A Comparison of Two Local Moment-Matching Nonlinear Filters: Local Particle Filter (LPF) and Local Nonlinear Ensemble Transform Filter (LNETF)

JIE FENG,[a] XUGUANG WANG,[a] AND JONATHAN POTERJOY[b,c]

[a] *School of Meteorology, University of Oklahoma, Norman, Oklahoma;* [b] *Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland;* [c] *Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*

ABSTRACT: The local particle filter (LPF) and the local nonlinear ensemble transform filter (LNETF) are two moment-matching nonlinear filters to approximate the classical particle filter (PF). They adopt different strategies to alleviate filter degeneracy. LPF and LNETF localize observational impact but use different localization functions. They assimilate observations in a partially sequential and a simultaneous manner, respectively. In addition, LPF applies the resampling step, whereas LNETF applies the deterministic square root transformation to update particles. Both methods preserve the posterior mean and variance of the PF. LNETF additionally preserves the posterior correlation of the PF for state variables within a local volume. These differences lead to their differing performance in filter stability and posterior moment estimation. LPF and LNETF are systematically compared and analyzed here through a set of experiments with a Lorenz model. Strategies to improve the LNETF are proposed. The original LNETF is inferior to the original LPF in filter stability and analysis accuracy, particularly for small particle numbers. This is attributed to both the localization function and particle update differences. The LNETF localization function imposes a stronger observation impact than the LPF for remote grids and thus is more susceptible to filter degeneracy. The LNETF update causes an overall narrower range of posteriors that excludes true states more frequently. After applying the same localization function as the LPF and additional posterior inflation to the LNETF, the two filters reach similar filter stability and analysis accuracy for all particle numbers. The improved LNETF shows more accurate posterior probability distribution but slightly worse spatial correlation of posteriors than the LPF.

KEYWORDS: Bayesian methods; Filtering techniques

1. Introduction

Data assimilation (DA) estimates the state of a dynamical system by combining information from short-range model forecasts (also called background or prior) and observations. In the Bayesian framework, the analysis (or posterior) accuracy relies on the quality of the estimated probability density function (PDF) for the prior and the observation likelihood.

For the widely used ensemble-based approach such as the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Mitchell 1998; Anderson 2001; Whitaker and Hamill 2002; Hunt et al. 2007) and ensemble-variational hybrid schemes (EnVar; Hamill and Snyder 2000; Lorenc 2003; Buehner 2005; Wang et al. 2007, 2008a,b, 2013; Wang and Lei 2014), the PDFs of the background and the observation likelihood are assumed to be Gaussian. Consequently, the probability distribution of the analysis states are derived via the estimated posterior mean and covariance. For practical applications, highly nonlinear model dynamics (Bocquet et al. 2010; Legrand et al. 2016) or nonlinear observation operators (Pires et al. 2010) could lead to non-Gaussian distributions of background forecast states and observation likelihoods. Such non-Gaussian behaviors of the background and the observation likelihood would yield a suboptimal performance of the EnKF and EnVar assimilation.

Different from EnKF and EnVar, the particle filter (PF; see Doucet et al. 2001 and van Leeuwen 2009 for a review) does not rely on the Gaussian assumption for the probability distribution of prior particles and observation likelihoods. The PF estimates the PDF of posterior particles (or ensemble) via weighting each prior particle conditioned on the assimilated observations. These weights are calculated by the likelihood of observations, which varies with the specific observation PDF.

Despite the theoretical benefits of PFs for non-Gaussian applications, various problems still constrain its potential use for high-dimensional dynamical systems. One of the most serious challenges is the so-called "filter degeneracy" (Doucet et al. 2001; Snyder et al. 2008). When using a limited number of particles to sample the prior distribution in a high-dimensional state space, most of the particles may not lie in the significant probability region determined by the likelihood of assimilated observations. During the update or resampling of particles, prior particles with small weights are typically removed, while others with much larger weights are duplicated (e.g., Gordon et al. 1993). After the successive assimilation of independent observations with relatively small errors, the weights will collapse onto a single particle, leading to the filter degeneracy. Previous studies have demonstrated that the smallest number of particles needed to prevent filter degeneracy roughly grow exponentially with the effective dimension of a system (Bengtsson et al. 2008; Bickel et al. 2008; Snyder et al. 2008). Therefore, it is impractical for the direct application of the classical PF for

problems such as operational atmospheric and oceanic DA, where the effective dimension is extremely high.

The filtering problem arising from a limited ensemble size is not unique to the PF. For the EnKF and EnVar, a small ensemble can produce large sampling errors (Anderson 2001) and the significant rank deficiency (e.g., Hamill and Snyder 2000) during the background covariance estimation, and thereby cause an unstable filter or even filter divergence. Therefore, various solutions are proposed to ensure the algorithms remain stable over longer time intervals. The most widely used solution is to limit the impact of observations in a local region (Mitchell et al. 2002). This process can be achieved by modulating the ensemble background error covariance with a function whose value decreases with an increasing distance (Houtekamer and Mitchell 2001; Whitaker and Hamill 2002; Holland and Wang 2013), or by inflating the observation errors when the observations are farther away from the grid point of interest (Ott et al. 2004; Szunyogh et al. 2005; Huang et al. 2019). Plenty of studies have demonstrated that the localization effectively removes spurious long-range error correlations (Oke et al. 2007), arriving at much improved filter stability.

Similar localization strategies have already been applied to several non-Gaussian filters built on the framework of ensemble-based Kalman filter. For example, the localization of background covariance in local-local Gaussian filter (Bengtsson et al. 2003) and the rank histogram filter (Anderson 2010), and the volume-based localization in the moment-matching filter (Lei and Bickel 2011). Unlike the aforementioned studies, PFs do not rely explicitly on prior covariances, but the similar localization concept is feasible for PFs. This study focuses on two recently proposed localized approaches to approximate PF solutions. One approach is termed as the local PF (LPF; Poterjoy 2016), the other is termed as the local nonlinear ensemble transform filter (LNETF; Tödter and Ahrens 2015; Tödter et al. 2016). Note that the localization alone does not fix the filter degeneracy issue for both approaches. Several additional treatments are applied to stabilize the filter. As a result, the posteriors from both methods only preserve limited moments of the PF,[1] similar to the moment-matching approach (e.g., Lei and Bickel 2011). The LPF processes observations sequentially. Like the traditional PFs, the LPF applies a bootstrap resampling of particles based on the weights determined by the observation likelihood calculations. In the classical PF, each prior state vector is assigned a scalar weight, i.e., each state element has a constant weight. To realize localization, the LPF first extends the scalar weights to spatially varying vector quantities. A localization coefficient is then applied to limit the impact of a given observation to a local physical region by gradually homogenizing the particle weights for more distant grid points. Posterior mean and variance are first calculated using the localized PF weights and the original prior particles. The particles are then updated by applying additional regularization and particle mixing steps (Poterjoy et al. 2019),

which are needed to stabilize the filter. These additional steps result in posterior particles that only match the mean and variance of the PF posterior density functions.

The LNETF adopts a similar framework as the local ensemble transform Kalman filter (LETKF, Hunt et al. 2007), which updates the ensemble at each grid point in parallel by considering only the observations within a local subdomain centered at the given grid. Like the LETKF, the LNETF applies a deterministic square root algorithm that uses transform matrices to update the prior. Unlike the LETKF, the weights used to transform prior particles are derived so that the posterior particles match the first two moments, namely the mean and covariance, of the PF-estimated posterior within each local subdomain. Similar to the LPF, additional steps are necessary to stabilize the filter in addition to the localization, such as the variance inflation for prior particles and the transformation of a random rotation matrix.

Previous studies have shown that both LPF and LNETF can be applied for high-dimensional systems without filter divergence even using a limited number of particles (Poterjoy and Anderson 2016; Poterjoy et al. 2019; Tödter et al. 2016). The LPF and LNETF, however, are distinct in various aspects. First, the former processes observations sequentially, while the latter assimilates observations simultaneously. Past EnKF comparisons suggest sequential and simultaneous updates can lead to different filter performance (e.g., Ehrendorfer 2007; Holland and Wang 2013). Second, they apply different functions for localization, although the common aim is to taper observational impact with an increasing distance. The third distinction, which may be the most significant, lies in their algorithmic differences in updating the particles to match limited moments of the PF. The impact of the above differences on filter stability and posterior moment estimation (e.g., mean, covariance, and the full PDF) is still unclear and of great interest. Additionally, the practical application of PFs for high-dimension systems is still in the early stage of exploration. An understanding of the differences of existing approaches that approximate the PF can advance the theoretical and practical development of the PF algorithm. Therefore, the goal of the present study is to systematically compare the two forms of localized nonlinear moment-matching methods, namely the LPF and the LNETF. Focus will be placed on the filter performance in the presence of differing localization functions and particle update algorithms. In addition, strategies to potentially improve the two methods are explored.

Note that the LPF applies a secondary adjustment to particles, after making the posteriors match the mean and variance. This step performs an additional univariate probability mapping of marginal state variables. The cumulative distribution function (CDF) of original PF weights is mapped to that of an equally weighted kernel estimate of the updated particles—denoted the kernel density distribution mapping (KDDM) in Poterjoy (2016). This step would, to some extent, account for higher moments. Given that the same step can be applied to the LNETF and our aforementioned foci in comparing the two methods, KDDM is not implemented in either filter for this comparison.

---

[1] Preserving a specific moment of the PF means adopting the same update algorithm as the PF to derive the posterior moment given particle weights.

The paper is organized in the following manner. Section 2 introduces the algorithms for the classical PF and its two approximations, the LPF and LNETF, with a focus on their specific means of alleviating filter degeneracy. Section 3 describes the configurations for the model and DA systems and discusses the experimental designs. Section 4 displays and analyzes experiment results, and section 5 provides conclusions and a discussion.

2. Methods

a. The particle filter

The foundation for the PF is Bayes's theorem, which quantifies the probability distribution of posterior particles given new pieces of observation information (Doucet et al. 2001). Consider an N_x-dimensional nonlinear dynamical system including $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_{N_x})^{\mathrm{T}}$, which is a random system state, and $\mathbf{y}$, which is an observational vector of length N_y. The observation relates to the true state $\mathbf{x}^t$ through the following:

$$\mathbf{y} = H[\mathbf{x}^t] + \boldsymbol{\varepsilon}, \tag{1}$$

where $H[]$ is an observation operator that maps the system states onto the observation space, and $\boldsymbol{\varepsilon}$ is the observational error. From Bayes's theorem, the PDF of the system state $\mathbf{x}$ conditioned on the observation $\mathbf{y}$ is calculated as follows:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{2}$$

where $p(\mathbf{y}|\mathbf{x})$ is the likelihood of observations, $p(\mathbf{x})$ is the PDF of prior particles and $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\,d\mathbf{x}$. Let $f(\mathbf{x})$ be any function of prior particles. Its posterior expectation given observation $\mathbf{y}$ is written as follows:

$$E[f(\mathbf{x}|\mathbf{y})] = \int f(\mathbf{x})p(\mathbf{x}|\mathbf{y})\,d\mathbf{x} = \int f(\mathbf{x})\frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}\,d\mathbf{x}. \tag{3}$$

Following Eq. (3), assuming $f(\mathbf{x}) = \mathbf{x}$, the first moment (i.e., the mean) of posterior particles can be obtained through a Monte Carlo approximation:

$$\overline{\mathbf{x}^a} = \sum_{n=1}^{N_e} w_n \mathbf{x}_n^f. \tag{4}$$

The superscripts a and f denote the posterior and prior particles, respectively, N_e is the number of particles, and $w_n = p(\mathbf{y}|\mathbf{x}_n^f)/\sum_{n=1}^{N_e} p(\mathbf{y}|\mathbf{x}_n^f)$ is the normalized observation likelihood or, in other words, the normalized weight of prior particle $\mathbf{x}_n^f$ given observation $\mathbf{y}$. The second moment (i.e., the covariance) of posterior particles $\mathbf{P}^a$ is computed given $f(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}^a})(\mathbf{x} - \overline{\mathbf{x}^a})^{\mathrm{T}}$:

$$\mathbf{P}^a = \sum_{n=1}^{N_e} w_n (\mathbf{x}_n^f - \overline{\mathbf{x}^a})(\mathbf{x}_n^f - \overline{\mathbf{x}^a})^{\mathrm{T}}. \tag{5}$$

Note that as shown in Eq. (4), given a new observation, the PF imposes a scalar weight on each prior state vector, which allows the observation to influence the entire state vector. For a limited number of particles, this will lead to inaccurate sampling of the posterior distribution for a high-dimensional system.

b. The local particle filter (LPF)

The LPF method used in this study followed the latest advances in Poterjoy et al. (2019). This section highlights various strategies applied in the LPF to alleviate the filter degeneracy.

