Ensemble Kalman Filtering with a Divided State-Space Strategy for Coupled Data Assimilation Problems*

XIAODONG LUO

International Research Institute of Stavanger, Bergen, Norway

IBRAHIM HOTEIT

King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

ABSTRACT

This study considers the data assimilation problem in coupled systems, which consists of two components (subsystems) interacting with each other through certain coupling terms. A straightforward way to tackle the assimilation problem in such systems is to concatenate the states of the subsystems into one augmented state vector, so that a standard ensemble Kalman filter (EnKF) can be directly applied. This work presents a divided state-space estimation strategy, in which data assimilation is carried out with respect to each individual subsystem, involving quantities from the subsystem itself and correlated quantities from other coupled subsystems. On top of the divided state-space estimation strategy, the authors also consider the possibility of running the subsystems separately. Combining these two ideas, a few variants of the EnKF are derived. The introduction of these variants is mainly inspired by the current status and challenges in coupled data assimilation problems and thus might be of interest from a practical point of view. Numerical experiments with a multiscale Lorenz 96 model are conducted to evaluate the performance of these variants against that of the conventional EnKF. In addition, specific for coupled data assimilation problems, two prototypes of extensions of the presented methods are also developed in order to achieve a trade-off between efficiency and accuracy.

1. Introduction

This work considers the data assimilation problem in coupled systems that consist of two subsystems. Examples in this aspect include, for instance, coupled ocean–atmosphere models (e.g., Russell et al. 1995), marine ecosystem models coupling physics and biology (e.g., Petihakis et al. 2009), coupled flow, and (contaminant) transport models (e.g., Dawson et al. 2004), to name a few.

In principle, data assimilation in coupled systems can be tackled by concatenating the states of the subsystems into one augmented state and treating the whole coupled system as a single dynamical system. After augmentation,

a conventional data assimilation method, such as the ensemble Kalman filter (EnKF), can be directly applied. In this work, we present a divided state-space estimation strategy in the context of ensemble Kalman filtering. Instead of directly applying the update formulae in the conventional EnKF, we consider the possibility to express the update formulae in terms of some quantities with respect to the subsystems themselves. In doing so, the updated formulae in the divided estimation framework introduce some extra "cross terms" to account for the effect of coupling between the subsystems.

The divided estimation framework is derived based on the joint estimation one; hence, in principle these two approaches are mathematically equivalent. The main purpose of this work is to investigate the possibility of using the divided estimation strategy as an alternative to its joint counterpart. Whenever convenient, we would advocate the use of the joint estimation strategy, since it is conceptually more straightforward. However, there might still be some aspects in which the divided estimation strategy may appear more attractive, for example, in terms of the flexibility of implementation in large-scale applications, to be further discussed later.

Corresponding author address: Ibrahim Hoteit, 4700 King Abdullah University of Science and Technology, Al-Khwarizmi Building (1), Room 4415, Thuwal 23955-6900, Saudi Arabia.
E-mail: ibrahim.hoteit@kaust.edu.sa

This work is organized as follows: Section 2 outlines the filtering step of the EnKF in the joint and divided estimation frameworks. Section 3 conducts numerical experiments with a multiscale Lorenz 96 model and verifies that the joint and divided estimation frameworks have close performance under the same conditions. Section 4 investigates two extensions of the divided estimation framework that aim to achieve a certain trade-off between computational efficiency and accuracy. Finally, section 5 concludes the work and discusses some potential future developments.

2. Joint and divided estimation strategies with the EnKF

In the literature there are many variants of the EnKF (see, e.g., Anderson 2001; Bishop et al. 2001; Evensen 1994; Burgers et al. 1998; Hoteit et al. 2002; Luo and Moroz 2009; Tippett et al. 2003; Whitaker and Hamill 2002). In this work, we use the ensemble transform Kalman filter (ETKF; Bishop et al. 2001) for illustration. The extension to other filters can be done in a similar way. The joint and divided estimation strategies mainly differ at the filtering step, which is thus our focus hereafter. For ease of notation, we drop the time indices of all involved quantities.

Suppose that the state vectors in the coupled subsystems are η and ξ, respectively, and the corresponding observation subsystems are given by $\mathbf{y}_\eta = \mathcal{H}_\eta \boldsymbol{\eta} + \mathbf{u}_\eta$ and $\mathbf{y}_\xi = \mathcal{H}_\xi \boldsymbol{\xi} + \mathbf{u}_\xi$, where $\mathcal{H}_\eta$ and $\mathcal{H}_\xi$ are some linear observation operators,[1] and $\mathbf{u}_\eta$ and $\mathbf{u}_\xi$ are the corresponding observation noise with zero means and covariances $\mathbf{R}_\eta$ and $\mathbf{R}_\xi$, respectively. In practice, it is possible that one of the subsystems (e.g., ξ) may not be observed. In this case, to overcome the technical problem in describing the unknown observation operator (e.g., $\mathcal{H}_\xi$), one can set the associated covariance matrix $\mathbf{R}_\xi$ of $\mathbf{y}_\xi$ to $+\infty$ so that $\mathbf{y}_\xi$ does not affect the update (Jazwinski 1970, p. 219). For convenience of discussion, we denote the dimensions of the vectors $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\mathbf{x}$, $\mathbf{y}_\eta$, $\mathbf{y}_\xi$, and $\mathbf{y}$ by m_η, m_ξ, m_x, m_η^{obv}, m_ξ^{obv}, and m_y, respectively, such that $m_\eta + m_\xi = m_x$ and $m_\eta^{\text{obv}} + m_\xi^{\text{obv}} = m_y$.

In the above setting we have assumed that the observation operators $\mathcal{H}_\eta$ and $\mathcal{H}_\xi$ for different subsystems are "separable," in the sense that the observation (say $\mathbf{y}_\eta$) of each subsystem only depends on the corresponding subsystem state (say $\boldsymbol{\eta}$). In some situations,

however, the observation with respect to one subsystem may depend on the state variables of both subsystems. In such cases, one may introduce a certain transform to the observation system augmented by the observations with respect to the subsystems [see Eq. (1)], so that the resulting augmented observation system (after the transform) has a diagonal or block diagonal observation operator and thus becomes separable.

One can concatenate the above observation subsystems and obtain

$$\mathbf{y} = \mathcal{H}\mathbf{x} + \mathbf{u}, \tag{1}$$

where $\mathcal{H}\mathbf{x} \equiv [(\mathcal{H}_\eta \boldsymbol{\eta})^{\text{T}}, (\mathcal{H}_\xi \boldsymbol{\xi})^{\text{T}}]^{\text{T}}$, and $\mathbf{u} = [\mathbf{u}_\eta^{\text{T}}, \mathbf{u}_\xi^{\text{T}}]^{\text{T}}$ is the augmented observation noise with zero mean and covariance $\mathbf{R}$. Here,

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_\eta & \mathbf{R}_{\eta\xi} \\ \mathbf{R}_{\eta\xi}^{\text{T}} & \mathbf{R}_\xi \end{pmatrix}, \tag{2}$$

with $\mathbf{R}_{\eta\xi}$ being the cross covariance between $\mathbf{u}_\eta$ and $\mathbf{u}_\xi$. Throughout this work, we assume the observation noise $\mathbf{u}_\eta$ and $\mathbf{u}_\xi$ are uncorrelated, such that $\mathbf{R}_{\eta\xi} = \mathbf{0}$. If, in addition, both $\mathbf{R}_\eta$ and $\mathbf{R}_\xi$ are diagonal, then the observation $\mathbf{y}$ can be assimilated serially through some scalar update formulae (Anderson 2003). For our deduction, though, we only need to assume that $\mathbf{R}$ is a block diagonal matrix. If this is not the case, that is, $\mathbf{R}_{\eta\xi} \neq \mathbf{0}$, one can still obtain results similar to those presented below (though in somewhat more complicated forms), following a procedure similar to the derivation in the appendix.

Let $\mathbf{X}^b = \{\mathbf{x}_i^b : \mathbf{x}_i^b = [(\boldsymbol{\eta}_i^b)^{\text{T}}, (\boldsymbol{\xi}_i^b)^{\text{T}}]^{\text{T}}, i = 1, \ldots, n\}$ be an n-member background ensemble consisting of the subsystem components $\boldsymbol{\Phi}^b \equiv \{\boldsymbol{\eta}_i^b, i = 1, \ldots, n\}$ and $\boldsymbol{\Xi}^b \equiv \{\boldsymbol{\xi}_i^b, i = 1, \ldots, n\}$. In addition, let

$$\bar{\mathbf{x}}^b = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^b, \tag{3a}$$

$$\mathbf{S}^b = \frac{1}{\sqrt{n-1}} [\mathbf{x}_1^b - \bar{\mathbf{x}}^b, \ldots, \mathbf{x}_n^b - \bar{\mathbf{x}}^b], \tag{3b}$$

where $\bar{\mathbf{x}}^b$ and $\mathbf{S}^b$ are the sample mean and a square root matrix of the sample covariance of $\mathbf{X}^b$, respectively. Here, $\bar{\mathbf{x}}^b$ consists of two components, $\bar{\boldsymbol{\eta}}^b$ and $\bar{\boldsymbol{\xi}}^b$, which are the sample means of the ensembles of $\boldsymbol{\Phi}^b$ and $\boldsymbol{\Xi}^b$, respectively. On the other hand, define

$$\mathbf{S}_\eta^b = \frac{1}{\sqrt{n-1}} [\boldsymbol{\eta}_1^b - \bar{\boldsymbol{\eta}}^b, \ldots, \boldsymbol{\eta}_n^b - \bar{\boldsymbol{\eta}}^b],$$

$$\mathbf{S}_\xi^b = \frac{1}{\sqrt{n-1}} [\boldsymbol{\xi}_1^b - \bar{\boldsymbol{\xi}}^b, \ldots, \boldsymbol{\xi}_n^b - \bar{\boldsymbol{\xi}}^b], \tag{4}$$

---

[1] In cases of nonlinear observation operators, one may either approximate them by some linear ones or adopt more sophisticated assimilation schemes (see, e.g., Hoteit et al. 2012; Luo et al. 2010; Luo and Hoteit 2014a; Van Leeuwen 2009; Zupanski 2005).

