

Improving vortex position accuracy with a new multiscale alignment ensemble filter

YUE YING^{1,2}, JEFFREY L. ANDERSON², AND LAURENT BERTINO¹

¹ Nansen Environmental and Remote Sensing Center, Bergen, Norway

² National Center for Atmospheric Research, Boulder, Colorado

ABSTRACT: A multiscale alignment (MSA) ensemble filtering method was introduced by Ying (2019) to reduce nonlinear position errors effectively during data assimilation. The MSA method extends the traditional ensemble Kalman filter (EnKF) to update states from large to small scales sequentially, during which it leverages the displacement vectors derived from the large-scale analysis increments to reduce position errors at smaller scales through warping of the model grid. This study stress-tests the MSA method in various scenarios using an idealized vortex model. We show that the MSA improves filter performance as number of scales (N_s) increases in the presence of nonlinear position errors. We tuned localization parameters for the cross-scale EnKF updates to find the best performance when assimilating an observation network. To further reduce the scale mismatch between observations and states, a new option called MSA-O is introduced to decompose observations into scale components during assimilation. Cycling DA experiments show that the MSA-O consistently outperforms the traditional EnKF at equal computational cost. A more challenging scenario for the MSA is identified when the large-scale background flow and the small-scale vortex are incoherent in terms of their errors, making the displacement vectors not effective in reducing vortex position errors. Observation availability for the small scales also limit the use of large N_s for the MSA. Potential remedies for these issues are discussed.

1. Introduction

Geophysical models have increasing prediction skill over the past decade thanks to improving model resolution, better representation of physical processes, as well as the increasing number and better use of observations that help better initialize the models. Data assimilation (DA) is the process of combining the model forecasts with observational information and finding the best estimate of the model states. DA is an important component in a prediction system, an inferior method can be the bottleneck in prediction skill even though better models and observing networks are already available. The ensemble Kalman filter (EnKF; Evensen 1994; Burgers et al. 1998) is the state-of-the-art DA method in many geophysical prediction problems (Houtekamer and Zhang 2016; Carrassi et al. 2018). However, the EnKF analysis becomes more suboptimal as nonlinearity increases in the system. Hoffman et al. (1995) showed that forecast error can be decomposed into three components: displacement, amplitude, and residual. A common source of nonlinearity is from large position errors (displacement) in geophysical features, such as atmospheric fronts, ocean eddies, sea ice edge, sunspots, and so on. Chen and Snyder (2007) showed in a simple two-dimensional vortex example that a linear error in vortex position manifests as nonlinear errors in the model states (winds), and they suggested including features (position, intensity, and shape) as observations in DA to correct model states. The feature-based observations improve the analyses when position errors are moderate. However, as position errors increase, more nonlinearity is

introduced in the relation between observed features and model states, making the EnKF updates suboptimal.

To address this issue, previous studies used two different strategies to improve DA performance. The first strategy is to use a nonlinear DA method, such as the particle filter (PF; Doucet et al. 2001; van Leeuwen 2009; Poterjoy 2016) and rank histogram filter (Anderson 2010, 2019), to obtain the best analysis in the presence of high nonlinearity. Some methods formulate the ensemble filter using non-Gaussian error distributions (moderate position errors will cause a skewed error distributions) (e.g., Bishop 2016; Hodgson et al. 2017; Poterjoy 2022), or transform the variables using Gaussian anamorphosis (Simon and Bertino 2009; Amezua and van Leeuwen 2014), to address the consequences of nonlinearity. Iterative methods are introduced to find a nonlinear solution with a sequence of linear updates (Sakov et al. 2012; Bocquet and Sakov 2014). These methods can handle nonlinearity in general, not only for position errors but also for nonlinear model dynamics and observation operators, but the drawback is the increased complexity in implementation and additional computational cost compared to the EnKF.

The second strategy for dealing with nonlinear position error is to explicitly account for displacements in DA (Ravera et al. 2007; Beezley and Mandel 2008) or to use the feature alignment technique (FAT; Nehrkorn et al. 2014, 2015; Stratman et al. 2018) prior to DA. For the particular case of hurricane DA, some studies suggested using a storm-relative framework (Aksoy 2013; Navarro and Hakim 2014; Lin et al. 2018) or vortex relocation (Liu et al. 2020) to avoid the negative impact from large displacements. The FAT is a more general solution to the

Corresponding author: Yue Ying, yue.ying@nersc.no

problem, which derives displacement vectors from observations to warp the model grid and reduce the misalignment between the observed and model simulated features, bringing the DA problem closer to a linear regime. Ying (2019) noticed that smoothing a feature to the large scale (low resolution) will naturally reduce the nonlinearity in its position errors, and introduced a multiscale alignment (MSA) method that derives displacement vectors more efficiently than the FAT. In the MSA, a model state is decomposed into several scale components (SCs) which are updated by the EnKF sequentially from the large to the small scales. After each update, displacement vectors are derived from the analysis increments and then used to warp the model grid to reduce the position errors at smaller scales (assuming that large-scale and small-scale displacement errors are coherent).

The multiscale approach is proven beneficial in large-dimensional DA problems (Zhang et al. 2009; Miyoshi and Kondo 2013; Buehner and Shlyaeva 2015; Caron and Buehner 2018). Decomposing the model state into SCs allows errors to be represented separately at each scale. Ying (2020) formulated a multiscale approach where both model states and observations are decomposed into SCs, allowing flexibility in the choice of filter parameters (localization, inflation, etc.). The observation SCs are akin to idea of smoothing or differentiating observations to extract information for DA (Weng and Zhang 2012; Bédard and Buehner 2020; Sodhi and Fabry 2022). In this study, we add the decomposition of observation SCs as a new option in the MSA, which is referred to as MSA-O.

The MSA/MSA-O method provides new means of representing the nonlinear error relation across scales through the alignment steps, i.e., using large-scale errors to reduce displacement errors at smaller scales. The existing multiscale DA approaches (e.g., Li et al. 2015; Wang et al. 2021) do not offer this capability since they only model the linear correlation between errors at different scales (off-diagonal terms in the error covariance matrix for SCs). In terms of treatment of position errors, the vortex relocation approach arbitrarily removes position uncertainties based on vortex position observation, while the FAT simultaneously reduces position errors with other errors in minimizing a nonlinear cost function. The MSA is similar to the FAT in finding displacements. However, unlike the FAT, the MSA uses an iteration over scales for better efficiency and to avoid local minima in minimization.

Although the MSA was shown to outperform the EnKF (Ying 2019), several questions still remain about its performance. First, will the MSA analysis be as good as the EnKF analysis in a quasi-linear regime? Second, is the MSA robust, i.e. does it guarantee a performance improvement for all kinds of prior error conditions? Finally, the MSA inevitably brings additional computational and tuning costs as the number of SCs increases. Does the improvement in performance justify the additional cost? In

this study, we design numerical experiments using a simplified two-dimensional vortex model, similar to the ones used in Chen and Snyder (2007) and Poterjoy (2022) (the Rankine vortex example in his section 5), to stress-test the MSA method and answer these questions. The model, DA methods, and numerical experiments are described in section 2 and results are presented in section 3. A summary of our findings is given in section 4.

2. Methodology

a. Model state and truth

The vortex model describes a non-divergent horizontal wind field, $\mathbf{v} = (u, v)$, defined on a doubly-periodic uniform grid $\mathbf{r} = (x, y)$. The initial wind field consists of a Rankine vortex and a background steering flow,

$$\mathbf{v}(\mathbf{r}) = \mathbf{v}_{\text{vort}}(\mathbf{r}; V_{\max}, R_{\text{mw}}, a) + \mathbf{v}_{\text{bkg}}(\mathbf{r}; \bar{V}_{\text{bkg}}), \quad (1)$$

where V_{\max} , R_{mw} , and a are the radius of maximum wind, maximum wind speed, and shape parameter for the Rankine vortex, and \bar{V}_{bkg} is the averaged background wind speed.

Model physics is based on two-dimensional vorticity dynamics, but we added a vorticity generation term to increase vortex intensity when the maximum wind speed is lower than a critical speed limit, simulating the rapid intensification process. Despite its simplicity, the vortex model captures the key physical processes and scale interactions in vortex dynamics, and the rapid intensification is particular interesting and challenging from a DA perspective (see Appendix A for details).

Let ψ be the state vector containing all model variables in physical space, i.e., $\mathbf{v}(\mathbf{r})$ flattened into a vector, the model advances from time t to $t+1$ as

$$\psi_{t+1} = m_{t \rightarrow t+1}(\psi_t; \xi, \kappa_g, \sigma_\kappa, V_c, \nu), \quad (2)$$

where m is the nonlinear model, ξ is the vorticity generation coefficient, κ_g and σ_κ control the spectral shape of the generation term, V_c is the critical speed limit, and ν is the dissipation coefficient.

For the truth (denoted with a star) ψ^* , the vortex is placed at the center of the domain, $V_{\max} = 35 \text{ m s}^{-1}$, $R_{\text{mw}} = 45 \text{ km}$, and $a = 1.5$, while the background flow $\bar{V}_{\text{bkg}} = 5 \text{ m s}^{-1}$. Model parameters are set as follows: the dissipation $\nu = 50 \text{ m}^2 \text{ s}^{-1}$, the generation term $\xi = 8.3 \times 10^{-7} \text{ s}^{-1}$, $\kappa_g = 8$, $\sigma_\kappa = 3$, and $V_c = 75 \text{ m s}^{-1}$. With this setup, the simulated vortex intensifies and reaches the speed limit in $\sim 9 \text{ h}$, then its intensity oscillates near this limit. However, adaptive time steps are required to achieve steady simulation of this oscillation (otherwise numerical instability occurs). For the sake of simplicity we choose to only run the model for 12 h and focus on DA experiments during the rapid intensification period.

b. Ensemble forecasts

An ensemble of model runs characterize the forecast errors prior to DA. Let $n = 1, \dots, N$ index the ensemble members. The initial ensemble $\psi_{n,t=0}$ is generated through perturbing the truth in several different ways to represent different sources of uncertainties. Figure 1 gives an example of vortex evolution in an ensemble forecast.

For member n , its initial wind field is

$$\mathbf{v}_n(\mathbf{r}) = \mathbf{v}_{\text{vort}}^*(\mathbf{r} + \mathbf{r}'_n; V_{\max,n}, R_{\text{mw},n}) + \mathbf{v}_{\text{bkg}}^*(\mathbf{r} + \mathbf{r}'_{\text{bkg},n}) + \mathbf{v}'_n. \quad (3)$$

Here, $\mathbf{r}'_n = (x'_n, y'_n)$ is a spatially constant vector field that displaces the wind field to represent position uncertainties, where x'_n and y'_n are randomly drawn from a normal distribution $\mathcal{N}(0, L_{\text{sprd}})$. The position spread L_{sprd} normalized by R_{mw} is a good indicator of the degree of nonlinearity in wind errors due to displacement, $L_{\text{sprd}}/R_{\text{mw}} = 0.2$ is a quasi-linear regime and $L_{\text{sprd}}/R_{\text{mw}} = 1$ enters a highly nonlinear regime.

Vortex structure uncertainties can be accounted for by randomly perturbing the V_{\max} and R_{mw} parameter for each member as

$$V_{\max,n} = V_{\max}^* + V'_n, \quad (4)$$

$$R_{\text{mw},n} = R_{\text{mw}}^* + R'_n, \quad (5)$$

where $V'_n \sim \mathcal{N}(0, V_{\text{sprd}})$ and $R'_n \sim \mathcal{N}(0, R_{\text{sprd}})$.

The background flow is perturbed in two ways: it is displaced along with the vortex ($\mathbf{r}'_{\text{bkg},n} = \mathbf{r}'_n$; in phase with the vortex position errors) and an additional additive noise is included to account for other error sources in the wind field (\mathbf{v}'_n). The additive noise is generated in the same manner as the background flow but with a lower magnitude than the background flow $\bar{V}' = 1.5 \text{ m s}^{-1}$. Another scenario is also considered where the background flow errors are out of phase with the vortex errors, where $\mathbf{r}'_{\text{bkg},n} = 0$ and \bar{V}' is increased to 3 m s^{-1} (in this case the coherence assumption is broken).

Model errors can be introduced by perturbing ξ . In a perfect-model scenario, the forecast model is the same as the true model ($\xi_n = \xi^* = 8.3 \times 10^{-7} \text{ s}^{-1}$). For imperfect models, ξ_n is randomly drawn from a uniform distribution $\mathcal{U}(3.3, 10) \times 10^{-7} \text{ s}^{-1}$, this range of values is low-biased to mimic the behavior of real hurricane models that tends to underestimate vortex intensity rather than to overestimate.

c. Observations

Let ϕ be a vector containing the observations, and N_o be the number of observed locations in the domain. Synthetic observations are generated from the truth as $\phi^o = h(\psi^*) + \epsilon^o$, where h is the observation operator and $\epsilon^o \sim \mathcal{N}(\mathbf{0}, \sigma_o^2 \mathbf{I})$ is the observation error. The observation errors are uncorrelated (diagonal covariance), and σ_o is the observation error standard deviation.