1) INFLATION OF OBSERVATIONAL ERROR

When assimilating observations with very small error standard deviation (Std) σ, the likelihood function spans a narrow region so that only few particles are given significant weights. It inevitably underestimates the ensemble spread in subsequent DA cycles and increases the risk of filter degeneracy. To overcome the problem, Poterjoy et al. (2019) applied the adaptive inflation of observational error variance (β inflation hereinafter) and found it dramatically improved the filter stability, especially for small observational errors.

The β inflation strategy estimates an inflation coefficient β ($\beta \geq 1$) of observational error variance to ensure that the ratio between the effective particle number N_{eff} and the actual number of prior particles N_e reaches a given threshold r_{eff} ($0 < r_{\mathrm{eff}} \leq 1$; hereinafter called the inflation parameter). Let w_{n,y_i} denote the normalized weight for the nth prior particle given the ith observation y_i, N_{eff} is then calculated by $N_{\mathrm{eff}} = \left(\sum_{n=1}^{N_e} w_{n,y_i}^2\right)^{-1}$. If observations have Gaussian likelihoods, w_{n,y_i} with β inflation is written as follows:

$$w_{n,y_i} = \exp\left\{-\frac{(y_i - H_i[\mathbf{x}_n^f])^2}{2\beta_i \sigma_{y_i}^2}\right\} \bigg/ \sum_{n=1}^{N_e} \exp\left\{-\frac{(y_i - H_i[\mathbf{x}_n^f])^2}{2\beta_i \sigma_{y_i}^2}\right\}, \tag{6}$$

where $\sigma_{y_i}^2$ represents the observational error variance of y_i and $H_i[]$ is the operator that generates the observation prior for y_i; β is calculated for each observation before assimilating any of them. More details concerning the β inflation can be referred to Poterjoy et al. (2019).

2) LOCALIZATION OF OBSERVATIONAL IMPACT

Considering the decreasing impact of observations with an increasing distance, the particle weights would become more homogeneous for remoter grids from a given observation. Mathematically, it is achieved as the following:

$$\omega_{n,y_i,j} = \left(w_{n,y_i} - 1/N_e\right) l_{y_i,j} + 1/N_e, \tag{7}$$

where $\omega_{n,y_i,j}$ is the localized particle weight as a function of not only the specific particle n and observation y_i but also the grid point ($j = 1, 2, \ldots, N_x$) in space. The localization effect is achieved by tapering the deviation of particle weights from their mean value $1/N_e$ with a correlation coefficient $l_{y_i,j}$ that decays with distance (hereinafter the loc-D scheme); $l_{y_i,j}$ usually uses a fifth-order polynomial function (Gaspari and Cohn 1999), which has a Gaussian-like shape. The localized particle weights are iteratively updated and normalized with observations sequentially assimilated:

$$\omega_{n,y_{1:i},j} = \omega_{n,y_{1:i-1},j} \, \omega_{n,y_i,j}. \tag{8}$$

The initial weights are homogeneous for each grid and particle, namely equal to $1/N_e$.

After calculating the weights $\omega_{n,y_{1:i},j}$, the LPF obtains the posterior mean and variance for the first i observations assimilated following the generic equations of the PF, i.e., Eqs. (4) and (5):

$$\overline{\mathbf{x}_{y_{1:i}}^a} = \sum_{n=1}^{N_e} \mathbf{\Omega}_{n,y_{1:i}} \circ \mathbf{x}_n^f \quad \text{and} \quad (9)$$

$$\mathbf{V}_{y_{1:i}}^a = \sum_{n=1}^{N_e} \mathbf{\Omega}_{n,y_{1:i}} \circ \left(\mathbf{x}_n^f - \overline{\mathbf{x}_{y_{1:i}}^a}\right) \circ \left(\mathbf{x}_n^f - \overline{\mathbf{x}_{y_{1:i}}^a}\right), \quad (10)$$

where $\mathbf{\Omega}_{n,y_{1:i}} = (\omega_{n,y_{1:i},1}, \omega_{n,y_{1:i},2}, \ldots, \omega_{n,y_{1:i},Nx})^{\mathrm{T}}$ and $\mathbf{V}_{y_{1:i}}^a = (V_{y_{1:i},1}, V_{y_{1:i},2}, \ldots, V_{y_{1:i},Nx})^{\mathrm{T}}$ composes the posterior variance for each grid point. As discussed in sections 1 and 2b(3), posterior particles are eventually rescaled to match the mean and variance in Eqs. (9) and (10) for each grid point. Similar moment-matching approach was proposed in Lei and Bickel (2011). Note that despite the sequential processing of observations,

per Eqs. (6)–(10), the LPF always uses the particle weights conditioned on the original prior particles [see Eq. (6)] to update the posterior mean and variance. In other words, the LPF posterior mean and variance do not rely on the order of assimilated observations or are not obtained sequentially. However, the LPF posteriors from a given observation are used as the priors for deriving the posterior particles given the next observation [see section 2b(3)]. The latter are then adjusted to match the posterior mean and variance obtained nonsequentially. We thus define the LPF as a partially sequential assimilation method.

3) RELAXATION OF POSTERIOR PARTICLES TO PRIORS AND ONLY MATCHING THE MEAN AND VARIANCE

To ensure the matching of posterior mean and variance calculated in Eqs. (9) and (10), the posterior particles given each new observation in the LPF are successively updated as follows. Let $\mathbf{x}_{n,y_{1:i-1}}^f$ be the nth prior particle after assimilating the first $i-1$ observations. The scalar weight of each particle $\mathbf{x}_{n,y_{1:i-1}}^f$ given the ith observation with Gaussian likelihood is $\tilde{w}_{n,y_i}$:

$$\tilde{w}_{n,yi} = \exp\left\{-\frac{\left(y_i - H_i\left[\mathbf{x}_{n,y_{1:i-1}}^f\right]\right)^2}{2\beta_i \sigma_{y_i}^2}\right\} \Bigg/ \sum_{n=1}^{N_e} \exp\left\{-\frac{\left(y_i - H_i\left[\mathbf{x}_{n,y_{1:i-1}}^f\right]\right)^2}{2\beta_i \sigma_{y_i}^2}\right\}. \quad (11)$$

Notice that $\tilde{w}_{n,y_i}$ is a scalar weight on which no localization is implemented, and it is used for the resampling procedure. Specifically, draw N_e particles from $(\mathbf{x}_{1,y_{1:i-1}}^f, \mathbf{x}_{2,y_{1:i-1}}^f, \ldots, \mathbf{x}_{N_e,y_{1:i-1}}^f)$ according to their respective scalar weights $\tilde{w}_{n,y_i}$, producing the resampled particles $(\mathbf{x}_{k_1,y_{1:i-1}}^f, \mathbf{x}_{k_2,y_{1:i-1}}^f, \ldots, \mathbf{x}_{k_{N_e},y_{1:i-1}}^f)$. The posterior particles given the first i observations are then computed by combining the resampled and prior particles as the following (Poterjoy 2016):

$$\mathbf{x}_{n,y_{1:i}}^a = \overline{\mathbf{x}_{y_{1:i}}^a} + \mathbf{r}_1 \circ (\mathbf{x}_{k_n,y_{1:i-1}}^f - \overline{\mathbf{x}_{y_{1:i}}^a}) + \mathbf{r}_2 \circ (\mathbf{x}_{n,y_{1:i-1}}^f - \overline{\mathbf{x}_{y_{1:i}}^a}), \quad (12)$$

where $\mathbf{r}_1 = (r_{1,1}, r_{1,2}, \ldots, r_{1,N_x})^{\mathrm{T}}$ and $\mathbf{r}_2 = (r_{2,1}, r_{2,2}, \ldots, r_{2,N_x})^{\mathrm{T}}$ are the weights of resampled and prior particles, respectively. The calculation of $\mathbf{r}_1$ and $\mathbf{r}_2$ can be referred to Poterjoy (2016). Near the observation location, $\mathbf{r}_1 = 1$ and $\mathbf{r}_2 = 0$, which means for those grid points that are near the observation the resampled particles are regarded as posterior particles. Therefore, the posterior particles near the observation approximately preserve all moments of the PF. For regions that are sufficiently distant from observations, $\mathbf{r}_1 = 0$, $\mathbf{r}_2 = 1$, the prior particles are used as the posterior particles. For regions in between (termed as gray zone), the resampled and the prior particles are blended and the weights $\mathbf{r}_1$ and $\mathbf{r}_2$ are deduced to ensure only the mean and variance of posterior particles are consistent with those obtained from the PF; i.e., following Eqs. (9) and (10).

In practical cycling DA experiments, despite the treatments above, the resampling procedure may still negatively impact the filter stability as it duplicates and removes particles, leading to insufficient diversity of posterior particles, especially with a small sample size. To mitigate the problem, Poterjoy et al. (2019) introduced an extra scalar mixing

parameter γ to further relax posterior particles to the prior particles (called γ relaxation):

$$\mathbf{x}_{n,y_{1:i}}^a = \overline{\mathbf{x}_{y_{1:i}}^a} + \gamma \mathbf{r}_1 \circ \left(\mathbf{x}_{k_n,y_{1:i-1}}^f - \overline{\mathbf{x}_{y_{1:i}}^a}\right) + [\gamma(\mathbf{r}_2 - 1) + 1] \circ \left(\mathbf{x}_{n,y_{1:i-1}}^f - \overline{\mathbf{x}_{y_{1:i}}^a}\right), \quad (13)$$

where γ is a tunable parameter between 0 and 1. The similar concept was first proposed in Zhang et al. (2004). The posterior particles obtained from Eq. (13) are then rescaled so that for each variable the mean and variance follow Eqs. (9) and (10). Notice that this further relaxation and rescaling allow the LPF to only maintain the mean and variance of the localized PF for all grids even at the regions near observations.

c. The local nonlinear ensemble transform filter (LNETF)

The algorithm of LNETF here follows Tödter and Ahrens (2015). This section again only highlights various strategies adopted by the LNETF to alleviate filter degeneracy.

1) INFLATION OF PRIOR PARTICLES

Different from the LPF, the LNETF in Tödter and Ahrens (2015) applies a prior inflation of particles that is widely used by EnKFs (e.g., Anderson and Anderson 1999) to deal with the collapse of particle weights and the underestimation of ensemble spread. The deviation of prior particles from their mean is inflated by a factor α as the following:

$$\mathbf{x}_n^f \leftarrow (\mathbf{x}_n^f - \overline{\mathbf{x}^f})\alpha + \overline{\mathbf{x}^f}, \quad (14)$$

where $\overline{\mathbf{x}^f}$ is the ensemble mean of prior particles.

2) LOCALIZATION OF OBSERVATIONAL IMPACT

The LNETF has two procedures to localize observational impact. One is similar to the LETKF that performs parallel assimilation for each grid point by incorporating only the observations within a local subdomain. The second is to modulate the particle weights so that for remote grids the deviations of the weights among different particles are reduced; see Eq. (15). Let assume that there are N_s ($N_s \leq N_y$) observations y_i ($i = 1, 2, \ldots, N_s$) within the localization radius r_s centered at grid j. Considering w_{n,y_i} as the normalized weight for the nth prior particle given observation y_i as in section 2b, the LNETF calculates the localized particle weights at grid j using the following function:

$$\omega_{n,y_i,j} = w_{n,y_i}^{l_{y_i,j}} \bigg/ \sum_{n=1}^{N_e} w_{n,y_i}^{l_{y_i,j}}, \tag{15}$$

where all terms have the same definitions as those in section 2b. A notable difference between Eqs. (7) and (15) is that the LNETF uses a different function to realize the localization. Specifically, the LNETF places the correlation coefficient $l_{y_i,j}$ as the power of the weight w_{n,y_i} (hereinafter the loc-P function) to smooth particle weights and to weaken observational impact for distant grid points. Such localization function holds for general observation likelihoods. The detailed comparison of the loc-D and loc-P functions will be given in section 4b. The accumulated weights for particles at grid j contributed by all N_s observations is then given as the following:

$$\omega_{n,j} = \prod_{i=1}^{N_s} \omega_{n,y_i,j}, \tag{16}$$

The localized weights $\omega_{n,j}$ for $j = 1, 2, \ldots, N_x$ are then normalized by the sum of all particles at each grid, i.e., $\omega_{n,j} \leftarrow \omega_{n,j} / \sum_{n=1}^{N_e} \omega_{n,j}$. The (local) posterior mean at grid j is computed by

$$\overline{x_j^a} = \overline{x_j^f} + \mathbf{x}_j^{f'T} \mathbf{\Omega}_j, \tag{17}$$

where $\mathbf{x}_j^{f'} = (x_{1,j}^f - \overline{x_j^f}, x_{2,j}^f - \overline{x_j^f}, \ldots, x_{N_e,j}^f - \overline{x_j^f})^T$ and $\mathbf{\Omega}_j = (\omega_{1,j}, \omega_{2,j}, \ldots, \omega_{N_e,j})^T$ contain the perturbations and weights of each prior particle at grid j, respectively. As shown in the next section the same localized particle weights are used to update particle perturbations.

