TU Delft

Delft University of Technology

Adaptive Correlation- and Distance-Based Localization for Iterative Ensemble Smoothers in a Coupled Nonlinear Multiscale Model

Vossepoel, Femke C.; Evensen, Geir; van Leeuwen, Peter Jan

DOI
10.1175/MWR-D-24-0269.1

Publication date
2025

Document Version
Final published version

Published in
Monthly Weather Review

Citation (APA)
Vossepoel, F. C., Evensen, G., & van Leeuwen, P. J. (2025). Adaptive Correlation- and Distance-Based Localization for Iterative Ensemble Smoothers in a Coupled Nonlinear Multiscale Model. *Monthly Weather Review*, *153*(11), 2593-2609. https://doi.org/10.1175/MWR-D-24-0269.1

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Adaptive Correlation- and Distance-Based Localization for Iterative Ensemble Smoothers in a Coupled Nonlinear Multiscale Model

FEMKE C. VOSSEPOEL,[a] GEIR EVENSEN,[b,c] AND PETER JAN VAN LEEUWEN[d]

[a] *Department of Geoscience and Engineering, Delft University of Technology, Delft, Netherlands*
[b] *Norwegian Research Center, Bergen, Norway*
[c] *Nansen Environmental and Remote Sensing Center, Bergen, Norway*
[d] *Department of Atmospheric Sciences, Colorado State University, Fort Collins, Colorado*

(Manuscript received 9 December 2024, in final form 19 July 2025, accepted 31 July 2025)

ABSTRACT: This paper extends the 2024 study of iterative ensemble smoothers by Evensen et al., who used a sizeable 1000-member ensemble configuration, to now using smaller, more affordable ensemble sizes with localization. As is well known, localization is needed to increase the effective ensemble size and avoid degradation of the smoother solutions by spurious correlations. As an alternative to the standard distance-based localization, we propose a reformulation of an adaptive correlation-based localization method that, in a local update, considers only those observations for which the absolute value of the correlation to the model counterpart is larger than a user-defined threshold. In the standard distance-based localization, we update model variables using only nearby observations in physical distance. In correlation-based localization, we update variables using only observations with small correlation distances. We define the correlation distance as one minus the absolute value of the ensemble correlation between a predicted measurement and the variable we are updating. Using the same formulation and implementation as in the 2024 Evensen et al. study, we compare the performance of the two localization strategies in a coupled nonlinear multiscale model and demonstrate the better or at least comparable performance of the adaptive correlation-based localization. We attribute this to an additional measurement error variance inflation for the measurements with a correlation distance close to the truncation distance, effectively leading to smoother updates. Furthermore, it solves the problem of space–time localization that is hard to solve using localization based on physical distance in ensemble smoothers over longer time windows. We also discuss strategies for the efficient implementation of the correlation-based approach.

KEYWORDS: Bayesian methods; Inverse methods; Coupled models; Data assimilation; Nonlinear models

1. Introduction

We use coupled ocean and atmosphere models that assimilate available observations of their state for weather and climate forecasts. In a preceding paper, we described how ensemble methods initially developed for parameter estimation in petroleum reservoir models have the potential for sequential data assimilation in coupled and multiscale unstable dynamical systems such as the ones used for weather and climate forecasting (Evensen et al. 2024). The overviews of data assimilation in coupled ocean–atmosphere models by Penny et al. (2017), Tondeur et al. (2020), and Miwa and Sawada (2024) have demonstrated the importance of simultaneously updating the state of both model components in coupled data assimilation. Miwa and Sawada (2024) highlight the sensitivity of the data assimilation performance to the values of hyperparameters (i.e., localization and inflation) and model imperfections (i.e., uncertain parameters) in the data assimilation.

In our previous paper, we used a simple representation of an Earth system by using a two-component coupled Kuramoto–Sivashinsky (KS) model to describe interactions between two components: one with a longer spatial scale (typically the atmosphere) and one with a shorter spatial scale (typically the ocean). Studying the covariance structures of this coupled model, we explored the interactions between the predominantly large and small spatial scales and their implications for data assimilation. Similar studies by Tondeur et al. (2020) and Miwa and Sawada (2024) have illustrated the information propagation between the two components and concluded that these cross-component feedbacks play essential roles both for the slow and the fast scales.

Compared to the Lorenz'96 models employed by Miwa and Sawada (2024), the KS model simulates realistic nonlinear coupled dynamics, reasonably representing the ocean–atmosphere interaction. While the KS model does not represent the quasigeostrophic ocean–atmosphere as in the model of De Cruz et al. (2016) employed by Tondeur et al. (2020), it does capture differences in spatial and temporal scales.

For the results of these studies to be meaningful for realistic applications and operational data assimilation, we must first test the performance of coupled data assimilation with limited ensemble sizes before dealing with different observational constraints in realistic applications, e.g., how to handle issues connected with varying measurement latency and density for the two model components. In Evensen et al. (2024), we used 1000 realizations in all experiments to study the assimilation method's actual performance without the influence of sampling errors. For smaller and more realistic ensemble sizes, we must use localization and possibly inflation to avoid ensemble

Denotes content that is immediately available upon publication as open access.

Corresponding author: Femke C. Vossepoel, f.c.vossepoel@tudelft.nl

DOI: 10.1175/MWR-D-24-0269.1
© 2025 American Meteorological Society. This published article is licensed under the terms of a Creative Commons Attribution 4.0 International (CC BY 4.0) License.

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

degeneracy. Without localization, ensemble data-assimilation methods require huge ensemble sizes to adequately represent the prior distribution's state space needed for the estimation problem (Evensen et al. 2022). Especially for large operational data assimilation systems that estimate $\mathcal{O}(10^8$–$10^{11})$ state variables and sometimes assimilate $\mathcal{O}(10^5$–$10^7)$ measurements in each update step, the typical choice for an ensemble size of 100 members provides insufficient samples of the prior to represent the fine scales and detailed information contained in the measurements. Through localization, we limit the area in which we relate the measurements to the model equivalents, which allows the model to fit the finer scales of these local measurements instead of trying to fit all observations at once, effectively increasing the local ensemble size.

Roughly speaking, we can implement the localization scheme based on a spatial distance (distance-based localization) or the correlations between the observations and model variables (adaptive correlation-based localization). One of the issues with distance-based localization in a coupled system with different length scales in the two components is our inability to define a proper truncation distance beforehand. For example, if we use an ocean observation to update the atmospheric state, is a truncation distance based on the dominant ocean or the atmospheric scales the most appropriate? As in our examples, this question also extends to time distances when we use an ensemble smoother to update the ensemble over a data assimilation window. This issue suggests that correlation-based localization might perform better as it automatically chooses which observations to use for the update of a model variable. However, the choice of the correlation cutoff will become more critical. More generally, the choice between these two approaches and their parameters will affect the performance of the coupled data assimilation. This paper aims to investigate the role of localization in coupled data assimilation, explore the advantages of two methods for truncating measurements, and assess the assimilation performance's sensitivity to the method's parameters. Understanding the effects of these choices in the idealized setting of the KS model will help the design and implementation of data assimilation approaches in operational coupled models of higher dimensions.

2. Coupled multiscale Kuramoto–Sivashinsky model

Evensen et al. (2024) introduced the coupled two-component multiscale KS model. The KS equation is a fourth-order partial differential equation initially used to describe diffusive thermal instabilities in laminar flame fronts (Kuramoto 1978; Sivashinsky 1977, 1980). We will not give a detailed account of the model here but refer to the discussion in Evensen et al. (2024). Several important properties of the model make it suitable as a testbed for studying our data assimilation methods. The model describes unstable and near-chaotic dynamics with nonlinear saturation of the linear instabilities, similar to oceanic and atmospheric behavior. It contains a one-dimensional and univariate space–time equation for each component of the coupled system, which makes it computationally efficient and easy to interpret results. In Evensen et al. (2024), we used the same time scale for both model components, but the spatial scales differed. We have also included and studied the impact of differences in time scales in the current study. The model system should resemble the behavior of coupled climate models, where the ocean and atmospheric components have vastly different spatial and temporal scales.

In our implementation, we couple an Atmos and an Ocean with the symbols A and O referring to their respective variables. The coupled model equations read

$$\tau_a \frac{\partial A}{\partial t} = -\frac{\lambda_a}{2}\frac{\partial A^2}{\partial x} - \lambda_a^2 \frac{\partial^2 A}{\partial x^2} - \frac{\lambda_a^4}{2}\frac{\partial^4 A}{\partial x^4} + \alpha_{oa}(O - A), \tag{1}$$

$$\tau_o \frac{\partial O}{\partial t} = -\frac{\lambda_o}{2}\frac{\partial O^2}{\partial x} - \lambda_o^2 \frac{\partial^2 O}{\partial x^2} - \lambda_o^4 \frac{\partial^4 O}{\partial x^4} + \omega_{ao}(A - O), \tag{2}$$

defined on the spatial domain $x \in [0{:}1024]$. Note that, contrary to the ad hoc approach used by Evensen et al. (2024), we introduced λ_a and λ_o as the spatial scales, and we introduced τ_a and τ_o for the temporal scales for Atmos and Ocean, respectively. In this manner, we made the spatial and temporal scales explicit in the equations. Multiplying the Ocean time derivative by the factor $\tau_o = 4$ and the Atmos time derivative by the factor $\tau_a = 1$, we ensure that the ocean dynamics evolve slower than the atmospheric ones. To introduce different spatial scales for the Atmos and Ocean equations, we scale the x coordinates in the two equations corresponding to a domain of length $L_a = 32$ for Atmos and $L_o = 256$ for Ocean as in Evensen et al. (2024). Transformed back to the standard x coordinate varying from 0 to 1024, we obtain the model Eqs. (1) and (2). We have used the factors $\lambda_a = 1024/32 = 32$ and $\lambda_o = 1024/256 = 4$, leading to a difference in the spatial scale of a factor of 8 between Atmos and Ocean.

We couple the two equations through the relaxation terms $\alpha_{oa}(O - A)$ and $\omega_{ao}(A - O)$, where we use coupling coefficients of 0.003 in both equations. Additionally, we halve the biharmonic damping of the Atmos variable to have more structures in the solutions.

Our multiscale KS model is suitable for conceptualizing data assimilation in coupled systems. The interaction between the different components occurs at multiple scales, which, as we will show, makes it difficult to apply distance-based localization. We can consider it a 1D analog to the 2D Navier–Stokes equations. The model has a unique solution given the initial conditions, but it has chaotic behavior and a finite-dimensional global attractor. The 1D KS model simplifies the analysis and interpretation of results, and we avoid using computationally expensive 2D or 3D models. The FORTRAN-90 implementation we use in all experiments is available from Evensen (2023).

3. Correlation and distance-based localization

Localization is a standard technique for reducing the sampling errors' effect on the quality of updates in ensemble-based data-assimilation methods.[1] When optimally applied to large-scale problems with small ensemble sizes, localization increases the solution space's degrees of freedom, allowing

[1] Section 3 is based on Evensen et al. (2025, chapter 7).

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

for a better match to data. Furthermore, localization reduces the spurious model updates caused by sampling errors and thereby the tendency for the ensemble to "collapse" when we condition on a large number of independent measurements. We refer to Evensen et al. (2022; chapter 10) for an overview of localization methods.

