Sigma-Point Kalman Filter Data Assimilation Methods for Strongly Nonlinear Systems

JAISON THOMAS AMBADAN AND YOUMIN TANG

Environmental Science and Engineering, University of Northern British Columbia, Prince George, British Columbia, Canada

ABSTRACT

Performance of an advanced, derivativeless, sigma-point Kalman filter (SPKF) data assimilation scheme in a strongly nonlinear dynamical model is investigated. The SPKF data assimilation scheme is compared against standard Kalman filters such as the extended Kalman filter (EKF) and ensemble Kalman filter (EnKF) schemes. Three particular cases—namely, the state, parameter, and joint estimation of states and parameters from a set of discontinuous noisy observations—are studied. The problems associated with the use of tangent linear model (TLM) or Jacobian when using standard Kalman filters are eliminated when using SPKF data assimilation algorithms. Further, the constraints and issues of SPKF data assimilation in real ocean or atmospheric models are emphasized. A reduced sigma-point subspace model is proposed and investigated for higher-dimensional systems.

A low-dimensional Lorenz 1963 model and a higher-dimensional Lorenz 1995 model are used as the test bed for data assimilation experiments. The results of SPKF data assimilation schemes are compared with those of standard EKF and EnKF, in which a highly nonlinear chaotic case is studied. It is shown that the SPKF is capable of estimating the model state and parameters with better accuracy than EKF and EnKF. Numerical experiments showed that in all cases the SPKF can give consistent results with better assimilation skills than EnKF and EKF and can overcome the drawbacks associated with the use of EKF and EnKF.

1. Introduction

Among sequential data assimilation methods, Kalman filters have been widely used in meteorology and oceanography. The standard Kalman Filter (KF) is a simplification of Bayesian estimation that provides sequential, unbiased, minimum error variance estimates based on a linear combination of all past measurements and dynamics (Welch and Bishop 1995). Since the introduction of the extended Kalman filter (EKF), the nonlinear extension to the standard KF, there have been many attempts to use the EKF in weather or climate prediction models. It has been shown that the EKF can be used in sequential data assimilation in strongly nonlinear systems (Miller et al. 1994). Unfortunately, the requirement of the Jacobian or tangent linear model (TLM) for the linearization of nonlinear functions limits the use of EKF for many real world problems. Another major drawback of EKF is that it only uses the first-order terms of the Taylor expansion of the nonlinear function. It is evident that this approximation often introduces large errors in the estimation of covariance matrices in highly nonlinear models (Miller et al. 1994). In other words, the inaccuracy of propagated means and covariances resulting from the linearization of the nonlinear model is one of the major drawbacks of the EKF data assimilation algorithm.

Another alternative to the standard KF is the ensemble Kalman filter (EnKF), introduced by Evensen (Evensen 1992; Houtekamer and Mitchell 1998), in which the error covariances are estimated approximately using an ensemble of model forecasts. The main concept behind the formulation of the EnKF is that if the dynamical model is expressed as a stochastic differential equation, the prediction error statistics, which are described by the Fokker–Plank equation, can be estimated using ensemble integrations (Evensen 1994, 1997); thus, the error covariance matrices can be calculated by integrating the ensemble of model states. The EnKF can overcome the EKF drawback that neglects the contributions from higher-order statistical moments in calculating the error covariance. The major strengths of the EnKF include the following: (i) there is no need

Corresponding author address: Youmin Tang, Environmental Science and Engineering, UNBC, 3333 University Way, Prince George, BC V2N 4Z9, Canada.
E-mail: ytang@unbc.ca

to calculate the tangent linear model or Jacobian of nonlinear models, which is extremely difficult for ocean (or atmosphere) general circulation models (GCMs); (ii) the covariance matrix is propagated in time via fully nonlinear model equations (no linear approximation as in the EKF); and (iii) it is well suited to modern parallel computers (cluster computing) (Keppenne 2000).

The finite ensemble size has major effects on the performance of the EnKF. A small ensemble size increases the residual errors and gives inaccurate statistical moments, and a large ensemble size is not computationally feasible in the case of atmospheric or ocean GCMs. Another disadvantage of the EnKF is that it assumes a linear measurement operator: if the measurement function is nonlinear, it has to be linearized in the EnKF. The nonlinear measurement functions appear in many situations: for example, the parameter estimation of nonlinear dynamical models, in which the measurement relationships between observations and parameters are nonlinear. Another example is satellite altimetry data assimilation, in which the observation (sea level height) is often nonlinearly related to the variable required for assimilation (e.g., temperature). Thus, the condition of linear measurement limits the use of the EnKF in some real world problems.

The sigma-point Kalman filters (SPKFs; van der Merwe et al. 2004) have recently been proposed in an attempt to address these drawbacks of EKF and EnKF. The SPKF is a *derivativeless* sequential optimal estimation method, using a novel deterministic sampling approach that eliminates the need for the calculation of TLM or the Jacobian of the model equations as needed by the standard KF (Julier et al. 1995; Nørgaard et al. 2000b; Ito and Xiong 2000; Lefebvre et al. 2002; Wan and van der Merwe 2000; Haykin 2001; van der Merwe 2004). It has been found that the expected error due to linearization is smaller than that of a truncated Taylor series linearization (Schei 1997; Lefebvre et al. 2002; van der Merwe and Wan 2001a). The SPKF algorithm has been successfully implemented in many areas such as robotics, artificial intelligence, natural language processing, and global positioning systems navigation (van der Merwe and Wan 2001a,b; Haykin 2001; van der Merwe 2004; van der Merwe et al. 2004; Wan and van der Merwe 2000). In this paper, we will show that SPKF, as an ensemble Kalman filter with a specific ensemble, has a great potential in the assimilation of nonlinear systems. This paper is meant as a major effort in exploring the possibility of applying SPKF in atmospheric and oceanic data assimilation.

This paper is structured as follows: Section 2 introduces the sigma-point methodology and section 3 describes SPKF implementation in the highly nonlinear Lorenz model. Section 4 describes the experimental details and gives a detailed comparison of SPKF with EKF and EnKF. Section 5 describes the SPKF implementation in higher-dimensional systems. Section 6 summarizes the conclusions.

2. Methodology: Sigma-point Kalman filters

In this section we will interpret the sigma-point concept and SPKF algorithms in detail. The so-called sigma-point approach is based on deterministic sampling of state distribution to calculate the approximate covariance matrices for the standard Kalman filter equations. The family of SPKF algorithms includes the unscented Kalman filter (UKF; Julier et al. 1995; Wan and van der Merwe 2000), the central difference Kalman filter (CDKF; Nørgaard et al. 2000b; Ito and Xiong 2000), and their square root versions (Haykin 2001; van der Merwe and Wan 2001a,b). Another interpretation of the sigma-point approach is that it implicitly performs a statistical linearization (Gelb 1974; Lefebvre et al. 2002) of the nonlinear model through a weighted statistical linear regression (WSLR) to calculate the covariance matrices (van der Merwe and Wan 2001a,b; van der Merwe et al. 2004). In SPKF, the model linearization is done through a linear regression between n number of points (called sigma points) drawn from a prior distribution of a random variable rather than through a truncated Taylor series expansion at a single point (van der Merwe et al. 2004). It has been found that this linearization is much more accurate than a truncated Taylor series linearization (Schei 1997; Lefebvre et al. 2002; van der Merwe and Wan 2001a).

Consider an L-dimensional dynamical system represented by a set of discretized state space equations,

$$\boldsymbol{\theta}_k = f(\boldsymbol{\theta}_{k-1}, q_{k-1}), \tag{1}$$

$$\boldsymbol{\psi}_k = h(\boldsymbol{\theta}_k, r_k), \tag{2}$$

where $\boldsymbol{\theta}_k$ represents the system state vector at time k, $f(\cdot)$ is the nonlinear function of the state, q_k is the random (white) model errors, $\boldsymbol{\psi}_k$ is the measured state, $h(\cdot)$ is the measurement function, and r_k is the zero-mean random measurement noise.

To understand how the sigma-point algorithm works, it is convenient to rewrite the standard Kalman filter optimal state update equation:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_k^- + \mathbf{K}_k(\boldsymbol{\psi}_k - \hat{\boldsymbol{\psi}}_k^-), \tag{3}$$

where the superscript negative sign represents the prior or forecast states. Here, $\mathbf{K}_k$ is the Kalman gain, which is optimally chosen such that it minimizes the weighted scalar sum of the diagonal elements of the error covari-

ance matrix $\mathbf{P}_{\theta_k}^-$ (Gelb 1974). The standard expression for the Kalman gain and the error covariance matrix is given by

$$\mathbf{K}_k = \mathbf{P}_{\theta_k}^- \mathbf{H}^T (\mathbf{H} \mathbf{P}_{\theta_k}^- \mathbf{H}^T + R)^{-1}, \qquad (4)$$

$$\mathbf{P}_{\theta_k}^- = E[(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)^T], \qquad (5)$$

where $\mathbf{H}$ is the linearized measurement operator, R is the observation error covariance matrix, and $E[\cdot]$ represents the mathematical expectation or the expected value. The error covariance update or the analysis covariance matrix, which represents the change in forecast error covariance when a measurement is employed, is given by

$$\mathbf{P}_{\theta_k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}_{\theta_k}^-, \qquad (6)$$

where $\mathbf{I}$ is the identity matrix. For EKF, the formulation of the forecast error covariance is given by

$$\mathbf{P}_{\theta_k}^- = \mathbf{A}_{k-1} \mathbf{P}_{\theta_{k-1}} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}, \qquad (7)$$

where $\mathbf{A}_{k-1}$ is the TLM of the nonlinear model (1) and (2) and $\mathbf{Q}_{k-1}$ is the model error covariance matrix. The TLM often introduces errors in highly nonlinear models and is extremely difficult to obtain for GCMs. Another major drawback of EKF is that it uses the linearized measurement operator $\mathbf{H}$ to calculate the Kalman gain and update error covariance. The linearization of nonlinear measurement is computationally difficult and may result in large estimation errors.