Several different types of observations are considered in this study. In section 3a, we assimilate an observation at one single location (randomly placed in the domain) to study its impact and the asymptotic behavior of the MSA method. Two types of observations are compared. The first observation measures directly the u and v wind at a given location, in this case, h consists mostly of linear interpolation coefficients, and we set $\sigma_o = 3 \text{ m s}^{-1}$. The second type is the vortex center position (x_c, y_c) , which is feature-based. The h operator in this case is nonlinear, involving a search for the maximum in the vorticity field. For position observations, we set $\sigma_o = 0.1 R_{\text{mw}} = 4.5 \text{ km}$. In section 3b, we assimilate several observations from a network. Two different network geometries are tested: first, a global network covering the entire domain ($N_o = 1000$ randomly placed locations, $\sim 8\%$ of which are within a 180 km radius of the vortex center); second, a more realistic targeted network that only has sparse observations within the vicinity of the vortex ($N_o = 60$ randomly placed locations, all within a 180 km radius of the vortex center). Finally, in section 3c, we conduct cycling DA experiments, assimilating the targeted observation networks (following the true vortex) that are available every 3 h during the 12-h period (in 3 analysis cycles).

d. The DA method and experiments

DA finds the optimal combination of the prior ensemble ψ_n^b with the observations ϕ^o , so that the analysis ensemble ψ_n^a has minimum error while its spread matches this error. Here we drop subscript t for simplicity since the analysis takes place for the same time (filtering). The EnKF first computes observation priors (observed values based on prior model states),

$$\Phi_n^b = h(\psi_n^b), \quad (6)$$

then compares it with the observations to obtain an innovation $\phi^o - \Phi_n^b$ which is used to update the state variables through the error covariance,

$$\text{cov}(\psi^b, \Phi^b) = \frac{1}{N-1} \sum_{n=1}^N \psi_n^b \Phi_n^{bT}. \quad (7)$$

The underlying ideas for the MSA method are: decomposing the model states into SCs and applying the EnKF updates separately and utilizing the large-scale analysis increments to derive displacement vectors that reduces position errors at smaller scales.

Appendix B describes methods for scale decomposition. Let N_s be the number of SCs. For scale s , the prior state becomes $\psi_{n,s}^b = \mathbf{F}_s \psi_n$, which is updated by the innovation through a cross-scale covariance $\text{cov}(\psi_s^b, \Phi^b)$. We

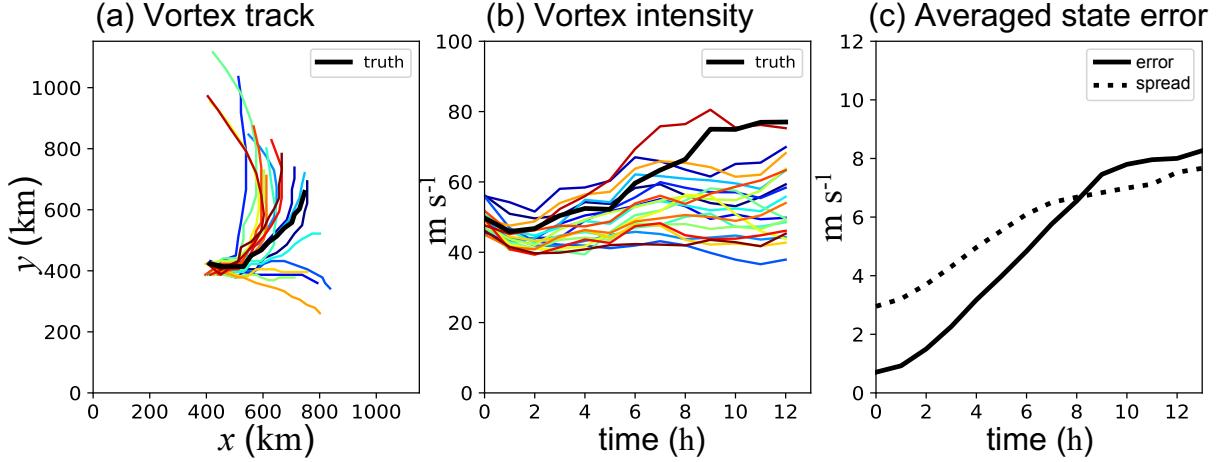


FIG. 1. Vortex evolution from a 20-member ensemble forecast with initial vortex position error $L_{\text{sprd}}/R_{\text{mw}} = 0.2$, no structure error, out-of-phase background flow errors and using the imperfect models. Vortex (a) tracks and (b) intensities are shown (colored lines for members, black line for the truth). (c) The domain-averaged model state error (solid black line) and the averaged ensemble spread (dotted line).

propose a new option for the MSA, which we call MSA-O (the letter O stands for observation), where the observations are also decomposed into SCs $\Phi_s^o = \mathbf{F}_s^o \Phi^o$ and used in correspondence with the state SCs, with covariances $\text{cov}(\Psi_s^b, \Phi_s^b)$, to reduce the scale mismatch during the EnKF updates. Figure 2 illustrates the scale decomposition of u -wind states and observations for the $N_s = 4$ case.

In this study, we choose the ensemble square root filter (EnSRF; Whitaker and Hamill 2002; Tippett et al. 2003) to perform the EnKF update. We refer to the MSA method using N_s SCs as EnSRF_MSA_ N_s , the MSA-O method as EnSRF_MSA-O_ N_s . Note that $N_s = 1$ reverts to the original EnSRF, and the largest N_s is 7 given the model grid resolution. Appendix C provides a detailed formulation of the MSA/MSA-O method.

To provide a benchmark for comparison, we run free model forecasts from the prior ensemble without assimilation, which is called the NoDA case. For single observation assimilation, a simple PF analysis based on importance sampling (see section 3 in van Leeuwen 2015) is computed to provide another benchmark. The PF applies Bayes rule directly. The prior weight for each member is $1/N$, which is multiplied by observation likelihood $\exp[-(\phi^o - \phi_n^b)^2 / (2\sigma_o^2)]$, then normalized across members to get the posterior weight. The probability distribution is then divided into N equal parts to obtain the analysis ensemble members with equal weights again, in which members with large weights are duplicated while those with small weights are removed. The PF analysis converges to the optimal solution as the ensemble size increases. However, it is not feasible for an observing network

with large N_o since with computationally viable ensemble sizes its weights quickly collapse causing filter degeneracy.

Table 1 summarizes all the test scenarios, in which the different filter methods are compared, and the next section describes the error metrics we used.

e. Evaluation metrics

The first error metric is the domain-averaged root-mean-square state variable error (domain-averaged error for short),

$$\varepsilon_{\text{domain-avg}} = \left[(\langle \Psi \rangle - \Psi^*)^T (\langle \Psi \rangle - \Psi^*) \right]^{\frac{1}{2}}, \quad (8)$$

where $\langle \cdot \rangle$ denotes an average over the ensemble members and $\overline{\langle \cdot \rangle}$ denotes an average over the state variables (over the domain).

Three feature-based metrics help further characterize errors in simulated vortex features: position, intensity, and size. The vortex position is found by searching for the maximum vorticity in the domain. The vortex intensity is defined as the maximum wind speed. The vortex size R_{size} characterizes the area of large wind speeds, it is derived from the azimuthal-average wind profile $V(R)$ relative to the diagnosed vortex center, so that $V(R_{\text{mw}} \leq R \leq R_{\text{size}}) > 15 \text{ m s}^{-1}$ (this value is arbitrarily chosen, for hurricanes the threshold is usually higher to relate to destructive winds). These diagnostic features are commonly used in the hurricane community (e.g., Zhang et al. 2009). Note that the size feature is not the same as R_{mw} , and it depends on V_{max} (intensity feature) since when a vortex intensifies its size also grows.

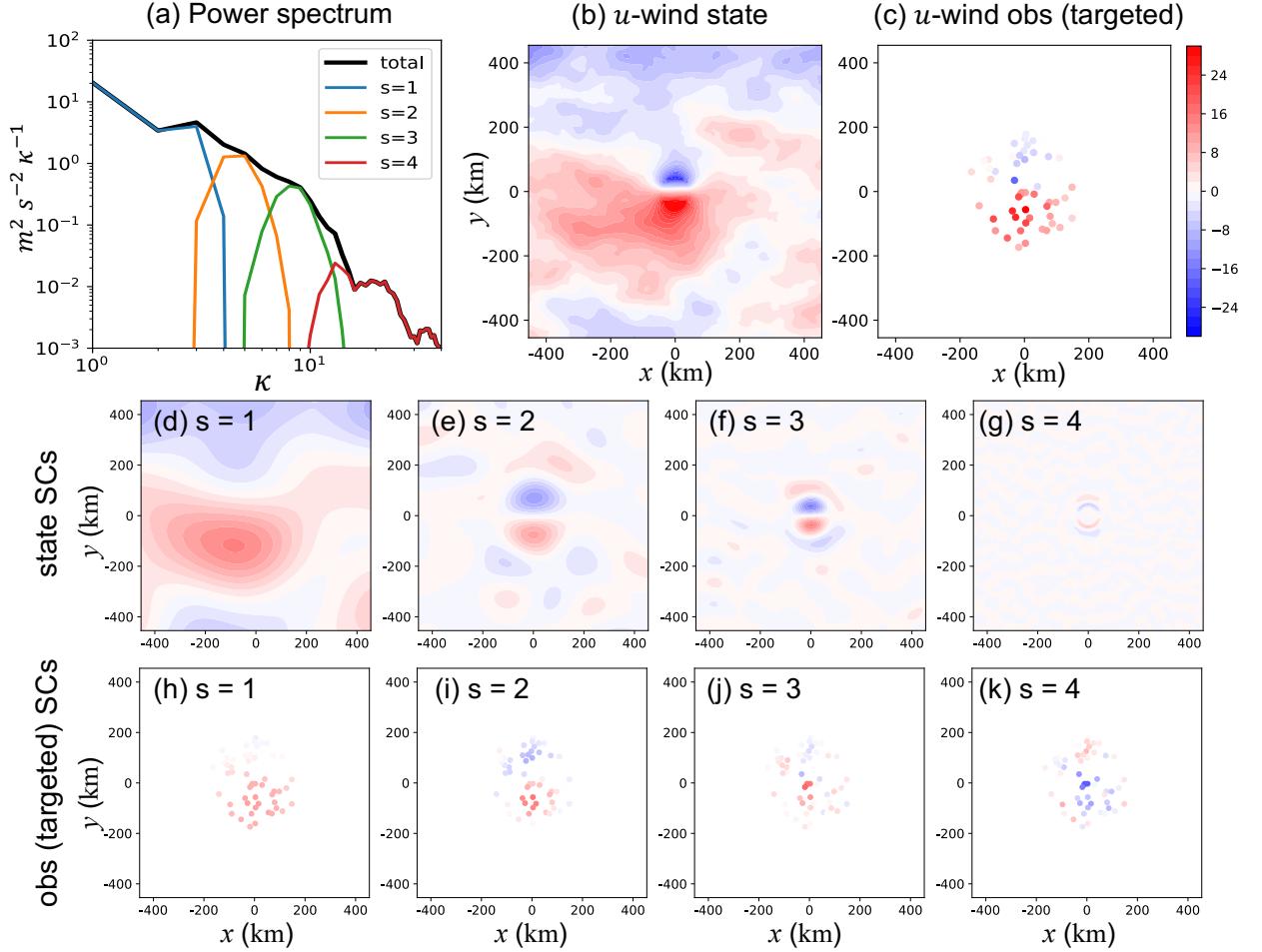


FIG. 2. Illustration of scale decomposition of model state and observation for $N_s = 4$: (a) kinetic energy spectra $\frac{1}{2}(u^2 + v^2)$ for the total state (black line) and for each SC (colored lines); (b) u -wind state, (c) observed u -wind from the targeted network, (d)–(g) u -wind state SCs for $s = 1$ –4, and (h)–(k) observation SCs for $s = 1$ –4.

Section	Scenario	$L_{\text{sprd}}/R_{\text{mw}}$	$V_{\text{sprd}}/V_{\text{max}}$	$R_{\text{sprd}}/R_{\text{mw}}$	Background flow	Obs type	Model
3a	Single_Wind_Obs	0.2, 0.4, 0.6, 0.8, 1	0	0	-	single (u, v)	-
	Position_Obs	0.2, 0.4, 0.6, 0.8, 1	0	0	-	(x_c, y_c)	-
3b	Baseline	0.2, 0.6, 1.0	0	0	in phase	global (u, v)	-
	Incoherent_BkgFlow	0.2, 0.6, 1.0	0	0	out of phase	global (u, v)	-
	Vmax_Error	0.6	0.06, 0.26	0	in phase	global (u, v)	-
	Rmw_Error	0.6	0	0.1, 0.2	in phase	global (u, v)	-
	Targeted_Network	0.2, 1.0	0	0	in phase	targeted (u, v)	-
3c	Baseline	0.6	0	0	in phase	targeted (u, v)	perfect
	Incoherent_BkgFlow	0.6	0	0	out of phase	targeted (u, v)	perfect
	Vort_Struct_Error	0.6	0.06	0.1	in phase	targeted (u, v)	perfect
	Imperfect_Model	0.6	0	0	in phase	targeted (u, v)	imperfect

TABLE 1. List of scenarios and their corresponding configurations in prior ensemble (spread in vortex position, intensity and size; background flow types), assimilated observation type and forecast model. Multiple values are listed for a parameter if they are tested in a sensitivity experiment. For the cycling experiments (section 3c), the vortex and background flow configurations are for the initial conditions.