So-called distance-based localization is the most commonly used localization method in meteorology and oceanography. In distance-based localization, the dynamical system's physical scales determine the truncation distance. The dominant scale in the ocean and atmosphere is the mesoscale, which is associated with the Rossby radius of deformation. The procedure for selecting a localization radius, typically a few Rossby radii, must ensure that measurements of, e.g., temperature, sea surface height, and pressure only influence the model's state variables in a region close to the measurements. For these problems, distance-based localization gives results almost equivalent to the infinite ensemble size limit for as long as the measurements are local and we have a well-defined decorrelation length. The first publications on distance-based localization introduced covariance localization (Hamill 2001; Houtekamer and Mitchell 2001; Bishop et al. 2001; Whitaker and Hamill 2002) or Kalman gain localization (Anderson 2003), where a tapering function damps long-range spurious correlations. In covariance localization, one tapers the covariance functions, while in Kalman gain localization, it is applied directly to the Kalman gain (Chen and Oliver 2017).

Covariance localization typically uses a damping operator to eliminate long-range spurious correlations in the state covariance matrix. Using a Schur (or Hadamard) product, the damping operator acts on the covariance through element-wise multiplication of two matrices: the ensemble covariance matrix with a damping correlation matrix. The operator applies a scaling factor equal to one at the location of the measurement and gradually reduces to zero further away. A commonly used damping function is the one by Gaspari and Cohn (1999), while Furrer and Bengtsson (2007) provide an alternative.

In cases when it is difficult to determine a fixed truncation distance in space and time for distance-based localization, adaptive correlation-based localization methods can help prevent ensemble collapse and reduce data mismatch. We will see below in Fig. 7 that the KS model poses some problems for distance-based localization due to the spatial periodicity of the solution. The KS model generates significant oscillating correlations for the Atmos variable extending to large distances, and using distance-based localization, it is not clear where to truncate. Adaptive correlation-based localization has proved very useful in petroleum reservoir applications, where it is challenging to determine a fixed localization distance as it would depend on the reservoir's geology and the placement of wells used to produce the reservoir.

The methods of Chen et al. (2009), Emerick and Reynolds (2011), Bishop and Hodyss (2007, 2009a,b), and Anderson (2012) provide localization based on the correlation value, typically depending on ensemble size. Because the adaptive techniques rely on the ability to separate spurious correlations from actual (large ensemble size) correlations, their performance is generally poorer than distance-based localization when distance based is appropriate. In statistics, the methodology is known as thresholding (see, e.g., Bickel and Levina 2008).

We will use a version of the so-called local analysis introduced by Haugen et al. (2002) and Evensen (2003) both for the distance-based and the adaptive correlation-based localization. Local analysis is a computational approach to efficiently update states or parameters in the data assimilation analysis by introducing individual update calculations for each local state variable at a grid point. Further details are provided below and can be found in Evensen (2003). While the local analysis of Haugen et al. (2002) updated all variables on a small cube of grid points using only nearby located measurements, we will in the following section introduce an adaptive correlation-based localization in combination with a local analysis scheme.

a. Adaptive correlation-based localization

The method of adaptive correlation-based localization is a straightforward implementation of a local analysis scheme where we update the parameters, or rows in the ensemble state matrix, sequentially one by one using only the measurements significantly correlated to the state variable or parameter at the location of the update. It uses the estimated correlations between "predicted measurements" and the state variables to determine if a measurement shall influence a state variable's update. In an update, we only retain the observations with the absolute value of a correlation above a certain threshold. Note that, in the remainder of the paper, when we refer to the magnitude of a correlation, we mean the absolute value of the correlation.

Adaptive localization sometimes has advantages over distance-based localization, e.g., when conditioning the model on non-local measurements when the truncation distance is unknown or when there are multiple distances at which the measurements correlate to the state variable or parameter. The main advantage is that we can implement adaptive correlation-based localization without referring to the spatial location of the updated variables. Thus, we only need the ensemble matrices of the predicted state and the predicted measurements to apply the localization. This is especially important in ensemble smoothers, where a physical-distance-based high correlation area becomes time dependent because observation information is advected with the flow (Brusdal et al. 2003). Adaptive correlation-based localization does not have this problem by construction. We will present a theoretical foundation for selecting the truncation value as a function of the ensemble size and discuss some possible remedies for further incremental improvement of the method.

Several publications have discussed various implementations of adaptive correlation-based localization, e.g., Evensen (2009, chapter 15.5), Bishop and Hodyss (2007, 2009a,b), Zhang and Oliver (2010), Anderson (2012, 2016), Luo and Bhakta (2020), Neto et al. (2021), Luo and Xia (2022), and Vishny et al. (2024). We will focus on the correlation truncation approach from Evensen (2009, chapter 15.5), which was also used by

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

Neto et al. (2021), Luo and Xia (2022), and Luo and Bhakta (2020).

An issue with adaptive correlation-based localization is that removing all measurements with an estimated correlation value less than the truncation value will also remove information from measurements with physical but low correlations. However, we would have similar problems when using a distance-based truncation. With correlation-based localization, we will include some measurements with spurious strong correlations. Similarly, with distance-based localization, we will retain some measurements with insignificant information located within the truncation distance.

b. Implementation strategy

There are several ensemble smoothers that we can choose from, such as the ensemble smoother (ES), the ES with multiple data assimilations (ESMDAs), and the iterative ensemble Kalman smoother (IES). We can write the global update steps using ES, ESMDA, and IES as

$$\mathbf{Z}^a = \mathbf{Z}^f\mathbf{T}, \tag{3}$$

where the posterior ensemble in $\mathbf{Z}^a \in \Re^{n\times N}$ is a linear combination of the prior ensemble in $\mathbf{Z}^f$. The transition matrix $\mathbf{T} \in \Re^{N\times N}$ determines the linear combination. Here, n is the state dimension, N is the number of realizations, and typically $n \gg N$. The matrix multiplication in Eq. (3) is the most expensive computation in the analysis scheme, requiring nN^2 floating point operations. However, Eq. (3) is highly parallelizable as we can compute each row in $\mathbf{Z}^a$, and hence each model variable, independently of the others.

The local analysis exploits the parallelization of the computation in Eq. (3), and we can write it as the n local updates for $l = 1 \ldots n$ computed from

$$\mathbf{Z}^a_l = \mathbf{Z}^f_l\mathbf{T}_l, \tag{4}$$

where the subscript l denotes the lth row of $\mathbf{Z}$. In the local analysis, we update each row $\mathbf{Z}_l$ using an individual transition matrix $\mathbf{T}_l$, which for each l may differ in the measurements used in its calculation.

The distance- and correlation-based localization implementations become very similar when writing the update as Eq. (4). The methods only differ in the selection procedure for selecting which observations to use in each local update. We can interpret correlation-based localization as applying another distance measure to define nearby and distant measurements.

A local update transforms the problem of finding one $\mathbf{T}$ matrix to n matrices $\mathbf{T}_l$, where n is the dimension of the state variable. The increased computational cost is an issue common for adaptive and distance-based localization. We must compute many small inversions in both approaches to obtain the n local $\mathbf{T}_l$ matrices. Still, for each $\mathbf{T}_l$, we will use a far lower number of measurements in the updates, reducing the local cost to $\mathcal{O}(m_l N^2)$ operations, where m_l is the number of observations used in each local analysis update. Note also that there may be no measurements for many state vector elements, in which case the transition matrix is just the identity matrix.

The local analysis allows updating each grid point's state variables one by one, as explained by Eq. (4). However, this is not a restriction, as we can update any subset of variables together using the same transition matrix. In many cases, it will be beneficial to define subsets of the state vector, e.g., the subsets of all the model's variables at the same grid points, as this ensures a joint and balanced update of the different model variables at each grid point. In some models, the subset can include all variables in a vertical grid column to ensure a balanced update respecting the model physics vertically. Defining such subsets of variables for each local update reduces the computational cost as we compute significantly fewer transition matrices $\mathbf{T}_l$.

In distance-based localization, we need to compute the Euclidian distance between each state variable and each measurement to a cost of order $\mathcal{O}(nm)$. It is often possible to significantly reduce this cost by using clever storage schemes for the measurements and by combining clusters of nearby grid points into one distance calculation.

A significant additional computation in the correlation-based localization is evaluating the correlation function between the state variables and the measurements, which fills a matrix $\mathbf{C}_{zy} \in \Re^{n\times m}$. Its computation requires nmN floating point operations to obtain the covariance $\mathbf{C}_{zy} = \mathbf{A}\mathbf{Y}^{\mathrm{T}}$, in which $\mathbf{A}$ is the forecast ensemble perturbation matrix and $\mathbf{Y}$ is the predicted measurement perturbation matrix. To obtain the correlation matrix, we must also normalize by the standard deviations of $\mathbf{A}$ and $\mathbf{Y}$. This computation significantly increases the cost of calculating which measurements to use for each state vector update. The total operational cost will be the accumulated costs of computing the correlation matrix (nmN), the n local inversions (nm_lN^2), and the final update equation, which is still nN^2 number of operations. We can write the total cost as $n(mN + m_lN^2 + N^2)$, showing that it is still linear in the state size n, but it depends on the total number of measurements and also the number of measurements retained in each local update.

Another primary concern is the storage of $\mathbf{C}_{zy} \in \Re^{n\times m}$, which becomes challenging for large models with many measurements. However, we expect many adjacent variables to use the same observations in the local update. Recall that the transition matrices depend only on the active measurements in the analysis. Thus, a better strategy is to compute the correlations for reasonably sized "grid blocks." By grid blocks, we mean a set of selected neighboring grid points that we update simultaneously. Using these grid blocks, we compute the correlations for blocks of variables in rows of $\mathbf{Z}$ (typically variables defined on neighboring grid points in an assimilation window), then identify which observations are active when updating each such block of variables, and calculate a joint transition matrix for this block of state variables. By dividing the state ensemble matrix $\mathbf{Z}$ into subblocks, we avoid the excessive storage of huge matrices. Furthermore, we can significantly reduce the number of inversions by reusing the transition matrices $\mathbf{T}_l$ for all state variables updated with the same set of measurements.

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

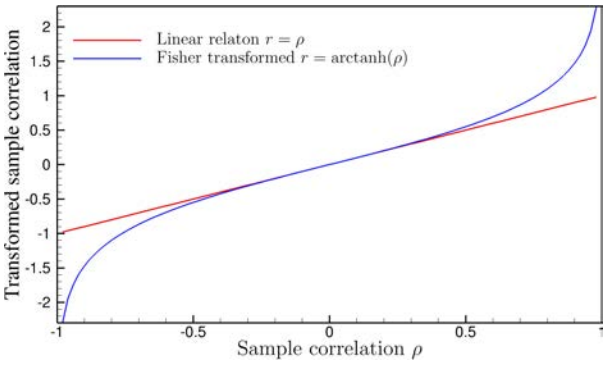

FIG. 1. The blue line in this plot shows the Fisher transformation from Eq. (5). Note that the difference of the transformation with the linear relation $r = \rho$ (in red) is negligible for correlations in the interval $[-0.6{:}0.6]$, where we apply the correlation-based localization. Thus, the Gaussian assumption on the distributions of sample correlations in Fig. 3 is valid.

c. *Selecting a correlation truncation value*

It is possible to derive a theoretical value for the sample-correlation standard deviation using the Fisher transform; see Guttman et al. (1982) and Flowerdew [2015; Eq. (9)]. It says that the variable r,

$$r_{ij} = \frac{1}{2}\ln\frac{1+\rho_{ij}}{1-\rho_{ij}} = \text{arc}\tanh(\rho_{ij}), \tag{5}$$

evaluated from the correlation between the measurement number j and the state variable number i will be approximately normally distributed with a standard deviation of $1/\sqrt{N-3}$. We need the sample standard deviation of the correlation ρ for adaptive correlation-based localization. First, the Fisher transform shows that the sampling error for high correlations (i.e., close to one) is much lower than for low correlations. Thus, data assimilation updates with measurements highly correlated to the state variables will have a lower sampling error than those with lower correlations. This result is a point in favor of correlation-based localization. Furthermore, from Fig. 1, we see that the standard deviation estimates of ρ and r are nearly identical for the low correlations we wish to truncate.

