A Local Unscented Transform Kalman Filter for Nonlinear Systems

KWANGJAE SUNG

Development Division, Korea Institute of Atmospheric Prediction Systems, Seoul, South Korea

HYO-JONG SONG

Department of Environmental Engineering and Energy, Myongji University, Yongin-si, South Korea

IN-HYUK KWON

Data Assimilation Team, Korea Institute of Atmospheric Prediction Systems, Seoul, South Korea

(Manuscript received 8 July 2019, in final form 18 May 2020)

ABSTRACT

This paper proposes an efficient data assimilation approach based on the sigma-point Kalman filter (SPKF). With a potential for nonlinear filtering applications, the proposed approach, designated as the local unscented transform Kalman filter (LUTKF), is similar to the SPKF in that the mean and covariance of the nonlinear system are estimated by propagating a set of sigma points—also referred to as ensemble members—generated using the scaled unscented transformation (SUT), while making no assumptions with regard to nonlinear models. However, unlike the SPKF, the LUTKF can reduce the influence of observations on distant state variables by employing a localization scheme to suppress spurious correlations between distant locations in the error covariance matrix. Moreover, while the SPKF uses the augmented state vector constructed by concatenating the model states, model noise, and measurement noise, the system state for the LUTKF is not augmented with the random noise variables, thereby providing an accurate state estimate with relatively few sigma points. In sensitivity experiments executed with a 40-variable Lorenz system, the LUTKF required only three sigma points to prevent filter divergence for linear/nonlinear measurement models. Comparisons of the LUTKF and the local ensemble transform Kalman filters (LETKFs) reveal the advantages of the proposed filter in situations that share common features with geophysical data assimilation applications. In particular, the LUTKF shows considerable benefits over LETKFs when assimilating densely spaced observations that are related nonlinearly to the model state and that have high noise levels—such as the assimilation of remotely sensed data from satellites and radars.

1. Introduction

One of the most widely used Bayesian filters for assimilating atmospheric and oceanic data is the ensemble Kalman filter (EnKF; Evensen 1992, 1994; Houtekamer and Mitchell 1998; Evensen and van Leeuwen 2000). Several variations of the EnKF have been introduced in the field of Earth science, such as the local ensemble Kalman filter (LEKF; Ott et al. 2004), the ensemble transform Kalman filter (ETKF; Bishop et al. 2001), and the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007).

For a nonlinear model, the EnKF methods can propagate the state estimate and its covariance in time

through fully nonlinear model equations without the linearization (e.g., calculation of the Jacobian or Hessian) of the nonlinear model as in the EKF (extended Kalman filter). However, the EnKF approaches assume a linear observation operator to obtain the standard Kalman gain and error covariance; to be more precise, they take into consideration only up to the first-order Taylor approximation of the nonlinear observation operator, instead of linearizing the nonlinear observation operator on the entire model space as in Eqs. (6.11), (6.14), and (6.15) of Hamill (2006). For this reason, it cannot provide an accurate estimate of up to the second-order moment in relation to the true states. Thus, if the EnKF algorithms are executed with highly nonlinear observation functions, they can produce large errors when estimating the states of the nonlinear observation

Corresponding author: Hyo-Jong Song, hjsong@mju.ac.kr

DOI: 10.1175/MWR-D-19-0228.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

[e.g., satellite altimetry data assimilation, where observations have a strong nonlinear relationship with the model state variables (Hamon et al. 2019)].

The choice of the ensemble size has a significant impact on the analysis performance of the EnKF algorithm. To determine which linear combination of ensemble members approaches the true state of the system through the analysis, the ensemble size of the EnKF should be large enough to approximately span, as fully as possible, the state space, although it can be reasonably small compared to Monte Carlo approaches. However, the use of excessive ensemble members may not be suitable for high-dimensional models, such as atmospheric and oceanic models, because it requires substantial computational resources. While the hybridization of background error covariance can be used for reducing the sampling noise with the aid of the static background error covariance, this study focuses on the identification of the worth of the new ensemble DA scheme. The background modeling method is set to be pure ensemble-oriented for the discussion to be simplified.

Recently, the sigma-point Kalman filter (SPKF), a new ensemble-based data assimilation method using the scaled unscented transformation (SUT) in nonlinear systems, has been proposed to overcome the aforementioned drawbacks associated with the EnKF approaches in the field of Earth science (Ambadan and Tang 2009; Luo and Moroz 2009; Tang et al. 2014). The most important difference between the SPKF and EnKF methods described above is in the generation of ensemble members and in the determination of the ensemble size. Unlike the EnKF methods, the sigma points in the SPKF are deterministically selected using the SUT during every data assimilation cycle. For an L -dimensional system, a set of $2L + 1$ sigma points in the SPKF is used to propagate the state estimate and its error covariance (uncertainty) through nonlinear model equations. Unlike the EnKF algorithms, the SPKF makes no assumptions regarding the linearized observation operator; therefore, it can estimate the error statistics of the model state through a fully nonlinear measurement operator by reinterpreting the calculation for the standard Kalman gain and analysis covariance (Ambadan and Tang 2009; Tang et al. 2014).

Despite this advantage offered by the SPKF over the EnKF methods, the SPKF approach is associated with some practical problems. The SPKF algorithm that uses $2L + 1$ sigma points for an L -dimensional system becomes prohibitively computationally expensive when dealing with a high-dimensional system due to the computational costs associated with more dimensions. Moreover, the SPKF approach uses the augmented state

vector that is constructed by concatenating the model states, model noise, and measurement noise. The augmented state dimension is the sum of the model state, model error, and measurement error dimensions. Hence, in the SPKF approach, it can be extremely difficult to assimilate a large number of observations for high-dimensional systems.

To solve this problem, some studies on SPKF have been made. One approach is to reduce sigma points so that only the most important sigma points are used. For example, the simplex SPKF (Julier and Uhlmann 2002) uses $L + 1$ sigma points that affect the evolution of error covariance in a nonlinear model with a large dimensionality. Another approach is the principal component analysis (PCA), which presents the principal component space with a fewer number of sigma-points while still having the main features of the full sigma-point space (Ambadan and Tang 2009). Although this scheme reduces the number of sigma points in full sigma-point space, it can approximate the error covariance with the same estimation accuracy as the original SPKF using sigma points in principal component space. An alternative solution to the same problem is to use the truncated singular value decomposition (TSVD) for the SPKF (Luo and Moroz 2009; Tang et al. 2014). Because the number of the sigma points decreases by truncating the analysis error covariance matrix, this approach can reduce the computational cost of the SPKF efficiently. However, for higher-dimensional systems (such as global atmospheric and oceanic models), the reduced SPKF schemes mentioned above are still computationally intractable.

The other potential solution to the problem with the SPKF is the localization, which is important for applications to high-dimensional systems in the EnKF algorithms. The localization scheme allows the analysis to be performed more efficiently for each model grid point with a moderate ensemble size as in low-dimensional models. However, since the SPKF uses the augmented state vector constructed by concatenating model state variables, model noise, and measurement noise for all model grid points, the model state estimate of the SPKF is executed for the global domain. Therefore, the localization approach in the SPKF may not be used to efficiently execute the data assimilation.

In this paper, we propose an efficient scheme for implementing the SPKF algorithm, called a local unscented transform Kalman filter (LUTKF). Unlike the SPKF, the LUTKF uses the localization scheme, which is the same as that used in the LETKF; that is, it executes the local data assimilation at each model grid point. Furthermore, the LUTKF does not use the augmented state vector; that is, it uses the state vector that consists

of only the state variables. This makes it easier to implement the local analysis in the LUTKF and enables the LUTKF to be more computationally efficient for nonlinear systems. In this study, the results from cycling data assimilation experiments executed through the 40-variable [Lorenz \(1996\)](#) model show that the LUTKF has great potential in performing efficient data assimilation with a small and reasonable ensemble size for nonlinear models. The performance benefits of the LUTKF were evaluated using the existing ensemble Kalman filters as benchmarks, including the SPKF (known as a global filter) and the LETKF (known as a local filter). As future work, we will explore the potential for high-dimensional models, such as atmospheric and oceanic models.

The rest of this paper is organized as follows: [section 2](#) presents the SPKF in its most basic form; [section 3](#) introduces the LUTKF algorithm, which incorporates the nonaugmented SPKF algorithm and the localization approach; [section 4](#) presents experiments to verify the validity of the LUTKF algorithm as an assimilation method; and finally, [section 5](#) summarizes results and discusses the potential of the LUTKF approach for high-dimensional problems.

2. Sigma-point Kalman filtering

The basic framework for the SPKF algorithm involves estimating the state of a nonlinear system by the dynamical model and the observation model with nonadditive noise as follows:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{w}_{k-1}), \quad (1)$$

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k), \quad (2)$$

where $\mathbf{x}_k$ denotes the system state vector at time k , $f(\cdot)$ is a nonlinear function of the state, $\mathbf{z}_k$ is the measurement vector, and $h(\cdot)$ is the nonlinear function that describes the relationship between the measurement and state. The process noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are assumed to be white noise and uncorrelated with the zero-mean and covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively.¹

In the SPKF, the process noise and measurement noise are treated as specific states to be estimated;

however, they are not actually estimated using the filter. Hence, the state vector for the SPKF is redefined as the concatenation of the original state, process noise, and measurement noise ([Ambadan and Tang 2009](#); [Tang et al. 2014](#)). The augmented state vector $\mathbf{x}_k^a$ and corresponding covariance $\mathbf{P}_k^a$ are given by

$$\mathbf{x}_k^a = [\mathbf{x}_k^T \quad \mathbf{w}_k^T \quad \mathbf{v}_k^T]^T, \quad (3)$$

$$\mathbf{P}_k^a = \begin{bmatrix} \mathbf{P}_k & 0 & 0 \\ 0 & \mathbf{Q}_k & 0 \\ 0 & 0 & \mathbf{R}_k \end{bmatrix}^T, \quad (4)$$

where $\mathbf{P}_k$ is the estimation error covariance of the original model state $\mathbf{x}_k$. The dimension of the augmented state vector is as follows:

$$L^a = L_x + L_w + L_v, \quad (5)$$

where L_x is the original state dimension, L_w is the process noise dimension, and L_v is the measurement noise dimension (i.e., $\mathbf{x}_k^a \in \mathbb{R}^{L_x + L_w + L_v}$).

The SPKF algorithm can estimate the statistical moments of the nonlinear model using the SUT ([Julier and Uhlmann 2004](#)). In the SPKF, the mean and covariance of the model state are calculated by propagating the deterministically chosen sigma points, also referred to as ensemble members, through the nonlinear model. The data assimilation procedure of the SPKF can be divided into two phases: sigma point selection and state estimation.

a. Sigma point selection

During this phase, $2L^a + 1$ sigma points are deterministically chosen using the SUT to enable them to capture the statistical properties of the nonlinear model ([Julier et al. 2000](#); [Ito and Xiong 2000](#); [Wan and van der Merwe 2000](#)). Thus,

$$\mathbf{x}_{k-1|k-1}^0 = \hat{\mathbf{x}}_{k-1|k-1}^a, \quad (6)$$

$$\mathbf{x}_{k-1|k-1}^i = \hat{\mathbf{x}}_{k-1|k-1}^a + \left[\sqrt{(L^a + \lambda) \mathbf{P}_{k-1|k-1}^a} \right]_i \quad \text{for} \\ i = 1, \dots, L^a, \quad \text{and} \quad (7)$$

$$\mathbf{x}_{k-1|k-1}^i = \hat{\mathbf{x}}_{k-1|k-1}^a - \left[\sqrt{(L^a + \lambda) \mathbf{P}_{k-1|k-1}^a} \right]_i \quad \text{for} \\ i = L^a + 1, \dots, 2L^a, \quad (8)$$

where the subscript $k-1|k-1$ means the posterior or analysis state at time $k-1$, $\hat{\mathbf{x}}_{k-1|k-1}^a$ is the mean of the augmented state, λ is a scaling parameter, and $\left[\sqrt{(L^a + \lambda) \mathbf{P}_{k-1|k-1}^a} \right]_i$ is the i th column (or row) vector of the matrix square root of the weighted covariance

¹ Since there is no general approach for determining the noise covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$ of the dynamical and observation models, these covariances, known as parameters of the Kalman filter ([Simon 2006](#)), are often obtained by sensitivity experiments or Monte Carlo methods ([Miller et al. 1994](#)) and by the estimation methods suggested by [Evensen \(2003\)](#).

matrix $(L^a + \lambda)\mathbf{P}_{k-1|k-1}^a$. For numerical efficiency, the Cholesky factorization is generally used to calculate the matrix square root.