3) DETERMINISTIC FILTER AND THE USE OF RANDOM ROTATION MATRIX FOR FILTER STABILIZATION

Different from the classical PF and the LPF, which attain posterior particles based on the resampling procedure, the LNETF adopts a deterministic square root filter. A transform matrix is derived to satisfy the first two moments, i.e., the mean and covariance, of posterior particles, consistent with those estimated by the PF. The prior particle perturbations are transformed into posterior perturbations as the following:

$$\mathbf{x}_j^{a'T} = \sqrt{n} \mathbf{x}_j^{f'T} \mathbf{T} \mathbf{\Lambda}, \tag{18}$$

$$\mathbf{T} = (\mathbf{W}_j - \mathbf{\Omega}_j \mathbf{\Omega}_j^T)^{1/2}, \tag{19}$$

where $\mathbf{x}_j^{a'} = (x_{1,j}^a - \overline{x_j^a}, x_{2,j}^a - \overline{x_j^a}, \ldots, x_{N_e,j}^a - \overline{x_j^a})^T$ is the perturbation vector of posterior particles $\mathbf{x}_j^a = (x_{1,j}^a, x_{2,j}^a, \ldots, x_{N_e,j}^a)^T$, and $\mathbf{W}_j = \mathrm{diag}(\mathbf{\Omega}_j)$. Here, $\mathbf{T}$ is the transform matrix that is derived from the particle weights considering the localized impact of observations, and $\mathbf{\Lambda}$ is an $N_e \times N_e$ mean-preserving random rotation matrix (Sakov and Oke 2008; Nerger et al. 2012), which has been demonstrated to be a very critical procedure to increase the diversity of particles and improve the filter stability. Note that, because of the use of local volumes, the LNETF only preserves the posterior covariance of the PF for state variables within each local volume. The covariance of state variables across local volumes is not guaranteed to be preserved.

Given the description of LPF and LNETF in sections 2b and 2c, respectively, it is noted that LPF and LNETF can preserve the posterior mean and variance of the PF. The LNETF additionally preserves the posterior correlation of state variables within the local volume. Moreover, it can be shown mathematically that given the same priors and localization, they produce the same posterior mean and variance, but may differ in other aspects of posteriors due to their different filter algorithms. This is further discussed in section 4d.

3. Models and experimental setups

a. Configurations of models and DA systems

In this study, we adopted the Lorenz (2005) model (denoted by Lorenz2005). Compared to the Lorenz96 model (Lorenz 1996), Lorenz2005 has higher spatial correlations better resembling the real atmosphere. It mimics the behavior of a meteorological variable along a constant latitude:

$$\dot{x}_j = -[x_{j-2K}][x_{j-K}] + ([x_{j-K}x_{j+K}]) - x_j + F, \tag{20}$$

where the square brackets denote an average of nearby grid points. We chose $K = 2$ {i.e., $[x_j] = (x_{j-1} + 2x_j + x_{j+1})/4$} and $F = 12$ following Rainwater and Hunt (2013). It has $N_x = 80$ grid points around a periodic domain; that is, the state vector is $\mathbf{x} = (x_1, x_2, \ldots, x_{80})^T$, where $x_0 = x_{80}, x_{-1} = x_{79}$. It is solved with a fourth-order Runge–Kutta scheme with a typical time step of $\Delta t = 0.05$ nondimensional units (corresponding roughly to 6 h).

We first run the model with $x_j = F$ ($j \neq 8$), $x_8 = 8.0001$ and integrate it for 500 days (i.e., 100 time units and 2000 time steps) to generate the state $\mathbf{x}^t(0)$ to initialize the nature (truth) run. From $\mathbf{x}^t(0)$ we integrate the model for another 500 days to generate a true trajectory for idealized DA experiments. Observations are then simulated by adding random noises to the true state at each grid (i.e., $N_y = 80$). The simulated observations are produced with a temporal frequency of 1 day, which means the DA interval is 24 h, and there are 500 consecutive analysis cycles. Two types of observations are tested: one has Gaussian distributed errors and the other has non-Gaussian double exponential distribution following Tödter and Ahrens (2015). The probability densities of observation errors are illustrated in Fig. 1. The observation errors have an expected value of zero and an Std of 1.0 (Fig. 1a) and 0.2 (Fig. 1b), respectively, representing two distinct observational accuracies.
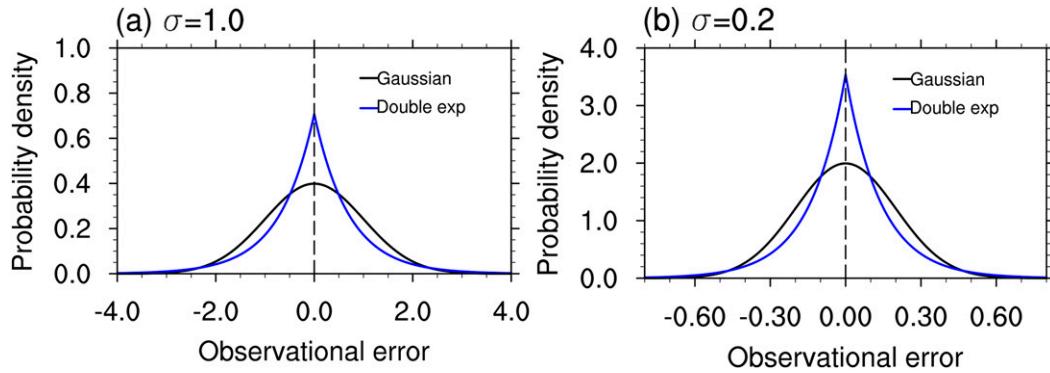
FIG. 1. Probability density of observational errors satisfying Gaussian (black) and double exponential (blue) distributions with the expected value zero and standard deviation (Std) of (a) $\sigma = 1.0$ and (b) $\sigma = 0.2$.

The comparison of skewness and kurtosis of probability distribution against those of a Gaussian distribution, that is, 0 and 3, respectively, signifies the level of non-Gaussianity (Morzfeld and Hodyss 2019). Our evaluation shows that the prior distribution in the current experimental setup has skewness (-0.02) that slightly deviates from zero and kurtosis (3.67) that is slightly larger than 3.0 and thus presents mild non-Gaussianity. The double exponential distribution of observations, despite the zero skewness, has kurtosis double the value of a Gaussian distribution (6 vs 3), characterized as medium non-Gaussianity.

The very first prior particles are produced by initializing 1-day ensemble forecasts from perturbed initial Gaussian distributed ensembles with Std 1. These initial conditions are centered at a state generated by adding uncorrelated Gaussian errors with mean 0 and Std 0.4 to the true state $\boldsymbol{x}^t(0)$. The performance of the LPF and LNETF are compared in terms of different particle numbers $N_e = 10, 20, 40$, and 80. For a fair comparison, the LPF and LNETF use the same true states and observations in each cycle and the common prior particles in the first analysis cycle. The prior particles in subsequent DA cycles are generated by integrating the same model with no model errors. The localization coefficient $l_{y_i,j}$ for the model grid j corresponding to the assimilation of observation y_i is determined by a Gaussian function $\exp[-d^2/(2r_d^2)]$, where d is the distance between grid point j and observation y_i, and r_d is the parameter that controls the range of observational impact. This Gaussian-like localization coefficient $l_{y_i,j}$ has no zero cutoff. For a fair comparison, the subdomain size for the LNETF assimilation is expanded to the globe (i.e., $r_s = N_x/2$ and $N_s = N_y$) to guarantee the same amount of observations as the LPF are considered for updating individual grid points. As shown in Eq. (13), the LPF adopts a relaxation factor. A value of $\gamma = 0.5$ is used since this value is overall optimal in our experiments

and also in Poterjoy et al. (2019). The configurations for the model and the parameters for experimental setups are summarized in Table 1.

b. Observation likelihood and inflation

The likelihood function for observations describes the plausibility of observations for certain true states (Berger and Wolpert 1984; i.e., each prior particle in this study). An accurate estimation of likelihood relies on a clear understanding of the PDF of observations. This study assumes that the PDF of observations is explicitly known, and reflected in the following two likelihood functions for Gaussian and non-Gaussian double exponential distributions:

$$p(y_i|\mathbf{x}_n) = \exp\left\{-\frac{(y_i - H_i[\mathbf{x}_n])^2}{2\sigma_{y_i}^2}\right\} \quad \text{and} \quad (21)$$

$$p(y_i|\mathbf{x}_n) = \exp\left(-\frac{|y_i - H_i[\mathbf{x}_n]|}{b_{y_i}}\right), \quad (22)$$

where the pair of vertical bars represents the absolute value. The error variances of the above two types of observations are $\sigma_{y_i}^2$ and $2b_{y_i}^2$, respectively. For a given particle, the likelihood $p(y_i|\mathbf{x}_n)$ then is normalized by its sum over all particles to approximate a particle weight.

Our tests show that the LNETF using β inflation [section 2b(1)] performs much better than with the standard prior inflation [section 2c(1)] in terms of posterior mean error for almost all particle numbers and observational error magnitudes (not shown). Consequently, the LNETF will adopt the β inflation as an alternative to the prior inflation in all the experiments of this study. Using the same inflation for the two filters allows for focusing our comparison primarily with respect to the localization and particle update strategies.

TABLE 1. Parameters of the model configuration and experimental setup for the Lorenz2005 model.

	State dimension N_x	Time step	Forecast length	Obs No. N_y	Std of obs error	Case No.
Lorenz2005	80	0.05	0.2	80	1.0; 0.2	500

TABLE 2. Descriptions of the algorithms involved in designed experiments.

Name	Localization	Relaxation to priors	Posterior inflation
LPF_loc-D	Loc-D	Yes, $\gamma = 0.5$	No
LPF_loc-P	Loc-P	Yes, $\gamma = 0.5$	No
LNETF_loc-P	Loc-P	No, $\gamma = 1$	No
LNETF_loc-D	Loc-D	No, $\gamma = 1$	No
LPF_loc-D_NoRlx	Loc-D	No, $\gamma = 1$	No
LNETF_loc-D_Rlx	Loc-D	Yes, $\gamma = 0.5$	No
LPF_loc-D_PoInfl	Loc-D	Yes, $\gamma = 0.5$	Yes
LNETF_loc-D_PoInfl	Loc-D	No	Yes
ETKF	Observation space localization	No	Yes

c. Experimental design

The LPF and LNETF present two major differences, one is the localization function and the other is the filter scheme. The LPF and LNETF adopt distinct localization functions [cf. Eqs. (7) and (15)] in their original forms that are the loc-D and loc-P, respectively. They have similar concept to localize observational impact, i.e., through homogenizing particle weights for remote grids, but differ in forms. Therefore, the localization functions can be used alternatively for individual filters. The differences of the LPF and LNETF schemes mainly lie in the different assimilation framework, i.e., the partially sequential and the simultaneous methods, respectively, and the different particle update algorithms, i.e., the resampling based and the deterministic transformation algorithms, respectively. To examine these differences between the LPF and LNETF and explore possible strategies to improve their performance, a set of experiments are designed (see the summary in Table 2).

- *Experiment 1.* These simulations compare the performance of the original LPF and LNETF, named LPF_loc-D and LNETF_loc-P, respectively.
- *Experiment 2.* The objective of this experiment is to apply both localization functions to the LPF and LNETF to examine the impact of loc-D and loc-P on filter performance. The experiments LPF_loc-D, LPF_loc-P, LNETF_loc-D, and LNETF_loc-P are conducted in both the cycling and noncycling modes. While applying the loc-D to LNETF or the loc-P to LPF, the normalized weights w_{n,y_i} of the nth prior particle conditioned on observation y_i are calculated in the same way but are homogenized for distant grids in different manners, using Eq. (7) for loc-D and Eq. (15) for loc-P.
- *Experiment 3.* To understand the role of the relaxation to prior particles used in the LPF (i.e., parameter γ), the performance of LPF_loc-D ($\gamma = 0.5$) is compared to that without using the relaxation (i.e., $\gamma = 1$; "LPF_loc-D_NoRlx"). Additionally, experiments that apply the similar relaxation [Eq. (13)] to the LNETF ("LNETF_loc-D_Rlx") are also performed. The variance of posterior particles after relaxation is then tuned to the same as that before the relaxation following the LPF [see section 2b(3)].
- *Experiment 4.* The impact of filter schemes on posterior moments is investigated by comparing the LPF and LNETF in noncycling experiments and by using a common localization function (i.e., LPF_loc-D and LNETF_loc-D).