In the case of physically uncorrelated measurements and state variables, the actual correlation will be zero, and the finite ensemble size will lead to spurious correlations with a normal distribution $\mathcal{N}[0, 1/(N-3)]$, i.e., with zero mean and a standard deviation of $1/\sqrt{N-3}$. To remove 99.7% of these spurious correlations, we can truncate all measurements with predicted correlations to a state variable of less than three standard deviations (see Fig. 2). The choice of three standard deviations makes sense since we wish to remove "all" spuriously correlated measurements, so, as a first-guess value, we set the truncation to

$$\rho_t = 3/\sqrt{N}. \tag{6}$$

In Fig. 3, we illustrate the truncation of measurements with different levels of physical, or true, correlation. We use an ensemble size of $N = 100$ and a truncation defined by Eq. (6) and assume its distribution is approximately Gaussian, which is a valid approximation in the range of correlation values we consider; see Fig. 1 and Flowerdew (2015; Fig. 1).

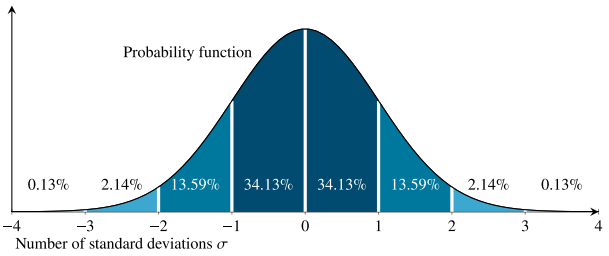

FIG. 2. The plot shows the probability density function of a Gaussian distribution with the percentages of samples falling in different parts of the pdf.

In the upper plot of Fig. 3, we assume zero physical correlation between a measurement and the state variable, as would be the case for a large portion of the variables in the model domain since the model domain's size is significantly larger than the characteristic length scales of the model variables. When using a truncation value defined by Eq. (6), we will remove an uncorrelated measurement in 99.7% of the cases when truncating based on the sample correlation (the red area). The middle plot of Fig. 3 illustrates that when the measurement has a low physical correlation of 0.2, we remove the measurement in 84.1% of the cases. An alternative interpretation is that with many measurements with a correlation of 0.2, we will retain 15.9% of them. In the case of the lower plot, where the correlation value is 0.3, we will keep half the measurements when the correlation equals the truncation value.

By applying this truncation scheme, we will also remove some measurements with a higher physical correlation located in the range $(\rho_t{:}2\rho_t)$, as we truncate these measurements if the

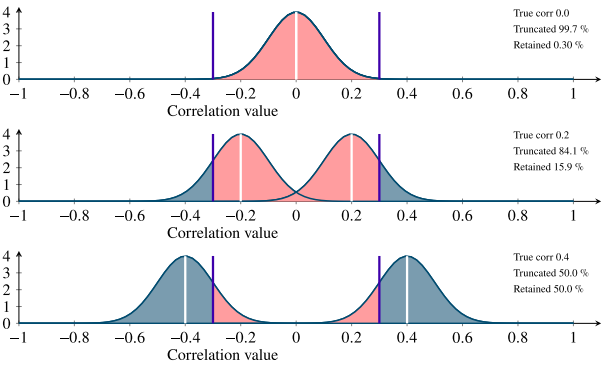

FIG. 3. The figure illustrates the truncation's impact for measurements with an actual physical correlation of 0.0, 0.2, and 0.4 in the three panels from top to bottom, as denoted by the white vertical line, in the case with sampling errors corresponding to using 100 realizations. The distribution's red part corresponds to the truncated measurements, while the blue part indicates the retained measurements. The symmetry in the plots accounts for both positive and negative correlations. The vertical blue lines indicate the theoretical truncation value from Eq. (6).

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

sampling correlations are below the truncation value ρ_t. We will truncate 15.9% of the measurements with a correlation equal to 0.4 and 2.27% of those with a truncation value equal to 0.5. There is barely any truncation for measurements with a correlation above 0.6.

In adaptive localization, we remove a fraction of measurements with significant physical correlations. We only remove those with a low sample correlation that would have had less impact on the update anyway. This implies that we will lose some information by truncating measurements with significant physical correlations, and we will include some measurements with small but substantial spurious correlations. Distance-based localization will also include measurements with negligible physical correlation and truncate measurements with high correlations outside the localization radius.

d. Tapering distant measurements

In standard covariance localization methods (Evensen et al. 2022, chapter 10), one uses a tapering of the covariance functions damping long-range spurious correlations. In local analysis, we update grid point by grid point and thereby risk using slightly different observations when computing the updates at two adjacent grid points. This approach will introduce small discontinuities in the updated realizations. Thus, to ensure smooth updates, we introduce tapering by inflating the errors of the most distant observations. This error inflation reduces these observations' impact on the analysis and effectively works as a tapering of the covariance functions.

We can follow a similar procedure for correlation-based localization. We define a correlation distance similar to the Euclidean distance used in distance-based localization as

$$d_c(l, j) = 1 - |\rho(l, j)| \in [0, 1], \tag{7}$$

where $\rho(l, j)$ is the ensemble correlation between state variable number l and measurement number j. If the correlation equals one or minus one, we have a minimum distance of zero; if the correlation is zero, we obtain a maximum distance of one. The truncation distance is then

$$d_t = 1 - \rho_t. \tag{8}$$

We use observation error inflation to reduce the impact of the most remotely located measurements within the truncation distance. We prescribe the error inflation factor as

$$E_{\text{inf}}(l, j) = \begin{cases} 1 & d_c(l, j) \le \beta d_t \\ \exp\left(\dfrac{d_c(l, j) - \beta d_t}{b}\right)^2 & d_c(l, j) > \beta d_t \end{cases}, \tag{9}$$

where $\beta \in (0, 1)$ defines the distance at which the measurement error inflation becomes active. If we require that the error inflation equals the maximum inflation, $E_{\text{inf}} = E_{\max}$, when $d_c(l, j) = d_t$, we obtain the following value for the parameter b in Eq. (9),

$$b = \frac{(1 - \beta) d_t}{\sqrt{\ln(E_{\max})}}. \tag{10}$$

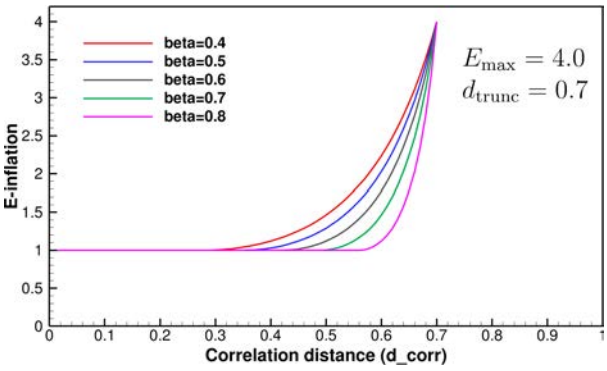

FIG. 4. This plot illustrates the error inflation $E_{\text{inf}}(d_c)$ for an example where we set $\rho_t = 0.3$, which gives $d_t = 0.7$, and the maximum inflation at d_t is $E_{\max} = 4.0$, and several values of β, which define at what correlation distance, βd_t, we start inflating.

So, we retain the original observation errors for measurements at a distance $d_c(l, j)$ closer to the grid point than βd_t. For observations located further away, we deliberately inflate the errors by the factor $E_{\text{inf}}(l, j)$ to reduce their impact on the update. This procedure ensures that when we introduce a new measurement to the local analysis when moving from one grid point to the next, the new measurement will have less impact on the analysis of this adjacent grid point. The effect of the new observation will gradually increase when its correlation with either one of the Ocean or the Atmos variables becomes more significant. It is straightforward to use the same tapering strategy in standard distance-based localization. Figure 4 shows the inflation factor for different values of β. In the following experiments, we set $\beta = 0.5$ and examine different values for $E_{\max}$. We introduce the observation error inflation by multiplying each row in $\mathbf{E}$ with the inflation factor E_{inf} corresponding to the measurement in question.

e. Pragmatic implementation

The transition matrix is obtained following Evensen et al. [2022, Eqs. (8.32)–(8.34)]:

$$\mathbf{T} = \mathbf{I} + \mathbf{W}/\sqrt{N - 1}, \tag{11}$$

where

$$\mathbf{W} = \mathbf{S}^{\mathrm{T}}(\mathbf{S}\mathbf{S}^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}})^{-1}[\mathbf{D} - \mathbf{g}(\mathbf{Z}^f)], \tag{12}$$

and

$$\mathbf{S} = \begin{cases} \mathbf{Y} & \text{for } n \ge N - 1 \\ \mathbf{Y}\mathbf{A}^{\dagger}\mathbf{A} & \text{for } n < N - 1. \end{cases} \tag{13}$$

In the equations above, the measurement matrix $\mathbf{D} \in \Re^{m \times N}$ contains all measurements at an update step, and we represent the measurement errors by the measurement perturbation matrix $\mathbf{E} \in \mathbf{R}^{m \times N}$. In the case of correlated observations, we specify correlated perturbations in $\mathbf{E}$. Similarly to the perturbation matrix $\mathbf{E}$, we have the predicted measurement anomalies in $\mathbf{Y} \in \Re^{m \times N}$.