The weight associated with the i th sigma-point is defined as

$$\omega_{(m)}^0 = \frac{\lambda}{L^a + \lambda}, \quad (9)$$

$$\omega_{(c)}^0 = \frac{\lambda}{L^a + \lambda} + (1 - \alpha^2 + \beta), \quad \text{and} \quad (10)$$

$$\omega_{(m)}^i = \omega_{(c)}^i = \frac{1}{2(L^a + \lambda)} \quad \text{for} \quad i = L^a + 1, \dots, 2L^a, \quad (11)$$

where $\omega_{(m)}^i$ and $\omega_{(c)}^i$ represent the weighting terms used to calculate the mean and covariance, respectively. The scaling parameter λ is determined by $\lambda = \alpha^2(L^a + \kappa) - L^a$. The value of α is a parameter that controls the size (spread) of the sigma-point distribution around the mean $\hat{\mathbf{x}}_{k-1|k-1}^a$ and should be set to a small positive value ($0 \leq \alpha \leq 1$). The scaling parameter κ is a nonnegative value ($\kappa \geq 0$) that guarantees the positive semidefiniteness of the covariance matrix, but it is not critical; therefore, the choice of $\kappa = 0$ is good as a default value. Another parameter β can be used to integrate the knowledge of the higher-order moments of the state distribution and is usually set to a nonnegative value ($\beta \geq 0$).²

b. Mean and covariance estimation

The sigma points are propagated through the fully nonlinear system [Eqs. (1) and (2)], while making no assumptions regarding the linearity of prediction and observation models:

$$\mathbf{x}_{(x),k|k-1}^i = f[\mathbf{x}_{(x),k-1|k-1}^i, \mathbf{x}_{(w),k-1|k-1}^i] \quad \text{for} \quad i = 0, \dots, 2L^a, \quad (12)$$

$$\mathbf{z}_k^i = h[\mathbf{x}_{(x),k|k-1}^i, \mathbf{x}_{(v),k-1|k-1}^i] \quad \text{for} \quad i = 0, \dots, 2L^a, \quad (13)$$

where $\mathbf{x}_{(x),k|k-1}^i$ is the prior or forecast sigma point state vector at time k , $\mathbf{x}_{(w),k-1|k-1}^i$ and $\mathbf{x}_{(v),k-1|k-1}^i$ are sigma point vectors that correspond to the model error and observation error, respectively, and $\mathbf{z}_k^i$ represents the transformed sigma point in observation space.

² The parameter β can be employed to control the error in the kurtosis that has an impact on the heaviness of the tails of the posterior distribution (Wan and van der Merwe 2000; van der Merwe 2004). If the state distribution of the dynamic model is Gaussian, then the optimal choice of β is 2.

Using the transformed sigma points obtained from Eqs. (12) and (13), the predicted state estimate (background mean) $\hat{\mathbf{x}}_{k|k-1}$, corresponding covariance $\mathbf{P}_{k|k-1}$, predicted measurement estimate $\hat{\mathbf{z}}_k$, and its error covariance $\mathbf{S}_k$ at time k can then be calculated by

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2L^a} \omega_{(m)}^i \mathbf{x}_{(x),k|k-1}^i, \quad (14)$$

$$\mathbf{P}_{k|k-1} = \sum_{i=0}^{2L^a} \omega_{(c)}^i [\mathbf{x}_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}] [\mathbf{x}_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}]^T, \quad (15)$$

$$\hat{\mathbf{z}}_k = \sum_{i=0}^{2L^a} \omega_{(m)}^i \mathbf{z}_k^i, \quad \text{and} \quad (16)$$

$$\mathbf{S}_k = \sum_{i=0}^{2L^a} \omega_{(c)}^i (\mathbf{z}_k^i - \hat{\mathbf{z}}_k)(\mathbf{z}_k^i - \hat{\mathbf{z}}_k)^T. \quad (17)$$

The SPKF expressions for calculating the optimal Kalman gain $\mathbf{K}_k$, the updated state estimate (analysis mean) $\hat{\mathbf{x}}_{k|k}$, and its covariance $\mathbf{P}_{k|k}$ are as follows (more detailed derivations for $\mathbf{K}_k$ and $\mathbf{P}_{k|k}$ can be found in the appendix):

$$\mathbf{K}_k = \mathbf{P}_{xz} \mathbf{S}_k^{-1}, \quad (18)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_k), \quad \text{and} \quad (19)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T, \quad (20)$$

where $\mathbf{P}_{xz}$ denotes the cross covariance between $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{z}}_k$ and can be computed by

$$\mathbf{P}_{xz} = \sum_{i=0}^{2L^a} \omega_{(c)}^i [\mathbf{x}_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}][\mathbf{z}_k^i - \hat{\mathbf{z}}_k]^T. \quad (21)$$

The estimates of the mean and covariance using SUT are accurate to the second-order (third order for true Gaussian priors) of the Taylor series expansion for a nonlinear function.³ However, the SPKF approach is computationally unfeasible in high-dimensional models, such as atmospheric or ocean models, due to the large dimensionality (L^a) of the augmented states. Another major drawback of the SPKF approach is that the sigma points of the SPKF are determined based on the global domain, making it impossible to use the localization scheme for generating the sigma points, which is a

³ A more detailed analysis for the estimation accuracy of the SUT can be found in van der Merwe (2004) and Simon (2006).

good strategy for overcoming issues with computational feasibility.

3. The local unscented transform Kalman filter

The LUTKF locally executes the data assimilation based on the nonaugmented SPKF algorithm (section 3a) where the spatial localization method (section 3b) is applied at each grid point.

a. Nonaugmented SPKF filtering

Suppose we have an L_x -dimensional discrete-time nonlinear system described by the dynamical model and the observation model with additive noise. Accordingly, Eqs. (1) and (2) can be rewritten, respectively, as

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}, \quad (22)$$

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{v}_k. \quad (23)$$

The process noise and measurement noise in the fully nonlinear system [Eqs. (1) and (2)] are injected in a nonlinear fashion. Hence, the state vector for the SPKF is redefined as the concatenation of the original state, process noise, and measurement noise so that the SPKF algorithm takes nonlinearities in the multiplicative noise into account. However, in the case of additive process and measurement noise, the system state vector does not need to be augmented with the random noise variables, since the SPKF does not need to consider nonlinearities in the noise model as multiplicative noise in Eqs. (1) and (2) (Julier and Uhlmann 1996; van der Merwe 2004). The use of the nonaugmented state vector can reduce the dimension of the sigma points as well as the total number of sigma points used. While the dimension of the augmented state vector in the SPKF is L^a , the dimension of the state vector in the nonaugmented SPKF is L_x . Hence, while the number of the sigma points required to estimate the true mean and covariance in the SPKF is $2L^a + 1$, the number of the sigma points used in the nonaugmented SPKF is $2L_x + 1$.

The sigma point selection and state estimation of the nonaugmented SPKF are similar to those of the SPKF, except for the use of the nonaugmented state vector; that is, the augmented state estimate $\hat{\mathbf{x}}^a$, covariance $\mathbf{P}^a$, and state dimension L^a in both phases of the SPKF [Eqs. (6)–(21)] are replaced with nonaugmented forms $\hat{\mathbf{x}}$, $\mathbf{P}$, and L_x for the nonaugmented SPKF, respectively. Furthermore, the sigma points are transformed through the dynamical model (22) and measurement model (23) as follows:

$$\chi_{(x),k|k-1}^i = f(\chi_{(x),k-1|k-1}^i) \quad \text{for } i = 0, \dots, 2L_x, \quad (24)$$

$$\mathbf{z}_k^i = h[\chi_{(x),k|k-1}^i] \quad \text{for } i = 0, \dots, 2L_x. \quad (25)$$

Since the process noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are assumed to be white noise with the zero-mean, there are not additive noise terms in Eqs. (14) and (16) for the nonaugmented SPKF. Therefore, the sigma points for noise vectors are not necessarily required in Eqs. (24) and (25) as in van der Merwe (2004).

Note that unlike the SPKF, the covariance matrices $\mathbf{Q}_{k-1}$ and $\mathbf{R}_k$ are added at the end of the prior error covariance $\mathbf{P}_{k|k-1}$ and measurement error covariance $\mathbf{S}_k$, respectively, in order to take account of the model error $\mathbf{w}_{k-1}$ and observation noise $\mathbf{v}_k$ as follows:

$$\mathbf{P}_{k|k-1} = \sum_{i=0}^{2L_x} \omega_{(c)}^i [\chi_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}][\chi_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}]^T + \mathbf{Q}_{k-1}, \quad (26)$$

$$\mathbf{S}_k = \sum_{i=0}^{2L_x} \omega_{(c)}^i (\mathbf{z}_k^i - \hat{\mathbf{z}}_k)(\mathbf{z}_k^i - \hat{\mathbf{z}}_k)^T + \mathbf{R}_k. \quad (27)$$

The nonaugmented SPKF algorithm has been successfully implemented and employed to analyze practical systems (Farina et al. 2002; Crassidis and Markley 2003; Bshara et al. 2010; Cui et al. 2017). A detailed description and deviation of this SPKF algorithm can be found in van der Merwe (2004).

b. Spatial localization

The LUTKF algorithm is analogous to carrying out the SPKF locally, with the nonaugmented form for each region of the model. The spatial localization has a great impact on the performance of the EnKFs for high-dimensional systems, such as atmospheric and oceanic models (Oke et al. 2005, 2008; Deng et al. 2010).

The localization requires enough ensemble members to represent the uncertainty in only the local space; therefore, the local analysis for LUTKF can efficiently perform the data assimilation with a smaller ensemble size for each model grid point, just as it is applied to the low-dimensional system. Localizing the analysis can suppress spurious correlations between distant positions in the background covariance $\mathbf{P}_{k|k-1}$, thereby providing better analysis results (Houtekamer and Mitchell 2001; Whitaker and Hamill 2002). Moreover, the local analysis can be computed independently in parallel. However, the localization can increase the imbalance of analyses (Greybush et al. 2011).

The localization approaches used in the EnKF algorithms can be classified into two general categories: B localization that addresses the background covariance $\mathbf{P}_{k|k-1}$; and R localization that deals with the

observation error covariance $\mathbf{R}_k$ (Greybush et al. 2011; Houtekamer and Zhang 2016).

The B localization multiplies the elements in the background covariance $\mathbf{P}_{k|k-1}$ by a distance-dependent function that decays to zero beyond a cutoff distance, and thus the observations do not influence the model state beyond the cutoff distance (Houtekamer and Mitchell 2001; Whitaker and Hamill 2002). The B localization can be applied to the EnKF algorithm by multiplying the elements of the background covariance matrix $\mathbf{P}_{k|k-1}$ by another matrix $\boldsymbol{\rho}$ whose elements denote the distance-dependent localization function using a Schur product; therefore, the ensemble Kalman gain in the EnKF is given by

$$\mathbf{K}_k = \boldsymbol{\rho} \circ \mathbf{P}_{k|k-1} \mathbf{H}^T [\mathbf{H}(\boldsymbol{\rho} \circ \mathbf{P}_{k|k-1}) \mathbf{H}^T + \mathbf{R}]^{-1}, \quad (28)$$

where $\mathbf{H}$ is the linearized operator (partial derivative) of the nonlinear measurement function h , and $\circ$ represents an element-wise product.

