The High-Rank Ensemble Transform Kalman Filter

BO HUANG AND XUGUANG WANG

School of Meteorology, University of Oklahoma, Norman, Oklahoma

CRAIG H. BISHOP

School of Earth Sciences, University of Melbourne, Melbourne, Australia

(Manuscript received 18 June 2018, in final form 31 May 2019)

ABSTRACT

The ensemble Kalman filter is typically implemented either by applying the localization on the background error covariance matrix (B-localization) or by inflating the observation error variances (R-localization). A mathematical demonstration suggests that for the same effective localization function, the background error covariance matrix from the B-localization method shows a higher rank than the R-localization method. The B-localization method is realized in the ensemble transform Kalman filter (ETKF) by extending the background ensemble perturbations through modulation (MP-localization). Specifically, the modulation functions are constructed from the leading eigenvalues and eigenvectors of the original B-localization matrix. Because of its higher rank than the classic R-localized ETKF, the B-/MP-localized ETKF is termed as the high-rank ETKF (HETKF). The performances of the HETKF and R-localized ETKF were compared through cycled data assimilation experiments using the Lorenz model II. The results show that the HETKF outperforms the R-localized ETKF especially for a small ensemble. The improved analysis in the HETKF is likely associated with the higher rank from the B-/MP-localization method, since its higher rank is expected to contribute more positively to alleviating the rank deficiency issue and thus improve the analysis for a small ensemble. The HETKF is less sensitive to the localization length scales and inflation factors. Furthermore, the experiments suggest that the above conclusion comparing the HETKF and R-localized ETKF does not depend on how the analyzed ensemble perturbations are subselected in the HETKF.

1. Introduction

The ensemble Kalman filter (EnKF; Evensen 1994) has been widely used in the atmospheric applications, since it was introduced as a Monte Carlo realization of the traditional Kalman filter (Kalman and Bucy 1961). In the EnKF, the background error covariances are estimated and evolved by cycling an ensemble of short-range forecasts and analyses. Compared to the three-dimensional variational (3DVar) method generally employing the static background error covariances, the EnKF embraces the advantage of accounting for the flow dependency of the forecast errors. The EnKF is therefore able to estimate the spatial, temporal, and multivariate error covariances in a more realistic fashion. Different variants of the EnKF have been developed for efficient implementation purposes (Houtekamer

and Mitchell 1998; Anderson 2001; Bishop et al. 2001, 2015, 2017; Whitaker and Hamill 2002; Wang and Bishop 2003; Wang et al. 2004; Hunt et al. 2007).

In the EnKF, the ensemble background error covariances, along with the observation error covariances, determine the pattern and magnitude of the corrections made on the model state variables by assimilating observations. Because of the computational constraints, the current operational EnKF systems generally run an ensemble with a size much smaller than the dimension of the numerical models (Houtekamer and Zhang 2016, their Table 1). This limited ensemble size causes sampling errors and rank deficiency in the estimated background error covariance matrix. If not properly treated, these issues will incur noisy analysis increments and even filter divergence (Hamill 2006). Directly increasing the ensemble size is able to improve the estimate of the background error covariances and thus the accuracy of the analyses and subsequent forecasts

Corresponding author: Xuguang Wang, xuguang.wang@ou.edu

DOI: 10.1175/MWR-D-18-0210.1

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

(Miyoshi et al. 2014; Lei and Whitaker 2017; Huang and Wang 2018). But the cost is very expensive. On the other hand, increasing the ensemble size in a cost-effective fashion in the ensemble-based data assimilation system is explored and demonstrated to improve the analyses and forecasts for the storm and global scales (e.g., Xu et al. 2008; Lorenc 2017; Huang and Wang 2018).

Alternatively, the covariance localization is commonly applied in the EnKF to deal with the aforementioned issues from running a small ensemble. Its general idea is to reduce or remove the correlations between two distant variables that are assumed to be physically small or spurious. On this basis, the distance-dependent localization is applied either on the background error covariance matrix (hereafter referred to as the B-localization method) or on the observation error covariance matrix (hereafter referred to as the R-localization method). The notations of the B-localization and R-localization methods are adapted from Greybush et al. (2011) and Holland and Wang (2013). The B-localization method is typically realized through a Schur product between the raw background error covariance matrix and a predefined distance-dependent localization matrix (Houtekamer and Mitchell 2001, 2005). The R-localization method is applied through inflating the observation error variances (Hunt et al. 2007). As a result, the corrections made by the distant observations are reduced or even removed after applying the localization. In general, the distance-dependent localization function is defined to be spatially homogeneous and temporally constant. Advanced localization methods were developed in the recent studies to account for the scale, spatial, or temporal dependency (Anderson 2007; Bishop and Hodyss 2007; Buehner and Charron 2007; Anderson and Lei 2013; Gasperoni and Wang 2015).

Miyoshi and Yamane (2007) and Greybush et al. (2011) found that in the assimilation of a single observation, the effective localization length scale in the R-localized Kalman gain was wider than that in the B-localized Kalman gain by applying the same localization function. It was also mentioned in these studies that the mathematical differences between the B-localization and R-localization methods were not straightforward to conclude in the assimilation of multiple observations. Sakov and Bertino (2011) compared the structures of the B-localized and R-localized Kalman gains at a single grid point, and suggested that both localization methods were expected to yield similar results in the practical applications.

While theoretical demonstrations of the mathematical differences between these two localization methods are limited, in the early studies, the performances of the B-localization and R-localization methods were usually

empirically evaluated and compared in terms of the analysis accuracy by running cycled data assimilation experiments. Janjić et al. (2011) and Nerger et al. (2012) using the Lorenz-96 model (Lorenz 1996) found that the B-localization method outperformed the R-localization method especially when the observation errors were much smaller compared to the background errors. Cycled data assimilation experiments in a simplified dynamical model in Greybush et al. (2011) showed that the B-localization and R-localization methods performed comparably, if both were optimally tuned. In these studies, the B-localization method was typically applied for the variants of the serial square root filter, and the R-localization method for the variants of the parallel implementation of the local ensemble filter. An exception was Janjić et al. (2011), which compared the B-localization and R-localization methods by performing a local analysis update using the same singular evolutive interpolated Kalman (SEIK) filter. Holland and Wang (2013) using a two-layer primitive equation model compared the B-localization and R-localization methods in both the same serial and same simultaneous square root filters. They found that the B-localization and R-localization methods resulted in different amounts of imbalance, which in turn affected the analysis accuracy.

This study contributes to the theoretical understanding of the differences between the B-localization and R-localization methods. A mathematical derivation is first provided with a focus on demonstrating the effective ranks of the background error covariance matrices by applying these two localization methods. The derivation does not rely on the assimilation of a single observation. Briefly, it is mathematically demonstrated in section 3 that for the same effective localization function, the B-localization method achieves a higher rank than the R-localization method in the localized background error covariance matrix. Meanwhile, the mathematical demonstration also shows that the B-localization method can be realized through extending and modulating the raw background ensemble perturbations (hereafter referred to as the MP-localization method). To reduce the computational cost, truncation of the eigenvectors from the B-localization matrix is applied to generate the modulation functions in the MP-localization method following Bishop et al. (2017). The MP-localized background error covariance matrix is thus consistent with that applying the traditional B-localization method.

The R-localization method is commonly applied in the ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003; Wang et al. 2004) by increasing the observation error variances with an increasing distance from the model state variable

(Hunt et al. 2007). In this study, the mathematical demonstration also shows that the R-localization method can be expressed in the form of the modulated background ensemble perturbations as in the B-/MP-localization method. This inspires the comparison of these two localization methods within the same ETKF algorithm through cycled data assimilation experiments. In contrast to most of the early studies that compared these two localization methods using different filters, such a comparison of the B-localization and R-localization methods within the same ETKF algorithm makes it more straightforward to link the resulting analysis performances with the localization differences.

To emphasize the mathematically derived higher-rank feature from the B-/MP-localization method, the B-/MP-localized ETKF in this study is interchangeably referred to as the high-rank ETKF (hereafter referred to as the HETKF), to distinguish it from the classic R-localized ETKF. In addition, two analysis ensemble perturbation subselection methods in Bishop et al. (2017) were implemented in the HETKF to investigate if such perturbation subselection methods affect the performances of the HETKF and R-localized ETKF.

The paper is organized as follows. Section 2 briefly introduces the B-localization and R-localization methods in the context of the generic EnKF update equations. Section 3 provides a mathematical derivation to demonstrate the rank differences of the B-localized and R-localized Kalman gains in the generic EnKF context. The ETKF algorithm and its R-localized form are briefly described in section 4. Section 5 describes the implementation of the B-/MP-localization method in the HETKF. The performances of the HETKF and R-localized ETKF are evaluated and compared using the Lorenz model II in section 6. The conclusions and discussion are presented in section 7.