Let h_{feature} denote the operator from model state to one of features from members does not equal the feature from the diagnostic features. Due to its nonlinearity, the mean

the ensemble mean. We thus define feature errors as

$$\varepsilon_{\text{feature}} = \left\langle [h_{\text{feature}}(\psi) - h_{\text{feature}}(\psi^*)]^2 \right\rangle^{\frac{1}{2}}. \quad (9)$$

In sections 3a and 3b, we present errors diagnosed from the posterior ensemble (final analysis) for each DA method. In section 3c (cycling DA) we present both the errors from the posterior ensemble at the analysis cycles and the ensemble forecasts to the end of DA period.

3. Results

a. Assimilation of single observations

1) ASYMPTOTIC BEHAVIOR OF THE MSA METHOD

In Single_Wind_Obs, we test how the MSA method performs as N_s increases. Since there is only one location observed in space, it is not possible to decompose observations into SCs (MSA-O is not tested here). We test the EnSRF_MSA with $N_s = 1-7$.

We first run the experiment once using a large ensemble size $N = 200$ without localization and visualize the filter updates in physical space in a moderate-nonlinearity scenario ($L_{\text{sprd}}/R_{\text{mw}} = 0.6$). Figure 3 compares wind contours from the analysis ensemble (members with different colors). The original EnSRF analysis suffers most from nonlinearity, vortices with larger position errors become deformed after the linear updates. On the other hand, the PF analysis perfectly preserves vortex structure by design. The EnSRF_MSA analyses improve as N_s increases in that the vortices are less deformed as they are nudged closer to the truth. The MSA method does not preserve vortex structure by design so there is still a slight deformation even for $N_s = 7$.

Figure 4 shows wind speed maps from one ensemble member that is displaced to the southeast in the prior. The EnSRF assimilates the observation at + and inserted a new vortex near the true position, but the linear update is not enough to completely remove the prior vortex, causing the analysis vortex to be elongated. A special case EnSRF_MS_5, where the MSA is applied without the alignment steps (similar to multiscale DA approaches), shows that multiscale DA alone does not improve the result. EnSRF_MSA_5 shows that displacement steps greatly contribute to the improvement of vortex structure. Figs. 4e–h illustrate the displacement vectors derived from the EnSRF update at $s = 1-4$, the warping of the grid gradually nudges the vortex towards truth, improving its position accuracy but at a slight cost of distortion of structure.

Figure 5 shows the bivariate relations between the observation (located at the + sign in Fig. 3) and one state variable (located at the \times sign in Fig. 3) during DA. The EnSRF makes linear updates so the assimilation of u and v observations shows as two sets of parallel line segments, which tend to push members away from the main cluster

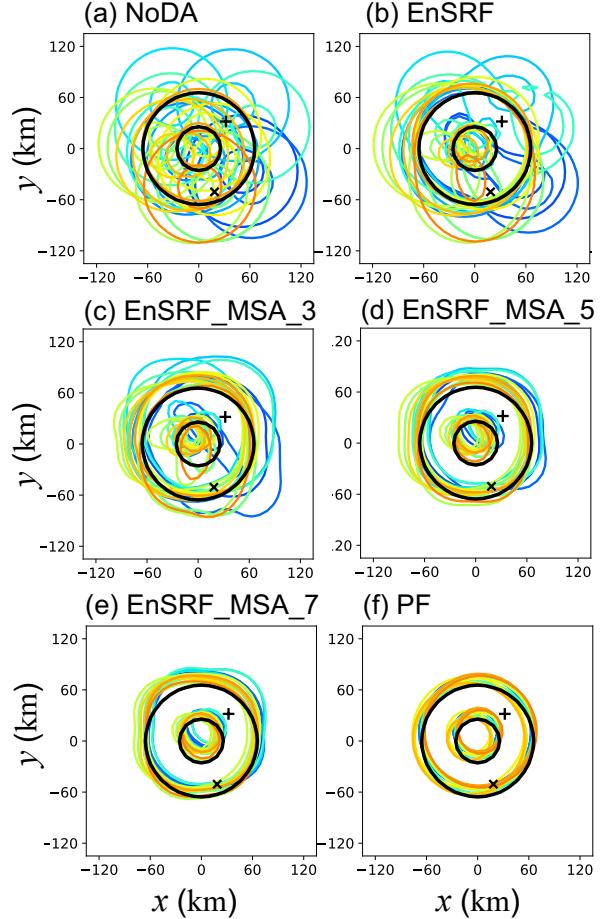


FIG. 3. Ensemble spaghetti plots of wind speed contours at 20 m s^{-1} for (a) NoDA, (b) EnSRF ($N_s = 1$), (c)–(e) EnSRF_MSA ($N_s = 3, 5$, and 7), and (f) PF for the Single.Wind.Obs scenario with $L_{\text{sprd}}/R_{\text{mw}} = 0.6$. The + sign marks the observation location and \times marks the state variable location for Fig. 5. The first 20 members are plotted in color and the true contour in black.

in the prior distribution. Although the final analysis ensemble mean gets much closer to the truth, some members have unphysical values in wind that correspond to the deformation of vortices. The PF update line segments are not parallel and the analysis members converge near the true values. The MSA method behaves more like the nonlinear PF update as N_s increases, the members are moved closer to the truth through the sequential updates over the SCs.

Now we test the robustness of this asymptotic behavior with 1000 random realizations (trials) using a smaller ensemble size ($N = 20$) for the EnSRF and EnSRF_MSA, but the PF benchmark is run with larger ensemble size ($N = 500$). Localization is not applied when assimilating the single observation. The left column of Fig. 6 summarizes the analysis errors averaged over these realizations.

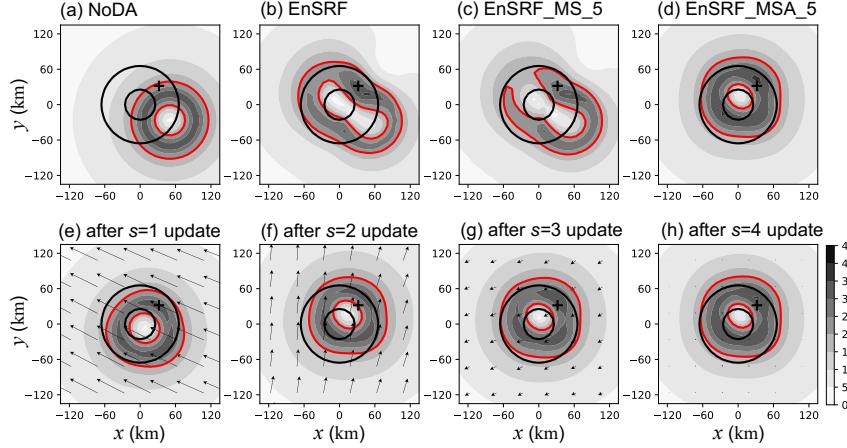


FIG. 4. Vortex from member 1 compared for (a) NoDA, (b) EnSRF, (c) multi-scale EnSRF without alignment (EnSRF_MS_5), and (d) EnSRF_MSA_5. (e)–(h) Intermediate stages after scale iteration $s = 1\text{--}4$ during EnSRF_MSA_5. Wind speed (m s^{-1}) is shown in gray shadings, the 20 m s^{-1} contour is highlighted in red and the true contour in black. The observation is located at the + sign. Displacement vectors derived for each scale are shown in (e)–(h).

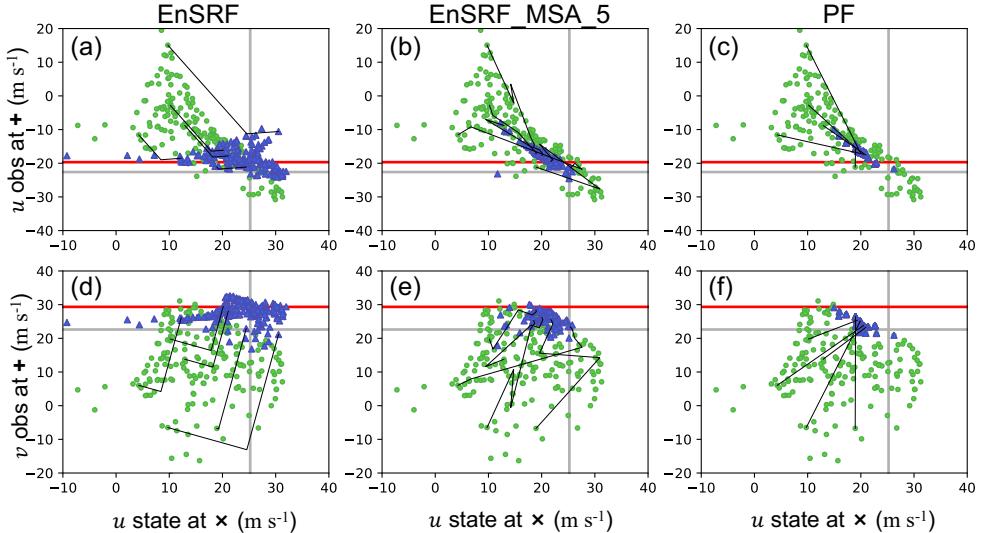


FIG. 5. Scatter plots showing the bivariate relation between the observed u/v variable (top/bottom row) with a u state variable being updated (in this case y axes shows the predictor and x axes shows the predictand). The position of the observation is located at the + sign in Fig. 3 and the state variable at the \times sign in Fig. 3. Three methods are shown: EnSRF (left column), EnSRF_MSA_5 (center column), and PF (right column). Gray lines mark the true values, red lines mark the observed values. Green circles are the prior ensemble, blue triangles are the analysis ensemble. Line segments are plotted to show the intermediate stages for 5 members: for the EnSRF case, one segment for assimilating u , then another segment for assimilating v ; for the EnSRF_MSA_5 case, one segment for each scale iteration $s = 1, \dots, 5$; for the PF case, only one segment is shown.

For domain-averaged errors (Fig. 6a), the EnSRF_MSA method using larger N_s is better than the EnSRF in non-linear regimes. As nonlinearity increases, a larger N_s is required to obtain the best analysis. In the quasi-linear regime ($L_{\text{sprd}}/R_{\text{mw}} = 0.2$), the EnSRF analysis is the best while the EnSRF_MSA analyses are worse. The degrada-

tion is due to vortex intensity errors (Fig. 6e) introduced by the alignment steps. The grid warping is a slightly diffusive process due to the use of an Eulerian grid. Since the Rankine vortex wind profile is nondifferentiable at the peak, it is difficult to preserve wind maxima when the displacements are not an integer number of grid spacings and

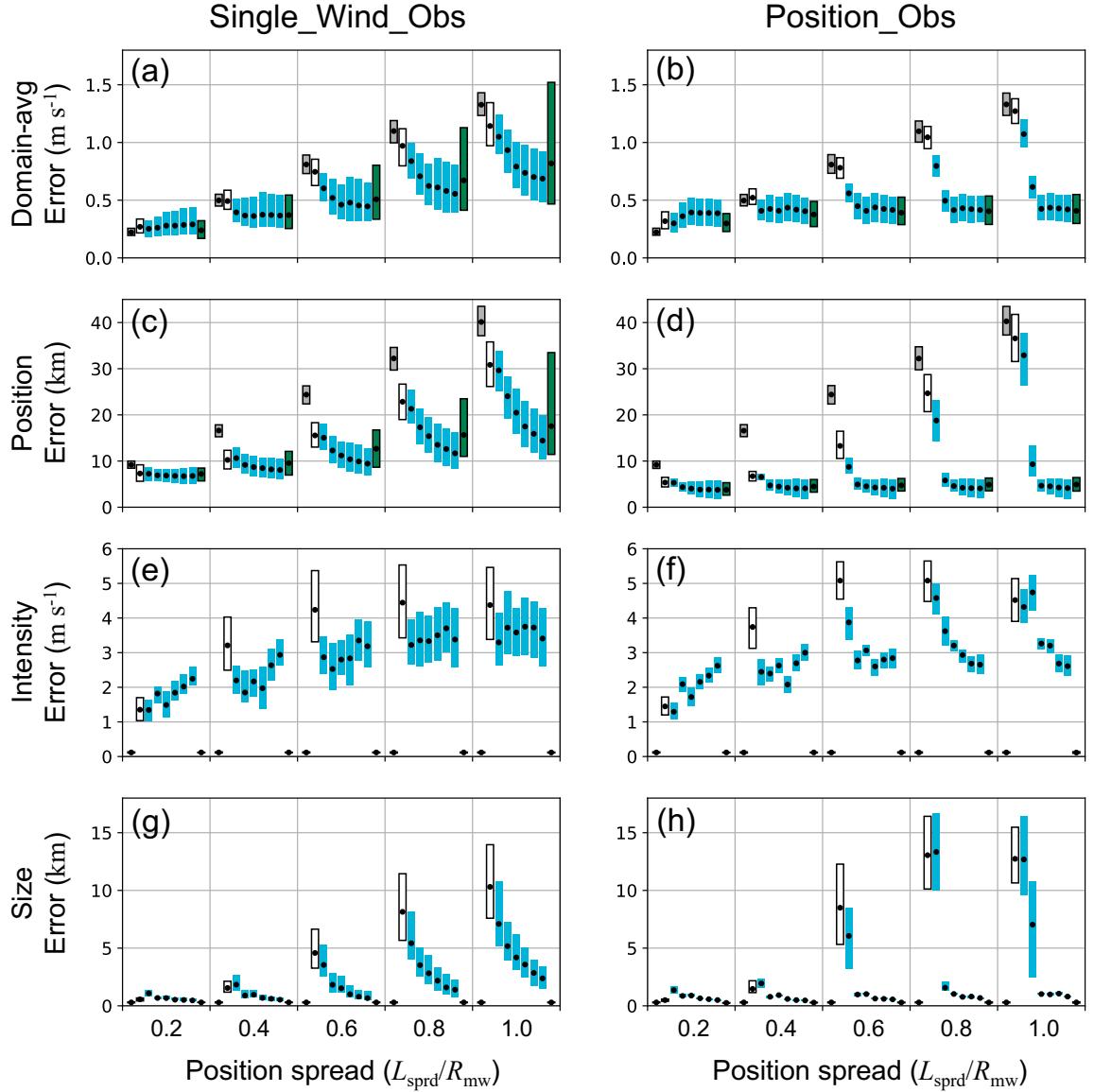


FIG. 6. Error boxplots showing the interquartile range of error metrics (domain-averaged, vortex position, intensity and size errors from top to bottom) obtained from 1000 trials. Two scenarios are compared: (left column) Single_Wind_Obs and (right column) Position_Obs. Cases with varying nonlinearity ($L_{\text{sprd}}/R_{\text{mw}}$ from 0.2 to 1.0) are tested. In each case, results are shown for NoDA (gray), EnSRF (white), EnSRF_MSA_7 (blue, $N_s = 1-7$ from left to right), and PF (dark green). Note that NoDA and PF results are hard to see in intensity and size errors since they are very close to zero.

