Iterative ensemble smoothers for data assimilation in coupled nonlinear multiscale models

Geir Evensen,[a,b] Femke C. Vossepoel,[c] Peter Jan van Leeuwen,[d]

[a] *Norwegian Research Center (NORCE), Bergen, Norway*
[b] *Nansen Environmental and Remote Sensing Center (NERSC), Bergen, Norway*
[c] *Department of Geoscience and Engeneering, Delft University of Technology, Delft, The Netherlands*
[d] *Department of Atmospheric Sciences, Colorado State University, Fort Collins, CO, USA*

ABSTRACT: This paper identifies and explains particular differences and properties of adjoint-free iterative ensemble methods initially developed for parameter estimation in petroleum models. Furthermore, we demonstrate the methods' potential for sequential data assimilation in coupled and multiscale unstable dynamical systems. To examine the data assimilation methods, we introduce a new nonlinear and coupled multiscale model based on two Kuramoto-Sivashinsky equations operating on different scales where a coupling term relaxes the two model variables towards each other. This model provides a convenient testbed for studying data assimilation in highly nonlinear and coupled multiscale systems. Results and discussions will provide an enhanced understanding of the ensemble methods' potential implementation and use in operational weather and climate-prediction systems. We show that the model coupling leads to cross-covariance between the two models' variables, allowing for a combined update of both models. The measurements of one model's variable will also influence the other and contribute to a more consistent estimate. Secondly, the new model allows us to examine the properties of iterative ensemble smoothers and assimilation updates over finite-length assimilation windows. We discuss the impact of varying the assimilation-window lengths relative to the model's predictability time scale. Furthermore, we show that iterative ensemble smoothers significantly improve the solution's accuracy compared to the standard ensemble-Kalman-filter update.

1. Introduction

Numerical weather prediction at national and international weather centers uses various data-assimilation approaches to initialize ocean and atmosphere models. de Rosnay et al. (2022) and Laloyaux et al. (2016) describe the current state-of-the-art coupled data-assimilation system at ECMWF, which combines the assimilation of ocean observations using a 3DVar scheme with the assimilation of atmospheric observations using a 4DVar method. The ECMWF data-assimilation system assumes a weak coupling between the two subsystems in which increments produced by independent data assimilation updates of each component are subsequently applied simultaneously to the initial model state of each component. After that, they perform a coupled model integration to provide the first guess for the next iteration in the cost-function minimization. GFDL uses a two-step ensemble-based filtering algorithm applied to a fully coupled climate model Chang et al. (2013). NCEP assimilates data into partially coupled earth-system components using a 3DVar-based scheme to assimilate land-surface, atmosphere, ocean, and sea-ice observations (Saha et al. 2010).

Penny et al. (2017) provide a comprehensive overview of coupled ocean-atmosphere models. While most operational approaches have a weak coupling between the two subsystems and so-called outer-loop coupling, more conceptual studies investigate the effect of a strong coupling. For example, Luo and Hoteit (2014) use a multiscale Lorenz-96 model to evaluate the performance of a state-space estimation strategy that assimilates data into each subsystem and correlated quantities from other coupled subsystems. In a slightly more realistic setting, Han et al. (2013) couple a Lorenz-63 model to a pycnocline ocean model.

Penny et al. (2019) compares several data-assimilation algorithms on a single quasi-geostrophic model. Tondeur et al. (2020) used the same model to study coupled data assimilation with the ensemble Kalman filter (EnKF) for a system consisting of a slow ocean coupled to a fast atmosphere. This study illustrates the mechanisms behind information propagation between the system's two components and concludes that cross-component effects are strong from the slow to the fast scale.

This study will use a simplified representation of two coupled earth-system components to explore the information propagation between a component with predominantly large spatial scales (the atmosphere) and a component with more minor spatial scales. In addition, we evaluate the efficiency of different data-assimilation algorithms in coupled data assimilation.

The coupled model uses the Kuramoto-Sivashinsky (KS) equations to describe interactions between two systems: one with a longer spatial scale (typically the atmosphere) and one with a shorter spatial scale (typically the ocean). We describe the model in the following section. After that, we introduce and discuss the adjoint-free data-assimilation methods before we run multiple experiments to examine the impact and value of coupled data assimila-

Corresponding author: Geir Evensen, geev@norceresearch.no

tion and various sensitivities to window lengths, number of iterations, etc.

2. Coupled multiscale Kuramoto-Sivashinsky model

We will now describe the nonlinear multiscale model used to study our data-assimilation methods. We needed a model with specific properties, including unstable and near-chaotic dynamics with nonlinear saturation of the linear instabilities, similar to oceanic and atmospheric behavior. For computational reasons and ease of interpretation, we searched for a one-dimensional and univariate equation for each component of the coupled system. Finally, we need a model that can inhibit different spatial and temporal scales, as we wish to study the assimilation of observations in coupled models with different scales. The model system should resemble the behavior of coupled climate models, where the ocean and atmospheric components have vastly different spatial and temporal scales. Our choice landed on a variant of the Kuramoto-Sivashinsky (KS) equation, from which we derived two model configurations operating on different spatial scales and with an interaction or relaxation term connecting the two models. In the current study, we only include different spatial scales between the model components, but it is possible to alter the temporal scale of the models in further studies.

The Kuramoto-Sivashinsky (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). The model can also describe the dynamics of fluid films on inclines (Shlang and Sivashinsky 1982; Tilley et al. 1994), flow in pipes (Chang 1986), and dynamics of chemical reactions. Kuramoto (1978) described how the coupling of an oscillation and a spatial inhomogeneity could produce spatio-temporal chaos and how one can obtain a balance between phase instability and amplitude instability. The KS equation is

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -\frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4}. \tag{1}$$

The equation reduces to the Burgers equation if we ignore the diffusion terms on the right-hand side, where the nonlinear term, uu_x, transfers energy between large and small scales and creates shocks. The harmonic diffusion term has a negative sign and acts to enhance any feature in the model solution. In contrast, the biharmonic diffusion term with a negative sign acts as a small-scale selective positive diffusion, thereby controlling any growing instabilities in the model.

Protas et al. (2004) introduced the KS equations for an evaluation of 4DVar data-assimilation methods. Jardak et al. (2010) and Chorin and Krause (2004) evaluated the performance of Bayesian filters using the KS-equations, while Azouani and Titi (2014); Lunasin and Titi (2017) used the equations as a testbed for the so-called continuous data-assimilation technique, and **?** used it for correlated observation error estimation.

We refer to https://online.kitp.ucsb.edu/online/transturb17/gibson/html/5-kuramoto-sivashinksy.html for a discussion of the numerical implementation and example codes. In short, we have used a Crank-Nicolson–Adams-Bashforth (CNAB) scheme for time stepping the model. The time-stepping method is second-order and implicit in time. We discretized space using a Fourier decomposition as this approach renders the inverse matrices in the scheme diagonal, allowing for a highly efficient implementation. Our Fortran-90 subroutine for the model integration is available from https://github.com/geirev/EnKF_MS/blob/main/src/m_model.F90. We have verified that our code exactly reproduces the solution from https://online.kitp.ucsb.edu/online/transturb17/gibson/html/5-kuramoto-sivashinksy.html.

In our implementation, we have used two KS models, and we refer to the two model solutions as *Atmos* and *Ocean* with the symbols A and O referring to their respective variables. The coupled model equations read

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

$$\frac{\partial O}{\partial t} = -\frac{1}{2}\frac{\partial O^2}{\partial x} - \frac{\partial^2 O}{\partial x^2} - \frac{\partial^4 O}{\partial x^4} + \omega_{ao}(A - O). \tag{3}$$

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

To introduce different spatial scales in the two models, we defined two different pseudo lengths of the model domains for Ocean and Atmos. We used a periodic model domain with 1024 grid points but assumed a physical size of 32 for the Atmos domain and 256 for the Ocean domain. A similar approach would introduce different time scales in the two model components.

Our multiscale KS model is suitable for conceptualizing data assimilation in coupled systems. We can consider it a 1D analog to the 2D Navier-Stokes equations. The model is well-posed; 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.

In nonlinear chaotic models, such as weather prediction models, we can define the model's *predictability time* as how long we can integrate the model before the model solution's uncertainty reaches the climatological error level. Typically, nonlinear unstable models initially experience exponential error growth before the errors saturate at the climatological variability due to nonlinear effects.

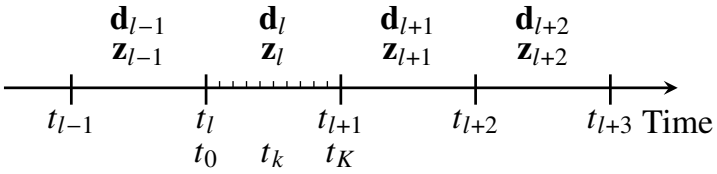

FIG. 1. The figure illustrates how we split time into discrete time windows. Within each time window, l, we integrate a model K time steps and assimilate the data $\mathbf{d}_l$.

3. Data-assimilation methods

We have adopted data-assimilation methods, notation, and formulations from our recent open-access textbook Evensen et al. (2022), where We split the time dimension into a sequence of time windows as illustrated in Fig. 1. The definition of finite-length assimilation windows facilitates easy assimilation of all data available within the time window in one single update. It makes it possible to use iterative ensemble smoothers to reduce the impact of model nonlinearity and to obtain superior results compared to standard EnKF solutions. The three methods considered in this paper are versions of the ensemble smoother (ES) that trace back to the ES of Van Leeuwen and Evensen (1996), the ensemble smoother with multiple data assimilation (ESMDA) proposed by Emerick and Reynolds (2012), and the ensemble randomized maximum likelihood method (EnRML) by Chen and Oliver (2012, 2013). We are using the ensemble subspace implementation of Evensen et al. (2019) for the EnRML, which we in the following refer to as the iterative ensemble smoother (IES). The term "randomized maximum likelihood" does not make sense as we are solving for randomized maximum a posteriori solutions. Evensen et al. (2022) discussed and explained the details of the algorithms. We also note that the first step in the IES algorithm becomes identical to the ES update if we choose a steplength of one. Thus, we can use the numerical IES implementation to compute the ES, ESMDA, and IES updates.

a. A theoretical basis for ensemble methods

The ensemble data-assimilation methods attempt to sample the posterior conditional pdf defined by Bayes' theorem

Bayes' theorem

$$f(\mathbf{z}|\mathbf{d}) = \frac{f\big(\mathbf{d}|\mathbf{g}(\mathbf{z})\big)\, f(\mathbf{z})}{f(\mathbf{d})}. \tag{4}$$

where we have introduced the composite model and measurement operator $\mathbf{g}$ in

$$\mathbf{y} = \mathbf{g}(\mathbf{z}), \tag{5}$$

which maps the state vector $\mathbf{z}$ to the predicted measurements $\mathbf{y}$.

Kitanidis (1995); Oliver et al. (1996) showed that for a Gaussian prior and likelihood, minimization of an ensemble of cost functions

Ensemble of cost functions

$$\begin{aligned}\mathcal{J}(\mathbf{z}_j) = &\frac{1}{2}\big(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}\big)^{\mathrm{T}}\mathbf{C}_{zz}^{-1}\big(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}\big) \\ &+ \frac{1}{2}\big(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\big)^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\big(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\big),\end{aligned} \tag{6}$$

results in an *approximate* sampling of the posterior Bayesian pdf. This *randomized maximum likelihood* (RML) sampling is exact in the case of a Gauss-linear model and measurement operator, and the significance of the approximation will depend on the level of nonlinearity of the model in Eq. (5). The cost functions are mutually independent for each realization j and use the random samples $\mathbf{z}_j^{\mathrm{f}} \sim \mathcal{N}(\mathbf{z}^{\mathrm{f}}, \mathbf{C}_{zz})$ for the prior $\mathbf{z}^{\mathrm{f}}$ and $\mathbf{d}_j \sim \mathcal{N}(\mathbf{d}, \mathbf{C}_{dd})$ for the perturbed measurements to represent the uncertainties. The covariances $\mathbf{C}_{zz}$ and $\mathbf{C}_{dd}$ are the error covariances for the prior state vector and the measurements.

Note that to solve the assimilation problems for a particular assimilation window, we must assume that measurements are independent between the different windows and that the model is a first-order Markov process. Additionally, we apply a filtering assumption by only updating the solution in the current assimilation window and ignoring any updates of the past windows. We can easily relax the the filtering assumption by using an ensemble Kalman smoother (EnKS) approach to update the solution in previous windows (Evensen and Van Leeuwen 2000).

We derive the ensemble methods by setting the gradient of the cost function in Eq. (6) to zero,

Ensemble of gradients set to zero

$$\mathbf{C}_{zz}^{-1}\big(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}\big) + \nabla_{\mathbf{z}}\mathbf{g}\big(\mathbf{z}_j\big)\mathbf{C}_{dd}^{-1}\big(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\big) = 0. \tag{7}$$

By introducing a linearization of $\mathbf{g}(\mathbf{z})$ around the prior estimate $\mathbf{z}_j^f$ we obtain an ensemble of Kalman filter update

equations.

$$\mathbf{z}_j^{\mathrm{a}} = \mathbf{z}_j^{\mathrm{f}} + \mathbf{C}_{zz}\mathbf{G}_j^{\mathrm{T}}\left(\mathbf{G}_j\mathbf{C}_{zz}\mathbf{G}_j^{\mathrm{T}} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{g}\left(\mathbf{z}_j^{\mathrm{f}}\right)\right). \quad (8)$$

However, as we noted in the previous chapter, these equations are only valid in the linear case or for modest updates in the nonlinear case. Alternatively, we can use the gradient and the Hessian

$$\nabla_{\mathbf{z}}\nabla_{\mathbf{z}}\mathcal{J}(\mathbf{z}_j) \approx \mathbf{C}_{zz}^{-1} + \nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}_j)\mathbf{C}_{dd}^{-1}\left(\nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}_j)\right)^{\mathrm{T}}. \quad (9)$$

in a Gauss-Newton iteration to minimize the ensemble of cost functions exactly. This latter approach corresponds to the ensemble 4DVar method where we use variational methods to minimize the ensemble of independent cost functions.

