomogeneous structure, yields a small but notable improvement over

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

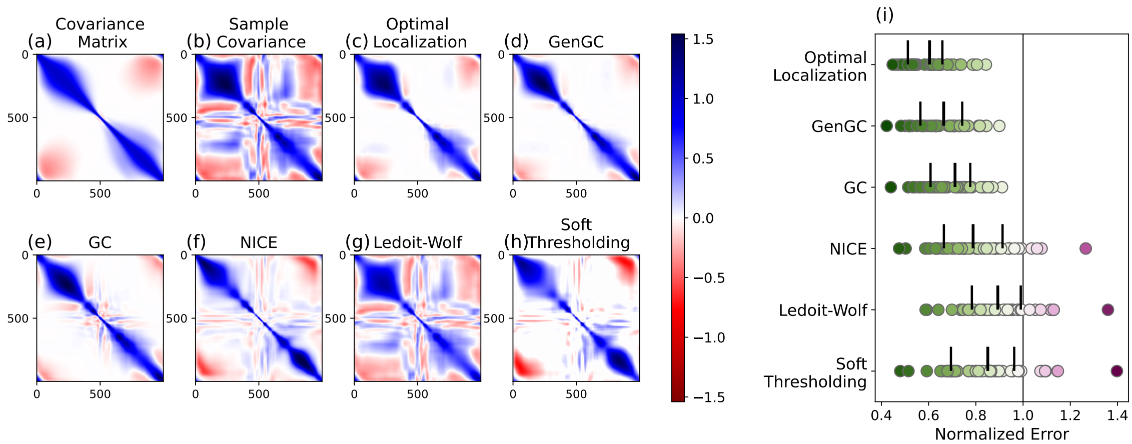

Figure 5. Covariance estimates and normalized errors for the estimation of $\mathbf{P}_1$. Panel (a): Covariance matrix $\mathbf{P}_1$ we wish to compute. Panel (b): Sample covariance matrix. Panel (c): covariance estimate from optimal localization. Panel (d): Distance-based localization using generalized Gaspari-Cohn (GenGC). Panel (e): Distance-based localization using Gaspari-Cohn (GC). Panel (f): Correlation-based localization with noise informed covariance estimation (NICE). Panel (e): Ledoit-Wolf covariance shrinkage. Panel (h): Soft thresholding. Panel (i): Covariance estimation errors in 50 experiments. Errors less than one (green) correspond to improvements over the sample covariance and errors larger than one (purple) mean that the covariance estimation is less accurate than the sample covariance. From left to right, vertical bars correspond to the 20th, 50th, and 80th quantile, respectively.

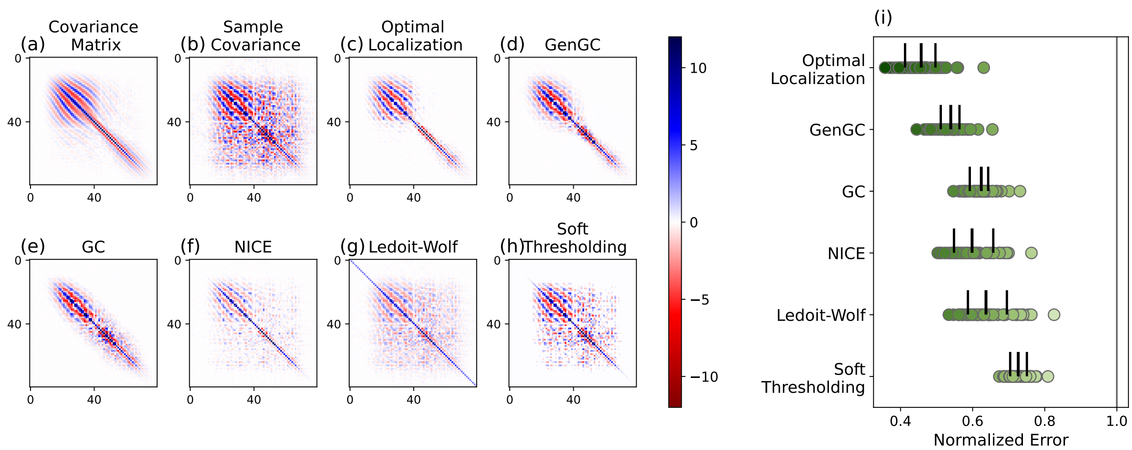

Figure 6. Same as Figure 5, but for $\mathbf{P}_2$. See text for experiment details and Figure 5 for figure description.

405 localization with GC. In contrast, NICE, thresholding, and Ledoit-Wolf are less accurate than both GC and GenGC localization, and in Figure 5 can produce estimates that are worse than the sample covariance. This is perhaps not surprising, since distance-based localization is most appropriate in this example, where covariances (and correlations) decay with distance, albeit in a complicated fashion. These simple examples thus reiterate the result that relatively simple localization is quite effective, and that capturing complex covariance/correlation structure with sophisticated localization techniques can lead to (marginally) improved covariance estimates, provided the dominating covariance structure is characterized by a spatial decay of covariance/

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

correlation (e.g., Morzfeld and Hodyss, 2023; Hodyss and Morzfeld, 2023). Finally, we note that thresholding and optimal localization produce non-PSD covariance estimates which can cause issues in cycling DA (Necker et al., 2023, Sec. 3.4).

5.2 Covariance estimation in cycling ensemble DA

The results of the static Gaussian examples in Section 5.1 highlight the differences between the various covariance estimation techniques, and in particular suggest that GenGC localization can outperform simpler distance-based localization, when the covariances being estimated exhibit spatially varying features. We now explore whether these results are maintained when applying the covariance estimation techniques within the ensemble Kalman filter (EnKF).