the wind maxima end up in between grid points and get weakened during interpolation. The vortices are weakened by $\sim 2 \text{ m s}^{-1}$ after warping in EnSRF_MSA. When we introduce a little nonlinearity by increasing position spread ($L_{\text{sprd}}/R_{\text{mw}} \leq 0.2$), the EnSRF analysis becomes more sub-optimal and EnSRF_MSA consistently outperforms the EnSRF. Vortex structures are disrupted in the EnSRF analyses due to the linear updates, causing large intensity and size errors. Note that the EnSRF analysis tends to overestimate

the wind maxima while the alignment steps underestimate them.

Even with very large ensemble size, the PF (perfect in the limit of infinite ensemble size) still produces larger domain-averaged errors than EnSRF_MSA_7 in the nonlinear regime ($L_{\text{sprd}}/R_{\text{mw}} = 1$). Although the PF perfectly preserves vortex features by design so their intensity and size errors are close to zero (not exactly zero because of the round-off errors in h_{feature}), for many trials large position

errors still occur in the PF analyses. This is because the probability for drawing a random particle that is close to the truth decreases rapidly as L_{sprd} increases. The next subsection further investigates the cause of large performance variance for the PF in nonlinear regimes.

2) SENSITIVITY TO OBSERVATION TYPE AND LOCATION

A single wind observation with $\pm 3 \text{ m s}^{-1}$ error is less informative than a position observation with $\pm 4.5 \text{ km}$ error. Especially when the wind observation is placed farther away from the vortex center, it only provides limited information about the vortex itself. To compare to Single_Wind_Obs, we test the direct assimilation of vortex position observations (Position_Obs) and results are shown in the right column of Fig. 6.

In nonlinear regimes, EnSRF_MSA_7 reaches the best performance in Position_Obs no matter how much nonlinearity arises from different L_{sprd} , while its performance degrades as L_{sprd} increases in Single_Wind_Obs. This shows that the position observations have more impact. However, assimilating position observations is quite challenging using the EnSRF (or when N_s is small for EnSRF_MSA). Because the mapping between the observation space (position) and the state space (winds) is highly nonlinear, the EnSRF is suboptimal and cause distortion of the vortex structures. These distortions are even worse (larger intensity and size errors) in Position_Obs comparing to Single_Wind_Obs, which agrees with Chen and Snyder (2007) as they documented vorticity distortion after assimilating position observations.

The information content of wind observations depends on their locations relative to the true vortex center R_{obs} . For observations near R_{mw} , the observed wind speeds are much higher ($\sim 40 \text{ m s}^{-1}$) than those from the vortex outskirt ($\sim 10 \text{ m s}^{-1}$). With the fixed instrument error of $\pm 3 \text{ m s}^{-1}$ these observations have lower percentage errors (we did not consider the case in real wind observations when observation errors increase with wind speeds). Figure 7 shows the domain-averaged errors (from Fig. 6a) again but with the 1000 trials further categorized into 5 bins with respect to observation location $R_{\text{obs}}/R_{\text{mw}}$. In the linear regime ($L_{\text{sprd}}/R_{\text{mw}} = 0.2$), observations farther away ($R_{\text{obs}}/R_{\text{mw}} > 1.2$) cause the degradation in EnSRF_MSA analyses. In the nonlinear regime ($L_{\text{sprd}}/R_{\text{mw}} = 1$), all observations reduce errors, however the biggest improvements come from assimilating observations near the maximum wind ($R_{\text{obs}} \approx R_{\text{mw}}$) and near the vortex center ($R_{\text{obs}} \approx 0$). The PF analysis suffers from ambiguity of observed low wind speeds near the vortex center ($R_{\text{obs}}/R_{\text{mw}} < 0.8$). Because both observations from the center and farther outside have low wind speeds, large weights are assigned accidentally to outliers that match the low wind observations, which degrades the analysis. On the other hand, the EnSRF_MSA does not suffer from this ambiguity due to

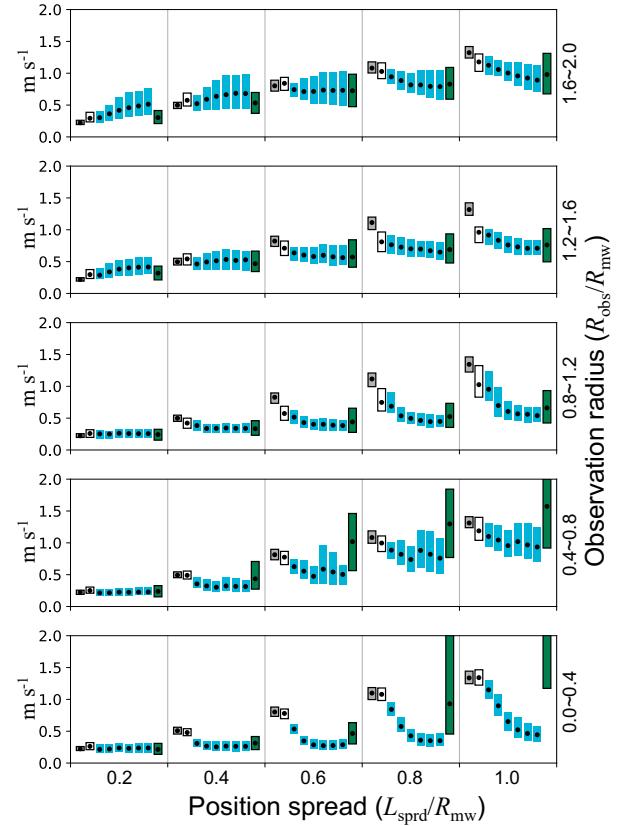


FIG. 7. Boxplots for domain averaged wind errors (same as Fig. 6a) but summarized individually with respect to position spread ($L_{\text{sprd}}/R_{\text{mw}}$, increasing from left to right) and vicinity of wind observation to the true vortex center ($R_{\text{obs}}/R_{\text{mw}}$, increasing from bottom up). With the 5 categories, there are about 200 trials in each.

the scale decomposition and the explicit alignment steps. For the highly nonlinear regime, even $N = 500$ is not a large enough ensemble size for the PF to produce optimal results. In this case, assimilating more observations using a localized PF (Poterjoy 2016) will reduce ambiguity and narrow down to the correct members.

b. Assimilation of observation networks

1) TUNING FOR BEST MULTISCALE LOCALIZATION

In this subsection, we perform manual tuning to select the localization parameters for a multiscale framework. Two parameters are tuned through empirical evaluation: the ROI and the amplitude factor α . The tested values are ROI = 8, 12, 16, 20, 24, 28, 32, 40, 48, 64 grid points and $\alpha = 0, 0.1, 0.2, \dots, 1$, which results in 110 cases in total. For each case, 100 trials are run where the global network is assimilated to update a prior ensemble ($N = 20$ members) with position spread $L_{\text{sprd}}/R_{\text{mw}} = 0.6$ and out-of-phase background flow errors.

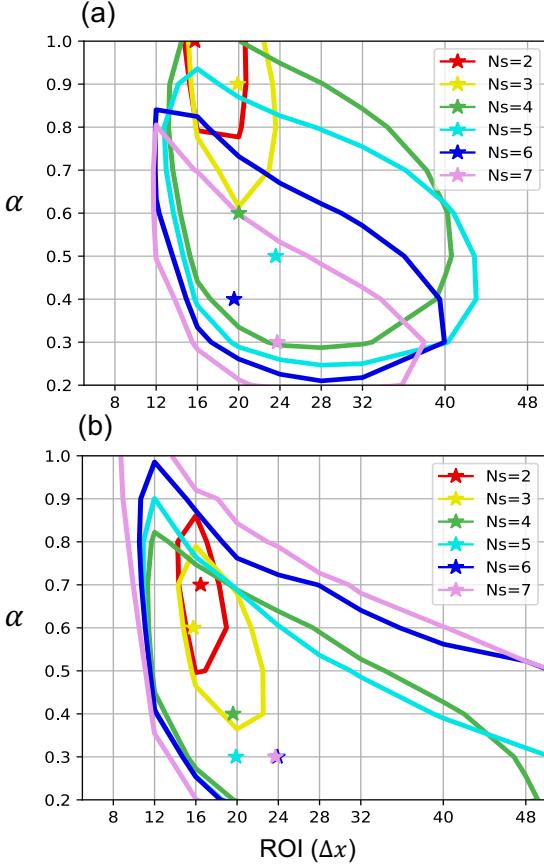


FIG. 8. Tuning of the localization ROI (in units of number of Δx) and amplitude parameter (α) for updating the first SC using (a) the MSA and (b) the MSA-O algorithms. Domain-averaged wind errors are obtained from 100 trials for each combination of parameters and tested for cases with N_s from 2 to 7 (color-coded). The best performance (minimum error) for each case is marked with a star, the contours indicate the range of parameters that achieve an error within 1 % of the best performance.

The filter update for only the first SC in EnSRF_MSA_ N_s is diagnosed for domain-averaged errors. Figure 8a shows the range of parameters that yields good performance (best performance marked as a star). As N_s increases, the first SC corresponds to larger scales and there is a bigger scale mismatch between the observations and the states. The best ROI increases with the underlying correlation scale, which is expected. On the other hand, the best α decreases as N_s increases, and the reasons are twofold. Because the bigger scale mismatch introduces more sampling noise that is not necessarily distance-dependent, reducing α helps reduce the impact of the noise. Also, as a larger ROI is favored for the larger scale, more observations are participating in the update, reducing α helps to prevent overfitting the observations.

Localization for the MSA-O algorithm is tuned in the same manner and results are shown in Fig. 8b. In this

TABLE 2. Tuned localization parameters (ROI_s and α_s for $s = 1, \dots, N_s$ from left to right) for the MSA method.

N_s	$ROI_s (\Delta x)$	α_s
1	16	1
2	16, 8	0.8, 0.9
3	18, 12, 8	0.7, 0.7, 0.8
4	20, 16, 12, 8	0.5, 0.5, 0.7, 0.8
5	22, 20, 16, 12, 8	0.4, 0.4, 0.6, 0.7, 0.8
6	24, 22, 20, 16, 12, 8	0.4, 0.4, 0.5, 0.7, 0.8, 0.8
7	26, 22, 20, 18, 16, 12, 8	0.3, 0.3, 0.4, 0.5, 0.7, 0.8, 0.8

case, since observations are also decomposed to SCs, the scale mismatch is much reduced in the update step, which requires less α adjustment (i.e., α is closer to 1). However, the best performance still comes from $\alpha < 1$ as in the MSA case, only a much wider range of parameters give relatively good performance (compare the purple lines, $N_s = 7$). The MSA-O algorithm requires smaller α than the MSA when updating smaller scales (compare red lines, $N_s = 2$), because the difference between the irregular observation network and the uniform model grid (see Fig. 2) further introduces discrepancies between the observation space and state space. Based on these findings, we configure the MSA in the following experiments with the best-performing localization parameters, which are listed in Table 2. The MSA-O configuration (not listed) is very similar to the MSA, only a few differences for small N_s .

2) SENSITIVITY TO UNCERTAINTIES OTHER THAN POSITION

We created scenarios with different sources of initial uncertainties, in background flow (Incoherent_BkgFlow) and in vortex structure (Vmax_Error and Rmw_Error), to test the robustness of the MSA performance. Figure 9 compares the EnSRF_MSA performance under these scenarios.