However, in this paper we will introduce another approximation that allows us to eliminate the models tangent linear operator. We will represent the model sensitivity by an ensemble-averaged sensitivity through the linear regression

$$\nabla_{\mathbf{z}}\mathbf{g}\left(\mathbf{z}_j\right) = \mathbf{C}_{zz}^{-1}\mathbf{C}_{zy}. \quad (10)$$

In Eq. (10), we replace all individual model sensitivities for the different realizations with a common average one. This approximation implies that we are no longer solving strictly for the minima of the cost functions in Eq. (6), but we approximate the estimates of the minima.

The final approximation is to represent all covariances using an ensemble of realizations, leading to the ensemble Kalman filter or smoother update formulated entirely in terms of ensemble matrices (Evensen 2003) as

$$\begin{aligned}\mathbf{Z}^{\mathrm{a}} &= \mathbf{Z}^{\mathrm{f}} + \mathbf{A}\mathbf{Y}^{\mathrm{T}}\left(\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}}\right)^{-1}\left(\mathbf{D} - \mathbf{g}\left(\mathbf{Z}^{\mathrm{f}}\right)\right) \\ &= \mathbf{Z}^{\mathrm{f}} + \mathbf{A}\mathbf{W}.\end{aligned} \quad (11)$$

Here we have defined the ensemble matrices

$$\mathbf{Z} = \left(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\right). \quad (12)$$

Furthermore, we define the projection $\mathbf{\Pi} \in \Re^{N\times N}$ as

$$\mathbf{\Pi} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)/\sqrt{N-1}, \quad (13)$$

where $\mathbf{1} \in \Re^N$ is a vector with all elements equal to one and $\mathbf{I}_N$ is the N-dimensional identity matrix. If we multiply an ensemble matrix with the orthogonal projection $\mathbf{\Pi}$, this subtracts the mean from the ensemble and scales the result with $1/\sqrt{N-1}$.

We can then define the zero-mean and scaled ensemble-anomaly matrix as

$$\mathbf{A} = \mathbf{Z}\mathbf{\Pi}, \quad (14)$$

and the ensemble covariance is

$$\overline{\mathbf{C}}_{zz} = \mathbf{A}\mathbf{A}^{\mathrm{T}}, \quad (15)$$

where the "overbar" denotes that we have an *ensemble*-covariance matrix.

Correspondingly, we can define an ensemble of perturbed measurements, $\mathbf{D} \in \Re^{m\times N}$, when given the real measurement vector, $\mathbf{d} \in \Re^m$, as

$$\mathbf{D} = \mathbf{d}\mathbf{1}^{\mathrm{T}} + \sqrt{N-1}\mathbf{E}, \quad (16)$$

where $\mathbf{E} \in \Re^{m\times N}$ is the centered measurement-perturbation matrix whose columns are sampled from $\mathcal{N}(0, \mathbf{C}_{dd})$ and divided by $\sqrt{N-1}$. We define the ensemble covariance matrix for the measurement perturbations as

$$\overline{\mathbf{C}}_{dd} = \mathbf{E}\mathbf{E}^{\mathrm{T}}. \quad (17)$$

The ensemble algorithms work both with a full-rank $\mathbf{C}_{dd}$ or the ensemble version of the measurement covariance represented by the perturbations in $\mathbf{E}$.

Finally, we define the ensemble of model-predicted measurements

$$\mathbf{\Upsilon} = \mathbf{g}(\mathbf{Z}), \quad (18)$$

with anomalies

$$\mathbf{Y} = \mathbf{\Upsilon}\mathbf{\Pi}, \quad (19)$$

where we have multiplied the model prediction by the projection $\mathbf{\Pi}$ to subtract the ensemble mean and divide the resulting anomalies by $\sqrt{N-1}$. From the second line in Eq. (11), we notice that the update becomes a linear combination of the prior ensemble anomalies.

Similarly to the EnKF update, we can write the IES equations using the ensemble matrices. Still, a better alternative is to use the EnRML's *ensemble subspace* variant developed by Evensen et al. (2019); Raanes et al. (2019). We refer to these papers and the open-access book by Evensen et al. (2022) for an in-depth discussion of this method.

b. Ensemble smoother

When formulated for a single data-assimilation window, the update computed by the ensemble smoother is a simple extension of the Ensemble Kalman Filter (EnKF) update by Evensen (1994); Burgers et al. (1998). In the standard form of EnKF, one updates the model solution instantly when measurements are available. *The ES provides a framework where we can update the solution at any time step or all time steps in an assimilation window, using simultaneously all measurements available within the window.* When measurements are available at the end of the window, the ES solution at the end of this window is identical to the EnKF solution. Furthermore, ES provides a means for using the EnKF formalism with measurements distributed over the assimilation window and updates computed at the end of

the window. Thus, we can view ES as an extension of the EnKF that adds an update in the space-time domain with measurements distributed in space and time over this domain.

c. Iterative ensemble smoother

While ES computes one linear step in the gradient direction to approximate the cost functions' minima, the IES uses Gauss-Newton iterations to search for the ensemble of cost functions' minima, defining the posterior ensemble. If the cost functions do not contain local minima, the iterative smoothers should converge to the global minimum of each cost-function realization. But note that we are using the ensemble-averaged model sensitivity from Eq. (10) that slightly changes the gradient Eq. (7) for each realization. Additionally, we use ensemble covariances, which constrain the posterior ensemble to the ensemble subspace spanned out by the prior ensemble of realizations.

In its standard formulation, IES attempts to estimate the initial conditions of the assimilation window. After updating the initial conditions, we must reintegrate the ensemble of model realizations over the window to compute the updated gradient in each iteration. Realizing, however, that the final converged transition matrix defines the posterior ensemble solution as a linear combination of the prior ensemble and that we do not need the gradient after the final update, we could use this final transition matrix to update the ensemble directly over the whole assimilation window. These two strategies would give identical results in the linear case without model errors.

For the iterative ensemble-smoother solutions, we will use the ensemble subspace implementation by Evensen et al. (2019) of the EnRML method by Chen and Oliver (2012).

A final note is that if we were to use the IES without introducing the ensemble-averaged model sensitivity and ensemble covariances, we would essentially have an ensemble-4DVar (En4DVar) method, which solves an ensemble of independent 4DVar problems using the adjoints to evaluate the model sensitivity. By introducing ensemble covariances in the En4DVar, we would obtain an iterative smoother with consistent error statistics that evolve and become updated at each analysis time. The resulting smoother would be more accurate than the IES as it does not replace the tangent-linear and adjoint operators by the approximate model sensitivity.

d. Ensemble smoother with multiple data assimilation

The ensemble smoother with multiple data assimilation (ESMDA) is an interesting alternative to IES, and it also attempts to approximately sample the posterior from Bayes. When requiring that finite number μ of coefficients α_i satisfy the condition

$$\sum_{i=1}^{\mu} \frac{1}{\alpha^i} = 1, \tag{20}$$

we can write Bayes' as

$$\begin{aligned} f(\mathbf{z}|\mathbf{d}) &\propto f\big(\mathbf{d}|\mathbf{g}(\mathbf{z})\big)\, f(\mathbf{z}) \\ &= f\big(\mathbf{d}|\mathbf{g}(\mathbf{z})\big)^{\left(\sum_{i=1}^{\mu} \frac{1}{\alpha^i}\right)} f(\mathbf{z}) \end{aligned} \tag{21}$$

The tapering of the likelihood allows for a gradual introduction of the measurements over a predefined number of MDA steps. For example, we could use two steps and set $\alpha_1 = \alpha_2 = 2$, which satisfies the condition in Eq. (20). In this case we would have to solve recursively the two updates

$$f(\mathbf{z}|\mathbf{d}) \propto f\big(\mathbf{d}|\mathbf{g}(\mathbf{z})\big)^{\frac{1}{2}} f(\mathbf{z}), \tag{22}$$

$$f(\mathbf{z}|\mathbf{d}) \propto f\big(\mathbf{d}|\mathbf{g}(\mathbf{z})\big)^{\frac{1}{2}} f(\mathbf{z}|\mathbf{d}). \tag{23}$$

The impact of raising a Gaussian likelihood to a power $1/\alpha$ is an inflation of the measurement error variance where the error covariance matrix becomes multiplied by α. Consequently, we can solve each MDA update using ES and the Eq. (11) but using the inflated measurement error covariance $\alpha \mathbf{C}_{dd} \approx \sqrt{\alpha}\mathbf{E}\sqrt{\alpha}\mathbf{E}^{\mathrm{T}}$. A word of caution is that it is necessary to resample measurement perturbations $\sqrt{\alpha}\mathbf{E}$ in each step to avoid introducing a bias from using dependent samples. During the stepwise updating using ESMDA, we must update the initial state of the assimilation window and then reintegrate the model ensemble over the window to obtain the "prior" ensemble of realizations for the next ESMDA step.

The advantage of ESMDA over ES is that while ES computes one large linear update, ESMDA computes a sequence of small linear updates, reducing the impact of the linearization in the ES scheme. We can view ESMDA as an Euler pseudo-time-stepping in state space with a short stepsize while ES computes the update over one large timestep equal to one. ESMDA has become one of the most popular data assimilation methods in petroleum applications, and several companies use it extensively and operationally for parameter estimation in large petroleum reservoir models. Due to the duality of the parameter- and state-estimation problems when updating the initial conditions of an assimilation window, ESMDA is equally applicable for sequential data assimilation. For linear dynamics and measurements, ESMDA and ES will result in the same solution with increasing ensemble size, independent of the number of ESMDA steps.

When used in sequential data assimilation, ESMDA updates the initial conditions of the assimilation window through a finite number of ES steps. In each ESMDA

step, we must rerun the model ensemble over the assimilation window to obtain the prior for the next update step. However, as for IES, we can choose two strategies for the final step in ESMDA. We can update the initial conditions of the assimilation window and rerun the model ensemble to obtain the solution over the window or update the ensemble directly over the whole assimilation window using the ES algorithm without rerunning the ensemble. Both these approaches are valid and consistent as the ESMDA steps are independent ES steps, and as for ES, we can freely choose the update strategy. A critical remark is that the prior ensemble for the last ESMDA step is the "nearly converged" ensemble from the previous update step. In the experiments below, we will show that, as for ES, it is an advantage to compute directly the final ES update over the whole assimilation window, particularly when the assimilation window becomes long compared to the predictability time of the model.

There is, however, a significant difference from IES if we decide to update the solution over the whole window in the final iteration. While in ESMDA, the prior for the last step is the posterior from the previous step, in IES, the posterior solution is a linear combination of the prior ensemble. Thus, the starting point for computing a global update over the window is different and typically less accurate in the IES than in the ESMDA. We will evaluate these two approaches in the experiments below.

There are other fundamental differences between IES and ESMDA. While IES is an iterative method, ESMDA computes a predefined number of update steps. IES minimizes the ensemble of initially defined RML cost functions. At the same time, ESMDA solves in each step for the minima of a new resampled ensemble of cost functions using a linear ES update. However, although the final ensemble solutions will be different using IES and ESMDA, both methods consistently attempt to sample the posterior pdf, and in the linear case with increasing ensemble size, they converge to the same pdf.

4. Importance of coupled data assimilation

We will now demonstrate the importance of computing fully coupled data-assimilation updates. By fully coupled, we understand that measurements of one model component will update all model components through their cross-covariances estimated from the ensemble of coupled model realizations. We start in the following section by examining the impact of the coupling term on the KS model and the cross-covariance functions. After that, in Sec. b, we will compare coupled and uncoupled data-assimilation experiments.

a. Ensemble predictions and covariances

We now present two ensemble prediction experiments to illustrate some properties of the coupled Kuramoto-Sivashinksy models with our choice of model parameters. Fig. 2 shows the results from experiment PRED0, where we run an ensemble of 1000 realizations of the two models in Eqs. (2) and (3) in an uncoupled mode with $\alpha oa = \omega_{ao} = 0$. Hence, the two models evolve independently of each other.

In Fig. 3, we present the coupled reference simulations where $\alpha oa = \omega_{ao} = 0.003$. In both the PRED0 and PRED1 experiments, we notice the apparent chaotic behavior of the two model variables and the differences in spatial scales between the Ocean and Atmos variables, as shown in the left panels of the Figs. 2 and 3. While the two models evolve independently in PRED0, we observe a significant impact of the coupling in PRED1, where the fine-scale Ocean features follow the large-scale structures in the Atmos solution, which also changes compared to the uncoupled simulation. The ensemble mean converges quickly towards zero for both compartments in both experiments, as shown in the center panels. Very importantly, the standard deviations shown in the right panels remain at the climatological level of around 1.25 and 1.75 for the two model components but appear to saturate at a slightly lower level in the coupled simulations than in the uncoupled one, suggesting that the coupling slightly stabilizes the model. The slightly higher standard deviation in the Atmos variable is likely a result of the lower biharmonic diffusion in this model and the difference in scales where the biharmonic diffusion has less effect on the Atmos variable's larger scales.

Fig. 4 shows the time evolution of the root-mean-square residuals between the ensemble mean and the reference solution as the solid lines. The soft lines are the root-mean-square of the ensemble standard deviations averaged over the spatial coordinate. The error growth saturates after an integration time of about 10-20 units of time, indicating the model's predictability time.

For coupled data assimilation, we are interested in the ensemble correlations between an observation of one model component's variable with itself and the variable of the other model. In Figs. 5 and 6, we show the space-time correlation functions between an observation of either the Ocean or the Atmos variables located in the center of the domain and the respective model components. The correlation function of an observation with the measured variable reflects the spatial scales of the variable. However, we note that there is also a correlation in time, which is essential when computing smoother solutions over an assimilation window. We also notice that the correlations from the uncoupled and coupled models differ slightly due to the change in dynamics introduced by the coupling term.

