

A Multiscale Alignment Method for Ensemble Filtering with Displacement Errors

YUE YING

Advanced Study Program, National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 29 May 2019, in final form 23 July 2019)

ABSTRACT

High-resolution models nowadays simulate phenomena across many scales and pose challenges to the design of efficient data assimilation methods that reduce errors at all scales. Smaller-scale features experience rapid nonlinear error growth that gives rise to displacement errors, which cause suboptimal ensemble filter performance. Previous studies have started exploring methods that can reduce displacement errors. In this study, a multiscale alignment (MSA) method is proposed for ensemble filtering. The MSA method iteratively processes the model state from the largest to the smallest scales. At each scale, an ensemble filter is applied to update the state with observations, and the analysis increments are utilized to derive displacement vectors for each member that align the ensemble at smaller scales before the next iteration. The nonlinearity in smaller-scale priors is reduced by removing larger-scale displacement errors. Because the displacement vectors are derived from analysis increments in the state space rather than the nonlinear observation-space cost function formulated in previous studies, this method provides a less costly and more robust way to solve for the displacement vectors. Observing system simulation experiments using a two-layer quasigeostrophic model were conducted to provide a proof of concept of the MSA method. Results show that the MSA method significantly improves the accuracy of posteriors compared to the existing ensemble filter methods with or without multiscale localization. Advantage of the MSA method are more evident when the ensemble size is relatively small and the cycling period is comparable to the average eddy turnover time of the dynamical system.

1. Introduction

Data assimilation (DA) seeks to reduce the errors in model initial conditions with information derived from observations. As model resolution increases, model states can resolve features across an increasing range of scales. High dimensionality and nonlinearity of model states pose challenges to the design of an efficient DA method that can simultaneously reduce errors at all resolved scales. One major difference between large- and small-scale errors is their different physical correlation scales in space and time. Smaller-scale features, such as individual convective systems in atmospheric models, tend to experience faster temporal evolution and more rapid error growth, therefore having shorter correlation lengths than larger-scale features. DA methods need to adapt to different error characteristics across scales, which motivated the investigation of multiscale methods in previous studies (e.g., Li et al. 2015). In ensemble-based DA methods such as the ensemble Kalman filter (Evensen 1994), localization

techniques are used to reduce the impact of sampling errors in correlations due to limited-size ensemble (Hamill et al. 2001). Multiscale localization methods were also proposed to adapt to different spatial scales in error covariance (Zhang et al. 2009; Miyoshi and Kondo 2013; Buehner and Shlyayeva 2015).

Another important difference in error characteristics between large and small scales is the fact that the small-scale errors are often non-Gaussian due to nonlinear error growth. The nonlinearity often manifests as displacement errors that are errors in the position of coherent features. Studies have shown that even if the errors in feature position are Gaussian, the resulting errors of the state can be non-Gaussian, which cause ensemble Kalman filters to perform suboptimally (Lawson and Hansen 2005; Chen and Snyder 2007). For relatively small-scale features in atmospheric models, such as hurricanes and mesoscale convective and cloud systems, these displacement errors can be challenging for DA. For example, position errors among prior ensemble members can lead to some members being cloud free while others being cloudy, which leads to a fat-tail prior error distribution (Geer and Bauer 2011).

Corresponding author: Yue Ying, mying@ucar.edu

DOI: 10.1175/MWR-D-19-0170.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

Substantial displacement errors due to model bias can limit the performance of convective-scale DA (e.g., Xue et al. 2014; Dawson et al. 2015; Yussouf et al. 2015).

To alleviate the negative impact of displacement errors, methods that reduce position errors from the prior ensemble have been investigated. For hurricane DA, the relocation of the hurricane vortex was proposed as a preprocessing step to reduce position errors (Hsiao et al. 2010). In the image processing literature, methods have been studied for deriving the displacement vectors, also known as an “optical flow,” between two frames in a movie (e.g., Horn and Schunck 1981). The optical flow is typically derived by minimizing the misfit between the target image and the source image displaced by the optical flow. Extending the optical flow idea to the problem of DA, the feature calibration and alignment (FCA) technique (Hoffman et al. 1995; Hoffman and Grassotti 1996; Nehrkorn et al. 2003) was introduced to partition prior errors into contributions from position and amplitude errors, which can then be separately evaluated. DA methods that explicitly estimate and reduce displacement errors were proposed (Brewster 2003; Ravela et al. 2007; Beezley and Mandel 2008), and implemented in the variational framework (Nehrkorn et al. 2014, 2015), and more recently in the ensemble-based method (Stratman et al. 2018). Typically, the FCA technique has a two-step formulation. First, the displacement vectors are estimated by minimizing a cost function, which contains a misfit term between the observations and the modeled observation from the displaced prior, as well as some penalty terms that impose physical constraints to the displacement vectors. Second, the displacement vectors are applied to each member to align the prior ensembles, and a normal DA step ensues.

The observation-space nonlinear cost function formulation currently adopted by the FCA technique faces some challenges when applying to atmospheric DA. Since the model state has large dimension and the small-scale features are highly nonlinear, the cost function can be ill-conditioned with complex topology and local minima. Observations that are sparse in space and nonlinearly related to the state, a common scenario in atmospheric DA, also complicate the nonlinear cost function. On the other hand, atmospheric DA methodology is much advanced in terms of dealing with the nonlinear relation between the state and observations. For example, the localized particle filters (e.g., Poterjoy 2016; Poterjoy et al. 2017) have demonstrated skill in handling nonlinear observations in high-dimensional systems. It will be favorable if the existing nonlinear filtering methods can be exploited

to solve the displacement problem. For displacements larger than the sizes of features that cause nonlinearity, a multiscale strategy was proposed as a solution in the image processing literature (Meinhardt-Llopis et al. 2013), where the optical flows are first derived using coarsened images, and iteratively refined with more details from images with gradually increasing resolution.

In this study, a multiscale alignment (MSA) method is proposed for reducing displacement errors in ensemble-based DA. The multiscale strategy is adopted to iteratively process the model state from large to small scales. At each scale, an ensemble filter is first applied to update the prior state, the resulting analysis increment is then utilized to derive the displacement vector for that scale. The larger-scale displacement vectors are applied to the smaller-scale state to align the prior ensemble before the next iteration. The ensemble filter handles the relation between the state and the observations, and the displacement vectors are derived in a similar fashion to the optical flow between two images, which is solving the problem in the state space as opposed to the nonlinear observation space formulation of FCA. The ability of large-scale displacement vectors to correct small-scale position errors depends on the physical relation between the prior errors at these scales, which is modeled as a cross-scale error covariance in multiscale DA. The alignment using large-scale displacements will reduce the small-scale errors if the prior errors are well correlated across scales, which is often the case for small-scale atmospheric features to have position errors due to systematic errors in large-scale background flows.

The paper is organized as follows. Section 2 provides a complete mathematical formulation of the proposed MSA method. Section 3 describes the design of numerical experiments to test the method and compare it to the existing multiscale localization and ensemble filtering methods. Section 4 shows some results as a proof of concept of the MSA method, and section 5 summarizes the findings of this study.

2. The multiscale alignment method

a. Ensemble filtering with treatment for displacement errors

Let \mathbf{X} be the state variables of the dynamical system of interest, and \mathbf{X}^* denotes the true state. The observation of the state is defined as

$$\mathbf{Y} = H(\mathbf{X}^*) + \boldsymbol{\varepsilon}^o, \quad (1)$$

where H is an observation operator that maps state space to observation space, and $\boldsymbol{\varepsilon}^o$ is an observation error that reflects uncertainties in measurements.

Here, the observation error is assumed to follow a Gaussian distribution, $\boldsymbol{\varepsilon}^o \sim \mathcal{N}(0, \mathbf{R})$, where \mathbf{R} is the observation error covariance. DA seeks to update a prior estimate of the state \mathbf{X}^b using the information obtained from the observation to provide a better estimate called the analysis \mathbf{X}^a . Here, the superscripts b and a denotes background (prior) and analysis (posterior) states, respectively. Ensemble filtering methods treat the DA problem as a collection of analysis cycles in time. At each cycle, the observations near that time are assimilated, and the resulting analysis is used to initialize model forecasts that provide the prior for the next cycle.

