

Systematic Comparison of Four-Dimensional Data Assimilation Methods With and Without the Tangent Linear Model Using Hybrid Background Error Covariance: E4DVar versus 4DEnVar

JONATHAN POTERJOY AND FUQING ZHANG

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 7 July 2014, in final form 13 January 2015)

ABSTRACT

Two ensemble formulations of the four-dimensional variational (4DVar) data assimilation technique are examined for a low-dimensional dynamical system. The first method, denoted E4DVar, uses tangent linear and adjoint model operators to minimize a cost function in the same manner as the traditional 4DVar data assimilation system. The second method, denoted 4DEnVar, uses an ensemble of nonlinear model trajectories to replace the function of linearized models in 4DVar, thus improving the parallelization of the data assimilation. Background errors for each algorithm are represented using a hybrid error covariance, which includes climatological errors as well as ensemble-estimated errors from an ensemble Kalman filter (EnKF). Numerical experiments performed over a range of scenarios suggest that both methods provide similar analysis accuracy for dense observation networks, and in perfect model experiments with large ensembles. Nevertheless, E4DVar has clear benefits over 4DEnVar when substantial covariance localization is required to treat sampling error. The greatest advantage of the tangent-linear approach is that it implicitly propagates a localized, full-rank ensemble covariance in time, thus avoiding the need to localize a time-dependent ensemble covariance. The tangent linear and adjoint model operators also provide a means of evolving flow-dependent information from the climate-based error component, which is found to be beneficial for treating model error. Challenges that need to be overcome before adopting a pure ensemble framework are illustrated through experiments estimating time covariances with four-dimensional ensembles and comparing results with those estimated with a tangent linear model.

1. Introduction

To simulate real weather events with an acceptable degree of accuracy, numerical weather prediction models rely on carefully specified initial conditions that consider all past and present information leading up to the initialization time. Data assimilation provides a means of initializing models with the most probable state, given prior statistics about the model and imperfect observations. For large dimensional systems, like atmospheric models, these methods typically assume Gaussian errors for the observations and prior probability density, and solve a linear system of equations for the posterior mean and covariance. Data assimilation

methods can take the form of filters, which consider information up to the time of the analysis, or smoothers, which consider both past and future information relative to the analysis time. The most commonly used filters in atmospheric science are the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Mitchell 1998) and 3D variational data assimilation methods (3DVar; Le Dimet and Talagrand 1986), and the most commonly used smoother is the 4D variational method (Thépaut and Courtier 1991). Ensemble-based approaches, like the EnKF, provide a sample representation of the posterior error covariance after each data assimilation cycle, which can be used for generating probabilistic forecasts and prior error statistics for future data assimilation cycles. On the other hand, variational methods provide an efficient means of assimilating large numbers of observations; the solution, however, is purely deterministic in that posterior covariances are not estimated. The framework for variational systems already exists for

Corresponding author address: Jonathan Poterjoy, Department of Meteorology, The Pennsylvania State University, 627 Walker Building, University Park, PA 16802.
E-mail: jpoterjoy@psu.edu

most models used operationally and for research; examples include thoroughly tested data processing and quality control procedures, efficient preconditioning methods, and physical balance constraints for the solution.

Another benefit of variational data assimilation is the straightforward implementation of a hybrid prior error covariance. Hybrid data assimilation methods combine flow-dependent, ensemble covariance with the stationary, climate-based covariances that are typically used in 3DVar and 4DVar. Hamill and Snyder (2000) found that mixing ensemble and static covariance produces better results than using the two components separately during data assimilation: the ensemble provides nonstationary error statistics, while the static covariance removes some of the filter sensitivity to ensemble size. Ensemble filters that use hybrid covariance are also less sensitive to model error, which introduces additional bias in the ensemble estimate (Etherton and Bishop 2004). Using a low-dimensional model, Bishop and Satterfield (2013) provide a mathematical justification for hybrid data assimilation by showing that a linear combination of ensemble and static covariance provides the optimal prior error covariance, given a distribution of ensemble variances and the likelihood of variances given the true conditional error variance.

Among the various types of hybrid data assimilation methods, recent studies find the best performance, in terms of analysis and forecast error reduction, from systems that incorporate ensemble information in variational smoothers (Tian et al. 2008, 2009; Zhang et al. 2009; Buehner et al. 2010b; Bishop and Hodyss 2011; Zhang and Zhang 2012; Zhang et al. 2013; Clayton et al. 2013; Kuhl et al. 2013; Poterjoy and Zhang 2014; Fairbairn et al. 2014; Wang and Lei 2014). One strategy is to introduce ensemble perturbations into pre-existing 4DVar systems, which will be denoted ensemble-4DVar (E4DVar) in this manuscript. A second strategy is to use a 4D ensemble to replace the tangent linear and adjoint model operators in 4DVar (Liu et al. 2008), which will be denoted 4D-ensemble-Var (4DEnVar). 4DEnVar requires the additional cost of running an ensemble forecast through the time window from which observations are collected, but avoids the coding and maintenance of a tangent linear model. More importantly, 4DEnVar does not require running the tangent linear model and its adjoint during each iteration of the cost function minimization, thus increasing the parallel capability of this algorithm over E4DVar. In this study, both E4DVar and 4DEnVar are implemented using the approach introduced in Zhang et al. (2009): from an ensemble forecast, the analysis state is found by minimizing the hybrid variational cost function, while an EnKF updates the ensemble perturbations around the hybrid solution.

Buehner et al. (2010a,b) provided the first comparison of E4DVar and 4DEnVar with other commonly used data assimilation methods for operational global models. Using a low-dimensional model, Fairbairn et al. (2014) performed a more systematic comparison of these methods, and investigated their sensitivity to observation density, ensemble size, model errors, and other system parameters. Both Buehner et al. and Fairbairn et al. applied a one-way coupling strategy in which ensemble perturbations are introduced into the variational system from a cycling EnKF, but the resulting variational solution is not fed back into the EnKF component. This approach facilitates a comparison of the two data assimilation methods using identical ensemble covariance, but does not allow the variational solution to affect the ensemble forecasts between cycles. In both of these studies, the authors applied either an ensemble-estimated covariance or a static covariance (no hybrid covariance) during data assimilation, and found E4DVar to provide more accurate results than 4DEnVar. Similar to Fairbairn et al. the current study applies a low-dimensional model to examine several aspects of E4DVar and 4DEnVar data assimilation systems. This study differs from previous studies that directly compare E4DVar and 4DEnVar in that it focuses primarily on hybrid implementations of the two systems, and uses a two-way coupling between EnKF and variational components, thus allowing E4DVar and 4DEnVar to evolve different ensemble forecast error covariance over time.

Hybrid data assimilation systems have known benefits for atmospheric models, but contain a number of user-specified parameters that must be tuned for optimal performance. These parameters include ensemble-specific values, such as ensemble size, covariance localization radius of influence, and covariance inflation coefficients, which are required for standard implementations of ensemble filters and smoothers. They also require the specification of a weighting coefficient for determining the amount of ensemble and stationary covariance used for the prior error statistics. To investigate the sensitivity of E4DVar and 4DEnVar to these tuning parameters, the two methods are compared over a range of configurations using the dynamical system introduced in Lorenz (1996, hereafter L96). The L96 model mimics the behavior of an atmospheric quantity over a latitudinal circle, but has a state dimension several orders of magnitude smaller than a typical atmospheric model, which allows for large numbers of assimilation experiments to be performed. Three main goals of these experiments are as follows: 1) investigate the role of the stationary covariance in hybrid data assimilation methods in the presence of preexisting ensemble covariance tuning parameters,

2) compare E4DVar with 4DEnVar systematically over a variety of data assimilation scenarios, and 3) examine the challenges that need to be overcome before replacing linearized models in 4DVar with a 4D ensemble.

The organization of this manuscript is as follows. We introduce the two data assimilation methods and forecast model in section 2. In section 3 we describe the implicit role of 4D covariance in variational smoothers, and show that differences in E4DVar and 4DEnVar estimations of time covariances determine their relative performance during data assimilation. Section 4 describes results from cycling data assimilation experiments, and section 5 provides a summary and conclusions.

2. Hybrid four-dimensional data assimilation

In this section, we present the mathematical framework for hybrid E4DVar and 4DEnVar, along with the dynamical system adopted for applying the two methods. We use lowercase boldface roman font to indicate vectors, uppercase boldface arial font to indicate matrices, and italic font to indicate scalars and nonlinear operators.

a. Hybrid E4DVar

The 4DVar method seeks a solution that minimizes the misfit of a control variable to the background state \mathbf{x}_0^b at $t = 0$ and observations \mathbf{y}_t at times $t = 0, 1, 2, \dots, \tau$. For the incremental form of 4DVar, the minimization is carried out with respect to increments $\delta\mathbf{x}_0$ from \mathbf{x}_0^b (Courtier et al. 1994). The function to minimize can be expressed as the sum of background (J_b) and observation (J_o) cost functions:

$$\begin{aligned} J(\delta\mathbf{x}_0) &= J_b(\delta\mathbf{x}_0) + J_o(\delta\mathbf{x}_0) \\ &= \frac{1}{2}\delta\mathbf{x}_0^T \mathbf{B}^{-1} \delta\mathbf{x}_0 \\ &\quad + \frac{1}{2} \sum_{t=0}^{\tau} [\mathbf{H}_t \mathbf{M}_t \delta\mathbf{x}_0 - \mathbf{d}_t]^T \mathbf{R}_t^{-1} [\mathbf{H}_t \mathbf{M}_t \delta\mathbf{x}_0 - \mathbf{d}_t], \end{aligned} \quad (1)$$

where \mathbf{B} is the background error covariance, \mathbf{R}_t is the observation error covariance, \mathbf{M}_t is the tangent linear model operator, and \mathbf{H}_t is the tangent linear observation operator. The vector \mathbf{d}_t contains the innovations at each time along a model trajectory from \mathbf{x}_0^b and is given by

$$\mathbf{d}_t = \mathbf{y}_t - H_t[M_t(\mathbf{x}_0^b)], \quad (2)$$

where M_t and H_t are the (possibly nonlinear) forecast model and observation operators, respectively.

The 4DVar cost function can be preconditioned for faster minimization by letting $\delta\mathbf{x}_0 = \mathbf{U}\mathbf{v}$, where \mathbf{U} is a square root of the background error covariance matrix:

$\mathbf{B} = \mathbf{U}\mathbf{U}^T$ (Lorenc et al. 2000). The cost function is then minimized with respect to the new control variable \mathbf{v} :

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T \mathbf{v} + \frac{1}{2} \sum_{t=0}^{\tau} [\mathbf{H}_t \mathbf{M}_t \mathbf{U}\mathbf{v} - \mathbf{d}_t]^T \mathbf{R}_t^{-1} [\mathbf{H}_t \mathbf{M}_t \mathbf{U}\mathbf{v} - \mathbf{d}_t]. \quad (3)$$

By taking the gradient of (3) with respect to \mathbf{v} ($\nabla_{\mathbf{v}}$) and setting $\nabla_{\mathbf{v}} J(\mathbf{v}) = 0$, the set of solutions that minimize the cost function must satisfy the system of linear equations:

$$\left[I + \sum_{t=0}^{\tau} \mathbf{U}^T \mathbf{M}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t \mathbf{M}_t \mathbf{U} \right] \mathbf{v} = \sum_{t=0}^{\tau} \mathbf{U}^T \mathbf{M}_t^T \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{d}_t. \quad (4)$$

In (4), \mathbf{M}_t^T and \mathbf{H}_t^T are the adjoint versions of the linearized model and observation operators, respectively, and the matrix on the left side of (4) that multiplies \mathbf{v} is the Hessian of $J(\mathbf{v})$. The control variable transformation forces the eigenvalues of the Hessian matrix to be greater than or equal to 1, owing to the fact that the covariance of \mathbf{v} has an inverse equal to the identity matrix. The resulting Hessian has a smaller condition number, which allows numerical solvers to find a solution to (4) in less iterations than would be required otherwise (Lorenc 1997). In this study, we apply the conjugate gradient method to solve (4) for \mathbf{v} , and then multiply the solution by \mathbf{U} to get the analysis increment.

Ensemble information is introduced into the 4DVar cost function by separating increments from the background state into two components:

$$\delta\mathbf{x}_0 = \sqrt{\beta} \delta\mathbf{x}_0^s + \sqrt{1-\beta} \delta\mathbf{x}_0^e, \quad (5)$$

where $\delta\mathbf{x}_0^s$ is the increment resulting from the static background covariance and $\delta\mathbf{x}_0^e$ is the increment resulting from the ensemble-estimated covariance. The coefficient β can range from 0 to 1 and controls the impact of the two increments on the final analysis. This form of hybrid data assimilation is equivalent to replacing the background error covariance in (1) with $\beta\mathbf{B}^s + (1-\beta)\mathbf{P}^e$, where \mathbf{P}^e is the ensemble-estimated error covariance and \mathbf{B}^s is a static covariance that does not change with time¹ (Wang et al. 2007).