The MSA method is built on the assumption that small-scale features and large-scale patterns are coherent so that the large-scale updates can be utilized to correct for errors in the small-scale features. However, it is debatable whether this assumption holds for all geophysical modeling scenarios. In Incoherent_BkgFlow, we test how the MSA method behaves when deviating from the coherent assumption. The incoherent background flow errors (see section 2b for details) are challenging for the MSA. Notice in Fig. 9f how the EnSRF_MSA analysis becomes worse than the EnSRF analysis as N_s increase (some improvement can be seen as N_s keep increasing, but still worse than $N_s = 1$). In the large-scale SC, there are both vortex-scale features projected to the large scale and the large-scale background flow uncertainties themselves. In Incoherent_BkgFlow, the alignment step can sometimes displace the vortex erroneously according to analysis increments after correcting

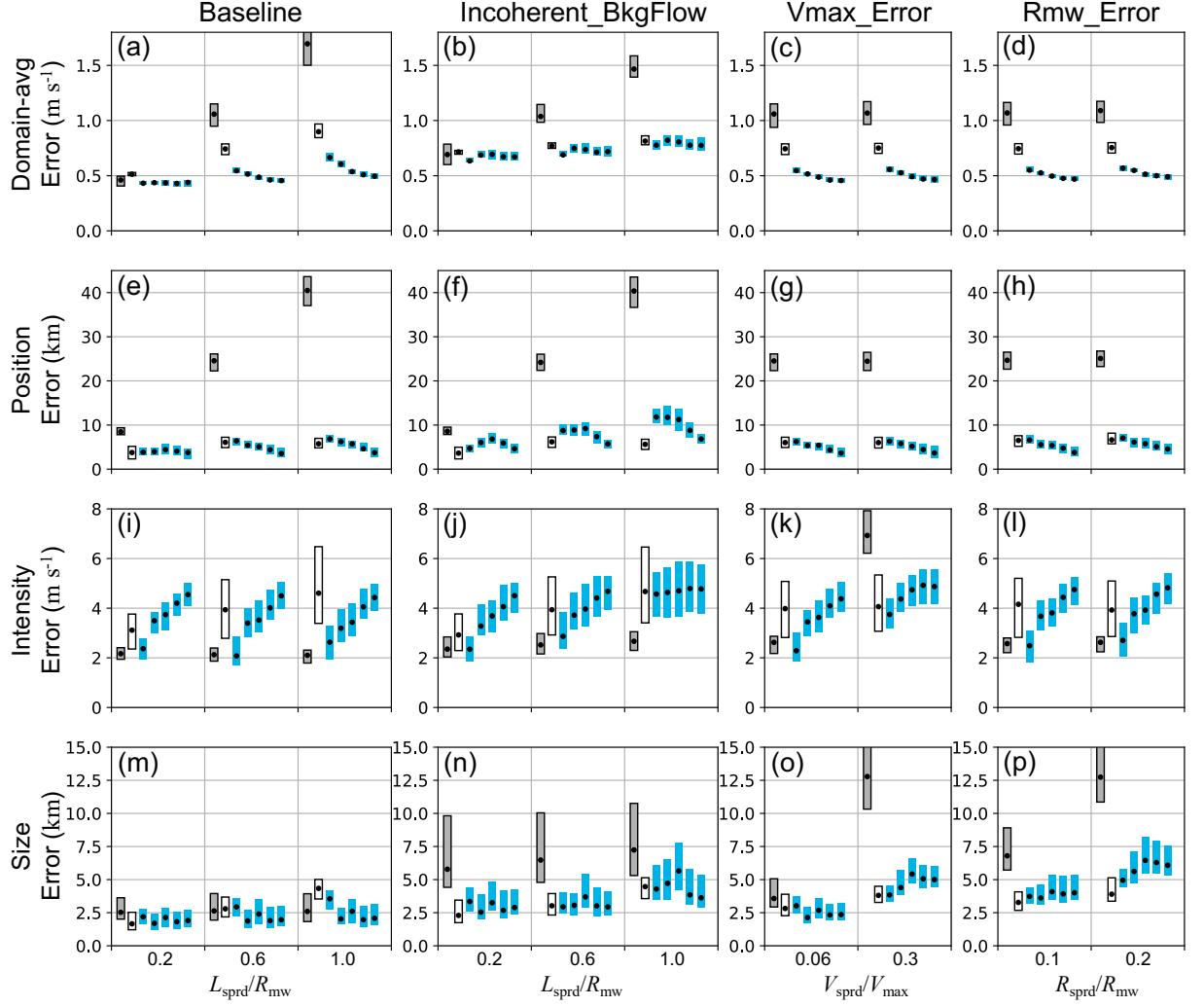


FIG. 9. Comparison of error boxplots (from 100 trials) among NoDA (gray), EnSRF (white), and EnSRF-MSA- N_s (blue; $N_s = 2$ –6 from left to right). Four scenarios with different initial condition uncertainties are tested: (from left to right columns) Baseline, Incoherent_BkgFlow, Vmax_Error, and Rmw_Error (See Table 1 for details). Four error metrics are shown from top to bottom rows.

the additive noise in the background flow, which increases the vortex position errors and degrades the analysis. In Baseline, the MSA method works as expected with best performance.

In Vmax_Error and Rmw_Error, the uncertainties in vortex structure are accounted for by varying the two parameters during generation of the initial Rankine vortex. Figs. 9g and 9h show that the additional uncertainties do not impact the performance of the MSA as the position errors are reduced as expected when N_s increase. Perturbing V_{max} increases both vortex intensity and size errors in the analysis (size also depends on V_{max} due to its definition). Perturbing R_{mw} increases vortex size errors in the analysis.

For all scenarios, the intensity errors first decrease then increase when N_s increases from 1 to 6 (Figs. 9i–l). En-

SRF ($N_s=1$) tends to overestimate the vortex intensity with occasional overfitting to local observations of peak wind speeds. On the other hand, EnSRF-MSA with large N_s tends to underestimate the vortex intensity because of the alignment and warping of the grid weakens the vortex (see discussion in subsection 3a.1). Some degradation in vortex size can be seen for larger N_s (Fig. 9o and 9p) also because there is more distortion due to warping of the grid.

3) SENSITIVITY TO OBSERVATION DENSITY AND COVERAGE

This subsection tests the sensitivity of the MSA/MSA-O performance to different observation density and coverage (i.e., its information content). Two types of observing network are considered: the global network is denser with

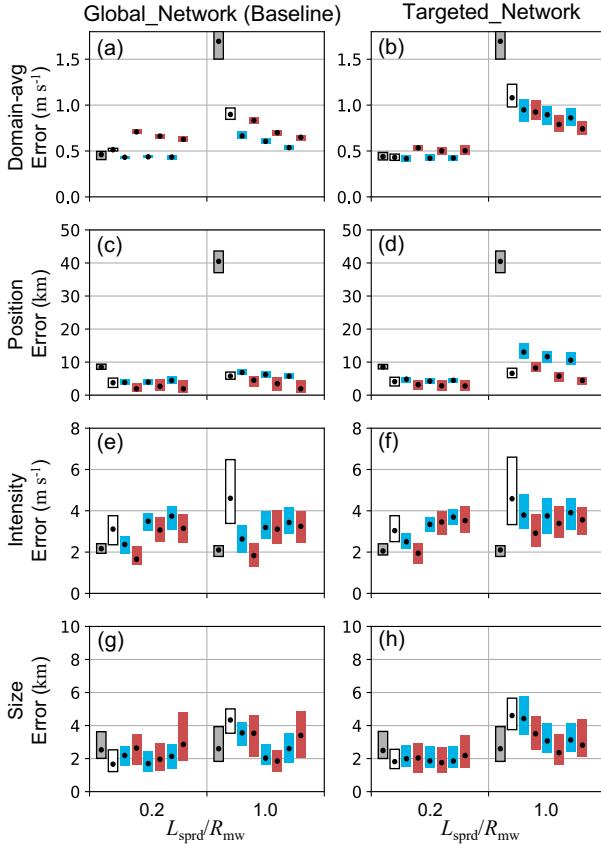


FIG. 10. Comparison of error boxplots (from 100 trials) among NoDA (gray), EnSRF (white), EnSRF_MSA (blue); $N_s = 2-4$ from left to right), and EnSRF_MSA-O (red; $N_s = 2-4$ from left to right). Two scenarios are tested: assimilating global network (Baseline; left column) and Targeted_Network (right column). For each scenario, three cases with $L_{\text{sprd}}/R_{\text{mw}} = 0.2, 0.6$, and 1.0 are tested. Four error metrics are shown from top to bottom.

coverage of the entire domain ($N_o = 1000$), while the targeted network is more sparse ($N_o = 60$) and is only deployed near the vortex center (within 180 km radius, see Fig. 2c for a visualization). Figure 10 compares the performance of EnSRF_MSA and EnSRF_MSA-O ($N_s = 1-4$) when assimilating these two types of observations.

Let us first look at the quasi-linear regime ($L_{\text{sprd}}/R_{\text{mw}} = 0.2$) in Global_Network. Decomposition of observation SCs in MSA-O improves position accuracy compared to MSA (Fig. 10c) but at a cost of degrading the domain-averaged error (Fig. 10a). Comparing the state SCs (Figs. 2d-g) with the observation SCs (Figs. 2h-k), we see that the smaller scales have significant mismatches due to differences in grid geometry. When N_o is not large enough, aliasing may occur. Although the MSA-O benefits from assimilating the large-scale observation SCs with better match to the state SCs, the inaccuracies in the scale decomposition (mostly at small scales) bring negative impacts.

As N_s increases, the MSA/MSA-O also tends to weaken the vortex intensity (Fig. 10e) due to warping of the grid (see subsection 3a.1). The size accuracy (Fig. 10g) does not monotonically improve as N_s increases, larger N_s degrades the size feature, probably because the MSA/MSA-O overfits the small-scale pattern emerging from the observations that are contaminated by aliasing noises. In the nonlinear regime ($L_{\text{sprd}}/R_{\text{mw}} = 1$), the behavior of MSA/MSA-O is the same. Since higher nonlinearity occurs in the prior errors the MSA/MSA-O consistently outperforms the single-scale EnSRF.

Using Global_Network as a baseline, now we compare the performance in Targeted_Network. Notice that the global network constrains the domain-averaged errors much better than the targeted network, thanks to its better information content. For Targeted_Network in the nonlinear regime (Fig. 10b), the benefit from using observation SCs (MSA-O) out-weights the negative impact from observation SC inaccuracies. Position errors (Fig. 10d) show that MSA-O performs better than MSA when assimilating the targeted network. Because the targeted network has smaller-scale vortex information dominating the larger-scale background flow information, the observation SCs help to extract information about the large scales in MSA-O, which helps to reduce position errors after assimilation.

c. Cycling data assimilation experiments

Finally, we test the MSA/MSA-O algorithms in a cycling DA setting. Four scenarios are tested: Baseline, Incoherent_BkgFlow, Vort_Struct_Error, and Imperfect_Model (see Table 1). For each scenario, the targeted observing network is assimilated every 3 h during a 12-h period, the cycling DA is performed for 100 trials over which the error metrics are evaluated from both the analysis ensemble and the ensemble forecasts to the end of the period.

One important constraint we place on the MSA/MSA-O algorithms is that a smaller ensemble size is used for larger N_s so that all methods are compared at equal computational cost. Ensemble size is 30 for NoDA and EnSRF, and 28, 25, and 22 for $N_s = 2, 3$, and 4 . A larger N_s effectively increases the state vector size (see Ying 2019) that is offset by reducing the ensemble size. The extra cost in the alignment step is also balanced by the reduced number of model runs in the ensemble forecast step.

Figure 11 compares the analysis and forecast error boxplots from different methods for each test scenario, and Fig. 12 shows the corresponding error time series, averaged over 100 trials. As discussed in subsection 3b.3, the MSA-O approach is better than the MSA when assimilating the targeted network. The errors are consistently smaller from EnSRF_MSA-O than EnSRF_MSA, and the EnSRF_MSA-O_3 method produces the best overall performance.