In the uncoupled case, the correlations between different variables are zero (within the sampling errors). Hence, for an uncoupled model system, an Ocean observation will only influence the Ocean variable, and similarly for an Atmos observation. However, for the coupled model, we obtain well-structured and significant cross-correlations

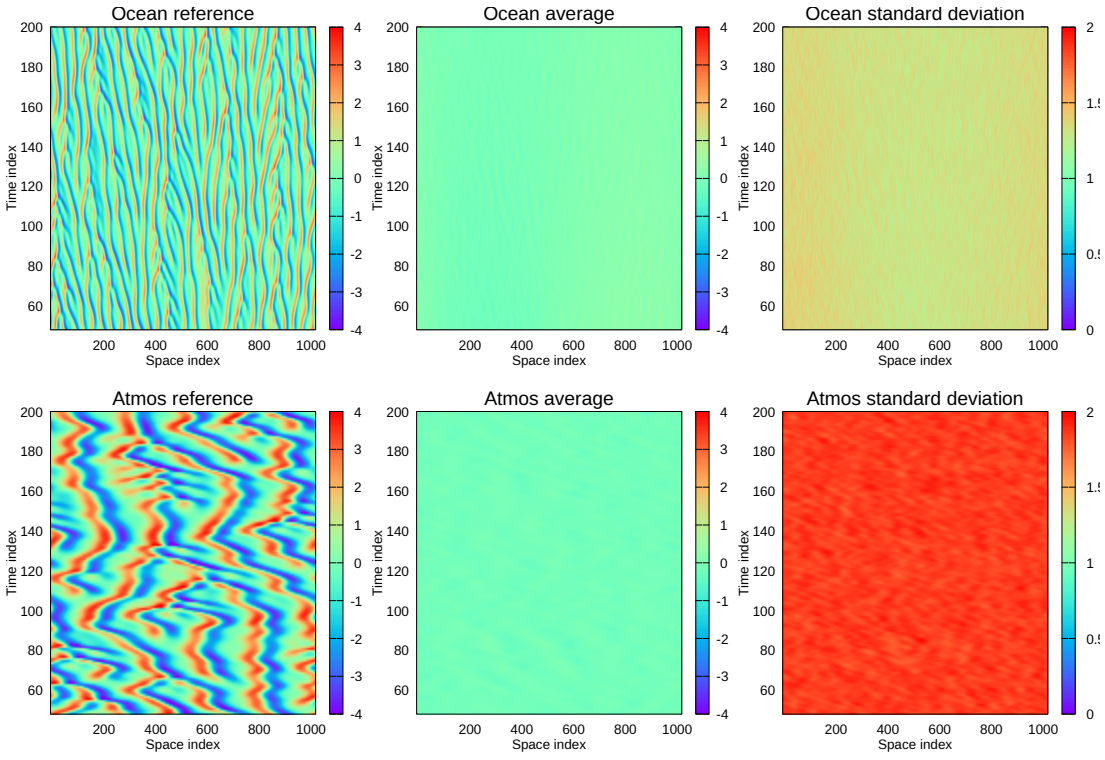

FIG. 2. Exp. PRED0: Uncoupled ensemble-prediction experiment without DA.

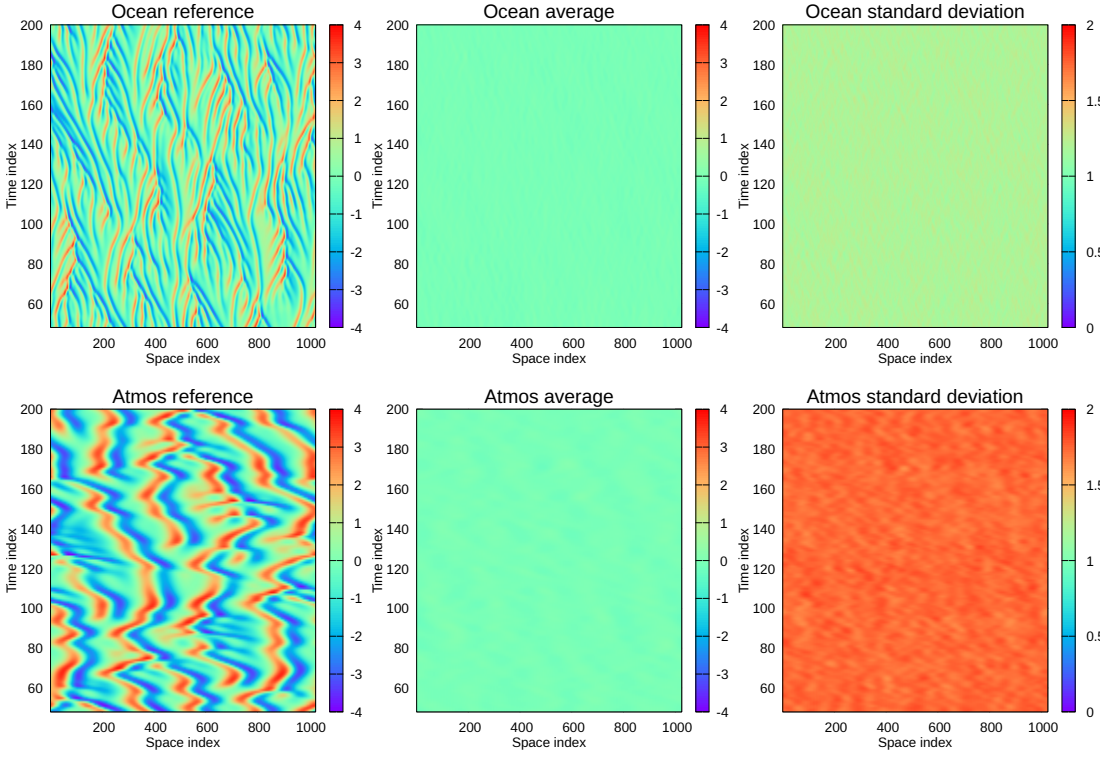

FIG. 3. Exp. PRED1: Coupled ensemble-prediction experiment without DA.

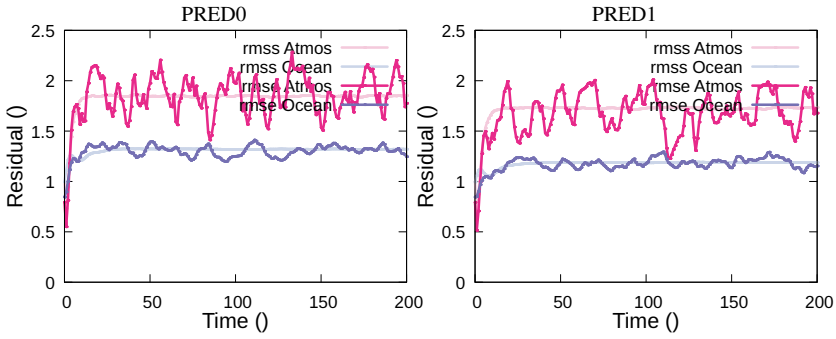

FIG. 4. The plots show the time evolution of the residuals for the ensemble predictions from the PRED0 and PRED1 experiments. The full lines indicate root-mean-square errors (RMSE) relative to the reference solution, while the soft lines are the ensemble-predicted RMSE.

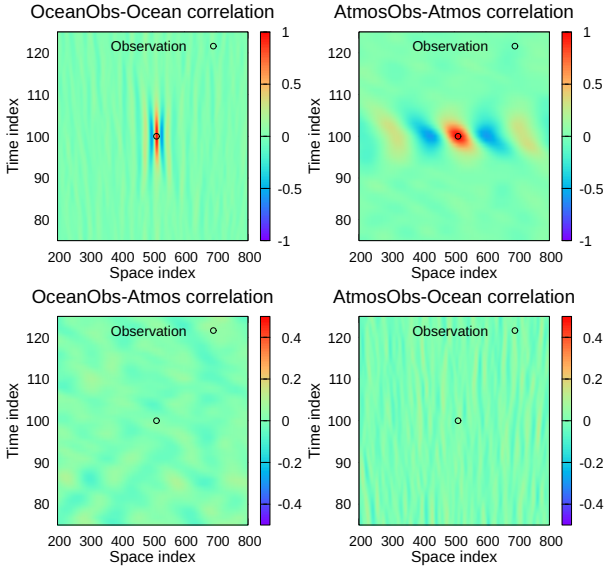

FIG. 5. Exp. PRED0: Correlations from uncoupled experiment.

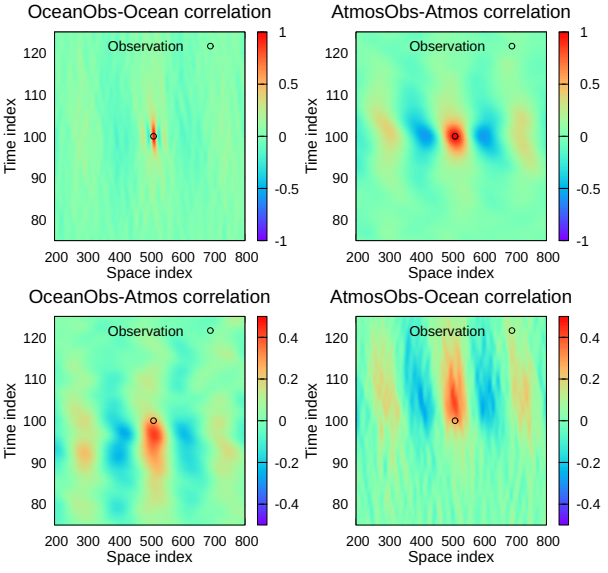

FIG. 6. Exp. PRED1: Correlations from coupled experiment.

between the variables, as seen in Fig. 6. These cross-correlations will cause the measurement of one model component to lead to an update of the other model component when computing an assimilation update. The cross-correlations indicate that an Ocean observation will influence the Atmos variable in the past while an Atmos observation predominantly influences the Ocean variable in the future. Also, with these cross-correlations, an Ocean observation will update the Atmos variable on the Atmos variable's spatial and time scale. In contrast, an Atmos observation will update the Ocean variable on multiple scales. I.e., the update introduces a prominent Atmos-scale feature and a finer Ocean-scale variability to the Ocean variable. From studying the space-time correlations, we observe that the Atmos model partly drives the Ocean model.

b. Coupled data-assimilation experiments

In the following experiments, we have used ESMDA with five steps and assimilation windows with a length of five units of time. On the other hand, we have varied the measurement densities to emphasize the results; see Tab. 1. Note that the first measurements become available at time $t = 50$.

We start by testing the impact of coupled assimilation of only Ocean data in Exp. KS-MDA-5-O and only Atmos data in Exp. KS-MDA-5-A, and we present the results in Figs. 7 and 9. Here, we jointly update the Ocean and Atmos variables when we assimilate the Ocean or the Atmos data. Complementary to these experiments, we have run Exps. KS-MDA-5-Osep and KS-MDA-5-Asep, using "uncoupled" data assimilation. We only update the Ocean variable using the Ocean data and only the Atmos variable when conditioning on the Atmos data; see Figs. 8 and 10.

Experiment	window length	num ocean obs	num atmos obs	Δt ocean obs	Δt atmos obs
KS-MDA-5-O	5	40	0	2	—
KS-MDA-5-Osep	5	40	0	2	—
KS-MDA-5-A	5	0	15	—	4
KS-MDA-5-Asep	5	0	15	—	4
KS-MDA-5	5	40	10	5	5
KS-MDA-5sep	5	40	10	5	5

TABLE 1. The table summarizes the coupled versus uncoupled data-assimilation experiments.

Assimilation of only Ocean data in experiment KS-MDA-5-O controls the Ocean variable well, as seen in Fig. 7. Interestingly, the information from the Ocean data also improves the Atmos variable significantly. After a few data-assimilation windows, we have recovered the Atmos variable as is clear from the residuals in Fig. 13. Note that we have a relatively high density of Ocean measurements in space and time to resolve the scales in the Ocean model variable. In this case, the Atmos solution's recovery results from the coupled data assimilation. In contrast, the results of Exp. KS-MDA-5-Osep illustrate that when we only update the Ocean variables when assimilating Ocean data, we see hardly any impact on the Atmos variable, and we also observe worsening Ocean results.

The situation differs from the alternative case, KS-MDA-5-A, where we only assimilate Atmos data. We obtained good convergence for the large-scale Atmos model, but with the correct spatial resolution in the data, we would need more information to control the Ocean model. We see some impact of the assimilation in the Ocean variable in Fig. 9 where we recover the Ocean's "large-scale" structures and somewhat reduce the estimated standard deviation. However, by comparing with Exp. KS-MDA-5-Asep in Fig. 10, this improvement in the Ocean variable is primarily due to the model coupling and dynamic interaction between the Atmos and Ocean variables.

The difference between the combined assimilation of Ocean and Atmos data in experiment KS-MDA-5 and the separate assimilation of Ocean and Atmos data in experiment KS-MDA-5sep emphasizes the value of the combined assimilation. Exp. KS-MDA-5 presented in Fig. 11 converges after a few assimilation windows, and we can control the further evolution of the model ensemble. Exp. KS-MDA-5sep shown in Fig. 12 converges much slower, and the resulting estimate has more significant errors for the Ocean and the Atmos variables. The residual plots in Fig. 13 also support the conclusion that combined assimilation of the Ocean and Atmos data captures the interaction between Ocean and Atmos more accurately and can lead to an improvement in the state estimate compared to the results of separate assimilation for the Ocean and Atmos variables.

5. Sensitivity study for ensemble DA methods

We will now further examine the properties of the iterative ensemble smoothers with our coupled model. The focus is not on the coupling but rather how to best configure the assimilation setup for the different assimilation methods in terms of the update strategy for the final iteration or step, the length of the data-assimilation window and impact of nonlinearity relative to the model's predictability time, the optimal number of steps in ESMDA, and the required number of iterations in IES. We note that the current versions of IES and ESMDA were previously only studied on a single window, while we will now study them in a recursive setting, where the assimilation updates in future assimilation windows will benefit from the updates from the previous assimilation steps, and thereby a reduced nonlinearity and a more Gaussian pdf of the prior ensemble.

a. Window update or model rerun

From Sec. 3, we noticed that we can update the solution at any instant over the assimilation window when using ES. Furthermore, for an assimilation window, the update uses the same linear combination of the prior ensemble realizations independently of the time step we are updating. Hence, for the ES update, we can choose between two strategies: update the solution over the whole window or update only the assimilation window's initial conditions and rerun the model to obtain the solution over the entire window. We will see that these two strategies lead to significantly different results for nonlinear dynamical models, while there would be no difference for linear models without model errors.