2. B-localization and R-localization methods in the generic EnKF

In this section, the notations in Ide et al. (1997) are used to illustrate the generic EnKF equations (Evensen 1994). The analysis $\mathbf{x}^a = [x_1^a, x_2^a, \dots, x_n^a]^T$ is obtained by correcting the background $\mathbf{x}^b = [x_1^b, x_2^b, \dots, x_n^b]^T$ with the observations $\mathbf{y}^o = [y_1^o, y_2^o, \dots, y_p^o]^T$ weighted by the Kalman gain $\mathbf{K}$:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y}^o - H(\mathbf{x}^b)], \quad (1)$$

and

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1}, \quad (2)$$

where $\mathbf{x}^a$ and $\mathbf{x}^b$ are the analysis and background vectors with a dimension of $n \times 1$, respectively; $\mathbf{y}^o$ is the observation vector with a dimension of $p \times 1$; H and $\mathbf{H}$ are the nonlinear and linearized observation operators, respectively; $\mathbf{P}^b$ denotes the background error covariance matrix with a dimension of $n \times n$; and $\mathbf{R} = \mathbf{I}r^2$ is the diagonal observation error covariance matrix with a dimension of $p \times p$. For simplicity, all the diagonal elements in $\mathbf{R}$ are set equal to r^2 . The Kalman gain matrix $\mathbf{K}$ has a dimension of $n \times p$. The superscripts a , b , and o denote the analysis, background, and observations, respectively. In the EnKF, $\mathbf{P}^b$ is estimated from a K -member ensemble of background forecasts:

$$\mathbf{P}^b = \frac{1}{K-1} \mathbf{X}^b (\mathbf{X}^b)^T = \mathbf{Z}^b (\mathbf{Z}^b)^T, \quad (3)$$

where K is the ensemble size; $\mathbf{X}^b = [\mathbf{x}_1^b, \mathbf{x}_2^b, \dots, \mathbf{x}_K^b] = [\mathbf{x}_1^b - \bar{\mathbf{x}}^b, \mathbf{x}_2^b - \bar{\mathbf{x}}^b, \dots, \mathbf{x}_K^b - \bar{\mathbf{x}}^b]$ is the background ensemble perturbation matrix with a dimension of $n \times K$ and each column represents the k th ensemble perturbation vector $\mathbf{x}_k^b$ with a dimension of $n \times 1$; $\bar{\mathbf{x}}^b = \sum_{k=1}^K \mathbf{x}_k^b / K$ is the background ensemble mean vector with a dimension of $n \times 1$; and $\mathbf{Z}^b = \mathbf{X}^b / \sqrt{K-1}$ is the background ensemble perturbation matrix normalized by a factor of $\sqrt{K-1}$.

As discussed in the introduction, the B-localization method is generally realized through a Schur product (denoted by “ $\circ$ ”) between the raw background error covariance matrix and a B-localization matrix $\mathbf{L}$ with the same dimension of $n \times n$:

$$\mathbf{P}_{\text{Bloc}}^b = \mathbf{P}^b \circ \mathbf{L}. \quad (4)$$

In practice, $\mathbf{L}$ is commonly defined by the Gaussian functions with its diagonal elements equal to 1.0 (see details in the next section). The B-localized $\mathbf{P}_{\text{Bloc}}^b$ is only calculated once and used to update the variables at all model grid points. For example, at the i th grid point, the B-localized Kalman gain calculated from $\mathbf{P}_{\text{Bloc}}^b$ is given by

$$(\mathbf{K}_{\text{Bloc}})_i = (\mathbf{P}_{\text{Bloc}}^b \mathbf{H}^T)_i (\mathbf{H} \mathbf{P}_{\text{Bloc}}^b \mathbf{H}^T + \mathbf{R})^{-1}. \quad (5)$$

To be consistent, in the rest of the paper, the subscript i outside the parentheses denotes the i th row of a matrix or the i th element of a vector. In Eq. (5), it denotes the i th row of the matrices.

In the R-localization method, to update the variables at the i th grid point, the diagonal elements in the original observation error covariance matrix $\mathbf{R}$ are inflated. A larger inflation coefficient is applied with an increasing distance away from the i th grid point:

$$\mathbf{R}_{\text{Rloc}}^i = [\text{diag}(\mathbf{g}_i)]^{-1} \mathbf{R} [\text{diag}(\mathbf{g}_i)]^{-1}. \quad (6)$$

In Eq. (6), the superscript i in $\mathbf{R}_{\text{Rloc}}^i$ denotes that the variables to be updated are at the i th grid point. The vector $\mathbf{g}_i$ with a dimension of $p \times 1$ is a distance-dependent monotonically decreasing function. It has the maximum value of 1.0 at the location of the i th grid point. The vector $\mathbf{g}_i$ is commonly defined by a Gaussian function (see details in the next section). Here “diag” is an operator that converts a vector to a diagonal matrix by aligning the elements of the vector along the diagonal. The R-localized Kalman gain at the i th grid point is given by

$$(\mathbf{K}_{\text{Rloc}})_i = (\mathbf{P}^b \mathbf{H}^T)_i (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}_{\text{Rloc}}^i)^{-1}, \quad (7)$$

where the subscript i outside the parentheses, as defined earlier, denotes the i th row in the matrices.

3. Mathematical demonstration of the higher rank of the B-localization method over the R-localization method

In this section, the B-localized and R-localized Kalman gains at the i th grid point shown in section 2 are reformulated to examine their differences. To make the derivations in both localization methods straightforward and consistent, two assumptions are made: (i) all the model grid points are observed (i.e., $n = p$), and (ii) the periodic boundary condition is applied.

First, the Gaussian function $\mathbf{g}_i$ at the i th grid point is defined. It determines the correlations between the i th grid point and the other grid points. The following describes the formation of the B-localization matrix $\mathbf{L}$ that is used to localize the full background error covariance matrix. Since the full background error covariance matrix is associated with all the model grid points, a Gaussian matrix $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \dots, \mathbf{g}_n]$ with a dimension of $n \times n$ is first formed, where the i th column is defined by the vector $\mathbf{g}_i$ with a dimension of $n \times 1$. Further following Eqs. (23) and (27) in Bishop et al. (2015), the Gaussian matrix $\mathbf{G}$ is explicitly calculated as

$$\mathbf{G} = \mathbf{F} \mathbf{\Phi} \mathbf{F}^T, \quad (8)$$

where $\mathbf{F}$ is an orthonormal matrix with a dimension of $n \times n$ ($\mathbf{F}^T \mathbf{F} = \mathbf{F} \mathbf{F}^T = \mathbf{I}$, $\mathbf{I}$ is the identity matrix) with each column representing an eigenvector of $\mathbf{G}$, and $\mathbf{\Phi}$ is a positive semidefinite diagonal matrix with a dimension of $n \times n$ with the diagonal elements representing the eigenvalues of $\mathbf{G}$. $\mathbf{F}$ is defined by a discrete Fourier basis of sine and cosine functions following Bishop et al. (2015). The i th element ϕ_{ii} in the diagonal matrix $\mathbf{\Phi}$ is calculated by

FIG. 1. Distribution of the Gaussian functions defined at every twenty grid points.

$$\phi_{ii} = n \sigma^2 \frac{\exp \left\{ - \left[\frac{s(i)}{d} \right]^2 \right\}}{\sum_{i=1}^n \exp \left\{ - \left[\frac{s(i)}{d} \right]^2 \right\}}, \quad (9)$$

where $s(i)$ is the wavenumber of the i th sinusoidal eigenfunction corresponding to ϕ_{ii} at the i th grid point, and the parameter d determines the width of the distribution of the Gaussian vector $\mathbf{g}_i$. Specifically, a larger d results in a tighter Gaussian distribution. In Eq. (9), $n \sigma^2$ is equal to the sum of all the eigenvalues ϕ_{ii} in the diagonal matrix $\mathbf{\Phi}$ or the sum of all the diagonal elements in the Gaussian matrix $\mathbf{G}$. Following the cycled DA experiments in section 6, $n = 240$ is chosen for illustration purposes in this section. Here $\sigma^2 = 1$ is selected to ensure that the peak value in the Gaussian vector $\mathbf{g}_i$ is equal to 1.0 at the i th grid point as required in Eq. (6). Figure 1 shows an example of the distribution of the Gaussian function $\mathbf{g}_i$ defined at every 20 grid points by selecting $d = 3$ in Eq. (9). The magnitude of $\mathbf{g}_i$ peaks at the i th grid point [e.g., $(\mathbf{g}_i)_i = 1$ where, as defined earlier, the subscript i outside the parentheses denotes the i th element in $\mathbf{g}_i$] and asymptotically decreases away from the i th grid point. Here, the assumption (ii) is applied to make the Gaussian functions periodically distributed.