We compare the performance of the different covariance estimation techniques in two sets of DA experiments with the EnKF:

1. State estimation for the modified storm-track Lorenz 96 model, Section 3.1, and
2. Joint state-parameter estimation for a modified Lorenz 96 model, Section 3.2.

We begin by describing the setups for the DA experiments followed by a discussion of the results.

5.2.1 DA setup for the modified storm-track Lorenz 96

We apply EnKFs with various covariance estimation techniques to the modified storm-track model (Section 3.1) and estimate the states x_j from observations $\mathbf{y}$. Twenty observations $\mathbf{y}$ are collected in the center of the domain (every other grid point between coordinates 20 and 60) at each model integration time step ($\Delta t = 0.05$). The observation error covariance $\mathbf{R}$ is the identity matrix. The initial ensemble ($n_e = 20$) for each EnKF are random draws from a free run of the model. We assimilate observations for a total of 4000 cycles and compute averaged root mean square errors (RMSE) in the analysis state estimates over the last 2000 cycles, where the first 2000 cycles are spin-up.

The various EnKFs differ in how they each estimate the forecast covariance matrix. Ledoit-Wolf shrinkage, soft thresholding, and NICE are implemented as described in Sections 2.3.1–2.3.3 and GC and GenGC localization is described in Section 4. For shrinkage and hybrid estimators, we construct the "average of the day" AoD background matrix $\mathbf{B}_{\text{AoD}}$, as described in Section 2.2.2. We choose one "day" to correspond to four integration time steps, i.e., $\mathbf{B}_{\text{AoD}}$ is recomputed from the sample covariance matrix every four assimilation cycles. We employ the $\mathbf{B}_{\text{AoD}}$ matrix for covariance shrinkage, Eq. (7), and for two different hybrid estimators, Eq. (6), one applying GC localization and the other GenGC localization. We also apply multiplicative inflation (Anderson and Anderson, 1999; Hamill et al., 2001) in all EnKFs and tune all hyperparameters (for covariance estimation and inflation) jointly to minimize RMSE in the analysis state estimate. In addition to the EnKFs with a small ensemble size, we also run an EnKF with a large ensemble of $n_e = 1600$ without localization/covariance estimation or inflation for comparison.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

5.2.2 DA setup for state and parameter estimation in the modified L96

We consider the modified L96 model of Section 3.2 and estimate both the states x_j and forcing F_j from observations $\mathbf{y}$. For the EnKF forecasting step, we use the discretized, stochastic model

$$\mathbf{x}^{k+1} = g(\mathbf{x}^k, \mathbf{F}^k) + \sigma_x \sqrt{\Delta t}\, \eta^k, \tag{24a}$$
$$\mathbf{F}^{k+1} = \mathbf{F}^k + \sigma_F \sqrt{\Delta t}\, \varepsilon^k, \tag{24b}$$

where $k = 0,1,2,\ldots$ is discrete time, $\mathbf{x} = (x_1,\ldots,x_{400})^{\mathrm{T}}$, $\mathbf{F} = (F_1,\ldots,F_{400})^{\mathrm{T}}$ are the state and parameter vector, $g(\cdot)$ is an explicit RK4 discretization of the right-hand-side of Eq. (15) with time step $\Delta t = 0.05$, and where η^k and ε^k are independent and identically distributed (IID) standard normal random vectors. The stochastic terms account for "model error" (Daley, 1991, Ch. 13.3) and to avoid an overdetermined problem (Evensen, 2009a, pp. 96-97; see also Evensen et al., 2007; Evensen, 2009b). The parameters $\sigma_x = 0.1$ and $\sigma_F = 0.1$ control the degree of stochasticity accounting for model error and the amount of artificial noise injected into the parameter model as regularization.

Motivated by Bishop et al. (2017), we consider observations of integrated state quantities, akin to averaged quantities of satellite observations. Thus, we have no direct observations of the forcing. Each integrated observation is an average of seven consecutive grid-points, located at the center grid point. There are a total of $n_y = 400$ observations at each cycle (i.e., one observation at each grid-point) with observation error variance 0.1. Observations are collected every 0.8 model time units (or equivalently 16 integration time steps).

We perform 500 cycles and discard the first 50 cycles as spin-up. We initialize the EnKFs ($n_e = 20$) with random draws from the free run for the states and use random trigonometric function to generate an initial ensemble of forcings:

$$\begin{aligned} F_j^{(i)} &= \alpha^{(i)} + \beta^{(i)} \sin\left(f^{(i)}\pi j/n_x + \delta^{(i)}\pi\right), \quad j = 1,2,\ldots,400, \quad i = 1,2,\ldots,n_e, \\ &\qquad \alpha^{(i)} \sim \mathcal{N}(8,1),\ \beta^{(i)} \sim \mathcal{N}(6,0.5),\ \delta^{(i)} \sim \mathcal{N}(0,0.5),\ f^{(i)} \sim \mathcal{N}(4,1). \end{aligned} \tag{25}$$

Covariance estimation techniques that make no assumptions about underlying spatial structures (NICE, thresholding, Ledoit-Wolf) can be directly applied to the forecast covariance matrix Eq. (19) of the extended state $\mathbf{z}^{\mathrm{T}} = (\mathbf{x}^{\mathrm{T}}, \mathbf{F}^{\mathrm{T}})$. Distance-based localization (GC and GenGC) for state-parameter estimation is described in Section 4.2 and similar ideas can be used for shrinkage and hybrid estimators. Specifically, we again define the "average of the day" (AoD) covariance matrices for the states and parameters ($\mathbf{B}_{\mathrm{xx}}$ and $\mathbf{B}_{\mathrm{FF}}$) as in Section 2.2.2. The cross-covariance AoD matrix $\mathbf{B}_{\mathrm{xF}}$ is the homogeneous, isotropic average of the cross-covariance between the state and parameter ensemble. We update the AoD covariances at the beginning of each assimilation cycle. As in the modified storm-track experiments, we use an AoD shrinkage estimator and use the AoD background matrices with two localization schemes (GC and GenGC) to obtain two hybrid estimators. As before, we apply multiplicative inflation to all EnKFs and tune all hyperparameters jointly for each covariance estimation method. Finally, we also consider a large ensemble EnKF ($n_e = 4000$) without localization/covariance estimation or inflation for comparison.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