On the other hand, the forecast error covariance for EnKF can be calculated by integrating the ensemble of model states, and is given by

$$\mathbf{P}_{\theta_k}^- \cong \overline{(\boldsymbol{\theta}_k - \bar{\boldsymbol{\theta}}_k)(\boldsymbol{\theta}_k - \bar{\boldsymbol{\theta}}_k)^T}, \qquad (8)$$

where the overbar denotes the ensemble average. Limited ensemble size often introduces errors in approximating the error covariance matrix. The EnKF also uses the linearized measurement operator $\mathbf{H}$ to calculate the Kalman gain. If the model involves nonlinear measurement functions, linearizing the nonlinear measurement functions might result in large errors.

The SPKF family addresses the above issues of EKF and EnKF. It uses a different approach in calculating the Kalman gain and the error covariance matrices. The technique employed in SPKF is to reinterpret the standard Kalman gain and covariance update equation in such a way that it does not need the TLM and the linearized measurement operator. This interpretation is explained below.

The first term $\mathbf{P}_{\theta_k}^- \mathbf{H}^T$ in Kalman gain Eq. (4) can be interpreted as the cross-covariance $\mathbf{P}_{\theta_k \tilde{\psi}_k}$ between the state and observation errors, and the remaining expression can be interpreted as the error covariance $\mathbf{P}_{\tilde{\psi}_k}$ of the difference between model and observation (Gelb

1974). Proof of this interpretation can be found in appendix A.[1] Therefore, the optimal gain or Kalman gain $\mathbf{K}_k$ can be rewritten as

$$\mathbf{K}_k = \mathbf{P}_{\theta_k \tilde{\psi}_k} \mathbf{P}_{\tilde{\psi}_k}^{-1}. \qquad (9)$$

Here, $\tilde{\boldsymbol{\psi}}_k$ is defined as the error between the noisy observation $\boldsymbol{\psi}_k$ and its prediction $\hat{\boldsymbol{\psi}}_k^-$ given by $\tilde{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k - \hat{\boldsymbol{\psi}}_k^-$. By using relation (9), the covariance update Eq. (6) can be rewritten as (see appendix B for details)

$$\mathbf{P}_{\theta_k} = \mathbf{P}_{\theta_k}^- - \mathbf{K}_k \mathbf{P}_{\tilde{\psi}_k} \mathbf{K}_k^T. \qquad (10)$$

Unlike the standard KF, the SPKF algorithm makes use of this new interpretation [Eqs. (9) and (10)], which avoids the use of the Jacobian while retaining consistency and accuracy.

In the standard KF the state error covariance is calculated during the time update process and is updated during the measurement update process. Updating the error covariance matrix is important because it represents the change in forecast error covariance when a measurement is performed. The EnKF implementation does not require the covariance update equation because it can directly calculate the updated error covariance matrix from a set of ensembles. Evensen (2003) has derived the analysis covariance equation, which is consistent with the standard KF error covariance update Eq. (6). But the true representation of the updated error covariance requires a large ensemble size, which is often computationally infeasible. The SPKF makes use of the reformulated error covariance to update Eq. (10) and chooses the ensembles deterministically in such a way that they can capture the statistical moments of the nonlinear model accurately; in other words, the forecast error covariance Eq. (5) is computed using deterministically chosen samples, called sigma points. In a broad sense, the SPKF algorithm implicitly uses the prior covariance update equation (or the analysis error covariance matrix) to calculate the forecast error covariance. Thus, SPKF is fully consistent with the time update and measurement update formulation of the Kalman filter algorithm. In the next subsection we will discuss each SPKF algorithm in detail.

a. Sigma-point unscented Kalman filter (SP-UKF)

The sigma-point unscented Kalman filter (SP-UKF) (Julier et al. 1995; Julier 1998; Wan and van der Merwe 2000) is an SPKF that can capture the statistical properties of the model state through a method known as scaled unscented transformation (SUT; Julier 2002).

---

[1] A more detailed statistical derivation and interpretation of these formulations can be found in Simon (2006).

Unlike the EKF, the SP-UKF uses the true nonlinear model and approximates the state distribution using a set of deterministically chosen states, known as sigma points, using SUT. In SP-UKF the state error covariance matrix is calculated from a set of particular ensembles that are generated by sigma points. Julier et al. (1995) have shown that for the nonlinear model given by (1), the number of sigma points needed to compute precisely the mean and covariance of the model state at time k, is $2L + 1$: thus, the *sigma-point state vector* is given by (Julier et al. 1995; Julier 2002; Wan and van der Merwe 2000)

$$\boldsymbol{\chi}_k = [\boldsymbol{\chi}_{k,0} \, \boldsymbol{\chi}_{k,i}^{+} \, \boldsymbol{\chi}_{k,i}^{-}] \quad i = 1, \ldots, L, \qquad (11)$$

where $\boldsymbol{\chi}_{k,0}$, $\boldsymbol{\chi}_{k,i}^{+}$, and $\boldsymbol{\chi}_{k,i}^{-}$ are the sigma-point vectors. The selection scheme for choosing the sigma points is based on the scaled unscented transformation that transforms the model state vector according to the following equations:

$$\boldsymbol{\chi}_{k,0} = \boldsymbol{\theta}_k \quad \text{is based on} \quad w_0^{(m)} = \frac{\lambda}{(L+\lambda)}; \qquad (12)$$

$$\boldsymbol{\chi}_{k,i}^{+} = \boldsymbol{\theta}_k + [\sqrt{(L+\lambda)\mathbf{P}_{\boldsymbol{\theta}_k}}]_i \quad i = 1, \ldots, L \quad \text{is based on}$$
$$w_0^{(c)} = \frac{\lambda}{(L+\lambda)} + (1 - \alpha^2 + \beta); \quad \text{and} \qquad (13)$$

$$\boldsymbol{\chi}_{k,i}^{-} = \boldsymbol{\theta}_k - [\sqrt{(L+\lambda)\mathbf{P}_{\boldsymbol{\theta}_k}}]_i, \quad \text{where}$$
$$i = (L+1), \ldots, 2L \quad \text{with corresponding weight,}$$
$$w_i^{(m)} = w_i^{(c)} = \frac{1}{2(L+\lambda)}, \quad \text{where} \quad i = 1, \ldots, 2L; \qquad (14)$$

where $[\sqrt{(L+\lambda)\,\mathbf{P}_{\boldsymbol{\theta}_k}}]_i$ is the ith row (or column) of the weighted matrix square root of the covariance matrix, $\mathbf{P}_{\boldsymbol{\theta}_k}$. $w_i^{(m)}$ is the weighting term corresponding to the mean, $w_i^{(c)}$ corresponds to the covariance, and $\lambda = \alpha^2 (L + \kappa) - L$ is a scaling parameter. The parameter α is set to a small positive value ($0 \leq \alpha \leq 1$) and determines the spread of the sigma points around the mean state $\boldsymbol{\theta}_k$. Another control parameter is κ, which guarantees the positive semidefiniteness of the covariance matrix and is usually set to a positive value ($\kappa \geq 0$); β is a nonnegative weighting term that can be used to incorporate any prior knowledge of the nature of the state distribution.[2]