then $\mathbf{S}^b = [(\mathbf{S}_\eta^b)^\mathrm{T}(\mathbf{S}_\xi^b)^\mathrm{T}]^\mathrm{T}$. Furthermore, let $\mathbf{Y}^b \equiv \{\mathbf{y}_i^b: \mathbf{y}_i^b = \mathcal{H}\mathbf{x}_i^b, i = 1, \ldots, n\}$ be the ensemble of forecasts of the projection of $\mathbf{X}^b$ onto the observation space (projection ensemble for short), then one can construct an $m_y \times n$ matrix (projection matrix for short):

$$\mathbf{S}^h = \frac{1}{\sqrt{n-1}}[\mathbf{y}_1^b - \overline{\mathbf{y}}^b, \ldots, \mathbf{y}_n^b - \overline{\mathbf{y}}^b], \qquad (5)$$

where

$$\overline{\mathbf{y}}^b = \frac{1}{n}\sum_{i=1}^n \mathbf{y}_i^b. \qquad (6)$$

Similarly, one can also decompose the projection ensemble $\mathbf{Y}^b = \{\mathbf{y}_i^b: \mathbf{y}_i^b = \mathcal{H}\mathbf{x}_i^b, i = 1, \ldots, n\}$ into two parts, $\mathbf{\Phi}^\mathrm{obv} \equiv \{\mathbf{y}_{\eta,i}^b: \mathbf{y}_{\eta,i}^b = \mathcal{H}_\eta \eta_i^b, i = 1, \ldots, n\}$ and $\mathbf{\Xi}^\mathrm{obv} \equiv \{\mathbf{y}_{\xi,i}^b: \mathbf{y}_{\xi,i}^b = \mathcal{H}_\xi \xi_i^b, i = 1, \ldots, n\}$, which satisfy $\mathbf{y}_i^b = [(\mathbf{y}_{\eta,i}^b)^\mathrm{T}(\mathbf{y}_{\xi,i}^b)^\mathrm{T}]^\mathrm{T}$ for $i = 1, \ldots, n$. Let the sample means of $\mathbf{\Phi}^\mathrm{obv}$ and $\mathbf{\Xi}^\mathrm{obv}$ be $\overline{\mathbf{y}}_\eta^b$ and $\overline{\mathbf{y}}_\xi^b$, respectively, such that $\overline{\mathbf{y}}^b = [(\overline{\mathbf{y}}_\eta^b)^\mathrm{T}(\overline{\mathbf{y}}_\xi^b)^\mathrm{T}]^\mathrm{T}$, then the projection matrix in Eq. (5) can also be decomposed as $\mathbf{S}^h = [(\mathbf{S}_\eta^h)^\mathrm{T}(\mathbf{S}_\xi^h)^\mathrm{T}]^\mathrm{T}$, where

$$\mathbf{S}_\eta^h = \frac{1}{\sqrt{n-1}}[\mathbf{y}_{\eta,1}^b - \overline{\mathbf{y}}_\eta^b, \ldots, \mathbf{y}_{\eta,n}^b - \overline{\mathbf{y}}_\eta^b],$$

$$\mathbf{S}_\xi^h = \frac{1}{\sqrt{n-1}}[\mathbf{y}_{\xi,1}^b - \overline{\mathbf{y}}_\xi^b, \ldots, \mathbf{y}_{\xi,n}^b - \overline{\mathbf{y}}_\xi^b]. \qquad (7)$$

a. Implementation of the ETKF in the joint estimation framework

In the joint estimation framework, the filtering step of the ETKF is given by

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}(\mathbf{y} - \mathcal{H}\overline{\mathbf{x}}^b), \qquad (8a)$$

$$\mathbf{S}^a = \mathbf{S}^b \mathbf{T}_{n-1} \mathbf{U}, \qquad (8b)$$

$$\mathbf{K} = \mathbf{S}^b(\mathbf{S}^h)^\mathrm{T}[\mathbf{S}^h(\mathbf{S}^h)^\mathrm{T} + \mathbf{R}]^{-1}. \qquad (8c)$$

In Eq. (8), $\mathbf{K}$ is the Kalman gain; $\mathbf{T}_{n-1}$ is the $n \times (n-1)$ transform matrix. Roughly speaking, $\mathbf{T}_{n-1}$ is an approximate square root of the matrix $\mathbf{\Lambda} = [\mathbf{I}_n + (\mathbf{S}^h)^\mathrm{T}\mathbf{R}^{-1}\mathbf{S}^h]^{-1}$ (with $\mathbf{I}_n$ being the n-dimensional identity matrix) and is constructed based on the $(n-1)$ leading eigenvalues of $\mathbf{\Lambda}$ and the associated eigenvectors (see Wang et al. 2004); the $(n-1) \times n$ matrix $\mathbf{U}$ (called centering matrix) satisfies $\mathbf{U}(\mathbf{U})^\mathrm{T} = \mathbf{I}_{n-1}$ and $\mathbf{U}\mathbf{1}_n = 0$ (Livings et al. 2008; Wang et al. 2004), where $\mathbf{1}_n$ is an n-dimensional vector with all its elements being 1. Readers are referred to Hoteit et al. (2002) and Wang et al. (2004) for the construction of such a centering matrix. Also note that it can be more convenient to use the square root update formula $\mathbf{S}^a = \tilde{\mathbf{T}}\mathbf{S}^b$,

with the transform matrix $\tilde{\mathbf{T}}$ in front of $\mathbf{S}^b$, when the ensemble size is larger than the dimension of the observation space (Posselt and Bishop 2012).

With $\overline{\mathbf{x}}^a$ and $\mathbf{S}^a$, the analysis ensemble $\mathbf{X}^a \equiv \{\mathbf{x}_i^a, i = 1, \ldots, n\}$ is generated by

$$\mathbf{x}_i^a = \overline{\mathbf{x}}^a + \sqrt{n-1}(\mathbf{S}^a)_i, \quad \text{for} \quad i = 1, \ldots, n, \qquad (9)$$

where $(\mathbf{S}^a)_i$ denotes the ith column of $\mathbf{S}^a$. Propagating $\mathbf{X}^a$ forward, one obtains a background ensemble at the next time step and a new assimilation cycle can begin.

b. Implementation of the ETKF in the divided estimation framework

In the divided estimation framework, we express all the quantities in the ETKF, for example, the mean, the square root matrices, and the Kalman gain, in terms of some quantities with respect to the subsystems, such that the divided estimation framework is mathematically equivalent to its joint estimation counterpart. In doing so, the mean update formulae of the ETKF in the divided estimation framework are given by

$$\overline{\eta}^a = \overline{\eta}^b + \mathbf{K}_{11}(\mathbf{y}_\eta - \mathcal{H}_\eta \overline{\eta}^b) + \mathbf{K}_{12}(\mathbf{y}_\xi - \mathcal{H}_\xi \overline{\xi}^b), \qquad (10a)$$

$$\overline{\xi}^a = \overline{\xi}^b + \mathbf{K}_{21}(\mathbf{y}_\eta - \mathcal{H}_\eta \overline{\eta}^b) + \mathbf{K}_{22}(\mathbf{y}_\xi - \mathcal{H}_\xi \overline{\xi}^b), \qquad (10b)$$

where

$$\mathbf{K}_{11} = \mathbf{S}_\eta^b \mathbf{T}_\xi (\mathbf{S}_\eta^h \mathbf{T}_\xi)^\mathrm{T}[(\mathbf{S}_\eta^h \mathbf{T}_\xi)(\mathbf{S}_\eta^h \mathbf{T}_\xi)^\mathrm{T} + \mathbf{R}_\eta]^{-1}, \qquad (11a)$$

$$\mathbf{K}_{12} = \mathbf{S}_\eta^b \mathbf{T}_\eta (\mathbf{S}_\xi^h \mathbf{T}_\eta)^\mathrm{T}[(\mathbf{S}_\xi^h \mathbf{T}_\eta)(\mathbf{S}_\xi^h \mathbf{T}_\eta)^\mathrm{T} + \mathbf{R}_\xi]^{-1}, \qquad (11b)$$

$$\mathbf{K}_{21} = \mathbf{S}_\xi^b \mathbf{T}_\xi (\mathbf{S}_\eta^h \mathbf{T}_\xi)^\mathrm{T}[(\mathbf{S}_\eta^h \mathbf{T}_\xi)(\mathbf{S}_\eta^h \mathbf{T}_\xi)^\mathrm{T} + \mathbf{R}_\eta]^{-1}, \qquad (11c)$$

$$\mathbf{K}_{22} = \mathbf{S}_\xi^b \mathbf{T}_\eta (\mathbf{S}_\xi^h \mathbf{T}_\eta)^\mathrm{T}[(\mathbf{S}_\xi^h \mathbf{T}_\eta)(\mathbf{S}_\xi^h \mathbf{T}_\eta)^\mathrm{T} + \mathbf{R}_\xi]^{-1}, \qquad (11d)$$

where $\mathbf{T}_\eta$ and $\mathbf{T}_\xi$ are some square root matrices of $[\mathbf{I} + (\mathbf{S}_\eta^h)^\mathrm{T}\mathbf{R}_\eta^{-1}\mathbf{S}_\eta^h]^{-1}$ and $[\mathbf{I} + (\mathbf{S}_\xi^h)^\mathrm{T}\mathbf{R}_\xi^{-1}\mathbf{S}_\xi^h]^{-1}$, respectively. The derivation of the above formulae is given in the appendix.

Based on Eq. (8b), the derivation of the square root update formulae in the divided estimation framework is relatively straightforward. Using the assumption $\mathbf{R} = \mathrm{diag}(\mathbf{R}_\eta, \mathbf{R}_\xi)$, one has $(\mathbf{S}^h)^\mathrm{T}\mathbf{R}^{-1}\mathbf{S}^h = (\mathbf{S}_\eta^h)^\mathrm{T}\mathbf{R}_\eta^{-1}\mathbf{S}_\eta^h + (\mathbf{S}_\xi^h)^\mathrm{T}\mathbf{R}_\xi^{-1}\mathbf{S}_\xi^h$ expressed in terms of the subsystem quantities. Therefore, the transform matrix $\mathbf{T}_{n-1}$ is now constructed based on the $(n-1)$ leading eigenvalues and the corresponding eigenvectors of $[\mathbf{I}_n + (\mathbf{S}_\eta^h)^\mathrm{T}\mathbf{R}_\eta^{-1}\mathbf{S}_\eta^h + (\mathbf{S}_\xi^h)^\mathrm{T}\mathbf{R}_\xi^{-1}\mathbf{S}_\xi^h]^{-1}$, and the square root update formulae become

$$\mathbf{S}_\eta^a = \mathbf{S}_\eta^b \mathbf{T}_{n-1}\mathbf{U}, \qquad (12a)$$

$$\mathbf{S}_\xi^a = \mathbf{S}_\xi^b \mathbf{T}_{n-1}\mathbf{U}, \qquad (12b)$$

where $\mathbf{U}$ is the same $(n-1) \times n$ centering matrix as previously discussed.

Accordingly, the analysis ensembles $\mathbf{\Phi}^a \equiv \{\eta_i^a, i = 1, \ldots, n\}$ and $\mathbf{\Xi}^a \equiv \{\xi_i^a, i = 1, \ldots, n\}$ are obtained from

$$\eta_i^a = \overline{\eta}^a + \sqrt{n-1}(\mathbf{S}_\eta^a)_i, \quad \text{for} \quad i = 1, \ldots, n, \qquad (13a)$$

$$\xi_i^a = \overline{\xi}^a + \sqrt{n-1}(\mathbf{S}_\xi^a)_i, \quad \text{for} \quad i = 1, \ldots, n. \qquad (13b)$$

Again, by propagating these two ensembles forward through the individual subsystems, one obtains the background ensembles for the next assimilation cycle.

The mean update formulae Eqs. (10a) and (10b) in the divided estimation framework are similar to that in Eq. (8a). However, they also exhibit clear differences. For instance, the correction terms in the divided estimation framework, say $\mathbf{K}_{11}(\mathbf{y}_\eta - \mathcal{H}_\eta\overline{\eta}^b)$ and $\mathbf{K}_{12}(\mathbf{y}_\xi - \mathcal{H}_\xi\overline{\xi}^b)$ in Eq. (10a), are associated with some gain matrices, say $\mathbf{K}_{11}$ and $\mathbf{K}_{12}$, that bear different forms from the Kalman gain $\mathbf{K}$ in Eq. (8c). There are certain similarities among these gain matrices as well. For instance, if one replaces $\mathbf{S}_\eta^b\mathbf{T}_\xi$ by $\mathbf{S}_\eta^b$ and $\mathbf{S}_\eta^h\mathbf{T}_\xi$ by $\mathbf{S}_\eta^h$, then the gain matrix $\mathbf{K}_{11}$ reduces to the Kalman gain with respect to the subsystem η. In this sense, the presence of the term $\mathbf{T}_\xi$ in $\mathbf{K}_{11}$ reflects the coupling between the subsystems η and ξ. Similar results can also be found for the other gain matrices $\mathbf{K}_{12}$, $\mathbf{K}_{21}$, and $\mathbf{K}_{22}$. The square root update formula, say Eq. (12a) for the subsystem η, has its transform matrix $\mathbf{T}_{n-1}$ as an approximate square root matrix of $[\mathbf{I} + (\mathbf{S}_\eta^h)^T\mathbf{R}_\eta^{-1}\mathbf{S}_\eta^h + (\mathbf{S}_\xi^h)^T\mathbf{R}_\xi^{-1}\mathbf{S}_\xi^h]^{-1}$, rather than $[\mathbf{I} + (\mathbf{S}_\eta^h)^T\mathbf{R}_\eta^{-1}\mathbf{S}_\eta^h]^{-1}$. The extra term $(\mathbf{S}_\xi^h)^T\mathbf{R}_\xi^{-1}\mathbf{S}_\xi^h$ also represents the effect of coupling between the subsystems.