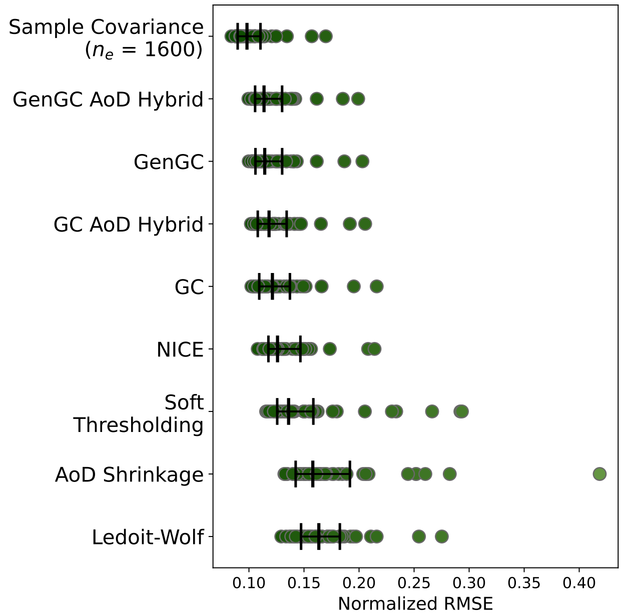

Figure 7. Normalized root mean square error (RMSE) in the analysis state estimate for 50 DA experiments for the modified storm-track model. Errors are normalized by the error of an EnKF that uses the sample covariance with the small ensemble. Therefore, covariance estimation methods that are an improvement have errors less than one (green) and methods that perform worse have errors greater than one (purple). From left to right, vertical bars for each method correspond to the 20th, 50th, and 80th quantiles, respectively.

5.2.3 Interpretation of the results from the numerical experiments

We evaluate the performance of the covariance estimation techniques using the RMSE in the analysis state estimates for 50 different DA experiments with different initial ensembles. Figures 7 and 8 show the RMSE in the analysis state normalized by the corresponding RMSE in the sample covariance with no estimation technique applied for the modified storm-track model and modified L96 model, respectively.

Overall, both DA experiments show that applying any covariance estimation technique to the sample covariance results in improved state estimates, with the exception of thresholding. We observe that NICE, Ledoit-Wolf, and soft thresholding are not competitive compared to distance-based localization and hybrid schemes, though NICE is more competitive in the modified storm-track experiments relative to the modified L96 experiments. Interestingly, the spatially-agnostic methods (NICE, thresholding, and Ledoit-Wolf) all perform worse in the joint state-parameter estimation experiments (Figure 8).

An advantage of the spatially-agnostic methods, such as NICE, is the flexibility gained by directly working with the joint state-parameter forecast error covariance. This flexibility, however, is not an advantage over localization and hybrid schemes, at least in the context of the modified L96 model where there is an underlying spatial structure. Instead, we see that across both experiments, applying localization already yields a major payoff in terms of reducing RMSE in the analysis estimates, with an additional bonus when incorporating more information via GenGC. Hybrid estimators with GenGC localization, which is the most "sophisticated" method we consider, outperforms all other estimation techniques. The difference in RMSE errors in the

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

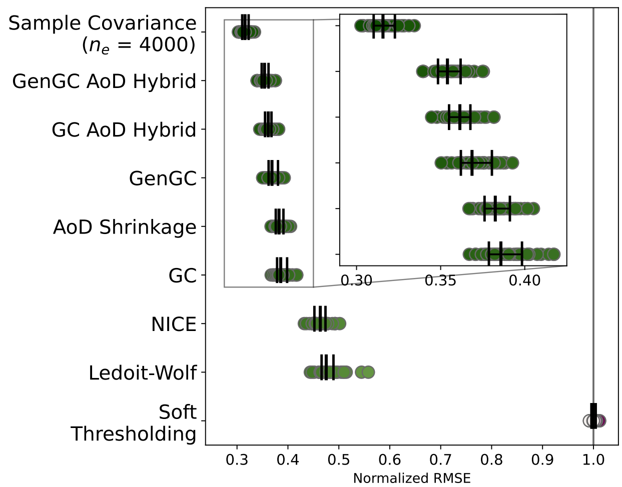

Figure 8. Same as Figure 7 for analysis state estimate of the modified L96 model.

485 GenGC hybrid estimators and the GC hybrid estimators, however, is small, but noticeable. While GenGC does not result in a drastic improvement over GC in these examples, small reductions in RMSE may be impactful in some applications.

The filter performance with soft thresholding changes dramatically between the modified storm-track experiments (Figure 7) and modified L96 experiments (Figure 8). While soft thresholding is comparable to NICE for the modified storm-track experiments, soft thresholding in the joint state-parameter experiments performs the same, if not worse than, the sample covariance estimator itself, and more importantly, the filter has diverged. The issues with thresholding can be partially traced back to the lack of positive definiteness in the forecast covariance matrix. Thresholding with small thresholds produces negative eigenvalues in the matrix $\mathbf{HP}^f\mathbf{H}^\mathrm{T}$. If these negative eigenvalues are greater than or equal in magnitude to the observation error variance, the matrix $\mathbf{HP}^f\mathbf{H}^\mathrm{T}+\mathbf{R}$ becomes ill-conditioned, which in turn results in a blow-up in the analysis ensemble and ultimately causes the filter to crash in the next forecast step. For thresholds that are large enough, the covariance estimates still have negative eigenvalues. However, these negative eigenvalues of $\mathbf{HP}^f\mathbf{H}^\mathrm{T}$ are smaller in magnitude than the observation error variance, and while the analysis ensemble does not blow up in this case, the filter still diverges. In the modified storm-track experiments, the observation error variance of one is large enough to "regularize" these negative eigenvalues produced by soft thresholding and yield reasonable state estimates. The smaller observation error variance of 0.1 in the modified L96 experiments leads to the divergence of the EnKF. The thresholding values that ensure stability of the EnKF (i.e., those used in Figure 8) are so large that essentially no thresholding occurs, i.e., the thresholded forecast covariance is nearly identical to the sample covariance.