However, applying the B localization to the LUTKF in order to remove spurious correlations in the background covariance $\mathbf{P}_{k|k-1}$ may be computationally intractable, since the background covariance matrix $\mathbf{P}_{k|k-1}$ is not represented explicitly in the expressions for calculating the background observation covariance $\mathbf{S}_k$ and Kalman gain $\mathbf{K}_k$ [Eqs. (17) and (18)]. Furthermore, since the B localization in model space yields full-rank background covariance matrix $\mathbf{P}_{k|k-1}$ (Manoj et al. 2014), the LUTKF may suffer from decomposing the

background covariance matrix. For this reason, the B localization may not be suitable for the LUTKF.

The simplest method for implementing the R localization takes into account the observations only within a cutoff radius surrounding the grid point, and it carries out the analysis separately for each spatial grid point of the model (Houtekamer and Mitchell 1998; Keppenne 2000; Ott et al. 2004; Hunt et al. 2007). However, the abrupt cutoff distance in the R localization can lead to a noisy estimate of the model state. To solve this problem, the smoothed R localization was proposed by Hunt et al. (2007). The localization can achieve an effect similar to the B localization by multiplying the elements of the observation error covariance $\mathbf{R}_k$ by a function that increases the uncertainty assigned to the observations gradually, as the distances of the observations from the analysis grid point increases (Houtekamer and Zhang 2016); therefore, the observations do not have an impact on the model state at analysis grid point beyond the cutoff distance c .

The localization can be applied to the LUTKF by dividing the diagonal elements of $\mathbf{R}_k$ in Eq. (27) by the Gaussian localization function $G(d, L) = \exp(-d^2/2L^2)$, where d is redefined as the distance between observation and analysis grid point, and L is reinterpreted as the localization distance for the observations.⁴ In practice, this study uses the following piecewise polynomial approximation of the Gaussian localization function with compact support, which is given by Eq. (4.10) of Gaspari and Cohn (1999) as follows:

$$C(d, c) = \begin{cases} 1 - \frac{20}{3} \left(\frac{d}{c}\right)^2 + 5 \left(\frac{d}{c}\right)^3 + 8 \left(\frac{d}{c}\right)^4 - 8 \left(\frac{d}{c}\right)^5, & 0 \leq d \leq \frac{1}{2}c, \\ \frac{8}{3} \left(\frac{d}{c}\right)^5 - 8 \left(\frac{d}{c}\right)^4 + 5 \left(\frac{d}{c}\right)^3 + \frac{20}{3} \left(\frac{d}{c}\right)^2 - 10 \left(\frac{d}{c}\right) + 4 - \frac{1}{3} \left(\frac{d}{c}\right)^{-1}, & \frac{1}{2}c \leq d \leq c, \\ 0, & c \leq d. \end{cases} \quad (29)$$

This equation facilitates addressing the localization values in terms of the cutoff distance $c = 2 \times (0.3)^{-0.5} L$ rather than the localization distance L . Considering only the diagonal elements of $\mathbf{R}_k$ for the localization can offer better computational efficiency compared to the localization technique that requires another matrix whose elements represent the localization function for all the elements, as in the B localization.

The LUTKF offers the following advantages over the SPKF: (i) the use of the nonaugmented state makes it easier to implement the localization scheme at each model grid point by considering only the original model

state, not including the process and measurement noise in the state vector, and it allows the LUTKF to be more computationally efficient by reducing the number of sigma points; and (ii) unlike the SPKF, the LUTKF can employ the localization method used to enhance the computational efficiency and to suppress spurious correlations between distant locations in the error covariance matrix.

⁴ The observation covariance matrix $\mathbf{R}$ is assumed to be diagonal. This is a reasonable assumption given that observation errors are white and uncorrelated.

c. Efficient algorithm implementation

In this subsection, we provide a step-by-step description for the LUTKF algorithm in aspect of ease of implementation and computational efficiency; therefore, we account for the simplest method for implementing the LUTKF algorithm in each step and demonstrate alternative methods for minimizing the amount of the computation in some steps.

By exploiting the localization method introduced in section 3b, the LUTKF can split the global analysis domain into a number of local analysis points (i.e., model grid points where the analyses are executed), and its data assimilation can be conducted independently for each local analysis point, as in the LETKF (Hunt et al. 2007; Greybush et al. 2011). Thus, if the state dimension of the dynamical model can be expressed as L_x at each model grid point, then the LUTKF based on the non-augmented state vector starts with an initial ensemble of $2L_x + 1$ L_x -dimensional model state vectors $\{\chi_{(x),0|0}^i\}_{i=0}^{2L_x}$ generated locally using the deterministic sampling based on the SUT at each grid point, just as it is applied to the L_x -dimensional system. For the LUTKF, the mean and covariance of the model state are initialized before any observations are available as follows:

$$\hat{\mathbf{x}}_{0|0} = E[\mathbf{x}_0], \quad (30)$$

$$\mathbf{P}_{0|0} = E[(\mathbf{x}_0 - \hat{\mathbf{x}}_{0|0})(\mathbf{x}_0 - \hat{\mathbf{x}}_{0|0})^T], \quad (31)$$

where $E[\cdot]$ denotes the expected value or the mathematical expectation, and $\mathbf{x}_0$ is the initial model state and can be chosen to properly approximate the error statistics of the model, as in Evensen (2003). By substituting L^a , $\hat{\mathbf{x}}_{k-1|k-1}^a$, and $\mathbf{P}_{k-1|k-1}^a$ with L_x , $\hat{\mathbf{x}}_{0|0}$, and $\mathbf{P}_{0|0}$ in Eqs. (6)–(11), respectively, the initial ensemble members $\{\chi_{(x),0|0}^i\}_{i=0}^{2L_x}$ and their associated weights ($\omega_{(m)}^i$ for the ensemble mean and $\omega_{(c)}^i$ for the ensemble covariance) are locally determined at each grid point.

For example, for the Lorenz (1996) model (hereafter L96) that mimics the time evolution of an unspecified scalar meteorological quantity x at N_x equidistant grid points along a latitude circle, the LUTKF starts with three ($2L_x + 1$, where $L_x = 1$) ensemble members chosen locally in a deterministic manner for each model grid point of the N_x -dimensional L96 model, just as it is applied to the one-dimensional model (i.e., $L_x = 1$) by the spatial localization at that model grid point, as shown in Fig. 1. That is, because the L96 model has one model variable x for one model grid point, three ensemble members are locally used for the LUTKF to solve the analysis at each model grid point. On the other hand, the LETKF starts with randomly generated ensemble members at each model grid point (see Fig. 1).

FIG. 1. Initial ensemble for the LUTKF and LETKF at each L96 model grid point with $L_x = 1$, when $\alpha = 1$, $\beta = 2$, $\kappa = 0$, $\hat{\mathbf{x}}_{0|0} = 0$, and $\mathbf{P}_{0|0} = 1$. The circle symbols with the blue solid line denote initial ensemble members of the LETKF sampled from the normal distribution with mean $\hat{\mathbf{x}}_{0|0} = 0$ and standard deviation $\sqrt{\mathbf{P}_{0|0}} = 1$, the probability density function of which is represented by the black solid curve. Three ensemble members of the LUTKF (red circles) generated by the deterministic sampling can statistical properties of the normal distribution.

After generating the initial ensemble, the LUTKF recursively estimates the model state using the following steps:

1) STEP 1

In this step, before the model state for each model grid point at time k is predicted, a global analysis ensemble of $2L_x + 1$ $L_{x,[g]}$ -dimensional model state vectors $\{\chi_{(x),k-1|k-1,[g]}^i\}_{i=0}^{2L_x}$ at time $k - 1$ is first obtained by the concatenation of the local analysis ensembles $\{\chi_{(x),k-1|k-1}^i\}_{i=0}^{2L_x}$ for each grid point of the model, where $[g]$ indicates that the ensemble member reflects the global model state. The global model state dimension $L_{x,[g]}$ is computed by multiplying local state dimension L_x by the number of the model grid points. Then, the global background ensemble $\{\chi_{(x),k|k-1,[g]}^i\}_{i=0}^{2L_x}$ at time k is obtained by applying the nonlinear function f to each global analysis ensemble member as follows:

$$\chi_{(x),k|k-1,[g]}^i = f[\chi_{(x),k-1|k-1,[g]}^i] \quad \text{for } i = 0, \dots, 2L_x. \quad (32)$$

The background ensemble mean $\hat{\mathbf{x}}_{k|k-1}$ at each model grid point is locally calculated using the local background ensemble $\{\chi_{(x),k|k-1}^i\}_{i=0}^{2L_x}$ for each grid point as follows:

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2L_x} \omega_{(m)}^i \chi_{(x),k|k-1}^i. \quad (33)$$

Also, the background error covariance $\mathbf{P}_{k|k-1}$ for each grid point is locally obtained by computing the weighted sum of the outer product for the local background ensemble perturbations and then adding the $L_x \times L_x$ local process noise covariance matrix $\mathbf{Q}_{k-1}$ at the end of the weighted sum, as in Eq. (26). The weighted sum of the outer product for vectors in

Eq. (26) can be replaced with a matrix multiplication as follows:

$$\mathbf{P}_{k|k-1} = \mathbf{X}_{k|k-1}^\omega \mathbf{X}_{k|k-1}^\top + \mathbf{Q}_{k-1}, \quad (34)$$

where $\mathbf{X}_{k|k-1}^\omega$ and $\mathbf{X}_{k|k-1}$ are $L_x \times (2L_x + 1)$ matrices whose i th column are $\omega_{(c)}^i [\chi_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}]$ and $\chi_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}$, respectively.

2) STEP 2

To estimate the model state in observation space, this step forms the global background observation ensemble $\{\mathbf{z}_{k,[g]}^i\}_{i=0}^{2L_x}$ by mapping $\{\chi_{(x),k|k-1,[g]}^i\}_{i=0}^{2L_x}$ into observation space through the observation operator h , as follows:

$$\mathbf{z}_{k,[g]}^i = h[\chi_{(x),k|k-1,[g]}^i] \quad \text{for } i = 0, \dots, 2L_x, \quad (35)$$

where $[g]$ indicates that the ensemble member reflects all available observations.

3) STEP 3

To perform the local analysis for each grid point in the LUTKF that requires R localization rather than B localization as mentioned in section 3b, local observations within a certain cutoff distance centered on each model grid point need to be selected from global observations, as shown in Fig. 2; that is, because observations beyond the cutoff distance at each grid point do not have an impact on the local analysis, they are not included for the local analysis, as in Houtekamer and Mitchell (1998). An optimal value of the cutoff distance for achieving the best data assimilation results can be determined empirically by trial and error.

Therefore, this step starts with the choice of the local observations required to perform the local analysis at each model grid.⁵ The L_x -dimensional local observation vector $\mathbf{z}_k$ and its error covariance matrix $\mathbf{R}_k$ at each grid point are determined by selecting elements used for the local analysis within a certain cutoff distance (local region) surrounding a given model grid point in the $L_{x,[g]}$ -dimensional global observation vector $\mathbf{z}_{k,[g]}$ and its error covariance matrix $\mathbf{R}_{k,[g]}$. For each grid point, a local ensemble of $2L_x + 1$ L_x -dimensional background observation vectors $\{\mathbf{z}_k^i\}_{i=0}^{2L_x}$ can be obtained by selecting elements corresponding to the local observation vector $\mathbf{z}_k$ in $\{\mathbf{z}_{k,[g]}^i\}_{i=0}^{2L_x}$.