Using a similar substitution described above, let $\delta\mathbf{x}_0^s = \mathbf{U}^s \mathbf{v}^s$ for $\mathbf{B}^s = \mathbf{U}^s \mathbf{U}^{sT}$, and $\delta\mathbf{x}_0^e = \mathbf{U}^e \mathbf{v}^e$ for $\mathbf{P}^e \circ \mathbf{C} = \mathbf{U}^e \mathbf{U}^{eT}$, where \circ indicates element wise multiplication, and \mathbf{C} is the correlation matrix used for localizing the ensemble covariance. As described in Buehner (2005), a square root of the localized ensemble covariance is

¹ The standard 4DVar uses $\mathbf{B} = \mathbf{B}^s$, where \mathbf{B}^s is estimated from climatological background errors.

given by $\mathbf{U}^e = (\mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_{N_e})$, where $\mathbf{e}_n = \sqrt{1/(N_e - 1)} \times \text{diag}(\mathbf{x}_{n,0}^f - \bar{\mathbf{x}}_0^f) \mathbf{C}^{1/2}$; Wang et al. (2007) show that this method is mathematically equivalent to the approach described in Lorenc (2003). Here, $\text{diag}(\mathbf{x}_{n,0}^f - \bar{\mathbf{x}}_0^f)$ is a matrix with the n th ensemble perturbation vector in the diagonal elements, and 0 in the nondiagonal elements. Equation (5) can then be written as

$$\begin{aligned} \delta \mathbf{x}_0 &= \sqrt{\beta} \mathbf{U}^s \mathbf{v}^s + \sqrt{1 - \beta} \mathbf{U}^e \mathbf{v}^e, \\ &= (\sqrt{\beta} \mathbf{U}^s \quad \sqrt{1 - \beta} \mathbf{U}^e) \begin{pmatrix} \mathbf{v}^s \\ \mathbf{v}^e \end{pmatrix}, \\ &= \mathbf{U}^{se} \mathbf{v}^{se}, \end{aligned} \quad (6)$$

where the hybrid increment is expressed as a product of the concatenated square root matrices \mathbf{U}^{se} and control variables \mathbf{v}^{se} . From (3), the E4DVar cost function is written as

$$\begin{aligned} J(\mathbf{v}^{se}) &= \frac{1}{2} \mathbf{v}^{seT} \mathbf{v}^{se} \\ &+ \frac{1}{2} \sum_{t=0}^{\tau} [\mathbf{H}_t \mathbf{M}_t \mathbf{U}^{se} \mathbf{v}^{se} - \mathbf{d}_t]^T \mathbf{R}_t^{-1} [\mathbf{H}_t \mathbf{M}_t \mathbf{U}^{se} \mathbf{v}^{se} - \mathbf{d}_t]. \end{aligned} \quad (7)$$

Likewise, the control variables that minimize the cost function are found by substituting \mathbf{U}^{se} for \mathbf{U} and \mathbf{v}^{se} for \mathbf{v} in (4) and solving for \mathbf{v}^{se} .

b. Hybrid 4DEnVar

In 4DEnVar, the role of the tangent linear and adjoint model in (4) is replaced by a four-dimensional ensemble (Liu et al. 2008). A hybrid 4DEnVar system can be formulated using the substitution described above in (6). Recall that E4DVar uses linear model and observation operators to approximate the transformation of $\mathbf{U}\mathbf{v}$ into observation space at time t :

$$\mathbf{M}_t \mathbf{U} \mathbf{v} = \sqrt{\beta} \mathbf{M}_t \mathbf{U}^s \mathbf{v}^s + \sqrt{1 - \beta} \mathbf{M}_t \mathbf{U}^e \mathbf{v}^e. \quad (8)$$

Using an ensemble forecast stored at each observation time in the assimilation window, $\mathbf{M}_t \mathbf{U}^e$ on the right-hand side of (8) can be written as

$$\begin{aligned} \mathbf{M}_t \mathbf{U}^e &= (\mathbf{M}_t \mathbf{e}_1 \mathbf{M}_t \mathbf{e}_2 \dots \mathbf{M}_t \mathbf{e}_{N_e}), \\ &= (\hat{\mathbf{e}}_{1,t} \quad \hat{\mathbf{e}}_{2,t} \dots \hat{\mathbf{e}}_{N_e,t}), \end{aligned} \quad (9)$$

where $\hat{\mathbf{e}}_{n,t} \approx \sqrt{1/(N_e - 1)} \text{diag}(\mathbf{x}_{n,t}^f - \bar{\mathbf{x}}_t^f) \mathbf{C}^{1/2}$ for $n = 1, 2, \dots, N_e$.² Here, $\hat{\mathbf{e}}_{n,t}$ approximates the evolution of

the n th ensemble perturbation from time 0 to t using the full nonlinear model. 4DEnVar requires a localization of the ensemble covariance at each observation time in the window, as opposed to E4DVar, which propagates a localized covariance forward implicitly using the tangent linear and adjoint models. The localization of time-dependent covariance introduces additional complexity in the 4DEnVar algorithm. To perform the localization, we adopt the strategy used in most previous studies, which is to use the same correlation matrix at each time (Liu et al. 2009; Buehner et al. 2010b; Liu and Xiao 2013; Fairbairn et al. 2014).

To remove the tangent linear model operator from the climatological term in (8), \mathbf{M}_t is assumed to be the identity matrix so $\mathbf{M}_t \mathbf{U}^s \approx \mathbf{U}^s$; this assumption is made in ‘‘first guess at the appropriate time’’ data assimilation, or 3DVar-FGAT (Fisher and Andersson 2001). Equation (8) can then be approximated by

$$\begin{aligned} \mathbf{M}_t \mathbf{U} \mathbf{v} &\approx \sqrt{\beta} \mathbf{U}^s \mathbf{v}^s + \sqrt{1 - \beta} (\hat{\mathbf{e}}_{1,t} \quad \hat{\mathbf{e}}_{2,t} \dots \hat{\mathbf{e}}_{N_e,t}) \mathbf{v}^e, \\ &= \mathbf{D}_t \mathbf{v}^{se}, \end{aligned} \quad (10)$$

where $\mathbf{D}_t = (\sqrt{\beta} \mathbf{U}^s \quad \sqrt{1 - \beta} \hat{\mathbf{e}}_{1,t} \quad \sqrt{1 - \beta} \hat{\mathbf{e}}_{2,t} \dots \sqrt{1 - \beta} \hat{\mathbf{e}}_{N_e,t})$. The 4DEnVar cost function is formed by substituting \mathbf{D}_t for $\mathbf{M}_t \mathbf{U}$ and \mathbf{v}^{se} for \mathbf{v} in (3):

$$\begin{aligned} J(\mathbf{v}^{se}) &= \frac{1}{2} \mathbf{v}^{seT} \mathbf{v}^{se} \\ &+ \frac{1}{2} \sum_{t=0}^{\tau} [\mathbf{H}_t \mathbf{D}_t \mathbf{v}^{se} - \mathbf{d}_t]^T \mathbf{R}_t^{-1} [\mathbf{H}_t \mathbf{D}_t \mathbf{v}^{se} - \mathbf{d}_t]. \end{aligned} \quad (11)$$

A similar substitution is made in (4) to solve for the \mathbf{v}^{se} that minimizes $J(\mathbf{v}^{se})$. The analysis increments at each observation time in the assimilation window are then given by $\delta \mathbf{x}_t = \mathbf{D}_t \mathbf{v}^{se}$. For most of this study, the increments and background state at the beginning of the window are added and propagated to the middle of the window to estimate the analysis. In section 4f, we provide a comparison of data assimilation experiments using the increments and background at the middle of the window to justify our choice for defining the analysis.

c. Coupled EnKF–variational data assimilation

The hybrid data assimilation methods discussed above use ensemble forecasts to provide flow-dependent statistics for the background state. In this study, we estimate the posterior mean using the E4DVar or 4DEnVar analysis, and adjust the ensemble perturbations at the middle of the assimilation window using an EnKF (Zhang et al. 2009). In doing so, the variational component assimilates observations at their correct times from $t = 0$ to $t = \tau$, while the EnKF uses ensemble information at $t = \tau/2$ and assumes all observations are

² Another option is to apply the full nonlinear observation operator on the forecast members and mean before calculating $\hat{\mathbf{e}}_{n,t}$ as is done in Liu et al. (2008).

valid at the same time. The hybrid analysis at $t = 0$ is propagated to $t = \tau/2$ with M_t , and added to the posterior perturbations to initialize the ensemble forecast for the next data assimilation cycle.

d. Treatment of sampling errors

The use of a small ensemble size relative to the state dimension, along with model error, nonlinearity, and other systematic errors in the data assimilation algorithm, can introduce bias in the estimation of the posterior mean and covariance. To treat noise in the ensemble error approximation, the prior ensemble covariance is localized using a Gaussian correlation function, with a tunable length scale parameter, or radius of influence (ROI). The localization function is applied using the same ROI in the EnKF and variational steps of the two-way coupled data assimilation systems examined in this study. The posterior covariance is also inflated after each cycle using a covariance relaxation method (Zhang et al. 2004), which replaces the posterior ensemble perturbations with a linear combination of posterior and prior perturbations:

$$\mathbf{x}_n^{la} \leftarrow (1 - \alpha)\mathbf{x}_n^{la} + \alpha\mathbf{x}_n^{lf}. \tag{12}$$

The α in (12) is called the ‘‘relaxation coefficient’’ and ranges from 0 to 1, where $\alpha = 0$ implies no inflation.

e. Forecast model

We use the L96 model to examine several aspects of the two data assimilation systems and explore the sensitivity of each method to tunable parameters. The model consists of variables x_i for $i = 1, 2, \dots, N_x$, which are equally spaced on a periodic domain. These variables are governed by the following set of equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \tag{13}$$

which are integrated forward numerically using the fourth-order Runge–Kutta method with a time step of 0.005 time units or 36 min.³ This dynamical system contains a quadratic advection term, a linear dissipation term, and a constant damping term. It exhibits chaotic behavior for $F > 5$ and $N_x > 12$, and produces a solution that has an error doubling time comparable to synoptic-scale features in atmospheric models for $F = 8$ (L96). For this study, we fix N_x at 40 and use $F = 8$ for most experiments, but allow F to vary when investigating the sensitivity of E4DVar and 4DEnVar to model error and

error growth rate. To distinguish between the F used by the true dynamical system, and the F used by the model in imperfect model experiments, we use F_t and F_m to denote the true and model values of F , respectively.

3. Four-dimensional covariance

The analysis increments that minimize the E4DVar and 4DEnVar cost functions depend implicitly on spatiotemporal covariances modeled in the assimilation window. For the case where a single observation at time $t = t_n$ is assimilated, Thepáut et al. (1996) show that the analysis increment at $t = 0$ that minimizes the 4DVar cost function is given by

$$\delta\mathbf{x}_0 = \mathbf{P}_{0,t_n} \mathbf{H}_{t_n}^T \frac{\mathbf{d}_{t_n}}{\sigma_b^2 + \sigma_o^2}, \tag{14}$$

where $\sigma_b^2 = \mathbf{H}_t \mathbf{M}_t \mathbf{B} \mathbf{M}_t^T \mathbf{H}_t^T$ is the background error covariance at $t = t_n$ projected into observation space, and σ_o^2 is the observation error variance. The matrix \mathbf{P}_{0,t_n} in (14) contains covariances between state variables at the beginning of the assimilation window and at $t = t_n$, and is given by $(\mathbf{M}_{t_n} \mathbf{B})^T$. E4DVar and 4DEnVar both rely implicitly on \mathbf{P}_{0,t_n} and σ_b^2 to calculate $\delta\mathbf{x}_0$. Because the spread of information from observations at future times to state variables at the beginning of the assimilation window is controlled by \mathbf{P}_{0,t_n} , we will discuss the means by which this covariance matrix is estimated by E4DVar and 4DEnVar.

The E4DVar calculation of \mathbf{P}_{0,t_n} follows the same form as 4DVar, but with \mathbf{B} replaced with a linear combination of the static and localized ensemble covariance:

$$\begin{aligned} \mathbf{P}_{0,t_n} &= \{\mathbf{M}_{t_n} [\beta \mathbf{B}^s + (1 - \beta)(\mathbf{P}^e \circ \mathbf{C})]\}^T, \\ &= \beta [\mathbf{M}_{t_n} \mathbf{B}^s]^T + (1 - \beta) [\mathbf{M}_{t_n} (\mathbf{P}^e \circ \mathbf{C})]^T. \end{aligned} \tag{15}$$

The hybrid \mathbf{P}_{0,t_n} comes from a weighted sum of static and ensemble covariances that are propagated in time by the tangent linear model.