3. Numerical experiments

a. Experiment settings

We consider the data assimilation problem in a multiscale Lorenz 96 (ms-L96) model [Lorenz 2006, his Eqs. (3.2) and (3.3)], whose governing equations are given by

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b}\sum_{j=1}^{K} z_{j,i},$$

$$\frac{dz_{j,i}}{dt} = cbz_{j+1,i}(z_{j-1,i} - z_{j+2,i}) - cz_{j,i} + \frac{hc}{b}x_i, \qquad (14)$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, K$, and F, c, b, and h are constant parameters. The state variables x_i and $z_{j,i}$ are cyclic as in the Lorenz 96 model (Lorenz and Emanuel 1998). For instance, one has $x_{m+1} = x_1$; $x_0 = x_m$; $z_{K+1,i} = z_{1,i}$; $z_{0,i} = z_{K,I}$, and so on. In the experiments, we let $m = 40$, $K = 1$, $F = 8$, $c = b = 10$, and $h = 0.8$. This results in an 80-dimensional dynamical system with 40 x_i variables and 40 $z_{1,i}$ variables. In the divided estimation framework, the two subsystems consist of the ordinary differential equations (ODEs) starting with dx_i/dt and $dz_{1,i}/dt$, respectively, that is, x_i and $z_{1,i}$ play the roles of η and ξ in section 2. For convenience, we call the component x_i fast mode (in terms of the rate of state change) and $z_{1,i}$ slow mode. Figure 1 plots the time series of some state variables in the ms-L96 model.

The dynamical system Eq. (14) is numerically integrated using the fourth-order Runge–Kutta–Fehlberg (RKF) method (Fehlberg 1970), and the system states are collected every 0.05 time unit (for brevity we call it an integration step). In the experiments, we run the system forward in time for 1500 integration steps and discard the first 500 steps to avoid a spinup period. In both the joint and divided estimation frameworks, data assimilation starts from step 501 until step 1500. The trajectory during this period is considered as the truth. Synthetic observations are generated by adding Gaussian white noise (with zero mean and unit variance) to the fast mode state variables $x_1, x_5, x_9, \ldots, x_{37}$ and to the slow mode ones $z_{1,1}, z_{1,5}, z_{1,9}, \ldots, z_{1,37}$ (i.e., every four state variables and) every four integration steps. Therefore, observations are available at 250 out of 1000 integration steps, from 20 out of 80 state variables. For convenience, we relabel the integration step 501 as the first assimilation step. An initial background ensemble with 20 ensemble members is generated by drawing samples from the 80-dimensional multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_{80})$ and then adding these samples to the true state at the first assimilation cycle.

In the experiments below, we consider an extra possibility in which the integration of the subsystems x_i and $z_{1,i}$ in Eq. (14) is also carried out in a "divided" way. This is achieved by temporally treating variables (say $z_{1,i}$) as constant parameters in the subsystem (say x_i) during the integration and vice versa. Such a parameterization may incur extra numerical errors during the integration steps. Our main motivation to consider this option is, however, for its potential usefulness in data assimilation practices. For instance, it could be a fast—although crude, and likely not the best possible—way to combine Earth's subsystem (ocean, atmosphere, etc.) models independently developed by different research groups and hence increase the reusability of existing resources. However, it is worthwhile to stress that running the
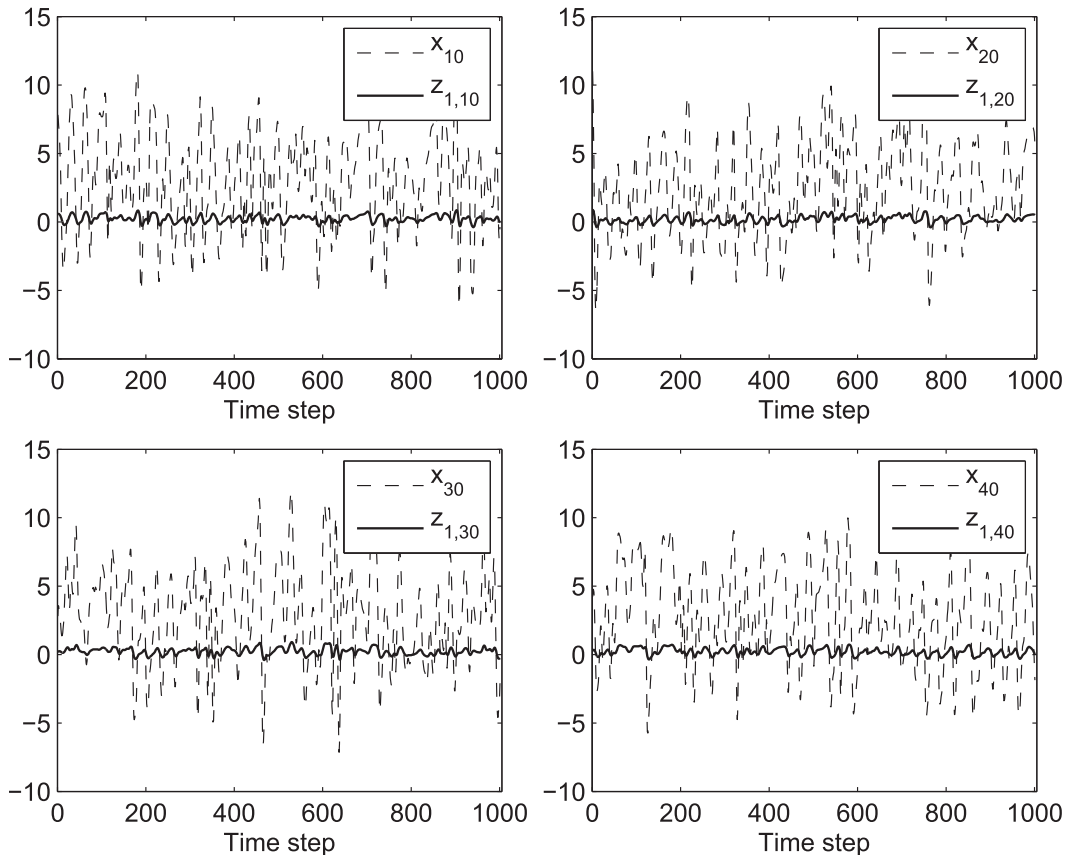
FIG. 1. Time series of some state variables in the multiscale Lorenz 96 model.

subsystems separately is not mandatory for the implementation of the divided estimation framework.

Therefore, in each experiment below we consider four possible scenarios, which differentiate from each other depending on whether they divide the dynamical system and/or the assimilation scheme. For convenience, we denote these scenarios by (DS-joint, DA-joint), (DS-joint, DA-divided), (DS-divided, DA-joint), and (DS-divided, DA-divided), respectively, where the abbreviations DS and DA stand for "dynamical system" and "data assimilation." Here, for instance, "DS-joint" means that the dynamical system is integrated as a whole, and "DA-divided" means that the divided estimation framework is adopted for data assimilation. Other terminologies are interpreted in a similar way.

For illustration, Fig. 2 outlines the main procedures in the scenario (DS-divided, DA-divided). Starting with an initial ensemble of the coupled system, we split the initial ensemble into two subensembles according to fast and slow modes and mark them by letters F and S, respectively. The subensemble F (S) acts as the input state vectors of the fast (slow) mode (denoted by solid arrow lines) and as the input "parameters" of the slow (fast) mode (denoted by

dotted arrow lines). With incoming observations, the background ensembles of the fast and slow modes are updated to their analysis counterparts as described in section 2b. Propagating the analysis ensembles forward, one starts a new assimilation cycle, and so on.

Below we compare the performance of the four scenarios through two sets of experiments. In the first set, we conduct the experiment in a plain setting, that is, without introducing covariance inflation (Anderson and Anderson 1999) or localization (Hamill et al. 2001) to the filter. In the second one, covariance inflation and localization are adopted and the details will be presented later. In all experiments, the different scenarios share the same truth, initial ensemble, and observations. For comparison, we use the root-mean-square error (RMSE) as a performance measure. For an m_x-dimensional system, the RMSE e_k of an analysis $\overline{\mathbf{x}}_k = (\overline{x}_{k,1}, \ldots, \overline{x}_{k,m_x})^{\mathrm{T}}$ with respect to the truth $\mathbf{x}_k = (x_{k,1}, \ldots, x_{k,m_x})^{\mathrm{T}}$ at time instant k is defined as

$$e_k = \frac{\|\overline{\mathbf{x}}_k - \mathbf{x}_k\|_2}{\sqrt{m_x}} = \sqrt{\frac{1}{m_x} \sum_{j=1}^{m_x} (\overline{x}_{k,j} - x_{k,j})^2}. \quad (15)$$
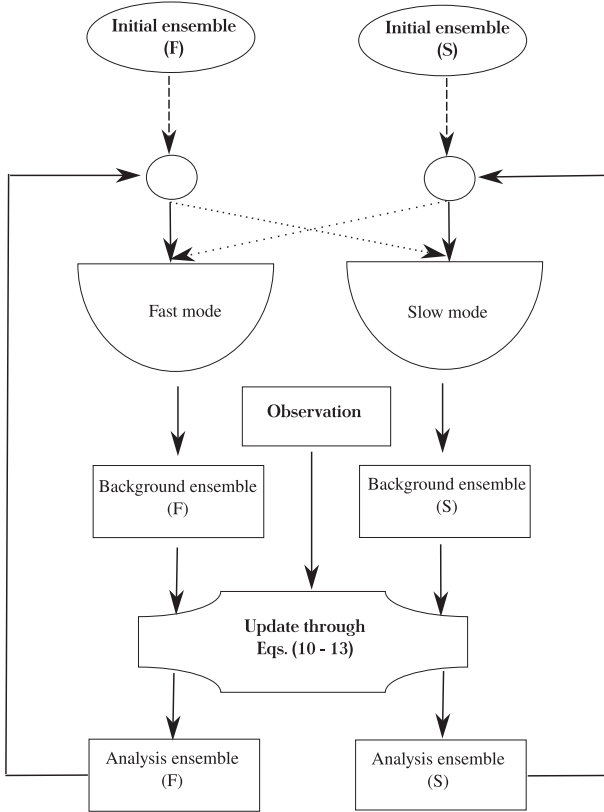
FIG. 2. Flowchart of the procedures in the (DS-divided, DA-divided) scenario in assimilating the multiscale Lorenz 96 system.
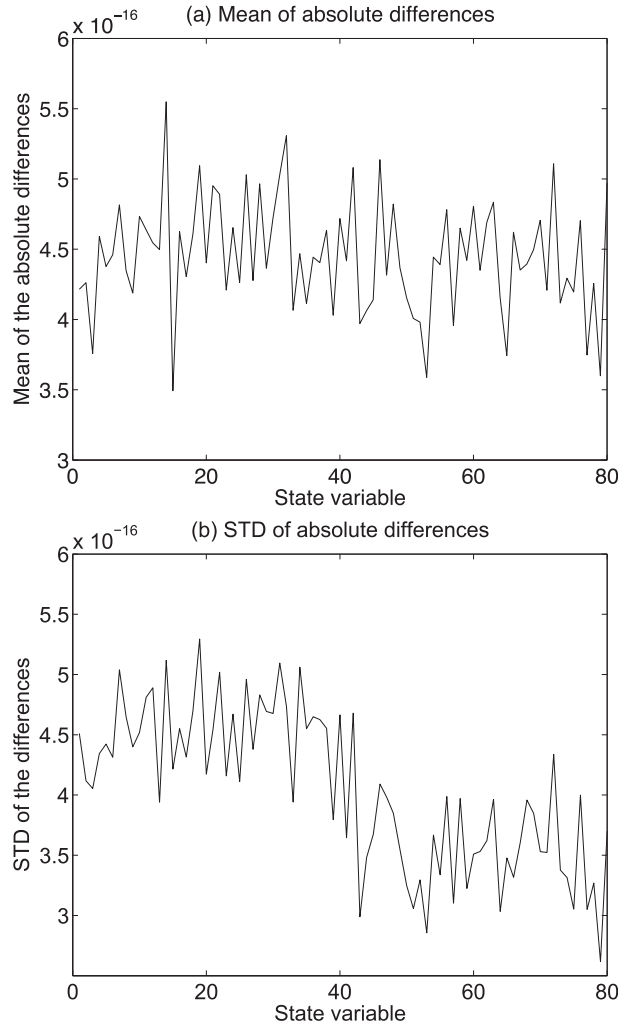