For ES, it is always better to update the whole window and, particularly, the end time of the window, as this estimate becomes the initial condition for the continued integration over the following window. The importance of updating the whole window depends on the length of the assimilation window relative to the model's predictability time. For short time windows, the error growth caused by the nonlinear and unstable model dynamics will not have time to impact the prediction significantly. Thus, whether we update the window's initial conditions or compute the linear ES update over the whole window in the short-window case does not lead to very different results.

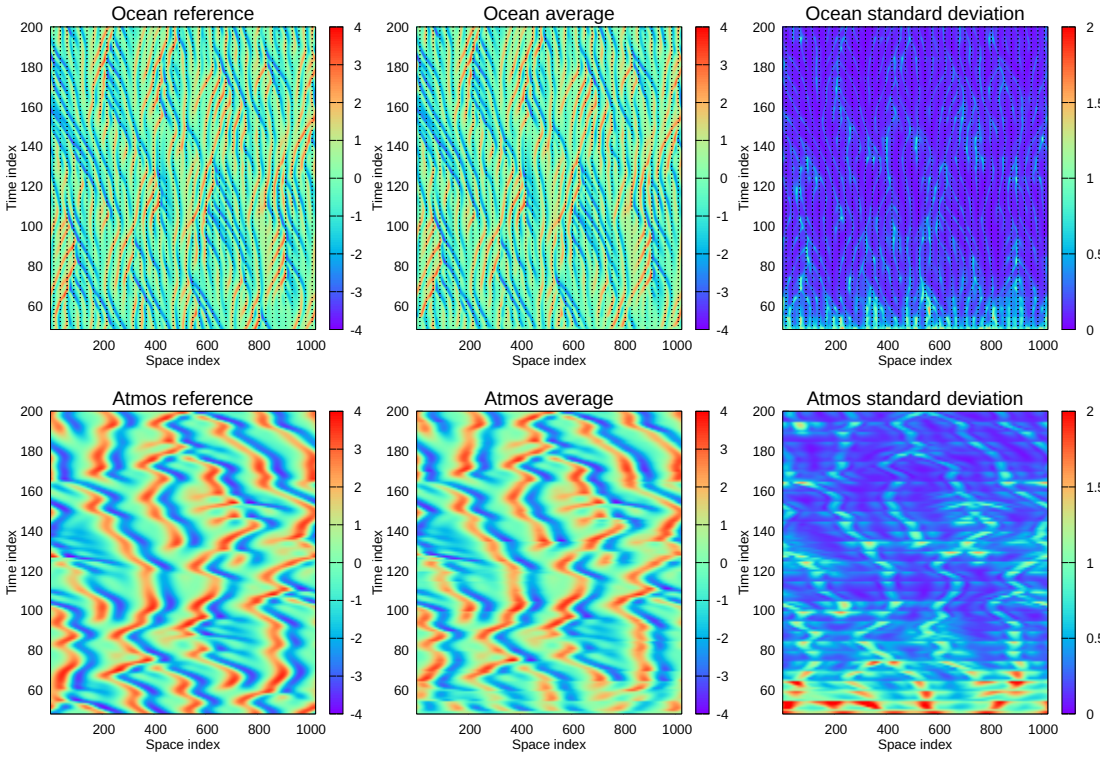

FIG. 7. Exp. KS-MDA-5-O: Assimilation of Ocean data and updating both the Ocean and Atmos variables.

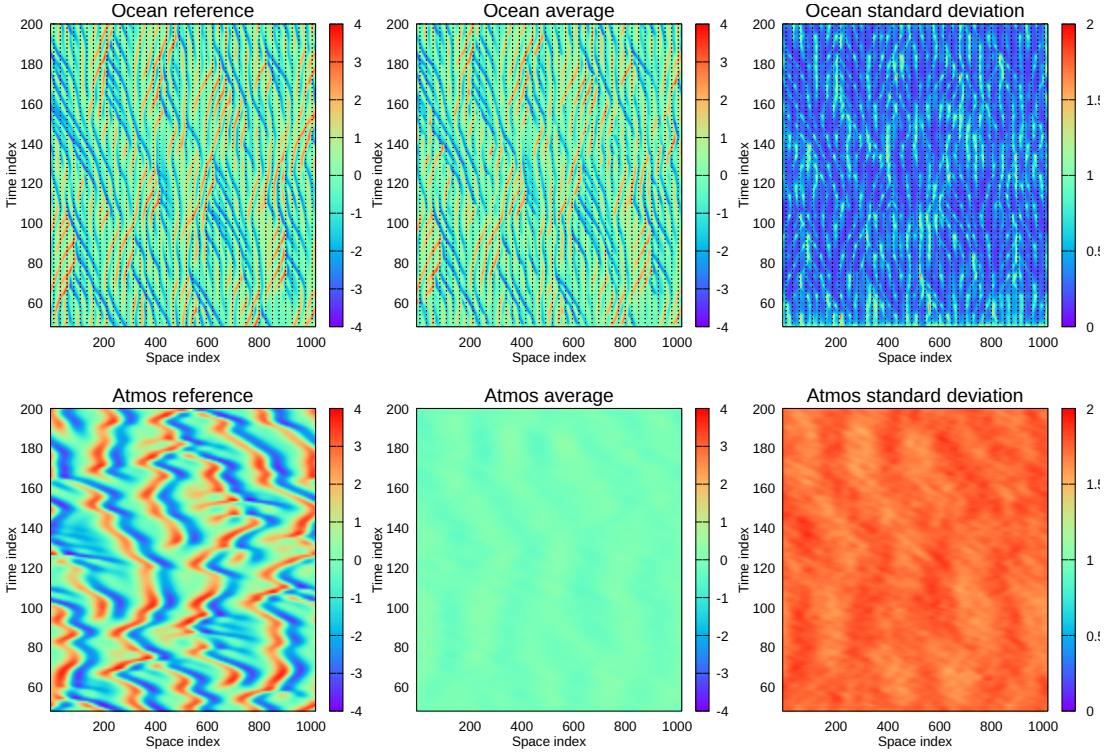

FIG. 8. Exp. KS-MDA-5-Osep: Same as Fig. 7 but with seperate assimilation where the Ocean data only update the Ocean variable.

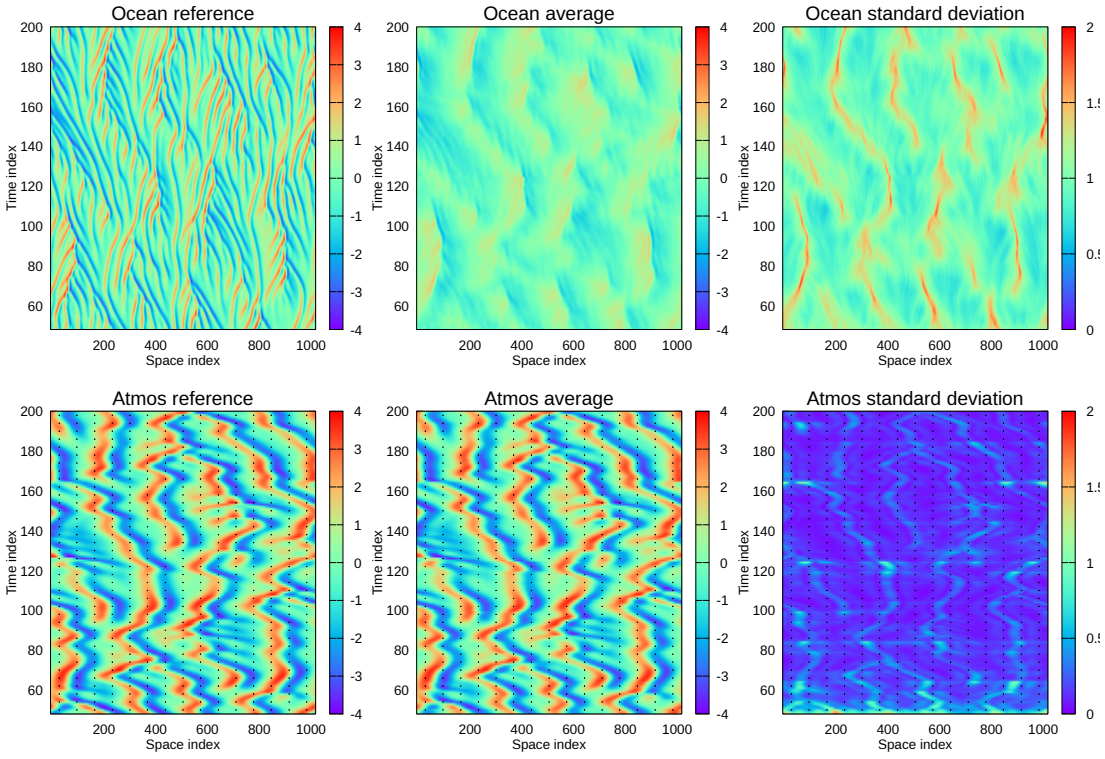

FIG. 9. Exp. KS-MDA-5-A: Assimilation of atmos data and updating both the Ocean and Atmos.

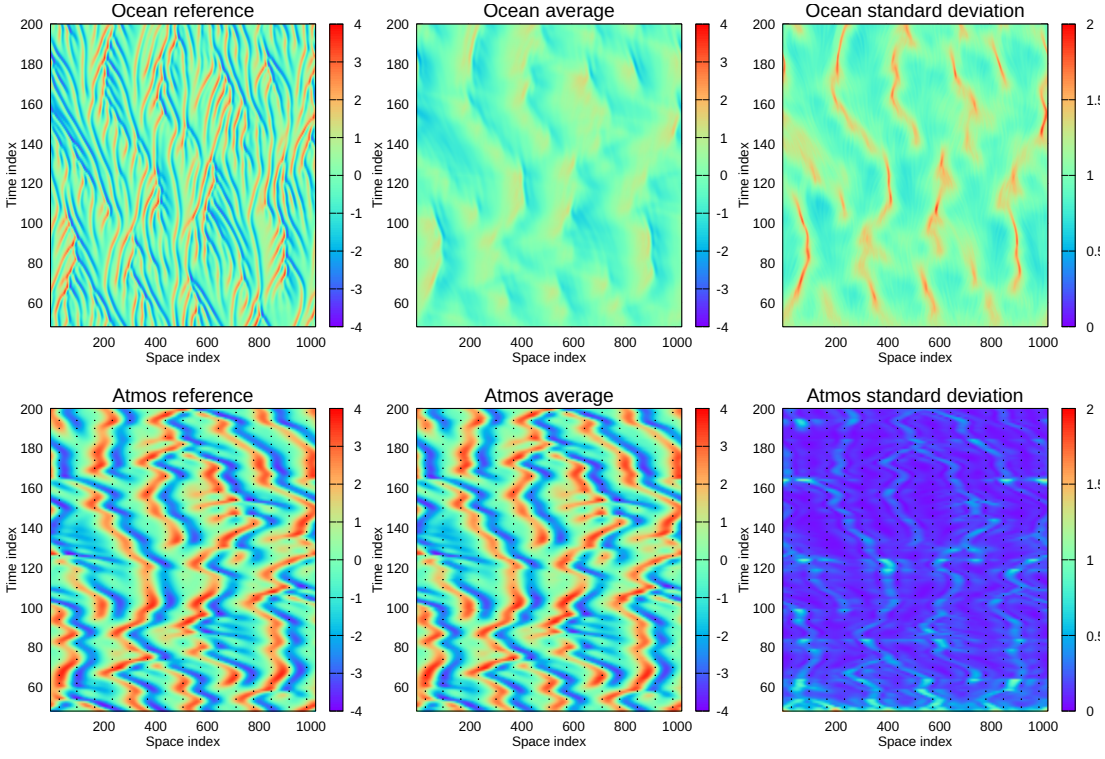

FIG. 10. Exp. KS-MDA-5-Asep: Same as Fig. 9 but with seperate assimilation where the atmos data only update the atmos variable.

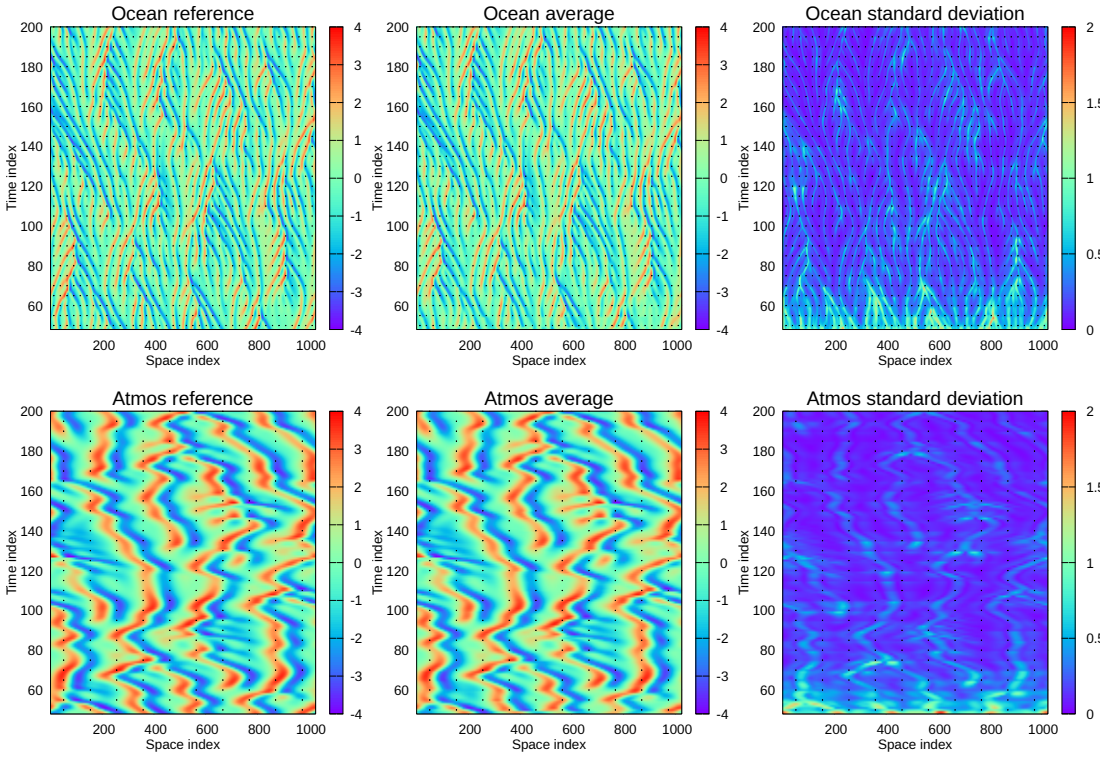

FIG. 11. Exp. KS-MDA-5: Assimilation of Ocean and Atmos data and updating both the Ocean and Atmos.