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

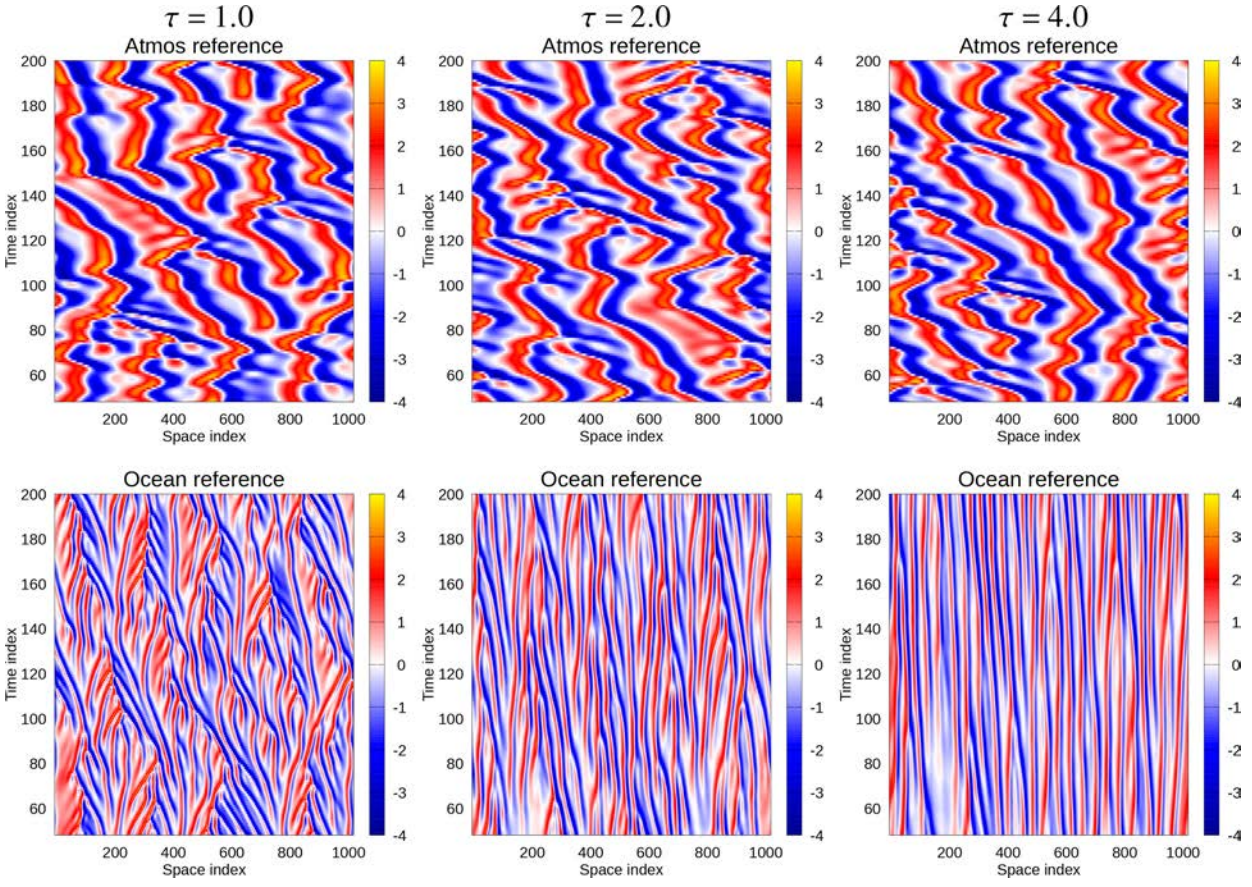

FIG. 5. The plots show the Hovmöller diagrams for (top) Atmos and (bottom) Ocean variables from coupled predictions using different time scales $\tau = 1.0$, 2.0, and 4.0 from left to right.

For each local update, we extract the rows corresponding to the measurements located within the truncation distance from $\mathbf{D}$, $\mathbf{Y}$, and $\mathbf{E}$ and store them in $\mathbf{D}_l \in \Re^{m_l \times N}$, $\mathbf{Y}_l \in \Re^{m_l \times N}$, and $\mathbf{E}_l \in \Re^{m_l \times N}$. We implement the tapering according to Eq. (9) by multiplying each row $\mathbf{E}_l$ by the corresponding inflation factor $E_{\text{inf}}(l, j)$.

To reduce the number of inversions and the computational cost, we define the state vector to contain both the Atmos and Ocean variables at the spatial grid point l for all times within the assimilation window. In the experiments below, we use an assimilation window of six time units, implying that we update 12 state variables (one for Ocean and one for Atmos for each time unit in the window) in each local update.

To select the active measurements for each local update, we compute the correlations between Ocean and Atmos with the predicted measurements at the final time of the assimilation window. If the correlation between the measurement and either the Ocean or the Atmos variable is larger than the truncation correlation, we include the measurement in the analysis. With this simplification, we reduce the number of required inversions by a factor of 12 and the number of computed correlations by a factor of 6. We can further reduce the computational cost by including larger subgroups of variables in each local update. However, our relatively low-dimensional KS model runs efficiently with this configuration.

4. Reference case and prediction experiments

We define a reference case for the coupled KS model using an Ocean component 4 times slower than the Atmos component, i.e., we set $\tau = 4$ in Eq. (2). Figure 5 shows examples of a free run of the coupled system for different values of τ. First, we notice the highly nonlinear model evolution and the difference in scales between the Ocean and Atmos variables, where the Ocean variable contains small-scale, slow features and the Atmos contains large-scale, relatively faster ones. It is clear that the coupling affects the behavior of both components and that the time-scale change for the Ocean component significantly impacts the evolution of both systems.

In Fig. 6, we show the time development of the ensemble standard deviation and the residual between a reference solution and the ensemble mean for the three values of τ. It is clear that the value of τ does not impact the climatological ensemble spread, only the time it takes to reach climatology for the Ocean component.

Finally, we present ensemble correlation functions from Ocean and Atmos observations with the model variables in

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

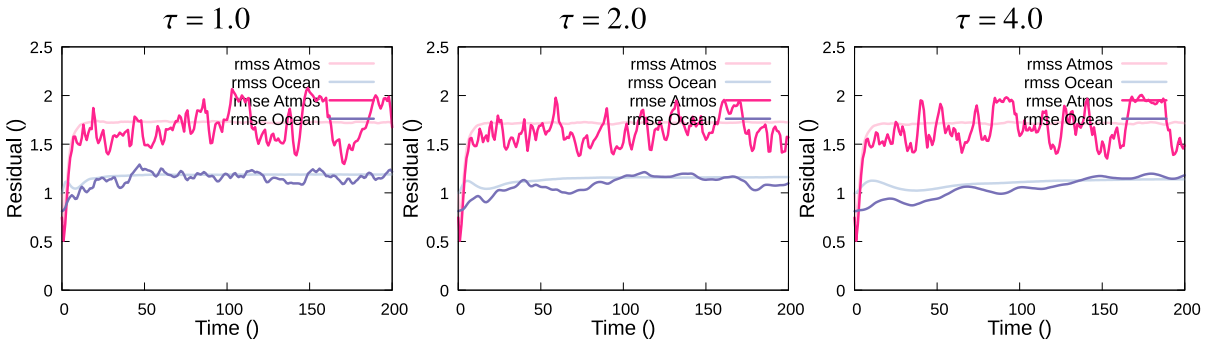

FIG. 6. The plots show the time evolution of the residuals for the ensemble predictions using different values of τ. The blue and red lines are for the Ocean and Atmos variables, respectively. The dark lines indicate RMSEs relative to the reference solution. The light lines are the ensemble-predicted RMS standard deviations denoted as RMSS.

Fig. 7. We observe the Ocean and the Atmos variables at one grid point in the middle of the domain and the middle of the simulation period. We notice the extension in time of the OceanObs–Ocean correlation when the Ocean component becomes slower with $\tau = 4$, and we also observe that the cross correlations between the Ocean and Atmos reduce in this case compared to the case of $\tau = 1$. The result is as expected when we acknowledge that in the case of the Ocean having an infinite time scale, there would be no covariation at all. A practical implication is that in a more realistic coupled ocean and atmosphere model, the correlation will be less than in the KS model with $\tau = 1$, at least for the dynamical variables describing the mesoscale dynamics. On the other hand, the variability of ocean mixed-layer temperature and the atmospheric temperature may have stronger correlations than the ones represented by the KS model, as, in reality, these are tightly connected.

5. Data assimilation experiments

We have run many data assimilation experiments to test different distance-based and adaptive correlation-based localization-scheme configurations. We examined the impact of the following parameters in these different configurations: the truncation length in the distance-based localization, the truncation correlation-distance value, and the error variance inflation for the most distant measurements in the adaptive correlation-based localization. We have also run experiments with different ensemble sizes, although we have focused on using 100 realizations in the ensemble.

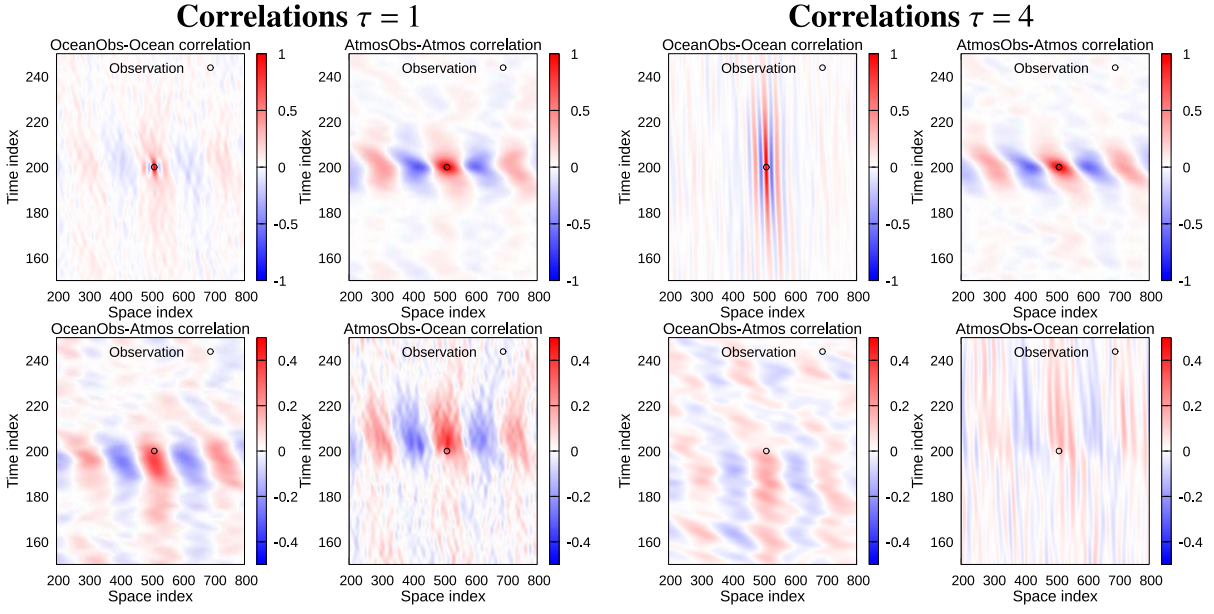

FIG. 7. Here, we show the correlations from coupled ensemble simulations for $\tau = 0$ and $\tau = 4$. (top) The correlations of an Ocean observation with the Ocean variable and an Atmos observation with the Atmos variable. (bottom) In contrast, the cross correlations between an Ocean observation with the Atmos variable and the Atmos observation with the Ocean variable.