FIG. 3. A comparison of the analyses of joint and divided estimation frameworks with a single update. The experiment is repeated 100 times. In each repetition, the background ensemble and the observation are drawn at random (so that in general they will change over different repetitions), and in each repetition, the joint and divided estimation frameworks share the same background ensemble and observation. (a) Mean value (over 100 repetitions) of the absolute differences between the analyses of the joint and divided estimation frameworks in each state variable. (b) Corresponding STD of the absolute differences.

b. Experiment results

1) RESULTS WITH THE PLAIN SETTING

First, we investigate whether the joint and divided estimation frameworks yield the same results. To this end, we compare the analyses obtained in both methods by conducting a single update step using identical background ensemble and observations. The experiment is repeated 100 times, each time the background ensemble and observations are drawn at random so that in general they will change over different repetitions. Figure 3 shows that the mean and standard deviation (STD) of the differences (in absolute values) between the state variables of the analyses of both estimation frameworks are in the order of 10^{-16}. Our computations are carried out with MATLAB (version R2012a), in which the numerical precision of floating point numbers is 2.2204×10^{-16}. This indicates that the tiny differences reported in Fig. 3 mainly stem from the numerical precision in computations.

Figure 4 depicts the time series of the RMSEs of the estimates obtained in the four different scenarios, (DS-joint, DA-joint), (DS-joint, DA-divided), (DS-divided, DA-joint), and (DS-divided, DA-divided), with a longer time horizon. These four assimilation scenarios have identical initial background ensembles and observations. However, the background ensembles in these four scenarios may (gradually) deviate from each other at subsequent time instants because of the chaotic nature of the ms-L96 model and the extra parameterization errors in the DS-divided scenarios. Therefore, in Fig. 4 one can see that, in the DS-joint scenarios, the differences between the estimates from the joint (Fig. 4a) and
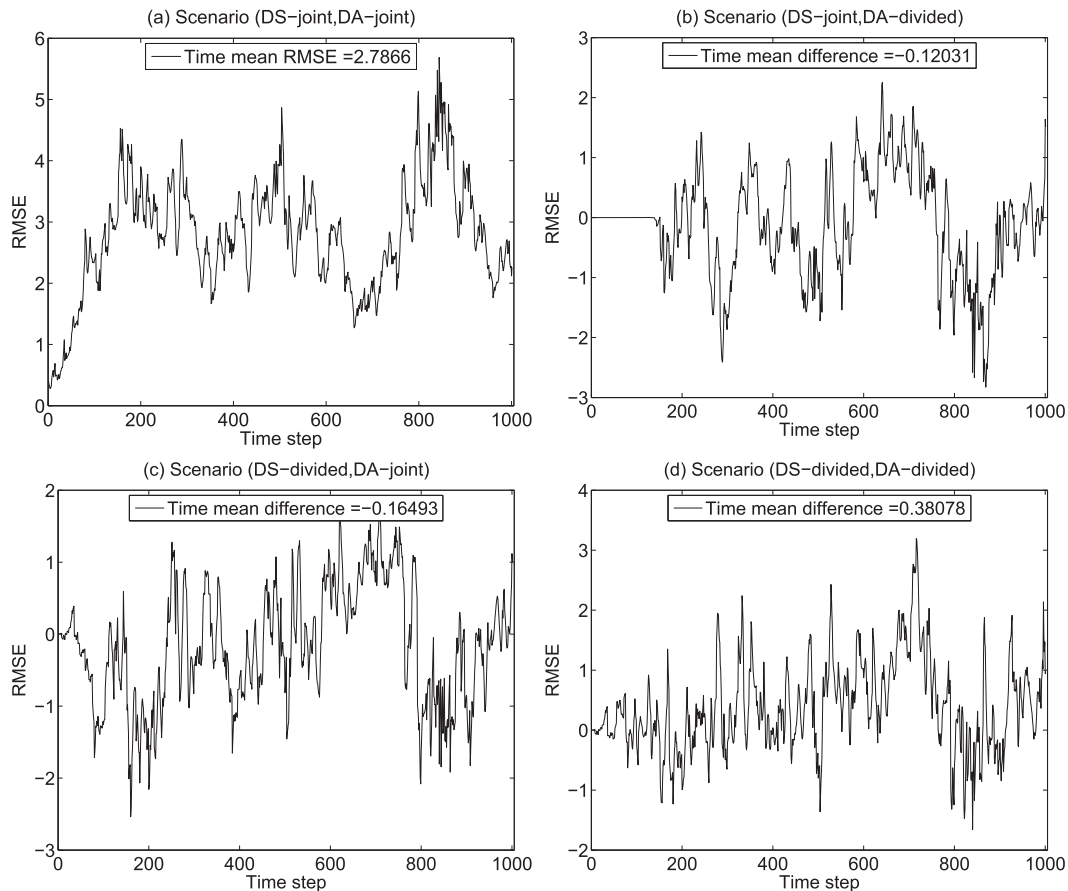
FIG. 4. (a)–(d) Time series of the RMSEs in four different scenarios. For comparison, in (b)–(d) we plot the differences of the RMSEs relative to those in (a) [obtained by subtracting the RMSEs in (b)–(d) from those in (a)]. In (b), the RMSEs overlap those in (a) up to around the first 130 integration steps. However, because of the chaotic nature of the ms-L96, tiny differences due to numerical precision are accumulated and amplified and eventually become noticeable. In (c) and (d), because of the extra numerical errors in the DS-divided scenarios, the RMSEs are indistinguishable from those in (a) up to the first few integration steps only, and become noticeably different afterward.

divided (Fig. 4b) estimation frameworks are nearly zero during the early assimilation period, but become more substantial over time. Meanwhile, in the DS-divided scenarios the estimates from either the joint (Fig. 4c) or the divided (Fig. 4d) estimation framework deviate from those in the (DS-joint, DA-joint) scenario (Fig. 4a) more quickly with the extra parameterization errors.

In terms of estimation accuracy, the time-mean RMSE in Fig. 4a is 2.7866. In contrast, the time-mean RMSEs in Figs. 4b–d are −0.1203 (lower), −0.1649 (lower), and +0.3808 (higher), respectively, relative to that in Fig. 4a. This seems to suggest that the extra numerical errors due to parameterization are not always harmful. For instance, the time-mean RMSE in Fig. 4c appears to be the lowest in these four tested scenarios. A possible explanation of this result is discussed later from the point of view of covariance inflation.

Because of the interactions of the forecast and update steps in assimilating the ms-L96 model, it is challenging to obtain an analytic description of the dynamics of the differences between the reference trajectory of the (DS-joint, DA-joint) scenario and those of the (DS-joint, DA-divided), (DS-divided, DA-joint), and (DS-divided, DA-divided) scenarios. For this reason, in what follows we adopt two statistical measures, namely, the boxplot (see the left column of Fig. 5) and the histogram (see the right column of Fig. 5), to characterize these differences.

A boxplot depicts a group of data through their quartiles. In this work, the boxplot is adopted to plot the differences at certain time instants. The differences are 80-dimensional vectors, obtained by subtracting the trajectory of the reference scenario (DS-joint, DA-joint) from those of the scenarios (DS-joint, DA-divided), (DS-divided, DA-joint), and (DS-divided, DA-divided) at
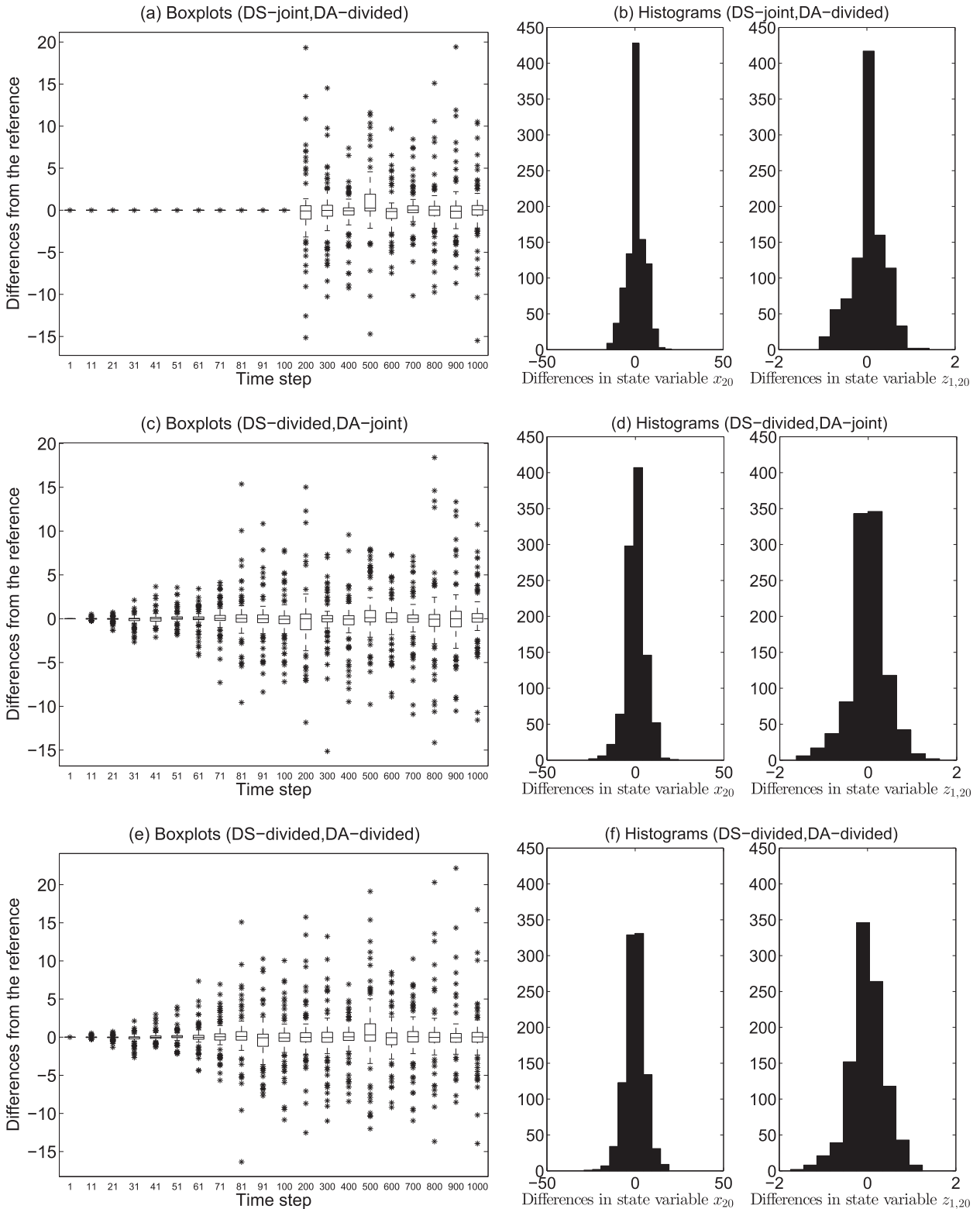
FIG. 5. (a),(c),(e) Boxplots and (b),(d),(f) histograms for the characterization of the trajectory differences of the (DS-joint, DA-divided), (DS-divided, DA-joint), and (DS-divided, DA-divided) scenarios from the reference trajectory in the (DS-joint, DA-joint) scenario.