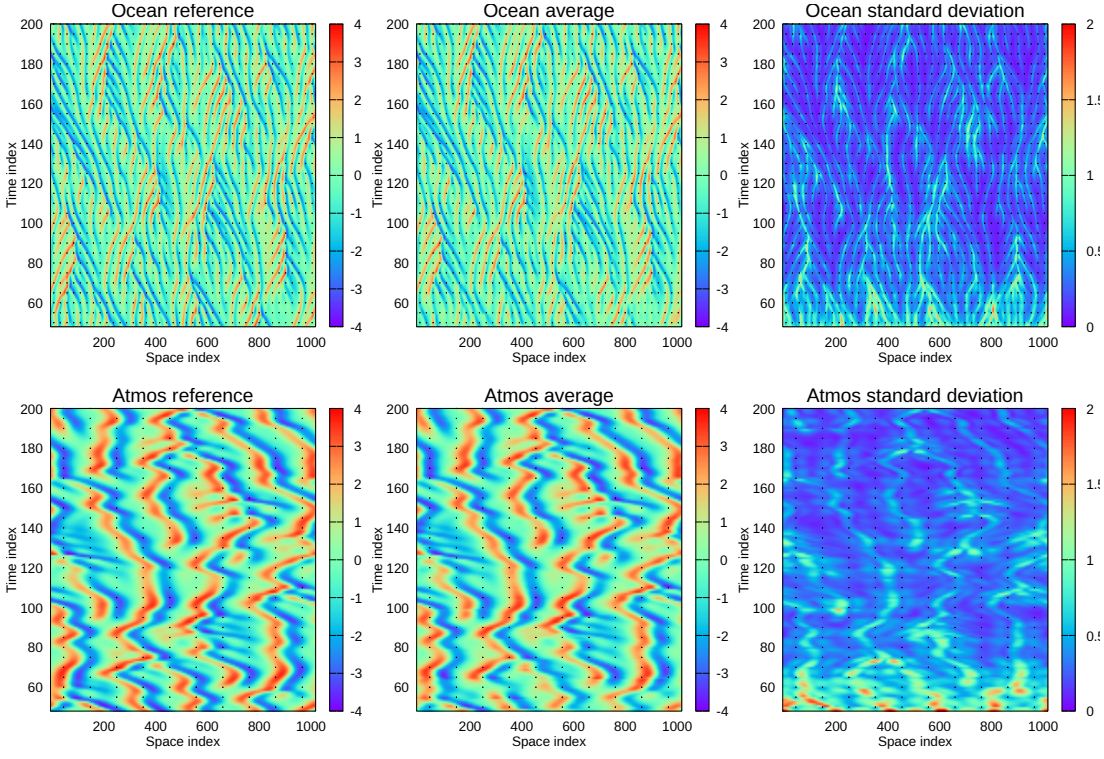

FIG. 12. Exp. KS-MDA-5sep: Same as Fig. 11 but with seperate assimilation for the Ocean and the Atmos variables.

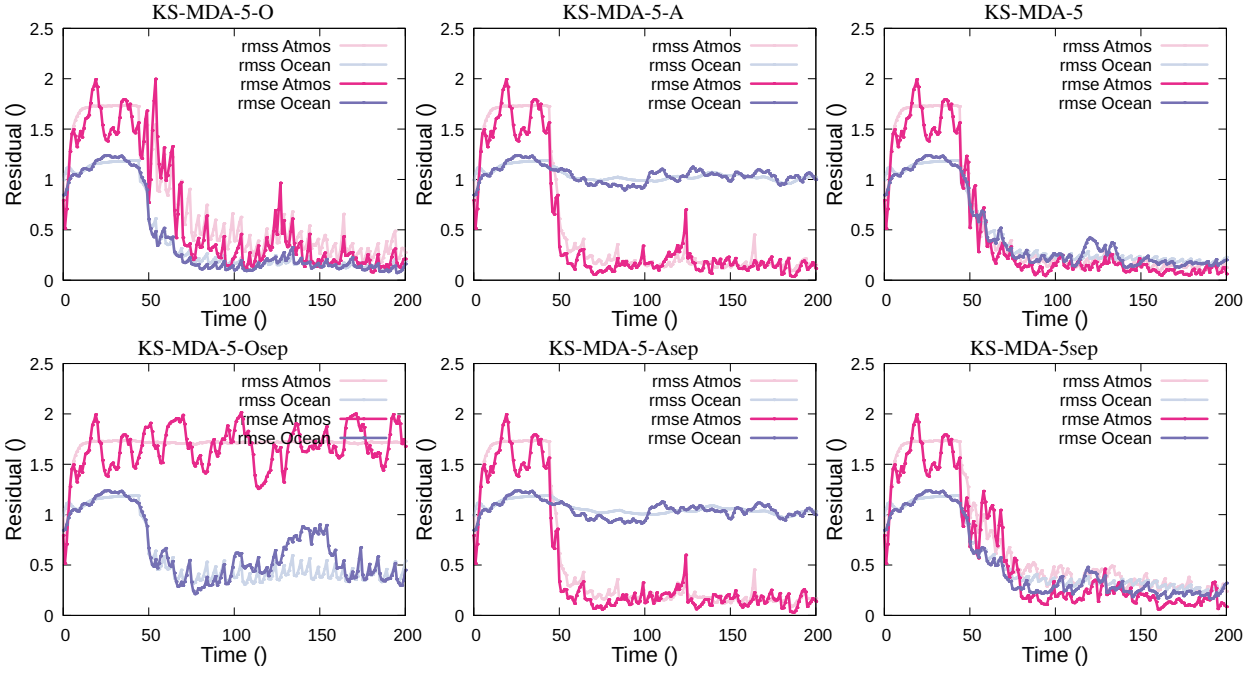

FIG. 13. The plots show the time evolution of the residuals for the KS-MDA-5 experiments listed in Tab. 1. The full lines indicate RMSE relative to the the reference solution while the soft lines are the ensemble predicted RMSE.

However, for longer time windows, the uncertainty at the end of the window obtained by integrating the model ensemble from updated initial conditions typically tends towards climatology, and the ensemble prediction in the next assimilation window starts all over again from a climatological level where the ensemble has forgotten all information from previously assimilated data.

In Figs. 14 and 15, we illustrate the difference between the two update strategies when using ES for a case with window lengths of six and measurements every second unit of time. We obtain a faster convergence and lower standard deviations when we update the ensemble over the whole window. Note that this case corresponds to computing an EnKF (or rather an EnKS) solution, where we have used measurements within an assimilation window to update the ensemble at the end of the assimilation window before continuing the integration.

ESMDA updates the assimilation window's initial conditions recursively, and following each update, we must rerun the ensemble over the window to obtain the solution for the current ESMDA step. This procedure increases the ensemble spread and uncertainty over the window due to using a nonlinear and unstable model. However, we can control some uncertainty growth at the final update step by calculating the ESMDA update over the whole window. This approach avoids the final ensemble integration. In the case of long assimilation windows, we obtain a better estimate of the solution at the end of the assimilation window and, thereby, a better estimate of the initial conditions of the following window. Figs. 16 and 17 illustrate how computing the solution over the whole window in the final ESMDA step improves the estimate. In particular, we notice a faster convergence for the Atmos variable, and we are better at controlling the dynamic instabilities that develop while integrating the model over the window. In an operational weather or climate prediction setting, we would typically update the solution at the end of the assimilation window at the final ESMDA step before predicting the solution over the following window.

The situation changes entirely in the IES as shown in Figs. 18 and 19. Updating the solution over the whole window in the final iteration significantly degrades the estimate. When using IES, we should update the initial condition of the assimilation window in the final integration. The reason ESMDA experiences an improvement, while the IES solution degrades, from updating the solution over the whole window in the final step or iteration is the following. *When computing the last update in ESMDA, the solution is a linear combination of the ensemble simulations of the previous MDA step. On the other hand, in IES, the solution is a linear combination of the prior ensemble.* Each step in ESMDA improves the initial conditions for the assimilation window, leading to an improved ensemble prediction over the assimilation window with lower uncertainty. Thus, in the final update step, the prior ensemble is already close to the posterior solution.

The summary in Fig. 20 clearly illustrates the above findings. The most surprising observation from this figure

Experiment	window length	num ocean obs	num atmos obs	Δt ocean obs	Δt atmos obs
KS-ES-6-2	6	40	10	2	2
KS-MDA-12-5	12	40	10	5	5
KS-IES-5-2	5	40	10	2	2
KS-ES-6-2X	6	40	10	2	2
KS-MDA-12-5X	12	40	10	5	5
KS-IES-5-2X	5	40	10	2	2

TABLE 2. The table summarizes the experiments used for testing whether to update initial conditions or the whole ensemble over the data assimilation window. The experiment names ending with an "X" update the initial conditions of the data-assimilation window and then rerun the model ensemble to obtain the final solution over the window.

is that the window length used for ESMDA significantly exceeds the one used with IES. We will see in the following discussion that the possibility of using a standard ES update for the final ESMDA step makes ESMDA less sensitive to the window length when used in sequential data assimilation. ESMDA also tolerates more significant errors in the initial conditions of the ensemble since we can partly correct these errors in the final ES-type update step. Furthermore, while ESMDA computes recursive linear regression updates, IES is a gradient-descent method and becomes sensitive to substantial nonlinearities. We will elaborate more on these topics in the following sections.

1) ENSEMBLE SMOOTHER SENSITIVITY EXPERIMENTS

We will now examine how the ensemble smoothers perform when we vary the length of the assimilation windows. In the case of ES, we repeat the experiments for window lengths increasing from one to nine units of time, and the time interval between the measurements is two units of time for all the experiments. We have ten and forty equally spaced measurements of the Atmos and Ocean models at each measurement time. We denote the experiments KS-ES-[1-9]-2[X], where KS denotes the Kuramoto-Sivashinksy model, ES means the ensemble smoother, and the numbers [1-9] define the length of the assimilation windows. Finally, an X at the end tells us we reran the model ensemble to obtain the solution. In contrast, the experiments without an X computed the ES update over the whole window.

Note that the ES and the EnKF will always have identical solutions at the end of the assimilation window; therefore, the prior for the following window will also be the same. We have already shown that we obtain better results from ES when we update the solution at the end of the assimilation window than when we update the window's initial conditions and rerun the ensemble over the window. From the upper plots in Fig. 20, we notice that ES generally leads to more accurate solutions when updating the solution over the window and not rerunning the ensemble. Updating the window's initial conditions results in poorer performance for all window lengths, and we can conclude that the standard EnKF with updates at the end of the assimilation window is the way to go.

Regarding the length of the assimilation window, the ES experiments perform well for a short window length of two units of time. Still, the performance deteriorates rapidly for longer assimilation windows. This worsening of the results with increasing window length comes from the model's nonlinearity, which results in a prior ensemble prediction over the window that approaches climatology for long windows. Then, the linear update computed by ES will break down, as previously discussed by Evensen and Van Leeuwen (2000) in an example using the Lorenz equations. From the upper left plot in Fig. 20, we notice how the standard deviation over the window grows until the simulation approaches climatology with long. The net effect is that at the end of the assimilation window, no predictive skill propagates forward into the following window. So, when using ES and EnKF, we should use short data-assimilation windows and update the solution at the end of, or the whole, window.

2) IES SENSITIVITY EXPERIMENTS

We now move on to examining IES in more detail. Similarly to the ES experiments in the previous section, we use the notation KS-IES-[1-15]-2[X] for the IES experiments. With IES, we could extend the length of the assimilation window, and we have run experiments with assimilation window lengths up to 15 units of time. In all the experiments, we used a maximum of 12 iterations, but in most cases, we observed no improvement after about five to eight iterations.

The plots in the second row of Fig. 23 summarize the residuals from using IES with different assimilation window lengths. IES performs equally well for window lengths of two to seven units of time, as seen from the right plot in the second row of Fig. 20, and the results are significantly better than those obtained from using ES (see upper left panel in the figure). Thus, an iterative method has the benefit of a more accurate estimate, but it comes at a higher computational cost.