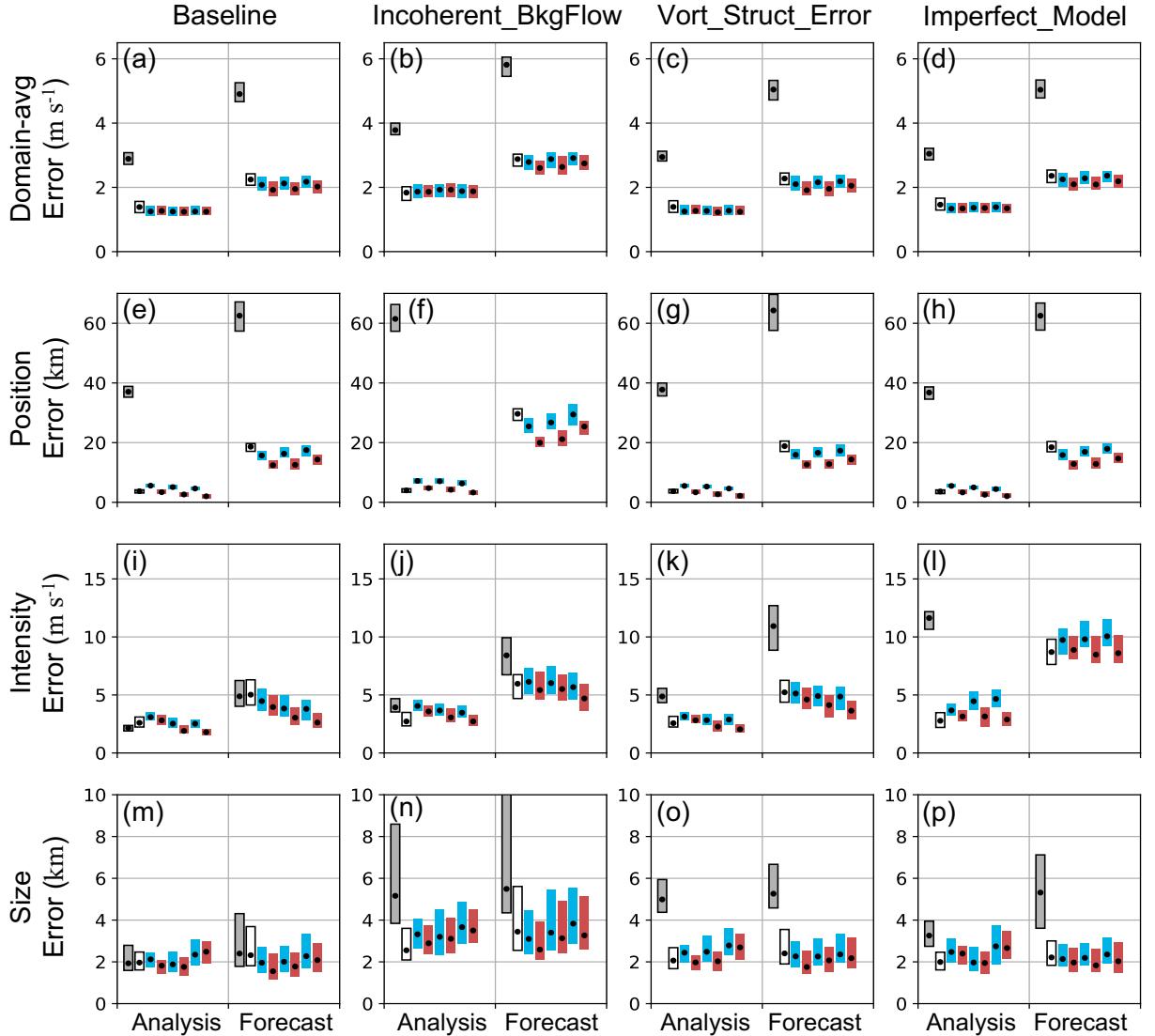


FIG. 11. Cycling DA experiments: comparison of error boxplots (from 100 trials) among NoDA (gray), EnSRF (white), EnSRF_MSA (blue), and EnSRF_MSA-O (red). Four scenarios are tested (from left to right columns): Baseline, Incoherent_BkgFlow, Vort_Struct_Error, and Imperfect_Model. Four error metrics are shown (from top to bottom) for the analysis (averaged over cycles at $t = 3, 6$, and 9 h) and forecast (averaged over the three forecasts at $t = 12$ h).

In the Baseline scenario, the analysis domain-averaged errors are smaller in EnSRF_MSA-O than in EnSRF, but the difference are small among different N_s , because the targeted network only updates near the vortex, which is a small portion of the entire domain. Clearer differences are seen for the feature errors: MSA-O with increasing N_s further reduces the position and intensity errors, but the improvement for size errors stops at $N_s = 3$ and starts increasing back for $N_s = 4$. For the forecasts, EnSRF_MSA-O_3 is the best-performing method. EnSRF_MSA-O_4 is struggling at vortex position and size in the forecasts. The

error time series suggest that error grows more rapidly after the EnSRF analyses (black lines) than the EnSRF_MSA-O_3 analyses (red lines), especially for the vortex position errors (Fig. 12e). Although the EnSRF analysis is already very accurate according to the domain-averaged errors, some physical imbalances (large local wind maxima that are not axisymmetric in the vortex) are introduced when overfitting local observations and overestimating intensity (see increased intensity error in Fig. 12i). The more rapid forecast error growth also indicates such imbalances. On the other hand, the MSA-O method is designed to

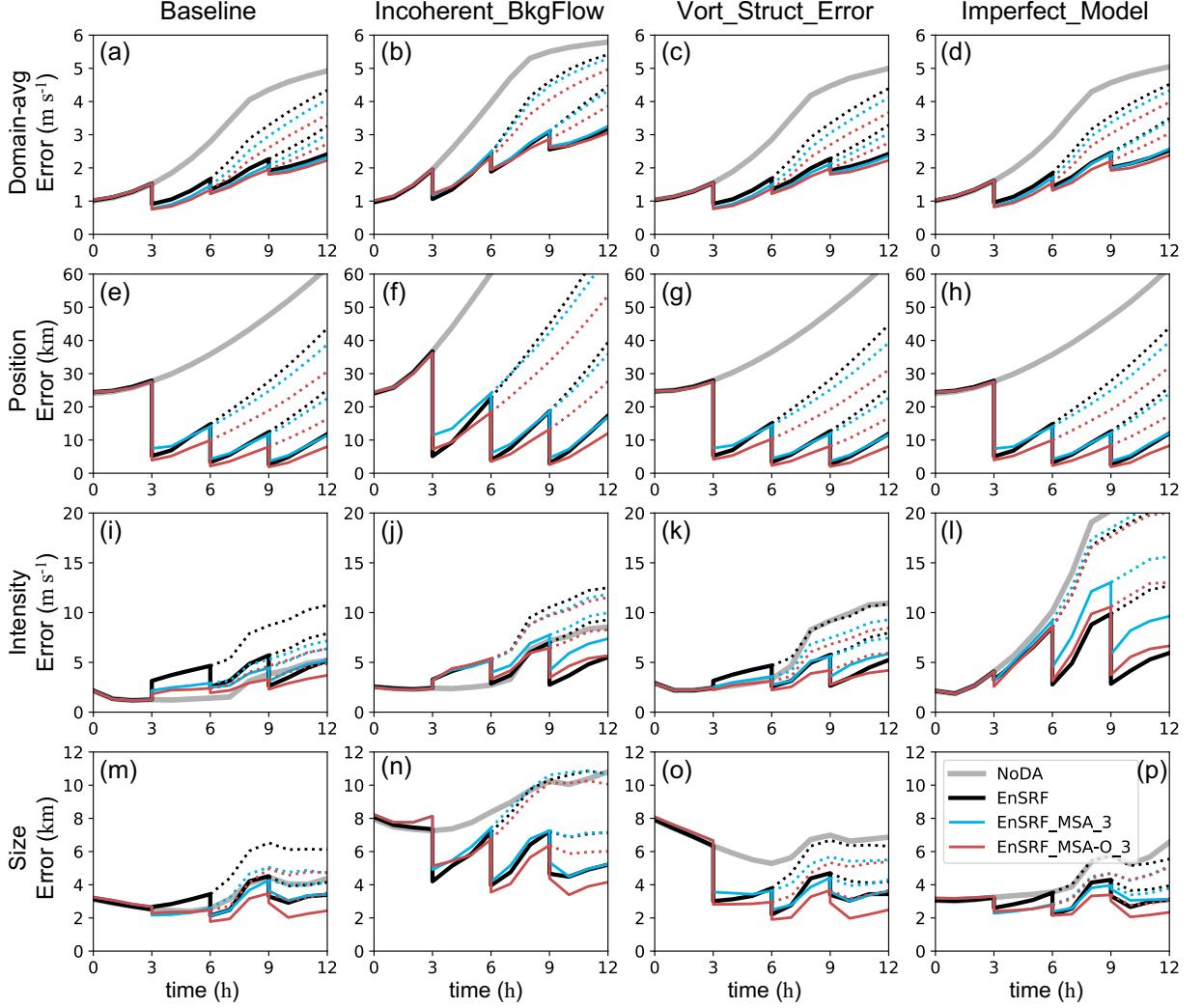


FIG. 12. Similar to Fig. 11 but showing averaged error time series from the 100 trials. In each panel, errors from NoDA (gray), EnSRF (black), EnSRF_MSA_3 (blue), and EnSRF_MSA_O_3 (red) are compared. Analysis cycles take place at $t = 3, 6$, and 9 h . Forecasts are run from the analysis time to the end of the period ($t = 12\text{ h}$) and marked as dotted lines with corresponding colors.

better preserve the vortex physical balance while reducing position errors at the same time. One drawback of the MSA-O with larger $N_s = 4$ is that it tends to overfit the large-scale SCs and collapse ensemble spread in position prematurely (not shown), so that when assimilating the small-scale SCs there is not enough prior spread to allow sufficient increments.

In the Incoherent_BkgFlow scenario, the larger initial uncertainties in background flow cause more rapid position error growth and overall larger vortex size errors than the baseline. NoDA position errors exceed 100 km (off chart in Fig. 11f). For the analysis, the EnSRF has the best performance. However, EnSRF_MSA_O_2 gives the overall best forecasts so it has an advantage over the EnSRF

(Fig. 11). The deviation from the coherent assumption (see subsection 3b.2) is also among the reasons why larger N_s does not further improve the MSA-O results. The behavior of the EnKF having better analysis but worse forecasts is also reported by other studies (e.g., Poterjoy et al. 2017; Poterjoy 2022). Inspection of the EnSRF ensemble spread (not shown) indicates that its forecast spread is also larger and its analysis spread is smaller than those from the EnSRF_MSA_O. This confirms that the EnSRF tends to overfit the observations, causing physical imbalances and more rapid error growth in the forecasts.

The Vort_Struct_Error scenario does not differ too much from the Baseline. With additional uncertainties in the initial vortex structure, the intensity and size errors (Figs.

11k, o) increase for NoDA. But the EnSRF is still able to reduce errors, even a bit more for the intensity and size errors thanks to the increased initial ensemble spread in vortex features. The MSA/MSA-O behave as expected in further improving over the EnSRF.

For the Imperfect_Model scenario, the additional model uncertainties cause much larger errors in vortex intensity than the Baseline, while the position and size errors remain low. NoDA has intensity error off the chart in Fig. 111. Despite this additional challenge, EnSRF_MSA_O_3 is still superior to the EnSRF for all error metrics except the vortex intensity. This result suggests that the alignment step works as long as patterns exist for the feature of interest (vortex structure in this case). When there are additional amplitude errors in the patterns being aligned, the alignment algorithm still works resiliently. These amplitude errors caused by model uncertainties, however, are independent of the position errors, they lead to higher vortex intensity errors, for which the MSA_EnSRF does not improve over the EnSRF.

4. Discussion and conclusions

In this study, we stress-test the MSA method of Ying (2019) in a simplified two-dimensional vortex model similar to that of Chen and Snyder (2007), which allows easier creation of various test scenarios and separation of different error sources (vortex position errors, structure errors, background flow errors and model errors). The asymptotic behavior of the MSA method is first demonstrated. As N_s increases, the EnSRF_MSA analysis errors decrease, approaching what can be achieved by a particle filter. We show that the best N_s depends on the degree of nonlinearity caused by initial position spread L_{spred} . As position errors increase a larger N_s is required to achieve the best performance. The observation type (u, v wind, or position observations) and the location of the observation relative to the vortex center (R_{obs}) change the information content for a single observation. The EnSRF_MSA method is less sensitive to observation location and can extract information from feature-based observations better than the original EnSRF.

A coherence assumption is made in the MSA method to reduce position errors in small scales according to large-scale analysis increments. A deviation from the coherence assumption is shown to degrade the performance of the MSA method. When background flow errors are independent (out of phase) of the vortex position errors, the interference from these background flow errors lead to misalignment of the vortices. It is unclear how frequently this incoherence between large-scale and small-scale features occurs in real geophysical prediction problems. To remedy this issue, we suggest further incorporating an adjustment term in the displacement vectors to scale their impact on the smaller scale. The adjustment term can be based on

cross-scale correlation (reflects the degree of coherence), or more ideally from a nonlinear regression process. Another option worth trying is to first decompose the model state into individual components that are incoherent with each other, then perform the MSA analysis separately for each. In general, disentangling the independent components and updating them separately seems a good strategy in DA to achieve better performance for each component.

When assimilating a network of observations, the additional decomposition of observation SCs is introduced as a new option, MSA-O, which is shown to perform better than MSA, especially when assimilating limited observation information (the targeted network), thanks to the reduced scale mismatch between observations and state variables during the filter update. The extraction of observation SCs (could be stated as a optimization problem itself) is not necessarily easy and is prone to aliasing errors, especially when observations are sparse and only with partial coverage. However, the benefits from assimilating observation SCs instead of the raw observation (e.g. Bédard and Buehner 2020; Sodhi and Fabry 2022) often outweigh these drawbacks. In our study, the MSA-O method has consistently better performance than the original EnSRF in all the test scenarios.