The sigma-point vector is then propagated through the nonlinear model (1) and (2) given by

$$\boldsymbol{\chi}_k^{\theta} = f(\boldsymbol{\chi}_{k-1}^{\theta}, \boldsymbol{\chi}_{k-1}^{q}), \qquad (15)$$

$$\mathcal{Y}_k^{\theta} = h(\boldsymbol{\chi}_k^{\theta}, \boldsymbol{\chi}_{k-1}^{r}), \qquad (16)$$

where $\boldsymbol{\chi}_k^{\theta}$ is the forecast sigma-point state vector, $\boldsymbol{\chi}_{k-1}^{q}$ is the sigma-point vector corresponding to the model error, and $\boldsymbol{\chi}_{k-1}^{r}$ corresponds to the observation error. The approximated mean, covariance, and cross-covariance for the calculation of Kalman gain are computed as follows (Julier et al. 1995; Julier 2002; Wan and van der Merwe 2000, 2001):

$$\hat{\boldsymbol{\theta}}_k^{-} \approx \sum_{i=0}^{2L} w_i^{(m)} \boldsymbol{\chi}_{k,i}^{\theta}, \qquad (17)$$

$$\hat{\boldsymbol{\psi}}_k^{-} \approx \sum_{i=0}^{2L} w_i^{(m)} \mathcal{Y}_{k,i}^{\theta}, \qquad (18)$$

$$\mathbf{P}_{\boldsymbol{\theta}_k}^{-} \approx \sum_{i=0}^{2L} w_i^{(c)} (\boldsymbol{\chi}_{k,i}^{\theta} - \hat{\boldsymbol{\theta}}_k^{-})(\boldsymbol{\chi}_{k,i}^{\theta} - \hat{\boldsymbol{\theta}}_k^{-})^{\mathrm{T}}, \qquad (19)$$

$$\mathbf{P}_{\tilde{\boldsymbol{\psi}}_k} \approx \sum_{i=0}^{2L} w_i^{(c)} (\mathcal{Y}_{k,i} - \hat{\boldsymbol{\psi}}_k^{-})(\mathcal{Y}_{k,i} - \tilde{\boldsymbol{\psi}}_k^{-})^{\mathrm{T}}, \quad \text{and} \quad (20)$$

$$\mathbf{P}_{\boldsymbol{\theta}_k \tilde{\boldsymbol{\psi}}_k} \approx \sum_{i=0}^{2L} w_i^{(c)} (\boldsymbol{\chi}_{k,i}^{\theta_k} - \hat{\boldsymbol{\theta}}_k^{-})(\mathcal{Y}_{k,i} - \hat{\boldsymbol{\psi}}_k^{-})^{\mathrm{T}}. \qquad (21)$$

The Kalman gain **K** can be calculated using Eq. (9) and the state covariance is updated using Eq. (10). A detailed description and derivation of the UKF algorithm and sigma-point formulation can be found in the above referenced literature.

b. Sigma-point central difference Kalman filter (SP-CDKF)

Unlike the SP-UKF, the SP-CDKF is based on Stirling's interpolation formulas (Ito and Xiong 2000; Nørgaard et al. 2000b; Press et al. 1992). In SP-CDKF the analytical derivatives in EKF are replaced by numerically evaluated central divided differences. The linearization in SP-CDKF is based on weighted statistical linear regression (Lefebvre et al. 2002). It has been shown that the SP-CDKF has superior performance to SP-UKF and EKF (Ito and Xiong 2000; Nørgaard et al. 2000b).[3] Using Sterling's polynomial interpolation,

---

[2] The weighting term corresponding to the zeroth sigma point directly affects the magnitude of errors in higher-order moments for symmetric distributions (Julier 2002; van der Merwe et al. 2000). The parameter β is thus introduced to minimize the higher-order errors.

[3] However, our numerical experiments show that the SP-CDKF does not always outperform SP-UKF. See the following discussions.

the nonlinear model given by Eq. (1) can be approximated as

$$f(\boldsymbol{\theta}_k) \approx \bar{f}(\bar{\boldsymbol{\varphi}}_k) + \tilde{D}_k + \frac{1}{2!}\tilde{D}_k^2, \qquad (22)$$

where $\tilde{f}(\bar{\varphi}_k)$ is the linearized model, and $\tilde{D}_k$ and $\tilde{D}_k^2$ are the central divided difference operators, which we will explain in the latter part in this section. Here the linearization of the nonlinear model is achieved by using a linear transformation that statistically decouples[4] the state vector $\boldsymbol{\theta}_k$ (Schei 1997). It has been shown that this approximation is always better than using the Jacobian matrix (Schei 1997). The linear transformation is based on the square root factorization of the model covariance matrix and is given by

$$\bar{\boldsymbol{\phi}}_k = \mathbf{S}_{\boldsymbol{\theta}_k}^{-1}\bar{\boldsymbol{\theta}}_k, \qquad (23)$$

$$\tilde{f}(\boldsymbol{\varphi}_k) \doteq f(\mathbf{S}_{\boldsymbol{\theta}_k}\bar{\boldsymbol{\varphi}}_k) = f(\bar{\boldsymbol{\theta}}_k). \qquad (24)$$

Here $\bar{\theta}_k$ is the mean state and $\mathbf{S}_{\boldsymbol{\theta}_k}$ is the Cholesky factor of the updated error covariance matrix (10) that satisfies the following condition:

$$\mathbf{P}_{\boldsymbol{\theta}_k} = \mathbf{S}_{\boldsymbol{\theta}_k}\mathbf{S}_{\boldsymbol{\theta}_k}^{\mathrm{T}}. \qquad (25)$$

The terms $\tilde{D}_k$ and $\tilde{D}_k^2$ are the first- and second-order central divided difference operators and can be written as (Ito and Xiong 2000; Nørgaard et al. 2000a; van der Merwe and Wan 2001a; Wan and van der Merwe 2001)

$$\tilde{D}_k = \left[\sum_{i=0}^{L}(\boldsymbol{\phi}_k - \bar{\boldsymbol{\phi}}_k)_i \boldsymbol{m}_i \boldsymbol{d}_i\right]\tilde{f}(\bar{\boldsymbol{\phi}}_k), \qquad (26)$$

$$\tilde{D}_k^2 = \left[\sum_{i=0}^{L}(\boldsymbol{\phi}_k - \bar{\boldsymbol{\phi}}_k)_i^2 \boldsymbol{d}_i^2 + \sum_{j=1}^{L}\sum_{\substack{q=0 \\ q \neq j}}^{L}(\boldsymbol{\phi}_k - \bar{\boldsymbol{\phi}}_k)_j(\boldsymbol{\phi}_k\right.$$
$$\left. - \bar{\phi}_k)_q(\boldsymbol{m}_j\boldsymbol{d}_j)(\boldsymbol{m}_q\boldsymbol{d}_q)\right]\tilde{f}(\bar{\boldsymbol{\phi}}_k), \qquad (27)$$

where $\boldsymbol{m}_i$, $\boldsymbol{d}_i$, and $\boldsymbol{d}_i^2$ are the mean, partial first-order, and partial second-order central divided difference operators respectively, defined as

$$\boldsymbol{m}_i\tilde{f}(\bar{\boldsymbol{\phi}}_k) = \frac{1}{2}[f(\bar{\boldsymbol{\theta}}_k + \delta\boldsymbol{s}_{\theta_i}) + f(\bar{\boldsymbol{\theta}}_k - \delta\boldsymbol{s}_{\theta_i})], \qquad (28)$$

$$\boldsymbol{d}_i\tilde{f}(\bar{\boldsymbol{\phi}}_k) = \frac{1}{2\delta}[f(\bar{\boldsymbol{\theta}}_k + \delta\boldsymbol{s}_{\theta_i}) - f(\bar{\boldsymbol{\theta}}_k - \delta\boldsymbol{s}_{\theta_i})], \quad \text{and} \qquad (29)$$

$$\boldsymbol{d}_i^2\tilde{f}(\bar{\boldsymbol{\phi}}_k) = \frac{1}{2\delta^2}[f(\bar{\boldsymbol{\theta}}_k + \delta\boldsymbol{s}_{\theta_i}) + f(\bar{\boldsymbol{\theta}}_k - \delta\boldsymbol{s}_{\theta_i}) - 2f(\bar{\boldsymbol{\theta}}_k)], \qquad (30)$$

where δ is the central difference step size and $\boldsymbol{s}_{\theta_i}$ is the ith column of the Cholesky factor of the covariance updated error covariance matrix (10)

$$\boldsymbol{s}_{\theta_i} = (\sqrt{\mathbf{P}_{\boldsymbol{\theta}_k}})_i. \qquad (31)$$

For implementing the SP-CDKF, augmented state vectors are constructed by concatenating the original model state, model, and observation error vectors. The augmented sigma-point state vectors are calculated using the following selection scheme:

$\boldsymbol{\chi}_{k,0} = \bar{\boldsymbol{\theta}}_k$　is based on　$w_0^{(m)} = \dfrac{\delta^2 - L}{\delta^2}$;

$\boldsymbol{\chi}_{k,i}^+ = \boldsymbol{\theta}_k + (\sqrt{\delta^2\mathbf{P}_{\boldsymbol{\theta}_k}})_i$,　where　$i = 1, \ldots, L$,　with corresponding weight,　$w_i^{(m)} = \dfrac{1}{2\delta^2}$,　where　$i = 1, \ldots, 2L$;

$\boldsymbol{\chi}_{k,i}^- = \boldsymbol{\theta}_k - (\sqrt{\delta^2\mathbf{P}_{\boldsymbol{\theta}_k}})_i$,　where　$i = (L+1), \ldots, 2L$,　with corresponding weight,　$w_i^{(c1)} = \dfrac{1}{4\delta^2}$,

　　where　$i = 1, \ldots, 2L$　and

$w_i^{(c2)} = \dfrac{\delta^2 - 1}{4\delta^4}$,　where　$i = 1, \ldots, 2L$. $\qquad (32)$