To gain further insight into results in Figure 7, we compare the forecast correlation matrices at the last cycle of one of the DA experiments shown in Figure 7 with the sample correlation from the large ensemble ($n_e = 1600$) experiment. We show correlations rather than covariances to more clearly identify the differences between the methods. Both GC and GenGC localization, in addition to their hybrid schemes, have correlations that better represent that of the large ensemble. The slight improvement

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

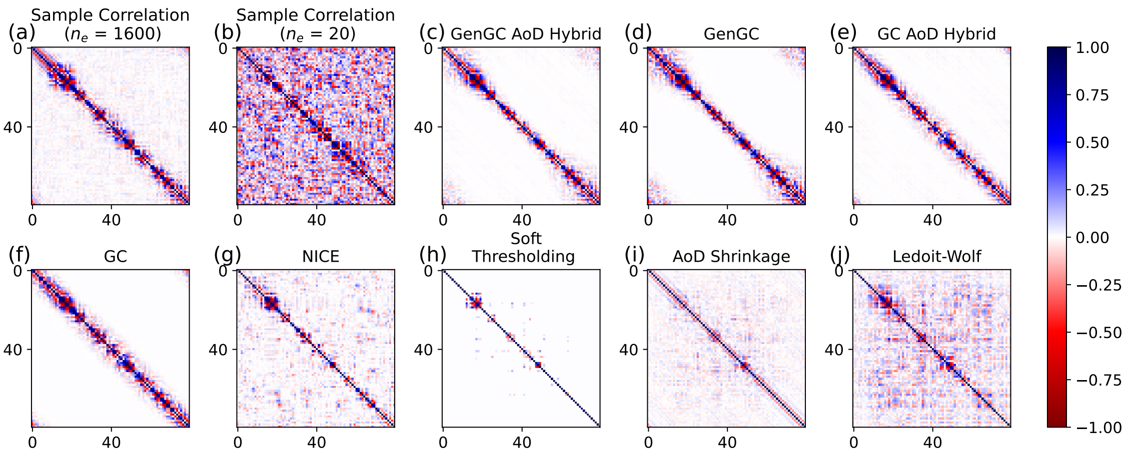

Figure 9. Forecast correlation matrices at the last cycle of one of the DA experiments with the modified storm-track model.

when implementing GenGC localization, either on its own or in the AoD hybrid scheme, relative to GC localization is likely due to the small differences in the correlation estimates (Figure 9c – f). The NICE scheme is somewhat competitive with the distance-based localization and AoD hybrid schemes, even though the correlation structures are generally tighter near the diagonal (Figure 9g). The worse performance of the shrinkage (AoD and Ledoit-Wolf) and soft thresholding is clear from their poor representation of the forecast correlations, particularly in the inability to capture correlation structures near the diagonal.

Finally, we can compare the performance of these methods when estimating the spatially varying parameter F_j in the modified L96 experiments. Figure 10(a) plots the percent error in the 2-norm in the parameter estimates, where for the parameter estimate $\tilde{\mathbf{F}}$ and true parameter $\mathbf{F}$, the percent error is defined as

$$||\tilde{\mathbf{F}} - \mathbf{F}||_2/||\mathbf{F}||_2 \times 100, \quad ||\mathbf{F}||_2 = \sqrt{\sum_{j=1}^{n_x} |F_j|^2}. \tag{26}$$

We compute the percent error in the parameter estimates for the same 50 DA experiments in Figure 8. Figure 10(i)–(x) plot analysis parameter estimates at the end of two of the DA experiments for each of the cases in Figure 10(a) compared to the exact parameter (solid black).

The main difference between the parameter errors and the analysis state RMSE is that the AoD shrinkage estimator performs well for parameter estimation, and in fact has the smallest percent errors. This is likely due to AoD shrinkage producing relatively smooth parameter estimates, as shown in the two examples in Figure 10(vi). Regions where the forcing F_j in Eq. (15) is large correspond to noisy states and often noisier parameter estimates, and is reflected in the parameter estimates of the large ensemble (Figure 10i) and hybrid estimators (Figure 10iii-iv). Localization by itself (Figure 10v-vii) produces noisier parameter estimates overall, suggesting that additional structure through the hybrid schemes can help reduce this noise, at the cost of requiring more data (synthetic or measured) to construct the target matrix $\mathbf{B}$. NICE, Ledoit-Wolf, and soft thresholding