Consider the coherent features in the state of a dynamical system (e.g., a cyclone, a line of organized convection), the uncertainties in the prior estimate can be viewed as due to *displacement errors* (errors in position) $\boldsymbol{\varepsilon}^d$ or *amplitude errors* in the feature $\boldsymbol{\varepsilon}^r$ (also called a residual error). The prior state can thus be written as

$$\mathbf{X}^b = \mathbf{X}^* + \boldsymbol{\varepsilon}^d + \boldsymbol{\varepsilon}^r. \quad (2)$$

To describe position errors of features in the state, a set of *displacement vectors* can be used. In this study, the displacement vectors $\mathbf{q} = (\mathbf{u}, \mathbf{v})$ are only considered in the two horizontal directions (x and y). The model state \mathbf{X} is defined on a spatially discretized model grid that has n_x grid points indexed by i . The i th grid point has horizontal coordinates (x_i, y_i) and displacement vector $q_i = (u_i, v_i)$. Let $\mathbf{X}(\mathbf{q})$ denote the state \mathbf{X} displaced by \mathbf{q} . The i th grid point is displaced from (x_i, y_i) to $(x_i - u_i, y_i - v_i)$, and the value for $\mathbf{X}(\mathbf{q})$ can be found by horizontal interpolation from the original \mathbf{X} . The displacement error can be expressed as $\boldsymbol{\varepsilon}^d = \mathbf{X}^b - \mathbf{X}^b(\mathbf{q})$, and the residual error $\boldsymbol{\varepsilon}^r = \mathbf{X}^b(\mathbf{q}) - \mathbf{X}^*$. The total prior error is $\boldsymbol{\varepsilon}^b = \boldsymbol{\varepsilon}^d + \boldsymbol{\varepsilon}^r$. The residual error $\boldsymbol{\varepsilon}^r$ is assumed to follow a Gaussian distribution, $\boldsymbol{\varepsilon}^r \sim \mathcal{N}[0, \mathbf{B}(\mathbf{q})]$, where its error covariance $\mathbf{B}(\mathbf{q})$ depends on the displacement \mathbf{q} . The displacement error $\boldsymbol{\varepsilon}^d$, however, is only nearly Gaussian when the size of displacement is small compared to the horizontal scale of displaced features. When displacement is large, $\mathbf{X}^b(\mathbf{q})$ becomes a nonlinear function of the displacement that gives rise to a non-Gaussian $\boldsymbol{\varepsilon}^d$ distribution, which will cause linear Gaussian filtering methods to be suboptimal.

Following Ravela et al. (2007), an ensemble filtering scheme that explicitly treats large displacement errors can be described as follows. According to Bayes's rule, the joint posterior distribution of \mathbf{X} and \mathbf{q} given the observations \mathbf{Y} can be written as

$$p(\mathbf{X}, \mathbf{q} | \mathbf{Y}) \propto p(\mathbf{Y} | \mathbf{X}, \mathbf{q}) p(\mathbf{X} | \mathbf{q}) p(\mathbf{q}). \quad (3)$$

The observation likelihood can be written as

$$p(\mathbf{Y} | \mathbf{X}, \mathbf{q}) = (2\pi)^{-n_y/2} |\mathbf{R}|^{-1/2} \exp\left\{-\frac{1}{2} \|\mathbf{Y} - H[\mathbf{X}(\mathbf{q})]\|_{\mathbf{R}}^2\right\}, \quad (4)$$

where n_y is the number of observations and $|\mathbf{R}|$ is the determinant of \mathbf{R} . Here, the notation $\|\mathbf{a}\|_{\mathbf{A}}^2$ is a shorthand for $\mathbf{a}^T \mathbf{A}^{-1} \mathbf{a}$, which is the squared norm of a vector \mathbf{a} normalized by a covariance matrix \mathbf{A} . The distribution $p(\mathbf{X} | \mathbf{q})$ characterizes the residual error $\boldsymbol{\varepsilon}^r$ given the displacement \mathbf{q} , which is also known as the *amplitude distribution*:

$$p(\mathbf{X} | \mathbf{q}) = (2\pi)^{-n_x/2} |\mathbf{B}(\mathbf{q})|^{-1/2} \exp\left\{-\frac{1}{2} \|\mathbf{X}(\mathbf{q}) - \mathbf{X}^b(\mathbf{q})\|_{\mathbf{B}(\mathbf{q})}^2\right\}. \quad (5)$$

The distribution $p(\mathbf{q})$ describes the uncertainty of the displacement, which is also called the *displacement distribution*. Typically, a function $L(\mathbf{q})$ can be formulated so that it increases as the displacement deviates from some physical constraints, and the displacement distribution can be expressed as $p(\mathbf{q}) \propto \exp[-L(\mathbf{q})]$. For example, a smoothness constraint is written as

$$L(\mathbf{q}) = \sum_{i=1}^{n_x} \alpha^2 \left[\left(\frac{\partial u_i}{\partial x}\right)^2 + \left(\frac{\partial u_i}{\partial y}\right)^2 + \left(\frac{\partial v_i}{\partial x}\right)^2 + \left(\frac{\partial v_i}{\partial y}\right)^2 \right], \quad (6)$$

where α^2 is a weighting factor that controls the strength of the constraint. With all terms in (3) explicitly defined, a cost function can be formulated by taking its negative logarithm and omitting the constant term, which yields

$$J(\mathbf{X}, \mathbf{q}) = \frac{1}{2} \|\mathbf{Y} - H[\mathbf{X}(\mathbf{q})]\|_{\mathbf{R}}^2 + \frac{1}{2} \|\mathbf{X}(\mathbf{q}) - \mathbf{X}^b(\mathbf{q})\|_{\mathbf{B}(\mathbf{q})}^2 + \frac{1}{2} \ln(|\mathbf{B}(\mathbf{q})|) + L(\mathbf{q}). \quad (7)$$

The analysis is $\mathbf{X}^a = \mathbf{X}'(\mathbf{q}')$, where \mathbf{X}' and \mathbf{q}' are found at the minimum of the cost function, which can also be expressed in terms of increments from the prior:

$$\mathbf{X}^a = \mathbf{X}^b(\mathbf{q}') + \delta \mathbf{X}^r = \mathbf{X}^b + \delta \mathbf{X}^d + \delta \mathbf{X}^r. \quad (8)$$

The *displacement increment* $\delta \mathbf{X}^d = \mathbf{X}^b(\mathbf{q}') - \mathbf{X}^b$ is the result from displacing the prior with displacement \mathbf{q}' , and the *amplitude increment* $\delta \mathbf{X}^r = \mathbf{X}'(\mathbf{q}') - \mathbf{X}^b(\mathbf{q}')$ is the additional amplitude adjustment made to the displaced field.

To obtain an estimate of the flow-dependent error covariance, a set of ensemble forecasts can be employed to compute sample estimates of statistics of the prior distribution (Evensen 1994). Each ensemble member

provides a different realization of the prior \mathbf{X}^b with different displacement and residual errors. Estimates of the first two moments of the prior distribution can be derived from the ensemble mean and covariance. The goal of ensemble filtering with displacement errors is to update each member with increments \mathbf{q}' and $\delta\mathbf{X}^r$ so that the resulting posterior ensemble becomes samples from the posterior distribution (3), and this can be done by minimizing (7) for each member with \mathbf{X}^b from the particular member and $\mathbf{B}(\mathbf{q})$ estimated from the ensemble.

b. Iterative cost function minimization with multiscale strategy

Direct minimization of the cost function in (7) is difficult, because $\mathbf{X}(\mathbf{q})$ is a highly nonlinear function of \mathbf{q} and the dependency of \mathbf{B} on \mathbf{q} complicates the calculation of gradients of the cost function. For high-dimensional systems, the linearization of the cost function around a reference point is necessary to allow efficient computation. When the global cost function minimum is too far away from the prior estimate, the analysis can be found iteratively with a linearization of the cost function around successive solutions during each iteration. Employing a multiscale strategy during iteration can potentially improve the convergence rate. Similar strategies are used in the incremental formulation of the four-dimensional variational method (Courtier et al. 1994), where the minimization of the cost function is performed at gradually increasing model resolutions.

The MSA method adopts a multiscale strategy that treats the state as a sum of different scales, and the solution for \mathbf{q}' and $\delta\mathbf{X}^r$ are found iteratively at scale $s = 1, 2, \dots, n_s$ from the largest to the smallest scales. Scale decomposition can be performed through a series of spectral filters, which are defined as transform matrices \mathbf{F}_n that map the state \mathbf{X} to its scale s component $\mathbf{X}_s = \mathbf{F}_s\mathbf{X}$. The sum of all \mathbf{F}_s gives the identity matrix. Similarly, $\mathbf{q}'_s = (\mathbf{F}_s\mathbf{u}', \mathbf{F}_s\mathbf{v}')$ are the displacement vectors at scale s .