The augmented sigma-point vectors are then propagated through the approximated nonlinear model (22), and the approximated mean model state vector can be computed as follows:

$$\hat{\boldsymbol{\theta}}_k^- \approx E\left[\tilde{f}(\bar{\boldsymbol{\phi}}_k) + \tilde{D}_k + \frac{1}{2!}\tilde{D}_k^2\right], \qquad (33)$$

---

[4] The linear transformation from the stochastic vector $\bar{\phi}_k$ to $\boldsymbol{\theta}_k$: $\boldsymbol{\theta}_k = \boldsymbol{S}_{\boldsymbol{\theta}_k}\phi_k$, decouples the fully coupled state vector $\boldsymbol{\theta}_k$ where the covariance of $\bar{\phi}_k$ is equal to the identity matrix. For computational reasons the square root matrix $\boldsymbol{S}_{\boldsymbol{\theta}_k}$ often remains triangular (Cholesky decomposition). More details on decoupling and its advantages in Kalman filters can be found in Ohmuro (1984), Baheti et al. (1990), and Daum and Fitzgerald (1983)

$$\approx \frac{\delta^2 - L}{\delta^2} f(\hat{\boldsymbol{\theta}}_{k-1}) + \frac{1}{2h^2} \sum_{i=1}^{L} [f(\hat{\boldsymbol{\theta}}_{k-1} + \delta\boldsymbol{s}_{\theta_i})$$

$$+ f(\hat{\boldsymbol{\theta}}_{k-1} + \delta\boldsymbol{s}_{\theta_i})], \tag{34}$$

$$\approx \sum_{i=0}^{2L} w_i^{(m)} \boldsymbol{\chi}_{k,i}^{\theta}. \tag{35}$$

Similarly, the measurement state, the forecast covariance, and the cross-covariance for the calculation of the Kalman gain are given by

$$\hat{\boldsymbol{\psi}}_k^- \approx \sum_{i=0}^{2L} w_i^{(m)} \mathcal{Y}_{k,i}^{\theta}, \tag{36}$$

$$\mathbf{P}_{\boldsymbol{\theta}_k}^- \approx \sum_{i=1}^{L} [w_i^{(c_1)}(\boldsymbol{\chi}_{k,i}^{\theta} - \boldsymbol{\chi}_{k,L+i}^{\theta})^2 + w_i^{(c_2)}(\boldsymbol{\chi}_{k,i}^{\theta} + \boldsymbol{\chi}_{k,L+i}^{\theta}$$

$$- 2\boldsymbol{\chi}_{k,0}^{\theta})^2], \tag{37}$$

$$\mathbf{P}_{\tilde{\boldsymbol{\psi}}_k}^- \approx \sum_{i=1}^{L} [w_i^{(c_1)}(\mathcal{Y}_{k,i}^{\theta} - \mathcal{Y}_{k,L+i}^{\theta})^2 + w_i^{(c_2)}(\mathcal{Y}_{k,i}^{\theta} + \mathcal{Y}_{k,L+i}^{\theta}$$

$$- 2\mathcal{Y}_{k,0}^{\theta})^2], \tag{38}$$

$$\mathbf{P}_{\boldsymbol{\theta}_k \tilde{\boldsymbol{\psi}}_k} \approx \sum_{i=0}^{L} w_i^{(m)}(\boldsymbol{\chi}_{k,i}^{\theta_k} - \hat{\boldsymbol{\theta}}_k^-)(\mathcal{Y}_{k,i} - \hat{\boldsymbol{\psi}}_k^-)^{\mathrm{T}} \tag{39}$$

$$\approx \sqrt{w_1^{(c_1)} \mathbf{P}_{\boldsymbol{\theta}_k}^-}(\mathcal{Y}_{k,1:L}^{\theta} - \mathcal{Y}_{k,L+1:2L}^{\theta})^{\mathrm{T}}. \tag{40}$$

One main advantage of SP-CDKF over SP-UKF is that it uses only one "control parameter" (δ) compared to three (λ, α, and κ) in UKF. For exact derivation and algorithmic details see Ito and Xiong (2000), Nørgaard et al. (2000a), van der Merwe and Wan (2001a), and Wan and van der Merwe (2001).

3. SPKF applied to the Lorenz model

In the field of data assimilation, the celebrated Lorenz (1963) model has served as a test bed for examining the properties of various data assimilation methods (Gauthier 1992; Miller et al. 1994; Evensen 1997) because the Lorenz model shares many common features with the atmospheric circulation and climate system in terms of variability and predictability (Palmer 1993). By adjusting the model parameters that control the nonlinearity of the system, the model can be used to simulate nearly regular oscillations or highly nonlinear fluctuations.

a. Lorenz 1963 model

The Lorenz (1963) model (hereafter Lorenz 1963) consists of nonlinear ordinary differential equations of three components, given by

$$\frac{dx}{dt} = \sigma(y - x) + q^x, \tag{41}$$

$$\frac{dy}{dt} = \rho x - y - xz + q^y, \quad \text{and} \tag{42}$$

$$\frac{dz}{dt} = xy - \beta z + q^z, \tag{43}$$

where variables x, y, and z are related to the intensity of convective motion and to the temperature gradients in the horizontal and vertical directions, respectively, and the parameters σ, ρ, and β will be referred to as dynamical parameters; q^x, q^y, and q^z represent the unknown model errors, assumed to be uncorrelated in time (white noise). Also, we assume that all the measurements or observations are linear functions of the nonlinear model states. The true data are created by integrating the model over 4000 time steps using the fourth-order Runge–Kutta scheme (Press et al. 1992), with parameters σ, ρ, and β set to 10.0, 28.0, and 8/3, respectively, and initial conditions set to 1.508870, −1.531271, and 25.46091 (Miller et al. 1994; Evensen 1997). The integration step is set to 0.01. The observation datasets are simulated by adding normal distributed noise to the true data. The assimilation process is completely subject to the model Eqs. (41)–(43) after the initial guesses are given; at each step of the integration, the initial conditions are the estimated model state from the previous step.

b. State estimation

To apply KF, we discretize the nonlinear Lorenz model (41)–(43) using the fourth-order Runge–Kutta method and write it in the form of state space equations given by (1) and (2), where $\boldsymbol{\theta}_k$ represents the system state vector (a column vector composed of x, y, and z), $f(\cdot)$ is the nonlinear function of the state, and $\boldsymbol{q}_k$ is the random (white) process noise vector (column vector composed of q^x, q^y, and q^z). The measured model state $\boldsymbol{\psi}_k$ required for the application of the KF is a function of the states according to Eq. (2), where $h(\cdot)$ is the measurement function and $\boldsymbol{r}_k$ is the random measurement noise vector.

To implement the SP-UKF, the state vector is redefined as the concatenation of the model states, model errors, and measurement errors. The augmented state

vector $\boldsymbol{\Theta}_k$ and the corresponding covariance matrix are given by the following equations:

$$\boldsymbol{\Theta}_k = \begin{bmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{q}_k \\ \boldsymbol{r}_k \end{bmatrix}, \tag{44}$$

$$\mathbf{P}_{\boldsymbol{\Theta}_k} = \begin{bmatrix} \boldsymbol{P}_{\theta_k} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_k \end{bmatrix}. \tag{45}$$

Therefore, the augmented state dimension is the sum of the original state dimension, model error dimension, and measurement error dimension given by

$$L_{\Theta} = L_{\theta} + L_q + L_r, \tag{46}$$

where L_{θ} is the dimension of the state, L_q is the dimension of the model error vector, and L_r is that of measurement errors. The augmented sigma points are found using the transformation Eqs. (12)–(14). The dimension of the augmented sigma-point vector is $2L_{\Theta} + 1$. For the Lorenz model discussed here, the augmented sigma-point vector dimension is 19. In other words the number of sigma points required to approximate the error statistics accurately is 19. The augmented sigma-point vector is then propagated through (15) and (16) and the optimal terms for the calculation of Kalman gain are computed according to Eqs. (17)–(21).