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

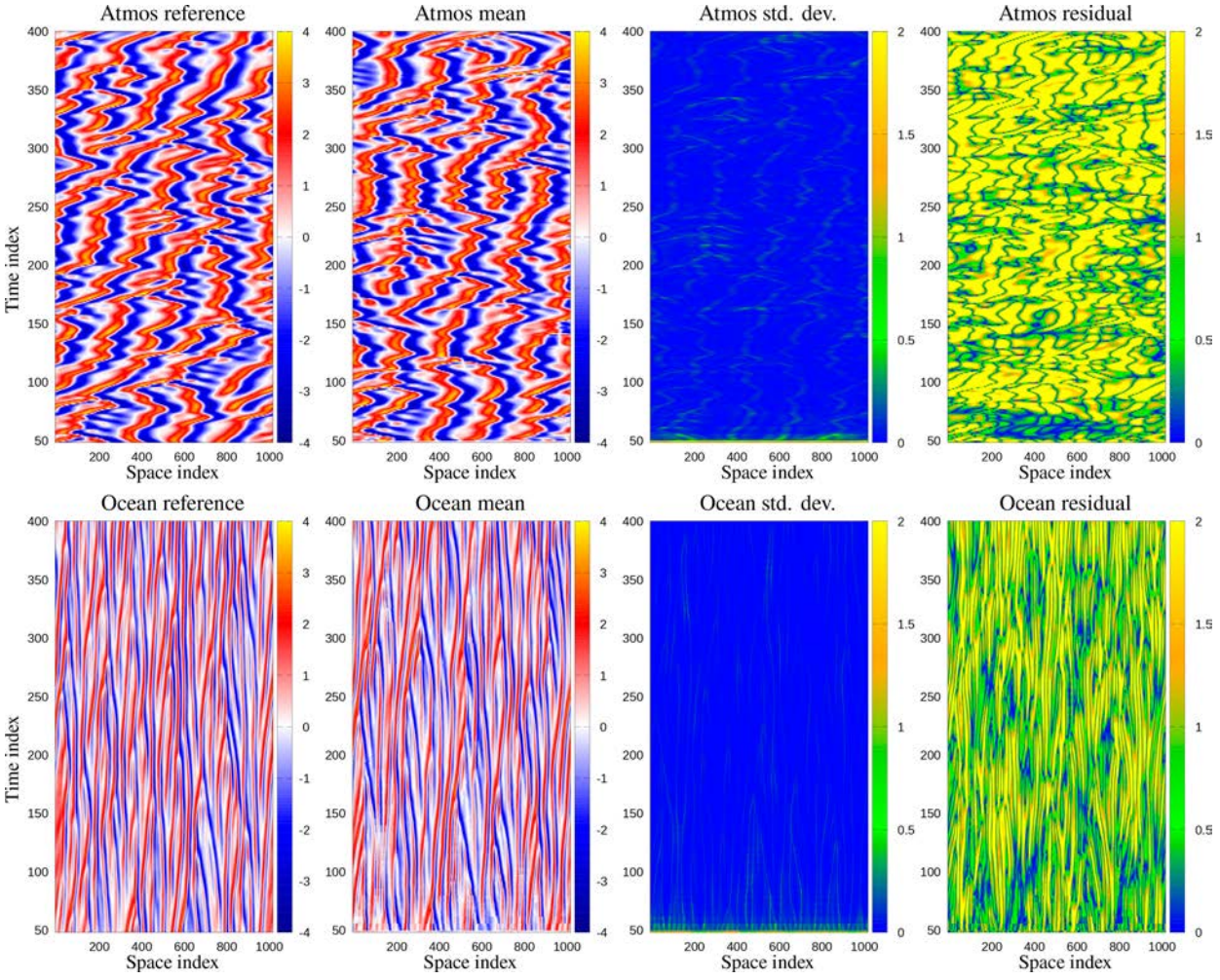

FIG. 8. The figure shows the Hovmöller plots from a global update with $N = 100$ realizations. (top) The Atmos variable and (bottom) the Ocean variable. We plot the solution for the period of the assimilation experiment from $t = 50$ to $t = 400$ and do not include the spinup ensemble integration from $t = 0$ to $t = 50$. The assimilation windows have a length of six units of time, and we plot the smoother estimate for all six units of time in every assimilation window. (first column) The respective variables' reference solution. (second column) The estimated posterior ensemble mean. (third column) The estimated ensemble standard deviations. (fourth column) The absolute value of the residual between the estimate and the reference solution.

We use a reference case where we run the model from time 0 to time 400. We have doubled the simulation length compared to the experiments in Evensen et al. (2024) since the Ocean is now 4 times slower, and we need a longer time integration to study the Ocean behavior. After 50 time units of spinup time, we sample the Ocean and Atmos variables to create observations. In all data assimilation experiments, we use the same measurement density, where we collect measurements every two time units, with a spatial resolution of every 15 grid points for the Ocean variable and every 45 grid points for the Atmos variable. Hence, there are three Ocean measurements for every Atmos measurement. Considering that the spatial scale ratio between Atmos and Ocean is eight, the Ocean variable has significantly fewer observations per length scale than the Atmos variable, i.e., 3/8 of the measurement density of the Atmos variable. As we will see below, the less dense sampling of the Ocean leads to a weaker control of the Ocean's evolution.

As the Ocean variable is evolving 4 times slower, the choice of $\tau = 4$ makes it easier to control as we have a 4 times higher relative sampling rate of the Ocean variable than the Atmos variable. We use the first period from $t = 0$ to $t = 50$ as an ensemble spinup period, and we do not generate any observations for this period. For the remainder of the simulation period, we simulate observational error in the assimilated observations by adding random noise from a normal distribution with a mean of zero and a standard deviation of 0.3 for both Ocean and Atmos measurements.

For the data assimilation experiments, we use the ESMDA as described in Evensen et al. (2024), where we found that the choice of four MDA steps is close to optimal for the current configuration. We defined an assimilation window of length

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

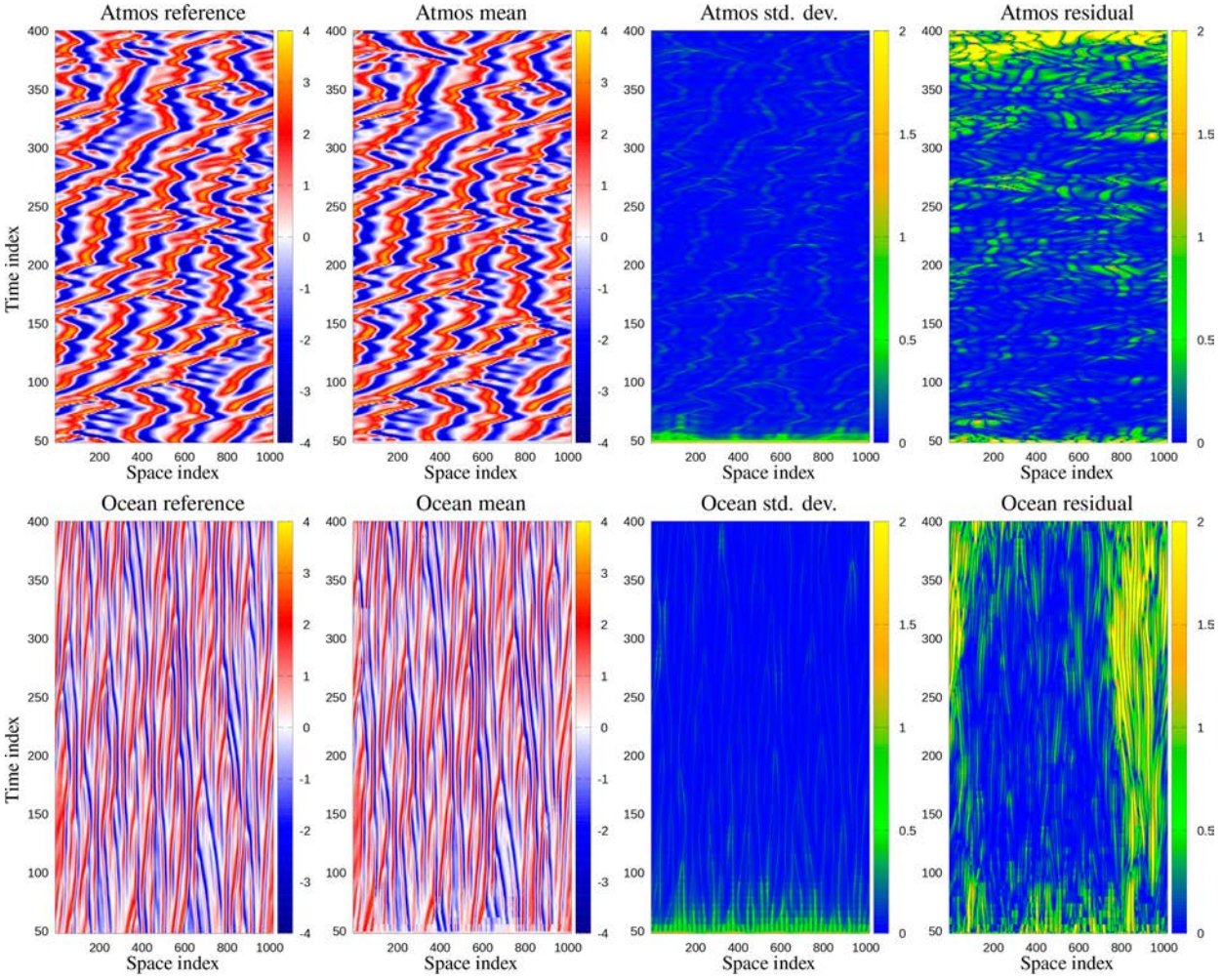

FIG. 9. The Hovmöller plots show the results from a global update with $N = 400$ realizations. See the description in Fig. 8.

equal to six units of time, which means that we have a grid of 15×3 measurements for the Atmos variable and 45×3 measurements for Ocean in each update step. Each MDA step is independent of the previous ones, and we recompute the correlations used to select measurements in each MDA step.

As discussed in Evensen et al. (2024), the optimal strategy when using ESMDA is to directly update the solution in the whole data assimilation window in the final update step rather than integrating the model over the window from updated initial conditions.

a. *Global updates*

Four experiments with global updates, as in Evensen et al. (2024), illustrate how ensemble size affects the analysis. Figures 8–10 give the results for the cases where we use 100, 400, and 1000 realizations. We name these reference experiments "global update experiments," in which we do not apply localization. When we use 100 model realizations, we have no skill in estimating the solution. We experience filter divergence, where the ensemble standard deviation does not correspond to the expected variance in the residuals. In the case with an ensemble size of 400 model realizations, the estimation improves, but there are remaining residuals, illustrated by the root-mean-square error (RMSE), that we are unable to reduce and correct. Moreover, filter divergence affects the Atmos variable estimate, where the residuals start growing at the end of the assimilation period while the ensemble standard deviation remains low, and it solves the space–time localization distance issue that has hindered ensemble smoothers for a long time. On the other hand, when using 1000 model realizations, we obtain low RMSE values and an ensemble standard deviation consistent with the expected variance in the residuals, suggesting we avoid filter divergence. Note that the global update results with $N = 400$ could benefit from adding inflation, but we did not pursue this idea in the current paper as we focus on localization methods.

b. *Local analysis*

When using local analysis, we must tune some parameters, whether we are using a distance- or correlation-based truncation of measurements. For distance-based localization, we need to determine the appropriate truncation distance.