Let $\mathbf{X}^{(s)}$ denote the state before the s th iteration. Initially, $\mathbf{X}^{(1)} = \mathbf{X}^b$ is set for each member. The analysis step at the s th iteration calculates increments \mathbf{q}'_s and $\delta\mathbf{X}_s^r$, and adjusts the state as

$$\mathbf{X}^{(s+1)} = \mathbf{X}^{(s)}(\mathbf{q}'_s) + \delta\mathbf{X}_s^r. \quad (9)$$

The state $\mathbf{X}^{(s)}$ has incorporated all the previous increments at scales 1 to $s - 1$. At scale s , for larger scales $\mathbf{X}_1^{(s)}, \dots, \mathbf{X}_{s-1}^{(s)}$, the displacement \mathbf{q}'_s can be omitted, because the resulting displacement increments are approximately Gaussian and can be accounted for by the amplitude increments instead. For smaller scales

$\mathbf{X}_{s+1}^{(s)}, \dots, \mathbf{X}_{n_s}^{(s)}$, however, the displacement \mathbf{q}'_s needs to be explicitly applied to the state through interpolation due to nonlinearity. For the $\mathbf{X}_s^{(s)}(\mathbf{q}'_s)$ term, since \mathbf{q}'_s and $\mathbf{X}_s^{(s)}$ are of the same scale, a first-order Taylor expansion gives

$$\mathbf{X}_s^{(s+1)} = \mathbf{X}_s^{(s)} + \nabla\mathbf{X}_s^{(s)} \cdot \mathbf{q}'_s + \delta\mathbf{X}_s^r. \quad (10)$$

The MSA method finds \mathbf{q}'_s and $\delta\mathbf{X}_s^r$ in two steps. The first step is a *DA step* that determines the combined increment $\delta\mathbf{X}_s = \nabla\mathbf{X}_s^{(s)} \cdot \mathbf{q}'_s + \delta\mathbf{X}_s^r$ through assimilating observations. The second step is an *alignment step* that uses $\delta\mathbf{X}_s$ to determine the corresponding displacement \mathbf{q}'_s . After finding the increments, the adjusted full state becomes

$$\mathbf{X}^{(s+1)} = \sum_{r=1}^s \mathbf{X}_r^{(s)} + \delta\mathbf{X}_s + \sum_{r=s+1}^{n_s} \mathbf{X}_r^{(s)}(\mathbf{q}'_s). \quad (11)$$

The DA step uses the observations \mathbf{Y} to adjust the state at scale s , which is equivalent to modifying the cost function in (7) to only allow changes in $\delta\mathbf{X}_s$ and holding everything else constant:

$$J(\delta\mathbf{X}_s) = \frac{1}{2}\|\mathbf{Y} - H(\mathbf{X}^{(s)} + \delta\mathbf{X}_s)\|_{\mathbf{R}}^2 + \frac{1}{2}\|\delta\mathbf{X}_s\|_{\mathbf{B}(\mathbf{q})}^2 + C. \quad (12)$$

Here, the error covariance $\mathbf{B}(\mathbf{q})$ characterizes the residual errors after removing displacement errors from the previous scales, which can be estimated from the ensemble of $\mathbf{X}_s^{(s)}$. The displacement increment $\nabla\mathbf{X}_s^{(s)} \cdot \mathbf{q}'_s$ is relatively Gaussian since \mathbf{q}'_s is of the same scale as the state $\mathbf{X}_s^{(s)}$. An ensemble filter will compute the analysis increment as

$$\delta\mathbf{X}_s = \text{cov}[\mathbf{X}_s^{(s)}, H(\mathbf{X}^{(s)})] \{ \text{cov}[H(\mathbf{X}^{(s)})] + \mathbf{R} \}^{-1} [\mathbf{Y} - H(\mathbf{X}^{(s)})], \quad (13)$$

where $\text{cov}\{\mathbf{X}_s^{(s)}, H[\mathbf{X}^{(s)}]\}$ is the covariance between the observation prior and the prior state filtered at scale s , and $\text{cov}\{H[\mathbf{X}^{(s)}]\}$ is the covariance between the observation prior and itself; both covariances are estimated from the ensemble. There are many variants of ensemble filters available for solving (13). More details on the particular choice of ensemble filter and its parameters will be provided in the next section.

The alignment step determines for each ensemble member the partitioning of the total analysis increment $\delta\mathbf{X}_s$ into displacement and amplitude parts. Since this partitioning does not change $\delta\mathbf{X}_s$, the observation misfit term (first term) in cost function in (7) remains unchanged. Both the residual error term (second term) and the third term in (7) will decrease as the displacement \mathbf{q}'_s

replaces part of $\delta \mathbf{X}_s$ and the residual error characterized by $\mathbf{B}(\mathbf{q})$ decreases. The modified cost function for finding \mathbf{q}'_s becomes

$$J(\mathbf{q}'_s) = \|\mathbf{X}_s^{(s+1)} - \mathbf{X}_s^{(s)}(\mathbf{q}'_s)\|^2 + L(\mathbf{q}'_s), \quad (14)$$

where the first term can be further expressed as

$$\|\delta \mathbf{X}_s - \nabla \mathbf{X}_s^{(s)} \cdot \mathbf{q}'_s\|^2 = \sum_{i=1}^{n_x} \left(\delta X_{s,i} - \frac{\partial X_{s,i}^{(s)}}{\partial x} u'_{s,i} - \frac{\partial X_{s,i}^{(s)}}{\partial y} v'_{s,i} \right)^2, \quad (15)$$

and the second term takes the form of (6). The same type of cost function was formulated for finding optical flows in Horn and Schunck (1981), and can be solved by the algorithm provided therein. Thanks to the comparable scale between \mathbf{q}'_s and $\mathbf{X}_s^{(s)}$ that allows the linearization of the $\mathbf{X}_s^{(s)}(\mathbf{q}'_s)$ term, the Euler–Lagrange equation for cost function in (14) results in a Poisson equation for \mathbf{q}'_s , which is less costly to solve than the original nonlinear cost function in (7).

After finishing all iterations, the final updated version of the full model state $\mathbf{X}^{(n_s+1)}$ becomes the analysis \mathbf{X}^a .

3. Numerical experiment design

A multiscale DA test system is built to facilitate numerical experiments that evaluate the performance of the proposed multiscale alignment method. The two-layer quasigeostrophic (QG) model (Smith et al. 2002) is chosen as the dynamical system. This model describes the two-dimensional vorticity dynamics for geophysical flows with baroclinic instability as a source for kinetic energy. Although much simpler than the atmospheric models used in real weather forecast, this nondimensionalized model can describe flows with motions across many scales and is suitable for the purpose of this study. The model domain has 128 grid points in each horizontal direction and two vertical layers, so that the maximum horizontal wavenumber is $k_{\max} = 63$. The Rossby deformation wavenumber is set to $k_d = 20$; the Rhines scale is set to $k_\beta = 4$; the velocity scale is $U = 0.2$ and the bottom drag coefficient is $r = 0.5$. An exponential cutoff filter is applied to $k > 40$ to remove energy buildup at the smallest scales. The same model configuration was used for the CNTL experiments in Ying et al. (2018) that evaluate the scale dependency of covariance localization; more detailed description of these model parameters can be found there. Changing parameters (k_d , k_β , U , and r) can alter the scaling of the model and therefore its behavior. For example, in Ying et al. (2018), the S_Scale experiment

sets parameters $k_d = 35$, $U = 0.6$, and $r = 20$, so that the model is much more energetic at small scales. For such a model dominated by small scale processes, the MSA method is expected to be less effective since the small scales are less related to the large-scale components.

Three spatial-scale ranges are defined according to the total horizontal wavenumber $k = \sqrt{k_x^2 + k_y^2}$ as large (L; $k \leq 5$), medium (M; $5 < k \leq 15$) and small (S; $k > 15$) scales. The three scales lead to three corresponding iterations for MSA algorithm. A modification made in the current study is that the temperature field θ is used as the state variable instead of the streamfunction ψ , because the temperature field has a similar energy spectral distribution to the horizontal winds, following the -3 power law. In spectral space, the two variables are related as $\theta_k = -k\psi_k$. Note that there is a one-to-one relation between the ψ and θ fields; therefore, using either one will fully describe the state. The θ field was chosen as the state variable so that there are more features in the smaller scales in contrast to using ψ .