On the other hand, for implementing SP-CDKF, we split the augmentation process between the time update and measurement update sections. For time update, the augmented state vector, covariance matrix, and sigma-point vector are given by

$$\hat{\boldsymbol{\Theta}}_{K-1} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{k-1} & \boldsymbol{q}_{k-1} \end{bmatrix}, \tag{47}$$

$$\mathbf{P}_{\boldsymbol{\Theta}_{k-1}} = \begin{bmatrix} \boldsymbol{P}_{\theta_{k-1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_{k-1} \end{bmatrix}, \quad \text{and} \tag{48}$$

$$\boldsymbol{\mathcal{X}}_{k-1} = \begin{bmatrix} \boldsymbol{\Theta}_{k-1} & \boldsymbol{\Theta}_{k-1} + \delta\sqrt{\boldsymbol{P}_{\boldsymbol{\Theta}_{k-1}}} \\ \boldsymbol{\Theta}_{k-1} - \delta\sqrt{\boldsymbol{P}_{\boldsymbol{\Theta}_{k-1}}} \end{bmatrix}, \tag{49}$$

and for measurement update they are given by

$$\hat{\boldsymbol{\Theta}}_{\hat{k}} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_k^- & \boldsymbol{r}_k \end{bmatrix}, \tag{50}$$

$$\mathbf{P}_{\boldsymbol{\Theta}_k} = \begin{bmatrix} \boldsymbol{P}_{\theta_k}^- & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_k \end{bmatrix}, \quad \text{and} \tag{51}$$

$$\boldsymbol{\mathcal{X}}_k = \begin{bmatrix} \boldsymbol{\Theta}_k & \boldsymbol{\Theta}_k + \delta\sqrt{\boldsymbol{P}_{\boldsymbol{\Theta}_k}} & \hat{\boldsymbol{\Theta}}_k - \delta\sqrt{\boldsymbol{P}_{\boldsymbol{\Theta}_k}} \end{bmatrix}. \tag{52}$$

The optimal terms for the calculation of Kalman gain are computed using (35)–(40).

c. Parameter estimation from noisy measurements

The model parameter estimation can be regarded as a special case of general state estimation in which the parameters are treated as specific states. Parameter estimation involves determining a nonlinear mapping

$$\mathbf{Y}_k = \mathcal{N}(\boldsymbol{\theta}_k, \boldsymbol{\Lambda}), \tag{53}$$

where the nonlinear map $\mathcal{N}(\cdot)$ may be the dynamical model $f(\cdot)$ or an empirical model parameterized by the vector $\boldsymbol{\Lambda}$. The state space representation of the parameter estimation problem for the Lorenz model can be written as

$$\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_{k-1} + \boldsymbol{q}_{k-1}^{\Lambda}, \tag{54}$$

$$\boldsymbol{\psi}_k = f(\boldsymbol{\theta}_k, \boldsymbol{\Lambda}_k) + \boldsymbol{r}_k^{\Lambda}, \tag{55}$$

where $f(\cdot)$ is the nonlinear measurement model given by the Lorenz Eqs. (1)–(3); $\boldsymbol{\Lambda}$ is the parameter vector that constitutes the dynamical parameters σ, ρ, and β; and $\boldsymbol{q}_k^{\Lambda}$ and $\boldsymbol{r}_k^{\Lambda}$ represent the model and measurement error vector respectively. The SPKF (SP-UKF and SP-CDKF) equations for the parameter estimation problem are similar to those of the state estimation formulation except that the state (here states are parameters) time evolution is linear [Eq. (54)] and the measurement function is nonlinear [Eq. (55)].

d. Joint estimation of parameters and states

The joint estimation approach is required when the "clean" state and parameters are not available. In general there are two approaches to estimating the model state and parameters simultaneously, namely the *dual estimation* and *joint estimation* approaches (Haykin 2001; Nelson 2000; van der Merwe 2004). In the dual estimation approach, two Kalman filters are running simultaneously for state and parameter estimation. On the other hand, in the joint estimation approach, the system state and parameters are concatenated into a single higher-dimensional joint state vector and only one Kalman filter is used to estimate the joint vector. For example, the joint state vector $\mathcal{J}_k$ for the SPKF data assimilation can be written as

$$\mathcal{J}_k = \begin{bmatrix} \boldsymbol{\theta}_k & \boldsymbol{\Lambda}_k \end{bmatrix}^{\mathrm{T}}. \tag{56}$$

In the joint estimation process, the SPKF schemes estimate the states using parameters that are estimated at every time step using the prior states. In this study, we will only present the joint estimation of parameters and states because it incorporates complete model states and parameters during assimilation cycles.
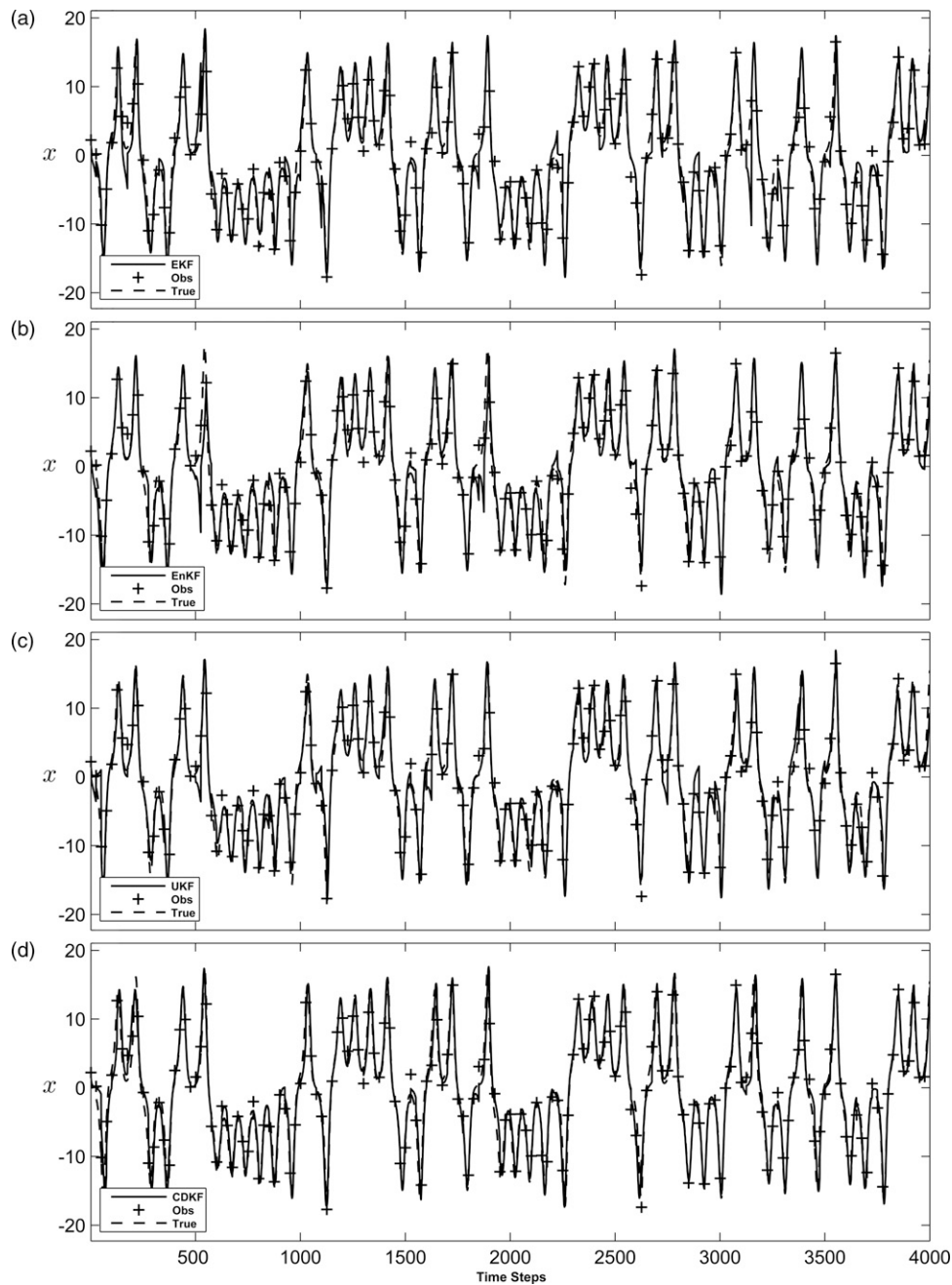
FIG. 1. Case 1—Assimilation solutions for the Lorenz model. The dashed curve is the true model,
+ indicates the noisy observations, and the solid trajectory indicates the assimilated solutions.

4. Experiments and results

In this section we demonstrate the feasibility of the SPKF algorithms as an effective data assimilation method for highly nonlinear models. The SPKF algorithms discussed in the previous sections will now be examined and compared with standard EKF and EnKF methods. To compare the SPKF algorithms with stan-dard EKF and EnKF, all experiments were designed almost identically to those of Miller et al. (1994) and Evensen (1997).

a. State estimation

The first set of experiments were carried out with initial conditions, parameters, and observation noise levels identical to those in Miller et al. (1994) and

Evensen (1997): the observations and initial conditions are simulated by adding normally distributed noise $N(0, \sqrt{2})$. Also the interval of observation is set to 25; that is, the observed states are assimilated to the nonlinear model at every 25 time steps.

For all the cases to be discussed, we assume that the model and observation errors are uncorrelated in both space and time. Because there is no general way to set the model error, usually the amount of model error to use in the KF is often determined experimentally by trial or by statistical methods such as Monte Carlo, which is computationally expensive (Miller et al. 1994). In our experiments, the model errors were intentionally designed in such a way that the model would not drift from the true state too much.[5] In detail, we set the model errors by calculating the expected errors in the state scaled by a decreasing exponential factor that is a function of the assimilation time; initially, the model covariance matrix is set to an arbitrary diagonal value and then anneals toward zero exponentially as the assimilation proceeds. For simulating model errors in the ensemble Kalman filter, we follow the method suggested by Evensen (2003). An ensemble of 1000 members was used in the EnKF as in Evensen's experiment (Evensen 1997).