While this study provides further proof of concept of the MSA method and its feasibility, we do expect some more challenges in real applications. Model biases can be much more severe in real models; features of interest can be completely missing in the forecasts compared to the observations. The MSA method is not expected to work in such extremely nonlinear scenarios. On the other hand, in the quasi-linear regime, the EnSRF method is already the best algorithm so the MSA method cannot improve. Thus, in practice a test for nonlinearity (e.g., Poterjoy and Anderson 2016) may be necessary to determine which DA method to use in the analysis. When applying the MSA method, the best choice for N_s may depend on how much nonlinearity there is in the system, it may also depend on the observing networks (sparse observations may prevent the use of large N_s due to lack of information at small scales).

Arguably, the simple vortex model results can be too optimistic since the model does not capture the intricate inner-core dynamics of a real three-dimensional vortex, especially how its features interact with its environmental conditions. Complications can arise in the implementation in a full-physics multi-variable model, some treatment of topography and lower boundary conditions in deriving displacement vectors can be found in Nehrkorn et al. (2014). Future studies shall also investigate the possibility to extend the two-dimensional displacement vector fields to three- or even four-dimensional (include the time dimension to reduce timing errors).

We pay extra attention to computational cost when comparing the MSA with the original EnSRF method. Our

results suggest that - at equal cost - the MSA-O algorithm consistently outperforms the EnSRF in nonlinear regimes, suggesting that it is scalable to larger problems. In practice, the cost of tuning filter parameters, such as localization and inflation, is also non-negligible. The manual tuning of localization parameters in this study would be infeasible in realistic models. The use of adaptive algorithms may thus be needed to provide online estimation of the best parameters, without the need for a priori tuning. Based on the localization literature (e.g., Ying et al. 2018, and references therein), the best localization distance depends on the physical correlation scale of the dynamical system, the ensemble size, and the observation density. For localization, adaptive algorithms (Zhen and Zhang 2014; Moosavi et al. 2018; Cheng et al. 2021) and empirical localization functions (e.g., Anderson and Lei 2013; Lei and Anderson 2014) can be used to automate the tuning. We did not apply any covariance inflation in our study for the sake of simplicity, but for numerous cycles and more complex models, we expect inflation to be necessary. Adaptive algorithms are also available (e.g., El Gharanti 2018) to automate the tuning of inflation. In realistic applications, the truth is unknown and errors are mixed from different sources, which pose challenges to the tuning of DA algorithms. Findings from this study should provide some initial guidance for future implementation of the MSA method in more challenging scenarios.

Acknowledgments. Yue Ying thanks the NCAR Advanced Study Program for supporting this study in its early stages, this study is also supported by the basic funding of NERSC through the Research Council of Norway (RCN). We appreciate the constructive comments from Jonathan Poterjoy and two anonymous reviewers that improved an earlier version of this manuscript.

Data availability statement. A Python code package is written to perform all the experiments shown in this study. The code is open-source and available at <https://doi.org/10.5281/zenodo.6477080>. The package includes the two-dimensional vortex model, data assimilation methods (ensemble filters and the multiscale algorithms), the scripts that run all the experiments and generate all the figures in this study. The experiment data can be easily reproduced by running the Python code and they are not separately archived. Error diagnostics data for reproducing the figures are archived at <https://doi.org/10.5281/zenodo.7020672>.

APPENDIX A

The two-dimensional vortex model

The model state can be defined by a non-divergent horizontal wind field, (u, v) , on a square domain with periodic boundary conditions and no rotation. The model can be

nondimensionalized and applied to vortex systems at different scales, but in this study we formulated the model with physical units commonly found in hurricane prediction to provide some context. The model grid uses Cartesian coordinates $\mathbf{r} = (x, y)$ with length L in each direction, it has 128×128 grid points and the grid spacing is set to $\Delta x = 9$ km. The governing equation is

$$\frac{\partial \zeta}{\partial t} = -\mathbf{v} \cdot \nabla \zeta + \gamma \zeta + \nu \nabla^2 \zeta, \quad (\text{A1})$$

where $\zeta = \partial v / \partial x - \partial u / \partial y$ is the vorticity, γ is the generation rate, and ν is a dissipation coefficient.

The equation is solved in spectral space using the fourth-order Runge-Kutta numerical scheme with a time step of 60 s. Let $\mathbf{k} = (k_x, k_y)$ denote the wavenumber in (x, y) directions that form the spectral space, $\kappa = |\mathbf{k}|$ is the total wavenumber. The generation rate is spatially varying and defined in spectral space as

$$\hat{\gamma}(\kappa) = \begin{cases} \xi e^{-\frac{(\kappa-\kappa_g)^2}{2\sigma_\kappa^2}} & \text{if } \max(V) < V_c \\ 0 & \text{if } \max(V) \geq V_c, \end{cases} \quad (\text{A2})$$

where $\hat{\gamma}$ is the Fourier coefficients of γ , ξ is a scalar generation coefficient, V_c is the critical wind speed that limits the generation, κ_g is the center wavenumber and σ_κ is the spectral bandwidth of the generation rate. The generation rate has a Gaussian spectral shape so that generation occurs mostly at wavenumbers $\kappa_g \pm \sigma_\kappa$.

The wind field consists of a Rankine vortex embedded in a background flow that both evolve with time. The initial Rankine vortex is created so that its center position is $\mathbf{r}_c = (x_c, y_c)$ and its radial wind profile follows

$$V_{\text{vort}}(R) = \begin{cases} V_{\text{max}} \frac{R}{R_{\text{mw}}} & \text{if } R \leq R_{\text{mw}} \\ V_{\text{max}} \left(\frac{R_{\text{mw}}}{R} \right)^a & \text{if } R > R_{\text{mw}}, \end{cases} \quad (\text{A3})$$

where $V_{\text{vort}} = |\mathbf{v}_{\text{vort}}|$, $R = |\mathbf{r} - \mathbf{r}_c|$ is the radius relative to the center, V_{max} is the vortex maximum wind speed, R_{mw} is the radius of maximum wind, and a is a shape parameter.

The background flow is initialized at $t = 0$ as a random wind field with a spectrum $\hat{V}_{\text{bkg}}(\kappa) \propto \kappa^{-3}$ and wind speed $V_{\text{bkg}} = |\mathbf{v}_{\text{bkg}}|$. The average wind speed \bar{V}_{bkg} (the bar denotes a spatial average) can be specified to control the relative strength of the initial background flow, for example $\bar{V}_{\text{bkg}} = 0$ turns off the background flow. To generate the random wind field, we first draw a random vorticity field $\hat{\zeta}_{\text{bkg}}(\kappa) \propto \kappa^{-1}$, convert the vorticity to wind so that the wind is non-divergent and has the correct -3 power law, then normalize and scale the wind speed to \bar{V}_{bkg} .

APPENDIX B

The scale decomposition method

Let N_s be the number of SCs. Given $L = 128\Delta x$, the largest meaningful wavenumber is $\kappa_{\max} = 16$ in this study ($8\Delta x$ resolves a complete sine wave). The characteristic wavenumber for each SC can be defined as $\kappa_s = \kappa_{\max}^{s/N_s}$, for $s = 1, \dots, N_s$. For the s th SC, a spectral lowpass response function is defined as

$$\hat{f}_s^L(\mathbf{\kappa}) = \begin{cases} 1 & \text{if } |\mathbf{\kappa}| < \kappa_s \\ \left[\cos\left(\frac{|\mathbf{\kappa}| - \kappa_s}{\kappa_{s+1} - \kappa_s}\right) \right]^2 & \text{if } \kappa_s \leq |\mathbf{\kappa}| \leq \kappa_{s+1} \\ 0 & \text{if } |\mathbf{\kappa}| > \kappa_{s+1}, \end{cases} \quad (\text{B1})$$

and a bandpass response function is then constructed by

$$\hat{f}_s = \begin{cases} \hat{f}_s^L & \text{if } s = 1 \\ \hat{f}_s^L - \hat{f}_{s-1}^L & \text{if } 1 < s < N_s \\ 1 - \hat{f}_{s-1}^L & \text{if } s = N_s, \end{cases} \quad (\text{B2})$$

In vector form, the spectral-space response function is $\hat{\mathbf{f}}_s = \hat{f}_s(\mathbf{\kappa})$, corresponding to $\mathbf{f}_s = f_s(\mathbf{r})$ in physical space. Since model states are defined on a uniform grid, it is easy to transform the states to spectral space, apply the response function in an element-wise product $\hat{\psi}_s = \hat{\mathbf{f}}_s \circ \hat{\psi}$, then transform back to get the s th state SC ψ_s .

Observation SCs are more difficult to derive because of the irregular grid. According to the Convolution Theorem, an element-wise product $\hat{\mathbf{f}}_s \circ \hat{\psi}$ in spectral space is equivalent to a convolution $\mathbf{f}_s * \psi$ in physical space. In this study, the convolution is done through a series of spatially weighted averages. The s th observation SC can be obtained by $\phi_s^o = \mathbf{f}_s * \phi^o$, whose i th element is

$$\phi_{s,i}^o = \sum_{j=1}^{N_o} f_s(\mathbf{r}_i - \mathbf{r}_j) \phi_j^o, \quad (\text{B3})$$

where \mathbf{r}_i is the position of ϕ_i^o . The low-pass filter can also be implemented more efficiently using a diffusion-based algorithm (e.g., Grooms et al. 2021) for the irregular grids. Following Buehner and Shlyaeva (2015), we allow some spectral overlapping between SCs by using a squared cosine transitioning from 1 to 0 in the response functions. This overlapping makes the SC computation more numerically stable, it also allows stronger error correlation among SCs which is more suitable for the MSA.

Hereafter we express the scale decomposition operation in matrix form as $\psi_{n,s} = \mathbf{F}_s \psi_n$ for the model states, $\phi_s^o = \mathbf{F}_s^o \phi^o$ for the observations, and $\phi_{n,s}^b = \mathbf{F}_s^o h(\psi_n^b)$ for the observation priors. Each row of \mathbf{F}_s contains the averaging kernel \mathbf{f}_s , and \mathbf{F}_s^o only differs from \mathbf{F}_s due to different geometry of the model grid and the observing network.

Spatial inhomogeneity in the observation grid will result in inaccuracies in the derived SCs. One can pad the data voids with fake observations (zeros; will not be assimilated) to achieve a relatively even distribution. Of course,

the same padding needs to be applied to the observation priors (simulated observations from model states) to derive a consistent innovation.

We use a reduced-dimension grid to represent large-scale SCs, which improves the efficiency of the multiscale approaches. The necessary grid dimension decreases with κ_s . The EnKF update is performed on the coarse grid and increments are then refined (with bilinear interpolation) and added to the native model grid, which is a common approach for multiscale/multigrid DA and in incremental 4DVar (Courtier et al. 1994).

APPENDIX C

The multiscale alignment (MSA) method

The EnKF update equation [Eqs. 13–14 in Burgers et al. (1998)] can be restated as

$$\Psi_n^a = \Psi_n^b + \mathbf{L} \circ \frac{\text{cov}(\psi^b, \phi^b)}{\text{cov}(\phi^b, \phi^b) + \mathbf{R}} (\phi^o - \phi_n^b), \quad (\text{C1})$$

for $n = 1, \dots, N$, where ψ_n^b is the prior state, ϕ^o is the observation, $\phi_n^b = h(\psi_n^b)$ is the observation prior, ensemble-estimated error covariances (cov) are computed according to (7), \mathbf{R} is the specified observation error covariance, and \mathbf{L} is the localization to remove spurious error covariances. The localization factor for assimilating the i th observation and updating the j th state variable is

$$L_{i,j} = \alpha \times \text{GC}(|\mathbf{r}_i - \mathbf{r}_j|, \text{ROI}), \quad (\text{C2})$$

where GC is the Gaspari and Cohn (1999) localization function with a parameter called radius of influence (ROI; the distance at which the covariance is tapered to zero), α is an additional amplitude parameter (similar to the one used in Lei and Whitaker 2017).

The MSA method (Ying 2019) applies the EnKF update (C1) sequentially for the SCs. Algorithm 1 provides a pseudo-code description. Note that each EnKF update makes an “analysis increment” $\Psi_{n,s}^a - \Psi_{n,s}^b$, while all the iterations lead to the “final analysis” Ψ_n^a .

We keep a general form of an EnKF update in the algorithm, indicating that different EnKF variants can be used. In this study, we choose to use the ensemble square root filter variant (EnSRF; Whitaker and Hamill 2002; Tippett et al. 2003), which assimilates observations one at a time assuming their errors to be uncorrelated $\mathbf{R} = \sigma_o^2 \mathbf{I}$. We only consider the scenario when σ_o is specified correctly during DA, and performed some manual tuning to optimize the localization parameters.

After obtaining the analysis increment, displacement vectors \mathbf{q} are derived by minimizing the cost function $\|\psi^b(\mathbf{q}) - \psi^a\|^2 + w \|\nabla \mathbf{q}\|^2$ using the Horn and Schunck (1981) optical flow algorithm. $\psi(\mathbf{q})$ is a shorthand to represent the warped model states, $u(\mathbf{r} + \mathbf{q})$ and $v(\mathbf{r} + \mathbf{q})$. Note

Algorithm 1 The MSA (if `decompose_obs = False`) and MSA-O (if `decompose_obs = True`) methods. Note that subscript n implies that the operation is repeated for $n = 1, \dots, N$.