As for ES, we repeated all 15 KS-IES-[1-15]-2X experiments by a set of experiments KS-IES-[1-15]-2 where we used the IES transition matrix to update the whole window in the final iteration. However, from the residuals in Fig. 23, it is clear that for IES, in contrast to ES, we should

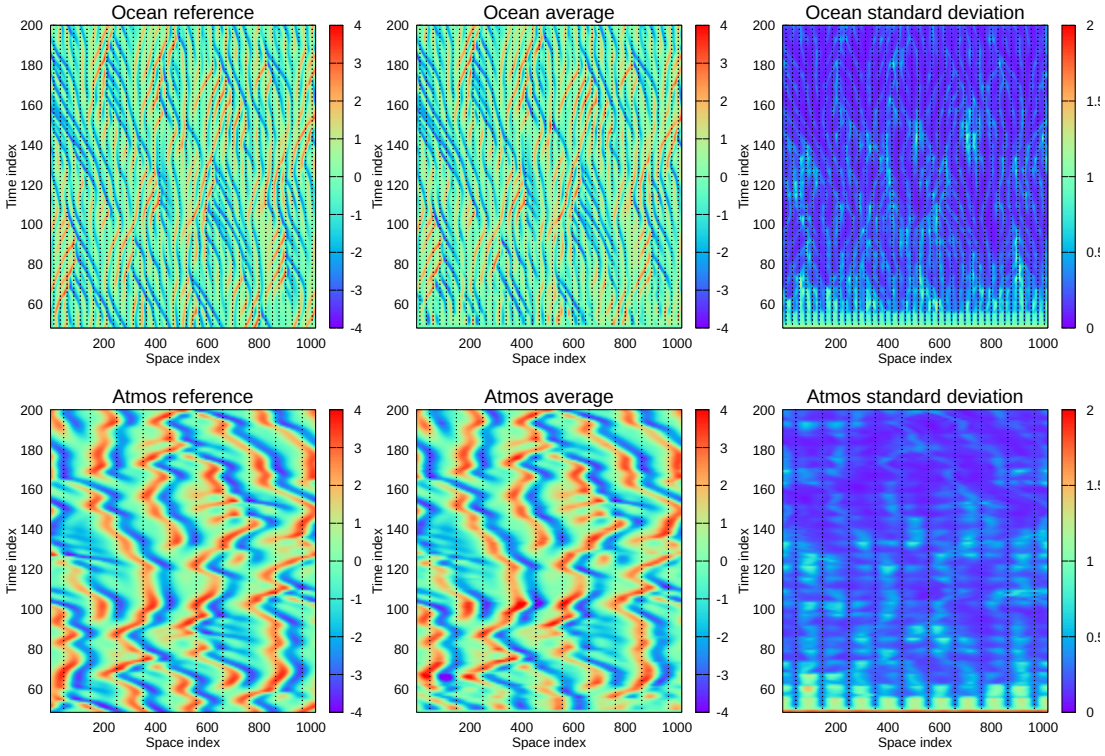

FIG. 14. Exp. KS-ES-6-2: ES with an assimilation window of length 6 time units and DA update of final solution over window.

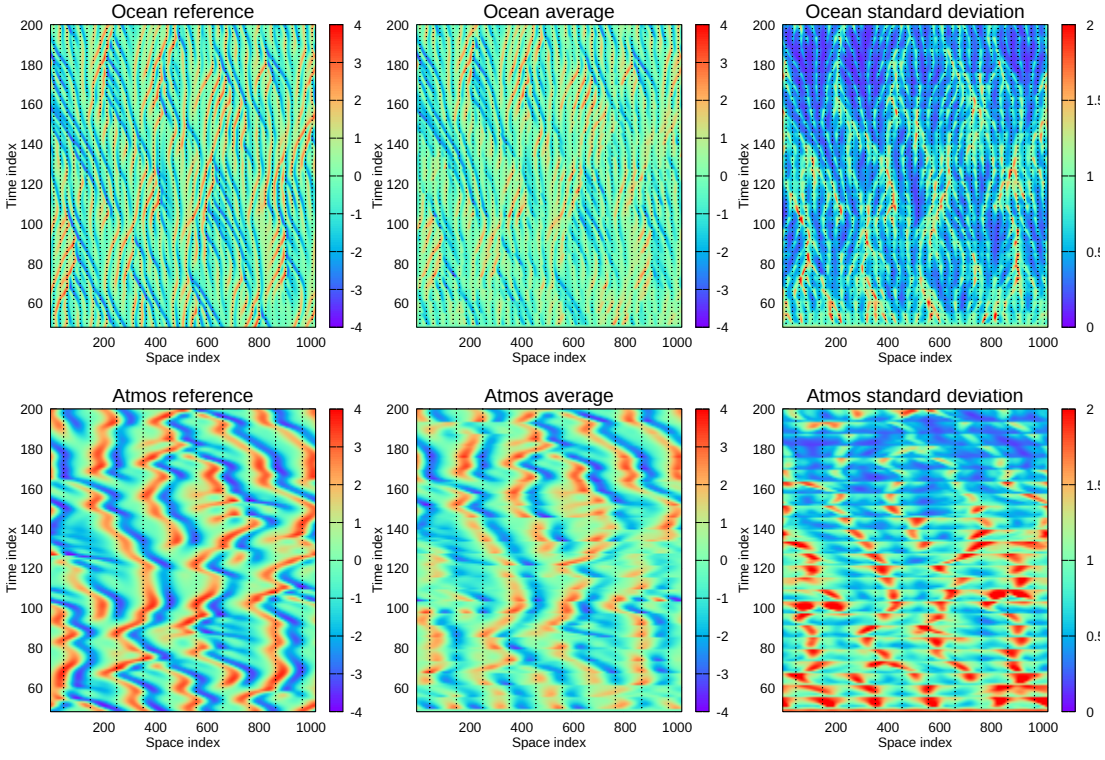

FIG. 15. Exp. KS-ES-6-2X: Same as Fig. 14 but with update of the window's initial condition.

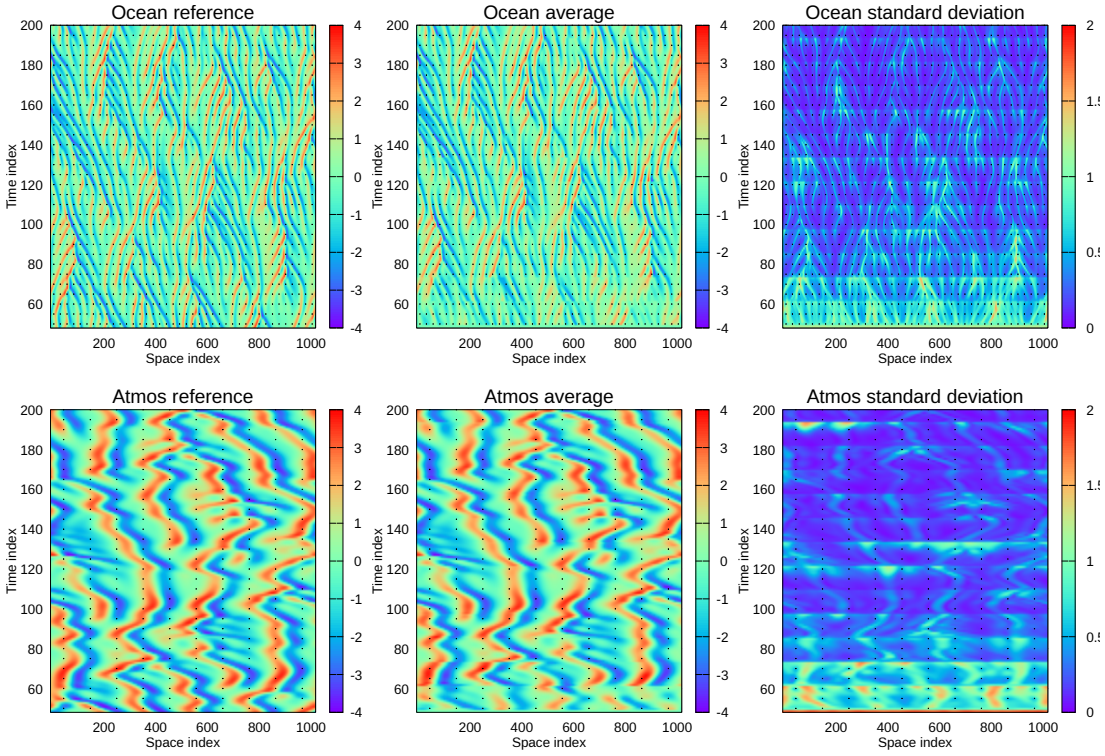

FIG. 16. Exp. KS-MDA-12-5: ESMDA with an assimilation window of length 12 time units and DA update of final solution over window.

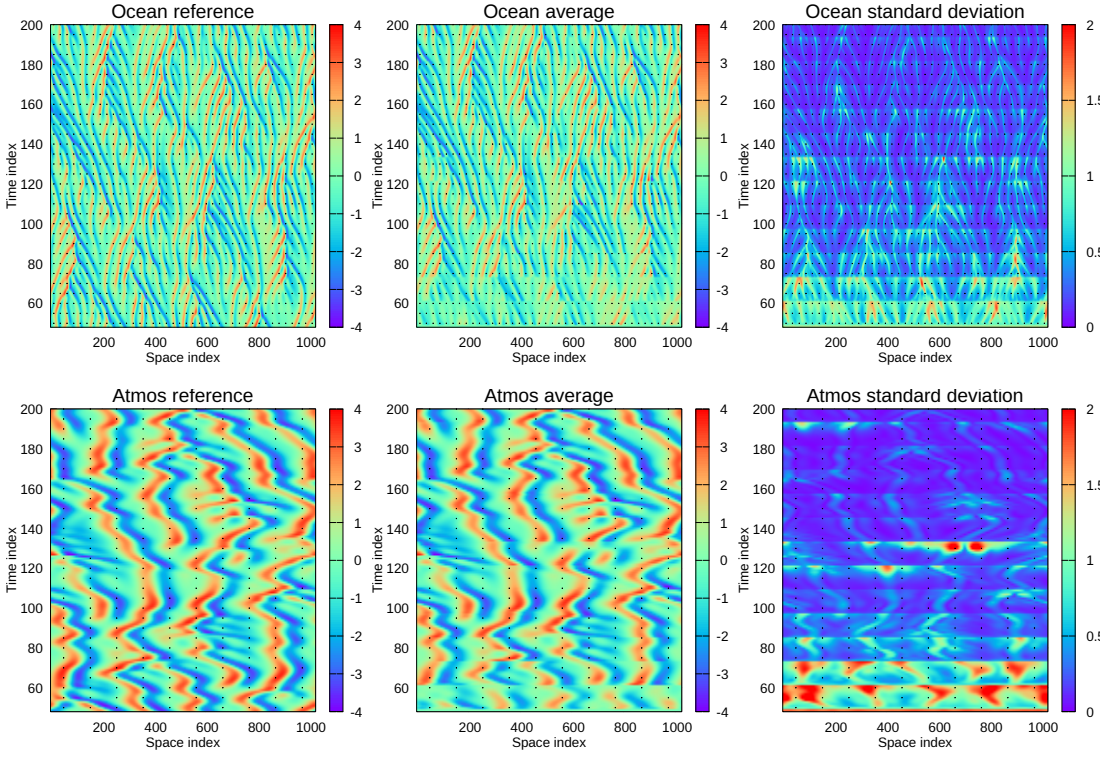

FIG. 17. Exp. KS-MDA-12-5X: Same as Fig. 16 but with update of the window's initial condition.

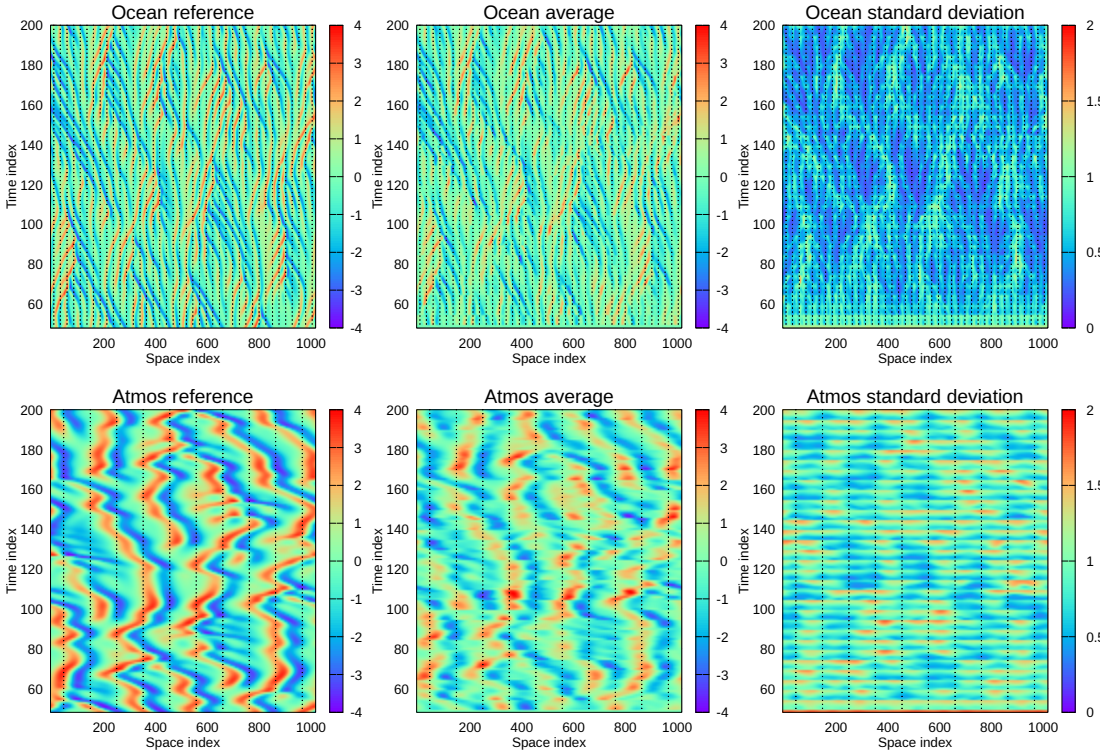

FIG. 18. Exp. KS-IES-5-2: IES with an assimilation window of length 5 time units and DA update of final solution over window.