- *Experiment 5.* The inflation of posterior particles is also applied to LPF_loc-D (denoted by LPF_loc-D_PoInfl) and LNETF_loc-D (denoted by LNETF_loc-D_PoInfl) to see if the filter stability and accuracy can be further improved. The posterior inflation step is the same as Eq. (14) but $\mathbf{x}_n^f$ and $\overline{\mathbf{x}^f}$ are replaced by $\mathbf{x}_n^a$ and $\overline{\mathbf{x}^a}$, respectively.
- *Experiment 6.* A Gaussian filter experiment is performed in this study as a benchmark to manifest the benefits of LPF and LNETF. The ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003; Wang et al. 2004) is adopted. The ETKF uses the observation space localization (R-localization) and posterior inflation strategies to stabilize the filter. The localization and inflation parameters of the ETKF are tuned for each σ and N_e to optimize the sample mean analysis RMSE for a fair comparison to those of other filters. The details of the algorithm and implementation to Lorenz2005 can be referred to Huang et al. (2019; i.e., their "R-D" method). Note that ETKF simply serves as a baseline and its detailed comparison with LPF and LNETF is beyond the scope of the study.

4. Results

In this section, the performance of the LPF and LNETF in assimilating both the Gaussian and non-Gaussian observations are evaluated, compared, and analyzed. The performances related to the filter stability and accuracy are assessed. A filter is considered stable if the sample mean root-mean-square error (RMSE) of analyses within a long experimental period is smaller than the Std σ of observational errors. In addition, the filter performance is evaluated by the range of parameters that can lead to a stable filter. A wider range of stable parameters indicates the filter is less sensitive to the parameter selection. The filter accuracy evaluates both the RMSE of posterior mean and higher moments of posterior particles. The results with Gaussian and non-Gaussian observations offer qualitatively similar conclusions, indicating the relative performance of LPF and LNETF could be insensitive to the observation likelihood functions considered. Therefore, only the results with non-Gaussian observations were shown throughout this study.

a. Comparing the LPF and LNETF in their original forms

We first compare the performance of the LPF and LNETF in their generic forms shown in sections 2b and 2c (i.e., LPF_loc-D versus LNETF_loc-P). Figure 2 shows the sample-mean RMSE
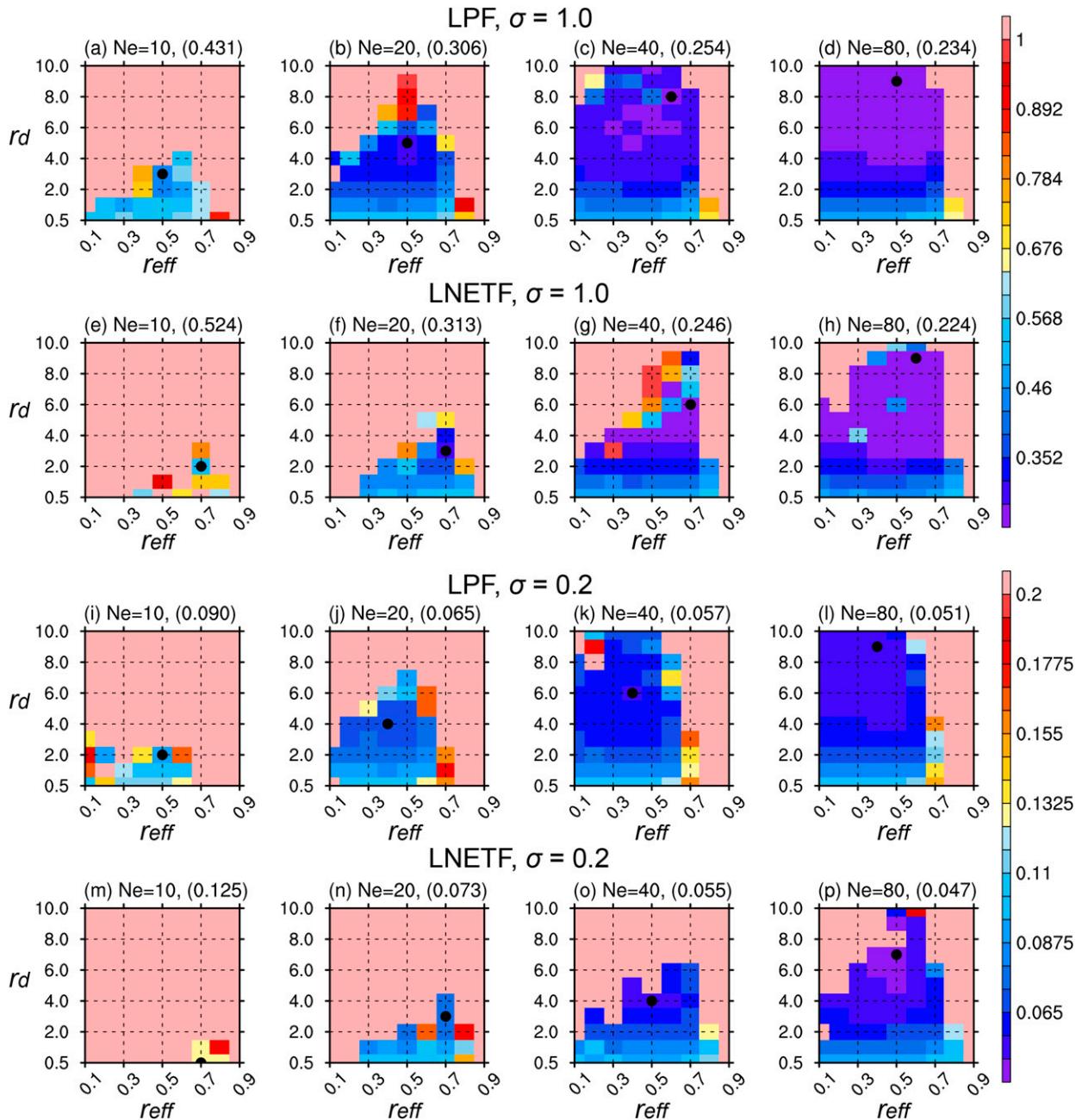
FIG. 2. Sample mean RMSE of analyses against the localization (r_d) and inflation (r_{eff}) parameters for the (a)–(d) LPF_loc-D and (e)–(h) LNETF_loc-P with $\sigma = 1.0$ in the cycling assimilation based on the Lorenz2005 model. (i)–(p) As in (a)–(h), but with $\sigma = 0.2$. The values in parentheses denote the minimum RMSE. The pink squares indicate the RMSE exceeds the Std of observational errors, which is considered to be an unstable filter (i.e., violating the stable filter definition in this study).

of analyses in cycling DA experiments for the LPF_loc-D and LNETF_loc-P as a function of the localization (r_d) and inflation (r_{eff}) parameters in the Lorenz2005 model. Note that in RMSE plots like Fig. 2, pink square indicates the filter is unstable (i.e., violating the stable filter definition in this study). For both the LPF_loc-D and LNETF_loc-P, the filter overall exhibits a narrower range of stable parameters when the ensemble size and/or the observation error is small. The former is

due to the larger sampling errors, and the latter is attributed to the increased risk of weight collapse with a narrower likelihood function. Additionally, a smaller ensemble size requires stronger localization (i.e., a shorter localization range) to maintain a stable filter, which is consistent with previous studies (Houtekamer and Mitchell 2001; Lorenc 2003). Noticeably, the LNETF_loc-P shows a much narrower range of stable parameters than the LPF_loc-D given the same particle number and observational

TABLE 3. The optimal localization (r_d) and inflation (r_{eff}) parameters and the corresponding sample-mean RMSE for the LPF_loc-D and LNETF_loc-P in Fig. 2. The results of the LNETF_loc-D in Fig. 4 and the ETKF are also listed for comparison. The same superscript number indicates that two RMSE values have statistically significant differences at a 0.05 level.

			$N_e = 10$	$N_e = 20$	$N_e = 40$	$N_e = 80$
$\sigma = 1.0$	r_d	LPF_loc-D	3.0	5.0	8.0	9.0
		LNETF_loc-P	2.0	3.0	6.0	9.0
		LNETF_loc-D	1.0	3.0	6.0	8.0
	r_{eff}	LPF_loc-D	0.5	0.5	0.6	0.5
		LNETF_loc-P	0.7	0.7	0.7	0.6
		LNETF_loc-D	0.6	0.7	0.7	0.6
	RMSE	LPF_loc-D	0.431[1]	0.306	0.254	0.234
		LNETF_loc-P	0.524[1,2]	0.313	0.246	0.224
		LNETF_loc-D	0.480[2]	0.311	0.239	0.222
		ETKF	0.322	0.315	0.292	0.290
$\sigma = 0.2$	r_d	LPF_loc-D	2.0	4.0	6.0	9.0
		LNETF_loc-P	0.5	3.0	4.0	7.0
		LNETF_loc-D	2.0	3.0	5.0	8.0
	r_{eff}	LPF_loc-D	0.5	0.4	0.4	0.4
		LNETF_loc-P	0.7	0.7	0.5	0.5
		LNETF_loc-D	0.5	0.5	0.4	0.4
	RMSE	LPF_loc-D	0.090[3]	0.065	0.057	0.051
		LNETF_loc-P	0.125[3,4]	0.073	0.055	0.047
		LNETF_loc-D	0.099[4]	0.069	0.053	0.047
		ETKF	0.067	0.071	0.066	0.064

error, especially for a small number of particles, such as N_e = 10 and 20.

The optimal parameters and minimum average RMSE for the LPF_loc-D and LNETF_loc-P in Fig. 2 are summarized in Table 3. Generally, the use of more particles requires less restrictive localization and less observational error inflation, due to the more accurate sampling of prior and posterior probability. Nevertheless, the optimal parameters r_d and r_{eff} for the LPF_loc-D are overall wider and smaller than those for the LNETF_loc-P, respectively, regardless of the observational accuracy. The former is because the loc-D function imposes stronger localization effect than the loc-P when particle weights deviate largely from each other (see discussion in section 4b). The latter indicates a larger inflation factor is needed for the LNETF_loc-P, probably due to its overall smaller range of posterior particles (i.e., the difference between the maximum and minimum posterior particles) as detailed in section 4d. In addition to the wider range of stable parameters (Fig. 2), the minimum average RMSE of the LPF_loc-D is also statistically significantly (at a 0.05 level) smaller than that of the LNETF_loc-P for small particle number N_e = 10 with both σ = 1.0 and 0.2. In contrast, for N_e = 20, 40, and 80, the minimum RMSEs are statistically indistinguishable between the two methods, even using a much longer period (e.g., 1000 DA cycles, not shown).

The LPF and LNETF are compared with the ETKF. Both nonlinear filters deliver smaller mean RMSE than the ETKF with an ensemble size 20 or more, indicating the benefit of considering the non-Gaussianity of prior particles and observation errors. However, the LPF and LNETF are inferior to the ETKF with N_e = 10, probably because of their insufficient sampling of posterior probability with a very small number of particles. The RMSE of ETKF apparently has less dependence on the ensemble size than that of LPF and LNETF. These results are consistent with earlier studies comparing linear and nonlinear ensemble filters (Lei and Bickel 2011; Tödter and Ahrens 2015; Shen et al. 2017).

b. Difference of the loc-D and loc-P functions

As discussed in sections 2b(2) and 2c(2), the original LPF and LNETF adopt different functions to realize localization, which may contribute to their performance difference shown in section 4a. To isolate the impact of localization function, this section implements the loc-D to LNETF. To better interpret the property of the two localization functions, we provide a simplified example with only two particles. Figure 3 shows the variation of two particle weights as a function of the standardized distance (i.e., $r_d = 1$) between the grids and the observation as shown in Eqs. (7) and (15). The two localization functions have almost the same variation of particle weights when the original weights are close ($w_1 = 0.6$ and $w_2 = 0.4$ in Fig. 3a). In contrast, for two significantly distinct original weights ($w_1 = 0.99$ and $w_2 = 0.01$ in Fig. 3b), their variation shows more apparent difference. Specifically, the loc-D smooths particle weights more and imposes stronger localization than the loc-P with the same distance parameter. This property of the loc-D function allows larger weight deviations (e.g., caused by a limited ensemble size) to be more effectively diminished, and thus reduces the risk of weight collapse and filter degeneracy. This difference between the loc-D and loc-P is consistent with the longer optimal localization parameter r_d for the LPF_loc-D (Table 3). Although the variation of the single-observation weights with distance could be more harmonized between the loc-D and loc-P by properly modifying the form of the localization function, a clear comparison of the two original functions in theory and practice is a necessary step to provide insight into an effective construction of localization functions.
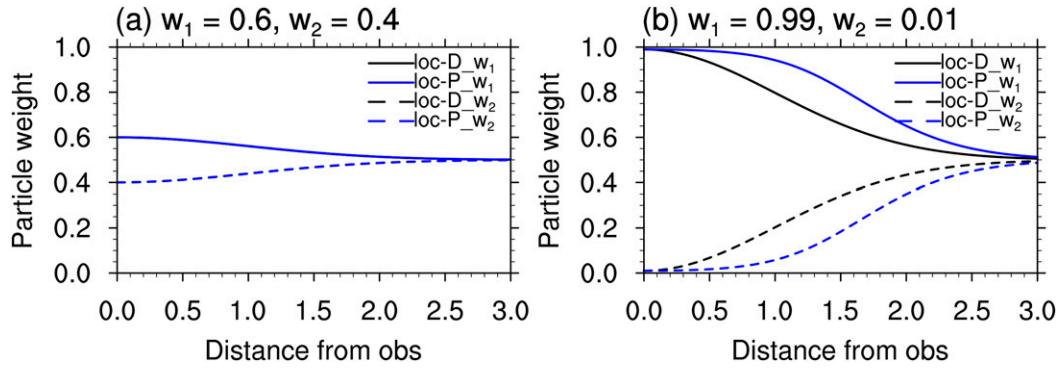
FIG. 3. Evolution of two particle weights as a function of the standardized distance from a given observation for the loc-D (black) and loc-P (blue) schemes. The weights at the observation position are (a) 0.6 and 0.4, and (b) 0.99 and 0.01, respectively.