An observing system simulation experiment (OSSE) is designed to test the proposed method with cycling DA trials, under the assumption that the model is perfect and the only source of errors is the initial condition. The QG model is first spun up to its quasi-steady climatology, then a period of 100 time units is sampled from this climatology as the truth trajectory. With this model configuration, the state features several positive and negative temperature anomalies associated with anticyclonic and cyclonic vortices that travel predominately westward. Doubly periodic boundary conditions allow these anomalies to evolve freely throughout their whole life cycles. The baroclinic instability due to difference between the two model layers generates small-scale vortices that interact with each other and the larger-scale vortices.

As a measure of the typical time scales of vortex life cycles, the eddy turnover time is on average 0.5, 0.3, and 0.2 time units for the L, M, and S scales, respectively. The smaller scales vortices evolve more rapidly and have shorter life cycles. The observing network is designed to provide information at every third grid point in each direction and is only available from the top model layer. Figure 1 shows a snapshot of the observation along with the truth model state. The identity observation operator is used (directly observing the θ field) for the sake of simplicity. Spatial inhomogeneity in the observing network is also not considered in this study. The observations are available every 0.05 time units. The θ fields have a climatological variability of ~ 10 (standard deviation). The observation errors, which are assumed to be uncorrelated in time and space, are

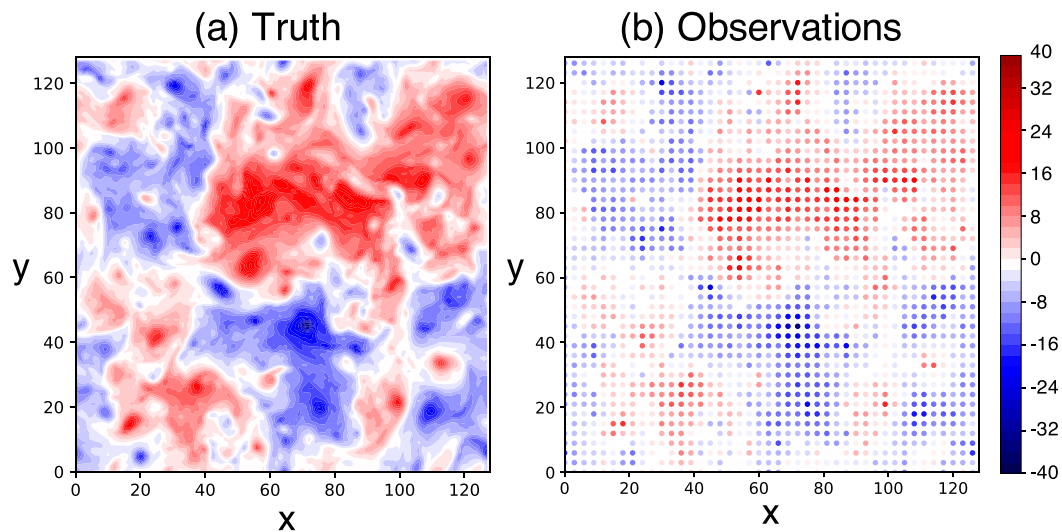


FIG. 1. A snapshot of top-layer model state (temperature field) from (a) truth, and (b) observing network used in the DA experiments.

generated by random draws from a Gaussian distribution with zero mean and standard deviation of 1.

The ensemble square root filter (EnSRF) is chosen as the ensemble filtering method for the DA step (13). The filter belongs to the deterministic square root family of ensemble Kalman filters (Anderson 2001; Whitaker and Hamill 2002; Tippett et al. 2003). The observations at a particular analysis cycle are assimilated serially, and a square root factor is applied to the increments to ensemble perturbations to account for the fact that the observation is not perturbed according to its uncertainty for each member. The initial ensemble is generated by randomly perturbing the true θ field with a tiny white noise, and spinning up with ensemble forecasts for 0.5 time units, allowing the flow-dependent errors to develop. To reduce the negative impact of sampling errors in the ensemble estimated error covariances, the Gaspari and Cohn (1999) localization function is applied to reduce the impact of spurious covariance. The localization distance, defined as radius of influence (ROI) where the covariance is tapered to zero, is specified differently for each scale. To achieve best filter performance, shorter ROIs should be specified for smaller scales due to their shorter physical correlation lengths, and shorter ROIs should also be specified when the ensemble size is smaller. For this study, the localization is manually tuned for best performance on the first cycle. Table 1 lists the specified ROIs for each scale given a range of ensemble sizes. To avoid filter divergence due to an underdispersive ensemble, an inflation factor is applied to the posterior ensemble perturbations, and the inflation factor is derived adaptively according to the

Desroziers et al. (2005) analysis error statistics, which is similar to the adaptive inflation ideas documented by previous studies (e.g., Anderson 2007; Li et al. 2009). The widely applied relaxation-to-prior-perturbation method (Zhang et al. 2004) was not chosen for inflation, because the mixing of prior and posterior perturbations potentially destroys the structure of aligned features in the analysis members. For the alignment step, the Horn and Schunck (1981) optical flow algorithm is used to solve (14) for each member. The strength of the smoothness constraint is set to $\alpha^2 = 100$. The displacement vectors are computed based on the top (observed) layer and applied to the model state on both layers.

Three methods will be compared: the multiscale alignment (MSA) method as described in section 2, the multiscale (MS) method that is similar to MSA but does not apply the alignment steps, and the single scale (SS) method that only performs EnSRF on the full state without iterating over different scales. The MS method is similar to the multiscale DA methods in previous

TABLE 1. Localization distance (ROI; number of grid points) specified for the L, M, and S scales in the MSA method for different ensemble sizes.

Ensemble size (N)	Localization distance (ROI)		
	L scale	M scale	S scale
5	12	8	5
10	18	12	7
20	24	16	10
40	30	22	15
80	30	22	15

studies (e.g., Li et al. 2015; Buehner and Shlyayeva 2015) but in an ensemble filtering formulation. The SS method is similar to the traditional EnSRF widely used in operational numerical weather prediction. The localization ROI for the M scale from the MS method is applied as the overall ROI in SS.

Two sets of sensitivity experiments are performed to test these methods under different conditions. First, with cycling period set to 0.1 time units, a range of ensemble sizes ($N = 5, 10, 20, 40,$ and 80) are tested. As ensemble size decreases, more severe sampling errors are present. Second, fixing ensemble size to $N = 20$, different cycling periods (0.05, 0.1, 0.15, 0.25, 0.4, and 0.5 time units) are tested. The shortest cycling period is shorter than the eddy turnover time for all scales, making the error growth mostly linear, while the longer cycling periods allow errors to grow more nonlinearly. As a result, the prior error distribution varies from nearly Gaussian (shorter cycling periods) to non-Gaussian (longer cycling periods).

4. Experiment results

a. Impact of alignment step on ensemble

The different stages of θ fields from one of the ensemble members is shown in Fig. 2 to illustrate the MSA method step by step. First, the prior state is decomposed into L, M, and S scale components over which the iterations will take place. The first iteration ($s = 1$) operates on the L-scale component. The prior \mathbf{X}_1^b is shown in Fig. 2a. Since there are no larger-scale displacements for the L scale, the aligned prior (Fig. 2b) is the same as the prior. Figure 2c shows the posterior $\mathbf{X}_1^a = \mathbf{X}_1^b + \delta\mathbf{X}_1$ from running the EnSRF at L scale (the DA step). Two highlighted contours from the prior and the truth reveal that the location differences of the contour lines are not as large as the size of θ features (warm and cool anomalies). Since the displacement errors are approximately linear, they are reduced by the ensemble filter in the posterior. The displacement \mathbf{q}_1^b derived from the analysis increment $\delta\mathbf{X}_1$ is overlaid on Fig. 2b. The second iteration ($s = 2$) operates on the M-scale component. The prior \mathbf{X}_2^b is shown in Fig. 2d, and the aligned prior $\mathbf{X}_2^b(\mathbf{q}_1^b)$ is shown in Fig. 2e. The position errors at this scale are comparable to the size of features. With L-scale displacement errors reduced, the aligned prior still has some displacement errors at the M scale. The DA step yields a posterior at M scale (Fig. 2f) as $\mathbf{X}_2^a = \mathbf{X}_2^b(\mathbf{q}_1^b) + \delta\mathbf{X}_2$, and the M-scale displacement \mathbf{q}_2^b derived from $\delta\mathbf{X}_2$ is overlaid on Fig. 2e. The third iteration ($s = 3$) operates on the S scale component. The prior \mathbf{X}_3^b is shown in Fig. 2g, and the aligned prior $\mathbf{X}_3^b(\mathbf{q}_1^b + \mathbf{q}_2^b)$ is shown in Fig. 2h. Because the S-scale

component has experienced 2–3 eddy turnover times during the forecast, the prior θ features have some structural errors in addition to being in the wrong location. However, with both the L- and M-scale displacements applied, the aligned prior has a pattern much closer to the truth than the original prior. The DA step yields the posterior at S scale (Fig. 2i) as $\mathbf{X}_3^a = \mathbf{X}_3^b(\mathbf{q}_1^b + \mathbf{q}_2^b) + \delta\mathbf{X}_3$. The final analysis field is the sum of the posteriors at all three scales.