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

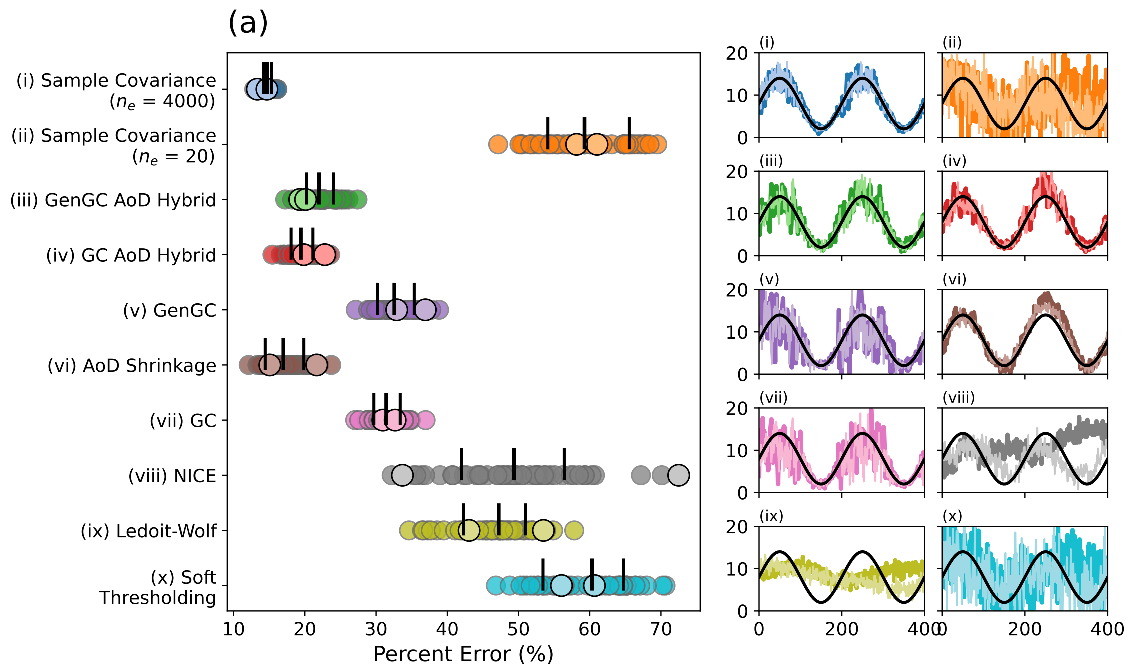

Figure 10. Percent errors in the parameter estimates and parameter examples from the joint state-parameter estimation experiments with the modified L96 model. Panel (a): Percent error of the analysis parameter estimates from the last cycle of the 50 DA experiments for several covariance estimation methods. The vertical bars denote the 20th, 50th, and 80th, quantiles (from left to right). Panels (i)–(x): Example analysis parameter estimates from the last cycle of two DA experiments (color). The exact parameter (solid black) is plotted for reference.

perform the worst, where NICE has significant variability in its parameter estimates (see Figure 10viii) that likely contributes to the degraded analysis state RMSE in Figure 8.

6 Conclusion

We study covariance estimation from a small number of samples as it applies to ensemble DA in the geosciences. We described and compared several existing techniques, including Gaspari-Cohn (GC) localization and hybrid estimators, and also introduced a new localization method with the Generalized Gaspari-Cohn (GenGC) correlation function of Gilpin et al. (2023). Specifically, we described two different approaches to construct GenGC localizations that leverage structural information from the two Lorenz 96 dynamical models. This work is the first to implement the GenGC correlation function for localization and hybrid covariance estimators and illustrates its potential in ensemble DA. These localization and hybrid covariance estimation techniques rely on an underlying assumption that correlation and covariance decay with (spatial) distance. Spatially-agnostic estimation techniques commonly found in the statistics literature offer more flexibility because these methods do *not* rely on a spatial decay of correlation. Such methods, however, are rarely tested in ensemble DA.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

We compared the performance of different covariance estimation techniques with distance-based GC and GenGC localization and hybrid estimators in systematic numerical experiments, including static covariance estimation and two sets of cycling DA experiments with the ensemble Kalman filter (EnKF). Across all experiments, we find that applying standard GC localization, with its assumption of a location-independent lengthscale, already yields major payoffs in terms of improved covariance estimates and thereby filter performance when the problem exhibits a spatial decay of correlations. Incorporating more intricate features during covariance estimation via GenGC localization and hybrid estimators does offer some additional improvements relative to standard GC localization. However, these "sophisticated" localization and hybrid schemes do come at a significant upfront cost, requiring additional hyperparameters and the construction of target matrices. In contrast, the spatially-agnostic schemes we consider (e.g., NICE, thresholding, Ledoit-Wolf), are not competitive with localization or hybrid estimators in all experiments. Although these methods are more flexible by construction, the lack of assumptions did not result in improved covariance estimates or filter performance. We find that making use of a spatial decay of correlations, which is present in all of our numerical examples, proves advantageous over the flexibility in the spatially-agnostic approaches. In addition, some estimation methods violate positive semi-definiteness (PSD) in the covariance estimates, which can cause filter divergence.

Our numerical experiments further demonstrate that encoding additional correlation structure via GenGC can improve DA performance relative to GC localization. While the additional improvement stemming from GenGC is not large, it is worth further study. This result is consistent with a linear (non-cycling) theory (Morzfeld and Hodyss, 2023; Hodyss and Morzfeld, 2023) that states that incorporating localization into covariance estimates leads to a large error reduction (compared to the sample covariance), while incorporating details of the correlation structure has only secondary effects. Moreover, we find that hybrid estimators are most accurate in our experiments, which is in line with, e.g., the work of Ménétrier et al. (2015).

Code availability. The code used in this paper will be made available on GitHub and Zenodo should the paper be accepted for publication.

Author contributions. SG contributed to the conceptualization, investigation, methodology, software development, visualization, and writing. MM and KL contributed to the conceptualization, funding acquisition, and writing.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. SG was supported in part by the National Science Foundation through the Data Driven Discovery Research Training Group at the University of Arizona, under grant DMS-1937229. KL was supported in part by the Simons Foundation under grant MP-TSM-00002687. MM is supported by the U.S. Office of Naval Research Grant N000142512298.