In 4DEnVar, the approximation of \mathbf{P}_{0,t_n} also contains a contribution from the static and ensemble covariance, except time covariance is not evolved from \mathbf{B}^s , and the ensemble component comes from a 4D ensemble forecast from $t = 0$ to $t = t_n$:

$$\mathbf{P}_{0,t_n} = \beta [\mathbf{B}^s]^T + (1 - \beta) \left\{ \frac{1}{N_e - 1} \sum_{n=1}^{N_e} \mathbf{x}'_{0,n} [(\mathbf{x}_{t_n,n})']^T \right\} \circ \mathbf{C}. \tag{16}$$

4DEnVar has the advantage of using the nonlinear model to estimate the ensemble component of \mathbf{P}_{0,t_n} , which is beneficial for cases when the tangent linear

³ As described in L96, the variables are scaled so that the time units in (13) are approximately equivalent to 5 days.

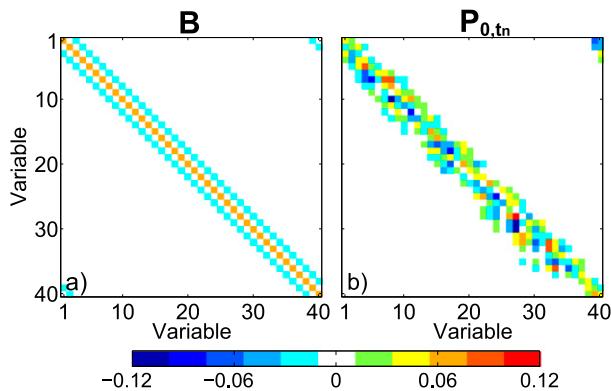


FIG. 1. (a) The climatological covariance, and (b) the true \mathbf{P}_{0,t_n} estimated for $t_n = 24$ h using a 10^5 size ensemble initialized from the climatological error distribution.

approximation is invalid, or when parts of M_t cannot be linearized. The latter case occurs for complex physical parameterization schemes used in atmospheric models (e.g., microphysics and cumulus schemes); therefore, \mathbf{M}_t is often linearized from a simplified version of M_t (e.g., Sun and Crook 1997; Honda et al. 2005; Huang et al. 2009).

In addition to not being able to evolve time covariance from \mathbf{B}^s , another disadvantage of 4DEnVar is that localization must be performed on a time covariance matrix. Comparing (15) and (16), E4DVar avoids this problem by localizing at $t = 0$, then applying the tangent linear model on the full-rank localized ensemble covariance. Recall from section 2b, the formulation of 4DEnVar tested in this study does not consider time when performing the localization. We examine the impact of the different localizations applied in E4DVar and 4DEnVar, by considering the case where the weight of static covariance is set to 0. To perform these experiments, 10 ensemble members are drawn from $N(\mathbf{x}_b, \mathbf{B}^c)$, where \mathbf{x}_b is taken as a 100-day forecast from arbitrary initial conditions, and \mathbf{B}^c (Fig. 1a) is a climatological covariance estimated during one of the cycling data assimilation experiments described in section 4. We apply the tangent linear and ensemble methods to estimate \mathbf{P}_{0,t_n} with and without covariance localization, then calculate the mean squared error (MSE) of \mathbf{P}_{0,t_n} for $t_n = 0$ –48 h. The localization applied for these cases takes the form of a Gaussian function with a ROI of 3; this value was chosen based on results from cycling experiments presented in section 4. The true value of \mathbf{P}_{0,t_n} (Fig. 1b) at each time is estimated from a 10^5 -member ensemble forecast from the nonlinear model, which produces results that are nearly indistinguishable from cases that use \mathbf{M}_t to evolve time covariances directly from \mathbf{B}^c (not shown). RMSEs are averaged over 1000 trials and plotted in Fig. 2 with error bars that represent the 95%

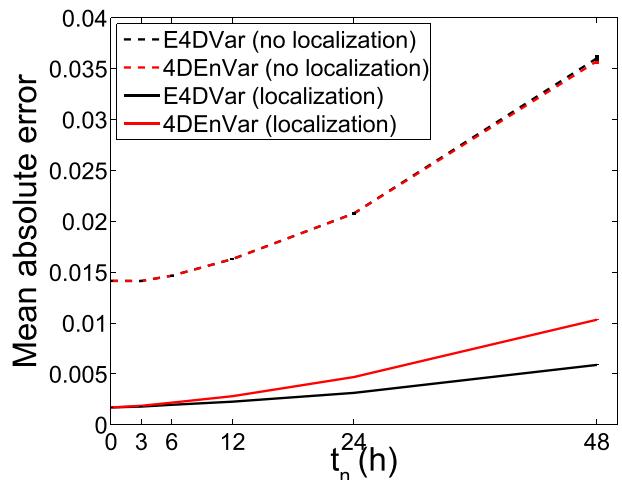


FIG. 2. Mean squared error of covariance estimated from the E4DVar method (black) and 4DEnVar method (red) without localization (dashed lines) and with localization (solid lines).

confidence interval for the calculations; owing to the large number of trials, the confidence interval at most times is narrower than the plotted line width.

When no localization is applied (dashed lines in Fig. 2), E4DVar and 4DEnVar provide similar errors for \mathbf{P}_{0,t_n} . For this case, the E4DVar covariance simplifies to

$$\mathbf{P}_{0,t_n} = \frac{1}{N_e - 1} \sum_{n=1}^{N_e} \mathbf{x}'_{0,n} [\mathbf{M}_{t_n} \mathbf{x}'_{0,n}]^T, \quad (17)$$

which is identical to the low-rank 4DEnVar approximation, but with ensemble perturbations evolved in time using \mathbf{M}_{t_n} . Equation (17) is no longer valid when localization is applied; in which case, E4DVar uses \mathbf{M}_{t_n} to evolve a full-rank covariance in time. The 4DEnVar estimate, however, remains low rank until localization is applied on the final solution. Applying the localization at $t = t_n$, rather than at the beginning of the assimilation window, results in significantly larger errors in the 4DEnVar representation of \mathbf{P}_{0,t_n} (solid lines in Fig. 2).

The effects of the different localization strategies are illustrated in Fig. 3 by applying both methods to estimate covariances between all state variables in \mathbf{x}_b at $t = 0$ h and variable 20 at $t_n = 0, 6, 12,$ and 24 h; these covariances come directly from the 20th column of \mathbf{P}_{0,t_n} . In the left column of Fig. 3, the true covariance at each time (black line) is plotted with values estimated from a 10-member ensemble (blue line). In the right column, the localized covariances are estimated by applying \mathbf{M}_{t_n} to evolve time covariance from $\mathbf{P}^e \circ \mathbf{C}$ (red line, indicating the E4DVar covariance) and by localizing the 10-member ensemble \mathbf{P}_{0,t_n} directly (dashed blue line, indicating the 4DEnVar covariance).

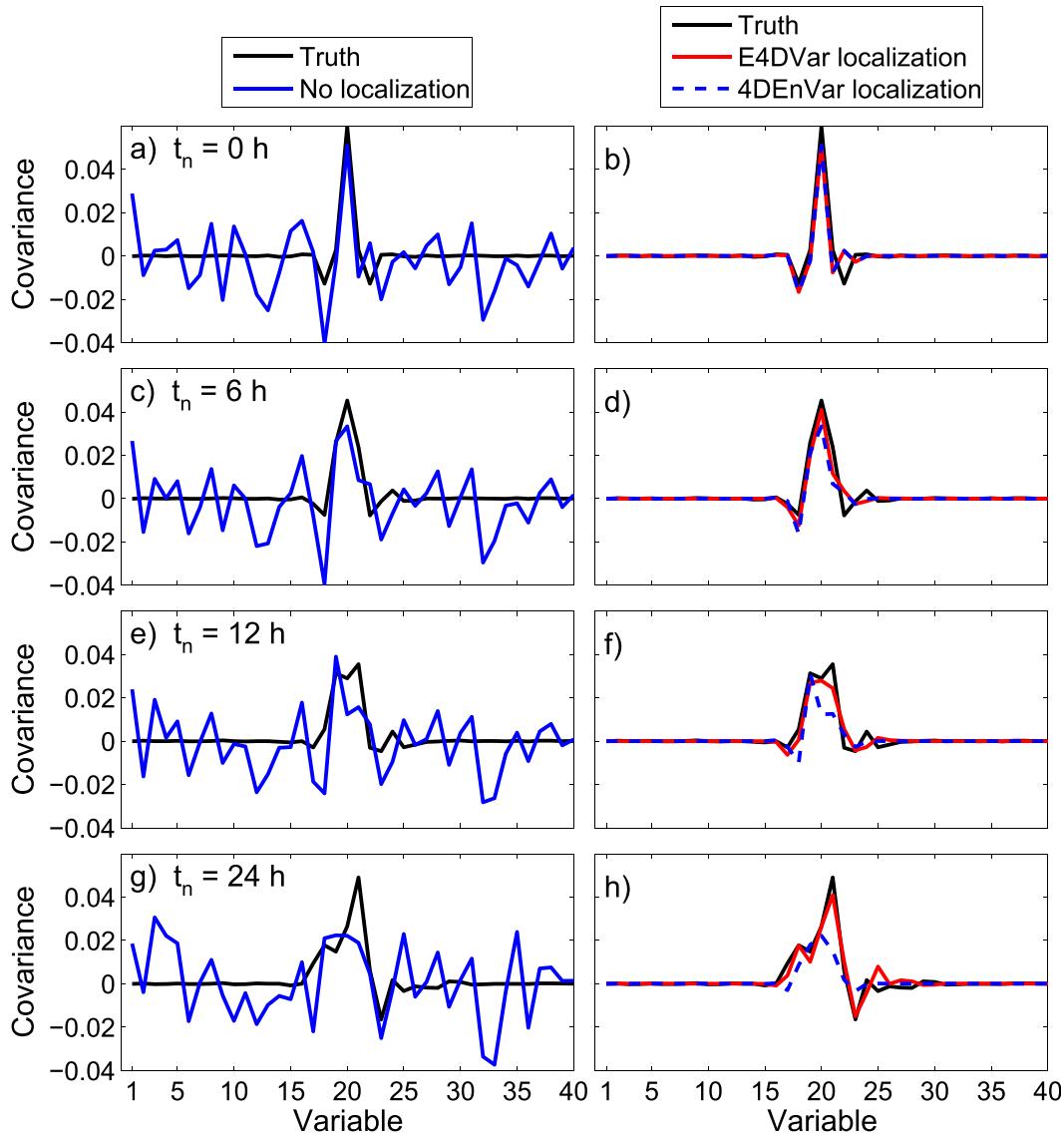


FIG. 3. Covariance between all 40 state variables at $t = 0$ h and variable 20 at (a),(b) $t_n = 0$ h; (c),(d) $t_n = 6$ h; (e),(f) $t_n = 12$ h; and (g),(h) $t_n = 24$ h. Values are estimated from the true covariance (black lines in first and second columns), a 10-member ensemble covariance (blue lines in first column), E4DVar localized covariance (red lines in second column), and 4DEnVar localized covariance (dashed blue lines in second column).

For t_n close to 0 h, both methods provide an accurate representation of the true covariance near variable 20. As t_n increases, the E4DVar covariance maintains the correct structure of the true signal, which propagates downstream and changes shape. The 4DEnVar covariance, however, becomes contaminated by spurious correlations, which propagate into the region where the true signal was originally represented well by the low-rank approximation. To match the true covariance by $t_n = 24$ h, the ensemble covariance in Fig. 3g requires the following: 1) the removal of distant spurious correlations, 2) a localization function that has a peak near the

downstream covariance maximum, and 3) a change in structure of the signal. The first and second requirements may be obtainable through adaptive localization methods that have been applied for similar applications (e.g., Bishop and Hodyss 2011), but the third requirement is unattainable by localization alone.

Though not shown, the localization strategy used by E4DVar to estimate single columns of \mathbf{P}_{0,t_n} (red lines in Fig. 3) can be implemented in an ensemble framework by localizing the ensemble perturbations around the variable of interest at $t = 0$ h before running the ensemble forecast to $t = t_n$ h. For example, each $\mathbf{x}_{0,n}$ can be

TABLE 1. Configuration of cycling data assimilation experiments.

Expt	N_e	α	ROI	T (h)	β	F_t	F_m	κ	Δt (h)
Ensemble size	5, 10, 20, 40	Variable	Variable	6	Variable	8	8	2	1.2
Window length	10	Variable	Variable	0, 6, 24	Variable	8	8	2	1.2
Obs network	20	Variable	Variable	6	0	8	8	1, 2, 3	1.2, 3
Model forcing	10	Variable	Variable	24	0	9, 10, 11	9, 10, 11	1, 3	1.2
Model error	20	Variable	Variable	24	Variable	8	7	2	1.2
Choice of analysis	10	Variable	5	24	Variable	8	8	2	1.2

replaced by $\mathbf{x}'_{0,m} \circ \mathbf{C}_{20}$, where \mathbf{C}_{20} is the 20th column of \mathbf{C} . The ensemble mean is then added to the localized perturbations, which are propagated forward with the nonlinear model to provide a sample representation of column 20 of \mathbf{P}_{0,t_n} . This approach avoids the use of a tangent linear model, but requires running the nonlinear model N_e times for each state variable.