Figure 3 displays the whole 20-member ensemble at S scale. The prior ensemble (Fig. 3a) shows a very scattered pattern among members. The domain averaged root-mean-square error (RMSE) of the ensemble mean with respect to the truth is 1.66, and the averaged pattern correlation (PC) between members and the truth is 0.31. The posterior from running the multiscale EnSRF (MS method) is shown in Fig. 3b; the RMSE is reduced to 1.31, and PC improved to 0.66. However, many noisy patterns remain. The aligned prior ensemble shown in Fig. 3c is the prior ensemble with L- and M-scale displacements removed from each member. The displacements alone reduce the RMSE to 1.51 and improve PC to 0.51. Note that there are still significant position and structure errors at the S scale left in the aligned prior, but the alignment step preconditioned the ensemble so that the nonlinearity caused by displacement is reduced. As a result, the posterior ensemble using the MSA method (Fig. 3d) shows better agreement with the truth (RMSE is 1.26 and PC is 0.69) compared to the MS method. The noisy signals that are contaminating the pattern-free regions in the MS posterior are much reduced in the MSA posterior, which is beneficial for the ensuing ensemble forecasts.

b. Cycling DA results

The performance of SS, MS, and MSA methods are further evaluated by running cycling DA trials with cycling period of 0.1 time units for each method with varying ensemble sizes of $N = 5, 10, 20, 40,$ and 80 . The domain-averaged RMSEs of the posterior and prior ensemble mean from 200 cycles are summarized in Figs. 4a and 4b. The average RMSE and consistency ratio (CR) are listed in Table 2. The CR is defined as the ratio between the ensemble spread and the RMSE of the ensemble mean, which is less than 1 if the ensemble is underdispersive and can potentially lead to filter divergence. Statistical significance of the improvement of one method over another is estimated by student t tests comparing the error reductions from the 200 cycles to zero. Temporal correlation coefficient of analysis errors between cycles is found to be less than 0.1, which means the results from each cycle can be treated as quasi-independent samples.

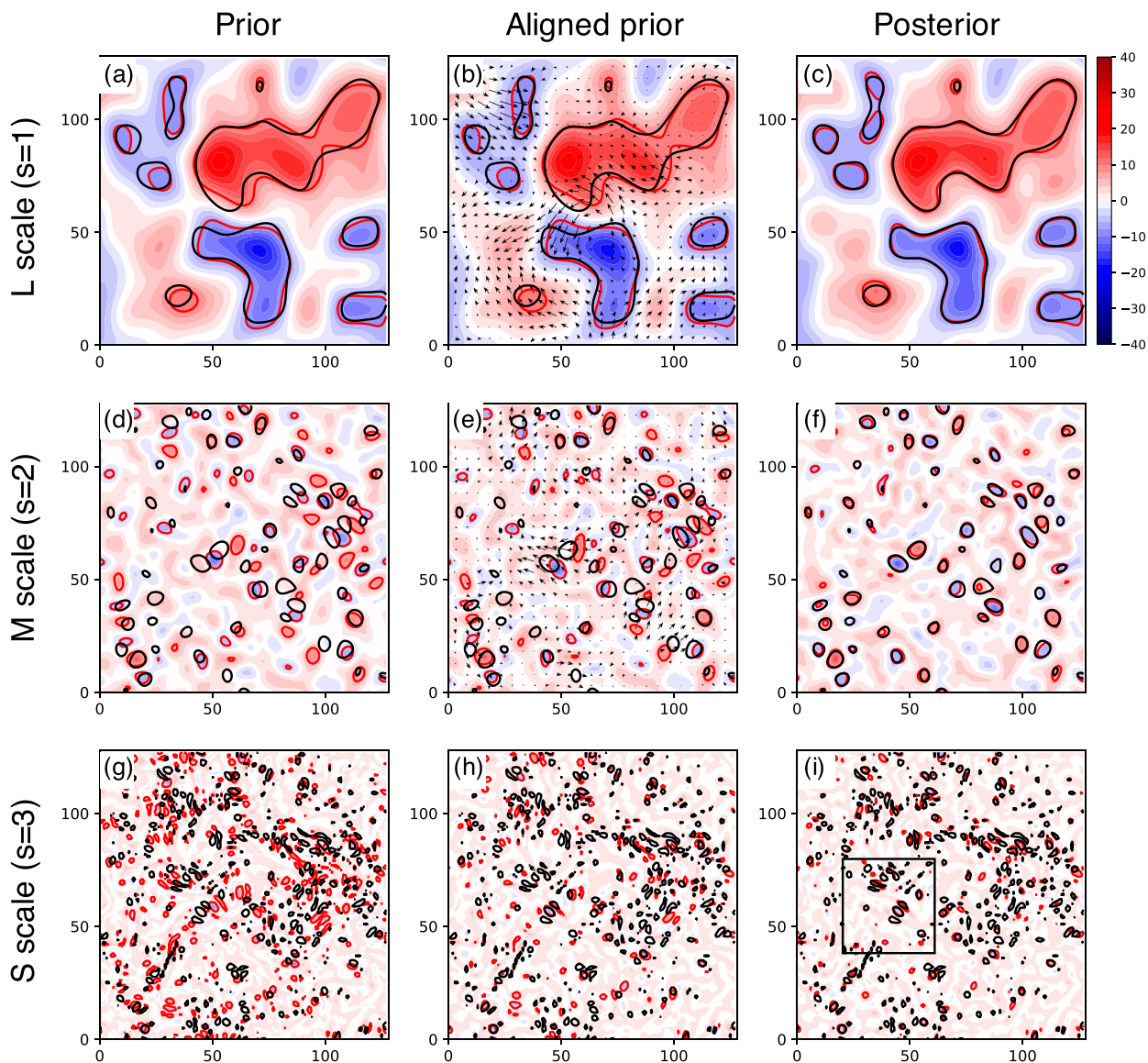


FIG. 2. Stages of the top-layer θ field from member 1 during the first MSA analysis cycle. (left) Priors before DA, (center) priors after the alignment step, and (right) posteriors after DA are shown for (top) large, (middle) medium, and (bottom) small scales. The color shadings are the θ field from -40 to 40 , one highlighted contour is shown in red for each field, and their corresponding truth contour is shown in black. Displacement vectors from the aligned priors to posteriors are overlaid on their corresponding aligned priors.

Smaller ensemble sizes have larger RMSEs due to the larger sampling errors in covariance estimates. For smaller ensemble sizes, significant improvements from SS to MS and from MS to MSA methods are found. For the $N = 5$ case, the MS method reduces the posterior errors from the SS method by 6.1%, and the MSA introduces an extra 10.9% error reduction over MS. The alignment technique improves the posteriors more than the multiscale scheme for ensemble sizes smaller than 40. The MSA also outperforms MS throughout the ensemble forecasts, as indicated by the prior RMSEs. The adaptive inflation has prevented filter divergence,

although the ensemble spread tends to be smaller than ideal after running the EnSRF ($CR < 1$). The multiscale localization in MS improves the consistency compared to the SS method. The alignment technique tends to reduce the ensemble spread, making the consistency of MSA comparable to the SS method for the posterior ensemble. Since the inflation factors used in these experiments are not spatially varying and scale dependent, a more carefully designed inflation scheme can potentially improve the consistency; this will be further investigated in future studies. For the largest ensemble ($N = 80$), the MS-SS and MSA-MS error