The B-localization matrix $\mathbf{L}$ is then formed by

$$\mathbf{W} = \mathbf{G} \mathbf{G}^T, \quad (10)$$

and

$$\begin{aligned} \mathbf{L} &= [\text{DIAG}(\mathbf{W})]^{-1/2} \mathbf{G} \mathbf{G}^T [\text{DIAG}(\mathbf{W})]^{-1/2} \\ &= [\text{DIAG}(\mathbf{W})]^{-1/2} [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n] [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n]^T \\ &\quad \times [\text{DIAG}(\mathbf{W})]^{-1/2}, \end{aligned} \quad (11)$$

where the matrix $\mathbf{W}$ has a dimension of $n \times n$ and the operator “DIAG” functions as only retaining the diagonal elements in a square matrix and setting the

off-diagonal elements equal to zero. The purpose of the left and right multiplication of $[\text{DIAG}(\mathbf{W})]^{-1/2}$ in Eq. (11) is to normalize the diagonal elements in the B-localization matrix $\mathbf{L}$ equal to 1.0. It can be further simplified by the periodic nature of the defined Gaussian functions:

$$\text{DIAG}(\mathbf{W}) = \sum_{j=1}^n [\text{diag}(\mathbf{g}_j)]^2 = \left\{ \sum_{j=1}^n [(\mathbf{g}_j)_i]^2 \right\} \mathbf{I} = w^2 \mathbf{I}, \quad (12)$$

where $w = w_{ii} = \sqrt{\sum_{j=1}^n [(\mathbf{g}_j)_i]^2}$. The term w_{ii} is independent of the index i because of the isotropic and periodic nature of the Gaussian functions.

By introducing Eqs. (3) and (11), the B-localized $\mathbf{P}_{\text{Bloc}}^b$ in Eq. (4) is rewritten as

$$\begin{aligned} \mathbf{P}_{\text{Bloc}}^b &= \{\mathbf{Z}^b (\mathbf{Z}^b)^T\} \circ \{[\text{DIAG}(\mathbf{W})]^{-1/2} [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n] [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n]^T [\text{DIAG}(\mathbf{W})]^{-1/2}\} \\ &= \sum_{j=1}^n \{[\text{DIAG}(\mathbf{W})]^{-1/2} [\text{diag}(\mathbf{g}_j)] \mathbf{Z}^b\} \{[\text{DIAG}(\mathbf{W})]^{-1/2} [\text{diag}(\mathbf{g}_j)] \mathbf{Z}^b\}^T \\ &= \sum_{j=1}^n \left\{ \left[\frac{1}{w} \mathbf{I} \right] [\text{diag}(\mathbf{g}_j)] \mathbf{Z}^b \right\} \left\{ \left[\frac{1}{w} \mathbf{I} \right] [\text{diag}(\mathbf{g}_j)] \mathbf{Z}^b \right\}^T. \end{aligned} \quad (13)$$

To simplify Eq. (13), we further define

$$\mathbf{V}_j^b = [\text{diag}(\mathbf{g}_j)] \mathbf{Z}^b. \quad (14)$$

Equation (13) then becomes

$$\begin{aligned} \mathbf{P}_{\text{Bloc}}^b &= \frac{1}{w^2} \sum_{j=1}^n (\mathbf{V}_j^b) (\mathbf{V}_j^b)^T \\ &= \frac{1}{w^2} [\mathbf{V}_1^b, \mathbf{V}_2^b, \dots, \mathbf{V}_n^b] [\mathbf{V}_1^b, \mathbf{V}_2^b, \dots, \mathbf{V}_n^b]^T. \end{aligned} \quad (15)$$

In Eq. (14), the matrix $\mathbf{V}_j^b$ with a dimension of $n \times K$ can be interpreted as modulating the raw background ensemble perturbation matrix by the Gaussian function $\mathbf{g}_j$

defined at the j th grid point. In particular, each column of the matrix $\mathbf{V}_j^b$ corresponds to a Schur product between a raw ensemble perturbation vector and the Gaussian vector $\mathbf{g}_j$ defined at the j th grid point. Equations (13)–(15) suggest that the B-localization method can be realized by an outer product of the expanded modulated ensemble perturbation matrix $1/(w) [\mathbf{V}_1^b, \mathbf{V}_2^b, \dots, \mathbf{V}_n^b]$ with a dimension of $n \times (nK)$. Since the B-localization method here is achieved through expanding and Modulating the raw ensemble Perturbation matrix, it is interchangeably termed as the MP-localization method.

Correspondingly, the Kalman gain at the i th grid point with the B-localized $\mathbf{P}_{\text{Bloc}}^b$ in Eq. (5) is reformulated by introducing Eq. (15):

$$\begin{aligned} (\mathbf{K}_{\text{Bloc}})_i &= \left\{ \left[\frac{1}{w^2} \sum_{j=1}^n (\mathbf{V}_j^b) (\mathbf{V}_j^b)^T \right] \mathbf{H}^T \right\}_i \left\{ \mathbf{H} \left[\frac{1}{w^2} \sum_{j=1}^n (\mathbf{V}_j^b) (\mathbf{V}_j^b)^T \right] \mathbf{H}^T + \mathbf{R} \right\}^{-1} \\ &= \left\{ \frac{1}{w^2} \sum_{j=1}^n (\mathbf{V}_j^b)_i (\mathbf{H} \mathbf{V}_j^b)^T \right\} \left\{ \frac{1}{w^2} \sum_{j=1}^n (\mathbf{H} \mathbf{V}_j^b) (\mathbf{H} \mathbf{V}_j^b)^T + \mathbf{R} \right\}^{-1}. \end{aligned} \quad (16)$$

For a more straightforward comparison with the R-localized Kalman gain form as will be introduced soon, Eq. (16) is further right multiplied by an identity

matrix with a dimension of $n \times n$ that is expressed as $\mathbf{I} = \sqrt{\sum_{j=1}^n [\text{diag}(\mathbf{g}_j)]^2} / \sqrt{\sum_{j=1}^n [(\mathbf{g}_j)_i]^2}$ by referring to Eq. (12):

$$(\mathbf{K}_{\text{Bloc}})_i = \left\{ \frac{1}{\sum_{j=1}^n [(\mathbf{g}_j)_i]^2} \sum_{j=1}^n (\mathbf{V}_j^b)_i (\mathbf{H} \mathbf{V}_j^b)^T \right\} \left\{ \frac{1}{\sum_{j=1}^n [(\mathbf{g}_j)_i]^2} \sum_{j=1}^n (\mathbf{H} \mathbf{V}_j^b) (\mathbf{H} \mathbf{V}_j^b)^T + \mathbf{R} \right\}^{-1} \frac{\sqrt{\sum_{j=1}^n [\text{diag}(\mathbf{g}_j)]^2}}{\sqrt{\sum_{j=1}^n [(\mathbf{g}_j)_i]^2}}. \quad (17)$$

In the R-localization method, given the assumption (i) that all the model grid points are observed (i.e., $n = p$), the Gaussian function $\mathbf{g}_i$ at the i th grid point is applied in

$$\begin{aligned} (\mathbf{K}_{\text{Rloc}})_i &= \{\mathbf{Z}^b \{\mathbf{H}\mathbf{Z}^b\}^T\}_i \{\mathbf{H}\mathbf{Z}^b (\mathbf{H}\mathbf{Z}^b)^T + [\text{diag}(\mathbf{g}_i)]^{-1} \mathbf{R} [\text{diag}(\mathbf{g}_i)]^{-1}\}^{-1} \\ &= (\mathbf{Z}^b)_i \{\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b]\}^T \{(\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b])\{\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b]\}^T + \mathbf{R}\}^{-1} \text{diag}(\mathbf{g}_i). \end{aligned} \quad (18)$$

By introducing the modulated background ensemble perturbation matrix in Eq. (14), Eq. (18) is simplified as

$$(\mathbf{K}_{\text{Rloc}})_i = (\mathbf{V}_i^b)_i (\mathbf{H}\mathbf{V}_i^b)^T [(\mathbf{H}\mathbf{V}_i^b)(\mathbf{H}\mathbf{V}_i^b)^T + \mathbf{R}]^{-1} \text{diag}(\mathbf{g}_i), \quad (19)$$

$$\begin{aligned} (\mathbf{K}_{\text{Rloc}})_i &= \frac{1}{[\mathbf{g}_i]_i^2} (\mathbf{V}_i^b)_i (\mathbf{H}\mathbf{V}_i^b)^T \left\{ \frac{1}{[\mathbf{g}_i]_i^2} (\mathbf{H}\mathbf{V}_i^b)(\mathbf{H}\mathbf{V}_i^b)^T + \mathbf{R} \right\}^{-1} \frac{\sqrt{[\text{diag}(\mathbf{g}_i)]^2}}{\sqrt{[\mathbf{g}_i]_i^2}} \\ &= \left\{ \frac{1}{\sum_{j=1}^n \delta_{ij} [\mathbf{g}_j]_i^2} \sum_{j=1}^n \delta_{ij} (\mathbf{V}_j^b)_i (\mathbf{H}\mathbf{V}_j^b)^T \right\} \left\{ \frac{1}{\sum_{j=1}^n \delta_{ij} [\mathbf{g}_j]_i^2} \sum_{j=1}^n \delta_{ij} (\mathbf{H}\mathbf{V}_j^b)(\mathbf{H}\mathbf{V}_j^b)^T + \mathbf{R} \right\}^{-1} \frac{\sqrt{\sum_{j=1}^n \{\delta_{ij} [\text{diag}(\mathbf{g}_j)]^2\}}}{\sum_{j=1}^n \delta_{ij} [\mathbf{g}_j]_i^2}, \end{aligned} \quad (20)$$

where,

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (21)$$

Equations (18)–(20) suggest that at the i th grid point, the localization effect by inflating the observation error variances in the R-localization method can be equivalently achieved by modulating the raw ensemble perturbation matrix with the Gaussian function $\mathbf{g}_i$ defined in Eq. (14). Such a reformulation assists in a direct mathematical comparison between the B-localization and R-localization methods.