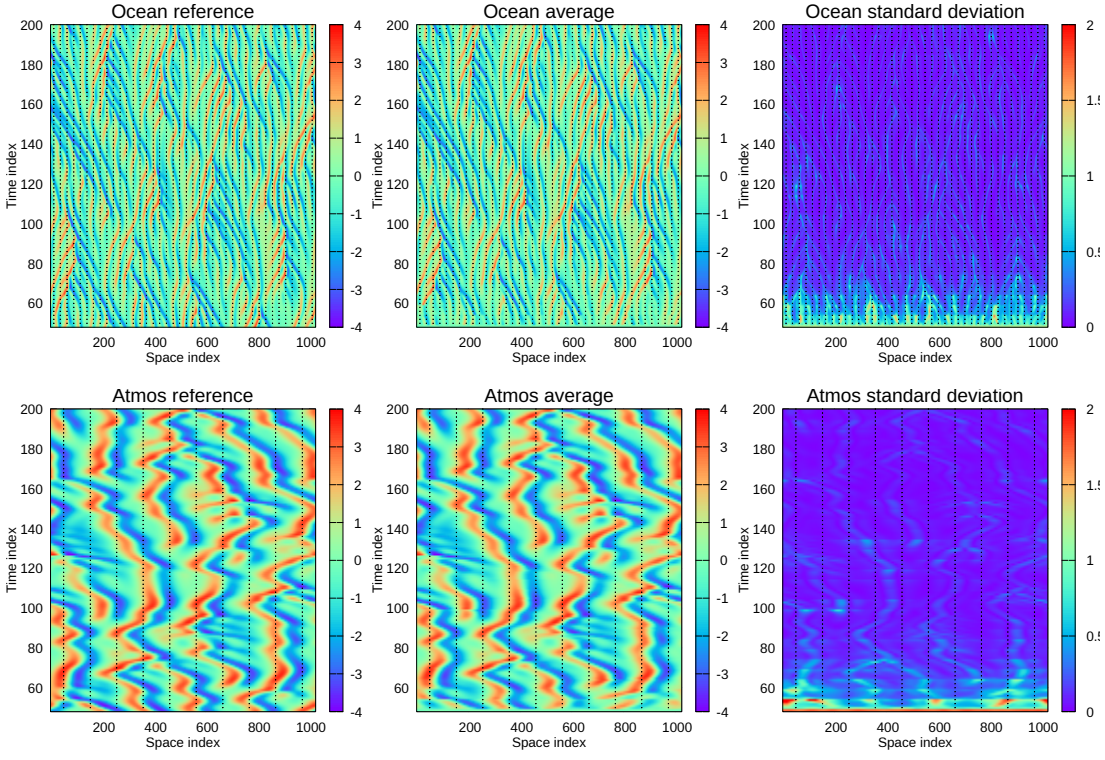

FIG. 19. Exp. KS-IES-5-2X: Same as Fig. 18 but with update of the window's initial condition.

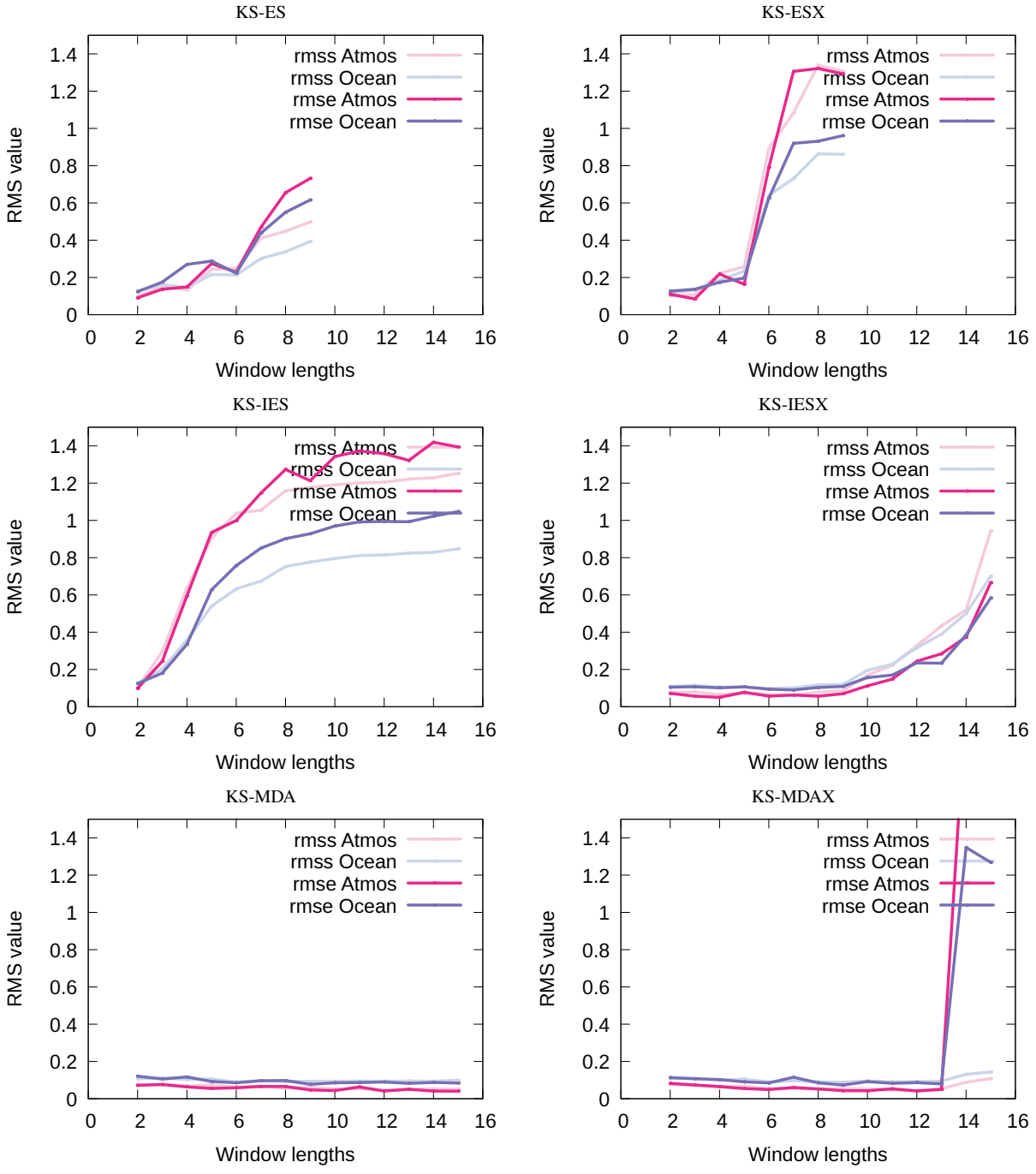

FIG. 20. The plots show the time-averaged residuals for the ES, ESMDA, and IES experiments listed in Tab. 2 as a function of window length for the period 101-200. The full lines indicate RMSE relative to the reference solution, while the soft lines are the ensemble predicted RMSE.

rerun the model ensemble in the final iteration. We explained in Sec. 3 that IES computes the updated ensemble of initial conditions as a linear combination of the prior ensemble initial conditions. Furthermore, we recall that the linear combination defined by the transition matrix leads to a posterior ensemble prediction that minimizes the ensemble of cost functions in Eq. (6). Thus, by computing the posterior ensemble over the data-assimilation window by multiplying the prior ensemble by the transition matrix, we lose the effect of the model evolution's nonlinearity and cannot improve the solution this way. Note, however, that

also here the estimates would become identical when using a linear model.

3) ESMDA sensitivity experiments

We now apply the ESMDA method in experiments KS-MDA-[1-15]-2[X], similar to those from the previous section. The ESMDA formulation for the recursive data-assimilation problem assumes that we update the initial state and rerun the model ensemble in each step to recompute the gradient. However, since ESMDA computes a sequence of independent ES updates, we can choose whether to update the initial conditions for the assimilation window or the whole solution over the window in the final ESMDA step.

From the plots in the bottom row of Fig. 20, we notice that ESMDA performs significantly better than ES in both the case with updating the whole window and when we rerun the model ensemble over the window. We can extend the window length significantly compared to when using ES, and there are hardly any differences between the ESMDA or IES solutions for window sizes up to six. We note, however, that the ESMDA residuals diverge from the ensemble standard deviations for window lengths greater than six. Thus, we cannot trust the predicted error estimates for longer assimilation windows.

We also note that computing an ES update of the whole window in the final step stabilizes the solution for longer windows and improves the estimate, particularly for long window lengths. This strategy also saves one ensemble integration.

We conclude that we are less impacted by the model's nonlinearity when using ESMDA than when using ES, and ESMDA has the added advantage that we can extend the data-assimilation window significantly. We also obtained a substantially better result with ESMDA than ES, although at a higher computational cost for the same window length.

b. Sensitivity to number of MDA steps

The number of ESMDA steps can significantly impact the solution (Evensen 2018; Evensen et al. 2020). Ideally, we would like to use as few steps as possible to reduce computational cost since every step implies a simulation of the whole ensemble. Using around four to eight ESMDA steps is common, but we currently have no general rule for how many steps we need to obtain the best estimate. We can only try a different number of steps, and when the solution does not change within the expected sampling errors, we can assume we have converged.

An issue is that every ESMDA step is an independent ES update and introduces a new set of perturbed observations. Another problem is that using a limited ensemble size introduces sampling errors. In the following experiments, we have used 1000 realizations to minimize the impact of sampling errors, and we have used data-assimilation windows of five, ten, and fifteen units of time. We ran nine experiments using one to nine ESMDA steps, and we plot the residuals in Fig. 21, averaged over the converged solution during the last 100 units of time. The experiments using a single step are equivalent to running ES, and it is clear that using more than one step improves the solution. Using two MDA steps leads to a significant improvement, and using two to nine steps improves the estimate for the shortest time window and results in nearly identical accuracy in the case with a window of five units of time. Thus, there is no benefit of using more than two to three steps in this case with modest nonlinearity. For the longer assimilation windows of 10 and 15 units of time, we obtain the most consistent solution using only three to seven MDA steps, all with similar performance. We also note that additional steps lead to divergence between the actual and estimated residuals. In the case with an assimilation window of 15 units of time and nine MDA steps, we experience substantial filter divergence.

Note that running experiments with a relatively small ensemble size makes it useless to run additional ESMDA steps if the method has already converged to a level lower than the sampling errors resulting from the limited ensemble size. A positive result from these experiments is that we can obtain estimates with an excellent accuracy using relatively few MDA steps.

6. Comparative performance of ensemble smoothers

Which ensemble method should we use? We can apply the ES with an update of the whole assimilation window or in a pure EnKF setting with an update of the final time step of the window. However, ESMDA and IES will improve upon the ES results in cases with significant non-linearity at an additional computational cost.

We like the consistency of the IES performance seen in the middle right plot of Fig. 20, where we obtain excellent results, and the converged solution shows consistency between actual and estimated residuals up to a window length of eight units of time. On the other hand, ESMDA converged in three to five steps for moderate assimilation window lengths and may be more computationally efficient than IES. ESMDA also works well with longer window lengths due to the final MDA step, where we compute an ES update over the whole window, which results in a higher accuracy and lower uncertainty of the initial conditions for the subsequent window.

In all the previous IES experiments, we ran a maximum of 12 iterations, but in most cases, IES converged in fewer iterations. We also realize that in a sequential data-assimilation system, the prior for each successive assimilation window is typically rather good, so we could reduce the number of iterations in IES without sacrificing the quality of the results. We also tested IES on the

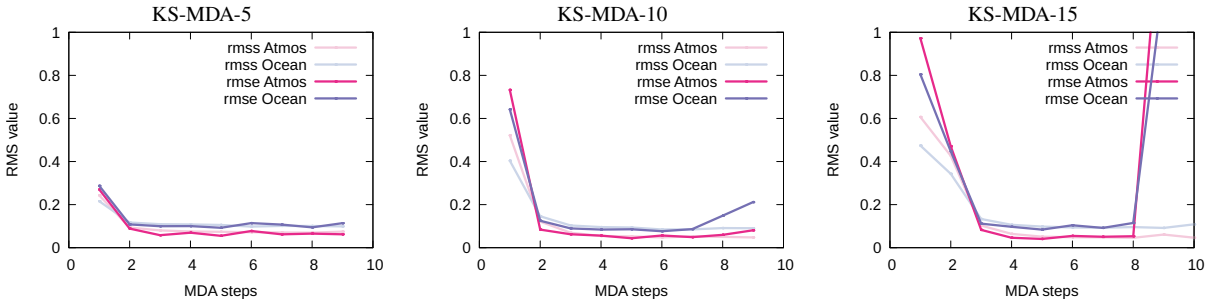

FIG. 21. The plots show time-averaged residuals for the ESMDA as a function of the number of MDA steps for three different window lengths of 5, 10, and 15 units of time for the period 101-200. The full lines indicate RMSE relative to the reference solution, while the soft lines are the ensemble predicted RMSE.

cases KS-IES-[2-6]-2X using a maximum of four iterations for each update, which in computational cost corresponds to running ESMDA with five steps since ESMDA avoids the final ensemble integration by updating the ensemble over the whole assimilation window in the last step. We note that the final ES update used in ESMDA helps control nonlinear instabilities and leads to an improved initial condition for the following window.

In Fig. 22, we compare the time series of residuals for the IES (KS-IES-[2-6]-2X), IES with four iterations (KS-IES4-[2-6]-2X), and the default case of ESMDA with five steps (KS-MDA-[2-6]-2) for the window lengths where we observed the methods to work well. We cannot conclude from these plots that one configuration is significantly better or worse than another. Furthermore, as weather prediction systems use relatively short data-assimilation windows compared to the weather's predictability time limit, we expect that both IES and ESMDA should converge to a sufficiently accurate solution with a modest number of iterations or steps. Fig. 23 compares the residuals from the different cases over the time interval 101-200, where the experiments have reached a quasi-steady accuracy level following the initial assimilation updates. Note that using a finite ensemble size introduces a small uncertainty. We could have repeated the experiments with different random seeds to obtain more robust estimates and plotted all residuals with an uncertainty estimate, but this becomes more relevant when using smaller ensemble sizes than the 1000 realizations used in all the experiments in this paper.

An additional sensitivity study would involve varying ensemble sizes. We have decided to exclude such a study in this paper since it would also include using localization methods. We have left such a study for future work.