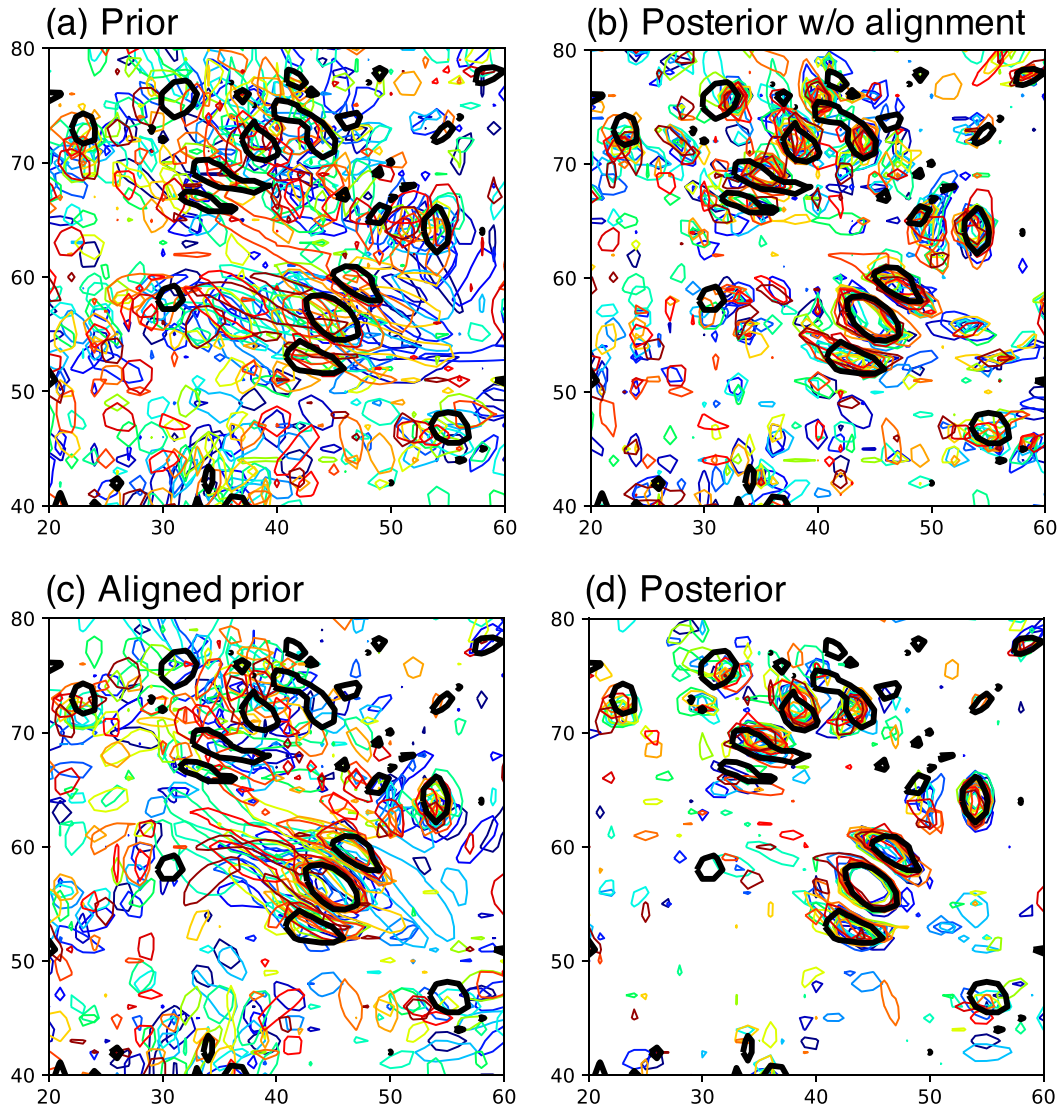


FIG. 3. Spaghetti plots of (a) prior ensemble, (b) posterior ensemble from running multiscale EnSRF without the alignment step (MS), (c) the prior ensemble after the alignment step, and (d) the posterior ensemble (MSA). All results are shown from the first analysis cycle. The contours for $\theta = \pm 3$ are shown from the truth (thick black lines) and each ensemble member (colored lines) filtered for the small scale ($k > 15$). Only part of the model domain, indicated by the box in Fig. 2i, is shown for clarity.

reductions become relatively small and no longer statistically significant.

The second set of sensitivity experiments is running 200 cycles of DA trials of the SS, MS, and MSA methods with ensemble size fixed at $N = 20$ and varying cycling periods of 0.05, 0.1, 0.15, 0.25, 0.4, and 0.5 time units. The posterior and prior RMSEs are summarized in Figs. 4c and 4d, and the averaged RMSE and CR are listed in Table 3. The overall RMSEs increase as the cycling period gets longer as expected. Significant MSA-MS and MS-SS improvements are found in posterior RMSEs for the whole range of cycling periods. The advantage of

using multiscale localization over single scale is increasing as the cycling period increases and the prior errors are increasingly dominated by larger scales. For the cycling period of 0.1, the MS method reduces the posterior errors from the SS method by 1.7%, and the MSA introduces an extra 5.6% error reduction over MS. The alignment technique reduces more posterior errors than the multiscale scheme for cycling periods 0.1, 0.15, and 0.25. These cases feature an error growth time that is similar to the eddy turnover time at S scale (0.2). This time is not so short that the errors are mostly linear nor so long that structural errors dominate displacement

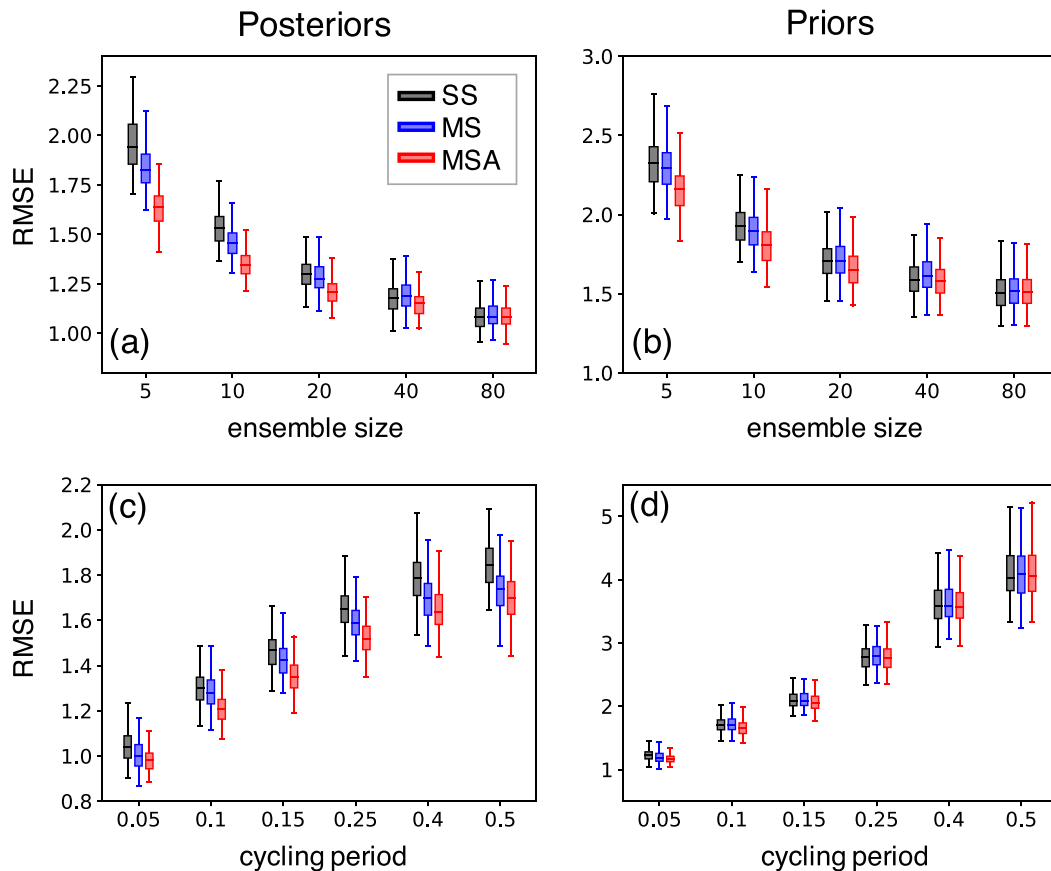


FIG. 4. Boxplots of domain-averaged RMSE of the ensemble mean θ fields with respect to truth from the 200 analysis cycles for (a),(c) posterior and (b),(d) prior ensembles. The SS, MS, and MSA cases are compared and plotted for two scenarios: (a),(b) fixing cycling period to 0.1 time units and vary ensemble sizes from $N = 5$ –80 members, and (c),(d) fixing ensemble size to $N = 20$ members and vary cycling periods from 0.05 to 0.5 time units. Note that the displayed error ranges are different for each panel.

errors. For the shortest cycling period (0.05), the prior errors are mostly linear because less than one eddy turnover time has passed for all scales. In this case, the prior errors are nearly linear and the EnSRF gives a

nearly optimal solution. On the other hand, for longer cycling periods (0.4 and 0.5), the S scale has experienced several eddy turnover times and the highly nonlinear prior errors pose challenges for all three methods.