By comparing Eqs. (17) and (20), it can be seen that the R-localization method can be regarded as a special case of the B-localization method when expressed using the Kronecker delta function in Eq. (21). The number of terms in the summations over the modulated ensemble perturbation matrix index j in the B-localization method in Eq. (17) is reduced to one in the R-localization method in Eq. (20). Specifically, in the B-localization method, a total of n modulated background ensemble perturbation matrices are involved in the calculation of the B-localized

Eq. (6) to calculate the localized $\mathbf{R}_{\text{Rloc}}^i$. The R-localized Kalman gain at the i th grid point is then obtained by introducing Eqs. (3) and (6) to Eq. (7):

where $(\mathbf{Z}^b)_i = (\mathbf{V}_i^b)_i$, given $(\mathbf{g}_i)_i = 1$.

To facilitate the comparison with the B-localized Kalman gain form in Eq. (17), Eq. (19) is further reformulated by introducing $[\mathbf{g}_i]_i^2 = 1$, $\text{diag}(\mathbf{g}_i) = \sqrt{[\text{diag}(\mathbf{g}_i)]^2} / \sqrt{[\mathbf{g}_i]_i^2}$, and the Kronecker delta function δ_{ij} in Eq. (21) below:

Kalman gain at the i th grid point. However, the R-localization method only includes the contribution from a single modulated ensemble perturbation matrix associated with the Gaussian function defined at the i th grid point. As a result, the rank of the B-localized Kalman gain is higher than that of the R-localization method. The above conclusion can also be drawn from a simple linear algebra analysis. Given $n > p > K$ in general, the rank of the original Kalman gain in Eq. (2) is $K - 1$. It is determined by the minimum of the ranks of $\mathbf{H}$, $\mathbf{P}^b$, and $\mathbf{R}$ in Eq. (2) that are p , $K - 1$, and p , respectively. Inflating the observation error variances of the original $\mathbf{R}$ as in Eq. (6) will not change its rank. The R-localized Kalman gain in Eq. (7) is thus of the same rank as the original Kalman gain. However, after applying the B-localization method, the maximum rank of $\mathbf{P}_{\text{Bloc}}^b$ in Eq. (4) can reach n due to its resultant block diagonal structure. The B-localized Kalman gain in Eq. (5) thus has an increased rank of p . Therefore, the linear algebra analysis also suggests a higher rank of the B-localized Kalman gain in contrast to the original and R-localized Kalman gains. This is consistent with the mathematical demonstration in this section.

The localization effects of the B-localization and R-localization methods are further isolated and compared by reformulating their localized

observation-space background error covariance matrices in Eqs. (17) and (20). For the B-localization method in Eq. (17):

$$\begin{aligned} \frac{1}{\sum_{j=1}^n [(\mathbf{g}_j)_i]^{2j=1}} \sum_{j=1}^n (\mathbf{H}\mathbf{V}_j^b)(\mathbf{H}\mathbf{V}_j^b)^T &= \frac{1}{w^2} \sum_{j=1}^n \{\mathbf{H}[\text{diag}(\mathbf{g}_j)\mathbf{Z}^b]\} \{\mathbf{H}[\text{diag}(\mathbf{g}_j)\mathbf{Z}^b]\}^T \\ &= \left\{ [\mathbf{H}\mathbf{Z}^b(\mathbf{H}\mathbf{Z}^b)^T] \circ \left[\frac{1}{w^2} \sum_{j=1}^n (\mathbf{g}_j\mathbf{g}_j^T) \right] \right\}. \end{aligned} \quad (22)$$

For the R-localization method in Eq. (20):

$$\begin{aligned} \frac{1}{\sum_{j=1}^n \delta_{ij} [(\mathbf{g}_j)_i]^{2j=1}} \sum_{j=1}^n \delta_{ij} (\mathbf{H}\mathbf{V}_j^b)(\mathbf{H}\mathbf{V}_j^b)^T &= \{\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b]\} \{\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b]\}^T \\ &= \{[\mathbf{H}\mathbf{Z}^b(\mathbf{H}\mathbf{Z}^b)^T] \circ (\mathbf{g}_i\mathbf{g}_i^T)\}. \end{aligned} \quad (23)$$

Equations (22) and (23) suggest that the effects of the localization applied on the observation-space background error covariance matrices are determined by the matrix $1/(w^2) \sum_{j=1}^n \mathbf{g}_j\mathbf{g}_j^T$ for the B-localization method and by the matrix $\mathbf{g}_i\mathbf{g}_i^T$ for the R-localization method, respectively. Figure 2 shows an example of the structures of these two matrices calculated at the 120th grid point (e.g., $i = 120$). The Gaussian matrix $\mathbf{G}$ in Eq. (8) is calculated by selecting $d = 3$ and $n = 240$ as in Fig. 1. The matrix $1/(w^2) \sum_{j=1}^n \mathbf{g}_j\mathbf{g}_j^T$ in Eq. (22) for the B-localization method (Fig. 2a) shows a symmetric structure with the magnitude equal to 1.0 on the diagonal and monotonically decreasing away from the diagonal. In contrast, the matrix $\mathbf{g}_i\mathbf{g}_i^T$ in Eq. (23) for the R-localization method (Fig. 2b) shows a localized circular structure centered at the element of (120, 120). For illustration purposes, the 120th row of these two matrices (e.g., the solid black line and the dashed black line in Figs. 2a and 2b) is used to determine the effective localization distance for the B-localization and R-localization methods in this study. Here the effective localization distance is defined as half of the interval where the localization coefficients taper to 0.01 (Bishop et al. 2015). As can be seen in Fig. 2c, the same localization parameter $d = 3$ in Eq. (9) results in a broader effective localization distance in the B-localization method than that in the R-localization method. This result seems inconsistent with the expectation that the tighter effective localization distance generally results in a higher rank of the localized background error

covariance matrix. However, Eqs. (17) and (20) suggest that the mathematically derived higher rank from the B-localization method is independent of the effective localization distance. Meanwhile, the effective localization distances in these two localization methods in Fig. 2c are caused by and consistent with the constructions of their localization matrices. Therefore, caution needs to be taken to relate the effective localization distance with the rank of the localized background error covariance matrix especially when different forms of localization are utilized.

To further verify the mathematical demonstration, the effective ranks resulting from both localization methods are calculated and compared in an example using $\mathbf{H} = \mathbf{I}$. In Fig. 2f, the raw observation-space background error covariance matrix is estimated from six members. Through a Schur product with the B-localization matrix (Fig. 2a), the B-localized observation-space background error covariance matrix retains the covariances near the diagonal, and reduces or even removes the covariances away from the diagonal (Fig. 2d). On the other hand, due to the “local” nature of the R-localization matrix (Fig. 2b), the R-localized observation-space background error covariance matrix (Fig. 2e) has more zero values on the diagonal in contrast to that applying the B-localization method. The E dimension (Patil et al. 2001; Oczkowski et al. 2005; Kuhl et al. 2007; Huang and Wang 2018) is used to evaluate the effective rank of the background error covariance matrix. The E dimension was

FIG. 2. Illustration of (a) the matrix $1/(w^2) \sum_{j=1}^n \mathbf{g}_j \mathbf{g}_j^T$ for the B-localization method in Eq. (22), (b) the matrix $\mathbf{g}_i \mathbf{g}_i^T$ ($i = 120$) for the R-localization method in Eq. (23) calculated with $d = 3$ and $n = 240$ in Eqs. (8) and (9), and (c) the localization coefficients from the 120th row of the localization matrix for the B-localization method (solid line) in (a) and the R-localization method (dashed line) in (b), respectively. The effective localization distance is defined as half of the interval where the correlation coefficients taper to 0.01. Illustration of the observation-space ensemble background error covariance matrix estimated from 6 members with (f) no localization, (d) the B-localization method, and (e) the R-localization method, and the size of the black solid and dashed squares in (d) and (e), respectively, doubles the effective localization distances as shown in (c). First five leading eigenvectors (colored solid lines) decomposed from (g) the B-localized and (h) the R-localized observation-space ensemble background error covariance matrices in (d) and (e), respectively. (i) Kalman gains of updating the 120th grid point from the B-localization method (solid line) and the R-localization method (dashed line).

calculated within a square as shown in Figs. 2d and 2e. The size of the square is defined as twice the effective localization distance. The B-localization method increases the E dimension from 4.21 to 10.33. This is consistent with the expansion of the dimension of the modulated ensemble perturbation matrix after applying the B-localization method in Eq. (15). In contrast, the R-localization method results in a reduced E dimension of 2.81 from 3.72. This is likely due to the occurrence of more zero values on the diagonal of the R-localized background error covariance matrix. Therefore, the mathematically derived higher rank in

the B-localized observation-space background error covariance matrix is consistently demonstrated by comparing their resulting effective ranks.