The analysis in Fig. 3 indicates the potential advantage of the loc-D in stabilizing filters in cycling data assimilation. To further examine the finding, additional cycling experiments are performed with the LNETF_loc-D approach and compared to those with the LNETF_loc-P. Figure 4 shows the average RMSE of the LNETF_loc-D as a function of the localization and inflation parameters. Replacing the loc-P with the loc-D in LNETF slightly widens the range of stable parameters as expected (cf. Figs. 4, 2). The minimum average RMSE of LNETF

is also improved for particle numbers $N_e = 10, 20, 40$, and 80 when the loc-D is used (cf. LNETF_loc-D and LNETF_loc-P in Table 3). Particularly, the reduced RMSE for $N_e = 10$ with nearly 8% and 20% for $\sigma = 1.0$ and 0.2, respectively (see Table 3), are statistically significant at the 0.05 level. The optimal localization and inflation parameters r_d and r_{eff} of the LNETF_loc-D also become closer to those of the LPF_loc-D for $\sigma = 0.2$ due to using the same localization function (Table 3). This finding is, however, not evident for $\sigma = 1.0$.
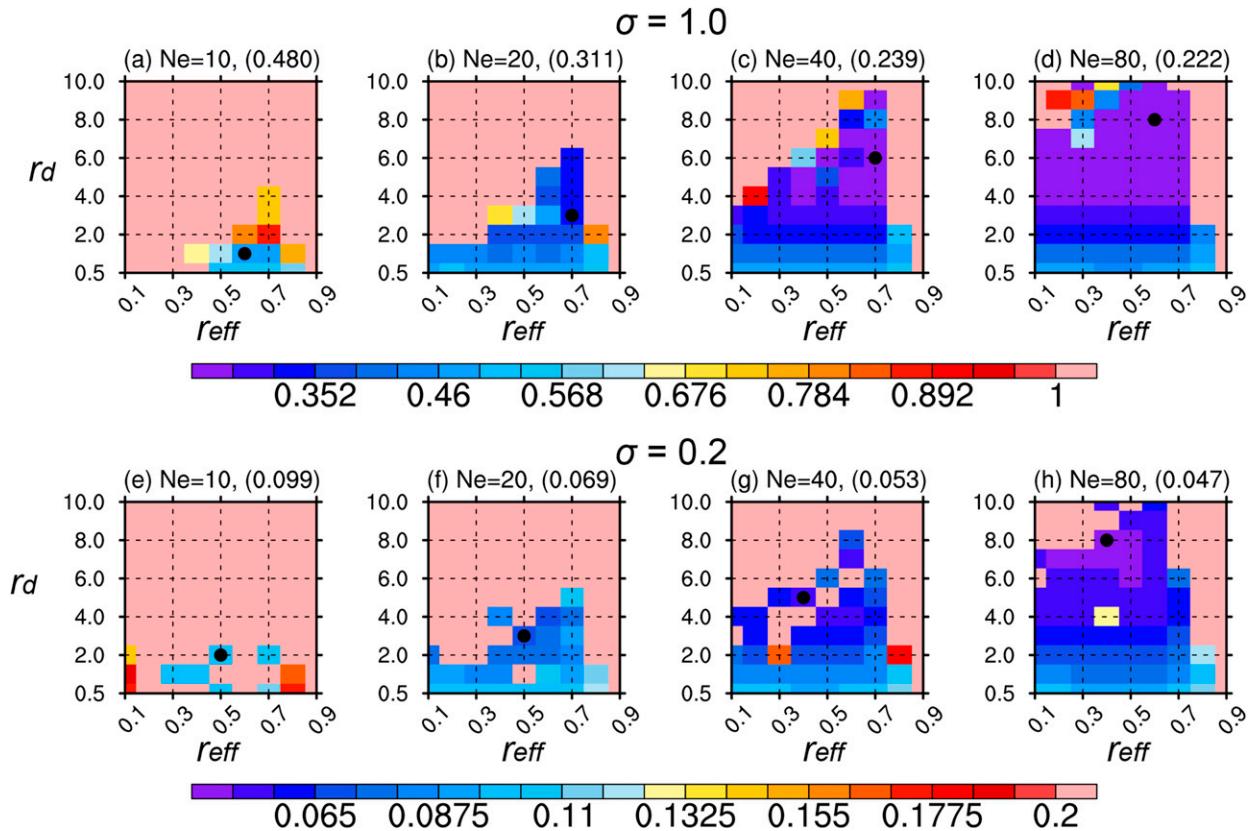

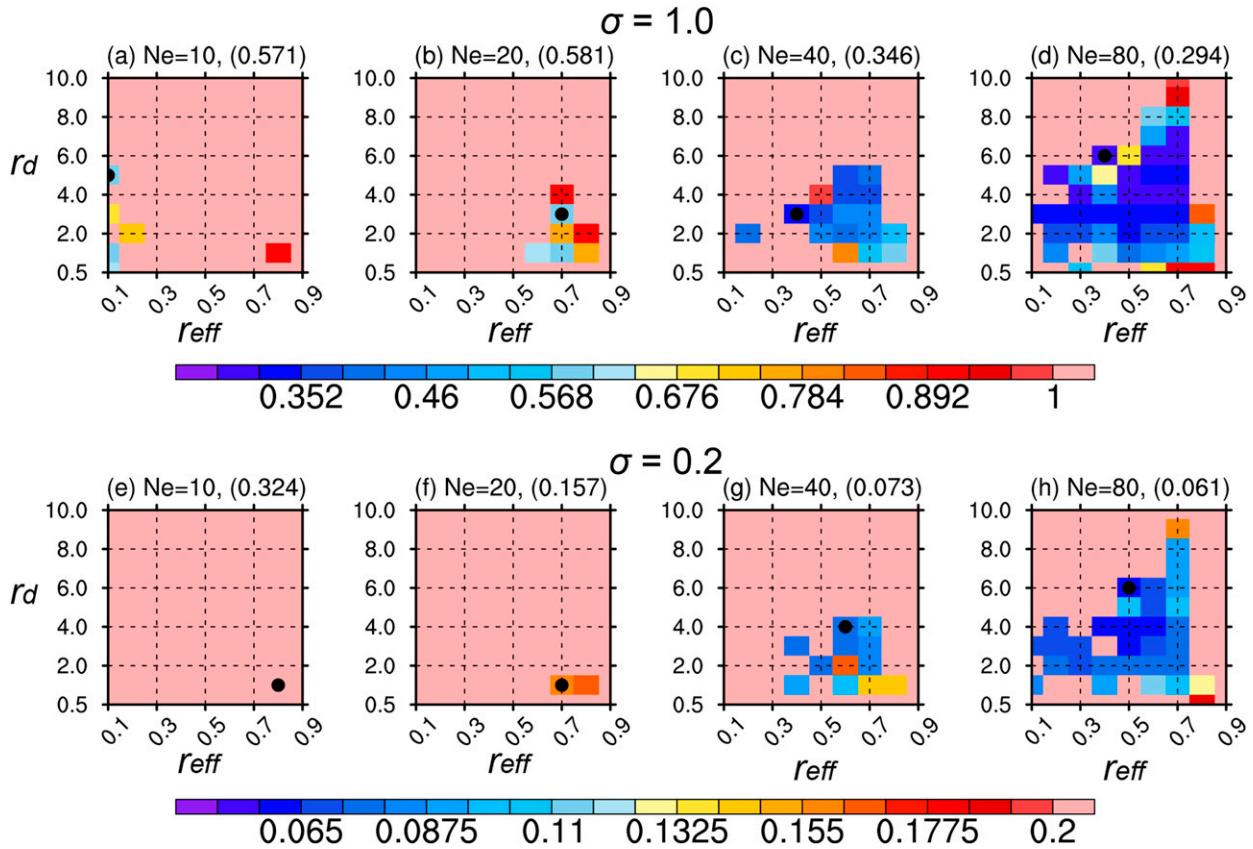FIG. 4. As in Fig. 2, but for the algorithm LNETF_loc-D.

FIG. 5. As in Fig. 2, but for the algorithm LPF_loc-D_NoRlx.

Note that in spite of the somewhat wider range of stable parameters for the LNETF_loc-D by using an alternative localization function, it still performs much worse than the LPF_loc-D, which should be attributed to their distinct update schemes of particles. Section 4e proposes and demonstrates additional means to further improve the LNETF to reach similar performance as the LPF.

To further diagnose the difference of the analysis accuracy between the LNETF_loc-D and LNETF_loc-P with optimal inflation and localization parameters, noncycling experiments are carried out. The two filters use the same prior particles in each DA cycle, which are from those of the cycled LPF_loc-D experiment using the optimal parameters (i.e., Figs. 2a–d,i–l). This experimental setup can clearly distinguish the impact of an individual localization function on the posterior mean and distribution. The results show that the mean analysis RMSE with $N_e = 10$ becomes much closer (0.46 vs 0.45 for $\sigma = 1.0$ and 0.09 vs 0.098 for $\sigma = 0.2$) between the LNETF_loc-D and LNETF_loc-P in the noncycling experiments given their respective optimal parameters in the cycling experiments (see Table 3). In other words, the difference between LNETF_loc-D and LNETF_loc-P is much reduced going from a cycling experiment to a noncycling experiment. Given the posterior difference gets accumulated in a cycling experiment, but not a noncycling experiment, we therefore deduce that the significantly smaller minimum RMSE of

LNETF_loc-D than LNETF_loc-P with $N_e = 10$ in the cycling experiment (Table 3) may be mainly attributed to the improved sampling of posterior distribution by using the localization function loc-D.

c. The role of relaxation to priors

One of the major differences of the LPF (Poterjoy et al. 2019) and LNETF (Tödter and Ahrens 2015) lie in the particle update algorithm. The former uses the relaxation of resampled posterior particles to the priors with γ coefficient [Eq. (13)]; called the γ relaxation. The LNETF on the other hand does not require the relaxation procedure. As discussed earlier, the diversity of particles in the classical PF and the LPF would be constrained by the resampling step, which duplicates particles with large weights and removes those with negligible weights. The γ relaxation aims to enhance the particle diversity through mixing posterior and prior particles in the LPF. It was demonstrated in Poterjoy et al. (2019) that the γ relaxation is particularly useful to stabilize the filter with small particle numbers, but is unnecessary for a sufficient number (i.e., $\gamma = 1$ for large particle numbers). In contrast, the LNETF updates particles with a deterministic filter instead of the resampling and does not use the relaxation step.

To clarify the impact of the relaxation procedure on filter performance, two experiments are implemented. One is to remove the γ relaxation from the LPF and compare the

LPF_loc-D_NoRlx and LPF_loc-D. Figure 5 illustrates the average RMSE of the LPF_loc-D_NoRlx against varying parameters in the Lorenz2005 model. The stable-parameter range of the LPF is significantly shrunk without using the γ relaxation for all particle numbers and observational accuracies (cf. Figs. 5, 2), especially for a small observational error ($\sigma = 0.2$) and ensemble size (e.g., $N_e = 10$ and 20). Even for a large number (e.g., $N_e = 80$), almost half of the parameters with stable filter originally in the LPF_loc-D (see Figs. 2d,l) show the filter divergence instead (see pink squares). Likewise, the minimum average RMSE of the LPF_loc-D_NoRlx for each ensemble size and observational accuracy is much worse than that of the LPF_loc-D and the LNETF_loc-D. The above results indicate that the γ relaxation is a critical procedure for the LPF to stabilize the filter, not only for a small ensemble size but also for relatively large ones.