The discussion in this section focuses mostly on the approximation of \mathbf{P}_{0,t_n} in the two data assimilation algorithms; however, the same positive and negative qualities of the two methods are equally true for the estimation of σ_b^2 . Discrepancies between the 4D covariance estimation in E4DVar and 4DEnvar will be used to interpret the relative performance of these two methods in the cycling data assimilation experiments described in section 4.

4. Cycling data assimilation experiments

In this section, we present results from experiments that examine the sensitivity of E4DVar and 4DEnVar to a variety of modeling and observing system configurations, and tuning parameters that control the background error statistics. Each experiment uses observations generated from a “truth” simulation with added noise drawn from $N(0, I)$ every κ variables (grid points) and Δt time units; unless otherwise noted in the manuscript, $\kappa = 2$ and $\Delta t = 1.2$ h, which means half the model variables in \mathbf{x} are observed every 1.2 h with an error variance of 1. Observations are assimilated over a 2600-day period, and root-mean-square errors (RMSEs) from the last 2500 days are used to quantify the accuracy of the posterior mean analyses. The first 100 days of data assimilation act as a spinup period to allow the ensemble covariance to evolve flow-dependent information for the given setup of the model and observation network.

We perform separate sets of data assimilation experiments to examine the sensitivity of E4DVar and 4DEnVar to ensemble size, window length, observation network, model forcing, and model error; the system parameters for each of these experiments are listed in Table 1. Results from each 2500-day assimilation period are summarized by averaging the analysis RMSEs over

all state variables and data assimilation cycles. Given the number of samples used to approximate these values, the largest margin of error for the RMSE calculations is found to be less than 0.01 with 95% confidence. For simplicity, RMSEs in each experiment are rounded to the nearest 0.001 units when the margin of error is smaller than 0.001, and 0.01 units when the margin of error is greater than 0.001.

The hybrid forms of E4DVar and 4DEnVar require two prior covariance matrices: \mathbf{B}^s and \mathbf{P}^e . As described in section 2c, \mathbf{P}^e is estimated from an ensemble forecast run from the previous data assimilation cycle, and \mathbf{B}^s comes from time-averaged forecast errors. To estimate \mathbf{B}^s , we first run the data assimilation systems with $\beta = 0$ for a given model setup, assimilation window length T , and observation network, and use differences between the forecast and true system state to estimate covariances for the model state variables. These covariances are then averaged to form a covariance matrix that contains the same diagonal and off-diagonal elements for each variable; an example \mathbf{B}^s is provided in Fig. 1a.

a. Ensemble size

For large modeling applications, computing resources often limit the number of ensemble members used to estimate \mathbf{P}^e . This sampling deficiency leads to an underestimation of the posterior error variance during every cycle unless covariance localization and inflation are used. In this section, we apply E4DVar and 4DEnVar with $N_e = 5, 10, 20$, and 40 to examine the sensitivity of each method to sampling errors caused by the finite ensemble size. These experiments use no static covariance (i.e., $\beta = 0$), a fixed window length of $T = 6$ h, and a range of ROI and α to tune the imperfect error statistics from the ensembles.

Figure 4 shows mean RMSEs calculated for each ensemble size and configuration of ROI and α near the optimal values, which are indicated by the black box in each subplot. As expected, both methods become more stable and require less localization (larger ROI) and less inflation (smaller α) as N_e increases. The optimal ROI and α for a given N_e are comparable for E4DVar and 4DEnVar, and are typically found near values that lead

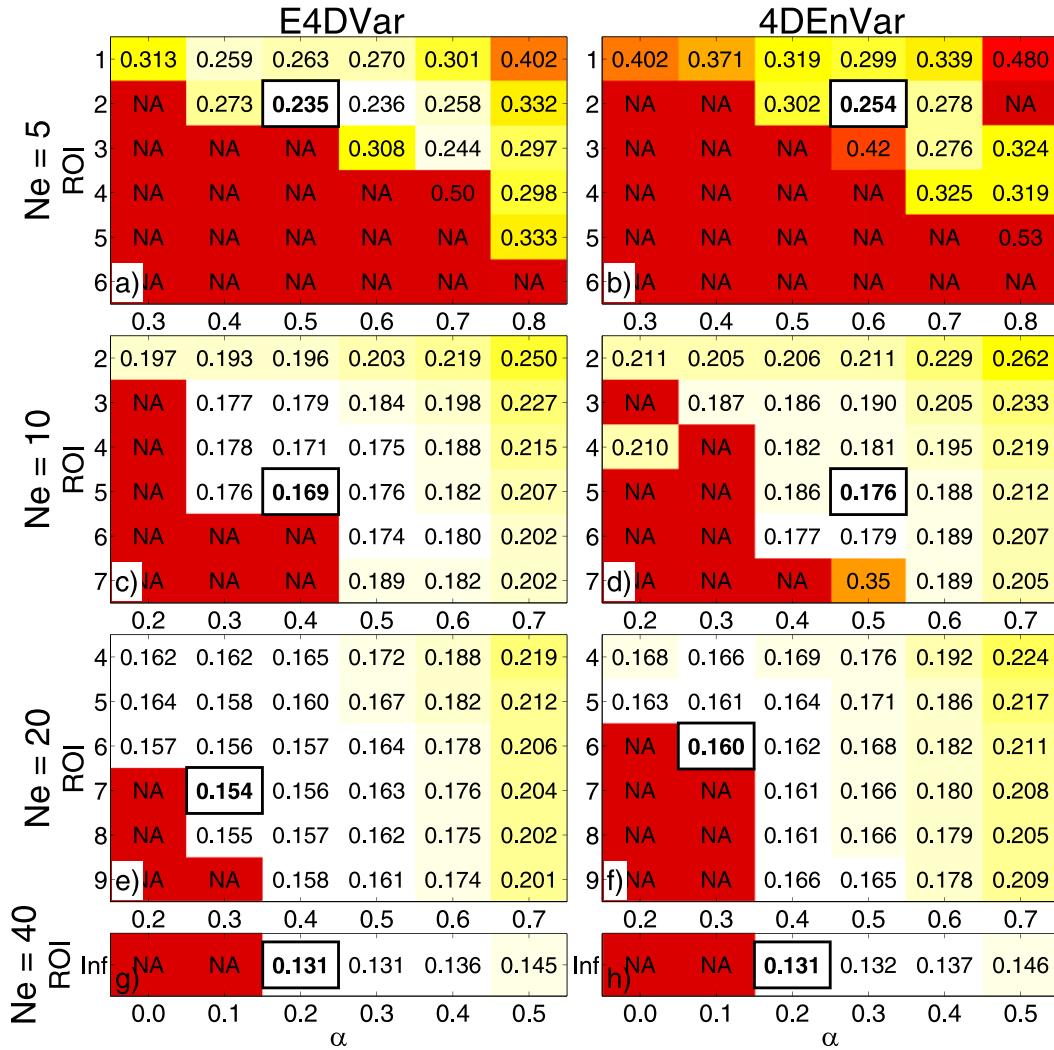


FIG. 4. Mean analysis RMSEs estimated for a range of α and ROI, and $N_e = 5, 10, 20,$ and 40 ; the range of α and ROI changes for each row. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

to filter divergence; we define filter divergence objectively by flagging configurations that produced 100-cycle-average RMSEs larger than 4 with NA for “not available” in the figures. In general, the configurations that modify the covariance the least while maintaining a stable solution produce the best results. E4DVar consistently outperforms 4DEnVar for cases with large sampling error (i.e., $N_e = 5, 10,$ and 20), owing to differences in localization. This result is verified by increasing N_e to 40 and turning the covariance localization off (Figs. 4g,h). As suggested by Fig. 2, the two methods produce comparable RMSEs when no modifications are made to the prior covariance; Fairbairn et al. (2014) found the same result using a slightly more complex version of the L96 model.

A hybrid covariance is used to compensate for sampling errors in cases that use $N_e = 5, 10,$ and 20 (Figs. 4a–f).

These experiments use a fixed ROI, taken from the optimal values found in Fig. 4, and assimilate the observations using a range of α and β values (Fig. 5). For $N_e = 5$ and 10 , E4DVar and 4DEnVar benefit from a configuration that uses a mix of about 10% static and 90% ensemble covariance. The application of a hybrid covariance also reduces the need for large variance inflation, as seen by the decrease in optimal α between Figs. 4a–d and 5a–d. The positive impact of using a hybrid covariance is less significant for $N_e = 20$, where the optimal β drops below 0.05 for the pair of data assimilation systems. The optimal α between Figs. 4e,f and 5e,f, remains fixed at 0.3, suggesting that covariance relaxation is more effective than the hybrid covariance at treating sampling errors when N_e increases.

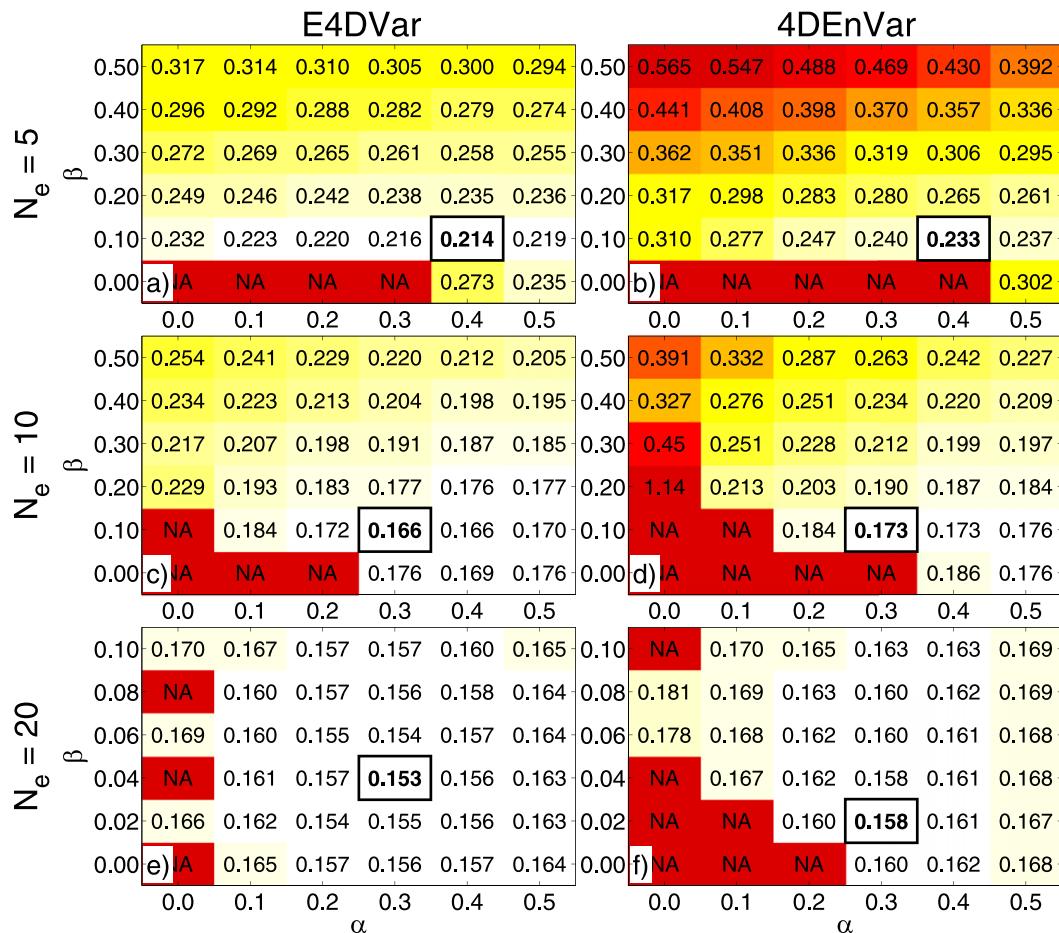


FIG. 5. Mean analysis RMSEs estimated for a range of α and β , with ROI fixed at the optimal values found in Fig. 4 and $N_e = 5, 10$, and 20 ; the range of β changes between the second and third rows. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

The relationship between β and ROI is examined by holding α constant at the optimal values found in Fig. 5. Mean RMSEs from this experiment (Fig. 6) show that the optimal ROI remains close to the values found when $\beta = 0$ in Fig. 4. The performance of both methods, however, depends less on the prescribed ROI for hybrid covariance, which is beneficial for cases in which the optimal ROI is unknown or changes significantly between observation type, location, and flow regime. While the results presented in Figs. 5 and 6 show E4DVar producing the lowest RMSEs over the range of parameters, it is significant to note that E4DVar and 4DEnVar respond equally to the use of a hybrid covariance to treat sampling errors.

b. Assimilation window length

This section examines the sensitivity of E4DVar and 4DEnVar to assimilation window length. The same observations are assimilated in all experiments, except T

values of 0, 6, and 24 h are used. Smaller values of T require more frequent data assimilation cycles, with $T = 0$ h resulting in an ensemble filter step each time observations are available ($\Delta t = 1.2$ h). Because the experiments with longer window lengths contain fewer assimilation cycles over the test period, RMSEs are averaged from analysis data every 24 h to match the longest T used in the experiments.