Then, the mean of the local background observation ensemble $\hat{\mathbf{z}}_k$ is locally calculated at each grid point as follows:

FIG. 2. Local analysis in the LUTKF. Square symbols represent model grid points (spatial positions at which the analyses are executed) and circle symbols denote observations. The model state at the model grid point denoted by the yellow star symbol is updated by assimilating observations (red circles) within the local region denoted by the green dashed line.

$$\hat{\mathbf{z}}_k = \sum_{i=0}^{2L_x} \omega_{(m)}^i \mathbf{z}_k^i. \quad (36)$$

The error covariance $\mathbf{S}_k$ for $\hat{\mathbf{z}}_k$ is locally determined by computing the weighted sum of the outer product for local background observation perturbations and then adding $\mathbf{R}_k$ at the end of the weighted sum, as in Eq. (27). The weighted sum of the outer product for vectors in Eq. (27) can be substituted with a matrix multiplication:

$$\mathbf{S}_k = \mathbf{Z}_k^\omega \mathbf{Z}_k^\top + \mathbf{R}_k, \quad (37)$$

where $\mathbf{Z}_k^\omega$ and $\mathbf{Z}_k$ are $L_x \times (2L_x + 1)$ matrices whose i th column are $\omega_{(c)}^i (\mathbf{z}_k^i - \hat{\mathbf{z}}_k)$ and $\mathbf{z}_k^i - \hat{\mathbf{z}}_k$, respectively. For the smoothed R localization, the diagonal elements of $\mathbf{R}_k$ in Eqs. (27) and (37) can be divided by the Gaussian localization function $C(d, c)$, as explained in section 3b. Similarly, the cross covariance $\mathbf{P}_{xz}$ between $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{z}}_k$ is locally calculated by

$$\begin{aligned} \mathbf{P}_{xz} &= \sum_{i=0}^{2L_x} \omega_{(c)}^i [\chi_{(x),k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}] (\mathbf{z}_k^i - \hat{\mathbf{z}}_k)^\top, \\ &= \mathbf{X}_{k|k-1}^\omega \mathbf{Z}_k^\top. \end{aligned} \quad (38)$$

4) STEP 4

In this step, the Kalman gain $\mathbf{K}_k$ is locally computed by multiplying $\mathbf{P}_{xz}$ by the inverse of $\mathbf{S}_k$ at each model grid point, as in Eq. (18). If the number of observations (L_x)

⁵ Since the local analysis needs much less data compared to the global analysis and can be separately executed for each model grid point in parallel, the localization enhances the computational efficiency of the data assimilation.

used for the local analysis of each model grid point is large compared to the ensemble size ($2L_x + 1$) and the number of model variables (L_x), this step may result in the highest computational cost among steps due to the inverse of $L_z \times L_z$ matrix $\mathbf{S}_k$.

Since the covariance matrix $\mathbf{S}_k$ is symmetric, then Eq. (18) can be expressed as a linear system

$$\mathbf{S}_k \mathbf{K}_k^T = \mathbf{P}_{xz}. \quad (39)$$

To avoid the matrix inversion in Eq. (18), determining $\mathbf{K}_k^T$ by solving the linear system given by Eq. (39) can reduce the numerical cost related to the calculation of $\mathbf{K}_k$, as in Krishnamoorthy and Menon (2013). Our simulations (not reported here), which used the 40-variable L96 model in which the average number of observations used for the local analysis is 50, showed that the computational time for this step from Eq. (18) is approximately 3.89 times as much as from Eq. (39). Therefore, it may be more efficient to compute $\mathbf{K}_k$ by solving the system of linear equations rather than inverting $\mathbf{S}_k$ in this step.

5) STEP 5

This step locally computes the local analysis ensemble mean $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ using Eqs. (19) and (20) at each model grid point, respectively. By substituting $\mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T$ with $\mathbf{P}_{xz} \mathbf{K}_k^T$ using Eq. (18), Eq. (20) can be written as

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P}_{xz} \mathbf{K}_k^T. \quad (40)$$

This is helpful for the efficient computation of $\mathbf{P}_{k|k}$ by reducing the number of times matrices are multiplied, compared to Eq. (20); to be more precise, the number of floating-point operations required by the calculation of $\mathbf{P}_{xz} \mathbf{K}_k^T$ is $L_x^2(2L_z - 1)$, while computing $\mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T$ requires $(L_x^2 + L_x L_z)(2L_z - 1)$ operations.

For each grid point, the local analysis ensemble members $\{\mathbf{x}_{(x),k|k}^i\}_{i=0}^{2L_x}$ and their associated weights at time k are locally obtained by replacing L^a , $\hat{\mathbf{x}}_{k-1|k-1}^a$, and $\mathbf{P}_{k-1|k-1}^a$ with L_x , $\hat{\mathbf{x}}_{k|k}$, and $\mathbf{P}_{k|k}$ in Eqs. (6)–(11), respectively. The ensemble members will be employed to compute the local background mean $\hat{\mathbf{x}}_{k+1|k}$ and covariance $\mathbf{P}_{k+1|k}$ in Step 1 of the next time $k + 1$.

d. Scalable algorithm for ensemble size

The LUTKF can estimate the state of the nonlinear system with a smaller number of deterministically chosen sigma points using both the nonaugmented form and localization. The localization in the LUTKF is used to remove the spurious long-distance correlations due to sampling error from finite ensemble size by suppressing

the influence of distant observation. Nonetheless, the use of the limited ensemble size in the LUTKF may still lead to the correlations that are deemed to be spurious, especially for higher-dimensional systems. Thus, the ensemble size needs to be large enough to approximately span the space of possible system states, because more ensemble members generally allow more observations to be used gainfully.

A possible solution to this problem is to increase the number of sigma points so that correlations between grid points that are far apart are suppressed, thus allowing the LUTKF to more accurately approximate the error correlation statistics of the model. We can devise several approaches to modify the generation of sigma points to make the LUTKF applicable for this problem. Here, as an example, we propose a variant of the LUTKF algorithm that controls the number of sigma points to be produced with the modifications, called a local scalable unscented transform Kalman filter (LSUTKF).

Similar to the LUTKF, the LSUTKF estimates the nonaugmented state of the nonlinear system [Eqs. (22) and (23)], in which the process noise and observation noise are purely additive. However, unlike the LUTKF, the LSUTKF employs a truncated normal distribution in selecting the sigma points to form the sigma point space that approximates the mean and error covariance of the model state more accurately, assuming that the system state vector $\mathbf{x}_k$ follows a Gaussian distribution. The truncated normal distribution is the probability distribution that represents the random variable x to be distributed normally with mean μ and variance σ^2 and to be truncated to the range (a, b) (i.e., $a < x < b$), where the values of a and b are defined over the domain of the standard normal distribution (Robert 1995; Chu et al. 2010). This distribution can be referred to as $\mathcal{N}_t(\mu, \sigma^2, a, b)$. To adjust the number of the sigma points according to the system, the LSUTKF generates the sigma points using both the deterministic sampling approach used in the LUTKF and the stochastic (random) sampling from the truncated normal distribution.

For an L_x -dimensional system, while the LUTKF uses $2L_x + 1$ sigma points, the number of sigma points required to approximate the error statistics of the model in the LSUTKF is $2qL_x + 1$, where q is a scaling factor that adjusts the number of sigma points. The initialization, forecast, and analysis steps of the LSUTKF are executed in the same way as those of the LUTKF except for the sigma point selection and the substitution of the value of L_x in the equations of the LUTKF with $\delta = qL_x$.

Given the mean $\hat{\mathbf{x}}_{k-1|k-1}$ and its covariance $\mathbf{P}_{k-1|k-1}$ in the sigma point selection phase of the LSUTKF, $2\delta + 1$

sigma points $\{\chi_{(x),k-1|k-1}^j\}_{j=0}^{2\delta}$ are locally generated at each model grid point as follows:

$$\chi_{(x),k-1|k-1}^0 = \hat{\mathbf{x}}_{k-1|k-1}, \quad (41)$$

$$\chi_{(x),k-1|k-1}^j = \begin{cases} \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{D}_m, & j = 2q(i-1) + m; m = 1, \dots, q-1; q \geq 2, \\ \hat{\mathbf{x}}_{k-1|k-1} + [\sqrt{(\delta + \lambda)\mathbf{P}_{k-1|k-1}}]_i, & j = 2q(i-1) + m; m = q; q \geq 1, \\ \hat{\mathbf{x}}_{k-1|k-1} - \mathbf{D}_m, & j = 2q(i-1) + q + m; m = 1, \dots, q-1; q \geq 2, \\ \hat{\mathbf{x}}_{k-1|k-1} - [\sqrt{(\delta + \lambda)\mathbf{P}_{k-1|k-1}}]_i, & j = 2q(i-1) + q + m; m = q; q \geq 1, \end{cases} \quad (42)$$

where $\mathbf{D}$ is the $L_x \times (q-1)$ matrix that consists of random vectors generated from the truncated normal distribution $\mathcal{N}_i\{0, (\sqrt{\mathbf{P}_{k-1|k-1}})_i, 0, [\sqrt{(\delta + \lambda)\mathbf{P}_{k-1|k-1}}]_i\}$, and $\mathbf{D}_m$ indicates the m th column vector of the matrix $\mathbf{D}$.

Similar to the LUTKF, the sigma points in the LSUTKF are deterministically chosen through the vector $[\sqrt{(\delta + \lambda)\mathbf{P}_{k-1|k-1}}]_i$ centered on the mean $\hat{\mathbf{x}}_{k-1|k-1}$ so that they can represent the mean and covariance of the posterior distribution. Moreover, to produce more sigma points that follow the Gaussian distribution, they are determined from the sigma point distribution $\mathbf{D}$ obtained from the truncated distribution around the mean $\hat{\mathbf{x}}_{k-1|k-1}$.

For example, when applied to the L96 model mentioned in section 3c, sigma points generated by the LSUTKF with mean $\hat{\mathbf{x}}_{k-1|k-1}$ and covariance $\mathbf{P}_{k-1|k-1}$ are indicated by circle symbols at the bottom of Fig. 3. The red circle represents the sigma point for the mean $\hat{\mathbf{x}}_{k-1|k-1}$, and the circle symbols with green and blue solid lines from both ends around this point denote sigma points obtained by the deterministic sampling of the LUTKF. The circle symbols that are located symmetrically in the light gray shaded region around the red circle symbol indicate sigma points obtained by the random sampling from the truncated normal distribution.

Based on this approach using the truncated normal distribution, we performed data assimilation tests with an enlarged ensemble size and identified its ability to eliminate the spurious long-distance correlations in the frame of the LUTKF. However, a more detailed description of the LSUTKF is beyond the scope of this study that aims to propose a localized form of the SPKF algorithm. Therefore, in addition to data assimilation results, more details on the methodology and implementation for the LSUTKF will be described in future work.

4. Cycling data assimilation experiments

a. Lorenz (1996) model

In this section, the LUTKF algorithm is applied to the L96 model. The L96 model includes N_x equally spaced

and for $i = 1, \dots, L_x$, the following procedures are executed as

variables $\{x_i\}_{i=1}^{N_x}$, which can be thought of as atmospheric variables at N_x equidistant grid points along a latitude circle, governed by the following differential equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad (43)$$

with cyclic boundary conditions (i.e., both x_{i+N_x} and x_{i-N_x} are equal to x_i).

For experiments executed in this study, the L96 model setup is similar to that of Lorenz and Emanuel (1998), where the number of model variables N_x is set to 40 and the magnitude of the forcing F is fixed at 8, resulting in chaotic behavior in the system dynamics. The L96 model assumes that a unit time $\Delta t = 1$ is equivalent to 5 days. Forward integration of Eq. (43) is carried out numerically through the fourth-order Runge–Kutta method with a time step of $\Delta t = 0.05$ (i.e., 6 h).

b. Design of the experiment

To investigate the data assimilation performance of the LUTKF and LETKF, several numerical experiments were performed using model and observation system configurations that mimic the features of realistic weather systems. Both the LUTKF and LETKF employ the correlation function with compact support [Eq. (29)] described in section 3b for the R localization technique, the Gaussian-type structure of which is modified by setting a localization length scale c to 1.1 and 3.7 for achieving the best data assimilation results for the LUTKF and LETKF, respectively. Note that the value of c in this experiment indicates the fraction of the distance between two adjacent grid points in the L96 model.