The different structures of the B-localized and R-localized observation-space background error covariance matrices further motivate an investigation of how many observations would literally influence their resulting analyses. The Kalman gains at the 120th grid point are thus calculated for these two localization methods. In general, the matrix inversion in Eq. (2) for the Kalman gain calculation can be solved by using the eigenvalues and eigenvectors decomposed from the

observation-space background error covariance matrix (Bishop et al. 2017). Figures 2g and 2h show the first five leading eigenvectors decomposed from the B-localized and R-localized observation-space background error covariance matrices in Figs. 2d and 2e, respectively. The eigenvectors from the B-localized observation-space background error covariance matrix cover the whole domain (Fig. 2g). However, all the five leading eigenvectors from the R-localized observation-space background error covariance matrix are confined in a local area centered at the 120th grid point. Their Kalman gains are then calculated using $\mathbf{R} = \mathbf{I}$ as shown in Fig. 2i. The B-localized Kalman gain at the 120th grid point extends over the whole domain with larger values at the grid points close to the 120th grid point. However, the R-localized Kalman gain is confined in a local area between the 60th and 180th grid points. This local nature of the R-localized Kalman gain is also shown in the curves representing the R-localization method in Figs. 2c and 2h. As a result, all the observations would contribute to updating a single grid point in the B-localization method, although the distant observations make less contribution. This also suggests that the optimal effective localization distance defined in this study does not explicitly suggest the number of the observations that influence the analysis in the B-localization method. In contrast, the analysis at a particular grid point in the R-localization method is influenced by limited observations that are close.

4. Implementation of the R-localization method in the ETKF

The mathematical demonstration in section 3 suggests that the traditional B-localization method in Eq. (4) can be realized by expanding and modulating the background ensemble perturbations through Eq. (15). This allows the implementation of the B-/MP-localization method in the ETKF that generally employs the R-localization method. This section first briefly describes the classic R-localized ETKF. The HETKF applying the MP-localization method will be discussed in section 5.

In the ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003; Wang et al. 2004) and its local form (LETKF; Hunt et al. 2007), the background ensemble perturbations are transformed to the analysis ensemble perturbations on the ensemble perturbation subspace by a transform matrix $\mathbf{T}$:

$$\mathbf{Z}^a = \mathbf{Z}^b \mathbf{T}, \quad (24)$$

where $\mathbf{Z}^a = \mathbf{X}^a / \sqrt{K-1}$ is the analysis ensemble perturbation matrix $\mathbf{X}^a$ with a dimension of $n \times K$ normalized by a factor of $\sqrt{K-1}$. As discussed in Bishop et al. (2001), Wang and Bishop (2003) and Wang et al. (2004), the transform matrix $\mathbf{T}$ is calculated to make sure that the analysis error covariances $\mathbf{P}^a = \mathbf{Z}^a (\mathbf{Z}^a)^T$ are updated by satisfying the optimal data assimilation theory $\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b$:

$$\mathbf{T} = \mathbf{C}(\mathbf{\Gamma} + \mathbf{I})^{-1/2} \mathbf{C}^T, \quad (25)$$

and

$$\mathbf{A} = (\mathbf{H}\mathbf{Z}^b)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{Z}^b) = \mathbf{C}\mathbf{\Gamma}\mathbf{C}^T, \quad (26)$$

where each column of the matrix $\mathbf{C}$ represents an eigenvector of the matrix $\mathbf{A}$ and the diagonal matrix $\mathbf{\Gamma}$ contains the corresponding eigenvalues.

The background ensemble mean is updated by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{Z}^b [\mathbf{C}(\mathbf{\Gamma} + \mathbf{I})^{-1} \mathbf{C}^T] (\mathbf{H}\mathbf{Z}^b)^T \mathbf{R}^{-1} [\mathbf{y}^o - \overline{H(\bar{\mathbf{x}}^b)}], \quad (27)$$

where the overbar denotes the ensemble mean.

The R-localization method in the ETKF is realized by following its implementation in the LETKF of Hunt et al. (2007). The update of the model state variables is performed independently at different model grid points. At the i th grid point, instead of using the original $\mathbf{R}$, the localized observation error covariance matrix $\mathbf{R}_{\text{Rloc}}^i$ defined in Eq. (6) is applied in Eqs. (24)–(27). Different from the LETKF that selects a subset of observations, the whole set of observations are selected here to update the variables at the i th grid point in the R-localized ETKF. This “global” analysis update is designed to assure a homogeneous comparison with the HETKF detailed in section 5.

5. Implementation of the high-rank ETKF (HETKF) by applying the MP-localization method

As shown in Eq. (15), the B-localized $\mathbf{P}_{\text{Bloc}}^b$ can be achieved by an outer-product of an expanded modulated background ensemble perturbation matrix $1/(w)[\mathbf{V}_1^b, \mathbf{V}_2^b, \dots, \mathbf{V}_n^b]$ with a dimension of $n \times (nK)$. This expression makes it possible to implement the B-localization method in the ETKF. However, the computational cost is very expensive, because it requires an eigen-decomposition of a matrix with a dimension of $(nK) \times (nK)$ in Eq. (26). To reduce the computational cost, following Bishop et al. (2017), the method of selecting the leading eigenvalues and eigenvectors

of the original B-localization matrix is implemented to reduce the number of the modulation functions and thus the size of the extended modulated background ensemble.

a. Specific implementation of the MP-localization method in the HETKF

Instead of directly using the columns of the Gaussian matrix $\mathbf{G}$ that forms the B-localization matrix $\mathbf{L}$ as the modulation functions in Eq. (14), the modulation functions in Bishop et al. (2017) are calculated from the leading eigenvalues and eigenvectors of the B-localization matrix $\mathbf{L}$. As a result, the size of the modulated background ensemble perturbation matrix is significantly reduced compared to the matrix $1/(w)[\mathbf{V}_1^b, \mathbf{V}_2^b, \dots, \mathbf{V}_n^b]$ in Eq. (15). The procedures are detailed as follows:

- (i) Calculate the eigenvalues and eigenvectors of the original B-localization matrix $\mathbf{L}$ and order them correspondingly from the largest to the smallest eigenvalue. In Eq. (28), the diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues of the B-localization matrix $\mathbf{L}$ that are sorted in a descending order and the columns of the matrix $\mathbf{E}$ represent the corresponding eigenvectors:

$$\mathbf{L} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T = (\mathbf{E}\mathbf{\Lambda}^{1/2})(\mathbf{E}\mathbf{\Lambda}^{1/2})^T. \quad (28)$$

- (ii) Calculate the modulation matrix $\hat{\mathbf{G}}$ by selecting and normalizing the first M leading eigenvalues and eigenvectors to form the localization matrix $\mathbf{L}_{\text{MP}}$. In this paper, the first M leading eigenvalues and eigenvectors are selected to account for more than 99% of the sum of all the eigenvalues following Bishop et al. (2017). Mathematically,

$$\mathbf{L}_{1-M} = (\mathbf{E}\mathbf{\Lambda}^{1/2})_{1-M}(\mathbf{E}\mathbf{\Lambda}^{1/2})_{1-M}^T, \quad (29)$$

$$\begin{aligned} \hat{\mathbf{G}} &= \{[\text{diag}(\mathbf{L}_{1-M})]^{-1/2}[(\mathbf{E}\mathbf{\Lambda}^{1/2})_{1-M}]\} \\ &= [\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \dots, \hat{\mathbf{g}}_M], \end{aligned} \quad (30)$$

and

$$\mathbf{L}_{\text{MP}} = \hat{\mathbf{G}}\hat{\mathbf{G}}^T. \quad (31)$$

- (iii) Generate an expanded modulated background ensemble perturbation matrix $\hat{\mathbf{X}}^b$ with a dimension of $n \times (MK)$ by a Schur product between each raw ensemble perturbation vector and each column of the modulation matrix $\hat{\mathbf{G}}$. Mathematically,

$$\hat{\mathbf{X}}^b = \sqrt{\frac{MK-1}{K-1}}[\text{diag}(\hat{\mathbf{g}}_1)\mathbf{X}^b, \text{diag}(\hat{\mathbf{g}}_2)\mathbf{X}^b, \dots, \text{diag}(\hat{\mathbf{g}}_M)\mathbf{X}^b], \quad (32)$$

and

$$\hat{\mathbf{Z}}^b = \frac{\hat{\mathbf{X}}^b}{\sqrt{MK-1}}. \quad (33)$$

The expanded modulated background ensemble can be obtained by adding the modulated ensemble perturbations in Eq. (32) to the original background ensemble mean. As such, the modulated background ensemble shares the same background ensemble mean as the original ensemble. The factor $\sqrt{MK-1}/\sqrt{K-1}$ in Eq. (32) is to ensure that the localized ensemble background error covariance matrices with and without the modulated perturbation form are equivalent:

$$\frac{\hat{\mathbf{X}}^b(\hat{\mathbf{X}}^b)^T}{MK-1} = \frac{\mathbf{X}^b(\mathbf{X}^b)^T}{K-1} \circ \mathbf{L}_{\text{MP}} \cong \frac{\mathbf{X}^b(\mathbf{X}^b)^T}{K-1} \circ \mathbf{L}. \quad (34)$$

For the computational concern, only M modulation functions are selected and used in the implementation of the MP-localization method, in contrast to using n modulation functions as in Eq. (15). In general, M is expected to be much smaller than n . But the M modulation functions are constructed to account for more than 99% of the variances of the original B-localization matrix. As suggested in Bishop et al. (2017), it is expected to cause minimum effects on the resulting effective rank of the localized background error covariance matrix and the cycled data assimilation experiment results in section 6. More importantly, the use of fewer modulation functions in the MP-localization method can significantly improve the computational efficiency.