Input: prior ensemble Ψ_n , observations Φ^o
Output: iteratively updated Ψ_n as final analysis ensemble
Parameters: number of scales N_s , localization function \mathbf{L}_s (ROI_s, α_s), observation error $\sigma_{o,s}^2$, and smoothness constraint w

```

1: for  $s$  in  $1, \dots, N_s$  do
2:    $\Psi_{n,s}^b = \mathbf{F}_s \Psi_n$ 
3:    $\Phi_n^b = h(\Psi_n)$ 
4:   if decompose_obs then
5:      $\Phi_s^o = \mathbf{F}_s^o \Phi^o$ 
6:      $\Phi_{n,s}^b = \mathbf{F}_s^o h(\Psi_n)$ 
7:      $\Psi_{n,s}^a = \Psi_{n,s}^b + \mathbf{L}_s \circ \frac{\text{cov}(\Psi_s^b, \Phi_s^b)}{\text{cov}(\Phi_s^b, \Phi_s^b) + \sigma_{o,s}^2 \mathbf{I}} (\Phi_s^o - \Phi_{n,s}^b)$ 
8:   else
9:      $\Psi_{n,s}^a = \Psi_{n,s}^b + \mathbf{L}_s \circ \frac{\text{cov}(\Psi_s^b, \Phi_s^b)}{\text{cov}(\Phi_s^b, \Phi_s^b) + \sigma_{o,s}^2 \mathbf{I}} (\Phi^o - \Phi_n^b)$ 
10:  end if
11:  if  $s < N_s$  then
12:     $\mathbf{q}_{n,s} = \underset{\mathbf{q}}{\text{argmin}} \|\Psi_{n,s}^b(\mathbf{q}) - \Psi_{n,s}^a\|^2 + w \|\nabla \mathbf{q}\|^2$ 
13:     $\Psi_n \leftarrow \Psi_n(\mathbf{q}_{n,s}) + \Psi_{n,s}^a - \Psi_{n,s}^b(\mathbf{q}_{n,s})$ 
14:  else
15:     $\Psi_n \leftarrow \Psi_n + \Psi_{n,s}^a - \Psi_{n,s}^b$ 
16:  end if
17: end for

```

that \mathbf{q} is a spatially varying vector field, unlike the constant position perturbations \mathbf{r}'_n in (3). The smoothness constraint is set to $w = 1$ [not tuned, see Horn and Schunck (1981) for a discussion of its effect]. The displacement vectors are then applied to warp the model state to reduce position errors at smaller scales, the warping is done through bi-linear interpolation to evaluate $\Psi(\mathbf{q})$.

We proposed and tested a new option to decompose observations into SCs corresponding to the state SCs, which we call MSA-O (`decompose_obs = True` in Algorithm 1). This option allows observations to be closer to the model states (less scale mismatch) during assimilation, improving the filter performance. Note that, for the MSA-O, the observation errors also undergo scale decomposition (B3) so that the observation error standard deviation for Φ_s^o is adjusted to

$$\sigma_{o,s} = \sigma_o f_s(\mathbf{0}), \quad (\text{C3})$$

i.e., large-scale observation error is reduced.

References

- Aksoy, A., 2013: Storm-relative observations in tropical cyclone data assimilation with an ensemble kalman filter. *Mon. Wea. Rev.*, **141** (2), 506–522, <https://doi.org/10.1175/MWR-D-12-00094.1>.
- Amezcuia, J., and P. J. van Leeuwen, 2014: Gaussian anamorphosis in the analysis step of the enkf: a joint state-variable/observation approach. *Tellus*, **66** (1), 23 493, <https://doi.org/10.3402/tellusa.v66.23493>.
- Anderson, J., and L. Lei, 2013: Empirical localization of observation impact in ensemble Kalman filters. *Mon. Wea. Rev.*, **141** (11), 4140–4153, <https://doi.org/10.1175/MWR-D-12-00330.1>.
- Anderson, J. L., 2010: A non-gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138** (11), 4186–4198, <https://doi.org/10.1175/2010MWR3253.1>.
- Anderson, J. L., 2019: A nonlinear rank regression method for ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **147** (8), 2847–2860, <https://doi.org/10.1175/MWR-D-18-0448.1>.
- Bédard, J., and M. Buehner, 2020: A practical assimilation approach to extract smaller-scale information from observations with spatially correlated errors: An idealized study. *Quart. J. Roy. Meteor. Soc.*, **146** (726), 468–482, <https://doi.org/10.1002/qj.3687>.
- Beezley, J. D., and J. Mandel, 2008: Morphing ensemble Kalman filters. *Tellus*, **60** (1), 131–140, <https://doi.org/10.1111/j.1600-0870.2007.00275.x>.
- Bishop, C. H., 2016: The gigg-enkf: ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Quart. J. Roy. Meteor. Soc.*, **142** (696), 1395–1412, <https://doi.org/10.1002/qj.2742>.
- Bocquet, M., and P. Sakov, 2014: An iterative ensemble Kalman smoother. *Quart. J. Roy. Meteor. Soc.*, **140** (682), 1521–1535, <https://doi.org/10.1002/qj.2236>.
- Buehner, M., and A. Shlyaeva, 2015: Scale-dependent background-error covariance localisation. *Tellus*, **67** (1), 28 027, <https://doi.org/10.3402/tellusa.v67.28027>.
- Burgers, G., P. J. Van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126** (6), 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2).
- Caron, J.-F., and M. Buehner, 2018: Scale-dependent background error covariance localisation: Evaluation in a global deterministic weather forecasting system. *Mon. Wea. Rev.*, **146** (5), 1367–1381, <https://doi.org/10.1175/MWR-D-17-0369.1>.
- Carrassi, A., M. Bocquet, L. Bertino, and G. Evensen, 2018: Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, **9** (5), e535, <https://doi.org/10.1002/wcc.535>.
- Chen, Y., and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.*, **135** (5), 1828–1845, <https://doi.org/10.1175/MWR3351.1>.
- Cheng, S., J.-P. Argaud, B. Iooss, A. Ponçot, and D. Lucor, 2021: A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping. *Mathematical Geosciences*, **53** (8), 1751–1780, <https://doi.org/10.1007/s11004-021-09951-z>.
- Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4d-var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120** (519), 1367–1387, <https://doi.org/10.1002/qj.49712051912>.
- Doucet, A., N. De Freitas, N. J. Gordon, and Coauthors, 2001: *Sequential Monte Carlo methods in practice*, Vol. 1. Springer, <https://doi.org/10.1007/978-1-4757-3437-9>.

- El Gharamti, M., 2018: Enhanced adaptive inflation algorithm for ensemble filters. *Mon. Wea. Rev.*, **146** (2), 623–640, <https://doi.org/10.1175/MWR-D-17-0187.1>.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162, <https://doi.org/10.1029/94JC00572>.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125** (554), 723–757, <https://doi.org/10.1002/qj.49712555417>.
- Grooms, I., N. Loose, R. Abernathy, J. Steinberg, S. D. Bachman, G. Marques, A. P. Guillaumin, and E. Yankovsky, 2021: Diffusion-based smoothers for spatial filtering of gridded geophysical data. *Journal of Advances in Modeling Earth Systems*, **13** (9), e2021MS002 552, <https://doi.org/10.1029/2021MS002552>.
- Hodyss, D., J. L. Anderson, N. Collins, W. F. Campbell, and P. A. Reinecke, 2017: Quadratic polynomial regression using serial observation processing: Implementation within dart. *Mon. Wea. Rev.*, **145** (11), 4467–4479, <https://doi.org/10.1175/MWR-D-17-0089.1>.
- Hoffman, R. N., Z. Liu, J.-F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.*, **123** (9), 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., and B. G. Schunck, 1981: Determining optical flow. *Artificial intelligence*, **17** (1-3), 185–203, [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2).
- Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144** (12), 4489–4532, <https://doi.org/10.1175/MWR-D-15-0440.1>.
- Lei, L., and J. L. Anderson, 2014: Empirical localization of observations for serial ensemble Kalman filter data assimilation in an atmospheric general circulation model. *Mon. Wea. Rev.*, **142** (5), 1835–1851, <https://doi.org/10.1175/MWR-D-13-00288.1>.
- Lei, L., and J. S. Whitaker, 2017: Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *Journal of Advances in Modeling Earth Systems*, **9** (2), 781–789, <https://doi.org/10.1002/2016MS000864>.
- 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** (9), 3804–3822, <https://doi.org/10.1175/MWR-D-14-00384.1>.
- Lin, K.-J., S.-C. Yang, and S. S. Chen, 2018: Reducing tc position uncertainty in an ensemble data assimilation and prediction system: A case study of typhoon fanapi (2010). *Weather and Forecasting*, **33** (2), 561–582, <https://doi.org/10.1175/WAF-D-17-0152.1>.
- Liu, Q., and Coauthors, 2020: Vortex initialization in the ncep operational hurricane models. *Atmosphere*, **11** (9), 968, <https://doi.org/10.3390/atmos11090968>.
- 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>.
- Moosavi, A., A. Attia, and A. Sandu, 2018: A machine learning approach to adaptive covariance localization. *arXiv preprint arXiv:1801.00548*, <https://doi.org/10.48550/arXiv.1801.00548>.
- Navarro, E. L., and G. J. Hakim, 2014: Storm-centered ensemble data assimilation for tropical cyclones. *Mon. Wea. Rev.*, **142** (6), 2309–2320, <https://doi.org/10.1175/MWR-D-13-00099.1>.
- Nehrkorn, T., B. Woods, T. Auligné, 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** (2), 686–702, <https://doi.org/10.1175/MWR-D-13-00164.1>.
- Nehrkorn, T., B. K. Woods, R. N. Hoffman, and T. Auligné, 2015: Correcting for position errors in variational data assimilation. *Mon. Wea. Rev.*, **143** (4), 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** (1), 59–76, <https://doi.org/10.1175/MWR-D-15-0163.1>.
- Poterjoy, J., 2022: Implications of multivariate non-gaussian data assimilation for multi-scale weather prediction. *Mon. Wea. Rev.*, <https://doi.org/10.1175/MWR-D-21-0228.1>.
- Poterjoy, J., and J. L. Anderson, 2016: Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Mon. Wea. Rev.*, **144** (5), 2007–2020, <https://doi.org/10.1175/MWR-D-15-0322.1>.
- Poterjoy, J., 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** (5), 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** (1-2), 127–145, <https://doi.org/10.1016/j.physd.2006.09.035>.
- Sakov, P., D. S. Oliver, and L. Bertino, 2012: An iterative enkf for strongly nonlinear systems. *Mon. Wea. Rev.*, **140** (6), 1988–2004, <https://doi.org/10.1175/MWR-D-11-00176.1>.
- Simon, E., and L. Bertino, 2009: Application of the gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the north atlantic with the enkf: a twin experiment. *Ocean Science*, **5** (4), 495–510, <https://doi.org/10.5194/os-5-495-2009>.
- Sodhi, J. S., and F. Fabry, 2022: Benefits of smoothing backgrounds and radar reflectivity observations for multiscale data assimilation with an ensemble Kalman filter at convective scales: A proof of concept study. *Mon. Wea. Rev.*, <https://doi.org/10.1175/MWR-D-21-0130.1>.
- Stratman, D. R., C. K. Potvin, and L. J. Wicker, 2018: Correcting storm displacement errors in ensembles using the feature alignment technique (fat). *Mon. Wea. Rev.*, **146** (7), 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** (7), 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., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137** (12), 4089–4114, <https://doi.org/10.1175/2009MWR2835.1>.
- van Leeuwen, P. J., 2015: *Nonlinear Data Assimilation for high-dimensional systems*, 1–73. Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-18347-3_1.

- Wang, X., H. G. Chipilski, C. H. Bishop, E. Satterfield, N. Baker, and J. S. Whitaker, 2021: A multiscale local gain form ensemble transform Kalman filter (mlgetkf). *Mon. Wea. Rev.*, **149** (3), 605–622, <https://doi.org/10.1175/MWR-D-20-0290.1>.
- Weng, Y., and F. Zhang, 2012: Assimilating airborne doppler radar observations with an ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). *Mon. Wea. Rev.*, **140** (3), 841–859, <https://doi.org/10.1175/2011MWR3602.1>.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130** (7), 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).
- Ying, Y., 2019: A multiscale alignment method for ensemble filtering with displacement errors. *Mon. Wea. Rev.*, **147** (12), 4553–4565, <https://doi.org/10.1175/MWR-D-19-0170.1>.
- Ying, Y., 2020: Assimilating observations with spatially correlated errors using a serial ensemble filter with a multiscale approach. *Mon. Wea. Rev.*, **148** (8), 3397–3412, <https://doi.org/10.1175/MWR-D-19-0387.1>.
- Ying, Y., F. Zhang, and J. L. Anderson, 2018: On the selection of localization radius in ensemble filtering for multiscale quasigeostrophic dynamics. *Mon. Wea. Rev.*, **146** (2), 543–560, <https://doi.org/10.1175/MWR-D-17-0336.1>.
- Zhang, F., Y. Weng, J. A. Sippel, Z. 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** (7), 2105–2125, <https://doi.org/10.1175/2009MWR2645.1>.
- Zhen, Y., and F. Zhang, 2014: A probabilistic approach to adaptive covariance localization for serial ensemble square root filters. *Mon. Wea. Rev.*, **142** (12), 4499–4518, <https://doi.org/10.1175/MWR-D-13-00390.1>.