Figures 1a–d show the state estimate using the EKF, EnKF, SP-UKF, and SP-CDKF, respectively. As can be seen, all four methods can generate the model states similar to true values, indicating good capability of these methods in estimating model states if the size of initial perturbation and observed noise are appropriate as given here. Figures 1a and 1b were also obtained by Miller et al. (1994) and Evensen (1997). It should be noted that the EKF and the EnKF can have good state estimates, but the former needs to construct TLM and the latter requires a large ensemble size of 1000. In contrast, the SP-UKF and the SP-CDKF only use 19 "particular" ensemble members (sigma points) here, showing their advantages over EKF and EnKF.

A comparison among the four methods is shown in Fig. 2: the variation of the error square (ES) with time step. The ES is defined here as the square of the difference between estimated state and true model state scaled by N, where N is a scalar quantity:[6]

$$\text{Error} = \frac{1}{N} \left(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^{\text{True}} \right)^2. \tag{57}$$

From Fig. 2, we can see that the SP-UKF and SP-CDKF assimilations have a smaller ES than EnKF at most times, although some assimilation steps have an opposite situation. These "peak" values of ES correspond to either overestimation or underestimation of model states, which are most probably related to random noise in the "observations" and to the chaotic nature of the Lorenz system. The state estimate is probably poor when a large noise is assimilated and when the state is a transition from one chaotic regime to the other (also see the discussions below).

The overall performance of each assimilation is measured by the root-mean-square error (RMSE) over all time steps, as shown in Table 1. As shown, SP-UKF and SP-CDKF have slightly smaller RMSE than others. The most impressive point in the table is that SPKF methods use only 19 sigma points (or in general 19 conditional ensembles) to estimate the statistical moments of the nonlinear model accurately. This turns out to be an advantage for the data assimilation problems in low-dimensional systems, but in the case of atmospheric or ocean GCMs the $2L + 1$ integration is not computationally feasible. More details on SPKF implementation, its limitations, and methods to overcome the limitations are described in detail in section 5.

For the sake of completeness, we performed an EnKF assimilation experiment with 19 ensembles compared to 1000 ensembles. The result of this experiment is shown in Fig. 3a and the corresponding ES and RMSE are shown in Fig. 3b and Table 1. These results show that errors of state estimate from the EnKF with 19 ensemble members are around 5–10 times as much as from SPKF. Thus, EnKF with only 19 members could not capture the mean and covariance of a highly nonlinear Lorenz model appropriately. On the other hand, with just 19 conditional ensembles (or sigma points), SPKF is able to capture the statistical moments of the highly nonlinear Lorenz model.

The assimilation experiments took place on a symmetric multiprocessor (SMP) machine with two AMD Optron 248 CPUs (Advanced Micro Devices 2007) with a clock speed of 2.2 GHz, running on Linux. MATLAB 7.3.0.298 (R2006) software (available online from Mathworks, Inc., at http://www.mathworks.com/products/matlab/index.html) was used to implement the model and data assimilation algorithm. Table 1 also compares the computation time required by each assimilation algorithm discussed above. To compare the computational efficiency, we use the same programming framework for implementing all the data assimilation methods discussed above. The computational cost is the least for EKF, followed by two SPKF methods. The EnKF that requires 1000 members for a good

---

[5] The model is considered to have a relatively large error at the initial time so the assimilation weighs more observation information. As such, the model prediction would not drift from the true value too much.

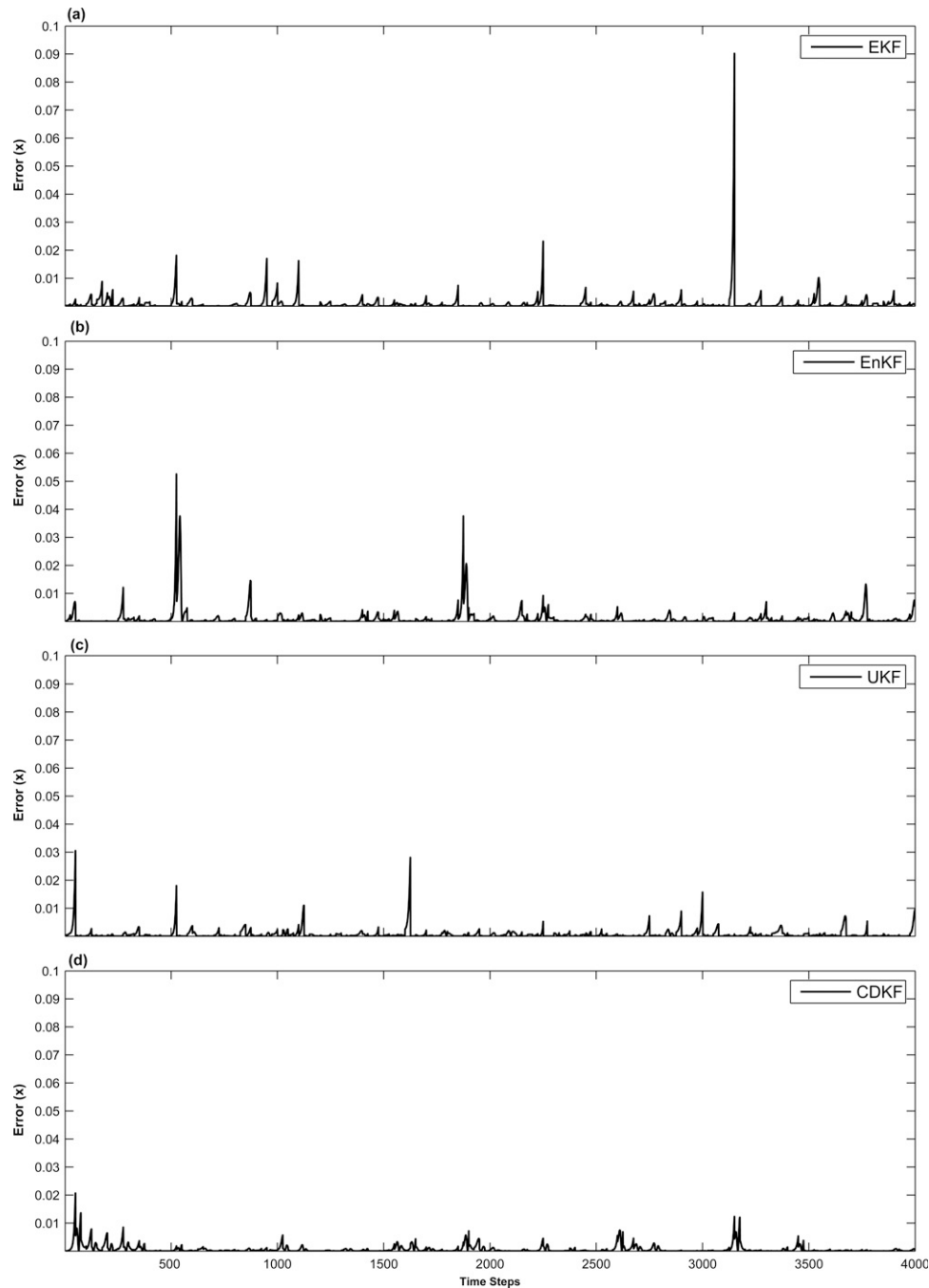[6] We choose N to be 4000, which is the total time step.

FIG. 2. Case 1—Assimilation errors for state estimation (corresponding to Fig. 1): (a) EKF, (b) EnKF, (c) SP-UKF, and (d) SP-CDKF.

estimate (see Fig. 1b) is the most expensive, around 50–80 times as much as SPKF.

The second set of experiments was carried out with a more realistic situation by increasing the observations' noise level tenfold: the observations and initial conditions are generated by adding normally distributed noise $N(0, \sqrt{20})$. The assimilation results are shown in

Fig. 4 and the corresponding ES is shown in Fig. 5. From Fig. 4 we can see the superior performance of SPKF algorithms. Among SPKFs, the SP-CDKF shows comparatively better assimilation skill than SP-UKF. The RMSE of the EKF, EnKF, SP-UKF, and SP-CDKF assimilation results were 5.39, 6.37, 4.25, and 4.56, respectively. From Figs. 4 and 5, it is evident that

TABLE 1. RMSE and computation time for case 1

Assimilation method	Computation time (s)	RMSE
EKF	37.04	1.812
EnKF (with 1000 ensembles)	7143.57	1.987
EnKF (with 19 ensembles)	132.77	6.123
SP-UKF	133.91	1.640
SP-CDKF	90.42	1.592

the SPKF assimilation schemes can give better estimation results even if the system is subjected to higher noise. Also, our numerical experiments showed that an EnKF assimilation with 1000 ensembles can give results (not shown here) as accurate as SPKF assimilation.