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

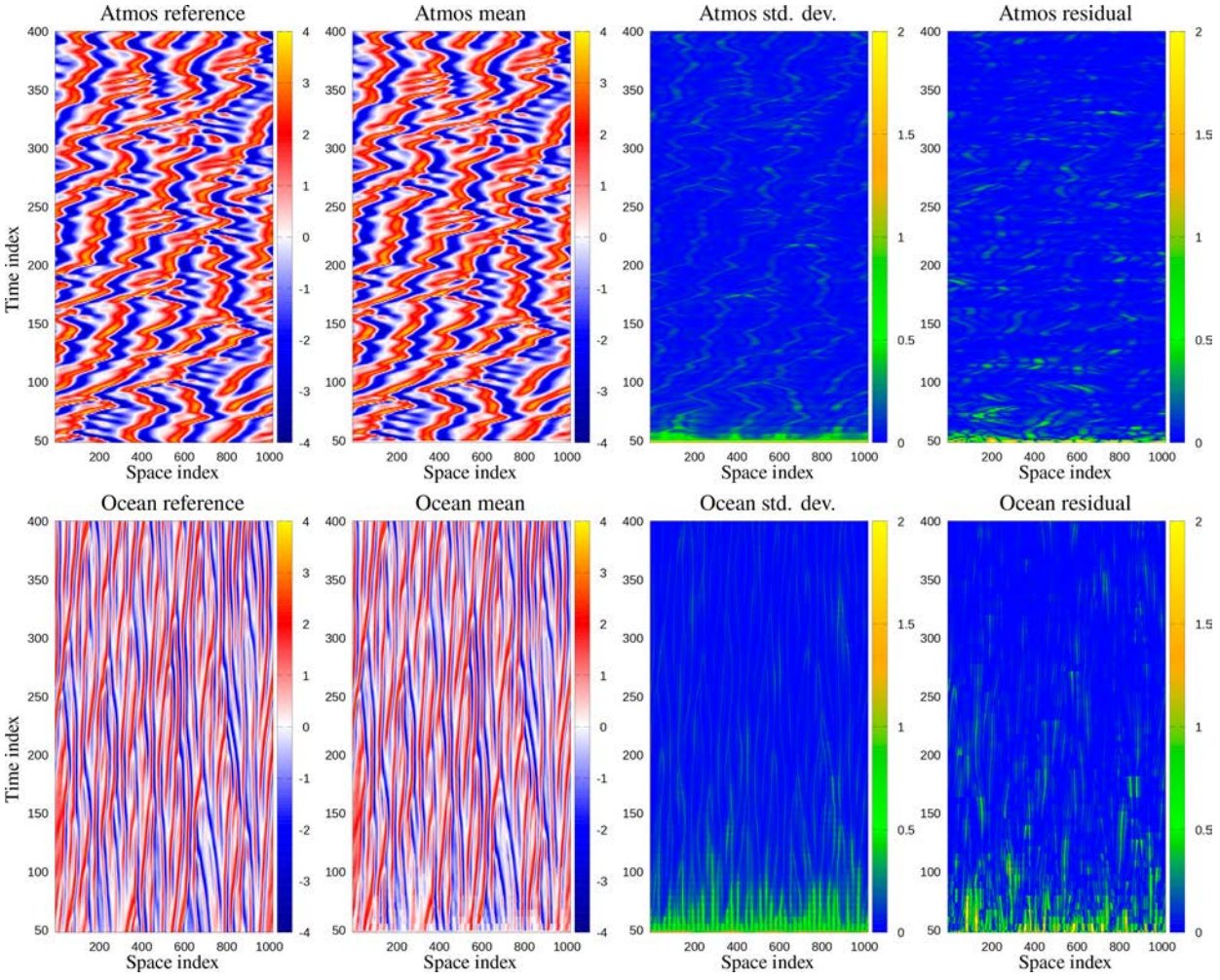

FIG. 10. The Hovmöller plots show the results from a global update with $N = 1000$ realizations. See the description in Fig. 8.

Typically, one would set the truncation distance large enough to include the characteristic length scales of the system. For the correlation-based localization discussed in section 1, we need to determine the truncation correlation distance where we have a theoretical value of $3/\sqrt{N-3}$ as a first guess. We can use the additional error inflation of the most remote measurements for both localization approaches, which requires setting a value for the maximum measurement error variance inflation $E_{\max}$ in Eq. (10). In the current application, we do not expect that the choice of β in Eq. (9) will have a significant impact, and we use $\beta = 0.5$. To assess the effect of these parameters, we run the experiments that we summarize in Fig. 11.

From Fig. 11, we observe that the distance-based localization performs best when using a localization truncation distance of about 50 grid points. It is interesting to note that the residuals for the Ocean variable are smaller for even shorter localization lengths, but for those localization lengths, the Atmos residual is larger. In additional experiments described below, we investigate the possibility of using different truncation distances for the Ocean and the Atmos variables. Additional measurement variance inflation for remote measurements has a negligible impact in the cases of Fig. 11 using distance-based localization. On the other hand, in the adaptive correlation-based localization, the error variance inflation has a considerable impact. We tried values of $E_{\max}$ equal to one, two, four, eight, and sixteen. We obtained the best results for distance-based localization using $E_{\max} = 4$. For correlation-based localization, $E_{\max} = 8$ gave the best results. It may be surprising to see the vast improvement we obtain by increasing $E_{\max}$ from one to eight, but it is clear that this additional error inflation is crucial for correlation-based localization. Figure 11 highlights that we obtain significantly better results using adaptive correlation-based localization than the physical distance-based one.

We furthermore observe that we obtain better results with correlation truncation values more significant than the correlation truncation value derived from Eq. (6). Based on the results of Fig. 11, we use truncation values between 0.3 and 0.4 corresponding to truncation distances d_t between 0.6 and 0.7. Thus, we can use a relatively strong truncation

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

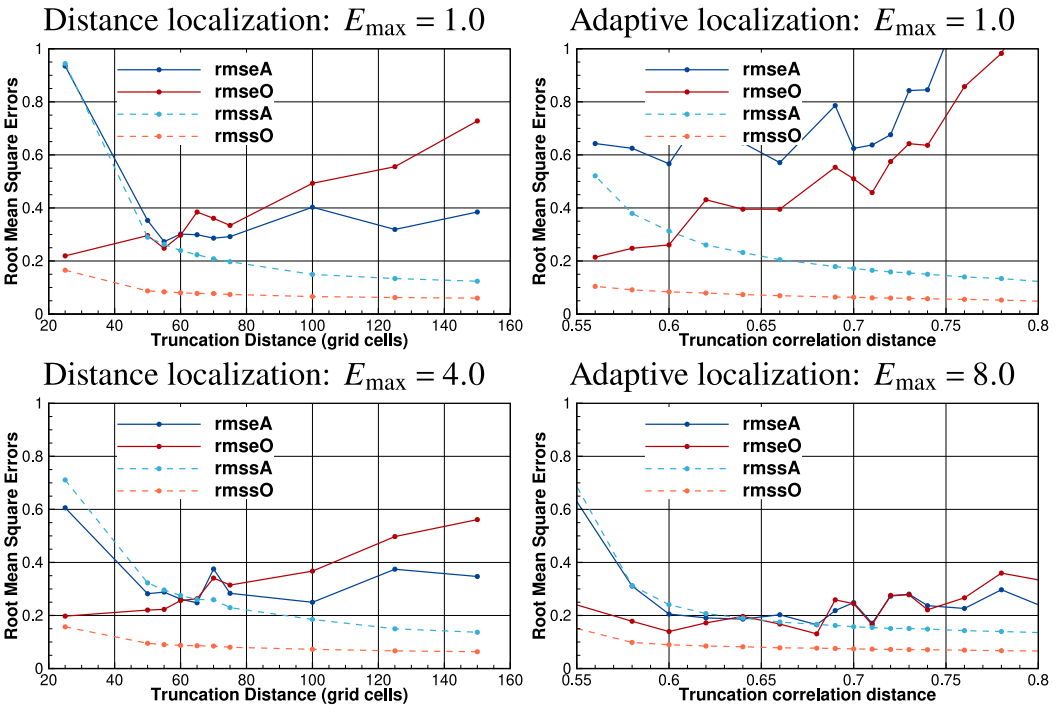

FIG. 11. The figure shows the RMSEs and ensemble standard deviations for different truncation distances for (left) distance-based localization and (right) correlation-based localization. (top) The RMS values when we do not use error variance inflation for remote measurements. (bottom) The results using the "best" values of $E_{max} = 4$ for physical distance-based localization and $E_{max} = 8$ when we use the correlation-based localization. Each point on the curves represents the averages over 50 experiments repeated with different random seeds. The ensemble size is $N = 100$.

that significantly reduces the number of measurements in each local update, reducing the computing time.

In Fig. 12, we show the results from the distance-based localization using a localization distance of 50 grid points and $E_{max} = 4$. In contrast, in Fig. 13, we present the results from the best configuration from Fig. 11 when using the adaptive correlation-based localization. The distance-based localization extends the solution space and allows the assimilation system to track the reference solution without any observable filter divergence efficiently. On the other hand, the adaptive correlation-based localization overall gives a slightly lower residual and better correspondence between the estimated ensemble standard deviations and the expected variance of the residuals. Note also that with correlation-based localization, we do not need to consider differences in truncation distances in the Atmos and Ocean variables, as is the case with distance-based localization. The apparent vertical lines seen in the Atmos standard deviations and residuals in Fig. 12 result from the distance-based localization using a relatively short truncation distance of 50 grid points, which creates small discontinuities in the updated ensemble realizations.

c. *Correlation-distance dependency*

The Hovmöller diagrams of the global update experiments in Figs. 8–10 illustrate the difficulty in obtaining accurate analyses with small ensemble sizes. Figures 12 and 13 give the Hovmöller diagrams for the case of distance-based localization and adaptive correlation-based localization, respectively, both for an ensemble size of $N = 100$. Ensemble sizes larger than 400 for these models give similar results with and without adaptive or distance-based localization. However, for smaller ensembles, both the adaptive correlation-based localization and the distance-based localization reduce the RMSE by a factor of 3 or more. The ensemble spread [root-mean-square standard deviation (RMSS), which is the square root of the temporal average of the squares of the spatial standard deviations] is hardly affected, suggesting no ensemble collapse in the localization cases. A comparison of the RMSS and RSME allows us to evaluate the estimated errors in comparison with the actual errors. However, we notice that the ensemble spread is systematically too low for the ocean variables in all localization experiments, indicating the need for including an additional inflation scheme. Since the Ocean dynamics are dominated by smaller spatial scales than the Atmos dynamics, there is no well-defined distance value for localization when we update the Atmos with Ocean observations, or vice versa. We performed many experiments with different localization distances for Atmos and Ocean, but all had their issues, as expected, given the outcomes of the experiments presented above.

The upper plots of Fig. 14 show the retained observations from the first assimilation time in the first assimilation analysis of ESMDA using distance-based localization and plotted at their respective distance–correlation location. In this example, we set the Euclidian truncation distance for distance-based