For the LETKF, the underestimation of forecast error covariance (i.e., uncertainty in its prior state estimate) can be caused by limited ensemble size N_e and systematic errors (Whitaker and Hamill 2002; Whitaker et al. 2008; Zhang et al. 2004). To avoid filter divergence due

FIG. 3. Sigma point selection for the LSUTKF at each L96 model grid point with $L_x = 1$, when $\alpha = 1$, $\beta = 2$, $\kappa = 0$, $q = 16$, $\hat{\mathbf{x}}_{k-1|k-1} = 0$, and $\mathbf{P}_{k-1|k-1} = 0.6$. The black solid curve represents the probability density function of the normal distribution with mean $\hat{\mathbf{x}}_{k-1|k-1} = 0$ and variance $\mathbf{P}_{k-1|k-1} = 0.6$. The red circle corresponds to the sigma point for the mean $\hat{\mathbf{x}}_{k-1|k-1}$. The circle symbols with green and blue solid lines in the light gray shading indicate the sigma points generated by minus ($-$) sigma point distribution $-\mathbf{D}$ and plus ($+$) sigma point distribution $+\mathbf{D}$, respectively. The circles from both ends centered on the red circle represent the sigma points chosen deterministically by matrix square root $\sqrt{(\delta + \lambda)\mathbf{P}_{k-1|k-1}}$.

to this problem, in this experiment, the LETKF uses the adaptive covariance inflation scheme of Whitaker and Hamill (2012) that is referred to as relaxation to prior spread (RTPS) in addition to the localization method. The tunable parameter for the RTPS method [the α in Eq. (3) of Whitaker and Hamill (2012)] was set to 0.4 for all the experiments. We found that this choice of α ensures that the LETKF provides a low value of the domain-averaged root-mean-square error (RMSE). As α is increased above the optimal value, better analysis results could not be obtained, because the background error covariance is increasingly overestimated. When α is decreased to 0, since the background error covariance can be systematically underestimated, the performance of the LETKF was degraded due to underweighted observations.

State estimation results of the LUTKF depend on the choices for the control parameters α , β , and κ , which were introduced in section 2. Trial simulations (not shown here) showed that values of $\alpha = 1$, $\beta = 2$, and $\kappa = 0$ lead to a high correlation between the true state and its estimated value by the LUTKF in the experiments executed in this study.

Considering that the model errors are white and uncorrelated, the model error covariance matrix $\mathbf{Q}$ is assumed to be diagonal. For our experiments, the diagonal elements of covariance matrix $\mathbf{Q}$ are set to 0.01. The true states are generated by integrating the system over 6000 cycles using the fourth-order Runge–Kutta method with the specified initial condition $\mathbf{x}_0$, which is generated by adding the noise $\mathbf{w} \sim \mathcal{N}(0, \mathbf{Q})$ to the magnitude of the forcing F . Furthermore, the covariance $\mathbf{Q}$ is incorporated into the assimilation by Eq. (26).

The experiments were carried out through three different types of the observation operators: the first experiment used the linear operator $h(\mathbf{x}) = \mathbf{x}$, where the model state is interpolated from the model space to observation space; and the second and third experiments exploited the nonlinear measurement operators, where the interpolated states are applied for $|\mathbf{x}|$ and $\ln(|\mathbf{x}|)$, respectively, introducing an additional nonlinear source that can limit the effectiveness of the LETKF. The features of each observation operator are described in Fig. 5 in Poterjoy (2016).

Observations N_z are created by adding the noise $\mathbf{v} \sim \mathcal{N}(0, \mathbf{R})$ to the true data using Eq. (23), where the diagonal elements of covariance matrix $\mathbf{R}$ are equal to 0.01.⁶ Unless stated otherwise, it is assumed that the default configuration is composed of a linear interpolation for h and no model error [e.g., $\mathbf{w}_{k-1}$ in Eqs. (1) and (22)] for the LETKF as in Hunt et al. (2007). Also, the value of N_z is set to 100 to prevent filter divergence during the cycling experiments performed by the assimilation methods according to three different observation operators and various ensemble sizes ranging from 2 to 1000. Trial simulations (not reported here) showed that this choice of N_z leads to a high degree of correlation between the true model state and its estimated value using the filters in the set experimental configurations. The time interval of observations was set to 6 h; thus, the observed states were assimilated to the nonlinear model every 6 h.

The spatial positions of the observations were fixed over time and determined randomly from a Gaussian distribution centered on the variable $\mathbf{x}_{20}$ with a standard deviation of one-third of the model domain length. As a result, most observations are located at the regions of the model domain surrounding the variable $\mathbf{x}_{20}$, and thus relatively unobserved domains exist, which is similar to environmental observation platforms. After a spinup period of 1000 cycles, the average values of prior domain-averaged RMSEs and the ensemble spread over 5000 cycles were calculated to verify the performance of the data assimilation algorithms.

The assimilation experiments were performed on a Cray XC40 system (available online from Cray, Inc., at <https://www.cray.com/products/computing/xc-series>) comprising 2904 dual-processor nodes, each with 128 GB of random access memory. Each processor in the node is a 2.60 GHz Intel Xeon CPU E5–2690 v3 with hyper-threading disabled, running on Linux. Twenty

⁶ The diagonal elements of $\mathbf{R}$ are much smaller than examined in most other studies in the literature. Additional experiments using diagonal elements that are set to 1 are addressed in section 4e.

computational nodes on the supercomputer system were employed for verification of the assimilation performance and computational efficiency of the filters. The L96 model and data assimilation methods were implemented with Python 2.7.3 software (available online from Python Software Foundation at <https://www.python.org/downloads/release/python-273>) and the Message Passing Interface (MPI) for the Python package (available online at <https://mpi4py.readthedocs.io/en/stable>) called mpi4py, used for parallel computing in Python. For the LUTKF, the data assimilation for each model grid point by using the localization scheme can be executed independently in parallel. Therefore, the Cray system was used to execute independently the local analysis for each grid point of the L96 model with $N_x = 40$ in parallel, using two processor compute cores per computational node; that is, data assimilation calculations of the LUTKF were parallelized by performing the local analysis using a compute core for each grid point.

c. Experiment with ensemble size and measurement operator

Figure 4 shows the average values of the prior ensemble mean RMSE and spread of the SPKF, LUTKF, and LETKF according to three different observation operators and ensemble size N_e . The SPKF required 361 [$2L^a + 1$, where L^a given by Eq. (5) is determined by $L_x = 40$, $L_w = 40$, and $L_v = 100$] ensemble members by the deterministic sampling and global analysis to provide satisfactory filtering solution for L96 model grid points. As described in section 3, the LUTKF algorithm based on the nonaugmented SPKF filtering and the localization approach can carry out the data assimilation with a small ensemble size, just as it is applied to the low-dimensional system. Since the L96 model has one model variable x for one model grid point, the LUTKF locally required only three ($2L_x + 1$, where $L_x = 1$) ensemble members by the deterministic sampling to provide a stable filtering solution at each grid point of the N_x -dimensional L96 model, just as it is applied to the one-dimensional model (i.e., $L_x = 1$) by the spatial localization. On the other hand, the LETKF provided satisfactory filtering results when $N_e \geq 10$ locally at each L96 model grid point; that is, its prior RMSE values shown in Fig. 4 decreased abruptly until the number of ensemble members reached 10 (i.e., the optimal ensemble size) and then converged to the values of about 0.114, 0.115, and 0.182 for the measurement operators $h(\mathbf{x}) = \mathbf{x}$, $h(\mathbf{x}) = |\mathbf{x}|$, and $h(\mathbf{x}) = \ln(|\mathbf{x}|)$, respectively. On average, its prior spread values matched the true forecast errors.

When $h(\mathbf{x}) = \mathbf{x}$ and $h(\mathbf{x}) = |\mathbf{x}|$, the results show that the LUTKF outperforms the prior estimation accuracy of the LETKF for small ensemble sizes ranging from 2 to

FIG. 4. Prior mean RMSE and spread of SPKF using $N_e = 361$, LUTKF using $N_e = 3$, and LETKF as a function of ensemble size with respect to the three different observation operators. The blue shading shows that the LUTKF can perform better than the LETKF when small ensembles are used, while the gray shading represents marginal accuracy improvements of LETKF over the LUTKF when enough ensemble sizes are used.

7 (blue shading in Figs. 4a,b), because the LETKF produces a noisy estimate for the forecast error distribution of the true state if a small value of N_e is used due to the random sampling.

When h is strongly nonlinear [i.e., $h(\mathbf{x}) = \ln(|\mathbf{x}|)$], the LUTKF can achieve about 91% higher accuracy than the LETKF for $N_e = 3$ (see blue shading in Fig. 4c), while it can reach about 46.21% and 48.74% higher accuracy for the observation types $h(\mathbf{x}) = \mathbf{x}$ and $h(\mathbf{x}) = |\mathbf{x}|$, respectively. Because the LETKF makes an incorrect linearization assumption that takes into account only the first-order term of the Taylor expansion even for the highly nonlinear function h , it is suboptimal for this observation operator. On the contrary, the LUTKF makes no assumption regarding h due to SUT, thereby providing up to at least a second-order accuracy.

Figure 4 also shows that the SPKF based on the global analysis and the augmented state vector can offer slightly more accurate estimation results with a larger ensemble size for the measurement operators $h(\mathbf{x}) = \mathbf{x}$,

FIG. 5. Rank histogram obtained from prior ensembles of SPKF using $N_e = 361$ for the three observation types. (from left to right) The verification is carried out for state variables 2, 10, 18, 26, and 34 and for all variables 1–40.

$h(\mathbf{x}) = |\mathbf{x}|$, and $h(\mathbf{x}) = \ln(|\mathbf{x}|)$, compared to the LUTKF based on the local analysis and nonaugmented state vector. However, the SPKF can be computationally unfeasible by requiring excessive ensemble members owing to the large dimensionality (L^a) of the augmented state vector. Meanwhile, the LUTKF is more computationally efficient for nonlinear systems by reducing the number of sigma points due to the use of the nonaugmented state vector and localization approach. The computational performance of the SPKF and LUTKF is discussed further in section 4e.

Figures 5–8 represent rank histograms obtained from prior ensembles of the SPKF, LUTKF, and LETKF during the experiments using the three observation types. The rank histograms were obtained by calculating the frequency with which the true model state lands in discrete bins formed by ranking prior ensemble members in ascending order. This histogram verification was performed for state variables 2, 10, 18, 26, and 34 and for all variables ranging from 1 to 40 in the model domain using the SPKF with the size $N_e = 361$ (Fig. 5), LUTKF with the size $N_e = 3$ (Fig. 6), and LETKF with the size $N_e = 3$ (Fig. 7) and $N_e = 10$ (Fig. 8) during cycles of data assimilation. Nonuniform histograms for the frequencies across the bins can indicate that the probabilistic representation of variables by the prior ensemble deviates from the true prior distribution of the system state.

Figures 5 and 6 show that the SPKF and LUTKF can achieve flat histograms for all the state variables and the three observation types. While Fig. 5 indicates that the SPKF based on the global domain can approximate the true prior distribution of the model state with a large ensemble size by using the augmented state vector, Fig. 6 demonstrates the advantage of the LUTKF in yielding reliable probabilistic representations close to the truth

even with a small number of ensemble members. As expected, the histograms in Fig. 7 reveal noticeable deficiencies in the probabilistic forecasts by the LETKF using a small ensemble size $N_e = 3$. Especially, the observation type $h(\mathbf{x}) = \ln(|\mathbf{x}|)$ contains substantial deviations, which come from the suboptimal analysis that has only the first-order accuracy even for the strongly nonlinear function as well as the use of a small number of ensemble members (see Figs. 7m–r). On the other hand, Fig. 8 shows that if the ensemble size is large enough to represent the true prior distribution of the model state, LETKF forecasts can provide relatively uniform histograms.

d. Experiment on correlation between state variables

In this subsection, several data assimilation experiments were performed to illustrate how the localizing approach has an impact on the prior correlation between neighboring state variables. The LUTKF and LETKF were run as local filters using the localization while the SPKF and ETKF were run as corresponding global filters not using the localization; afterward, the localization performance was compared.

Figure 9 shows prior correlations between state variable 20 (i.e., $\mathbf{x}_{20}$) and all other variables in the background covariance, which were determined by the local filters and global filters; the assumed true correlations obtained by the particle filter (PF; Doucet et al. 2001; van Leeuwen 2009) using 10 000 ensemble members are illustrated by black lines for reference.⁷ For simplicity,

⁷ The PF makes no assumptions regarding the linearity of prediction and observation models and about the error distributions needed to estimate and sample from the posterior density.