In the third set of experiments, we increased the observation noise level as well as the interval between consecutive observations; the interval between observations was increased from 25 to 40 and the observations and initial conditions were generated by adding normally distributed noise $N(0, \sqrt{20})$. The assimilation results and corresponding ES are shown in Figs. 6 and 7 respectively.

Figure 7 shows some divergence in some time steps of the assimilation track among the four methods. For example, the errors (ES) vary almost steadily in SP-UKF whereas SP-CDKF has a relatively significant variation of ES with time steps. Compared with SPKF, the variation of ES is more striking in EKF and EnKF. The significant variation of ES might be related to the chaotic nature of the Lorenz system and the capability

of individual algorithm in capturing the observation information. The chaotic Lorenz attractor is known to have a butterfly shape with two wings. For a good estimate of the transition state from one wing to the other, the assimilation should be able to characterize the information of both wings of the Lorenz attractor. Obviously this depends on two issues: the observation itself and the assimilation algorithm. If the observation is more frequently assimilated (i.e., the interval between observations is small), sufficient data allow the coverage of more information of both chaotic regimes in assimilation. This is the reason why there are many more abnormal'' values of ES in Fig. 7 than in Fig. 2, in which the observations are more frequent. On the other hand, if one assimilation algorithm has a better capacity to mix observation and model information to characterize transitions, it would have better estimates for transition states. In many cases, it highly depends on the model and observation error covariances. When the observation error covariance is usually predescribed, the model error covariance is updated at each assimilation step in the family of Kalman filters, depending on the algorithm. Thus, Fig. 7 suggests that SPKF is probably better than EKF and EnKF in the assimilation of some transition states using noisy observations.

Again, we repeated the EnKF assimilation (for case 3) with 1000 ensembles and the result is shown in Fig. 8. The result is not as good as SPKF assimilation and seems noisier. This is probably because the observation
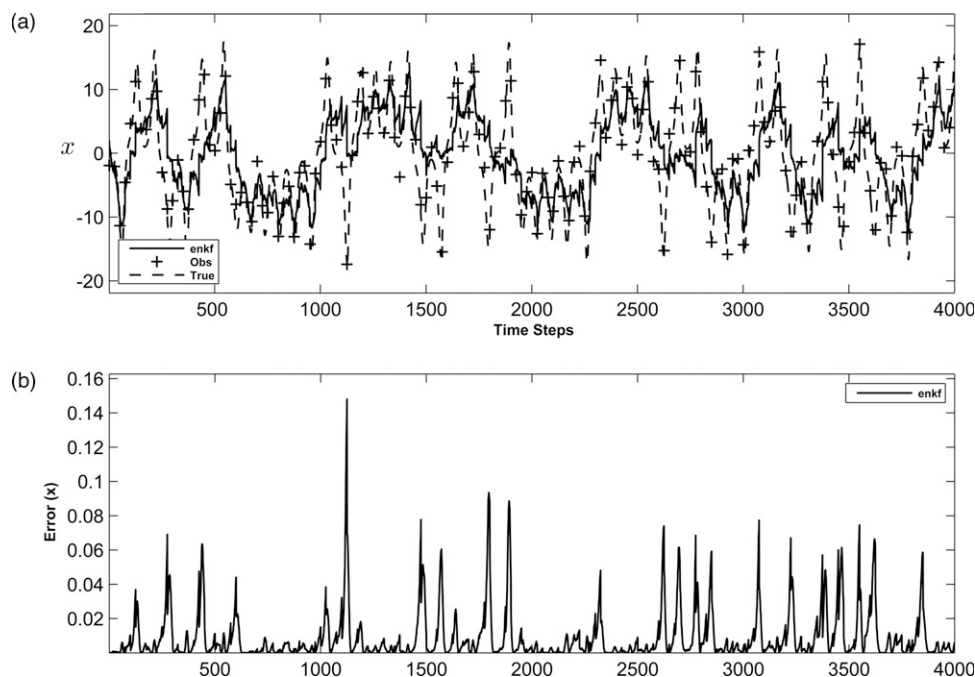

FIG. 3. (a) EnKF data assimilation solutions for the Lorenz model with 19 ensembles (symbols as in Fig. 1). (b) Assimilation errors.
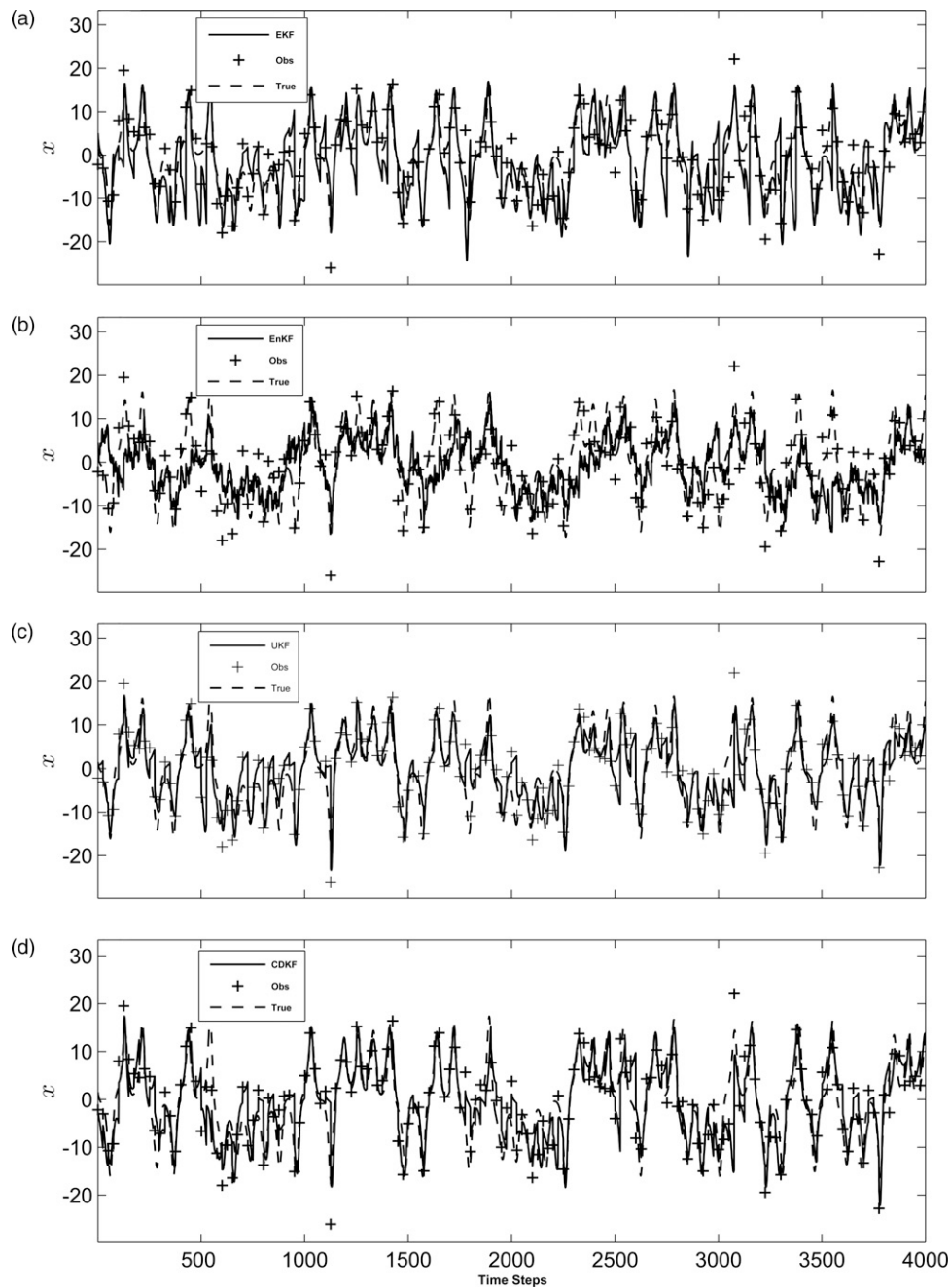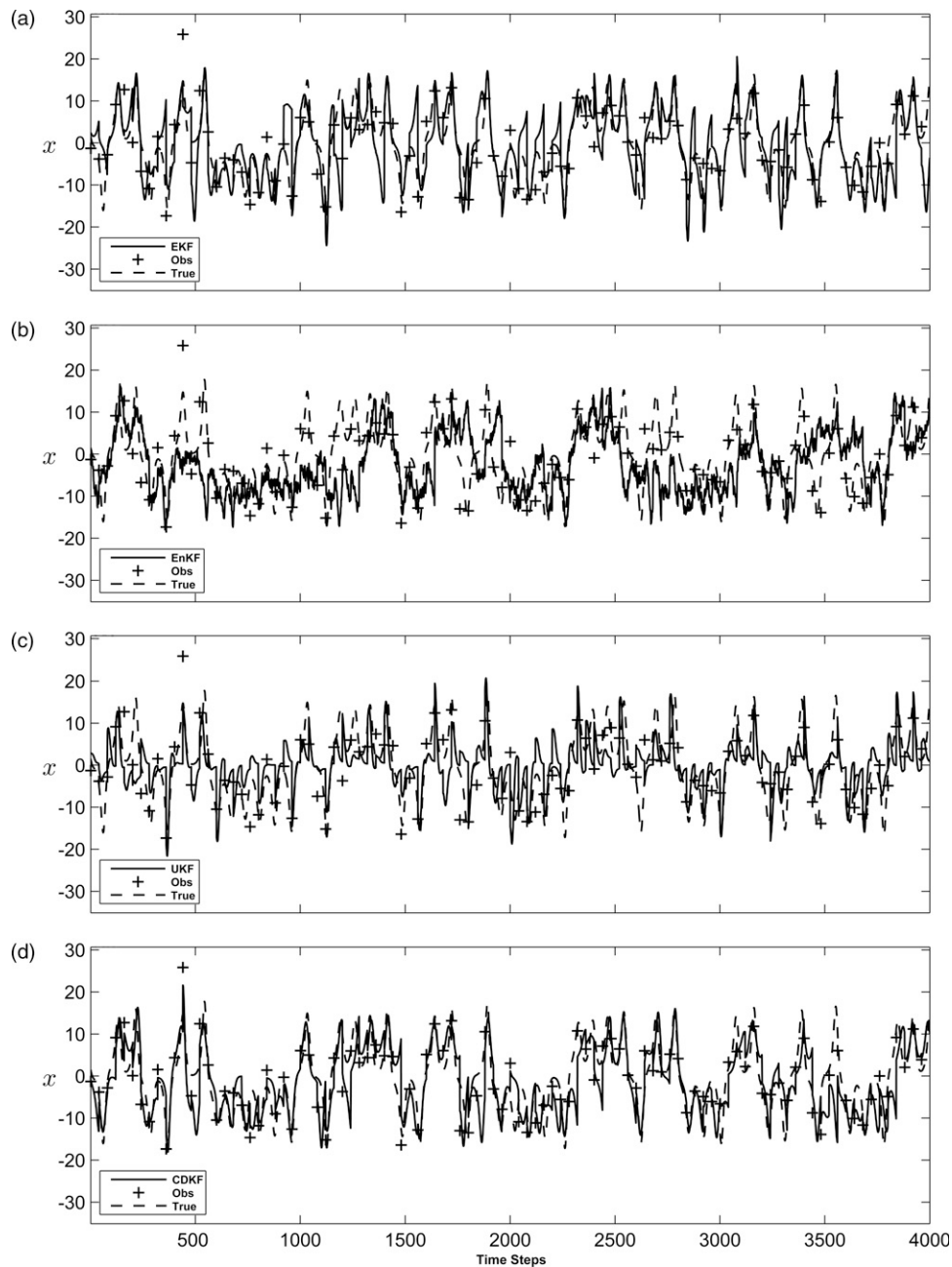
FIG. 4. Case 2—Assimilation solutions for the Lorenz model with a tenfold increase in the observation noise levels (symbols as in Fig. 1). (Here the EnKF simulation is performed with 19 ensembles).