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

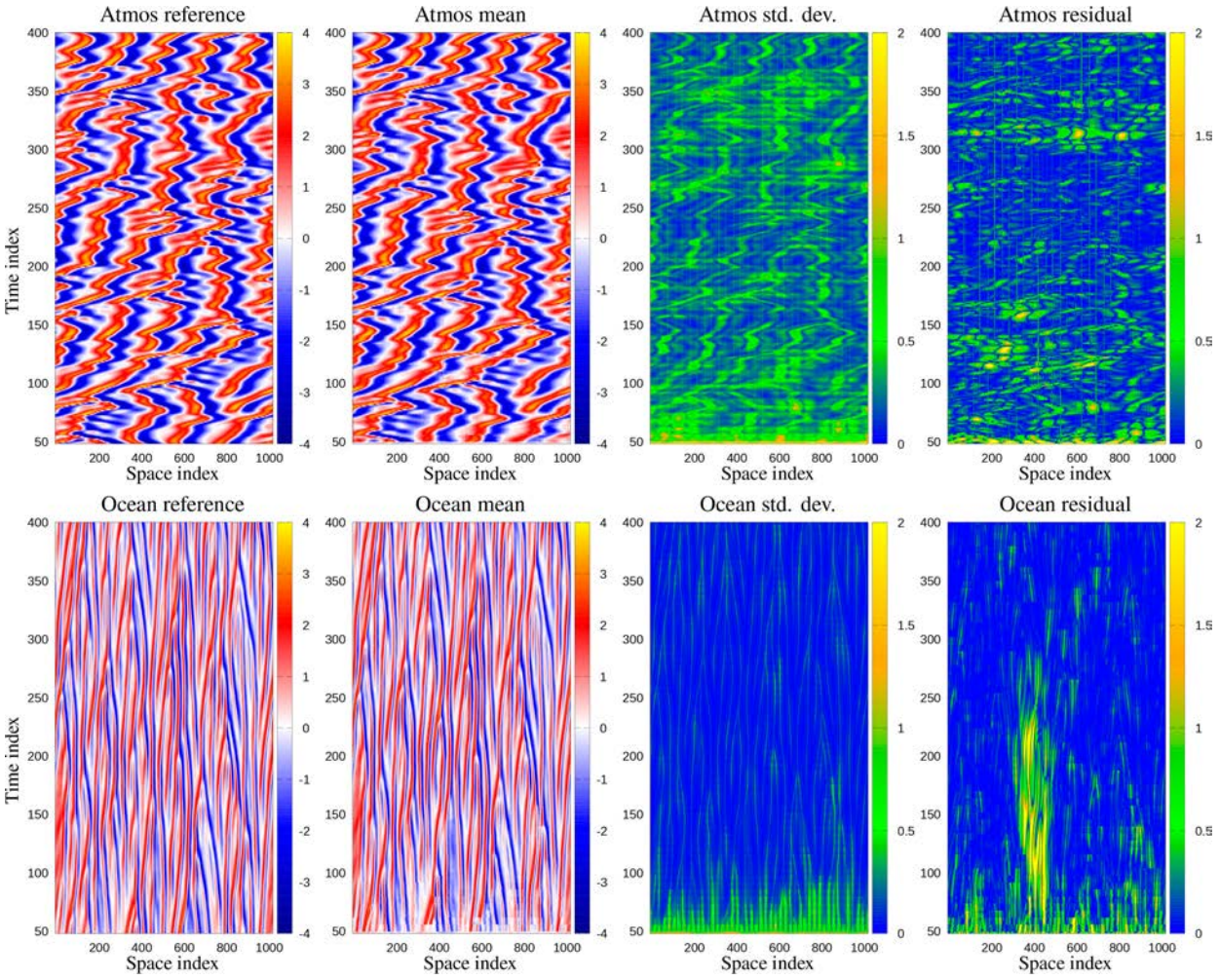

FIG. 12. The Hovmöller plots show the results from an update using distance-based localization, with a localization distance of 50 grid points, $E_{\max} = 4.0$, and $N = 100$ realizations in the ensemble. See the description in Fig. 8.

localization to 150 grid points, hence the absence of points at greater distances. The "innovation," i.e., the difference between the observation and the model equivalents, determines the color of each observation. It is interesting to see that many correlations are very small, even at very small distances. One would expect these correlations to have high uncertainties, leading to less accurate updates. We observe relatively more observation–model pairs that differ more than three units for the Ocean component than the Atmos component, suggesting that we need a more significant update for the Ocean variable. The upper plots of Fig. 14 also illustrate that the Atmos variable has high correlations between observation and model at distances exceeding 75, which we do not see for the Ocean variable. We can interpret this difference in results from the correlation structures from the free runs in Fig. 7 with $\tau = 4$, where Ocean correlation length scales drop off at about 75 grid points.

In the lower plots of Fig. 14, we set the truncation scale in the correlation-based localization to 0.3. Some model–data pairs with a lower correlation are visible in the plots. In cases where the correlation for one component is higher than the threshold and lower for the other component, we still include the model–observation pair. One could argue that only the component with a correlation higher than 0.3 should be updated by that observation, but we did not implement that. We find recurring strong correlations for the Atmos variable every 120 grid points, consistent with the patterns in Fig. 7. Figure 7 also explains the "bow like" structures seen in the upper plots of Fig. 14 via the typical Atmos correlation "blob" size of about 75 grid points. Interestingly, the coupling through the data assimilation allows Ocean to inherit some of the longer correlation distances at about 120 gridpoint distance from Atmos, a desirable feature related to the physical interactions between the two components enforced by the data assimilation.

6. Summary and outlook

In this paper, we discussed different approaches to localization. We demonstrated how adaptive correlation-based localization in the case of a coupled Kuramoto–Sivashinsky model

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

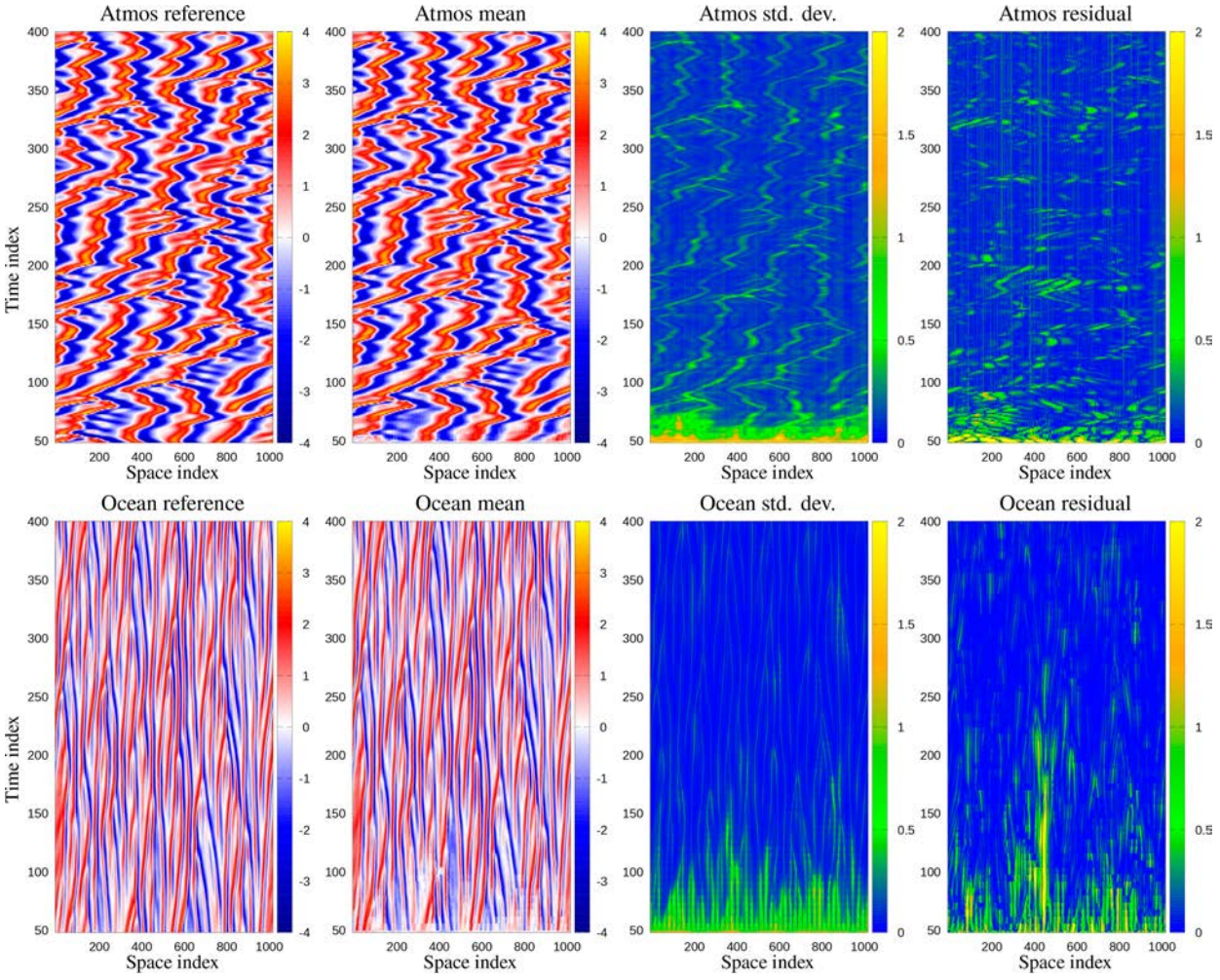

FIG. 13. The Hovmöller plots show the results from an update using correlation-based localization, with a correlation truncation distance of $d_t = 0.68$, $E_{\max} = 8.0$, and $N = 100$ realizations in the ensemble. See the description in Fig. 8.

resulted in lower RMSE values than distance-based localization. Moreover, it offers an approach that considers the different spatial scales of the two coupled systems, and it solves the space–time localization distance issue that has hindered ensemble smoothers for a long time.

A fixed truncation distance in distance-based localization will exclude observations strongly correlated with the analyzed variable and located relatively far from the analysis point. Thus, distance-based localization is less effective in incorporating the information contained in relevant observations. The differences in spatial scales between the coupled systems make the tuning of distance-based localization difficult: While the correlation scales between Ocean observations and the modeled Ocean variable may be most prominent on short spatial scales, the correlation between Ocean observations and Atmos model variables, or between Atmos observations and Ocean model variables, also involves larger spatial scales. So, there is not one unique scale that presents itself as a sensible distance truncation scale, which complicates the tuning of a distance-based localization scheme.

Our approach differs from the standard adaptive covariance and Kalman gain localization presented in previous studies in two distinct ways. First, we compute a local analysis, updating the model solution grid block by grid block using a subset of measurements for each grid block. Second, we introduce a measurement error-variance inflation for measurements with a significant Euclidian or correlation distance approaching the truncation value (see Fig. 4 and its impact in Fig. 11). The measurement error variance inflation leads to smoother updates in space and time and significantly reduces the error residuals, as the results from Fig. 11 demonstrate. We propose that this measurement error variance inflation explains the success of the adaptive correlation-based localization in this paper's experiments.

Our results have focused on smoothing via ESMDA, while many applications use filters such as variants of the EnKF. We found that adaptive correlation-based localization achieves a better overall performance than distance-based localization. Figure 14 contains information on why that might be the case: longer-range correlations, especially in the atmosphere, contain

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

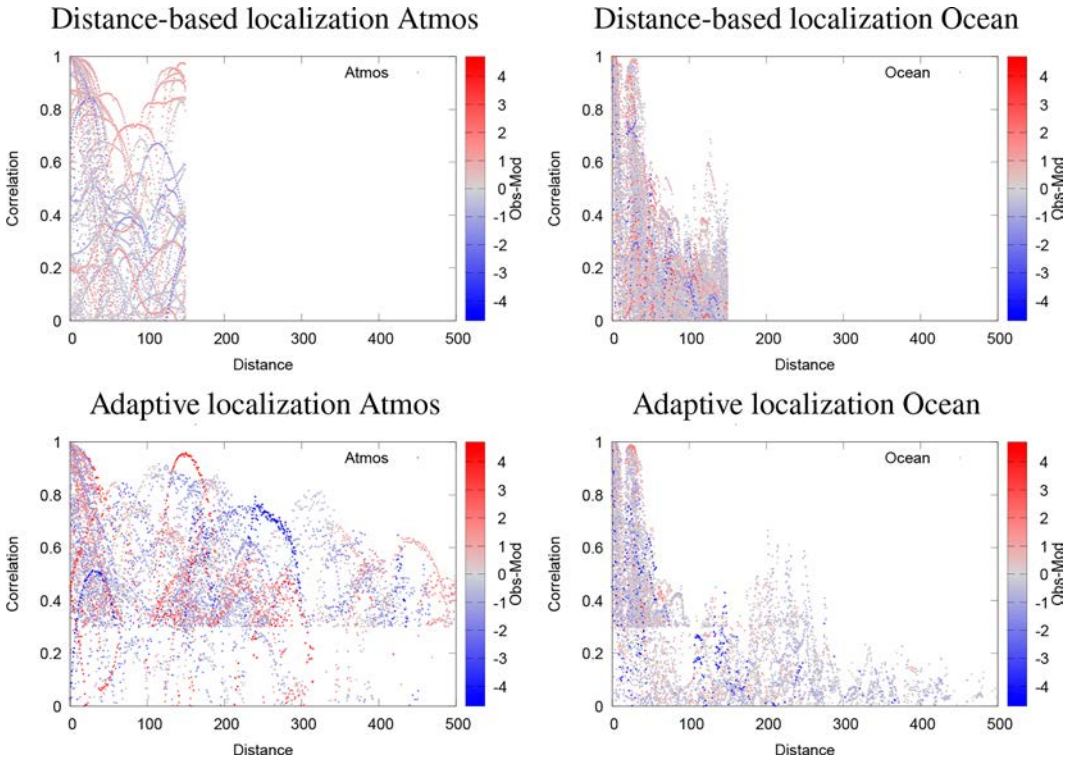

FIG. 14. The figure illustrates the selection of observations in distance-based and adaptive correlation-based localization. (top) All the observations accounted for in the distance-based localization truncated at 150 grid points and plotted at their respective distance–correlation locations. (bottom) All the observations accounted for in the adaptive correlation-based localization truncated at an absolute correlation value of 0.3. The colors indicate the differences between the observations and their model equivalents. (left) The Atmos observations and (right) the Ocean observations.

useful information, and observations with small local correlations omitted by adaptive localization will not lead to significant updates. We expect that this will also hold for any filter implementation of adaptive correlation-based localization.