FIG. 6. Rank histogram obtained from prior ensembles of LUTKF using $N_e = 3$ for the three observation types. (from left to right) The verification is carried out for state variables 2, 10, 18, 26, and 34 and for all variables 1–40.

the Gaspari and Cohn (1999) correlation function [Eq. (29)] used in the local filters for the localization is marked as “GC” in this figure.

For $N_e = 100$, the ETKF offers relatively accurate correlations near $\mathbf{x}_{20}$ (red lines in Figs. 9a–c). However, the use of N_e that is insufficient to describe nonzero uncertainty in the global space can lead to a noisy estimate of the correlation between state variables far from $\mathbf{x}_{20}$ in the background covariance (green lines in Figs. 9a–c). The resultant inaccurate correlations can overestimate the impact of the observations from one location on the analysis in distant locations.

When $N_e = 10$, the LETKF provides comparable results as those produced by the ETKF using $N_e = 100$ (red lines in Figs. 9d–f). This is because the localization

requires an ensemble whose size represents the nonzero uncertainty in only the local region and removes spurious correlations between distant positions in the background covariance by reducing the influence of observations on distant state variables.

Nonetheless, if N_e is too small for the local analysis, then the correlation results of the LETKF are relatively noisy in regions far from $\mathbf{x}_{20}$ (green lines in Figs. 9d–f). Furthermore, the correlation in the LETKF experiments using $N_e = 3$ with the highly nonlinear $h(\mathbf{x}) = \ln(|\mathbf{x}|)$ was more inaccurate compared to experiments with the observation types $h(\mathbf{x}) = \mathbf{x}$ and $h(\mathbf{x}) = |\mathbf{x}|$ (green line in Fig. 9f). This result is caused by the incorrect assumption of linearity in the LETKF for the nonlinear function h , in addition to sampling errors due to a small

FIG. 7. Rank histogram obtained from prior ensembles of LETKF using $N_e = 3$ for the three measurement operators. (from left to right) The verification is carried out for state variables 2, 10, 18, 26, and 34 and for all variables 1–40.

FIG. 8. As in Fig. 7, but for LETKF using $N_e = 10$.

N_e . For the LETKF, a larger value of N_e results in an increase in the localization radius c to obtain more observations that is effective in conducting the state estimate, as shown in Figs. 9d–f.

On the other hand, the LUTKF with the smaller ensemble size than the LETKF using $N_e = 10$ can estimate

relatively accurate prior correlations [even for $h(\mathbf{x}) = \ln(|\mathbf{x}|)$], since it makes no assumption with regard to h , owing to the SUT (red lines in Figs. 9j–l).

Also, the LUTKF provides results comparable to the SPKF with 361 [$2L^a + 1$, where L^a given by Eq. (5)] is determined by $L_x = 40$, $L_w = 40$, and $L_v = 100$

FIG. 9. Prior correlations between state variable 20 and all other variables in the background error covariance, which are estimated using the global filters [(a)–(c) ETKF and (g)–(i) SPKF] and local filters [(d)–(f) LETKF and (j)–(l) LUTKF] for three different types of observations.

FIG. 10. Case A: (a)–(d) Prior state estimates (dashed curves) and (e) corresponding errors (solid trajectories) for the 40-variable L96 model and $h(\mathbf{x}) = \ln(|\mathbf{x}|)$ with densely spaced measurements (i.e., $N_z = 200$): (a) LETKF for $N_e = 3$, (b) LETKF for $N_e = 10$, (c) LUTKF for $N_e = 3$, and (d) SPKF for $N_e = 561$. The solid black trajectory in (a)–(d) indicates the true state. Assimilation errors for data assimilation methods in (e) are measured by the domain-averaged prior RMSEs with time step.

ensemble members (red lines in Figs. 9g–i). These results demonstrate that the LUTKF can capture major components of the state estimate using a much smaller N_e than would be required otherwise, thereby providing a good approximation when many ensemble members are not available.

e. Potential advantages of the LUTKF

For data assimilation problems in realistic settings, such as atmospheric and oceanic systems, remotely sensed data from radars (Johnson et al. 2015) and satellites (Hocking et al. 2011) need to be assimilated to perform weather prediction and analysis. In this case, the filters should process densely spaced observations that are related nonlinearly to the model state, which poses a challenge for the LETKF, given the incorrect linearization assumption of h . For a situation analogous to this problem, the first set of experiments (hereafter Case A) was performed for the L96 system by assimilating a network of 200 randomly located observations every 6 h using $h(\mathbf{x}) = \ln(|\mathbf{x}|)$.

Figures 10a–d represent the prior state estimates obtained from the LETKF with 3 and 10 members,

TABLE 1. RMSE and computational time of each assimilation method for Case A.

Assimilation method	RMSE	Computational time (s)
LETKF (using 3 ensembles)	0.490	5.37
LETKF (using 10 ensembles)	0.162	9.54
LUTKF (using 3 ensembles)	0.171	6.70
SPKF (using 561 ensembles)	0.154	151.73

LUTKF for $N_e = 3$, and SPKF using $N_e = 561$ [$2L^a + 1$, where L^a given by Eq. (5) is determined by $L_x = 40$, $L_w = 40$, and $L_v = 200$] over time steps ranging from 3000 to 4000 during the cycling experiments for a much clearer comparison of the performance of the methods. The results show that all three filters, except LETKF with $N_e = 3$, can provide the state estimates close to true state, indicating that they perform well in estimating the model state provided that the magnitude of the observation noise is appropriate. However, the LETKF needs to assume the linearity of observation models and the SPKF requires a large ensemble size of 561. On the contrary, the LUTKF uses only three ensemble members, indicating its benefit over the LETKF and SPKF. The variation in the prior ensemble mean RMSEs over the 40 variables with time steps for the three assimilation schemes is shown in Fig. 10e.

Table 1 compares the overall assimilation performance of each filter through the averaged prior RMSE over all the variables and time steps during the cycling experiments. As seen in Fig. 10e and Table 1, the LUTKF can accomplish forecast results that are as accurate as both the LETKF using $N_e = 10$ and SPKF.

For the sake of completeness, Fig. 11 represents the prior state estimates and their domain-averaged RMSEs for Case A experiments using $N_z = 100$ as performed in section 4c. The figure shows that the model states are fairly well estimated by the LUTKF, providing prior RMSE values that are comparable to those of the LETKF using $N_e = 10$ and SPKF using $N_e = 361$ [$2L^a + 1$, where L^a given by Eq. (5) is calculated by $L_x = 40$, $L_w = 40$, and $L_v = 100$] at most time steps, although the opposite situations occur at some assimilation time steps. Although not shown, the four assimilation methods in this case yield higher prior mean RMSEs than assimilation results presented in Table 1, due to a smaller number of observations; the prior mean RMSE values of the LETKF for $N_e = 3$, LETKF for $N_e = 10$, LUTKF, and SPKF assimilation results are 2.685, 0.182, 0.213, and 0.171, respectively.

The frequent peak values of prior mean RMSEs in Figs. 10e and 11e results from either overestimation or underestimation of model states, as shown in Figs. 10a and 11a, respectively. These estimation errors are most

FIG. 11. As in Fig. 10, but for $N_z = 100$.

probably related to random measurement noise and to the chaotic characteristics of the L96 system for transition states. Thus, the assimilation methods can lead to poor results in estimating the model state when using observations with a large amount of noise during the transition phase period. More details on the performance of the filters when noisier observations are assimilated were examined in detail through the second set of experiments.

Table 1 also shows the computational time required for each assimilation approach to estimate the model state. To compare the computational efficiency, the same programming framework described in section 4b is used to implement all the assimilation schemes. In addition, the local analysis for each model grid point of the LUTKF and LETKF can be computed independently in parallel by using a parallel programming library such as the mpi4py package. As a result, even though the LUTKF yields no computational benefits over the LETKF when $N_e = 3$, the LUTKF consumes less computational cost compared to the LETKF with $N_e = 10$, as shown in Table 1. The SPKF that requires 561 members for a good state estimate is the most computationally expensive, costing 12.31 times as much as LUTKF.

To mimic a more realistic situation, the second set of experiments (hereafter Case B) was executed by increasing the noise level (error standard deviation) of the observations tenfold; that is, the diagonal entries of the

FIG. 12. Case B: (a)–(d) Prior state estimates (dashed curves) and (e) corresponding errors (solid trajectories) for the 40-variable L96 model and $h(\mathbf{x}) = \ln(|\mathbf{x}|)$ with densely spaced observations (i.e., $N_z = 200$) and a tenfold increase in the observation noise levels (error standard deviations): (a) LETKF for $N_e = 3$, (b) LETKF for $N_e = 10$, (c) LUTKF for $N_e = 3$, and (d) SPKF for $N_e = 561$. The solid black trajectory in (a)–(d) indicates the true state. Assimilation errors for data assimilation methods in (e) are measured by the domain-averaged prior RMSEs with time step.

observation error covariance $\mathbf{R}$ were set to 1. The diagonal entries of the model error covariance $\mathbf{Q}$ were also fixed to 1. Figure 12 presents the assimilation results for the state estimation methods and corresponding RMSEs over assimilation time steps ranging from 3000 to 4000. Comparing Figs. 12a–c reveals that the LUTKF estimates the state better than the LETKF with $N_e = 3$ or $N_e = 10$. Figure 12e shows some divergent tracks over time steps of cycling assimilation among the four estimation approaches. For example, the RMSE for the SPKF changes almost steadily, while the LUTKF has a relatively significant change in RMSE with time steps. In contrast to the LUTKF, the change in RMSE is more abrupt and frequent in the two LETKF methods. From Fig. 12e, we can also see that the LUTKF can provide more reliable and accurate estimation results than the two LETKF approaches, although the system is subject to higher noise.

The prior mean RMSE of assimilation results for each estimation scheme in Case B is shown in Table 2. The results show that the estimation accuracy of the LUTKF

TABLE 2. RMSE and correlation of each assimilation method for Case B.

Assimilation method	RMSE	Correlation
LETKF (using 3 ensembles)	3.007	0.649
LETKF (using 10 ensembles)	1.129	0.951
LUTKF (using 3 ensembles)	1.003	0.963
SPKF (using 561 ensembles)	0.894	0.969

is as good as the SPKF, and the LUTKF provides better results than the two LETKF methods. Table 2 also compares the correlation between the true and estimated trajectories from the assimilation methods. The LUTKF using $N_e = 3$ can provide correlation results that are as high as the SPKF assimilation. This means that the LUTKF has a superior ability to capture the observation information effectively when estimating the model state using noisy observations. For Case B experiments with $N_z = 100$ (not reported here), the LETKF with $N_e = 3$ and $N_e = 10$, LUTKF, and SPKF yield prior mean RMSE values of 4.243, 1.518, 1.636, and 1.146, respectively, and their correlations are 0.313, 0.911, 0.902, and 0.949, respectively.

Although not shown, the data assimilation methods in Case B are performed at a similar computational cost as Case A and result in accurate estimates of the error statistics of the model. To provide a more complete assessment of filter performance when h is nonlinear, Fig. 13 shows assimilation results and corresponding RMSEs for the additional experiment using $h(\mathbf{x}) = |\mathbf{x}|$ in Case B. From Fig. 13, we can see that the LUTKF can provide state estimation results that are as accurate as the LETKF with $N_e = 10$ and SPKF. Although not shown, the prior mean RMSE values of the LETKF for $N_e = 3$, LETKF for $N_e = 10$, LUTKF, and SPKF assimilation results are 0.826, 0.331, 0.397, and 0.329, respectively. These results suggest that LUTKF continues to provide reliable performance, even if observations relate nonlinearly to the model state, provided that sufficient observations are available.