Figure 7 shows mean RMSEs from experiments that use $\beta = 0$, $N_e = 10$, and a range of ROI and α values near the optimal pair of ensemble covariance parameters. The methods are equivalent when run in filtering mode ($T = 0$), and they both exhibit a decrease in RMSEs as T is increased to 24 h. Smoothers use increasingly more data to approximate the analysis state as window length is extended, which leads to a smaller sensitivity to random errors in the ensemble and observation network, and an overall reduction in analysis RMSEs. For the same reasoning, the longer window length also

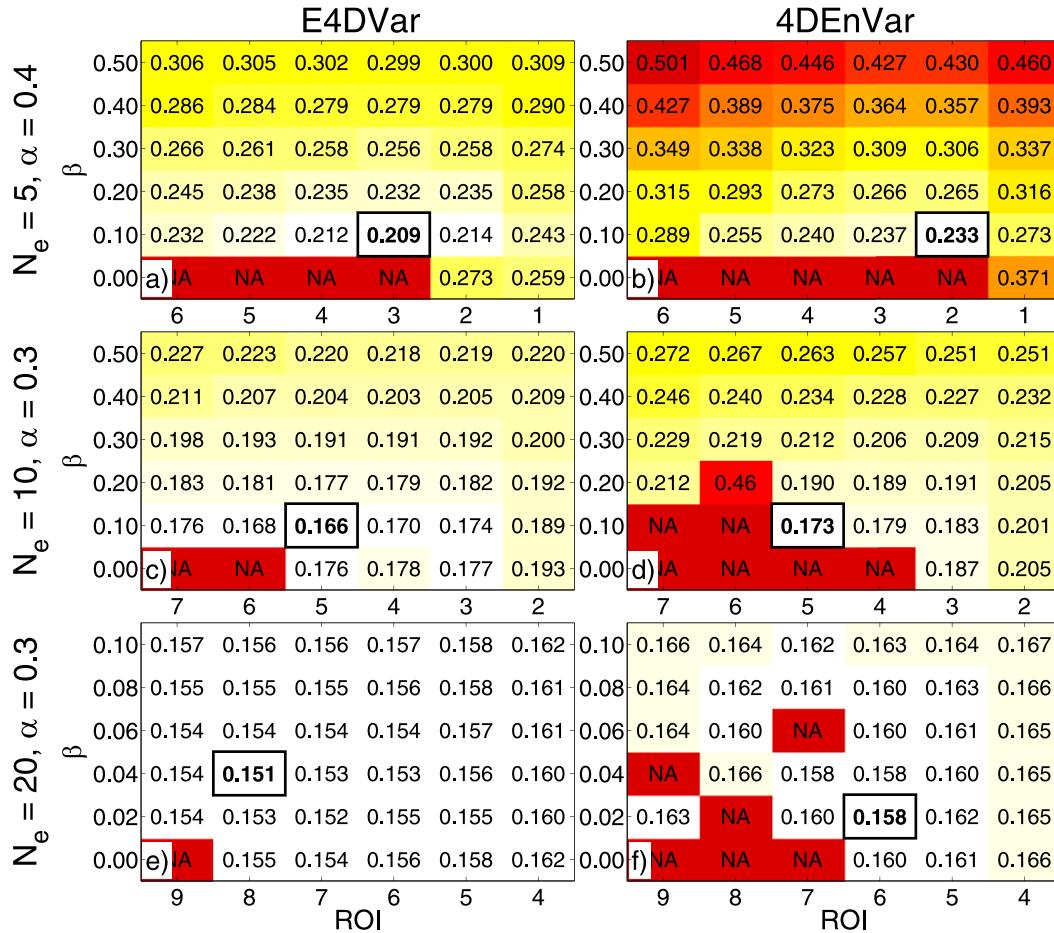


FIG. 6. Mean analysis RMSEs estimated for a range of α and ROI, and $N_e = 5, 10,$ and 20 ; the range of β and ROI changes between each row. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

decreases the amount of relaxation needed to prevent filter divergence (Fig. 7). Though not shown, the performance of both methods continues to increase with T until reaching a point where nonlinearity in the model dynamics introduces errors in the solution. Figure 7 also verifies that differences in covariance localization cause E4DVar to outperform 4DEnVar as window length increases; the inability of 4DEnVar (with localization) to produce analyses as accurate as E4DVar for large T is consistent with the error estimation of \mathbf{P}_{0,t_n} for large t_n in Fig. 2.

We introduce hybrid covariance in the cost function to examine how T affects the optimal weight of static covariance for $N_e = 10$. Figure 8 shows that the optimal β decreases with T in a manner similar to α (Fig. 7). This result provides additional evidence that both 4D data assimilation methods are less sensitive to sampling noise as assimilation window length increases. Furthermore, the inability of 4DEnVar to propagate \mathbf{B}^s in the

assimilation window does not appear to decrease the benefit of using static covariance to treat sampling errors induced by the small ensemble size—this conclusion follows from the near-equal decrease in RMSEs when hybrid covariance is applied in E4DVar and 4DEnVar. Because of the compact shape of the climatological covariance in this dynamical system (cf. Figs. 1 and 3), \mathbf{B}^s serves mostly to reduce the weight of spurious sample correlations that remain after localization.

c. Observation network

We examine the sensitivity of the data assimilation methods to spatial and temporal observation density by changing the observation parameters κ and Δt . In doing so, we adopt the model and data assimilation setup used to produce Figs. 4e and 4f by fixing the system parameters at $N_e = 20, T = 6$ h, and $\beta = 0$; the ensemble size is increased from previous experiments to reduce the effects of random sampling errors on the data assimilation.

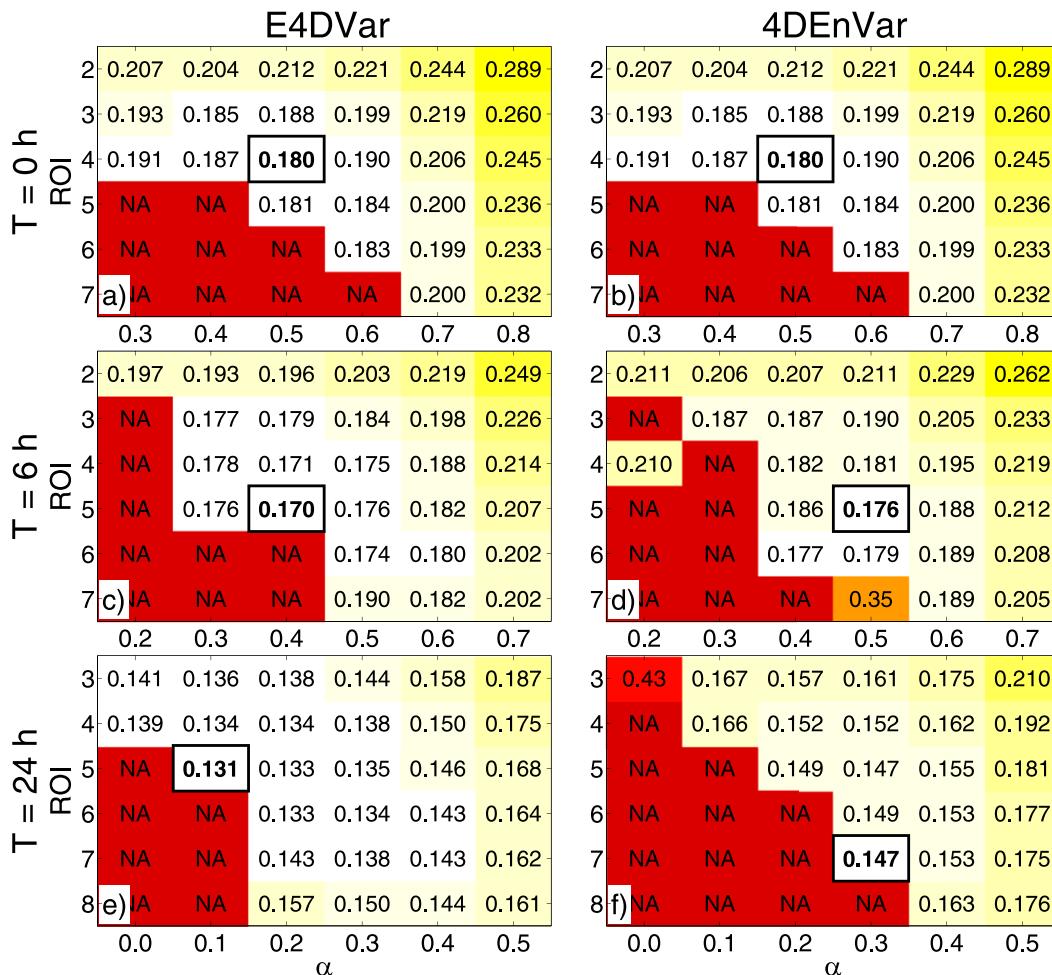


FIG. 7. Using an ensemble size of 10, mean analysis RMSEs are estimated for a range of α and ROI, and $T = 0, 6,$ and 24 h; the range of α values changes between each row. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

We first decrease κ from 2 to 1 (Figs. 9a,b), which doubles the spatial density of the observations. The additional observations reduce the differences in analysis accuracy between E4DVar and 4DEnVar, and allow both methods to become stable over a larger number of ensemble parameters (cf. Figs. 4c,d). Because the system has full observability in this experiment, the ensemble members remain close to the true solution between data assimilation cycles, which reduces the sensitivity of analysis RMSEs to covariance localization and inflation.

We then make the observation network sparse by increasing κ to 3 (Figs. 9c,d). With κ fixed at 3, we also decrease the temporal frequency of observations by increasing Δt from 1.2 to 3 h (Figs. 9e,f). These experiments show clear benefits of E4DVar over 4DEnVar in terms of stability and accuracy when fewer observations are used. For sparse observation networks, the number

of possible model solutions that can fit the observations in the assimilation window increases. The process of modeling an accurate background error covariance in the assimilation window then becomes more challenging, owing to the larger variance required to match the increase in uncertainty. As a result, the solution depends more on the choice of ROI and α for estimating background error covariance given limited information.

We perform a second set of experiments to examine the impact of using a hybrid covariance to improve the background error estimation for sparse observation networks. Figure 10 shows the sensitivity of E4DVar and 4DEnVar to parameters α and β when data are assimilated from the three different observation networks using a fixed ROI of 7. For the densely spaced observations (Figs. 10a,b), the optimal β is found near 0.02, which amounts to only a small decrease in E4DVar and 4DEnVar analysis RMSEs compared to the case in

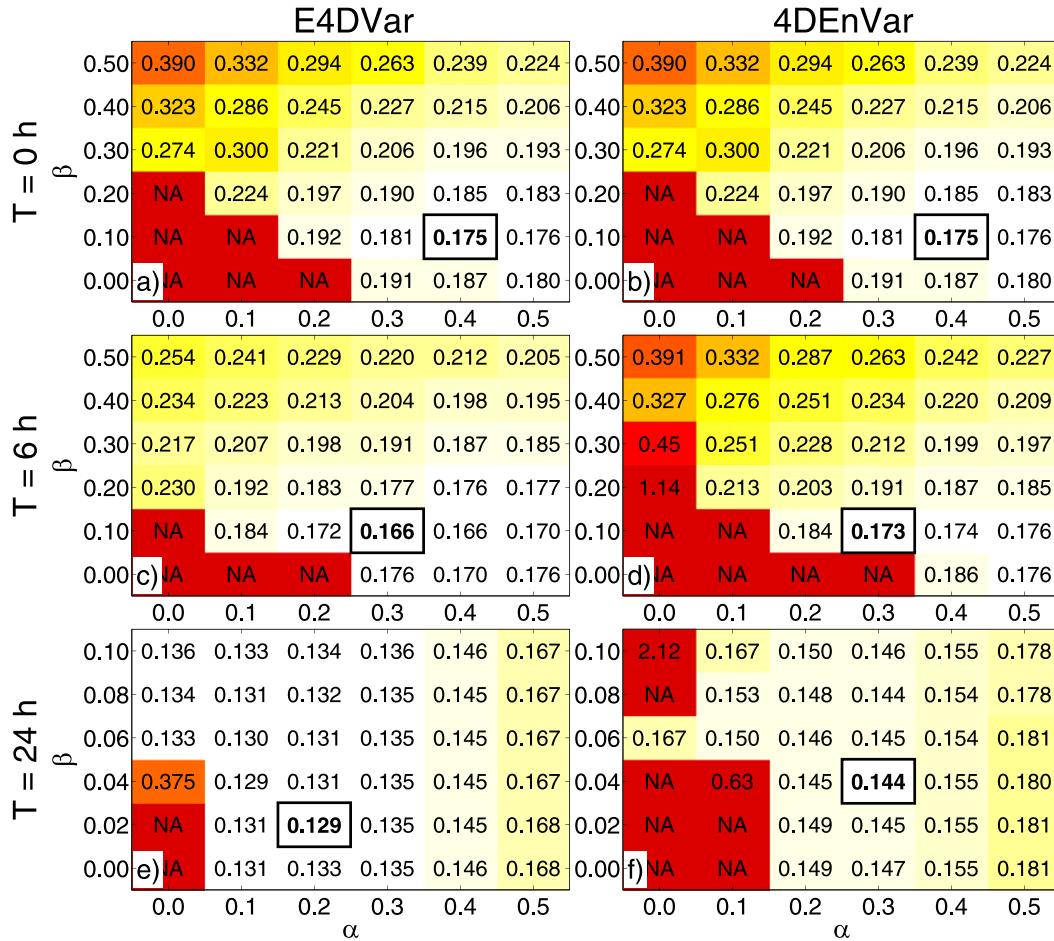


FIG. 8. Mean analysis RMSEs estimated for a range of α and β , and $T = 0, 6$, and 24 h; the range of β values changes between each row. Ensemble size and ROI are fixed at 10 and the optimal value found in Fig. 7, respectively. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

which no static covariance is used. The optimal weight of the static covariance, however, increases when the observation network is made more sparse (Figs. 10c–f), which is indicative of less accurate ensemble covariance when fewer observations are available to constrain the solution. While E4DVar continues to produce smaller analysis RMSEs than 4DEnVar, we find that the two methods benefit equally from the use of a hybrid covariance in these experiments.