some particular time instants. A boxplot is used here to indicate the spatial distribution of the 80 elements in a difference vector at a particular time instant. For ease of visualization, we only plot the boxes at time steps {1:10:91} and {100:100:1000}, where $v_{\text{ini}}:\Delta v:v_{\text{final}}$ stands for an array of scalars that grow from the initial value v_{ini} to the final one v_{final}, with an even increment Δv each time. Our boxplot setting follows the custom in MATLAB (version R2012a): on each box, the band inside the box denotes the median, the bottom and top of the box represent the 25th and 75th percentiles, the ends of the whiskers indicate the extension of the data that are considered nonoutliers, while outliers are marked individually as asterisks in Fig. 5. Note that, in the (DS-joint, DA-divided) scenario, because the differences from the reference trajectory are very tiny at the early assimilation stage, the boxes appear to collapse during this period (e.g., from time steps 1 to 91), which is consistent with the results in Fig. 4b. As time moves forward, the trajectory of the (DS-joint, DA-divided) scenario gradually deviates from the reference. Therefore, as indicated in Fig. 5a, the spreads of the differences become larger from time step 200 on, compared to those at earlier time steps. In addition, more outliers (asterisks) are seen after time step 200, while the medians of the differences appear to remain close to zero at all time steps. Similar phenomena can also be observed in the (DS-divided, DA-joint) and (DS-divided, DA-divided) scenarios, except that the periods in which the boxplots collapse are much shorter compared to that in the (DS-joint, DA-divided) scenario, which is also consistent with the results in Fig. 4.

The histogram is also used here to depict the distribution of an element in a difference vector during the whole assimilation time window. In the right column of Fig. 5 we show the twentieth and sixtieth elements, which correspond to the trajectory differences in the state variables x_{20} and $z_{1,20}$, respectively, in the scenarios (DS-joint, DA-divided), (DS-divided, DA-joint), and (DS-divided, DA-divided). In the (DS-joint, DA-divided) scenario, the histogram of the differences in state variable x_{20} appears to have a single peak at zero, while its support is inside the interval $[-15, 15]$. The histogram of the differences in state variable $z_{1,20}$ also has a single peak at zero, but its support is narrower, being inside the interval $[-1.5, 1.5]$ instead. Similar phenomena are also observed in the (DS-divided, DA-joint) and (DS-divided, DA-divided) scenarios, although the heights of the peaks tend to be lower, and the corresponding supports tend to be wider.

Overall, the results in Figs. 4 and 5 seem to suggest that the trajectories of the (DS-joint, DA-divided), (DS-divided, DA-joint), and (DS-divided, DA-divided)

scenarios tend to oscillate around the reference trajectory of the (DS-joint, DA-joint) scenario, although they may also substantially deviate from the reference one at many time instants.

2) RESULTS WITH BOTH COVARIANCE INFLATION AND LOCALIZATION

Covariance inflation (Anderson and Anderson 1999) and localization (Hamill et al. 2001) are two important auxiliary techniques that can be used to improve the performance of an EnKF. Since the EnKF is a Monte Carlo implementation of the Kalman filter, when the ensemble size is relatively small, certain issues may arise, including, for instance, systematic underestimation of the variances of state variables, overestimation of the correlations of different state variables, and rank deficiency in the sample error covariance matrix.

Covariance inflation (Anderson and Anderson 1999) is introduced to tackle the variance underestimation problem by artificially increasing the sample error covariance to some extent. In relation to the results in the previous experiment, one possible explanation of the result there is that the extra numerical errors due to parameterization may have acted as some additive noise in the dynamical model, which is not always bad for a filter's performance. Indeed, as has been reported in some earlier works (e.g., Gordon et al. 1993; Hamill and Whitaker 2011), introducing some artificial noise to the dynamical model may improve filter performance. In the context of EnKF, this may be considered an alternative form of covariance inflation (Hamill and Whitaker 2011), which may enhance the robustness of the filter from the point of view of H_∞ filtering theory (Luo and Hoteit 2011; Altaf et al. 2013; Triantafyllou et al. 2013). One may also introduce artificial noise in a more sophisticated way, for example, through a certain nonlinear regression model, such that the statistical effect of the regression model mimics that of the dynamical model (Harlim et al. 2014).

How to optimally conduct covariance inflation is an ongoing research topic in the data assimilation community. Some recent developments include, for example, adaptive covariance inflation techniques (see, e.g., Anderson 2007, 2009) and covariance inflation from the point of view of residual nudging (Luo and Hoteit 2014b, 2013, 2012), among many others. For our purpose here, it appears sufficient to conduct covariance inflation by simply multiplying the analysis sample error covariance by a factor δ^2 ($\delta \geq 1$), as originally proposed in Anderson and Anderson (1999). The values of δ in the experiment are {1:0.05:1.3}.

Covariance localization (Hamill et al. 2001) is adopted to deal with the overestimation of the correlations and rank deficiency. In practice, different methods are
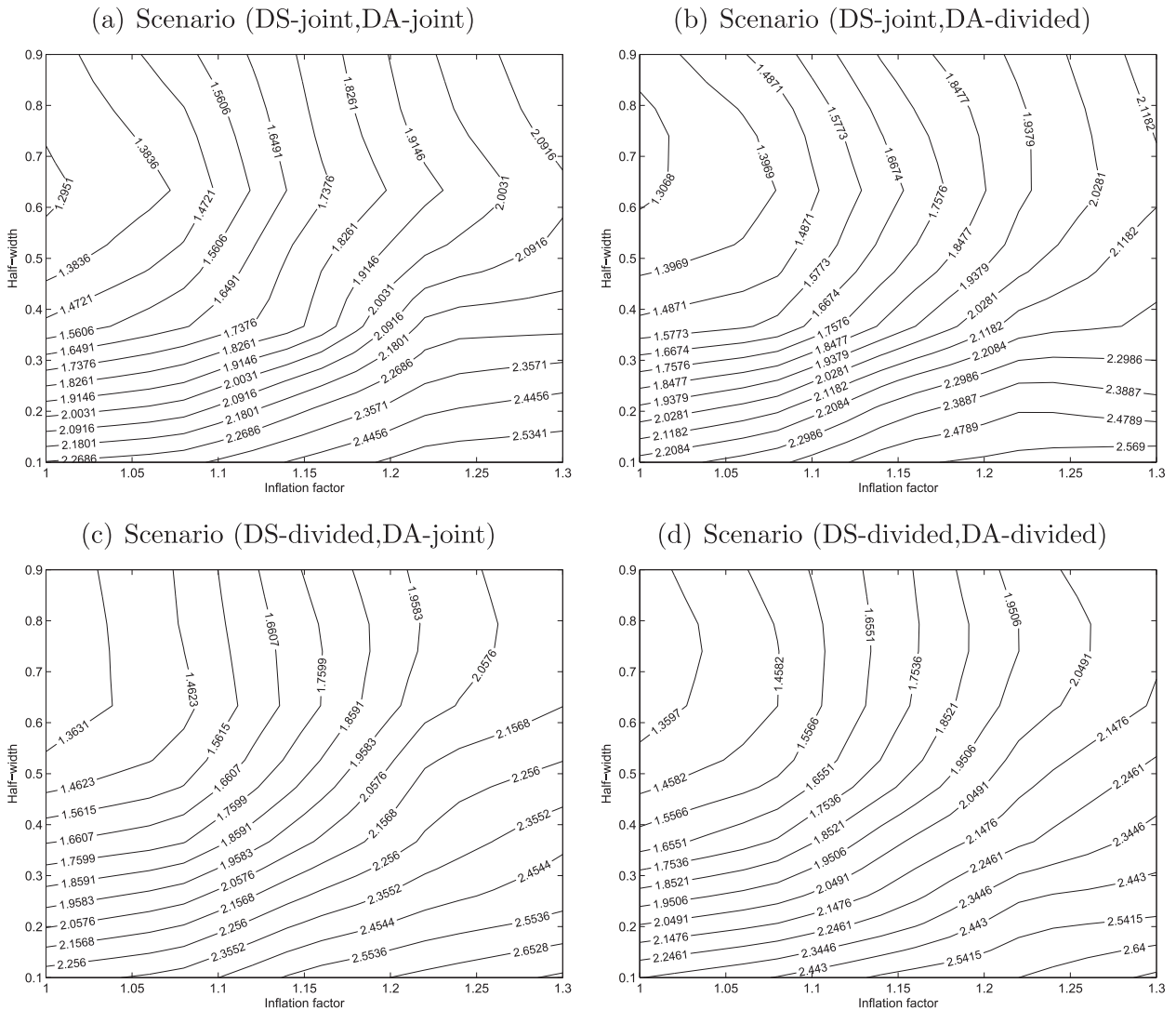
FIG. 6. Time-mean RMSEs of four different scenarios with both covariance inflation and localization adopted.

proposed to conduct localization (see, e.g., Anderson 2007, 2009; Clayton et al. 2013; Kuhl et al. 2013; Wang et al. 2007). In our experiments, localization is directly applied to the gain matrices. We assume that $z_{1,i}$ and x_i are located at the same grid point i. Covariance localization thus follows the settings in Anderson (2007), in which a parameter l_c, called the half-width (or length scale of localization), controls the degree of correlation tapering. We use the same half-width for the fast and slow components of the ms-L96 model, with l_c being chosen from the set {0.1:0.2:0.9}. In general, for both the joint and divided estimation frameworks, one may use different half-widths for different components (e.g., ocean and atmosphere) of a coupled system. In such circumstances, it could be more efficient to use an adaptive localization approach (see, e.g., Bishop and Hodyss 2007, 2009a,b, 2011).