Figure 3 shows an example of the modulation functions in the matrix $\hat{\mathbf{G}}$ and the associated localization matrix $\mathbf{L}_{\text{MP}}$ calculated from the original B-localization matrix $\mathbf{L}$ in Fig. 2a. In this example, the first eight leading eigenvalues and corresponding eigenvectors are selected to account for more than 99% of the sum of all the eigenvalues of the original B-localization matrix $\mathbf{L}$. The resulting localization matrix $\mathbf{L}_{\text{MP}}$ (Fig. 3b) almost recovers the original B-localization matrix $\mathbf{L}$ (Fig. 2a). In particular, the size of the modulated background ensemble perturbation matrix is only increased by a factor of 8 using Eq. (32) instead of a factor of 240 using Eq. (15) in this example. This significantly improves the computational efficiency in the HETKF. For an isotropic B-localization matrix as in Fig. 2a, the associated modulation function used in

FIG. 3. Illustration of (a) the eight modulation functions (colored solid lines) in $\hat{\mathbf{G}}$ and (b) the associated localization matrix $\mathbf{L}_{\text{MP}}$ calculated from the B-localization matrix in Fig. 2a.

the MP-localization method has a larger wavenumber with an increasing eigenvalue (Fig. 3a).

b. Ensemble mean and perturbation update in the HETKF

Following section 4, Eqs. (24)–(27) are used for the HETKF ensemble mean and perturbation update. Instead of using $\mathbf{Z}^b$ with a dimension of $n \times K$, $\mathbf{Z}^b$ with a dimension of $n \times (MK)$ is applied in these equations.

During the ensemble perturbation update, directly applying $\hat{\mathbf{Z}}^b$ with a dimension of $n \times (MK)$ in Eq. (24) would produce MK analysis perturbations in the HETKF. In the practical applications, the K analysis perturbations need to be selected to initialize a K -member ensemble of background forecasts before advancing to the next DA cycle. To have a robust comparison of the B-/MP-localization and R-localization methods, following Bishop et al. (2017), two methods were implemented and examined to subselect the

analysis perturbations during the ensemble perturbation update in the HETKF.

The first perturbation subselection method is defined as deterministic to distinguish it from the second, stochastic method. In the deterministic perturbation subselection method, the first K columns of $\hat{\mathbf{Z}}^a$ are selected. To remove the modulation effect, a demodulation procedure in Eq. (35) is applied by left-multiplying each column with a diagonal matrix $[\text{diag}(\hat{\mathbf{g}}_1)]^{-1}$ associated with the first modulation function:

$$\hat{\mathbf{Z}}_{\text{MP-D}}^a = \{[\text{diag}(\hat{\mathbf{g}}_1)]^{-1}\hat{\mathbf{z}}_1^a, [\text{diag}(\hat{\mathbf{g}}_1)]^{-1}\hat{\mathbf{z}}_2^a, \dots, [\text{diag}(\hat{\mathbf{g}}_1)]^{-1}\hat{\mathbf{z}}_K^a\}, \quad (35)$$

where the letter “D” in the subscript “MP-D” stands for “deterministic”. Finally, the K -member analysis ensemble perturbation matrix $\mathbf{X}_{\text{MP-D}}^a$ is recovered by referring to the relation between $\hat{\mathbf{X}}^b$ and $\hat{\mathbf{Z}}^b$ in Eqs. (32) and (33):

$$\mathbf{X}_{\text{MP-D}}^a = \sqrt{K-1} \hat{\mathbf{Z}}_{\text{MP-D}}^a. \quad (36)$$

In Bishop et al. (2017), this deterministic perturbation subselection method was compared with a more robustly derived selection approach termed as the gain-form ETKF. It was found that when the B-localization matrix was anisotropic, the gain-form ETKF method showed more advantages in generating the analysis perturbations. However, our further examination using the Lorenz model II (see the next section for details) showed that for an isotropic B-localization matrix $\mathbf{L}$, the performances of these two perturbation selection methods were statistically indistinguishable (not shown here). So in this study we use this deterministic perturbation subselection method described in Eq. (35) for the cycled data assimilation experiments. A similar subselection procedure was adopted in Kretschmer et al. (2015) to select the K analysis ensemble perturbations updated from a background ensemble formed by a mixture of the flow-dependent and climatological perturbations. This HETKF implementation and the classic R-localized ETKF described in section 4 are hereafter referred to as “MP-D” and “R-D”.

The second perturbation subselection method, defined as stochastic, is based on the idea of updating each member with different sets of perturbed observations (Houtekamer and Mitchell 1998). Specifically, the K sets of perturbed observations are generated and assimilated to update the raw ensemble members in the HETKF. This HETKF implementation is denoted as “MP-S” in this study, where the letter “S” stands for stochastic. To have a homogeneous comparison of the B-/MP-localization and R-localization methods, the same perturbed observation approach is also applied for the R-localized ETKF, which is denoted as “R-S” hereafter. This stochastic approach avoids the analysis perturbation subselection issue in the HETKF. It deviates from the idea of updating the perturbations through the transform. Nevertheless, it provides an additional avenue to further reveal if the differences between the B-localization and R-localization methods for the analysis update will be dependent upon the perturbation subselection methods.

6. Experiments with the Lorenz model II

a. Lorenz model II

In this section, the Lorenz model II (Lorenz 2005; Rainwater and Hunt 2013; Fairbairn et al. 2014; Bishop et al. 2015) was used for the cycled data assimilation experiments to compare the performances of

the HETKF and R-localized ETKF. Compared to the Lorenz Model I (also known as the Lorenz-96 model; Lorenz 1996), the Lorenz Model II produces spatially smoothed model trajectory. The equation for the Lorenz Model II is given by

$$\frac{dX_n}{dt} = [X, X]_{K,n} - X_n + F, \quad (37)$$

where

$$[X, X]_{K,n} = \sum_{j=-J}^J \sum_{i=-J}^J '(-X_{n-2K-i} X_{n-K-j} + X_{n-K+j-i} X_{n+K+j}) / K^2. \quad (38)$$

To be noted first, the usage of the symbols and letters defined in Eqs. (37) and (38) is restricted within this subsection for illustration purposes. They are not associated with the previous sections. A total of N variables are evenly distributed on a latitude cycle. Each variable X is indexed by n ($n = 0, 1, 2, \dots, N-1$). The periodic boundary condition is applied. F is the forcing term. The smoothing parameter K , chosen much smaller than N , is used to define $J = (K-1)/2$ if K is odd and $J = K/2$ if K is even. The modified summation sign Σ' functions similarly as the regular summation sign Σ except that the first and last terms are multiplied by a factor of 0.5. In Eq. (38), Σ' is used if K is even, otherwise, Σ' is replaced by Σ if K is odd. Following the suggestions in Lorenz (2005), the parameters of the Lorenz Model II are set as $N = 240$, $F = 15$ and $K = 8$ in our experiments. The model is integrated using the fourth-order Runge–Kutta scheme. A nondimensional time step is chosen to be 0.025 (which is equivalent to about 18 min in the real atmosphere).

b. Experiment design

First, the Lorenz model II was continuously integrated for a total of 80 000 time steps by selecting 240 random numbers as the initial condition. The model trajectories between the time steps of 15 001 and 30 000 serve as the simulated model climatology. The model trajectories over the last 50 000 steps are treated as the “truth” for verification. The initial ensemble is randomly drawn from the simulated model climatology. The observations are assimilated every five time steps. There are 10 000 data assimilation cycles in total. At each analysis time, the integral observations mimicking the satellite radiances are generated by first averaging the “true” state variables $\mathbf{x}^t = [x_1^t, x_2^t, \dots, x_n^t]^T$ over its adjacent 21 grid points and then adding a random noise ε drawn from a Gaussian distribution $N(0, r^2 = 1.32)$. The observation standard deviation r is 20% that of

FIG. 4. An example of the 6-member background ensemble (black), the background ensemble mean (blue), the simulated nonperturbed observations (red), and the “true” state (green) in the first data assimilation cycle.

the simulated model climatology (Wang et al. 2007). For example, the simulated integral observation y_i^o at the i th grid point is calculated by

$$y_i^o = \frac{1}{21} \sum_{j=i-10}^{j=i+10} x_j^t + \varepsilon. \quad (39)$$

In the experiments with the stochastic perturbation subselection method (“MP-S” and “R-S”), the K sets of perturbed observations are further generated by adding random noises drawn from the same Gaussian distribution $N(0, r^2 = 1.32)$ to the simulated observations. Figure 4 shows an example of the “true” state, nonperturbed observations and background ensemble in the first data assimilation cycle.