The second experiment examines the impact of applying the particle mixing step to the LNETF by relaxing posterior particles in LNETF_loc-D to prior ones with $\gamma = 0.5$ (i.e., LNETF_loc-D_Rlx) and maintain the original posterior mean and variance similarly as in the LPF_loc-D. It is found that the γ relaxation has only very small impact on the LNETF in terms of the average RMSE (not shown). The much less reliance of the LNETF on the relaxation approach is likely because the LNETF adopts a deterministic update without resampling and uses the random matrix in the transformation [see section 2c(3)]. Both to some extent enhance the particle diversity.

d. Difference in the particle update schemes of LPF and LNETF

The LPF and LNETF have different filter algorithms, leading to the differences in posterior moments, which can dramatically impact their filter performance. The LPF relies on the resampling algorithm during the sequential processing of observations, while the LNETF applies a deterministic transformation to derive posterior particles based on the simultaneous assimilation. To further isolate the impact of filter schemes, we compare the LPF_loc-D and LNETF_loc-D assimilation in noncycling experiments with the same prior particles as in section 4b.

Our results show that the LPF_loc-D and LNETF_loc-D in noncycling experiments give equal posterior mean and variance for each variable at each grid point when using the same localization and β inflation parameters and the same prior particles in each cycle. This result is consistent with those discussed in section 2c, indicating that the implementation of the two nonlinear filters has similar effects on posterior mean and variance. Specifically, for both LPF and LNETF, the weights used to obtain posterior mean and variance are calculated from the original prior members and then modulated by the localization function to derive the localized weight vectors. This means that the accumulative localized particle weights given all observations are identical for the two filters once the same localization function and prior particles are used. The sequential or simultaneous procedure does not affect the calculation of posterior mean and variance. The different filter schemes of LPF and LNETF only lead to the difference of posterior property beyond the mean and variance.

This section compares the probability distribution of posterior particles for the LPF_loc-D and LNETF_loc-D with their respective optimal parameters in noncycling experiments. Given the same mean and variance, such a difference comes from their distinct particle update algorithms in approximating moments beyond the mean and variance.

Figure 6 illustrates the posterior probability distribution averaged over all grids and cases for the LPF_loc-D (blue) and LNETF_loc-D (green) as well as their differences. The posterior particles at each grid in each cycle are normalized before calculating the PDF. Such calculation provides meaningful statistics because the effects of different posterior mean and variance among different grids on PDF are eliminated by the normalization and grid points of the Lorenz2005 model present statistically identical dynamical properties. It shows that the LPF_loc-D and LNETF_loc-D have basically the similar symmetric structure of probability distribution around zero. However, the differences in the shapes of their posterior probability distribution can be observed. Such differences are qualitatively preserved for all particle numbers and observational errors tested. The LPF_loc-D has about 0.8% lower and 1.2% higher probability near the 1-Std value and the mean value, respectively. In other words, the LPF_loc-D tends to have a sharp peak and flat tails around 1 Std from the mean. Notably, beyond 2 Stds, the LPF_loc-D has slightly (up to 0.3%) higher probability than the LNETF_loc-D. This difference is seemingly small but may potentially influence the filter stability and accuracy in the successive DA cycles (see more discussions on Table 4).

We used three metrics to quantify the differences of posterior probability distribution between the LPF_loc-D and LNETF_loc-D (see Table 4). One is the percentage of cases where the range of posterior particles from one filter is larger than the other ("L_range"). Note the sum of L_range values for the LPF_loc-D and LNETF_loc-D should be one. It is found that the posterior particles of the LPF_loc-D apparently have a larger percentage (more than double) to span a wider range than those of the LNETF_loc-D albeit having the same variance. Further diagnostics suggest that the larger range of posterior particles by the LPF_loc-D is largely attributed to the γ relaxation. The second metric is relevant to the L_range but measures the percentage of the true state out of the range of posterior ensemble ("Tru_out") for all cases and grids. It is found that the true states fall outside of posterior particles more frequently for the LNETF_loc-D. Although their difference of Tru_out only ranges from 1.5% to 3%, the higher frequency of the true state falling outside of the ensemble is more likely to cause larger sampling error and weight collapse. The third metric is the continuous ranked probability score (CRPS) that measures the deviation between the forecasted and observed probability for ranked ensemble states (e.g., Hersbach 2000). A lower CRPS score means a more accurate probability distribution. The LNETF_loc-D produces a slightly better CRPS than the LPF_loc-D for all particle numbers and observational errors. The worse probability distribution of the LPF_loc-D measured by CRPS is most likely due to the relaxation of prior particles, which loses certain observational information contained in posterior particles.
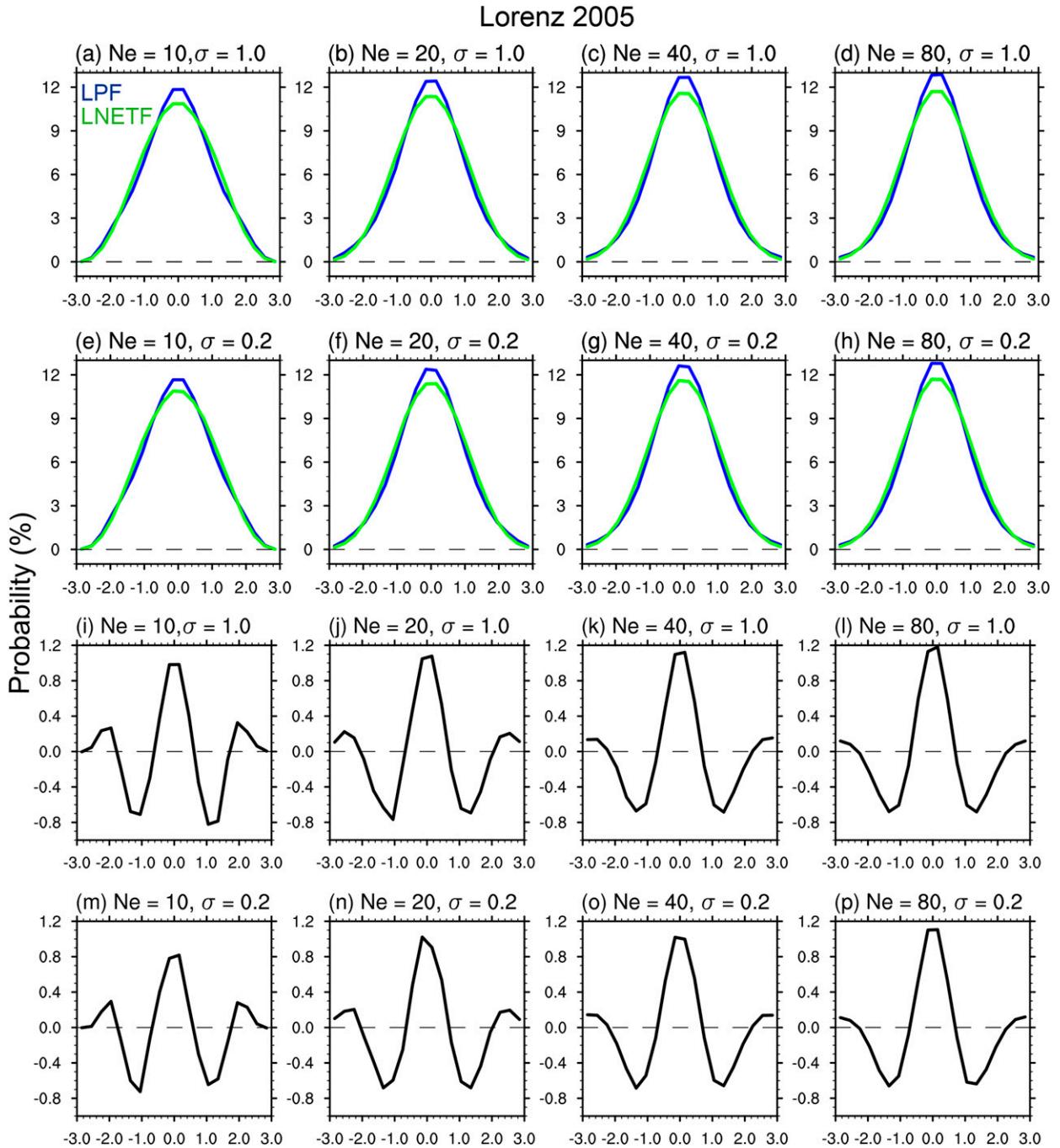
FIG. 6. Averaged probability distribution of posterior particles over all grids and cases in noncycling experiments for the LPF_loc-D (blue) and LNETF_loc-D (green) with (a)–(d) $\sigma = 1.0$ and (e)–(h) $\sigma = 0.2$ in the Lorenz2005 model. The posterior particles at each grid in each cycle are normalized before calculating the probability distribution. (i)–(p) The probability difference between the LPF_loc-D and LNETF_loc-D.

e. Improvement of the LNETF

The diagnostic results in noncycling experiments in section 4d imply that, despite their same posterior mean and variance, the LNETF_loc-D overall gives a narrower range of posterior particles relative to the LPF_loc-D, increasing the possibility of the true state falling outside of the posterior ensemble. Such a

difference likely explains the performance difference of the two filters seen in the cycling experiments (see Table 3 and cf. Figs. 2 and 4). To test this hypothesis, we apply the posterior inflation (Whitaker and Hamill 2002) to the LNETF_loc-D (i.e., the LNETF_loc-D_PoInfl) to explore if the filter stability can be improved.

TABLE 4. Evaluation of posterior probability distribution for the LPF_loc-D and LNETF_loc-D (in parentheses) with $\sigma = 1.0$ and 0.2. L_range is the percentage of a larger range of posterior particles. Tru_out is the percentage of the true state out of the range of posterior ensemble. CRPS is the continuous ranked probability score. The three metrics are statistics for all grids and cases.

		$N_e = 10$	$N_e = 20$	$N_e = 40$	$N_e = 80$
$\sigma = 1.0$ LPF (LNETF)	L_range (%)	66.4 (33.6)	65.9 (34.1)	70.9 (29.1)	75.4 (24.6)
	Tru_out (%)	18.8 (21.8)	8.2 (10.0)	3.8 (5.1)	1.2 (2.0)
	CRPS	0.192 (0.191)	0.176 (0.172)	0.170 (0.165)	0.164 (0.159)
$\sigma = 0.2$ LPF (LNETF)	L_range (%)	66.7 (33.3)	64.8 (35.2)	68.0 (32.0)	72.4 (27.6)
	Tru_out (%)	14.4 (17.6)	7.4 (9.0)	3.4 (4.5)	1.6 (2.4)
	CRPS	0.181 (0.180)	0.173 (0.170)	0.168 (0.164)	0.165 (0.161)

The first step is to determine the optimal inflation factor α of posterior particles. It is selected by tuning the localization parameter r_d and the posterior inflation factor α for optimal performance (in terms of RMSE) given $r_{\mathrm{eff}} = 0.2$. $\alpha \in \{1.00, 1.01, \ldots, 1.15\}$ as in Tödter and Ahrens (2015). We specify $r_{\mathrm{eff}} = 0.2$, since the LNETF is especially unstable for small values of r_{eff} (see Figs. 2, 4). For a fair comparison, a range of inflation factor α is tested for the LPF. The optimal α for the LPF_loc-D_PoInfl and LNETF_loc-D_PoInfl in the Lorenz2005 model is listed in Table 5. Overall, α tends to be smaller with more particles and becomes close to zero with $N_e = 80$. The optimal α for the LNETF_loc-D_PoInfl is larger than that for the LPF_loc-D_PoInfl as expected, especially for $N_e = 10$ and 20.

We specified the optimal posterior inflation factor α for each ensemble size and observational accuracy. Since the additional posterior inflation does not reduce the minimum RMSE of posterior mean of the LPF_loc-D and LNETF_loc-D with $N_e = 40$ and 80, it will only be used in the experiments with $N_e = 10$ and 20. The average RMSE as a function of the localization and inflation parameters r_d and r_{eff} in cycling experiments is shown in Fig. 7. Relative to Fig. 4, the stable-parameter range of the LNETF_loc-D_PoInfl apparently becomes wider for small particle numbers $N_e = 10$ and 20, reaching similar performance as the LPF_loc-D_PoInfl. The additional inflation of posterior particles also apparently reduces the minimum RMSE for the LNETF_loc-D with $N_e = 10$. In contrast, the additional posterior inflation has no impact on the analysis RMSE of the LPF (cf. Figs. 2, 7).