We investigate the filter performance in the afore-mentioned four scenarios by combining different values of the inflation factor δ and the half-width l_c. The corresponding results, in terms of time-mean RMSEs (the averages of the RMSEs over the assimilation time window) are reported in Fig. 6. In the experiments, the filters' performance is improved in most of the cases, in comparison with the results in Fig. 4. In Fig. 6, the best filter performance is obtained with $l_c \approx 0.7$, while with localization, covariance inflation does not seem to help improve the estimation accuracy,[2] similar to the findings of Penny (2014). The above results, however, may strongly depend on the experimental settings. For

---

[2] When covariance localization is excluded, inflation may improve the filters' performance (results not shown).

instance, in the context of the hybrid local ETKF, Penny (2014) found that the best filter performance is achieved at relatively small l_c values (e.g., ≈0.2).

Figure 6 also indicates that, for a given model integration scenario (either DS-joint or DS-divided), the joint and divided estimation frameworks yield very close results. On the other hand, for a given estimation framework (either DA-joint or DA-divided), integrating the subsystems separately tends to deteriorate filter performance. In general, the performance deterioration is not severe, less than 10% in all cases with the same values of δ and l_c.

4. Two extensions from the practical point of view

In this section, we present two extensions of the aforementioned frameworks. These are largely motivated by the current status and challenges of conducting data assimilation in coupled ocean–atmosphere models (Bishop et al. 2012). These two extensions are illustrated within the (DS-divided, DA-divided) scenario. The extensions to the other scenarios can be implemented in a similar way.

a. Different ensemble sizes in the subsystems

Here, we consider the possibility of running the filter with different ensemble sizes in the fast and slow modes. This may be considered as an example in which one wants to gain certain computational efficiency by running fewer ensemble members in one of the subsystems, but possibly at the cost of certain loss of accuracy. To this end, let the ensemble sizes of the fast and slow modes be n_f and n_s, respectively. In the experiments, we consider four different cases, with ($n_f = 20$, $n_s = 20$), ($n_f = 20$, $n_s = 15$), ($n_f = 15$, $n_s = 20$), and ($n_f = 15$, $n_s = 15$) at the prediction step, and the targeted ensemble size is 20 for both modes at the filtering step. To apply the filter update formulae, the ensemble sizes of both modes should be equal. Therefore, dimension mismatch will arise when $n_f \neq n_s$. This issue is addressed through a conditional sampling scheme discussed in the online supplemental material.

In each of the above cases, we investigate the filter's performance when (i) neither covariance inflation nor covariance localization is applied (the plain setting) and (ii) both covariance inflation and covariance localization are adopted. In setting ii, the covariance inflation factor is 1.15 for both the fast and slow modes, and the half-width for covariance localization is 0.75.

Figure 7 plots the time series of the RMSEs for the above four different cases. In each case, when the filter is equipped with both covariance inflation and localization, its time-mean RMSE tends to be lower than that of the plain setting (with neither inflation nor localization). On the other hand, if one takes the case ($n_f = 20$, $n_s = 20$) with both covariance inflation and localization as the reference, then it is clear that reducing the ensemble size of either the fast or slow mode degrades the filter performance in terms of RMSE. Also, comparing Figs. 7b and 7c, one can see that reducing the ensemble size of the fast mode appears to have a larger (negative) impact than reducing the ensemble size of the slow one, which may be because the fast mode appears to dominate the dynamics of the ms-L96 model [see Fig. 8 later as well as the similar results in Hoteit and Pham (2004)]. On the other hand, comparing Figs. 7c and 7d, it seems better to simply reduce the ensemble sizes of both the fast and slow modes, in contrast to the case that reduces the ensemble size of the fast mode only. This may also be because the fast mode is the dominant part to the dynamics of the ms-L96 model; therefore, the extra errors due to the sampling scheme may be significant to the filter performance. However, a comparison between Figs. 7b and 7d suggests that if one only reduces the ensemble size of the slow mode, then the filter performance can be better than that resulting from reducing the ensemble sizes of both modes. Similar results are also observed with the plain setting, except that with the plain setting, the case ($n_f = 20$, $n_s = 15$) seems to perform slightly better than the one with ($n_f = 20$, $n_s = 20$).

b. Incorporating the ensemble optimal interpolation into the divided estimation framework

If one subsystem of the coupled model (e.g., the ocean in the coupled ocean–atmosphere model) exhibits relatively slow changes, then it may be reasonable to assume that this subsystem has a (almost) constant background covariance over a short assimilation time window (Hoteit et al. 2002).[3] As a result, optimal interpolation (OI; see, e.g., Cooper and Haines 1996) could be a reasonable assimilation scheme for such a slow-varying subsystem model because of its simplicity in implementation and significant savings in computational cost. The ensemble optimal interpolation (EnOI; see, e.g., Counillon and Bertino 2009) is an ensemble implementation of the OI scheme. It has an update step similar to that of the EnKF, but computes the associated background covariance (or square root matrix) based on a "historical" ensemble (Counillon and Bertino 2009).

---

[3] In the context of meteorological applications, the extension described here mainly targets short-term (e.g., subseasonal) time scales, while for seasonal or longer time scale applications (e.g., climate studies), the small-variation assumption (e.g., in the ocean component) may not be valid.
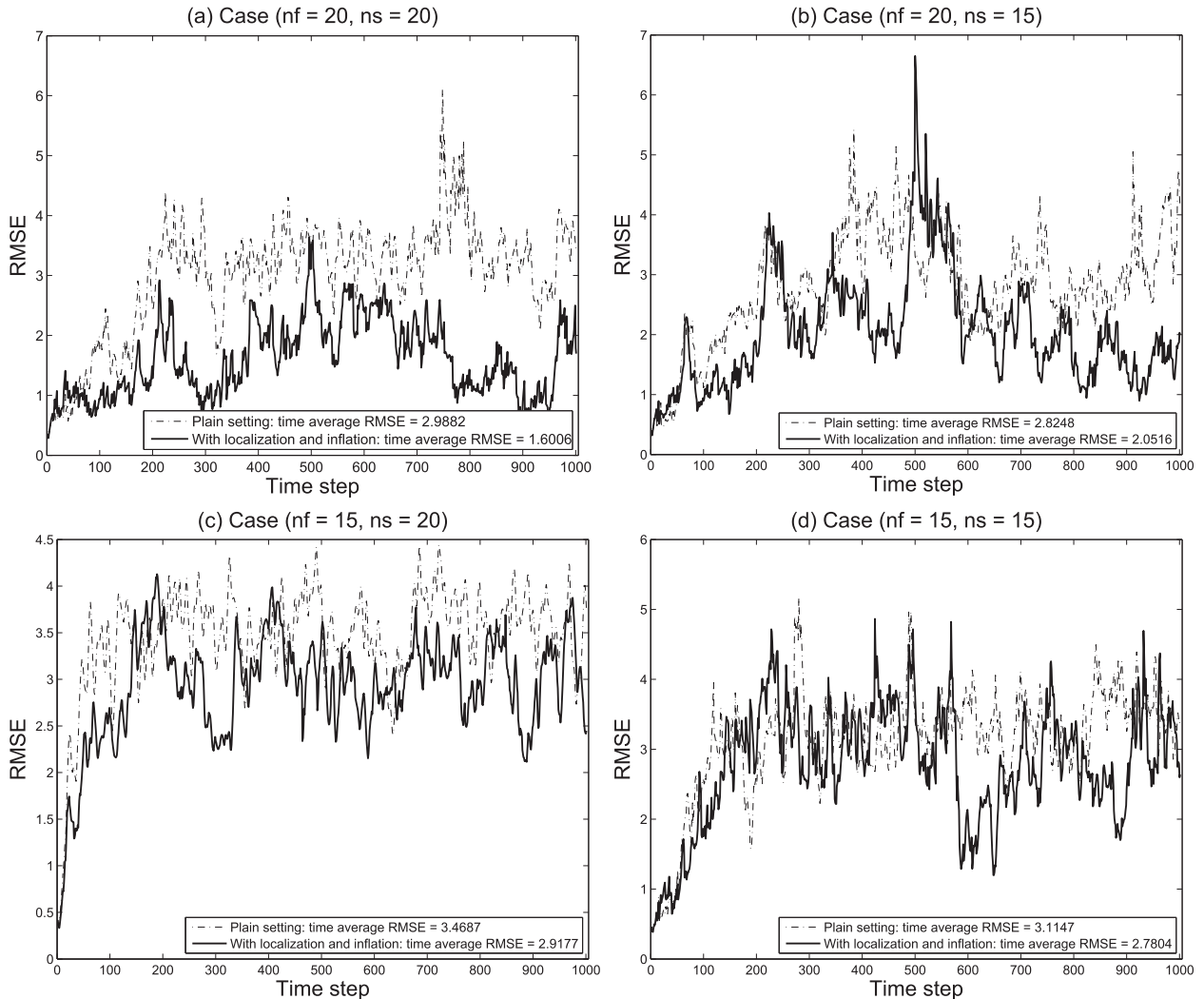
FIG. 7. Time series of the RMSEs of four cases with different ensemble sizes in the fast and slow modes.

At the prediction step, the EnOI only propagates the analysis mean forward to obtain a background mean at the next assimilation cycle. This is computationally much cheaper than propagating the whole analysis ensemble forward as in the EnKF and hence appears attractive for certain applications (e.g., oceanography; see Hoteit et al. 2002; Bishop et al. 2012).

Here, we consider the possibility to tailor the divided estimation framework so as to incorporate the EnOI into one of the subsystems. Such a modification is largely motivated by the current status and challenges of operational data assimilation in coupled ocean–atmosphere models, in which, because of the limitations in computational resource, one may use OI or three-dimensional variational data assimilation (3D-Var; or their ensemble implementations) for the ocean model and a more sophisticated scheme such as four-dimensional variational

data assimilation (4D-Var) or EnKF for the atmosphere model. Therefore, combining these different assimilation systems becomes a challenge in practice (Bishop et al. 2012).

In our investigation below, to incorporate the EnOI into the divided estimation framework, some modifications are introduced. (i) At the prediction step, the slow mode only propagates forward the analysis mean of the corresponding subensemble and uses the analysis mean with respect to the fast mode as the "parameters" in the numerical integrations of the slow mode. On the other hand, the fast mode propagates forward the corresponding analysis subensemble [updated through Eqs. (12) and (13)] and uses the update of the historical ensemble [also through Eqs. (12) and (13)] of the slow mode as the parameters in the numerical integrations of the fast mode. (ii) At the filtering step, the background
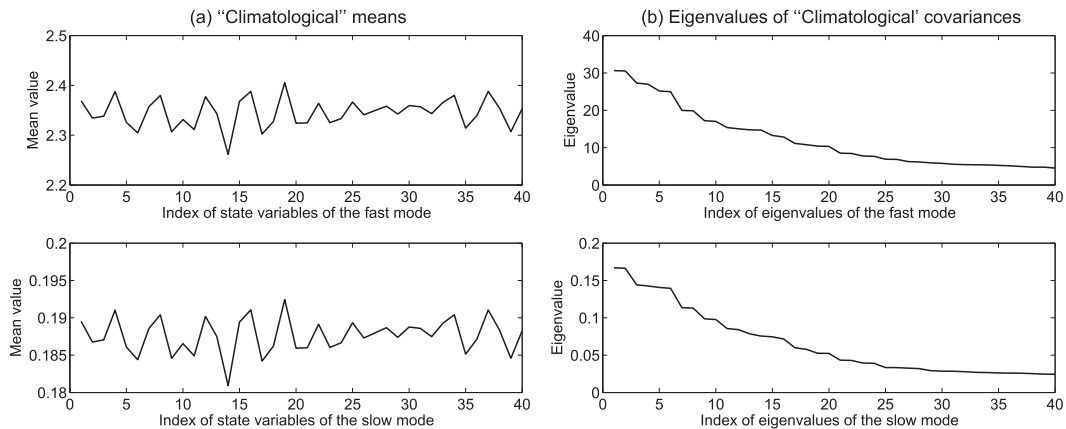
FIG. 8. Climatological means and eigenvalues of the climatological covariances of the fast and slow modes.

subensemble of the fast mode is the propagation of the analysis subensemble from the previous assimilation cycle, while the background subensemble of the slow mode is the historical ensemble generated by drawing a specified number of samples from a Gaussian distribution whose mean and covariance are equal to the "climatological" mean and covariance of the slow mode, respectively. This historical ensemble is produced once for all and does not change over the assimilation window. However, at each assimilation cycle, when a new observation is available, the historical ensemble is updated according to Eqs. (12) and (13) and is used as the parameters of the fast mode. In doing so, the cross covariance between the historical ensemble of the slow mode and the flow-dependent subensemble of the fast mode may not accurately represent the true correlations between both modes.