TABLE 2. Domain-averaged RMSE of posterior and prior ensemble means with respect to truth averaged over the 200 analysis cycles shown for cases with ensemble sizes from $N = 5$ to 80 with cycling period of 0.1 time units. Corresponding consistency ratios are shown for each case in parentheses. Reduction of errors from SS to MS and from MS to MSA are also shown. A student's t test is performed on the MS-SS and MSA-MS differences against zero, and the bold fonts indicate statistical significance at $p < 0.01$ level.

	Ensemble size	Error of ensemble mean (consistency ratio)			Error reduction		
		SS	MS	MSA	MS-SS	MSA-MS	MSA-SS
Posterior	5	1.96 (0.83)	1.84 (1.01)	1.64 (0.88)	-0.119	-0.200	-0.319
	10	1.54 (0.82)	1.46 (0.98)	1.35 (0.85)	-0.076	-0.112	-0.188
	20	1.31 (0.84)	1.29 (0.97)	1.21 (0.83)	-0.022	-0.072	-0.094
	40	1.19 (0.88)	1.20 (0.94)	1.15 (0.83)	0.010	-0.045	-0.035
	80	1.09 (0.93)	1.10 (0.94)	1.09 (0.91)	0.009	-0.004	0.005
Prior	5	2.33 (1.05)	2.31 (1.04)	2.16 (0.93)	-0.022	-0.146	-0.168
	10	1.94 (0.95)	1.91 (0.98)	1.82 (0.89)	-0.033	-0.090	-0.123
	20	1.72 (0.94)	1.73 (0.96)	1.66 (0.87)	0.010	-0.062	-0.052
	40	1.61 (0.95)	1.63 (0.94)	1.60 (0.87)	0.026	-0.037	-0.011
	80	1.52 (0.97)	1.53 (0.97)	1.53 (0.95)	0.012	-0.005	0.007

TABLE 3. As in Table 2, but showing cases with varying cycling periods from 0.05 to 0.5 time units, all cases using an ensemble size of $N = 20$ members.

	Cycling period	Error of ensemble mean (consistency ratio)			Error reduction		
		SS	MS	MSA	MS-SS	MSA-MS	MSA-SS
Posterior	0.05	1.05 (0.82)	1.01 (0.91)	0.99 (0.81)	-0.039	-0.022	-0.061
	0.1	1.31 (0.84)	1.29 (0.97)	1.21 (0.83)	-0.022	-0.072	-0.094
	0.15	1.47 (0.85)	1.43 (1.01)	1.35 (0.87)	-0.041	-0.071	-0.112
	0.25	1.65 (0.85)	1.59 (1.08)	1.52 (0.91)	-0.058	-0.068	-0.126
	0.4	1.79 (0.84)	1.70 (1.13)	1.65 (0.92)	-0.094	-0.050	-0.144
	0.5	1.85 (0.81)	1.74 (1.12)	1.71 (0.90)	-0.111	-0.031	-0.142
Prior	0.05	1.24 (0.88)	1.21 (0.92)	1.18 (0.82)	-0.031	-0.023	-0.054
	0.1	1.72 (0.94)	1.73 (0.96)	1.66 (0.87)	0.010	-0.062	-0.052
	0.15	2.10 (0.98)	2.11 (1.00)	2.07 (0.92)	0.008	-0.039	-0.031
	0.25	2.78 (1.02)	2.80 (1.05)	2.77 (0.97)	0.024	-0.030	-0.006
	0.4	3.62 (1.06)	3.63 (1.08)	3.61 (1.00)	0.013	-0.026	-0.013
	0.5	4.11 (1.04)	4.10 (1.05)	4.09 (1.00)	-0.011	-0.001	-0.012

For prior errors (forecasts), the improvement from MS becomes insignificant, while the MSA method only brings significant improvement for cycling periods 0.1 and 0.15. The ensemble spread behaves similarly to that from the previous set of experiments; MS method has the best consistency while SS and MSA methods are slightly underdispersive.

5. Discussion and conclusions

This study proposes a multiscale alignment (MSA) method for ensemble filtering to reduce the impact of displacement errors due to nonlinear error growth. The MSA method is based on a multiscale ensemble filtering scheme that iteratively processes the different scales of a model state from the largest to the smallest. At each scale, the traditional ensemble filter is applied to update the state with observations. The analysis increments are then utilized to derive sets of displacement vectors through the Horn and Schunck (1981) optical flow algorithm for each member. The larger-scale displacements are used in an alignment step for the smaller-scale prior ensemble states before the DA step takes place. This alignment step removes the larger-scale displacement errors from the prior, and conditions the prior to have more linear residual errors to be further corrected by the ensuing ensemble filter step. Cycling DA trials with 200 cycles using the two-layer quasigeostrophic (QG) model show that the proposed MSA method significantly improves the accuracy of posteriors compared to existing multiscale localization methods. The MSA method is most advantageous when the ensemble size is small and the cycling period is comparable to the average eddy turnover time of the dynamical system. The current study provides a proof-of-concept test of the proposed MSA method in a simplified QG system;

the application of MSA to real atmospheric prediction models is currently being investigated.

In terms of computational cost, the EnSRF has a cost of $O(n_X n_Y N)$ during the DA step, where n_X is the size of state, n_Y is the number of observations, and N is the ensemble size. Note that n_s is the number of scales. The optical flow algorithm during the alignment step costs $O(n_X n_i)$ for each member, where n_i is the number of iterations used in the Horn and Schunck (1981) algorithm ($n_i \sim 20$ is usually enough). Therefore, the total cost of the MSA method, as prototyped and tested in this paper, is $O[n_s n_X (n_Y + n_i) N]$. As discussed in section 2, the optical flow algorithm is solving a quasi-linear system of the displacement, which is much cheaper than the nonlinear cost function formulation of the same problem (e.g., Ravela et al. 2007; Nehrkorn et al. 2014; Stratman et al. 2018). Potentially, the number of state variables for larger scales can be reduced to further save computational cost. Assume that, for a one-dimensional domain with n_X variables, the resolution is reduced by a factor of $2^{(n_s - s)}$ for scale s . The total number of state variables becomes $[1 + (1/2) + \dots + (1/2^{n_s - 1})] n_X < 2n_X$. This setup will result in a total cost that is no more than $O[2n_X (n_Y + n_i) N]$ no matter how large n_s is. An optimal scale decomposition method for a particular model state is by itself an interesting research problem.

Since the computational cost of the MSA method increases with the number of scales, it is desirable to decompose the state into as few scales as possible. The necessity of the alignment step is determined by the amount of displacement errors in the prior ensemble, which also depends on the observation frequency (cycling period). The cutoff of the largest scale is recommended to be chosen so that the state contains mostly linear prior errors with displacements smaller than the sizes of features; the next scale can be selected so that

displacement errors start to introduce nonlinearity (cycling period comparable to the averaged eddy turnover time); the smallest scale can be the highly nonlinear error growth regimes without enough constraints from observations. Observations with different spatial density can be applied to the DA steps for different scales too. For dense observations, super observations can be used for the large-scale iteration to further reduce computational cost.

The MSA method uses the same observation to update model state at different scales. For a given scale, the observation information is used multiple times: several times in the displacement vector calculation from preceding scales and another time during the DA step of the current scale. Assimilating duplicate information can potentially cause issues such as overfitting the observations. In this study, adaptive covariance inflation is applied to avoid this issue. Using the observation more than once is a common issue in iterative DA methods. For example, the incremental 4DVar (Courtier et al. 1994) reevaluates the cost function in between iterations (outer loops) to include more details in the smaller scales, and assimilates the observation again during the next minimization step. Inflating the observation error variance during iterative assimilation is another method to avoid overfitting, which is also known as “tempering of the observation likelihood” (see section 3.2 in van Leeuwen et al. 2018). Perturbations can also be added to the displacement vectors. These potential treatments in MSA method can be further investigated in future studies.

The QG model state is defined on a doubly periodic domain, therefore boundary condition is not considered during the calculation of displacement vectors. For regional models, more constraints need to be considered for the displacement vectors. For example, displacement vectors can only be zeros along the lateral boundaries, and the dynamic and thermodynamic fields need to remain in physical balance after displacement. One can refer to Nehr Korn et al. (2014) for additional treatments necessary for displacement vectors when applied to a hurricane case using the Weather Research and Forecasting (WRF) Model.