For a more rigorous evaluation, we extended the cycling DA experiments for the LPF_loc-D_PoInfl and LNETF_loc-D_PoInfl with their respective optimal parameters to 2500 cycles (i.e., 10 000 model steps). Figure 8 summarizes the 2500-cycle mean analysis RMSE, spread, and CRPS of posterior particles as a function of particle number in the cycling experiments. The LPF_loc-D_PoInfl and LNETF_loc-D_PoInfl after

optimization using posterior inflation have very similar RMSE, which are statistically indistinguishable at a 0.05 level for all particle numbers and observational errors. The ensemble spreads of the two methods both overestimate the analysis error amplitude. However, the LNETF_loc-D_PoInfl has 10%–40% larger ensemble spread than the LPF_loc-D_PoInfl for $N_e = 10$ and 20, likely due to the larger analysis inflation factor α used in the former (see Table 5). In terms of the CRPS metric, the LNETF_loc-D_PoInfl performs better than the LPF_loc-D_PoInfl. It is similar to those of the noncycling experiments (see Table 4), except the superiority of LNETF_loc-D_PoInfl is more significant in the cycling experiments.

In addition to evaluating the accuracy of the posteriors in a univariate sense using mean RMSE, spread, and CRPS, the accuracy of the spatial correlation of posterior particles estimated by the LPF and LNETF are compared. We compute the respective posterior spatial correlation of the two methods with a large ensemble size $N_e = 400$ and use their mean as the referencing truth. Their respective optimal localization (i.e., loc-D) and inflation (i.e., β inflation) parameters are tested and adopted. It is assumed that $N_e = 400$ is sufficient for the LPF_loc-D and LNETF_loc-D to provide a reliable estimation of posterior correlation. Figure 9 gives an example of the correlation matrix of posterior particles from grid 1 to 21 for the LPF_loc-D and LNETF_loc-D with $N_e = 400$ in the Lorenz2005 model. It can be seen that the structure of spatial correlation for the two methods given the same observation error is very similar. The average absolute deviations of the spatial correlation of the LPF_loc-D and LNETF_loc-D from their mean value range from 0.033 to 0.040 for grid intervals from 1 to 3. Such average deviation will be used to decide if the errors of estimated spatial correlation with a much smaller N_e have statistically significant differences between the two filters.

Figure 10 shows the 2500-sample and 80-grid mean absolute errors of posterior spatial correlation between grids with intervals 1, 2, and 3 for the LPF_loc-D_PoInfl and LNETF_loc-D_PoInfl

TABLE 5. The optimal posterior inflation factor α for the LPF_loc-D_PoInfl and LNETF_loc-D_PoInfl in the Lorenz2005 model.

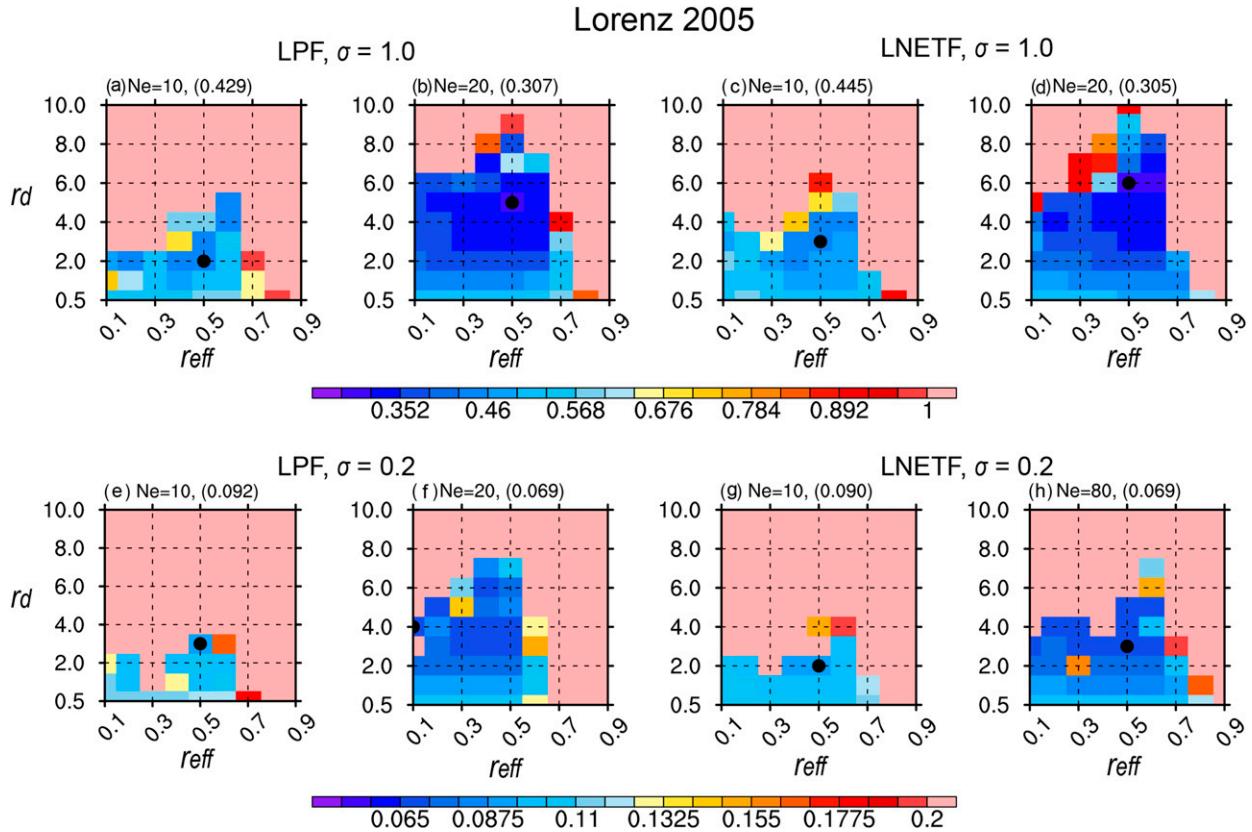|		$N_e = 10$	$N_e = 20$	$N_e = 40$	$N_e = 80$
LNETF_loc-D_PoInfl	$\sigma = 1.0$	1.13	1.11	1.06	1.01
	$\sigma = 0.2$	1.10	1.06	1.05	1.01
LPF_loc-D_PoInfl	$\sigma = 1.0$	1.05	1.03	1.03	1.03
	$\sigma = 0.2$	1.02	1.04	1.02	1.01

FIG. 7. Sample mean RMSE of analyses against the localization (r_d) and inflation (r_{eff}) parameters for the (a),(b) LPF_loc-D_PoInfl and (c),(d) LNETF_loc-D_PoInfl with $\sigma = 1.0$ in the Lorenz2005 model. (e)–(h) As in (a)–(d), but for $\sigma = 0.2$.

in the Lorenz2005 model. It is found that the LPF_loc-D_PoInfl presents more accurate posterior spatial correlation than the LNETF_loc-D_PoInfl, but only statistically significant (i.e., the error difference larger than the uncertainty bar) for $N_e = 80$. As discussed in section 2b(3), the LPF relies on the relaxation to priors to stabilize and improve the filter performance. We hypothesize that blending resampled particles with prior particles maintains a larger portion of the flow-dependent prior information that may be favorable for the spatial correlation estimation. The LNETF, if solved globally, preserves both the mean and covariance of the classical PF for all variables at all grid points [see section 2c(3)]. However, the implementation of the local-volume filter allows the LNETF to only preserve the posterior covariance of the PF for variables within the local volume, i.e., that at the central grid of each local volume in this study, probably degrading its performance in posterior spatial correlation among the grids across local volumes.

5. Conclusions and discussion

The classical PF can theoretically provide more accurate solutions than the EnKF and EnVar data assimilation strategies when priors and observation likelihoods have non-Gaussian distributions. Its application to high-dimensional systems, however, is restricted by the filter degeneracy issue associated with a requirement that the number of particles must increase exponentially with the state dimension. Among

the various attempts to apply PFs to high-dimensional systems with a limited ensemble size, some strategies, e.g., the localization, are adopted to stabilize the filter. However, these strategies inevitably increase the filter complexity and
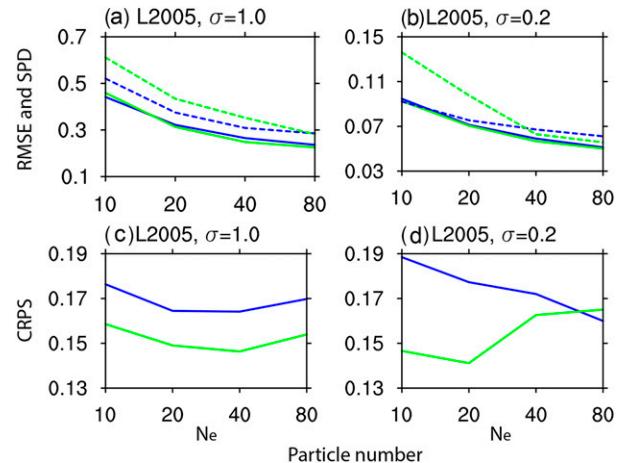

FIG. 8. Sample mean (a),(b) analysis RMSE (solid lines) and spread (dashed lines) and (c),(d) CRPS of posterior particles as a function of particle number for the LPF_loc-D_PoInfl (blue) and LNETF_loc-D_PoInfl (green) in the Lorenz2005 model.
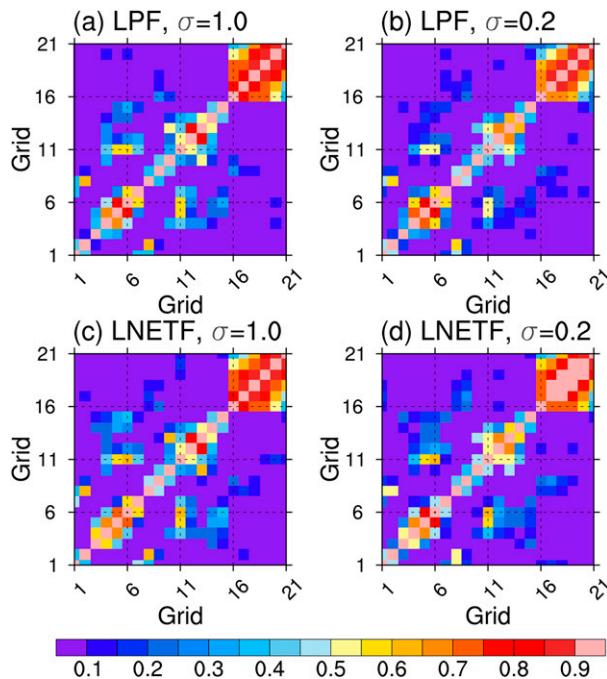
FIG. 9. Spatial correlation of posterior particles from grid 1 to 21 for the (a),(b) LPF_loc-D and (c),(d) LNETF_loc-D with σ = (left) 1.0 and (right) 0.2 in the Lorenz2005 model.

some ''cure'' may deviate from the original motivation of the Monte Carlo method. The two nonlinear filters, LPF and LNETF, compared in this study belong to this category. The LPF and LNETF adopt the partially sequential and the simultaneous assimilation, respectively. Both filters localize observational impacts to alleviate filter degeneracy but using different localization functions. In addition, they adopt distinct schemes to update posterior particles. The final result of all these ''cures'' is that both methods can preserve the posterior mean and variance of the PF. The LNETF additionally preserves the posterior correlation of variables within each local volume. The current study systematically compares the two methods in the widely used Lorenz (2005) model to reveal the impacts coming from their differences. Methods to improve the LNETF to reach indistinguishable performance relative to the LPF are proposed and tested.

The results showed that the original LPF displays a much larger stable parameter range than the original LNETF with varying localization and inflation parameters for all particle numbers when the same adaptive observational error inflation scheme (i.e., β inflation) is used. The minimum average RMSE of the original LPF is significantly (at the 0.05 level) smaller than that of the original LNETF with very small ensemble size like $N_e = 10$. Their RMSE scores become similar with larger particle numbers. Such performance is found to be closely related to their distinct original localization functions (loc-D versus Loc-P) and particle update algorithms. The localization function used in the LPF (i.e., loc-D), when the weights assigned on the particles are largely different, imposes stronger localization effects than the localization function used in the

LNETF (loc-P). Such property of the loc-D potentially reduces the risk of weight collapse and improves the filter stability. Our experiments have shown that the performance of the LNETF is improved when it uses the same localization function (i.e., loc-D) as the LPF.