Furthermore, two sets of experiments were designed to more thoroughly evaluate their performances. The first set, termed as K6PX ($X = 30, 60, 120$, and 240), runs a 6-member ensemble ($K = 6$) but assimilates an increasing number of observations ($p = 30, 60, 120$, and 240 correspondingly). The other set, termed as KYP240 ($Y = 3, 6$, and 9), assimilates a total of 240 observations ($p = 240$) but runs ensembles with an increasing size ($K = 3, 6$, and 9). In both sets of experiments, a range of localization and inflation factors are tuned for the cycled DA experiments. Specifically, the degree of localization is determined by the parameter d in Eq. (9). A larger d results in stronger localization. The inflation is realized by multiplying the analysis perturbations with a factor larger than 1.0 before continuing to the next DA cycle. The root-mean-square error (RMSE) between the analysis and the “truth” is calculated and averaged from the last 8000 cycles to quantify the analysis accuracy. The percentage of the RMSE reduction (PRR) of the MP-localization method over the R-localization method is further defined as

$$\text{PRR} = \frac{\text{RMSE}(\text{R-loc}) - \text{RMSE}(\text{MP-loc})}{\text{RMSE}(\text{R-loc})} \times 100\%. \quad (40)$$

To eliminate the random seed effects, all the experiments were repeated with eight trials by selecting different sets of random seeds (Bishop et al. 2015; Janjić et al. 2011).

c. Experiment results

1) SENSITIVITY OF THE FOUR FILTERS TO LOCALIZATION AND INFLATION FACTORS

To obtain the minimum analysis error, extensive tuning tests were performed for each of the eight trials in each filter by combining different sets of localization and inflation factors. Figure 5 shows the analysis RMSE of “R-D,” “MP-D,” “R-S,” and “MP-S” in K6P240 as a function of the localization and inflation factors. RMSE in each filter in Fig. 5 is averaged from all the eight trials. With the optimal localization and inflation factors (denoted by the red asterisk), the MP-localization method outperforms the R-localization method in both the deterministic and stochastic perturbation subselection methods. In addition, compared to the R-localization method, the MP-localization method shows less sensitivity to the localization and inflation factors. This feature is characterized by the broader blue areas in the MP-localization method in Figs. 5b and 5d. For a given localization method, compared to the stochastic perturbation subselection method, the deterministic perturbation subselection method achieves smaller minimum analysis error with less localization and inflation. The less accurate analysis in the stochastic perturbation subselection method could owe to the additional sampling errors from perturbed observations (Whitaker and Hamill 2002). Overall, the MP-localization method using the deterministic

FIG. 5. Analysis RMSE averaged over all the eight trials in the experiment of K6P240 as a function of the localization factors on the vertical axis and the inflation factors on the horizontal axis for (a) “R-D,” (b) “MP-D,” (c) “R-S,” and (d) “MP-S.” Note different color scales are applied in (a),(b) and (c),(d). Red asterisk indicates the optimal combination of the localization and inflation factors that gives the minimum analysis RMSE in each filter.

perturbation subselection method (“MP-D”) shows the most accurate analysis in K6P240.

2) FILTER PERFORMANCE AS A FUNCTION OF THE OBSERVATION NUMBER

In this subsection, the K6PX ($X = 30, 60, 120$, and 240) experiments are examined. The top panel in Fig. 6 shows the minimum analysis RMSE calculated from the optimal combination of the localization and inflation factors for each of the eight trials of “R-D,” “MP-D,” “R-S,” and “MP-S.” For both the deterministic and stochastic perturbation subselection methods, the MP-localization method significantly outperforms the R-localization method in all the four experiments. In general, the percentage of the RMSE reduction of the MP-localization method over the R-localization method tends to be slightly reduced with an increasing number of the observations (Fig. 8a). This is likely due to the overall improved analysis through the cycled assimilation of a larger number of observations. In the noncycled experiments, the percentage of the RMSE reduction of

the MP-localization method over the R-localization method increases with an increasing number of the observations (not shown). The latter result is more consistent with the expectation that the superiority associated with the higher rank in the estimated background error covariances becomes more pronounced in the assimilation of a larger number of the observations. For a given localization method, the deterministic perturbation subselection method shows smaller minimum analysis error compared to the stochastic perturbation subselection method.

Following section 3, the matrix $1/(w^2) \sum_{j=1}^n \mathbf{g}_j \mathbf{g}_j^T$ in Eq. (22) for “MP-D” and “MP-S”, and the matrix $\mathbf{g}_i \mathbf{g}_i^T$ in Eq. (23) for “R-D” and “R-S” are calculated at the 120th grid point to reveal their effective localization distances (shown in the bottom panel of Fig. 6). Here, these two matrices are calculated by applying the optimal localization factors. In both the deterministic and stochastic perturbation subselection methods, the optimal effective localization distance in the MP-localization method is tighter than that in the R-localization method.

FIG. 6. (a)–(d) Minimum analysis RMSE with the optimal combination of localization and inflation factors for “R-D” (blue dashed), “MP-D” (blue solid), “R-S” (red dashed), and “MP-S” (red solid), over the eight trials labeled on the horizontal axis in the experiments of (a) K6P30, (b) K6P60, (c) K6P120, and (d) K6P240. (e)–(h) Localization coefficients corresponding to the 120th row of the matrix $1/(w^2) \sum_{j=1}^n \mathbf{g}_j \mathbf{g}_j^T$ as shown in Eq. (22) for “MP-D” (blue solid) and “MP-S” (red solid) and of the matrix $\mathbf{g}_j \mathbf{g}_j^T$ for as shown in Eq. (23) for “R-D” (blue dashed) and “R-S” (red dashed) by applying the optimal tuned localization factor d defined in Eq. (9) and listed at the top of the figures of (e) K6P30, (f) K6P60, (g) K6P120, and (h) K6P240.

But this does not mean fewer observations will influence the analysis in the B-/MP-localization method than that in the R-localization method as discussed in section 3. For a given localization method, the stochastic perturbation subselection method requires stronger localization than the deterministic perturbation subselection method to obtain the minimum analysis error. As discussed in the last subsection, this could be caused by the additional sampling errors from the perturbed observations in the stochastic perturbation subselection method (Whitaker and Hamill 2002). The bottom panel of Fig. 6 shows that the optimal effective localization distance in each of “R-D,” “MP-D,” “R-S,” and “MP-S” is not very sensitive to an increasing number of the observations. This result may be associated with the assimilation of the integral observations in our cycled data assimilation experiments. As shown in Eq. (39), the integral observations that are close could contain highly correlated information. Increasing the observation densities may not efficiently add additional degrees of freedom of the information provided by the integral observations. The effective localization distance is therefore likely less sensitive to an increasing number of the integral observations as designed in our cycled data assimilation experiments.

3) FILTER PERFORMANCE AS A FUNCTION OF THE ENSEMBLE SIZE

Figure 7 shows the results for the KYP240 ($Y = 3$ and 9) experiments. When the ensemble size is reduced

from 6 to 3, the relative performance of the MP-localization and R-localization methods (Fig. 7a) is similar to that of K6P240. Specifically, in both the deterministic and stochastic perturbation subselection methods, the MP-localization method significantly outperforms the R-localization method. For a given localization method, the deterministic perturbation subselection method shows smaller minimum analysis error than the stochastic perturbation subselection method. In addition, in K3P240, the MP-localization method using the stochastic perturbation subselection method (“MP-S”) even shows more accurate analysis than the R-localization method using the deterministic perturbation subselection method (“R-D”).

By increasing the ensemble size to 9 (K9P240), for both perturbation subselection methods, the MP-localization and R-localization methods perform comparably and their difference is statistically insignificant in most of the eight trials. Consistently, the percentage of the RMSE reduction of the MP-localization method over the R-localization method is reduced with an increasing ensemble size in both perturbation subselection methods (Fig. 8b). This is within the expectation that the higher rank from the B-/MP-localization method would contribute more positively to alleviating the rank deficiency issue and thus improving the analysis for a small ensemble. This may further suggest that the improved analysis in the B-/MP-localization method is likely associated with its higher rank as demonstrated in section 3.

FIG. 7. As in Fig. 6, but for the experiments of (a),(c) K3P240 and (b),(d) K9P240.

The effective localization distances calculated from the optimal localization factors are shown for the KYP240 experiments in the bottom panel of Fig. 7. For a given perturbation subselection method, the optimal effective localization distance in the MP-localization method is tighter than that of the R-localization method. For both localization methods, the optimal effective

localization distance from the stochastic perturbation subselection method is tighter than the deterministic perturbation subselection method except for the K3P240 experiment with the R-localization method. Figure 7 also demonstrates as expected that the effective localization distances in each of “R-D,” “MP-D,” “R-S,” and “MP-S” becomes wider with a larger ensemble size.

FIG. 8. Percentage of the RMSE reduction of the MP-localization method over the R-localization method in the deterministic (blue) and stochastic (red) perturbation subselection methods in the experiments of (a) K6PX and (b) KYP240.