References

Al-Ghattas, O., Chen, J., Sanz-Alonso, D., and Waniorek, N.: Covariance operator estimation: Sparsity, lengthscale, and ensemble Kalman filters, Bernoulli, 31, 2377–2402, https://doi.org/10.3150/24-bej1809, 2025.

Anderson, J.: An ensemble adjustment Kalman filter for data assimilation, Mon. Weather Rev., 129, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129%3C2884:aeakff%3E2.0.co;2, 2001.

Anderson, J.: Localization and sampling error correction in ensemble Kalman filter data assimilation, Mon. Weather Rev., 140, 2359–2371, https://doi.org/10.1175/MWR-D-11-00013.1, 2012.

Anderson, J. L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, Physica D: Nonlinear Phenomena, 230, 99–111, https://doi.org/10.1016/j.physd.2006.02.011, 2007.

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2, 1999.

Asch, M., Bocquet, M., and Nodet, M.: Data Assimilation, Society for Industrial and Applied Mathematics, https://doi.org/10.1137/1.9781611974546, 2016.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, Q. J. Roy. Meteor. Soc., 134, 1971–1996, https://doi.org/10.1002/qj.340, 2008.

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, Quarterly Journal of the Royal Meteorological Society, 143, 607–633, https://doi.org/https://doi.org/10.1002/qj.2982, 2017.

Bickel, P. J. and Levina, E.: Covariance regularization by thresholding, The Annals of Statistics, 36, 2577 – 2604, https://doi.org/10.1214/08-aos600, 2008a.

Bickel, P. J. and Levina, E.: Regularized estimation of large covariance matrices, The Annals of Statistics, 36, 199 – 227, https://doi.org/10.1214/009053607000000758, 2008b.

Bishop, C. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models, Tellus A, 61, 84–96, https://doi.org/10.3402/tellusa.v61i1.15533, 2009a.

Bishop, C. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere, Tellus A, 61, 97–111, https://doi.org/10.3402/tellusa.v61i1.15536, 2009b.

Bishop, C., Whitaker, J., and Lei, L.: Gain form of the ensemble transform Kalman filter and its relevance to satellite data assimilation with model space ensemble covariance localization, Mon. Weather Rev., 145, 4575–4592, https://doi.org/10.1175/MWR-D-17-0102.1, 2017.

Bishop, C. H. and Hodyss, D.: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation, Q. J. Roy. Meteor. Soc., 133, 2029–2044, https://doi.org/10.1002/qj.169, 2007.

Bishop, C. H. and Satterfield, E. A.: Hidden Error Variance Theory. Part I: Exposition and Analytic Model, Mon. Weather Rev., 141, 1454 – 1468, https://doi.org/10.1175/MWR-D-12-00118.1, 2013.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 140, 1521–1535, https://doi.org/10.1002/qj.2236, 2014.

Buehner, M.: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting, Q. J. Roy. Meteor. Soc., 131, 1013–1043, https://doi.org/https://doi.org/10.1256/qj.04.15, 2005.

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

Buehner, M., Mourneau, J., and Charette, C.: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction, Nonlin. Processes Geophys., 20, 669–682, https://doi.org/10.5194/npg-20-669-2013, 2013.

Burgers, G., van Leeuwen, P., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, https://doi.org/10.1175/1520-0493(1998)126%3C1719:asitek%3E2.0.co;2, 1998.

Cai, T. and Liu, W.: Adaptive Thresholding for Sparse Covariance Matrix Estimation, J. Am. Stat. Assoc., 106, 672–684, https://doi.org/10.1198/jasa.2011.tm10560, 2011.

Caron, J.-F. and Buehner, M.: Implementation of Scale-Dependent Background-Error Covariance Localization in the Canadian Global Deterministic Prediction System, Weather and Forecasting, 37, 1567–1580, https://doi.org/10.1175/waf-d-22-0055.1, 2022.

Chen, Y. and Oliver, D. S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother, Mathematical Geosciences, 44, 1–26, https://doi.org/10.1007/s11004-011-9376-z, 2012.

Chen, Y. and Oliver, D. S.: Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, Computational Geosciences, 17, 689–703, https://doi.org/10.1007/s10596-013-9351-5, 2013.

Clayton, A. M., Lorenc, A. C., and Barker, D. M.: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, Q. J. Roy. Meteor. Soc., 139, 1445–1461, https://doi.org/10.1002/qj.2054, 2012.

Courtier, P. and Talagrand, O.: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results, Q. J. Roy. Meteor. Soc., 113, 1329–1347, https://doi.org/10.1002/qj.49711347813, 1987.

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, 1991.

Destouches, M., Montmerle, T., Michel, Y., and Ménétrier, B.: Estimating optimal localization for sampled background-error covariances of hydrometeor variables, Q. J. Roy. Meteor. Soc., 147, 74–93, https://doi.org/10.1002/qj.3906, 2021.

Emerick, A. A. and Reynolds, A. C.: History matching a field case using the ensemble smoother with multiple data assimilation, SPE Journal, 17, 1139–1153, https://doi.org/10.2118/141758-PA, 2012.

Emerick, A. A. and Reynolds, A. C.: Ensemble smoother with multiple data assimilation, Computers & Geosciences, 55, 3–15, https://doi.org/10.1016/j.cageo.2012.03.011, 2013.

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

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer, 2nd edn., 2009a.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Systems, 29, 83–104, https://doi.org/10.1109/mcs.2009.932223, 2009b.

Evensen, G.: Analysis of iterative ensemble smoothers for solving inverse problems, Computational Geosciences, 22, 885–908, https://doi.org/10.1007/s10596-018-9731-y, 2018.

Evensen, G., Hove, J., Meisingset, H., Reiso, E., Seim, K. S., and Espelid, Ø.: Using the EnKF for Assisted History Matching of a North Sea Reservoir Model, in: Proceedings of SPE Reservoir Simulation Symposium, RSS, Society of Petroleum Engineers, https://doi.org/10.2523/106184-ms, 2007.