For Case A and Case B experiments, densely spaced observations that are close to variable $\mathbf{x}_{20}$ were assimilated. The LUTKF with $N_e = 3$ can be very advantageous in these experiments. On one hand, the filter degeneration due to a small ensemble size can be offset by assimilating numerous observations on the same location. On the other hand, $N_z = 100$ and $N_z = 200$ suggest that a substantial number of observations can be assimilated into nearly the same location.

To examine the impact of less dense observations on the assimilation performance of the filters, the third set of experiments (hereafter Case C) was performed as in Case B, but densely spaced observations ($N_z = 200$) are

FIG. 13. As in Fig. 12, but for $h(\mathbf{x}) = |\mathbf{x}|$.

replaced with 40 observations ($N_z = 40$) that are equally spaced on the grid points; that is, all but the observation network among experimental configurations set in Case B were applied to Case C. For the best assimilation results in Case C, the localization length scale c of the Gaussian localization function given by Eq. (29) is set to 1.1, 1.4, and 3.7 for the LUTKF using $N_e = 3$, LETKF with $N_e = 3$, and LETKF with $N_e = 10$, respectively.

Figure 14 shows the prior state estimates of the LETKF with $N_e = 3$ and $N_e = 10$, LUTKF using $N_e = 3$, and SPKF using $N_e = 241$ [$2L^a + 1$, where L^a given by Eq. (5) is determined by $L_x = 40$, $L_w = 40$, and $L_v = 40$] and their corresponding domain-averaged RMSEs in Case C. Figure 15 represents the prior state estimation errors from the true state and the posterior state estimates minus prior state estimates (analysis increments) as a function of time (data assimilation cycle) and space (L96 model state variable) for the LETKF with $N_e = 10$, LUTKF using $N_e = 3$, and SPKF using $N_e = 241$ in Case C. As seen in Figs. 14, 15a, 15c, and 15e, the LUTKF can achieve estimation results that are better than the LETKF with $N_e = 3$ and comparable to the LETKF with $N_e = 10$, although it cannot provide estimation results that are as accurate as the SPKF, since assimilated observations are insufficient. Figures 15b, 15d, and 15f show that the continuity of the analysis increment of the LUTKF is comparable to that of the LETKF. This occurs because the LUTKF that performs the spatial

FIG. 14. Case C: (a)–(d) Prior state estimates (dashed curves) and (e) corresponding errors (solid trajectories) for the 40-variable **L96** model and $h(\mathbf{x}) = \ln(|\mathbf{x}|)$ with uniformly distributed (less densely spaced) measurements (i.e., $N_z = 40$): (a) LETKF for $N_e = 3$, (b) LETKF for $N_e = 10$, (c) LUTKF for $N_e = 3$, and (d) SPKF for $N_e = 241$. The solid black trajectory in (a)–(d) indicates the true state. Assimilation errors for data assimilation methods in (e) are measured by the domain-averaged prior RMSEs with time step.

localization as in the LETKF can result in the overlap of assimilated observations within the local regions between neighboring model grid points, which helps to ensure the spatial continuity of the analysis at adjacent grid points (Hunt et al. 2007).

Table 3 shows the prior mean RMSE, correlation between the true and estimated states, and computational time for the assimilation schemes in Case C. From Fig. 14e and Table 3, we can see that the estimation accuracy of the LUTKF is similar to that of the LETKF with $N_e = 10$, providing RMSE values and correlation results comparable to the LETKF for $N_e = 10$, even though the LUTKF provides no estimation accuracy improvements over the SPKF with $N_e = 241$. This suggests that the LUTKF can capture the observation information as effectively as the LETKF for $N_e = 10$, when estimating the model state using a smaller number of observations.

Furthermore, because fewer observations than Case A and Case B are assimilated in Case C, the LETKF with $N_e = 3$, LETKF with $N_e = 10$, LUTKF, and SPKF in Case C yield higher prior mean RMSEs and correlations when compared to those in Table 1 and

Table 2, and have about 60.29%, 105.16%, 74.93%, and 76.82% lower computational times than those in Table 1, respectively, as shown in Table 3.

To examine the assimilation performance of the filters when applied to the **L96** system with more chaotic behavior, Case C experiments using the **L96** system with $F = 8.1$ were performed. In general, the behavior of the **L96** system is affected by the magnitude of F (Lorenz 1996). For very small values of F , all solutions of Eq. (43) will converge to a steady state, while, for somewhat larger values of F , most solutions are periodic. However, when the value of F is large enough, the solutions are chaotic.

The maximal Lyapunov exponent (MLE) that can provide a reasonable indication of the magnitude of the chaos of the system was calculated; a negative MLE indicates that the system is exponentially stable, while a positive MLE means that the system is chaotic and unpredictable. The MLE values of **L96** models for $F = 8.0$ and $F = 8.1$ were 0.212 and 0.218, respectively. This suggests that the **L96** model with $F = 8.1$ displays somewhat more chaotic behavior compared to the model using $F = 8.0$.

Figure 16 shows the prior state estimates and corresponding RMSEs of the filters for Case C experiments using the **L96** system with $F = 8.1$. Similar to assimilation results of the **L96** model using $F = 8.0$ in Fig. 14, the LUTKF using as few as three ensemble members can achieve the estimation performance that is as accurate as the LETKF with $N_e = 10$ and can outperform the LETKF using $N_e = 3$, although it can yield no significant benefits over the SPKF using $N_e = 241$.

From Table 3, we can see that LETKF with $N_e = 3$, LETKF with $N_e = 10$, LUTKF, and SPKF using the **L96** model with $F = 8.1$ provide about 1.55%, 4.81%, 4.59%, and 5.52% higher RMSE values, respectively, and 1.74%, 1.79%, 1.34%, and 1.54% lower correlations, respectively, compared to experiments using the **L96** model with $F = 8.0$. This result occurs because the filters can result in poor estimation accuracy in estimating the system state that is more chaotic and unstable by using a larger value of F . Table 3 also denotes that the filters regardless of the value of F have a similar computational cost, since the magnitude of F does not have a great impact on the computational time of the filters.

5. Conclusions

This paper presents an efficient data assimilation method called localized unscented transform Kalman filter (LUTKF) for nonlinear filtering applications. Unlike the EnKF that uses a random sampling strategy and makes assumptions regarding the linearity of measurement operators, the LUTKF estimates the error statistics of the model state (mean and covariance)

FIG. 15. Case C: (a),(c),(e) Prior state estimation errors from the true state and (b),(d),(f) posterior state estimates minus prior state estimates (analysis increments) as a function of time (data assimilation cycle) and space (model state variable) for the 40-variable L96 model and $h(\mathbf{x}) = \ln(|\mathbf{x}|)$ with uniformly distributed (less densely spaced) measurements (i.e., $N_z = 40$): (a),(b) LETKF for $N_e = 10$, (c),(d) LUTKF for $N_e = 3$, and (e),(f) SPKF for $N_e = 241$.

through deterministically chosen samples (called sigma points) using the scaled unscented transformation (SUT) without making any assumptions for nonlinear observation operators.

In a broad sense, the LUTKF can be thought of as a variant of the localized EnKF (such as the LETKF) with the specific sample selection approach (i.e., SUT). Similar to localized EnKFs, the LUTKF can reduce the impact of observations on distant state variables by using the localization approach to suppress spurious correlations between distant positions in the background

covariance; meanwhile, the model state estimate for the SPKF is executed for the global domain. The spatial local analysis within neighboring regions of each model state variable makes it possible to use far fewer measurements than the SPKF, and it enables the LUTKF to perform data assimilation with a smaller ensemble size than the SPKF.

In addition to avoiding the global analysis of the SPKF, the LUTKF is designed to carry out local analysis for the nonlinear dynamical model and observation model with additive noise. Therefore, while the system state for the SPKF is based on an augmented form in

TABLE 3. RMSE, correlation, and computational time of each assimilation method for Case C.

Assimilation method	$F = 8.0$			$F = 8.1$		
	RMSE	Correlation	Computational time (s)	RMSE	Correlation	Computational time (s)
LETKF (using 3 ensembles)	4.712	0.172	3.35	4.785	0.169	3.32
LETKF (using 10 ensembles)	2.369	0.781	4.65	2.483	0.767	4.58
LUTKF (using 3 ensembles)	2.592	0.743	3.83	2.711	0.733	4.06
SPKF (using 241 ensembles)	1.974	0.843	85.81	2.083	0.830	86.72

which the model state vector is redefined as the concatenation of the original model state, model error, and observation error, the LUTKF does not need to augment the model state with the random noise variables. The use of the nonaugmented state vector can reduce the number of dimensions of the sigma points as well as the total number of sigma points used.

Because the LUTKF uses both the nonaugmented form and the localization scheme, another advantage of the LUTKF is that it is computationally affordable for nonlinear filtering applications, and its data assimilation calculations can be easily parallelized. In terms of computing time, the LUTKF can provide substantial benefits over the SPKF that executes a global analysis.

The LUTKF approach introduced in this manuscript may be problematic when applied to geophysical data assimilation applications. Similar to localized EnKFs, the local analysis does not ensure that physical balances are preserved in the posterior model state. This issue can be worsened as the localization length scale decreases, as demonstrated by Mitchell et al. (2002). Moreover, the LUTKF assumes that observations are uncorrelated in order to localize a posterior solution in a manner similar to covariance localization in localized EnKFs, such as the LETKF approach used here. This assumption may limit the effectiveness of the filters in assimilating remotely sensed data, such as satellite retrievals (Stewart et al. 2008). In the case of infrared satellite radiance data, off-diagonal components of observation error covariance can be significant (Bormann et al. 2016); in this situation, further consideration is required regarding the feasibility of the LUTKF approach.

Results from the cycling data assimilation experiments with the 40-variable Lorenz model show that the LUTKF requires only three ensemble members (i.e., sigma points) to accurately estimate the error statistics of the model. With the linear observation operators, the LUTKF yields lower prior mean RMSEs than the LETKF (with localization and adaptive inflation) for ensemble sizes smaller than eight. The LUTKF is most beneficial in applications in which dense observations related nonlinearly to the model state are assimilated. In this case, the LUTKF offers considerable benefits over the LETKF and results that are comparable to the

SPKF. These results suggest that the LUTKF has the potential to become an effective method to assimilate observations into realistic geophysical systems that require strongly nonlinear observation operators; for example, the assimilation of remotely sensed data, such as satellite radiances or radar reflectivity.

We have already begun examining the feasibility of the LUTKF in larger models. Although not discussed in this study, the LUTKF requires a much smaller ensemble size than the LETKF and SPKF to estimate the error statistics of the model in cycling data assimilation experiments carried out with a realistic atmospheric model that includes 70 762 328 variables. Furthermore, we have already studied the LUTKF algorithm that can scale the ensemble size to more accurately approximate the error statistics of the realistic model, called a local scalable unscented transform Kalman filter (LSUTKF). Data assimilation experiments using this model will be the topic of a future study that will explore the estimation

FIG. 16. As in Fig. 14, but for $F = 8.1$.

accuracy, numerical stability, and computational difficulty for high-dimensional problems.

Acknowledgments. We thank Dr. Jeffrey Anderson for helpful and insightful comments on this work. This work was part of the R&D project on the development of global numerical weather prediction systems of the Korea Institute of Atmospheric Prediction Systems (KIAPS) funded by the Korea Meteorological Administration (KMA) and supported by 2019 Research Fund of Myongji University.