In addition to the actual performance issues, there could also be preferences regarding the theoretical formulations of the different techniques, i.e., direct RML sampling of the posterior in IES and gradual transport of the prior towards the posterior using RML sampling and ES for the short linear update steps in ESMDA. IES introduces significant errors through the RML sampling in nonlinear systems. Could we assume that ESMDA solves for minuscule ES updates that introduce Gaussianity into the next window's prior? The recursive-in-time updating process in sequential data assimilation may also reduce the RML sampling approximation.

7. Summary

In this study, we have demonstrated the possibility of using adjoint-free iterative ensemble methods for sequential data assimilation. We defined a coupled multiscale model system based on the nonlinear Kuramoto-Sivashinsky model. The model system allows for examining data assimilation in multiscale dynamical systems and models with unstable dynamics. An advantageous property of this model is its saturation of uncertainty at a climatologic level, similar to what we observe in atmospheric and ocean models. In addition, the model is one-dimensional in space, allowing us to represent spatial variability in the model. In addition to demonstrating the importance and value of coupled data-assimilation updates, we have studied the properties of iterative ensemble smoothers. The joint assimilation of data for a model operating at different physical scales and with collective updates of both model components yields superior results compared to treating the model components and data independently. In particular, we demonstrate the importance of assimilating measurements from the component with the small spatial scales into the coupled system for reconstructing the component with the large spatial scales. We also explain the similarity between ensemble 4DVar and our iterative smoothers. The main differences include the replacement of the tangent-linear operator with an ensemble-averaged model sensitivity, which eliminates the use of adjoint models and backward integrations. Additionally, we represent all covariance matrices with an ensemble of model realizations. We demonstrated the effectiveness and efficiency of ESMDA and IES in dealing with nonlinearities in a coupled model. Furthermore, we illustrated the effect of updating over the assimilation window versus updating the window's initial conditions in ensemble smoothers. We

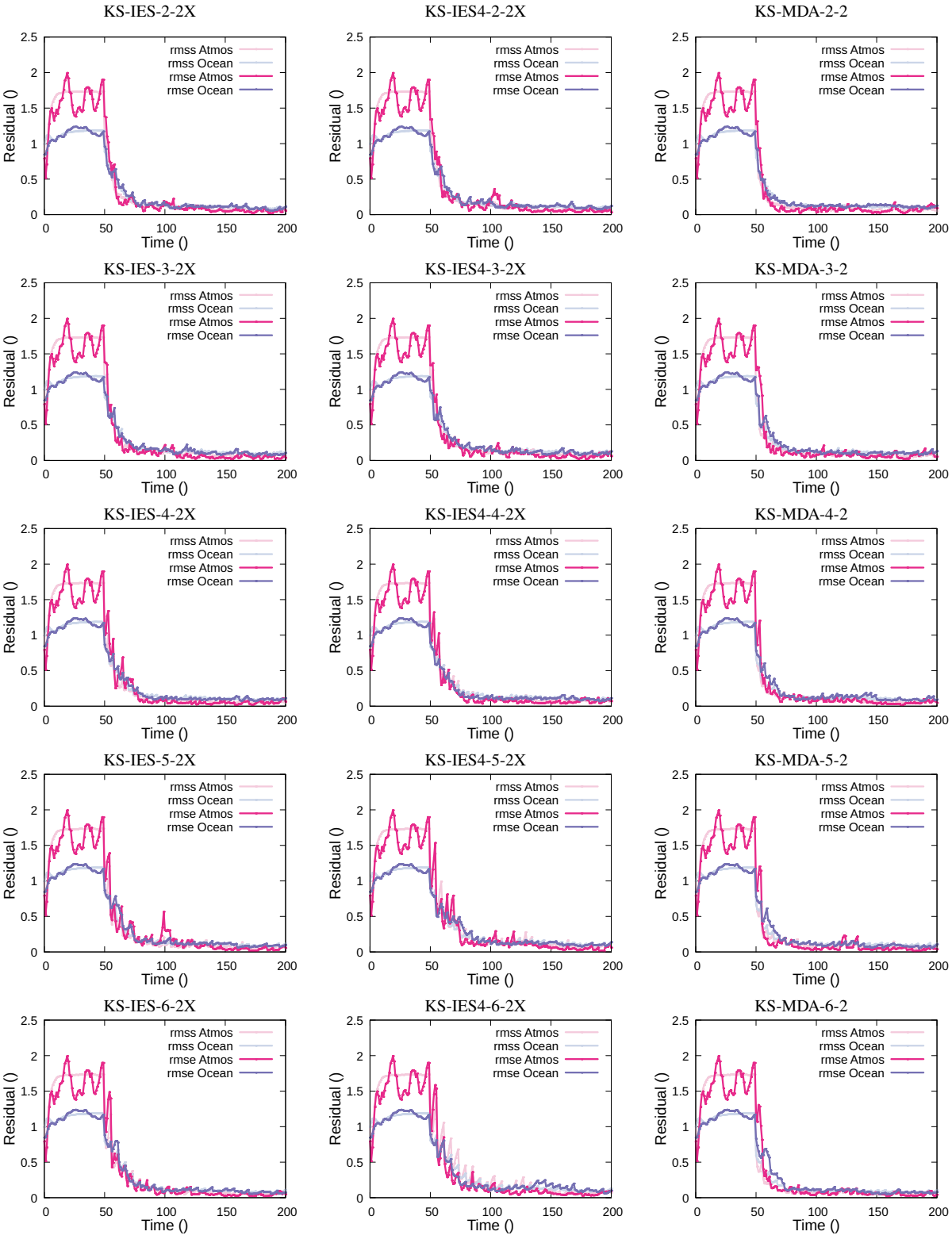

FIG. 22. The plots show the time evolution of residuals for the ensemble prediction using different window lengths for a converged IES (IES), an IES using only four iterations (IES4), and an ESMDA with five steps (MDA5). The full lines indicate RMSE relative to the reference solution, while the soft lines are the ensemble predicted RMSE.

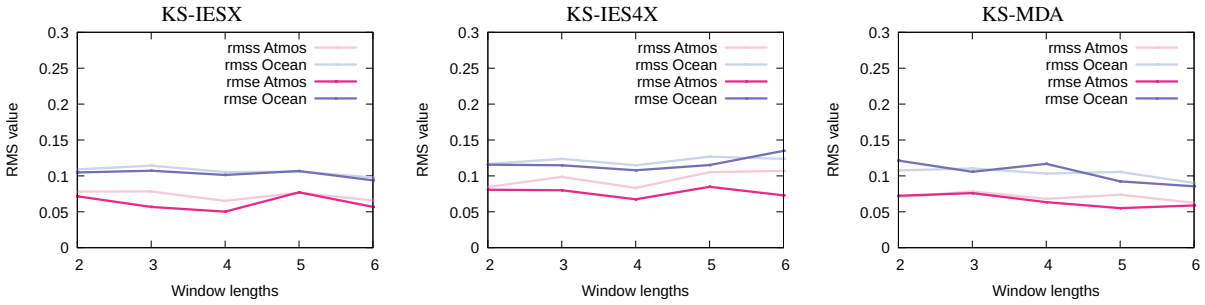

FIG. 23. The plots show time-averaged residuals for the IESX, IES4X, and ESMDA experiments as a function of window length for the period 101-200. The full lines indicate RMSE relative to the reference solution, while the soft lines are the ensemble predicted RMSE.

believe the methods and formulations discussed in the paper, formulated initially for petroleum applications, will be valuable for future data-assimilation systems for atmospheric, oceanic, and other coupled earth-system models.

References

Azouani, A., and E. S. Titi, 2014: Feedback control of nonlinear dissipative systems by finite determining parameters - a reaction-diffusion paradigm. *Evolution Equations and Control Theory*, **3 (4)**, 579–594, doi:10.3934/eet.2014.3.579.

Burgers, G., P. J. Van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.*, **126**, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Chang, H.-C., 1986: Nonlinear waves on liquid film surfaces—i. flooding in a vertical tube. *Chemical Engineering Science*, **41 (10)**, 2463–2476, doi:10.1016/0009-2509(86)80032-X.

Chang, Y. S., S. Zhang, A. Rosati, T. L. Delworth, and W. F. Stern, 2013: An assessment of oceanic variability for 1960-2010 from the gfdl ensemble coupled data assimilation. *Climate Dynamics*, **40**, 775–803, doi:10.1007/S00382-012-1412-2/FIGURES/16.

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

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

Chorin, A. J., and P. Krause, 2004: Dimensional reduction for a bayesian filter. *Proceedings of the National Academy of Sciences*, **101 (42)**, 15 013–15 017, doi:10.1073/pnas.0406222101.

de Rosnay, P., and Coauthors, 2022: Coupled data assimilation at ecmwf: current status, challenges and future developments. *Quarterly Journal of the Royal Meteorological Society*, **148 (747)**, 2672–2702, doi:https://doi.org/10.1002/qj.4330.

Emerick, A. A., and A. C. Reynolds, 2012: History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computat Geosci*, **16 (3)**, 639–659, doi:10.1007/S10596-012-9275-5.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10,143–10,162, doi:10.1029/94JC00572.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.

Evensen, G., 2018: Analysis of iterative ensemble smoothers for solving inverse problems. *Computat Geosci*, **22 (3)**, 885–908, doi:10.1007/s10596-018-9731-y.

Evensen, G., P. N. Raanes, A. S. Stordal, and J. Hove, 2019: Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Frontiers in Applied Mathematics and Statistics*, **5**, 47, doi:10.3389/fams.2019.00047.

Evensen, G., and P. J. Van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.*, **128**, 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Evensen, G., F. C. Vossepoel, and P. J. Van Leeuwen, 2022: *Data Assimilation Fundamentals: A Unified formulation for State and Parameter Estimation*. Springer, 254 pp., doi:10.1007/978-3-030-96709-3, Open access.

Evensen, G., and Coauthors, 2020: An international initiative of predicting the sars-cov-2 pandemic using ensemble data assimilation. *Foundations of Data Science*, 65, doi:10.3934/fods.2021001.

Han, G., X. Wu, S. Zhang, Z. Liu, and W. Li, 2013: Error covariance estimation for coupled data assimilation using a lorenz atmosphere and a simple pycnocline ocean model. *Journal of Climate*, **26 (24)**, 10 218 – 10 231, doi:10.1175/JCLI-D-13-00236.1.

Jardak, M., I. M. Navon, and M. Zupanski, 2010: Comparison of sequential data assimilation methods for the kuramoto–sivashinsky equation. *International Journal for Numerical Methods in Fluids*, **62 (4)**, 374–402, doi:10.1002/fld.2020.

Kitanidis, P. K., 1995: Quasi-linear geostatistical therory for inversing. *Water Resources Research*, **31 (10)**, 2411–2419, doi:10.1029/95WR01945.

Kuramoto, Y., 1978: Diffusion-Induced Chaos in Reaction Systems. *Progress of Theoretical Physics Supplement*, **64**, 346–367, doi:10.1143/PTPS.64.346.

Laloyaux, P., M. Balmaseda, D. Dee, K. Mogensen, and P. Janssen, 2016: A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **142 (694)**, 65–78, doi:https://doi.org/10.1002/qj.2629.

Lunasin, E., and E. S. Titi, 2017: Finite determining parameters feedback control for distributed nonlinear dissipative systems—a computational study. *Evolution Equations and Control Theory*, **6 (4)**, 535–557, doi:10.3934/eect.2017027.

Luo, X., and I. Hoteit, 2014: Ensemble Kalman filtering with a divided state-space strategy for coupled data assimilation problems. *Monthly Weather Review*, **142 (12)**, 4542 – 4558, doi:10.1175/MWR-D-13-00402.1.

Oliver, D. S., N. He, and A. C. Reynolds, 1996: Conditioning permeability fields to pressure data. *ECMOR V-5th European Conference on the Mathematics of Oil Recovery*, 11, doi:10.3997/2214-4609.201406884.

Penny, S. G., E. Bach, K. Bhargava, C.-C. Chang, C. Da, L. Sun, and T. Yoshida, 2019: Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, **11 (6)**, 1803–1829, doi:10.1029/2019MS001652.

Penny, S. G., and Coauthors, 2017: Coupled data assimilation for integrated earth system analysis and prediction: Goals, challenges and recommendations. Special Report WWRP 2017-3, WMO, Geneva, Switzerland.

Protas, B., T. R. Bewley, and G. Hagen, 2004: A computational framework for the regularization of adjoint analysis in multiscale pde systems. *Journal of Computational Physics*, **195 (1)**, 49–89, doi:10.1016/j.jcp.2003.08.031.

Raanes, P. N., A. S. Stordal, and G. Evensen, 2019: Revising the stochastic iterative ensemble smoother. *Nonlin. Processes Geophys*, **26**, 325–338, doi:10.5194/npg-2019-10.

Saha, S., and Coauthors, 2010: The NCEP climate forecast system reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015 – 1058, doi:10.1175/2010BAMS3001.1.

Shlang, T., and G. Sivashinsky, 1982: Irregular flow of a liquid film down a vertical column. *Journal de Physique*, **43**, 459–466, doi:10.1051/jphys:01982004303045900.

Sivashinsky, G., 1977: Nonlinear analysis of hydrodynamic instability in laminar flames—i. derivation of basic equations. *Acta Astronautica*, **4 (11)**, 1177–1206, doi:10.1016/0094-5765(77)90096-0.

Sivashinsky, G. I., 1980: On flame propagation under conditions of stoichiometry. *SIAM Journal on Applied Mathematics*, **39 (1)**, 67–82, doi:10.1137/0139007.

Tilley, B. S., S. H. Davis, and S. G. Bankoff, 1994: Nonlinear long-wave stability of superposed fluids in an inclined channel. *Journal of Fluid Mechanics*, **277**, 55–83, doi:10.1017/S0022112094002685.

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, doi:10.1007/s10955-020-02525-z.

Van Leeuwen, P. J., and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.*, **124**, 2898–2913, doi:10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.