Fan, J. and Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties, J. Am. Stat. Assoc., 96, 1348 – 1360, https://doi.org/10.1198/016214501753382273, 2001.

Fisher, M.: Background error covariance modelling, in: Seminar on recent development in data assimilation for atmosphere and ocean, pp. 45–63, Shinfield Park, Reading, 2003.

Friedman, J., Hastie, T., and Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso, Biostatistics, 9, 432–441, 2008.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

Gaspari, G. and Cohn, S. E.: Construction of Correlation Functions in Two and Three Dimensions, Q. J. Roy. Meteor. Soc., 125, 723–757, https://doi.org/10.1002/qj.49712555417, 1999.

Gaspari, G., Cohn, S. E., Guo, J., and Pawson, S.: Construction and Application of Covariance Functions with Variable Length-Fields, Q. J. Roy. Meteor. Soc., 132, 1815–1838, https://doi.org/10.1256/qj.05.08, 2006.

Gilpin, S.: A Generalized Gaspari-Cohn (GenGC) Correlation Function, https://github.com/sgilpin4/gengc, 2025.

Gilpin, S., Matsuo, T., and Cohn, S. E.: A Generalized, Compactly-Supported Correlation Function for Data Assimilation Applications, Q. J. Roy. Meteor. Soc., 149, 1953–1989, https://doi.org/10.1002/qj.4490, 2023.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R.: Balance and Ensemble Kalman Filter Localization Techniques, Mon. Weather Rev., 139, 511–522, https://doi.org/10.1175/2010mwr3328.1, 2011.

Gu, Y. and Oliver, D. S.: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation, SPE Journal, 12, 438–446, https://doi.org/10.2118/108438-PA, 2007.

Hamill, T. M. and Snyder, C.: A hybrid ensemble Kalman Filter-3D variational analysis scheme, Mon. Weather Rev., 128, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000.

Hamill, T. M., Whitaker, J., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Mon. Weather Rev., 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129%3C2776:ddfobe%3E2.0.co;2, 2001.

Hamill, T. M., Whitaker, J. S., Anderson, J. L., and Snyder, C.: Comments on "Sigma-Point Kalman Filter Data Assimilation Methods for Strongly Nonlinear Systems", Journal of the Atmospheric Sciences, 66, 3498 – 3500, https://doi.org/10.1175/2009JAS3245.1, 2009.

Hodyss, D. and Morzfeld, M.: How Sampling Errors in Covariance Estimates Cause Bias in the Kalman Gain and Impact Ensemble Data Assimilation, Mon. Weather Rev., 151, 2413 – 2426, https://doi.org/10.1175/MWR-D-23-0029.1, 2023.

Horn, R. and Johnson, C. R.: Matrix Analysis, Cambridge University Press, 1985.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 128, 123–137, https://doi.org/10.1175/1520-0493(2001)129%3C0123:asekff%3E2.0.co;2, 2001.

Iglesias, M. A., Law, K., and Stuart, A. M.: Ensemble Kalman methods for inverse problems, Inverse Problems, 29, 045 001, 2013.

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., and Raynaud, L.: Ensemble of data assimilations at ECMWF, Tech. rep., European Centre for Medium-Range Weather Forecasting, https://doi.org/10.21957/OBKE4K60, 2010.

Kalman, R.: A new approach to linear filtering and prediction problems, Journal of Basic Engineering, 82, 35–45, https://doi.org/10.1115/1.3662552, 1960.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge university press, 2003.

Kuhl, D. D., Rosmond, T. E., Bishop, C. H., McLay, J., and Baker, N. L.: Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework, Mon. Weather Rev., 141, 2740–2758, https://doi.org/10.1175/mwr-d-12-00182.1, 2013.

Ledoit, O. and Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal., 88, 4254–4278, https://doi.org/10.1016/S0047-259X(03)00096-4, 2004.

Lei, L. and Anderson, J. L.: Comparisons of empirical localization techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model, Mon. Weather Rev., 142, 739–754, https://doi.org/10.1175/mwr-d-13-00152.1, 2014.

Lorenc, A.: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var, Q. J. Roy. Meteor. Soc., 129, 3183–3203, https://doi.org/10.1256/qj.02.132, 2003a.

Lorenc, A. C.: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var, Q. J. Roy. Meteor. Soc., 129, 3183–3203, https://doi.org/10.1256/qj.02.132, 2003b.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

Lorenz, E.: Predictability: A problem partly solved, In: Proceedings of the ECMWF Seminar on Predictability, 1, 1–18, https://doi.org/10.1017/cbo9780511617652.004, 1996.

Lorenz, E. N. and Emanuel, K. A.: Optimal sites for supplementary weather observations: Simulation with a small model, Journal of the Atmospheric Sciences, 55, 399–414, https://doi.org/10.1175/1520-0469(1998)055%3C0399:osfswo%3E2.0.co;2, 1998.

Luo, X. and Bhakta, T.: Automatic and adaptive localization for ensemble-based history matching, Journal of Petroleum Science and Engineering, 184, 106 559, https://doi.org/10.1016/j.petrol.2019.106559, 2020.

Luo, X., Bhakta, T., and Nævdal, G.: Correlation-Based Adaptive Localization With Applications to Ensemble-Based 4D-Seismic History Matching, SPE Journal, 23, 396–427, https://doi.org/10.2118/185936-PA, 2018a.

Luo, X., Lorentzen, R. J., Valestrand, R., and Evensen, G.: Correlation-Based Adaptive Localization for Ensemble-Based History Matching: Applied To the Norne Field Case Study, SPE Reservoir Evaluation and Engineering, 22, 1084–1109, https://doi.org/10.2118/191305-PA, 2018b.