d. Model forcing

In section 3, 4D covariances modeled using ensemble and tangent-linear approaches show that the accuracy of \mathbf{P}_{0,t_n} decreases faster with time in 4DEnVar compared to E4DVar (Fig. 2). One problem with the ensemble representation of \mathbf{P}_{0,t_n} is that spurious correlations—which are typically removed at the beginning of the time window in E4DVar—propagate forward and degrade the

estimation of the true covariance signal (Fig. 3). The model forcing term F can control the rate at which small initial-condition errors grow in time between $t = 0$ and the time of observations at $t = t_n$. For example, the error doubling time of the model forecast when F is increased from 8 to 10 decreases from 2.1 to 1.5 days (Lorenz and Emanuel 1998). Likewise, the model forcing term is expected to influence the period over which time covariance can be estimated effectively by the two data assimilation methods.

Using a perfect model, we perform data assimilation experiments with F set to 9, 10, and 11 to test E4DVar and 4DEnvar on dynamical systems that are intrinsically less predictable. These experiments are similar to those presented in Figs. 7e and 7f in that we use $N_e = 10$, $T = 24$ h, and $\beta = 0$, which means the performance of each method depends greatly on how \mathbf{P}_{0,t_n} is estimated over a long window using statistics from a small

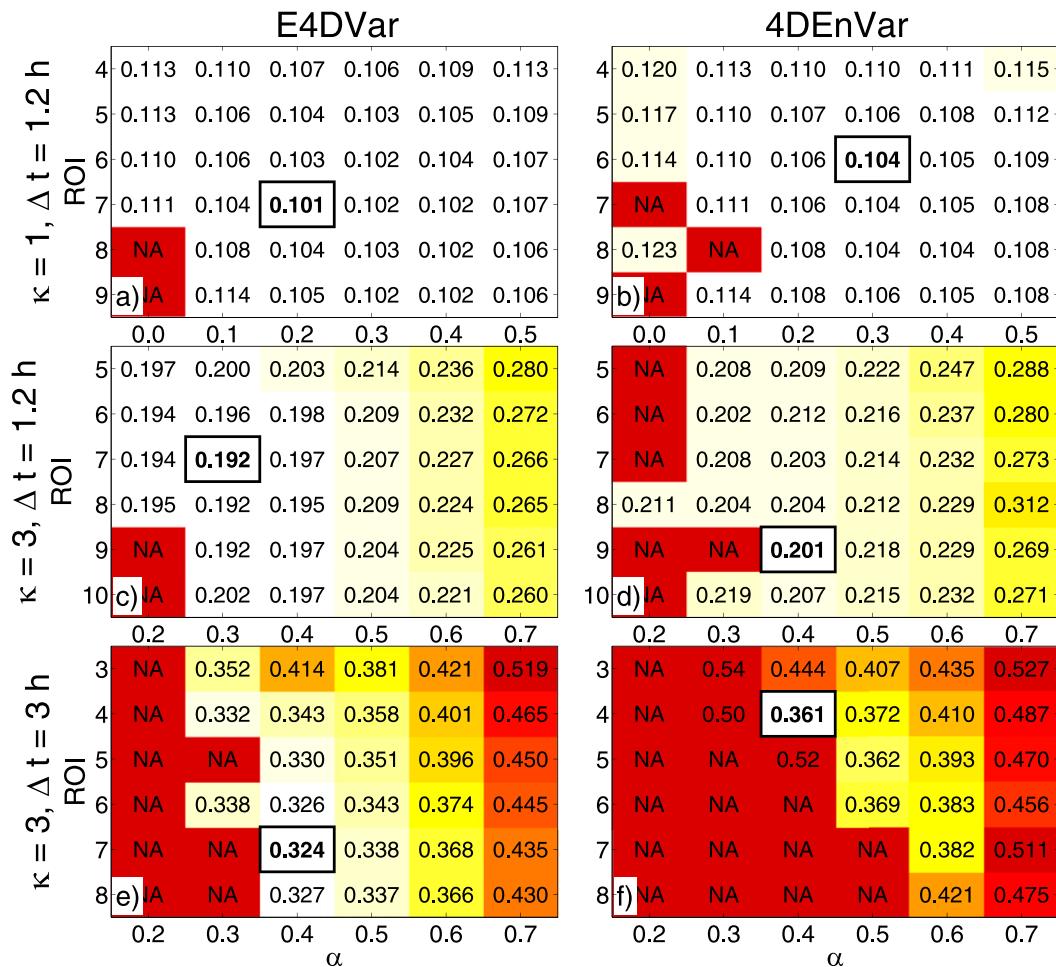


FIG. 9. Using an ensemble size of 20 and $T = 6 \text{ h}$, mean analysis RMSEs are estimated for a range of α and ROI for experiments assimilating observations (a),(b) of every variable ($\kappa = 1$); (c),(d) every third variable ($\kappa = 3$); and (e),(f) $\kappa = 3$ and every 3 h ($\Delta t = 3 \text{ h}$); the range of α and ROI values changes between each row. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

ensemble. Results from these experiments (Fig. 11) show that increasing F leads to larger analysis RMSEs in both data assimilation systems. The faster error growth also causes both methods to become increasingly less stable for large α ; significant posterior variance inflation is no longer needed to increase the prior variance for future data assimilation cycles. As hypothesized, the 4DEnVar analyses suffer a larger drop in accuracy, owing to faster error growth in the assimilation window.

e. Model error

In this section, we test E4DVar and 4DEnVar with an imperfect forecast model by keeping the true forcing term fixed at $F_i = 8$ and reducing the forcing term in the model used during prediction and assimilation steps to $F_m = 7$. For these experiments, we use an ensemble size of $N_e = 20$ and a window length of $T = 24 \text{ h}$, so applying

an incorrect forecast model over a long assimilation window causes the dominant system errors. In section 4b, the same configuration is applied with a perfect model to produce Figs. 7e and 7f. Figures 12a and 12b show results from experiments that apply E4DVar and 4DEnVar with $\beta = 0$ to find the optimal ROI and α for tuning the ensemble covariance without hybrid covariance. Both methods produce the lowest analysis RMSEs when a relatively small ROI (around 3–4) is used with an α of 0.6. Model error reduces the accuracy of 4D covariances approximated in the assimilation window, and covariance localization acts as the only mechanism for improving its estimation. The issues discussed in section 3 regarding time localization in 4DEnVar become more severe, owing to the use of a shorter ROI to manage sampling errors related to the imperfect model. As a result, E4DVar provides more

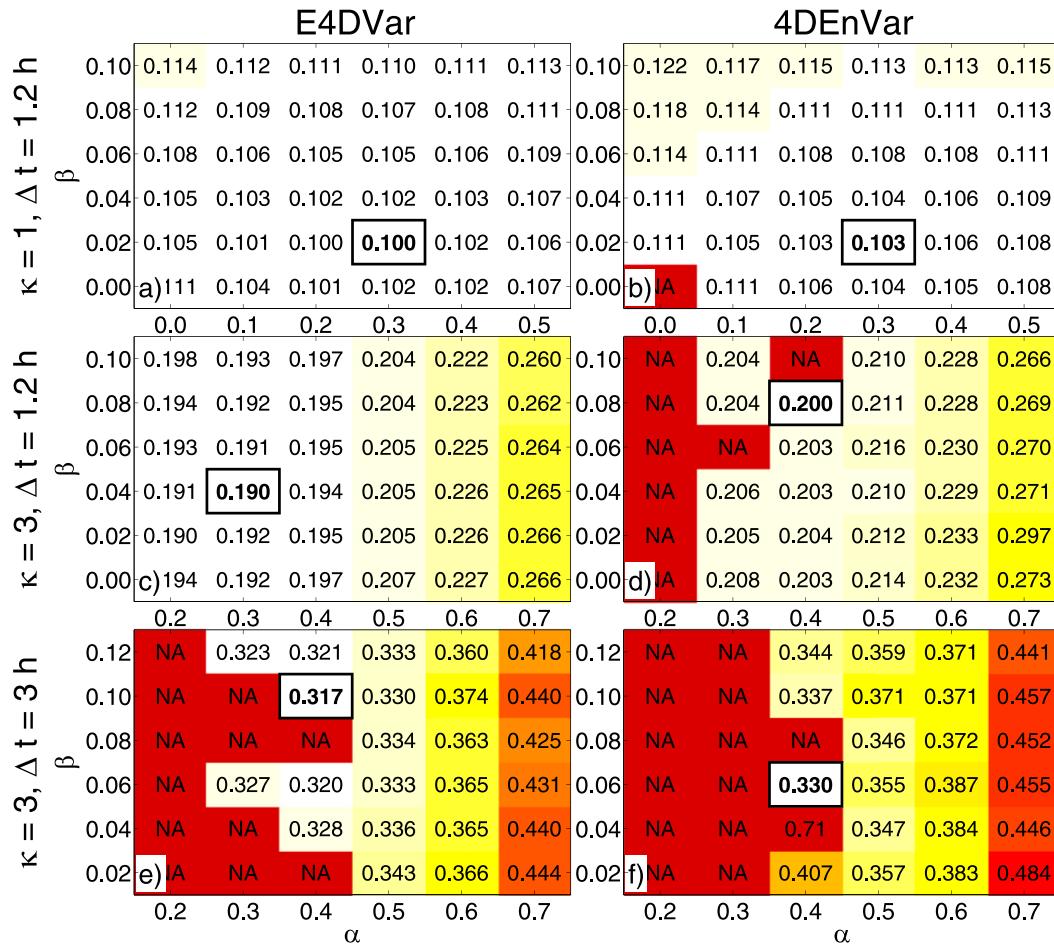


FIG. 10. Using an ensemble size of 20 and $T = 6 \text{ h}$, mean analysis RMSEs are estimated for a range of α and β for experiments assimilating observations (a),(b) of every variable ($\kappa = 1$); (c),(d) every third variable ($\kappa = 3$); and (e),(f) $\kappa = 3$ and every 3 h ($\Delta t = 3 \text{ h}$); the range of β and α values change between the first and second rows. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

accurate 4D covariances and lower analysis RMSEs than 4DEnVar when the system is subjected to model error. Fairbairn et al. (2014) reached a similar conclusion from their experiments comparing E4DVar and 4DEnVar with an imperfect model.

When a hybrid covariance is introduced to treat model error during data assimilation, E4DVar and 4DEnVar respond very differently than in experiments where sampling errors related to ensemble size is most dominant (see section 4a). Increasing β from 0.0 to 0.1 in the E4DVar experiment leads to a 30% reduction in analysis RMSEs (Fig. 12c), and removes most of the sensitivity to α . The hybrid covariance in E4DVar serves a similar purpose as additive inflation for EnKFs; \mathbf{B}^s introduces orthogonal directions for errors to grow in the assimilation window that would otherwise be missing from the ensemble component. 4DEnVar analyses,

however, show a much smaller decrease in RMSE when the static covariance is introduced (Fig. 12d). Similar to the nonhybrid cases, the hybrid implementation of 4DEnVar requires a large α for optimal performance, which suggests that the data assimilation benefits less from \mathbf{B}^s when it is fixed in time. Because we generate \mathbf{B}^s using exact forecast errors estimated from previous nonhybrid data assimilation experiments, both hybrid methods benefit from the inclusion of model error in the climatological covariance estimate. Our approach for generating a climatological covariance differs from atmospheric science applications, where \mathbf{B}^s is generated without any prior knowledge of the true state. As a result, we expect our data assimilation systems to yield more accurate analyses than would be found for a system with an unknown true state.