APPENDIX

Kalman Gain and Analysis Covariance for the SPKF

The error in the posterior state estimation is defined as

$$\begin{aligned}
 \mathbf{P}_{k|k} &= E[\mathbf{e}_{k|k} \mathbf{e}_{k|k}^T], \\
 &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T], \\
 &= E\{[\mathbf{x}_k - (\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{z}}_k)][\mathbf{x}_k - (\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{z}}_k)]^T\}, \\
 &= E[(\mathbf{e}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_k)(\mathbf{e}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_k)^T], \\
 &= E[(\mathbf{e}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_k)(\mathbf{e}_{k|k-1}^T - \tilde{\mathbf{z}}_k^T \mathbf{K}_k^T)], \\
 &= E[\mathbf{e}_{k|k-1} \mathbf{e}_{k|k-1}^T - \mathbf{e}_{k|k-1} \tilde{\mathbf{z}}_k^T \mathbf{K}_k^T - \mathbf{K}_k \tilde{\mathbf{z}}_k \mathbf{e}_{k|k-1}^T + \mathbf{K}_k \tilde{\mathbf{z}}_k \tilde{\mathbf{z}}_k^T \mathbf{K}_k^T], \\
 &= E[\mathbf{e}_{k|k-1} \mathbf{e}_{k|k-1}^T] - E[\mathbf{e}_{k|k-1} \tilde{\mathbf{z}}_k^T] \mathbf{K}_k^T - \mathbf{K}_k E[\tilde{\mathbf{z}}_k \mathbf{e}_{k|k-1}^T] + \mathbf{K}_k E[\tilde{\mathbf{z}}_k \tilde{\mathbf{z}}_k^T] \mathbf{K}_k^T,
 \end{aligned} \tag{A3}$$

where $\mathbf{e}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$ and $\tilde{\mathbf{z}}_k = \mathbf{z}_k - \hat{\mathbf{z}}_k$.

Similar to the posterior estimation error $\mathbf{e}_{k|k}$, assuming that both the prior estimation error $\mathbf{e}_{k|k-1}$ and measurement residual $\tilde{\mathbf{z}}_k$ are unbiased, the cross covariance matrices $\mathbf{P}_{xz}$ (between $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{z}}_k$) and $\mathbf{P}_{zx}$ (between $\hat{\mathbf{z}}_k$ and $\hat{\mathbf{x}}_{k|k-1}$) can be expressed as

$$\mathbf{P}_{xz} = E[\mathbf{e}_{k|k-1} \tilde{\mathbf{z}}_k^T], \tag{A4}$$

$$\mathbf{P}_{zx} = E[\tilde{\mathbf{z}}_k \mathbf{e}_{k|k-1}^T]. \tag{A5}$$

Substitution of Eqs. (A4) and (A5) into Eq. (A3) yields the following analysis covariance:

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P}_{xz} \mathbf{K}_k^T - \mathbf{K}_k \mathbf{P}_{zx} + \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T. \tag{A6}$$

Next we find the Kalman gain matrix $\mathbf{K}_k$ that minimizes the trace of the covariance $\mathbf{P}_{k|k}$. The trace of the analysis covariance $\text{Tr}(\mathbf{P}_{k|k})$ is minimized when its matrix derivative with respect to the gain matrix is

$$\mathbf{e}_{k|k} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}. \tag{A1}$$

Because the Kalman filter is a minimum mean square error (MMSE) estimator, we want to minimize the mathematical expectation of the square of the magnitude of the estimation error, which is given by $E[\|\mathbf{e}_{k|k}\|_2^2]$.⁸ This is the same as minimizing the trace of the posterior estimation error covariance (analysis covariance) $\mathbf{P}_{k|k}$ calculated by

$$\begin{aligned}
 \mathbf{P}_{k|k} &= E[(\mathbf{e}_{k|k} - E[\mathbf{e}_{k|k}])(\mathbf{e}_{k|k} - E[\mathbf{e}_{k|k}])^T], \\
 &= E[\mathbf{e}_{k|k} \mathbf{e}_{k|k}^T] - E[\mathbf{e}_{k|k}] E[\mathbf{e}_{k|k}]^T.
 \end{aligned} \tag{A2}$$

Assuming that the estimation error $\mathbf{e}_{k|k}$ is unbiased (i.e., $E[\mathbf{e}_{k|k}] = 0$), the analysis covariance $\mathbf{P}_{k|k}$ is rewritten as

zero. Using the gradient matrix rules, this can be represented by

$$\frac{\partial \text{Tr}(\mathbf{P}_{k|k})}{\partial \mathbf{K}_k} = -2\mathbf{P}_{xz} + 2\mathbf{K}_k \mathbf{S}_k = 0. \tag{A7}$$

From Eq. (A7), the optimal Kalman gain that yields MMSE estimates is calculated by

$$\mathbf{K}_k = \mathbf{P}_{xz} \mathbf{S}_k^{-1}. \tag{A8}$$

By replacing the Kalman gain in Eq. (A6) with the optimal value derived above, the optimal expression for analysis covariance is given by

$$\begin{aligned}
 \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{xz}^T, \\
 &= \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T.
 \end{aligned} \tag{A9}$$

⁸ Means the two-norm of a vector, also known as the Euclidean norm, and is defined as the length or magnitude of a vector as $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{x_1^2 + \dots + x_n^2}$.

The use of the optimal Kalman gain in Eq. (A8) and analysis covariance in Eq. (A9) ensure that the SPKF estimates the model state using the Kalman filter equations regardless of the nonlinearity of the measurement model.

REFERENCES

- Ambadan, J. T., and Y. Tang, 2009: Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems. *J. Atmos. Sci.*, **66**, 261–285, <https://doi.org/10.1175/2008JAS2681.1>.
- 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](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2).
- Bormann, N., M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteor. Soc.*, **142**, 1767–1780, <https://doi.org/10.1002/qj.2774>.
- Bshara, M., U. Orguner, F. Gustafsson, and L. V. Biesen, 2010: Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on WiMAX networks. *IEEE Trans. Veh. Technol.*, **59**, 283–294, <https://doi.org/10.1109/TVT.2009.2030504>.
- Chu, P.-S., X. Zhao, C.-H. Ho, H.-S. Kim, M.-M. Lu, and J.-H. Kim, 2010: Bayesian forecasting of seasonal typhoon activity: A track-pattern-oriented categorization approach. *J. Climate*, **23**, 6654–6668, <https://doi.org/10.1175/2010JCLI3710.1>.
- Crassidis, J. L., and F. L. Markley, 2003: Unscented filtering for spacecraft attitude estimation. *J. Guid. Control Dyn.*, **26**, 536–542, <https://doi.org/10.2514/2.5102>.
- Cui, B., X. Chen, and X. Tang, 2017: Improved cubature Kalman filter for GNSS/INS based on transformation of posterior sigma-points error. *IEEE Trans. Signal Process.*, **65**, 2975–2987, <https://doi.org/10.1109/TSP.2017.2679685>.
- Deng, Z., Y. Tang, and G. Wang, 2010: Assimilation of Argo temperature and salinity profiles using a bias-aware localized EnKF system for the Pacific Ocean. *Ocean Modell.*, **35**, 187–205, <https://doi.org/10.1016/j.ocemod.2010.07.007>.
- 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.
- Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97**, 17 905–17 924, <https://doi.org/10.1029/92JC01972>.
- , 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>.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- , and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867, [https://doi.org/10.1175/1520-0493\(2000\)128<1852:AEKSFN>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2).
- Farina, A., B. Ristic, and D. Benvenuti, 2002: Tracking a ballistic target: Comparison of several nonlinear filters. *IEEE Trans. Aerosp. Electron. Syst.*, **38**, 854–867, <https://doi.org/10.1109/TAES.2002.1039404>.
- 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>.
- 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, <https://doi.org/10.1017/CBO9780511617652.007>.
- Hamon, M., E. Greiner, P.-Y. Le Traon, and E. Remy, 2019: Impact of multiple altimeter data and mean dynamic topography in a global analysis and forecasting system. *J. Atmos. Oceanic Technol.*, **36**, 1255–1266, <https://doi.org/10.1175/JTECH-D-18-0236.1>.
- Hocking, J., P. Rayer, R. Saunders, M. Matricardi, A. Geer, and P. Brunel, 2011: RTTOV v10 users' guide. EUMETSAT Satellite Application Facility on Numerical Weather Prediction, 92 pp.
- 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 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>.
- 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>.
- Ito, K., and K. Xiong, 2000: Gaussian filters for nonlinear filtering problems. *IEEE Trans. Autom. Control*, **45**, 910–927, <https://doi.org/10.1109/9.855552>.
- Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, <https://doi.org/10.1175/MWR-D-14-00345.1>.
- Julier, S. J., and J. K. Uhlmann, 1996: A general method for approximating nonlinear transformations of probability distributions. Tech. Rep., Robotics Research Group, Department of Engineering Science, University of Oxford, Oxford, United Kingdom, 27 pp.
- , and —, 2002: Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. *Proc. 2002 American Control Conf.*, Anchorage, AK, IEEE, 887–892, <https://doi.org/10.1109/ACC.2002.1023128>.
- , and —, 2004: Unscented filtering and nonlinear estimation. *Proc. IEEE*, **92**, 401–422, <https://doi.org/10.1109/JPROC.2003.823141>.
- , J. Uhlmann, and H. F. Durrant-Whyte, 2000: A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Autom. Control*, **45**, 477–482, <https://doi.org/10.1109/9.847726>.
- Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971–1981, [https://doi.org/10.1175/1520-0493\(2000\)128<1971:DAIAPE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1971:DAIAPE>2.0.CO;2).
- Krishnamoorthy, A., and D. Menon, 2013: Matrix inversion using Cholesky decomposition. *Proc. IEEE Signal Processing*

- Algorithms Architectures, Arrangements, and Applications*, Poznan, Poland, IEEE, 70–72.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Shinfield Park, Reading, ECMWF, 1–18.
- , and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, [https://doi.org/10.1175/1520-0469\(1998\)055<0399:OSFSWO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2).
- Luo, X., and I. M. Moroz, 2009: Ensemble Kalman filter with the unscented transform. *Physica D*, **238**, 549–562, <https://doi.org/10.1016/j.physd.2008.12.003>.
- Manoj, K. K., Y. Tang, Z. Deng, D. Chen, and Y. Cheng, 2014: Reduced-rank sigma-point Kalman filter and its application in ENSO model. *J. Atmos. Oceanic Technol.*, **31**, 2350–2366, <https://doi.org/10.1175/JTECH-D-13-00172.1>.
- Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1056, [https://doi.org/10.1175/1520-0469\(1994\)051<1037:ADAISN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<1037:ADAISN>2.0.CO;2).
- 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](https://doi.org/10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2).
- Oke, P. R., A. Schiller, D. A. Griffin, and G. B. Brassington, 2005: Ensemble data assimilation for an eddy-resolving ocean model of the Australian region. *Quart. J. Roy. Meteor. Soc.*, **131**, 3301–3311, <https://doi.org/10.1256/qj.05.95>.
- , G. B. Brassington, D. A. Griffin, and A. Schiller, 2008: The Bluelink ocean data assimilation system (BODAS). *Ocean Modell.*, **21**, 46–70, <https://doi.org/10.1016/j.ocemod.2007.11.002>.
- 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>.
- 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>.
- Robert, C. P., 1995: Simulation of truncated normal variables. *Stat. Comput.*, **5**, 121–125, <https://doi.org/10.1007/BF00143942>.
- Simon, D., 2006: *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 526 pp.
- Stewart, L. M., S. L. Dance, and N. K. Nichols, 2008: Correlated observation errors in data assimilation. *Int. J. Numer. Methods Fluids*, **56**, 1521–1527, <https://doi.org/10.1002/fld.1636>.
- Tang, Y., Z. Deng, K. K. Manoj, and D. Chen, 2014: A practical scheme of the sigma-point Kalman filter for high-dimensional systems. *J. Adv. Model. Earth Syst.*, **6**, 21–37, <https://doi.org/10.1002/2013MS000255>.
- van der Merwe, R., 2004: Sigma-point Kalman filters for probabilistic inference in dynamic state-space models. Ph.D. thesis, Oregon Health & Science University, 397 pp.
- van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114, <https://doi.org/10.1175/2009MWR2835.1>.
- Wan, E. A., and R. van der Merwe, 2000: The unscented Kalman filter for nonlinear estimation. *Proc. 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (ASSPCC)*, Lake Louise, AB, IEEE, 153–158, <https://doi.org/10.1109/ASSPCC.2000.882463>.
- 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).
- , and —, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, <https://doi.org/10.1175/MWR-D-11-00276.1>.
- , —, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463–482, <https://doi.org/10.1175/2007MWR2018.1>.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, [https://doi.org/10.1175/1520-0493\(2004\)132<1238:IOIEAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2).