Ménétrier, B. and Auligné, T.: Optimized Localization and Hybridization to Filter Ensemble-Based Covariances, Mon. Weather Rev., 143, 3931–3947, https://doi.org/10.1175/MWR-D-15-0057.1, 2015.

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, Mon. Weather Rev., 143, 1622–1643, https://doi.org/10.1175/mwr-d-14-00157.1, 2015.

Morzfeld, M. and Hodyss, D.: A theory for why even simple covariance localization is so useful in ensemble data assimilation, Mon. Weather Rev., 151, 717–736, https://doi.org/10.1175/mwr-d-22-0255.1, 2023.

Necker, T., Hinger, D., Griewant, P., Miyoshi, T., and Weissmann, M.: Guidance on how to improve vertical covariance localization based on a 1000-member ensemble, EGUSphere, pp. 1–23, https://doi.org/10.5194/npg-30-13-2023, 2022.

Necker, T., Hinger, D., Griewank, P. J., Miyoshi, T., and Weissmann, M.: Guidance on how to improve vertical covariance localization based on a 1000-member ensemble, Nonlin. Processes Geophys., 30, 13–29, https://doi.org/10.5194/npg-30-13-2023, 2023.

Nino-Ruiz, E., Guzman, L., and Jabba, D.: An ensemble Kalman filter implementation based on the Ledoit and Wolf covariance matrix estimator, Journal of Computational and Applied Mathematics, 384, 113 163, https://doi.org/10.1016/j.cam.2020.113163, 2021.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus A, 56, 415, https://doi.org/10.3402/tellusa.v56i5.14462, 2004.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, Mon. Weather Rev., 120, 1747–1763, https://doi.org/10.1175/1520-0493(1992)120%3C1747:tnmcss%3E2.0.co;2, 1992.

Poterjoy, J. and Zhang, F.: Systematic comparison of four-dimensional data assimilation methods with and without the tangent linear model using hybrid background error covariance: E4DVar versus 4DEnVar, Mon. Weather Rev., 143, 1601–1621, https://doi.org/10.1175/mwr-d-14-00224.1, 2015.

Pourahmadi, M.: High-Dimensional Covariance Estimation, Wiley, 2013.

Rothman, A. J., Levina, E., and Zhu, J.: Generalized thresholding of large covariance matrices, Journal of the American Statistical Association, 104, 177–186, https://doi.org/10.1198/jasa.2009.0101, 2009.

Satterfield, E. A., Hodyss, D., Kuhl, D. D., and Bishop, C. H.: Observation-Informed Generalized Hybrid Error Covariance Models, Mon. Weather Rev., 146, 3605–3622, https://doi.org/10.1175/MWR-D-18-0016.1, 2018.

Schur, J.: Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen, Journal für die reine und angewandte Mathematik, 1911, 1–28, 1911.

https://doi.org/10.5194/egusphere-2026-5032
Preprint. Discussion started: 21 September 2026

Stanley, Z., Grooms, I., and Kleiber, W.: Multivariate localization functions for strongly coupled data assimilation in the bivariate Lorenz '96 system, Nonlin. Processes Geophys., 28, 656–683, https://doi.org/10.5194/npg-28-565-2021, 2021.

Talagrand, O. and Courtier, P.: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, Q. J. Roy. Meteor. Soc., 113, 1311–1328, https://doi.org/10.1002/qj.49711347812, 1987.

Tippett, M., Anderson, J., Bishop, C., Hamill, T., and Whitaker, J.: Ensemble square root filters, Mon. Weather Rev., 131, 1485–1490, https://doi.org/10.1175/1520-0493(2003)131%3C1485:esrf%3E2.0.co;2, 2003.

Tong, X. T. and Morzfeld, M.: Localized ensemble Kalman inversion, Inverse Problems, 39, 064 002, https://doi.org/10.1088/1361-6420/accb08, 2023.

van Leeuwen, P. J.: A consistent interpretation of the stochastic version of the Ensemble Kalman Filter, Q. J. Roy. Meteor. Soc., 146, 2815–2825, https://doi.org/10.1002/qj.3819, 2020.

Vishny, D., Morzfeld, M., Gwirtz, K., Bach, E., Dunbar, O. R., and Hodyss, D.: High-dimensional covariance estimation from a small number of samples, Journal of Advances in Modeling Earth Systems, 16, 1–30, https://doi.org/10.1029/2024MS004417, 2024.

Vossepoel, F. C., Evensen, G., and van Leeuwen, P. J.: Adaptive correlation- and distance-based localization for iterative ensemble smoothers in a coupled nonlinear multiscale model, Mon. Weather Rev., https://doi.org/10.1175/mwr-d-24-0269.1, 2025.

Wainwright, M. J.: High-Dimensional Statistics: A Non-Asymptotic Viewpoint, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, https://doi.org/10.1017/9781108627771, 2019.

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), Mon. Weather Rev., 149, 605–622, https://doi.org/10.1175/mwr-d-20-0290.1, 2021.

Wang, Z., Lei, L., Anderson, J. L., Tan, Z.-M., and Zhang, Y.: Convolutional Neural Network-Based Adaptive Localization for an Ensemble Kalman Filter, Journal of Advances in Modeling Earth Systems, 15, 1–18, https://doi.org/10.1029/2023ms003642, 2023.

Whitaker, J. S. and Hamill, T. M.: Ensemble Data Assimilation without Perturbed Observations, Mon. Weather Rev., 130, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:edawpo>2.0.co;2, 2002.

Zhang, F., Zhang, M., and Hansen, J. A.: Coupling ensemble Kalman filter with four-dimensional variational data assimilation, Advances in Atmospheric Sciences, 26, 1–8, https://doi.org/10.1007/s00376-009-0001-8, 2009.