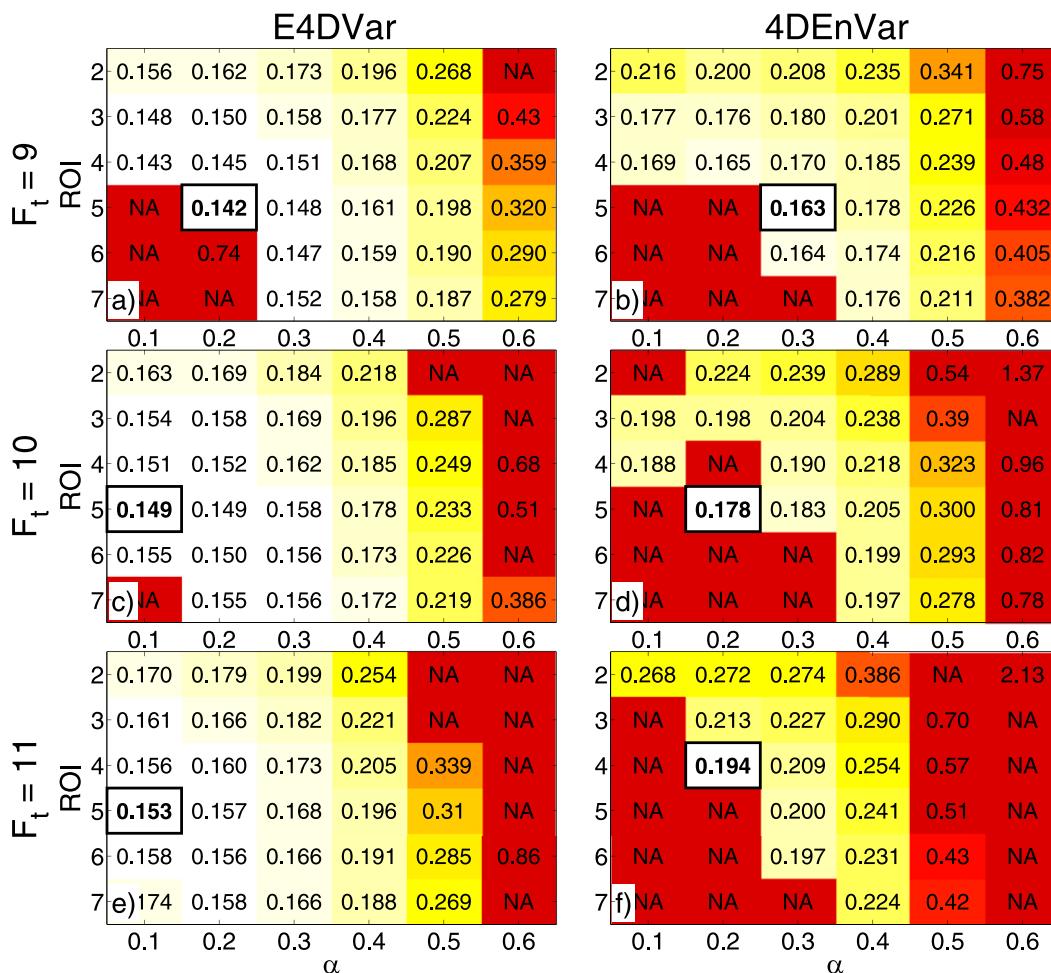


FIG. 11. Using an ensemble size of 10 and $T = 24$ h, mean analysis RMSEs are estimated for a range of α and ROI, and $F_e/F_i = 9, 10$, and 11. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

One means of propagating \mathbf{B}^s through the assimilation window in 4DEnVar is to replace the ensemble forecast perturbations at $t = 0$ with hybrid perturbations:

$$\mathbf{x}_{n,0}^{ff} \leftarrow \beta \mathbf{x}_{n,0}^{fs} + (1 - \beta) \mathbf{x}_{n,0}^{ff}, \quad (18)$$

where each $\mathbf{x}_{n,0}^{fs}$ is drawn randomly from $N(0, \mathbf{B}^s)$. This approach introduces additional noise in the covariance estimate, but the static contribution is no longer fixed in time. Figures 12e and 12f show results from E4DVar and 4DEnVar experiments in which the climatological covariance is introduced through hybrid perturbations, rather than using the full-rank \mathbf{B}^s matrix in the cost function. 4DEnVar clearly benefits from the hybrid perturbation method, as seen by the decrease in analysis RMSEs and smaller reliance on α . Nevertheless, E4DVar performs worse with the hybrid perturbations, owing to the fact that \mathbf{B}^s is replaced by a sample

estimate. This example demonstrates the advantage of using \mathbf{M} , to evolve \mathbf{B}^s through the assimilation window in E4DVar, and suggests that 4DEnVar may benefit more from other means of treating model error, such as additive inflation or stochastic physics in the ensemble.

f. Choice of analysis

For the data assimilation experiments described up to this point in the manuscript, the analysis is formed using the nonlinear model to propagate the analysis increment plus background field at $t = 0$ forward to the middle of the window at $t = \tau/2$:

$$\mathbf{x}_{\tau/2}^a \equiv M_{\tau/2}(\mathbf{x}_0^b + \delta \mathbf{x}_0). \quad (19)$$

E4DVar and 4DEnVar apply different strategies for propagating increments forward in time when minimizing the cost function. In E4DVar, increments at $t > 0$

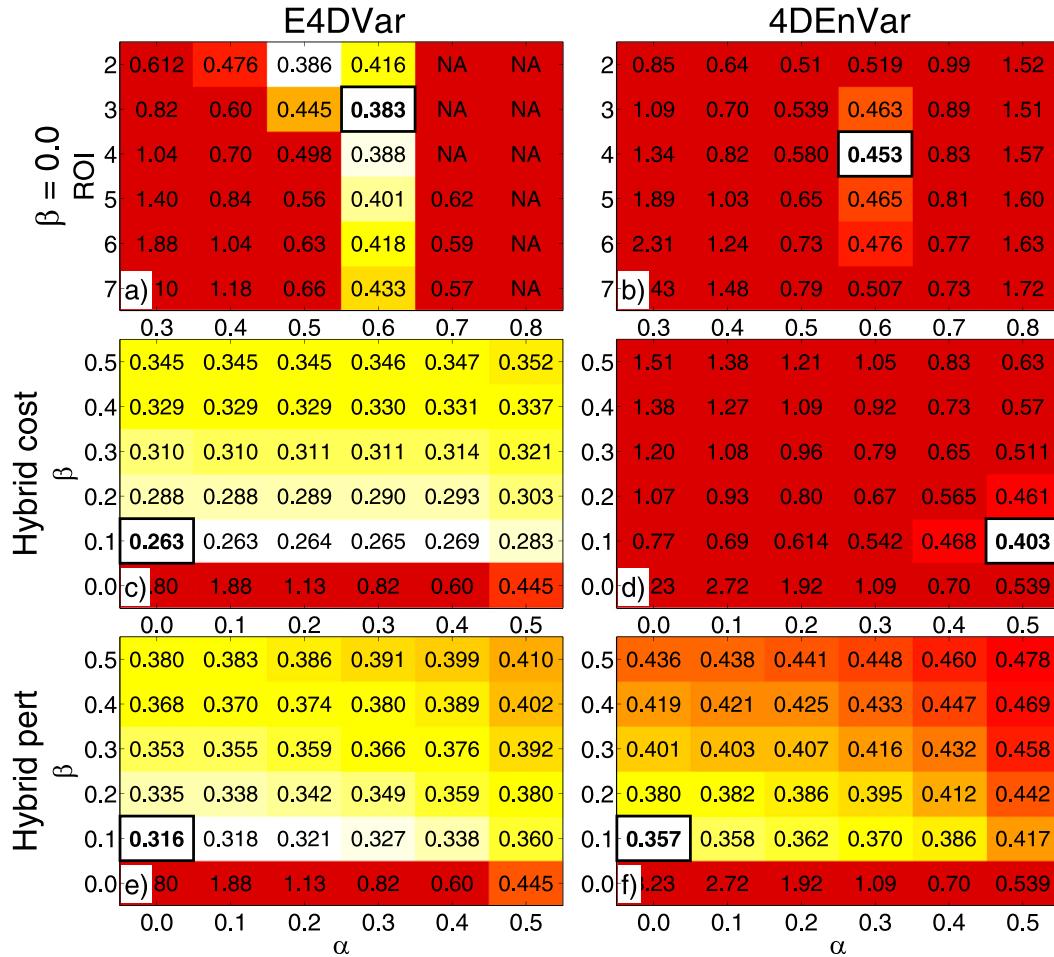


FIG. 12. (a),(b) Mean analysis RMSEs are estimated for a range of α and ROI for $N_e = 20$, $T = 24$ h, and an incorrect model forcing of $F_m = 7$. (c)–(f) ROI is fixed at 3, and RMSEs are estimated for ranges of α and β . Hybrid covariance is used in the standard cost function form in (c) and (d) and with ensemble perturbations in (e) and (f). Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

in the assimilation window are evolved in time using the tangent linear model, so the analysis at the middle of the window is given by

$$\mathbf{x}_{\tau/2}^a \equiv M_{\tau/2}(\mathbf{x}_0^b) + \mathbf{M}_{\tau/2} \delta \mathbf{x}_0. \quad (20)$$

The difference between (19) and (20) reflects the linear approximation made in the incremental 4DVar algorithm used in this study. In 4DEnVar, analysis increments at each time in the window come from multiplying the control variables by a matrix of evolved ensemble perturbations (i.e., $\delta \mathbf{x}_t = \mathbf{D}_t \mathbf{v}^{se}$). Therefore, the 4DEnVar analysis at the middle of the window is given by

$$\mathbf{x}_{\tau/2}^a \equiv M_{\tau/2}(\mathbf{x}_0^b) + \mathbf{D}_{\tau/2} \mathbf{v}^{se}. \quad (21)$$

Liu et al. (2009) formulate their 4DEnVar cost function so that the control variables are valid at the middle of the

assimilation window. This approach is identical to minimizing (11), then applying (21) to form the analysis.

To examine the sensitivity of E4DVar and 4DEnVar to the choice of analysis calculation, we revisit the experiments presented in Figs. 8e and 8f, which test the performance of both methods over a range of α and β values and a 24-h assimilation window. We apply E4DVar and 4DEnVar using the analyses defined by (20) and (21), respectively, and compare the results with experiments using (19). Mean analysis errors are plotted in Fig. 13 using “ $t = 0$ increments” to denote experiments that propagate the analysis from the beginning of the window using the nonlinear model, and “ $t = 12$ h increments” to denote experiments that add increments from the middle of the window to the background at this time. Figures 13a and 13c show a minor advantage of using the nonlinear model to propagate the analysis

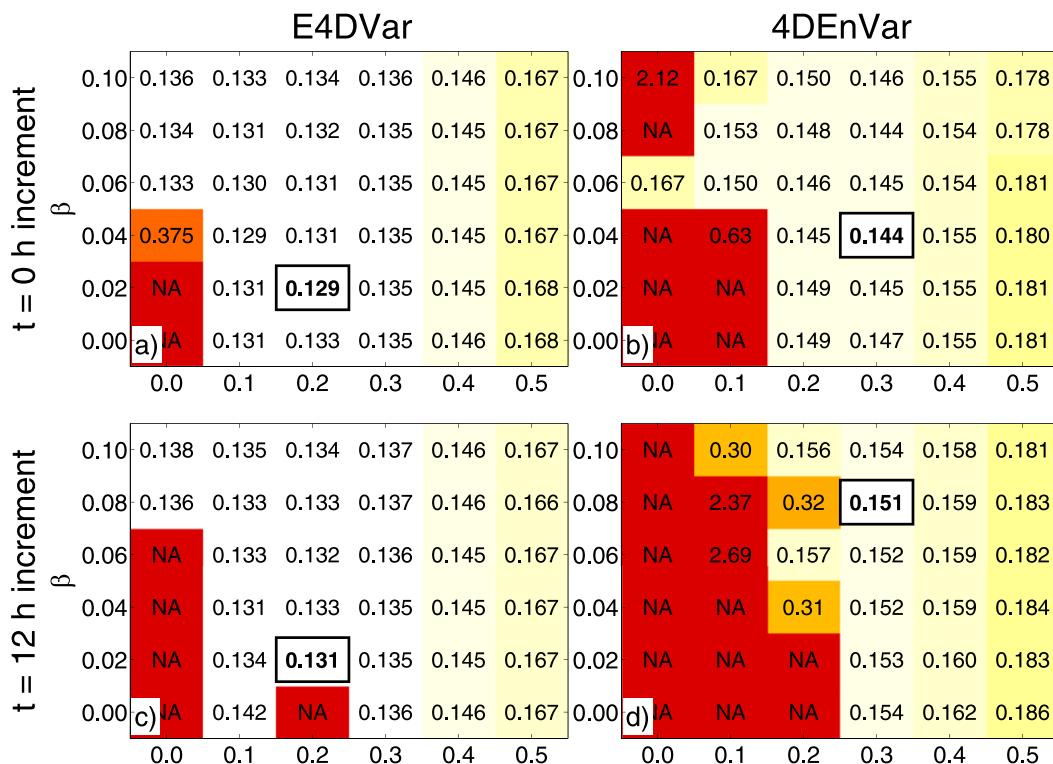


FIG. 13. Mean analysis RMSEs estimated for a range of α and β , and $T = 24$ h using (a),(b) analysis increments propagated forward with the nonlinear model from $t = 0$ h, and (c),(d) analysis increments estimated at $t = 12$ h and added to the background at this time. Ensemble size and ROI are fixed at 10 and the optimal value found in Fig. 7, respectively. Red shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box.