For the case when observations are sparse, the cost function used in FCA method that is formulated in observation space can be ill-conditioned. The state-space cost function in the optical flow algorithm will allow a more robust solution, since there will always be a pair of full images from the prior and posterior states, and displacement vectors are only derived where observations are available. In this study, the displacement vectors are determined by the top-layer temperature field increments and applied to both layers, because the QG

model state is mostly barotropic. For more complex three-dimensional models with different vertical modes, the displacement vectors can potentially be derived separately for each level according to their analysis increments. For models with several state variables, the ensemble filter will update each variable according to the cross-variable covariance estimated from the ensemble. The displacement vectors can potentially be derived separately from the increments in each variable, which will result in a more physically consistent solution.

A Python code package for the MSA method described in this study, along with the QG model code (courtesy of Dr. Shafer Smith) and the configuration of experiments conducted in this study, can be found at https://github.com/myying/QG_Multiscale_DA. Currently, the MSA method is being implemented in the Data Assimilation Research Testbed (DART) framework, and real-data experiments are being conducted to further test its utility.

Acknowledgments. This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977. The author appreciates the support from the Advanced Study Program. The author thanks Jeffrey Anderson for his guidance at various stages of this research. This study also benefited greatly from the discussion with Mohamad El Gharamti, Chris Snyder, and Jonathan Poterjoy. The author also appreciates the comments from three anonymous reviewers that helped in improving an earlier version of the paper.

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2).
- , 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99–111, <https://doi.org/10.1016/j.physd.2006.02.011>.
- Beezley, J. D., and J. Mandel, 2008: Morphing ensemble Kalman filters. *Tellus*, **60A**, 131–140, <https://doi.org/10.1111/j.1600-0870.2007.00275.x>.
- Brewster, K. A., 2003: Phase-correcting data assimilation and application to storm-scale numerical weather prediction. Part I: Method description and simulation testing. *Mon. Wea. Rev.*, **131**, 480–492, [https://doi.org/10.1175/1520-0493\(2003\)131<0480:PCDAAA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0480:PCDAAA>2.0.CO;2).
- Buehner, M., and A. Shlyayeva, 2015: Scale-dependent background-error covariance localisation. *Tellus*, **67A**, 28027, <https://doi.org/10.3402/tellusa.v67.28027>.
- Chen, Y., and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 1828–1845, <https://doi.org/10.1175/MWR3351.1>.
- Courtier, P., J. N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an

- incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, <https://doi.org/10.1002/qj.49712051912>.
- Dawson, D. T., M. Xue, J. A. Milbrandt, and A. Shapiro, 2015: Sensitivity of real-data simulations of the 3 May 1999 Oklahoma City tornadic supercell and associated tornadoes to multimoment microphysics. Part I: Storm- and tornado-scale numerical forecasts. *Mon. Wea. Rev.*, **143**, 2241–2265, <https://doi.org/10.1175/MWR-D-14-00279.1>.
- Desroziers, G., L. Berre, B. Chapnick, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, <https://doi.org/10.1256/qj.05.108>.
- 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, <https://doi.org/10.1029/94JC00572>.
- Gaspari, G., and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, <https://doi.org/10.1002/qj.49712555417>.
- Geer, A. J., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteor. Soc.*, **137**, 2024–2037, <https://doi.org/10.1002/qj.830>.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, [https://doi.org/10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2).
- Hoffman, R. N., and C. Grassotti, 1996: A technique for assimilating SSM/I observations of marine atmospheric storms. *J. Appl. Meteor.*, **35**, 1177–1188, [https://doi.org/10.1175/1520-0450\(1996\)035<1177:ATFASO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1996)035<1177:ATFASO>2.0.CO;2).
- , Z. Liu, J. F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.*, **123**, 2758–2770, [https://doi.org/10.1175/1520-0493\(1995\)123<2758:DROFE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<2758:DROFE>2.0.CO;2).
- Horn, B. K. P., and B. G. Schunck, 1981: Determining optical flow. *Artif. Intell.*, **17**, 185–203, [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2).
- Hsiao, L. F., C. S. Liou, T. C. Yeh, Y. R. Guo, D. S. Chen, K. N. Huang, C. T. Terng, and J. H. Chen, 2010: A vortex relocation scheme for tropical cyclone initialization in advanced research WRF. *Mon. Wea. Rev.*, **138**, 3298–3315, <https://doi.org/10.1175/2010MWR3275.1>.
- Lawson, W. G., and J. A. Hansen, 2005: Alignment error models and ensemble-based data assimilation. *Mon. Wea. Rev.*, **133**, 1687–1709, <https://doi.org/10.1175/MWR2945.1>.
- Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523–533, <https://doi.org/10.1002/qj.371>.
- Li, Z., J. C. McWilliams, K. Ide, and J. D. Farrara, 2015: A multiscale variational data assimilation scheme: Formulation and illustration. *Mon. Wea. Rev.*, **143**, 3804–3822, <https://doi.org/10.1175/MWR-D-14-00384.1>.
- Meinhardt-Llopis, E., J. S. Perez, and D. Kondrermann, 2013: Horn–Schunck optical flow with a multi-scale strategy. *Image Process. Online*, **3**, 151–172, <https://doi.org/10.5201/ipol.2013.20>.
- Miyoshi, T., and K. Kondo, 2013: A multi-scale localization approach to an ensemble Kalman filter. *SOLA*, **9**, 170–173, <https://doi.org/10.2151/sola.2013-038>.
- Nehrkorn, T., R. N. Hoffman, C. Grassotti, and J. F. Louis, 2003: Feature calibration and alignment to represent model forecast errors: Empirical regularization. *Quart. J. Roy. Meteor. Soc.*, **129**, 195–218, <https://doi.org/10.1256/qj.02.18>.
- , B. K. Woods, T. Auligne, and R. N. Hoffman, 2014: Application of feature calibration and alignment to high-resolution analysis: Examples using observations sensitive to cloud and water vapor. *Mon. Wea. Rev.*, **142**, 686–702, <https://doi.org/10.1175/MWR-D-13-00164.1>.
- , —, and R. N. Hoffman, 2015: Correcting for position errors in variational data assimilation. *Mon. Wea. Rev.*, **143**, 1368–1381, <https://doi.org/10.1175/MWR-D-14-00127.1>.
- Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76, <https://doi.org/10.1175/MWR-D-15-0163.1>.
- , R. A. Sobash, and J. L. Anderson, 2017: Convective-scale data assimilation for the Weather Research and Forecasting Model using the local particle filter. *Mon. Wea. Rev.*, **145**, 1897–1918, <https://doi.org/10.1175/MWR-D-16-0298.1>.
- Ravela, S., K. Emanuel, and D. McLaughlin, 2007: Data assimilation by field alignment. *Physica D*, **230**, 127–145, <https://doi.org/10.1016/j.physd.2006.09.035>.
- Smith, K. S., G. Boccaletti, C. C. Henning, I. Marinov, C. Y. Tam, I. M. Held, and G. K. Vallis, 2002: Turbulent diffusion in the geostrophic inverse cascade. *J. Fluid Mech.*, **469**, 13–48, <https://doi.org/10.1017/S0022112002001763>.
- Stratman, D. R., C. K. Potvin, and L. J. Wicker, 2018: Correcting storm displacement errors in ensemble using the feature alignment technique (FAT). *Mon. Wea. Rev.*, **146**, 2125–2145, <https://doi.org/10.1175/MWR-D-17-0357.1>.
- 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, [https://doi.org/10.1175/1520-0493\(2003\)131<1485:ESRF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2).
- van Leeuwen, P. J., H. Künsch, L. Nerger, R. Potthast, and S. Reich, 2018: Particle filters for high-dimensional geoscience applications: A review. arXiv preprint arXiv:1807.10434.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- Xue, M., M. Hu, and A. D. Schenkman, 2014: Numerical prediction of the 8 May 2003 Oklahoma City tornadic supercell and embedded tornado using ARPS with the assimilation of WSR-88D data. *Wea. Forecasting*, **29**, 39–62, <https://doi.org/10.1175/WAF-D-13-00029.1>.
- Ying, Y., F. Q. Zhang, and J. L. Anderson, 2018: On the selection of localization radius in ensemble filtering for multiscale quasi-geostrophic dynamics. *Mon. Wea. Rev.*, **146**, 543–560, <https://doi.org/10.1175/MWR-D-17-0336.1>.
- Yussouf, N., D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, <https://doi.org/10.1175/MWR-D-14-00268.1>.
- Zhang, F. Q., C. Snyder, and J. Z. 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, [https://doi.org/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).
- , Y. H. Weng, J. A. Sippel, Z. Y. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, <https://doi.org/10.1175/2009MWR2645.1>.