In a filter, just as in a smoother, we typically assimilate all observations in an assimilation window, which means that even in a filter, we use the space–time correlations of the observations and the model to update the solution. Now in a filter, the update occurs at the end of the assimilation window using observations within the window, whereas an iterative smoother updates the assimilation window's initial conditions. This means that the influence of the localization will be different. In a filter, inaccuracies due to localization will be propagated with the forecast to the next assimilation time. In contrast, in an iterative smoother localization inaccuracies will be propagated through the assimilation window. The iterative nature of the smoother will act to remove some of these inaccuracies. However, the final update will be at the start of the window, and the reduced inaccuracies from the iterations will grow with the final propagation over the window. Hence, it is hard to predict which method will suffer less from localization inaccuracies. This will depend, among others, on the window length. As an aside, if we implement the ensemble smoother (so without the multiple data assimilation) with an update at the end of the window, the method is identical to an EnKF, so the adaptive correlation-based localization in that case will have similar performance as in the case of an EnKF implementation.

The time of the computation of the correlations will influence the result of the adaptive correlation-based localization. In our implementation, the correlations are computed at the end of the assimilation window. If we implemented the adaptive correlation-based localization in a filter, this would be the same.

While the multitude of spatial scales does not plague adaptive correlation-based localization, it also requires tuning. In addition to the inflation factor, the adaptive correlation-based localization has as tuning parameters the correlation threshold and the observation error tapering. We provided a theoretical estimate for the correlation threshold value, which should serve well as a first-guess value in more complex models. The tapering is very similar to that used in the standard distance-based localization. In our experiments, we found that stronger tapering toward larger observation errors is beneficial for adaptive localization compared to distance-based localization. This is not surprising because the physical distances over which the correlations are large can be much larger in the adaptive approach than the distance-based localization cutoff, see e.g., Fig. 14. Meaningful rules of thumb require more

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

experimentation in more realistic systems. Adaptive inflation methods are likely to be most efficient, and a combination of distance-based and adaptive localization may be an attractive and practical approach in realistic applications.

Adaptive correlation-based localization tends to include more observations in the analysis than distance-based localization. While this increases the influence of the observations on the data assimilation analysis, it may result in relatively high computational costs. Also here, an interesting idea to avoid this issue is a combination of distance-based and correlation-based localization, in which we first limit the number of observations by applying a wide spatial truncation, followed by a truncation of the correlation distance. Such an approach not only reduces computational costs but can offer a useful balance between the well-known stability of distance-based localization and the adaptivity of the correlation-based approach.

We performed the experiments in this paper using a highly idealistic representation of two Earth systems bearing limited similarity to the actual ocean and atmosphere. Further research should include the application of more realistic models of coupled systems. It will be interesting to see the performance of the adaptive correlation-based localization when coupling models of different Earth components. We can imagine that for some Earth system components, the coupling will be very nonlocal, very complex, and possibly non-Gaussian. In those cases, a localization based on correlations rather than on distance measures might better capture the relevant correlations than a localization based on physical distance. This would need to be investigated. In reality, the frequency and location of observations will not be as favorable as in the experiments presented here, and follow-up studies should consider this fact as well. An exciting application for adaptive correlation-based localization would be the assimilation of "nonlocal" satellite radiances sensitive to deep atmospheric layers as discussed by Campbell et al. (2010).

In conclusion, this study illustrates the benefits, sensitivities, and promise that adaptive correlation-based localization methods hold, and further research should consider these methods in more realistic systems.

Acknowledgments. Geir Evensen was supported by the Research Council of Norway through the REMEDY project, Grant 336240. Peter Jan van Leeuwen was supported by the National Science Foundation, Grant 2220201. The authors thank two anonymous reviewers and the editor, Jeffrey Anderson, for their valuable suggestions which improved the manuscript.

Data Availability statement. The code and input data with instructions to compile and run all experiments and plot the results presented in this paper are available from the GitHub repository (Evensen 2023).

REFERENCES

Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

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

——, 2016: Reducing correlation sampling error in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **144**, 913–925, https://doi.org/10.1175/MWR-D-15-0052.1.

Bickel, P. J., and E. Levina, 2008: Covariance localization by thresholding. *Ann. Stat.*, **36**, 2577–2604, https://doi.org/10.1214/08-AOS600.

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

——, and ——, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus*, **61A**, 84–96, https://doi.org/10.1111/j.1600-0870.2008.00371.x.

——, and ——, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus*, **61A**, 97–111, https://doi.org/10.1111/j.1600-0870.2008.00372.x.

——, B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Brusdal, K., J. M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P. J. Van Leeuwen, E. Dombrowsky, and J. Verron, 2003: A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.*, **40–41**, 253–289, https://doi.org/10.1016/S0924-7963(03)00021-6.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.*, **138**, 282–290, https://doi.org/10.1175/2009MWR3017.1.

Chen, Y., and D. S. Oliver, 2017: Localization and regularization for iterative ensemble smoothers. *Comput. Geosci.*, **21**, 13–30, https://doi.org/10.1007/s10596-016-9599-7.

——, ——, and D. Zhang, 2009: Efficient ensemble-based closed-loop production optimization. *SPE J.*, **14**, 634–645, https://doi.org/10.2118/112873-PA.

De Cruz, L., J. Demaeyer, and S. Vannitsem, 2016: The Modular Arbitrary-Order Ocean-Atmosphere Model: MAOOAM v1.0. *Geosci. Model Dev.*, **9**, 2793–2808, https://doi.org/10.5194/gmd-9-2793-2016.

Emerick, A., and A. Reynolds, 2011: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Comput. Geosci.*, **15**, 251–269, https://doi.org/10.1007/s10596-010-9198-y.

Evensen, G., 2003: The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

——, 2009: *Data Assimilation: The Ensemble Kalman Filter.* 2nd ed. Springer, 280 pp., https://doi.org/10.1007/978-3-642-03711-5.

——, 2023: EnKF_MS. GitHub, accessed 15 January 2024, https://github.com/geirev/EnKF_MS.

——, F. C. Vossepoel, and P. J. Van Leeuwen, 2022: *Data Assimilation Fundamentals: A Unified formulation of the State and Parameter Estimation Problem.* Springer, 245 pp., https://doi.org/10.1007/978-3-030-96709-3.

——, ——, and ——, 2024: Iterative ensemble smoothers for data assimilation in coupled nonlinear multiscale models. *Mon. Wea. Rev.*, **152**, 1277–1301, https://doi.org/10.1175/MWR-D-23-0239.1.

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC

——, D. S. Oliver, and R. G. Hanea, 2025: *Ensemble History Matching: Conditioning Reservoir Models on Dynamic Data.* Springer, 254 pp.

Flowerdew, J., 2015: Towards a theory of optimal localisation. *Tellus*, **67A**, 25257, https://doi.org/10.3402/tellusa.v67.25257.

Furrer, R., and T. Bengtsson, 2007: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivar. Anal.*, **98**, 227–255, https://doi.org/10.1016/j.jmva.2006.08.003.

Gaspari, G., and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

Guttman, I., S. Wilks, and J. Hunter, 1982: *Introductory Engineering Statistics.* John Wiley and Sons, 580 pp.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Haugen, V. E., O. M. Johannessen, and G. Evensen, 2002: Indian Ocean: Validation of the Miami Isopycnic Coordinate Ocean Model and ENSO events during 1958–1998. *J. Geophys. Res.*, **107**, 3043, https://doi.org/10.1029/2000JC000330.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Kuramoto, Y., 1978: Diffusion-induced chaos in reaction systems. *Prog. Theor. Phys. Suppl.*, **64**, 346–367, https://doi.org/10.1143/PTPS.64.346.

Luo, X., and T. Bhakta, 2020: Automatic and adaptive localization for ensemble-based history matching. *J. Pet. Sci. Eng*, **184**, 106559, https://doi.org/10.1016/j.petrol.2019.106559.

——, and C.-A. Xia, 2022: Continuous Hyper-parameter OPtimization (CHOP) in an ensemble Kalman filter. *Front. Appl. Math. Stat.*, **8**, 1021551, https://doi.org/10.3389/fams.2022.1021551.

Miwa, N., and Y. Sawada, 2024: Strongly versus weakly coupled data assimilation in coupled systems with various inter-compartment interactions. *J. Adv. Model. Earth Syst.*, **16**, e2022MS003113, https://doi.org/10.1029/2022MS003113.

Neto, G. M. S., R. V. Soares, G. Evensen, A. Davolioa, and D. J. Schiozer, 2021: Subspace ensemble randomized maximum likelihood with local analysis for time-lapse-seismic-data assimilation. *SPE J.*, **26**, 1011–1031, https://doi.org/10.2118/205029-PA.

Penny, S. G., and Coauthors, 2017: Coupled data assimilation for integrated Earth system analysis and prediction: Goals, challenges, and recommendations. Special Rep. WWRP 2017-3, WMO, 45 pp., https://repository.library.noaa.gov/view/noaa/28431.

Sivashinsky, G. I., 1977: Nonlinear analysis of hydrodynamic instability in laminar flames—I. Derivation of basic equations. *Acta Astronaut.*, **4**, 1177–1206, https://doi.org/10.1016/0094-5765(77)90096-0.

——, 1980: On flame propagation under conditions of stoichiometry. *SIAM J. Appl. Math.*, **39**, 67–82, https://doi.org/10.1137/0139007.

Tondeur, M., A. Carrassi, S. Vannitsem, and M. Bocquet, 2020: On temporal scale separation in coupled data assimilation with the ensemble Kalman filter. *J. Stat. Phys.*, **179**, 1161–1185, https://doi.org/10.1007/s10955-020-02525-z.

Vishny, D., M. Morzfeld, K. Gwirtz, E. Bach, O. R. A. Dunbar, and D. Hodyss, 2024: High-dimensional covariance estimation from a small number of samples. *J. Adv. Model. Earth Syst.*, **16**, e2024MS004417, https://doi.org/10.1029/2024MS004417.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Zhang, Y., and D. S. Oliver, 2010: Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Math. Geosci.*, **42**, 327–345, https://doi.org/10.1007/s11004-010-9267-8.

Brought to you by TU DELFT | Unauthenticated | Downloaded 11/19/25 10:29 AM UTC