7. Conclusions and discussion

A mathematical demonstration is first provided to compare the B-localization and R-localization methods. It is shown that when the same effective localization function is applied, the B-localization method achieves a higher rank than the R-localization method in the localized background error covariance matrix. The mathematical demonstration is further illustrated and validated using a simple example. Further examination suggests that all the observations will contribute to updating a single grid point in the B-localization method. However, the analysis at a particular grid point in the R-localization method is influenced by limited observations that are close. Meanwhile, the mathematical demonstration also shows that the B-localization method can be realized through extending and modulating the raw background ensemble perturbations or the MP-localization method. Specifically, in the MP-localization method, each raw ensemble perturbation vector is modulated through an element-wise multiplication with each of the modulation functions. To improve the computational efficiency, the modulation functions are calculated from the leading eigenvalues and eigenvectors of the original B-localization matrix. The resulting MP-localized background error covariance matrix is thus consistent with that applying the traditional B-localization method. In the mathematical demonstration, it proves that the R-localization method can also be expressed in the form of the modulated ensemble perturbations as in the B-localization method. The B-/MP-localization method is then implemented in the ETKF and further compared with the R-localization method using the same ETKF algorithm. Because of the higher rank from the B-localization method as derived in the mathematical demonstration, the B-/MP-localized ETKF is termed as the high-rank ETKF (HETKF) to distinguish it from the classic R-localized ETKF.

Extensive cycled data assimilation experiments were conducted to compare the performances of the HETKF and R-localized ETKF using the Lorenz model II. Using the same ETKF algorithm warrants a homogeneous comparison between these two localization methods, so that it is more straightforward to relate their resulting analysis performances with the localization differences. The results show that the HETKF significantly and consistently improves the analysis over the R-localized ETKF especially for a small ensemble. Since the higher rank from the HETKF is expected to contribute more positively to mitigating the rank deficiency issue for a small ensemble, the improved analysis of the HETKF over the R-localized ETKF is likely associated with the higher rank from the B-/MP-localization method.

In addition, the advantage of the HETKF over the R-localized ETKF tends to be slightly reduced with the increasing number of the observations. This result could be attributed to the improved accuracy of the system through the cycled assimilation of a larger number of observations. Furthermore, the HETKF is less sensitive to the localization length scales and inflation factors than the R-localized ETKF. In all the experiments, the HETKF shows tighter optimal effective localization distance than the R-localized ETKF. The above conclusion of comparing the HETKF and R-localized ETKF does not rely on the perturbation subselection methods in the HETKF.

It is also found that in both the HETKF and R-localized ETKF, the stochastic perturbation subselection method shows larger analysis error than the deterministic perturbation subselection method. In addition, in both filters, the stochastic perturbation subselection method generally requires stronger localization and larger inflation than the deterministic perturbation subselection method to obtain the minimum analysis error except for the experiment with the R-localized ETKF and a very small ensemble (e.g., K3P240). This can be attributed to the sampling errors by perturbing the observations in the stochastic perturbation subselection method (Whitaker and Hamill 2002).

In this study, the improved analysis from the B-/MP-localization method over the R-localization method is demonstrated using the same ETKF algorithm in the Lorenz model II. This is consistent with the results in Janjić et al. (2011) and Nerger et al. (2012) that adopted different EnKF variants for comparison. To implement the HETKF in the real modeling systems, additional treatments are likely needed for the computational concerns. For example, a parallel, patch-based implementation like the LETKF can be adopted to improve the computational scalability. Further diagnostics (not shown) showing the analysis errors calculated from the full 10 000 cycles (i.e., including those cycles before the errors get stabilized) indicate that the HETKF requires less time to converge compared to the R-localized ETKF. This feature and the less sensitivity of the HETKF to the localization length scales and inflation factors are attractive for the real model applications.

Acknowledgments. This study is primarily supported by ONR Grant N00014-18-1-2666 and NOAA Grant NA15NWS4680022. Craig Bishop would like to acknowledge the support from the NRL base program PE0601153N. Computational resources provided by the OU Supercomputing Center for Education and Research at the University of Oklahoma were used for this study.

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2).
- , 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99–111, <https://doi.org/10.1016/j.physd.2006.02.011>.
- , and L. Lei, 2013: Empirical localization of observation impact in ensemble Kalman filters. *Mon. Wea. Rev.*, **141**, 4140–4153, <https://doi.org/10.1175/MWR-D-12-00330.1>.
- Bishop, C. H., and D. Hodyss, 2007: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 2029–2044, <https://doi.org/10.1002/qj.169>.
- , 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](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2).
- , B. Huang, and X. Wang, 2015: A nonvariational consistent hybrid ensemble filter. *Mon. Wea. Rev.*, **143**, 5073–5090, <https://doi.org/10.1175/MWR-D-14-00391.1>.
- , J. S. Whitaker, and L. Lei, 2017: Gain form of the ensemble transform Kalman filter and its relevance to satellite data assimilation with model space ensemble covariance localization. *Mon. Wea. Rev.*, **145**, 4575–4592, <https://doi.org/10.1175/MWR-D-17-0102.1>.
- Buehner, M., and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 615–630, <https://doi.org/10.1002/qj.50>.
- 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>.
- Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone, 2014: A comparison of 4DVar with ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.*, **140**, 281–294, <https://doi.org/10.1002/qj.2135>.
- Gasparoni, N. A., and X. Wang, 2015: Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. *Mon. Wea. Rev.*, **143**, 1981–2000, <https://doi.org/10.1175/MWR-D-14-00272.1>.
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522, <https://doi.org/10.1175/2010MWR3328.1>.
- Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.
- 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](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](https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2).
- , and —, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289, <https://doi.org/10.1256/qj.05.135>.
- , and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, <https://doi.org/10.1175/MWR-D-15-0440.1>.
- Huang, B., and X. Wang, 2018: On the use of cost-effective valid-time-shifting (VTS) method to increase ensemble size in the GFS hybrid 4DEnVar system. *Mon. Wea. Rev.*, **146**, 2973–2998, <https://doi.org/10.1175/MWR-D-18-0009.1>.
- Hunt, B. R., E. J. 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>.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan*, **75**, 181–189, https://doi.org/10.2151/jmsj1965.75.1B_181.
- Janjić, T., L. Nerger, A. Albertella, J. Schröter, and S. Skachko, 2011: On domain localization in ensemble-based Kalman filter algorithms. *Mon. Wea. Rev.*, **139**, 2046–2060, <https://doi.org/10.1175/2011MWR3552.1>.
- Kalman, R. E., and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic Eng.*, **83**, 95–108, <https://doi.org/10.1115/1.3658902>.
- Kretschmer, M., B. R. Hunt, and E. Ott, 2015: Data assimilation using a climatologically augmented local ensemble transform Kalman filter. *Tellus*, **67A**, 26617, <https://doi.org/10.3402/tellusa.v67.26617>.
- Kuhl, D., and Coauthors, 2007: Assessing predictability with a local ensemble Kalman filter. *J. Atmos. Sci.*, **64**, 1116–1140, <https://doi.org/10.1175/JAS3885.1>.
- Lei, L., and J. S. Whitaker, 2017: Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *J. Adv. Model. Earth Syst.*, **9**, 781–789, <https://doi.org/10.1002/2016MS000864>.
- Lorenc, A. C., 2017: Improving ensemble covariances in hybrid variational data assimilation without increasing ensemble size. *Quart. J. Roy. Meteor. Soc.*, **143**, 1062–1072, <https://doi.org/10.1002/qj.2990>.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Shinfield Park, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 1–18.
- , 2005: Designing chaotic models. *J. Atmos. Sci.*, **62**, 1574–1587, <https://doi.org/10.1175/JAS3430.1>.
- Miyoshi, T., and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135**, 3841–3861, <https://doi.org/10.1175/2007MWR1873.1>.
- , K. Kondo, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophys. Res. Lett.*, **41**, 5264–5271, <https://doi.org/10.1002/2014GL060863>.
- Nerger, L., T. Janjić, J. Schröter, and W. Hiller, 2012: A regulated localization scheme for ensemble-based Kalman filters. *Quart. J. Roy. Meteor. Soc.*, **138**, 802–812, <https://doi.org/10.1002/qj.945>.
- Oczkowski, M., I. Szunyogh, and D. J. Patil, 2005: Mechanisms for the development of locally low-dimensional atmospheric dynamics. *J. Atmos. Sci.*, **62**, 1135–1156, <https://doi.org/10.1175/JAS3403.1>.

- Patil, D. J., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878–5881, <https://doi.org/10.1103/PhysRevLett.86.5878>.
- Rainwater, S., and B. Hunt, 2013: Mixed-resolution ensemble data assimilation. *Mon. Wea. Rev.*, **141**, 3007–3021, <https://doi.org/10.1175/MWR-D-12-00234.1>.
- Sakov, P., and L. Bertino, 2011: Relation between two common localisation methods for the EnKF. *Comput. Geosci.*, **15**, 225–237, <https://doi.org/10.1007/s10596-010-9202-6>.
- 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](https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2).
- , —, 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:WIBAE0>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1590:WIBAE0>2.0.CO;2).
- , T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A Comparison of Hybrid Ensemble Transform Kalman Filter–Optimum Interpolation and Ensemble Square Root Filter Analysis Schemes. *Mon. Wea. Rev.*, **135**, 1055–1076, <https://doi.org/10.1175/MWR3307.1>.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- Xu, Q., L. Wei, H. Lu, C. Qiu, and Q. Zhao, 2008: Time-expanded sampling for ensemble-based filters: Assimilation experiments with a shallow-water equation model. *J. Geophys. Res.*, **113**, 1–12, <https://doi.org/10.1029/2007JG000450>.