forward from $t = 0$ in the E4DVar case. The validity of the tangent linear approximation in this model leads to analysis differences that are statistically indistinguishable for many of the configurations tested. On the other hand, we find the analysis accuracy in the 4DEnVar case to be degraded when increments are taken from the middle of the window and added to the background at this time (Figs. 13b,d). For 4DEnVar, the advantage of using (19) to define the analysis instead of (21) is realized by substituting $\mathbf{D}_0 \mathbf{v}^{se}$ for $\delta \mathbf{x}_0$ in (19), and applying a first-order Taylor series approximation for the nonlinear model:

$$\mathbf{x}_{\tau/2}^a \equiv M_{\tau/2}(\mathbf{x}_0^b + \mathbf{D}_0 \mathbf{v}^{se}), \quad (22)$$

$$\approx M_{\tau/2}(\mathbf{x}_0^b) + \mathbf{M}_{\tau/2} \mathbf{D}_0 \mathbf{v}^{se}. \quad (23)$$

The term $\mathbf{M}_{\tau/2} \mathbf{D}_0$ in (23) is similar to the \mathbf{D}_t used in (21), except the square root of the background error covariance is integrated from $t = 0$ to $t = \tau/2$ using the tangent linear model instead of an ensemble forecast from the nonlinear model. Evidence provided in section 3 suggests that $\mathbf{M}_{\tau/2} \mathbf{D}_0$ provides a better approximation

of the background errors than $\mathbf{D}_{\tau/2}$, owing to differences in covariance localization and the inability of 4DEnVar to propagate the static covariance forward in time. Because the method that uses the nonlinear model to propagate an analysis from the beginning to middle of the window implicitly models the evolution of perturbations more accurately in time, it produces better analyses than the alternative approach.

5. Conclusions

Using a low-dimensional dynamical system, this study investigates the behavior of two hybrid 4D data assimilation systems currently being tested for operational use and for research applications in atmospheric science. One method, called E4DVar, is an extension of the traditional 4DVar method, and uses tangent linear and adjoint models to implicitly evolve a hybrid prior error covariance in the assimilation window. A second method, called 4DEnVar, uses a four-dimensional ensemble in place of the linearized models in 4DVar. Because of the absence of tangent linear and adjoint

models in 4D EnVar , the climatological component of the background error covariance matrix is fixed in time; therefore, the 4D EnVar method in this study operates as a hybrid between 4D EnVar and 3D Var-FGAT .

A fundamental difference between E4D Var and 4D EnVar is how temporal covariances are estimated in the assimilation window. 4D EnVar uses an ensemble forecast from the full nonlinear model to estimate the time evolution of prior error covariance. This approach is beneficial for situations where the tangent linear approximation is invalid, as is the case for physical parameterization schemes used frequently in atmospheric models. It also provides a means of performing 4D data assimilation in a manner that is highly parallel, unlike E4D Var , which requires an integration of the tangent linear and adjoint models during each iteration of the cost function minimization. Nevertheless, the ensemble representation of time covariances in 4D EnVar has disadvantages when a hybrid covariance is applied, or when localization is needed to treat sampling errors.

The L96 model is used to examine the performance of E4D Var and 4D EnVar under a number of model, data assimilation, and observing system configurations. Experiments performed using ensemble covariance during data assimilation yield results that agree with the findings of Buehner et al. (2010b) and Fairbairn et al. (2014) in that E4D Var consistently produces the lowest analysis errors. The two methods provide comparable analysis accuracy when the forecast model is perfect, and when dense observation networks and large ensembles are available; however, E4D Var performs significantly better for sparse observations, and in cases where small ensemble sizes and model error decrease the optimal localization length scale. E4D Var also produces significantly lower analysis errors than 4D EnVar when either the assimilation window length or model forcing term are increased; in both cases, the pure ensemble representation of covariances in the assimilation window is found to be less accurate than applying linearized models to evolve a localized covariance matrix in time.

The E4D Var method has additional advantages when a hybrid prior covariance is used in the cost function for imperfect model experiments. The ability of E4D Var to evolve time-dependent covariance from the climatological component in the assimilation window has the same effect as applying additive inflation in EnKFs, except without the extra sampling errors. We find that 4D EnVar benefits more from using hybrid perturbations—comprising samples from the ensemble forecast and climatological error distribution—rather than using the full climatological covariance matrix in the cost function. Hybrid perturbations allow the ensemble to evolve errors from the climatological covariance in time, but

introduce additional sampling noise. The inability of 4D EnVar to evolve the full static covariance forward in time, however, only appears to be an issue when the forecast model is imperfect. In perfect model experiments, E4D Var and 4D EnVar benefit equally from using static covariance in the cost function to improve the prior error covariance estimated from small ensembles, or in cases where observations are sparse.

One drawback of E4D Var not addressed in this study is the necessary use of a simplified forecast model when coding the tangent linear and adjoint operators for weather applications. Data assimilation experiments performed at national forecast centers, such as Environment Canada (Buehner et al. 2010b) and the Met Office (Lorenc et al. 2014), suggest that the positive qualities of E4D Var outweigh the negative effects of simplified linear operators in global models. For high-resolution limited-area models, Gustafsson and Bojarova (2014) provide evidence that suggests the opposite to be true. The relative performance of E4D Var and 4D EnVar for convective-scale applications, where errors in physical parameterization schemes are nontrivial, remains an open question. Despite the potential errors from simplified linear models in E4D Var , this method has advantages over 4D EnVar when short localization length scales are required, when forecast errors grow quickly in time, and when a hybrid covariance is necessary to compensate for model errors. Furthermore, alternative approaches to the hybrid methods tested in this study can be considered more effective for weather applications (e.g., replacing 3D Var-FGAT in 4D EnVar with 4D Var or E4D Var). These strategies would also combine the most computationally expensive components of each system, and require the coding of tangent linear and adjoint models.

Acknowledgments. This work was supported by the Office of Naval Research Grant N000140910526, the National Science Foundation Grant AGS-1305798, and NASA Grant NNX12AJ79G. The authors thank Steven Greybush from The Pennsylvania State University for many helpful discussions during the course of this work, and three anonymous reviewers for their insightful comments.

REFERENCES

- Bishop, C. H., and D. Hodyss, 2011: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Wea. Rev.*, **139**, 1241–1255, doi:10.1175/2010MWR3403.1.
- , and E. A. Satterfield, 2013: Hidden error variance theory. Part I: Exposition and analytic model. *Mon. Wea. Rev.*, **141**, 1454–1468, doi:10.1175/MWR-D-12-00118.1.
- Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in

- a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, doi:10.1256/qj.04.15.
- , P. L. Houtekamer, C. Charette, H. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550–1566, doi:10.1175/2009MWR3157.1.
- , —, —, —, and —, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.*, **138**, 1567–1586, doi:10.1175/2009MWR3158.1.
- Clayton, A. M., A. C. Lorenc, and D. M. 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.
- Courtier, P., J.-N. Thepát, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, doi:10.1002/qj.49712051912.
- Etherton, B. J., and C. H. Bishop, 2004: Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.*, **132**, 1065–1080, doi:10.1175/1520-0493(2004)132<1065:ROHDAS>2.0.CO;2.
- 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.
- Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone, 2014: A comparison of 4DVar with ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.*, **140**, 281–294, doi:10.1002/qj.2135.
- Fisher, M., and E. Andersson, 2001: Developments in 4D-Var and Kalman filtering. ECMWF Tech. Memo. 347, ECMWF, 36 pp.
- Gustafsson, N., and J. Bojarova, 2014: Four-dimensional ensemble variational (4D-En-Var) data assimilation for the High Resolution Limited Area Model (HIRLAM). *Nonlinear Processes Geophys.*, **21**, 745–762, doi:10.5194/npg-21-745-2014.
- Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKVF>2.0.CO;2.
- Honda, Y., M. Nishijima, K. Koizumi, Y. Ohta, K. Tamiya, T. Kawabata, and T. Tsuyuki, 2005: A pre-operational variational data assimilation system for a non-hydrostatic model at the Japan Meteorological Agency: Formulation and preliminary results. *Quart. J. Roy. Meteor. Soc.*, **131**, 3465–3475, doi:10.1256/qj.05.132.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Huang, X., and Coauthors, 2009: Four-dimensional variational data assimilation for WRF: Formulation and preliminary results. *Mon. Wea. Rev.*, **137**, 299–314, doi:10.1175/2008MWR2577.1.
- 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.
- Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110, doi:10.1111/j.1600-0870.1986.tb00459.x.
- Liu, C., and Q. Xiao, 2013: An ensemble-based four-dimensional variational data assimilation scheme. Part III: Antarctic applications with Advanced Research WRF using real data. *Mon. Wea. Rev.*, **141**, 2721–2739, doi:10.1175/MWR-D-12-00130.1.
- , —, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136**, 3363–3373, doi:10.1175/2008MWR2312.1.
- , —, and —, 2009: An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing System Simulation Experiments with Advanced Research WRF (ARW). *Mon. Wea. Rev.*, **137**, 1687–1704, doi:10.1175/2008MWR2699.1.
- Lorenc, A. C., 1997: Development of an operational variational assimilation scheme. *J. Meteor. Soc. Japan*, **75**, 339–346.
- , 2003: The potential of the ensemble Kalman filter for NWP: A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, doi:10.1256/qj.02.132.
- , and Coauthors, 2000: The Met Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **126**, 2991–3012, doi:10.1002/qj.49712657002.
- , N. Bowler, A. Clayton, and S. Pring, 2014: Development of the Met Office's 4D-EnVar system. *Proc. Sixth EnKF Workshop*, Buffalo, NY, Met Office. [Available online at hftp.psu.edu/fuz4/EnKF2014/EnKF-Day1/Lorenc_4D-EnVar.pptx.]
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Vol. 1, Reading, United Kingdom, ECMWF, 1–18.
- , 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.
- Poterjoy, J., and F. Zhang, 2014: Intercomparison and coupling of ensemble and four-dimensional variational data assimilation methods for the analysis and forecasting of Hurricane Karl (2010). *Mon. Wea. Rev.*, **142**, 3347–3364, doi:10.1175/MWR-D-13-00394.1.
- Sun, J., and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.*, **54**, 1642–1661, doi:10.1175/1520-0469(1997)054<1642:DAMRFD>2.0.CO;2.
- Thepát, J.-N., and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, **117**, 1225–1254, doi:10.1002/qj.49711750206.
- , —, G. Belaud, and G. Lamaître, 1996: Dynamical structure functions in a four-dimensional variational assimilation: A case study. *Quart. J. Roy. Meteor. Soc.*, **122**, 535–561, doi:10.1002/qj.49712253012.
- Tian, X., Z. Xie, and A. Dai, 2008: An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.*, **113**, D21124, doi:10.1029/2008JD010358.
- , —, —, C. Shi, B. Jia, F. Chen, and K. Yang, 2009: A dual-pass variational data assimilation framework for estimating soil moisture profiles from AMSR-E microwave brightness temperature. *J. Geophys. Res.*, **114**, D16102, doi:10.1029/2008JD011600.
- Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble-variational (4D-EnVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, doi:10.1175/MWR-D-13-00303.1.
- , C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble-3DVar hybrid

- analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227, doi:[10.1175/MWR3282.1](https://doi.org/10.1175/MWR3282.1).
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:[10.1175/1520-0493\(2004\)132<1238:IOIEAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2).
- , M. Zhang, and J. A. Hansen, 2009: Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **26**, 1–8, doi:[10.1007/s00376-009-0001-8](https://doi.org/10.1007/s00376-009-0001-8).
- , —, and J. Poterjoy, 2013: E3DVar: Coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar. *Mon. Wea. Rev.*, **141**, 900–917, doi:[10.1175/MWR-D-12-00075.1](https://doi.org/10.1175/MWR-D-12-00075.1).
- Zhang, M., and F. Zhang, 2012: E4DVar: Coupling an ensemble Kalman filter with four-dimensional data assimilation in a limited-area weather prediction model. *Mon. Wea. Rev.*, **140**, 587–600, doi:[10.1175/MWR-D-11-00023.1](https://doi.org/10.1175/MWR-D-11-00023.1).