In addition to the localization function, the filter performance difference among the original LPF and LNETF is also attributed to their distinct filter algorithms. Our experiments suggest that the relaxation of posterior particles to priors (γ relaxation) is critical for the resampling step of the LPF to enhance the particle diversity and maintain the filter stability. In contrast, the LNETF does not require the γ relaxation since it incorporates a random rotation of perturbations to the deterministic square root transformation of priors to stabilize the filter.

Our experiment results also demonstrate that the posterior mean and variance of the LPF and LNETF are identical given the same priors and localization function. It is because both filters use the localized particle weights on original prior particles to estimate posterior mean and variance regardless of the serial or simultaneous processing of observations. For the LPF, its posteriors from a given observation are used as the priors for deriving the posterior particles given the next observation before the latter are adjusted to match the posterior mean and variance (i.e., the defined partially sequential assimilation). With regard to the posterior distribution, the posterior particles of the LNETF exhibit an overall narrower range when using the same prior particles, observations, and localization function (i.e., loc-D) as the LPF, resulting in a higher percentage of true state out of posterior ensemble. Therefore, additional inflation of posterior particles was applied to the LNETF. As expected, it significantly improves the filter stability and analysis accuracy of the LNETF, especially with fewer particles, reaching similar performance as the LPF.

Several other aspects of the posterior particles are evaluated and compared. The LNETF has 10%–40% larger posterior ensemble spread than the LPF due to using a larger posterior inflation factor, but both algorithms overestimate the posterior spread on average. The LNETF exhibits a more accurate probability distribution of posterior particles than the LPF in terms of the CRPS. The inferior CRPS by the LPF is likely attributed to the relaxation of the LPF to the prior particles. This relaxation process may cause the loss of certain observational information in posterior particles. On the other hand, the LPF exhibits a slightly improved estimation of the spatial correlation of posterior particles relative to the LNETF when the ensemble size is relatively large.

It is worthwhile to note that while this study addresses the differences between the LPF and LNETF, methods that further improve the LNETF are proposed. Tödter and Ahrens (2015) demonstrated that the LNETF does not work with very small ensemble sizes such as 10. Our study has shown that the filter stability of LNETF with such a small particle number in the same model can be improved by using the β inflation, loc-D localization function, and posterior inflation. After optimization, the two nonlinear filters ultimately reach very similar levels of analysis accuracy. These initial experiments, despite conducted in a simple model, can provide implications for the
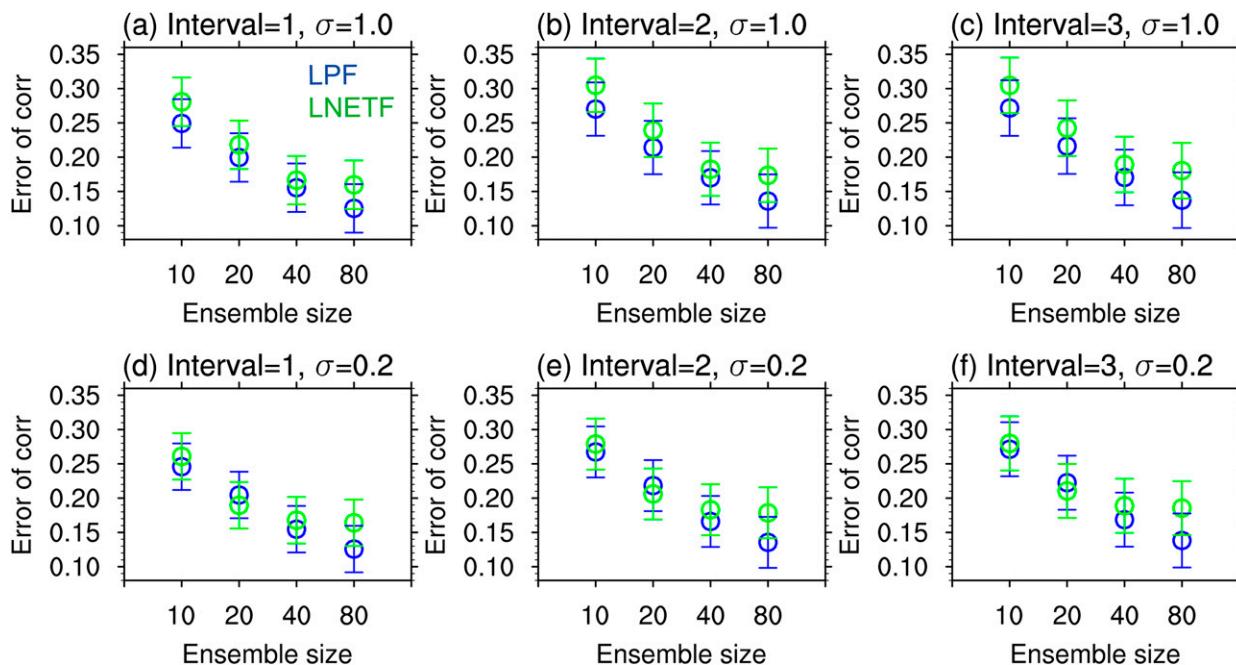
FIG. 10. Average absolute errors of posterior spatial correlation between grids with intervals (a),(d) 1; (b),(e) 2; and (c),(f) 3 for the LPF_loc-D_PoInfl (blue circles) and LNETF_loc-D_PoInfl (green circles) with σ = (top) 1.0 and (bottom) 0.2 in the Lorenz2005 model. Vertical bars represent the uncertainty of references.

development of the classical PF and nonlinear DA in high-dimensional systems. It is noted that the LPF translates into the classical PF as the particle number gets larger and its various parameters are relaxed. From our findings, we also emphasize that caution must be taken when comparing DA algorithms, since generic strategies for maintaining filter stability (such as choices of localization function and inflation) can have a major impact on conclusions. Comparing the two nonlinear filters in a more realistic high-dimensional model and/or considering model deficiencies is left for future study. The physical balance in analysis fields also deserves to be carefully investigated for the two localized particle filters. For future experiments and for a fair comparison with the LNETF adopting a subdomain, the LPF should be designed to use the Gaspari–Cohn like function (Gaspari and Cohn 1999).

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.

——, 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186–4198, https://doi.org/10.1175/2010MWR3253.1.

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758, https://doi.org/10.1175/1520-0493(1999)127%3C2741:AMCIOT%3E2.0.CO;2.

Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.*, **108**, 8775, https://doi.org/10.1029/2002JD002900.

——, P. Bickel, and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, D. Nolan and T. Speed, Eds., Vol. 2, Institute of Mathematical Statistics, 316–334.

Berger, J. O., and R. L. Wolpert, 1984: *The Likelihood Principle*. Institute of Mathematical Statistics, 208 pp.

Bickel, P., B. Li, and T. Bengtsson, 2008: Sharp failure rates for the bootstrap particle filter in high dimensions. *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, B. Clarke and S. Ghosal, Eds., Vol. 3, Institute of Mathematical Statistics, 318–329.

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

Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, **138**, 2997–3023, https://doi.org/10.1175/2010MWR3164.1.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, https://doi.org/10.1256/qj.04.15.

Doucet, A., N. de Freitas, and N. Gordon, Eds., 2001: An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 3–14, https://doi.org/10.1007/978-1-4757-3437-9_1.

Ehrendorfer, M., 2007: A review of issues in ensemble-based Kalman filtering. *Meteor. Z.*, **16**, 795–818, https://doi.org/10.1127/0941-2948/2007/0256.

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. E. 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.

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

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Holland, B., and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **139**, 758–770, https://doi.org/10.1002/qj.2006.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Huang, B., X. G. Wang, and C. H. Bishop, 2019: The high-rank ensemble transform Kalman filter. *Mon. Wea. Rev.*, **147**, 3025–3043, https://doi.org/10.1175/MWR-D-18-0210.1.

Hunt, B. R., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, https://doi.org/10.1016/j.physd.2006.11.008.

Legrand, R., Y. Michel, and T. Montmerle, 2016: Diagnosing non-Gaussianity of forecast and analysis errors in a convective-scale model. *Nonlinear Processes Geophys.*, **23**, 1–12, https://doi.org/10.5194/npg-23-1-2016.

Lei, J., and P. Bickel, 2011: A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Mon. Wea. Rev.*, **139**, 3964–3973, https://doi.org/10.1175/2011MWR3553.1.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP: A comparison with 4D-var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, https://doi.org/10.1256/qj.02.132.

Lorenz, E. N., 1996: Predictability—A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.

——, 2005: Designing chaotic models. *J. Atmos. Sci.*, **62**, 1574–1587, https://doi.org/10.1175/JAS3430.1.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808, https://doi.org/10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.

Morzfeld, M., and D. Hodyss, 2019: Gaussian approximations in filters and smoothers for data assimilation. *Tellus*, **71A**, 1600344, https://doi.org/10.1080/16000870.2019.1600344.

Nerger, L., T. Janjić, J. Schroter, and W. Hiller, 2012: A unification of ensemble square root Kalman filters. *Mon. Wea. Rev.*, **140**, 2335–2345, https://doi.org/10.1175/MWR-D-11-00102.1.

Oke, P. R., P. Sakov, and S. P. Corney, 2007: Impacts of localization in the EnKF and EnOI: Experiments with a small model. *Ocean Dyn.*, **57**, 32–45, https://doi.org/10.1007/s10236-006-0088-8.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428, https://doi.org/10.3402/tellusa.v56i5.14462.

Pires, C. A., O. Talagrand, and M. Bocquet, 2010: Diagnosis and impacts of non-Gaussianity of innovations in data assimilation. *Physica D*, **239**, 1701–1717, https://doi.org/10.1016/j.physd.2010.05.006.

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.

——, 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**, 2007–2020, https://doi.org/10.1175/MWR-D-15-0322.1.

——, L. Wicker, and M. Buehner, 2019: Progress toward the application of a localized particle filter for numerical weather prediction. *Mon. Wea. Rev.*, **147**, 1107–1126, https://doi.org/10.1175/MWR-D-17-0344.1.

Rainwater, S., and B. R. Hunt, 2013: Ensemble data assimilation with an adjusted forecast spread. *Tellus*, **65A**, 19929, https://doi.org/10.3402/tellusa.v65i0.19929.

Sakov, P., and P. R. Oke, 2008: Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Wea. Rev.*, **136**, 1042–1053, https://doi.org/10.1175/2007MWR2021.1.

Shen, Z., Y. Tang, and X. Li, 2017: A new formulation of vector weights in localized particle filters. *Quart. J. Roy. Meteor. Soc.*, **143**, 3269–3278, https://doi.org/10.1002/qj.3180.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640, https://doi.org/10.1175/2008MWR2529.1.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the national centers for environmental prediction global model. *Tellus*, **57A**, 528–545, https://doi.org/10.3402/tellusa.v57i4.14721.

Tödter, J., and B. Ahrens, 2015: A second-order exact ensemble square root filter for nonlinear data assimilation. *Mon. Wea. Rev.*, **143**, 1347–1367, https://doi.org/10.1175/MWR-D-14-00108.1.

——, P. Kirchgessner, L. Nerger, and B. Ahrens, 2016: Assessment of a nonlinear ensemble transform filter for high-dimensional data assimilation. *Mon. Wea. Rev.*, **144**, 409–427, https://doi.org/10.1175/MWR-D-15-0073.1.

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

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158, https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

——, and T. Lei, 2014: GSI-based four dimensional ensemble-variational (4DEnsVar) data assimilation: Formulation and single resolution experiments with real data for NCEP global forecast system. *Mon. Wea. Rev.*, **142**, 3303–3325, https://doi.org/10.1175/MWR-D-13-00303.1.

——, C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex

ensemble? *Mon. Wea. Rev.*, **132**, 1590–1605, https://doi.org/
10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2.
——, C. Snyder, and T. M. Hamill, 2007: On the theoretical
equivalence of differently proposed ensemble/3D-Var hybrid
analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227, https://
doi.org/10.1175/MWR3282.1.
——, D. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid
ETKF-3DVAR data assimilation scheme for the WRF model.
Part I: Observing system simulation experiment. *Mon. Wea. Rev.*,
136, 5116–5131, https://doi.org/10.1175/2008MWR2444.1.
——, ——, ——, and ——, 2008b: A hybrid ETKF-3DVAR data
assimilation scheme for the WRF model. Part II: Real obser-
vation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147, https://
doi.org/10.1175/2008MWR2445.1.

——, D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-
based ensemble–variational hybrid data assimilation for
NCEP Global Forecast System: Single-resolution experi-
ments. *Mon. Wea. Rev.*, **141**, 4098–4117, https://doi.org/10.1175/
MWR-D-12-00141.1.
Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data as-
similation without perturbed observations. *Mon. Wea.
Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)
130<1913:EDAWPO>2.0.CO;2.
Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial esti-
mate 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%3C1238:IOIEAO%3E2.0.CO;2.