To generate the historical ensemble of the slow mode, we run the ms-L96 model forward in time for 100 000 integration steps, where the step size is 0.05. The climatological statistics are then taken as the temporal mean and covariance of the generated trajectory. Figure 8 shows the values of the climatological means and the eigenvalues of the climatological covariances of the fast and slow modes. These results suggest that the fast mode dominates the slow one in magnitudes, consistent with the results in Fig. 1.

In the experiments below, the ensemble sizes of the fast and slow modes are both 20. For distinction, hereafter we refer to the extended assimilation scheme with the EnOI as "DA-divided-exEnOI" and that without the EnOI as DA-divided. We also consider two settings. In the plain setting neither covariance inflation nor localization is conducted, while in the other setting both auxiliary techniques are applied, with an inflation factor of 1.15 for the fast and slow modes and a half-width of 0.7.

Figure 9 plots the time series of the RMSEs for the DA-divided and DA-divided-exEnOI. When neither

covariance inflation nor localization is adopted, the magnitudes of the trajectories of DA-divided and DA-divided-exEnOI are comparable at many time instants, although substantial differences are also spotted in some cases (e.g., the interval between time steps 100 and 200). On the other hand, when covariance inflation and localization are applied, both DA-divided and DA-divided-exEnOI schemes tend to yield lower time-mean RMSEs. In addition, with covariance inflation and localization, the difference (in time-mean RMSE) between DA-divided and DA-divided-exEnOI is narrowed from around 0.06 to around 0.01. Although the relative performance of the DA-divided and DA-divided-exEnOI schemes may in general change from case to case, the above experiment suggests—at least for the ms-L96 model—the potential of incorporating the EnOI into the divided estimation framework to reduce the computational cost.

5. Discussion and conclusions

We consider the data assimilation problem in coupled systems composed of two subsystems. A straightforward method to tackle this problem is to augment the state vectors of the subsystems. In contrast, the divided estimation framework reexpresses the update formulae in the joint estimation framework in terms of some quantities with respect to the subsystems themselves. We also consider the option of running the subsystems separately, which may bring flexibility and efficiency to data assimilation practices in certain situations, but possibly at the cost of larger discretization errors during model integrations.

We use a multiscale Lorenz 96 model to evaluate the performance of four different data assimilation scenarios, combining different options of joint/divided subsystems and joint/divided estimation frameworks. In addition, we also consider two possible extensions that
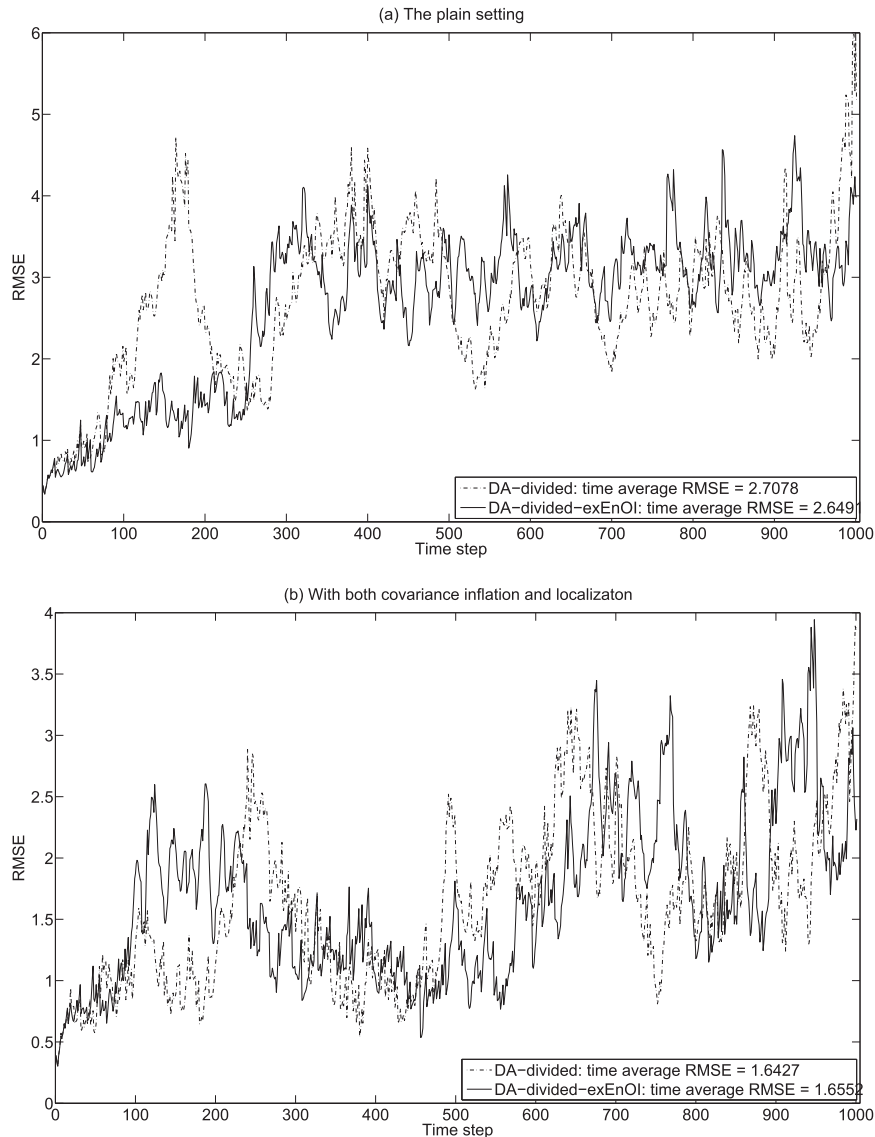
FIG. 9. Time series of the RMSEs in cases of DA-divided and DA-divided-exEnOI. (a) Neither covariance inflation nor localization is applied; (b) both covariance inflation and localization are conducted.

may be relevant for certain coupled data assimilation problems. The experiment results suggest that (i) with identical background ensemble and observation, the joint and divided estimation frameworks yield the same estimate within the machine's numerical precision; (ii) running the subsystems separately may bring in extra flexibility in practice, but at the cost of reduced estimation accuracy in certain circumstances; and (iii) for the approximations used in the extension schemes of section 4, provided that the assimilation schemes are properly configured, one might still obtain reasonable estimates, especially when both covariance inflation and localization are applied.

The current work mainly serves as a proof-of-concept study. In real applications, for instance, data assimilation in coupled ocean–atmosphere general circulation models (OAGCM), model balance and the generation of initial background ensemble are among the issues that require special attention (Saha et al. 2014; Zhang et al. 2007). Additional challenges (e.g., different time scales between ocean and atmosphere components) may also arise when coupled data assimilation is extended to longer time scales (e.g., in the context of climate studies). In this case, certain configurations in the current work may need to be modified, including (but not limited to), for instance, the way to generate the initial background

ensemble and to conduct the conditional sampling (see the online supplemental material). This study may be considered a complement to some existing works in the literature (e.g., Zhang et al. 2007) in terms of the data assimilation schemes in use. In light of the mathematical equivalence between the joint and divided estimation frameworks, we envision that existing techniques (see, e.g., Saha et al. 2014; Zhang et al. 2007) and their future developments used to tackle the aforementioned challenges can also be applied in a similar way within the divided estimation framework.

One may also extend the present work to the situations where the coupled system consists of more than two components. This extension may be of interest in certain situations, for instance, when the interactions of land, ocean, and atmosphere are in consideration or when the domain of a global model is divided into a number of subdomains such that data assimilation is conducted in a set of regional models, similar to the scenario considered in the local ensemble Kalman filter (Ott et al. 2004). In such cases, the corresponding update formulae may become more complicated when adopting the divided estimation framework. This topic will be investigated in the future.

APPENDIX

Gain Matrices in the Divided Estimation Framework

In the divided estimation framework, the most cumbersome part lies in the expansion of the Kalman gain $\mathbf{K}$ in Eq. (8c). Here, we split the deduction into a few steps. First of all, we compute the component $\mathbf{S}^b(\mathbf{S}^h)^{\mathrm{T}}$, which reads

$$\mathbf{S}^b(\mathbf{S}^h)^{\mathrm{T}} = \begin{bmatrix} \mathbf{S}_\eta^b(\mathbf{S}_\eta^h)^{\mathrm{T}} & \mathbf{S}_\eta^b(\mathbf{S}_\xi^h)^{\mathrm{T}} \\ \mathbf{S}_\xi^b(\mathbf{S}_\eta^h)^{\mathrm{T}} & \mathbf{S}_\xi^b(\mathbf{S}_\xi^h)^{\mathrm{T}} \end{bmatrix}. \qquad (A1)$$

Next, we consider the component $[\mathbf{S}^h(\mathbf{S}^h)^{\mathrm{T}} + \mathbf{R}]^{-1}$, which can be expanded to

$$[\mathbf{S}^h(\mathbf{S}^h)^{\mathrm{T}} + \mathbf{R}]^{-1} = \begin{bmatrix} \mathbf{S}_\eta^h(\mathbf{S}_\eta^h)^{\mathrm{T}} + \mathbf{R}_\eta & \mathbf{S}_\eta^h(\mathbf{S}_\xi^h)^{\mathrm{T}} \\ \mathbf{S}_\xi^h(\mathbf{S}_\eta^h)^{\mathrm{T}} & \mathbf{S}_\xi^h(\mathbf{S}_\xi^h)^{\mathrm{T}} + \mathbf{R}_\xi \end{bmatrix}^{-1}. \qquad (A2)$$

Applying the matrix inversion lemma (Simon 2006, p. 11) on the right-hand side of Eq. (A2), we have

$$[\mathbf{S}^h(\mathbf{S}^h)^{\mathrm{T}} + \mathbf{R}]^{-1} = \begin{pmatrix} \mathbf{C}_\eta^{-1} & -\mathbf{A}_{\eta\xi}\mathbf{C}_\xi^{-1} \\ -\mathbf{A}_{\xi\eta}\mathbf{C}_\eta^{-1} & \mathbf{C}_\xi^{-1} \end{pmatrix}, \qquad (A3)$$

where

$$\mathbf{A}_{\eta\xi} = [\mathbf{S}_\eta^h(\mathbf{S}_\eta^h)^{\mathrm{T}} + \mathbf{R}_\eta]^{-1}[\mathbf{S}_\eta^h(\mathbf{S}_\xi^h)^{\mathrm{T}}],$$
$$\mathbf{A}_{\xi\eta} = [\mathbf{S}_\xi^h(\mathbf{S}_\xi^h)^{\mathrm{T}} + \mathbf{R}_\xi]^{-1}[\mathbf{S}_\xi^h(\mathbf{S}_\eta^h)^{\mathrm{T}}], \qquad (A4)$$

and

$$\begin{aligned} \mathbf{C}_\eta &= \mathbf{S}_\eta^h(\mathbf{S}_\eta^h)^{\mathrm{T}} + \mathbf{R}_\eta - [\mathbf{S}_\eta^h(\mathbf{S}_\xi^h)^{\mathrm{T}}]\mathbf{A}_{\xi\eta} \\ &= \mathbf{R}_\eta + \mathbf{S}_\eta^h[\mathbf{I} - (\mathbf{S}_\xi^h)^{\mathrm{T}}[\mathbf{S}_\xi^h(\mathbf{S}_\xi^h)^{\mathrm{T}} + \mathbf{R}_\xi]^{-1}\mathbf{S}_\xi^h](\mathbf{S}_\eta^h)^{\mathrm{T}} \\ &= \mathbf{R}_\eta + \mathbf{S}_\eta^h[\mathbf{I} + (\mathbf{S}_\xi^h)^{\mathrm{T}}\mathbf{R}_\xi^{-1}\mathbf{S}_\xi^h]^{-1}(\mathbf{S}_\eta^h)^{\mathrm{T}} \\ &= \mathbf{R}_\eta + (\mathbf{S}_\eta^h\mathbf{T}_\xi)(\mathbf{S}_\eta^h\mathbf{T}_\xi)^{\mathrm{T}}. \end{aligned} \qquad (A5)$$

The equality between the second and third lines of Eq. (A5) is derived based on the Sherman–Morrison–Woodbury identity (Sherman and Morrison 1950) such that

$$(\mathbf{I} + \mathbf{M}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M})^{-1} = \mathbf{I} - \mathbf{M}^{\mathrm{T}}(\mathbf{M}\mathbf{M}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{M}. \qquad (A6)$$

In the last line of Eq. (A5), $\mathbf{T}_\xi$ is a square root of $[\mathbf{I} + (\mathbf{S}_\xi^h)^{\mathrm{T}}\mathbf{R}_\xi^{-1}\mathbf{S}_\xi^h]^{-1}$ and is equivalent to the transform matrix of the ETKF, with respect to the subsystem ξ (Bishop et al. 2001). Similarly, we have

$$\mathbf{C}_\xi = \mathbf{R}_\xi + (\mathbf{S}_\xi^h\mathbf{T}_\eta)(\mathbf{S}_\xi^h\mathbf{T}_\eta)^{\mathrm{T}}, \qquad (A7)$$

with $\mathbf{T}_\eta$ being a square root of $[\mathbf{I} + (\mathbf{S}_\eta^h)^{\mathrm{T}}\mathbf{R}_\eta^{-1}\mathbf{S}_\eta^h]^{-1}$.

Combining Eqs. (8c), (A1), and (A3), and with some algebra, we obtain the Kalman gain

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{pmatrix}, \qquad (A8)$$

where

$$\mathbf{K}_{11} = \{\mathbf{S}_\eta^b(\mathbf{S}_\eta^h)^\mathrm{T} - \mathbf{S}_\eta^b(\mathbf{S}_\xi^h)^\mathrm{T}[\mathbf{S}_\xi^h(\mathbf{S}_\xi^h)^\mathrm{T} + \mathbf{R}_\xi]^{-1}\mathbf{S}_\xi^h(\mathbf{S}_\eta^h)^\mathrm{T}\}\mathbf{C}_\eta^{-1} = \mathbf{S}_\eta^b\mathbf{T}_\xi(\mathbf{S}_\eta^h\mathbf{T}_\xi)^\mathrm{T}[(\mathbf{S}_\eta^h\mathbf{T}_\xi)(\mathbf{S}_\eta^h\mathbf{T}_\xi)^\mathrm{T} + \mathbf{R}_\eta]^{-1}. \qquad \text{(A9)}$$

The deduction of the last line of Eq. (A9) is similar to that in Eq. (A5) and hence we omit the details. The other elements of **K** can be obtained in a similar way and are summarized in Eq. (11).

REFERENCES

Altaf, U. M., T. Butler, X. Luo, C. Dawson, T. Mayo, and H. Hoteit, 2013: Improving short-range ensemble Kalman storm surge forecasting using robust adaptive inflation. *Mon. Wea. Rev.*, **141,** 2705–2720, doi:10.1175/MWR-D-12-00310.1.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

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

——, 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus,* **59A,** 210–224, doi:10.1111/j.1600-0870.2006.00216.x.

——, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus,* **61A,** 72–83, doi:10.1111/j.1600-0870.2008.00361.x.

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

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, doi:10.1002/qj.169.

——, and ——, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus,* **61A,** 84–96, doi: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, doi:10.1111/j.1600-0870.2008.00372.x.

——, and ——, 2011: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Wea. Rev.,* **139,** 1241–1255, doi:10.1175/2010MWR3403.1.

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

——, and Coauthors, 2012: Joint GODAE OceanView—WGNE workshop on short- to medium-range coupled prediction for the atmosphere-wave-sea-ice-ocean: Status, needs and challenges: Data assimilation—Whitepaper. GODAE OceanView Rep., 14 pp. [Available online at https://www.godae-oceanview.org/outreach/meetings-workshops/task-team-meetings/coupled-prediction-workshop-gov-wgne-2013/white-papers/.]

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

Clayton, A., A. Lorenc, and D. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.,* **139,** 1445–1461, doi:10.1002/qj.2054.

Cooper, M., and K. Haines, 1996: Altimetric assimilation with water property conservation. *J. Geophys. Res.,* **101,** 1059–1077, doi:10.1029/95JC02902.

Counillon, F., and L. Bertino, 2009: Ensemble optimal interpolation: Multivariate properties in the Gulf of Mexico. *Tellus,* **61A,** 296–308, doi:10.1111/j.1600-0870.2008.00383.x.

Dawson, C., S. Sun, and M. F. Wheeler, 2004: Compatible algorithms for coupled flow and transport. *Comput. Methods Appl. Mech. Eng.,* **193,** 2565–2580, doi:10.1016/j.cma.2003.12.059.

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

Fehlberg, E., 1970: Classical fourth and lower order Runge-Kutta formulas with stepsize control and their application to heat transfer problems (in German). *Computing,* **6,** 61–71, doi:10.1007/BF02241732.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proc. F Radar Signal Process.,* **140,** 107–113, doi:10.1049/ip-f-2.1993.0015.

Hamill, T. M., and J. S. Whitaker, 2011: What constrains spread growth in forecasts initialized from ensemble Kalman filters? *Mon. Wea. Rev.,* **139,** 117–131, doi:10.1175/2010MWR3246.1.

——, ——, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129,** 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Harlim, J., A. Mahdi, and A. J. Majda, 2014: An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models. *J. Comput. Phys.,* **257,** 782–812, doi:10.1016/j.jcp.2013.10.025.

Hoteit, I., and D. T. Pham, 2004: An adaptively reduced-order extended Kalman filter for data assimilation in the tropical Pacific. *J. Mar. Syst.,* **45,** 173–188, doi:10.1016/j.jmarsys.2003.11.004.

——, ——, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *J. Mar. Syst.,* **36,** 101–127, doi:10.1016/S0924-7963(02)00129-X.

——, X. Luo, and D. T. Pham, 2012: Particle Kalman filtering: An optimal nonlinear framework for ensemble Kalman filters. *Mon. Wea. Rev.,* **140,** 528–542, doi:10.1175/2011MWR3640.1.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 400 pp.

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

Livings, D. M., S. L. Dance, and N. K. Nichols, 2008: Unbiased ensemble square root filters. *Physica D,* **237,** 1021–1028, doi:10.1016/j.physd.2008.01.005.

Lorenz, E. N., 2006: Predictability—A problem partly solved. *Predictability of Weather and Climate,* T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 40–58.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.,* **55,** 399–414, doi:10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

Luo, X., and I. M. Moroz, 2009: Ensemble Kalman filter with the unscented transform. *Physica D,* **238,** 549–562, doi:10.1016/j.physd.2008.12.003.

——, and I. Hoteit, 2011: Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Wea. Rev.,* **139,** 3938–3953, doi:10.1175/MWR-D-10-05068.1.

——, and ——, 2012: Ensemble Kalman filtering with residual nudging. *Tellus,* **64A,** 17 130, doi:10.3402/tellusa.v64i0.17130.

——, and ——, 2013: Covariance inflation in the ensemble Kalman filter: A residual nudging perspective and some implications. *Mon. Wea. Rev.,* **141,** 3360–3368, doi:10.1175/MWR-D-13-00067.1.

——, and H. Hoteit, 2014a: Ensemble Kalman filtering with residual nudging: An extension to state estimation problems with nonlinear observation operators. *Mon. Wea. Rev.,* **142,** 3696–3712, doi:10.1175/MWR-D-13-00328.1.

——, and I. Hoteit, 2014b: Efficient particle filtering through residual nudging. *Quart. J. Roy. Meteor. Soc.,* **140,** 557–572, doi:10.1002/qj.2152.

——, I. M. Moroz, and I. Hoteit, 2010: Scaled unscented transform Gaussian sum filter: Theory and application. *Physica D,* **239,** 684–701, doi:10.1016/j.physd.2010.01.022.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus,* **56A,** 415–428, doi:10.1111/j.1600-0870.2004.00076.x.

Penny, S. G., 2014: The hybrid local ensemble transform Kalman filter. *Mon. Wea. Rev.,* **142,** 2139–2149, doi:10.1175/MWR-D-13-00131.1.

Petihakis, G., G. Triantafyllou, K. Tsiaras, G. Korres, A. Pollani, and I. Hoteit, 2009: Eastern Mediterranean biogeochemical flux model—Simulations of the pelagic ecosystem. *Ocean Sci.,* **5,** 29–46, doi:10.5194/os-5-29-2009.

Posselt, D. J., and C. H. Bishop, 2012: Nonlinear parameter estimation: Comparison of an ensemble Kalman smoother with a Markov chain Monte Carlo algorithm. *Mon. Wea. Rev.,* **140,** 1957–1974, doi:10.1175/MWR-D-11-00242.1.

Russell, G., J. Miller, and D. Rind, 1995: A coupled atmosphere-ocean model for transient climate change studies. *Atmos.–Ocean,* **33,** 683–730, doi:10.1080/07055900.1995.9649550.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate,* **27,** 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Sherman, J., and W. J. Morrison, 1950: Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann. Math. Stat.,* **21,** 124–127, doi:10.1214/aoms/1177729893.

Simon, D., 2006: *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches.* Wiley-Interscience, 552 pp.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

Triantafyllou, G., I. Hoteit, X. Luo, K. Tsiaras, and G. Petihakis, 2013: Assessing a robust ensemble-based Kalman filter for efficient ecosystem data assimilation of the Cretan Sea. *J. Mar. Syst.,* **125,** 90–100, doi:10.1016/j.jmarsys.2012.12.006.

Van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.,* **137,** 4089–4114, doi:10.1175/2009MWR2835.1.

Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered simplex ensemble. *Mon. Wea. Rev.,* **132,** 1590–1605, doi:10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2.

——, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A comparison of hybrid ensemble transform Kalman filter–optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.,* **135,** 1055–1076, doi:10.1175/MWR3307.1.

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

Zhang, S., M. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.,* **135,** 3541–3564, doi:10.1175/MWR3466.1.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.,* **133,** 1710–1726, doi:10.1175/MWR2946.1.