assimilated is noisier and less frequent; thus, an ensemble size of 1000 is probably not enough to capture the statistical moments accurately.

b. Parameter estimation

Estimating uncertain dynamical model parameters is one of the important tasks in data assimilation, where the measurement function is usually nonlinear. The re-

quirement of the tangent linear measurement operator **H** in the optimal gain term given by Eq. (10) makes the EKF and EnKF assimilation schemes inaccurate and inappropriate for the parameter estimation in nonlinear dynamical systems. It has been shown that the EnKF data assimilation gave poor results in estimating the dynamical parameter of the Lorenz model (Kivman 2003). The SPKF methods should be better alternatives
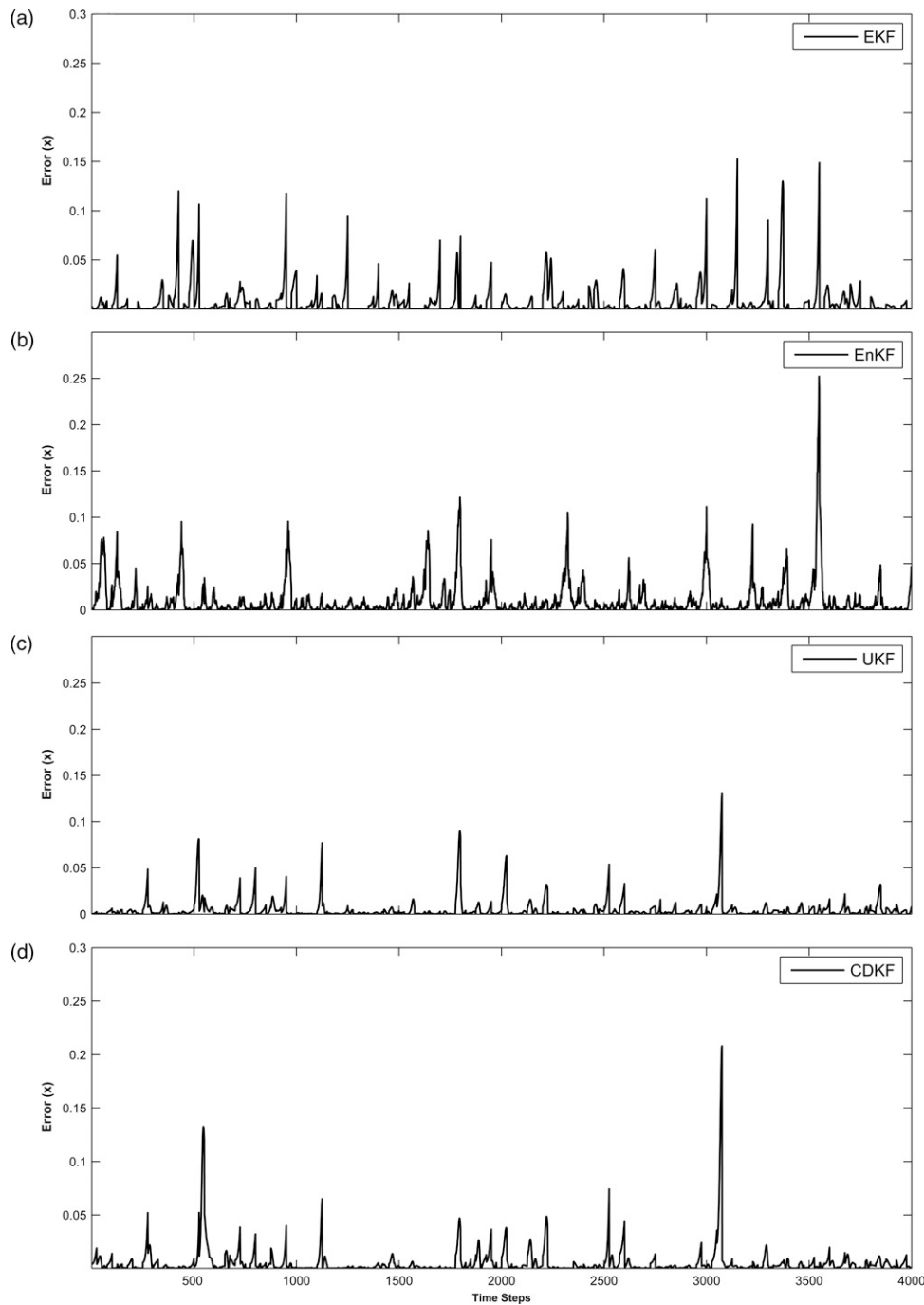
FIG. 5. Case 2—Assimilation errors for state estimation (corresponding to Fig. 4): (a) EKF, (b) EnKF, (c) SP-UKF, and (d) SP-CDKF.

for parameter estimation because they do not need to linearize the nonlinear measurement function.

The experimental setup is identical to that of the first case of the state estimation problem discussed in the above subsection. To simulate a more realistic situation, the initial guesses of the parameters are generated by adding normal distributed noise of covariance 100 to the true parameters. In the first case, we assume that only one parameter (say β) is uncertain. Thus, our task is to estimate the correct value of β from infrequent observations contaminated by noise. Figure 9 shows the SPKF parameter estimation results. Figures 9a and 9b

FIG. 6. Case 3—Assimilation solutions for the Lorenz model with fewer observations and a tenfold increase in the observation noise levels (symbols as in Fig. 1). Here the EnKF simulation is performed with 19 ensembles.

show the parameter estimation using SP-UKF and SP-CDKF, respectively.

From these figures it is clear that SPKF assimilation methods can retrieve dynamical parameters well from noisy observations. In the above experiment, even though the initial parameter was far from the true value (the standard deviation is 10), the SPKF method is still able to estimate the parameter accurately. In general,

our experiments suggest a faster convergence for SP-CDKF algorithm. This might be due to the algorithm tuning problem, because SP-CDKF uses only one control parameter (δ) compared to three (λ, α, and κ,) in SP-UKF.

In the second case we assume that two dynamical parameters (say ρ and β) are uncertain. This situation is more difficult than the first case because inaccuracy in
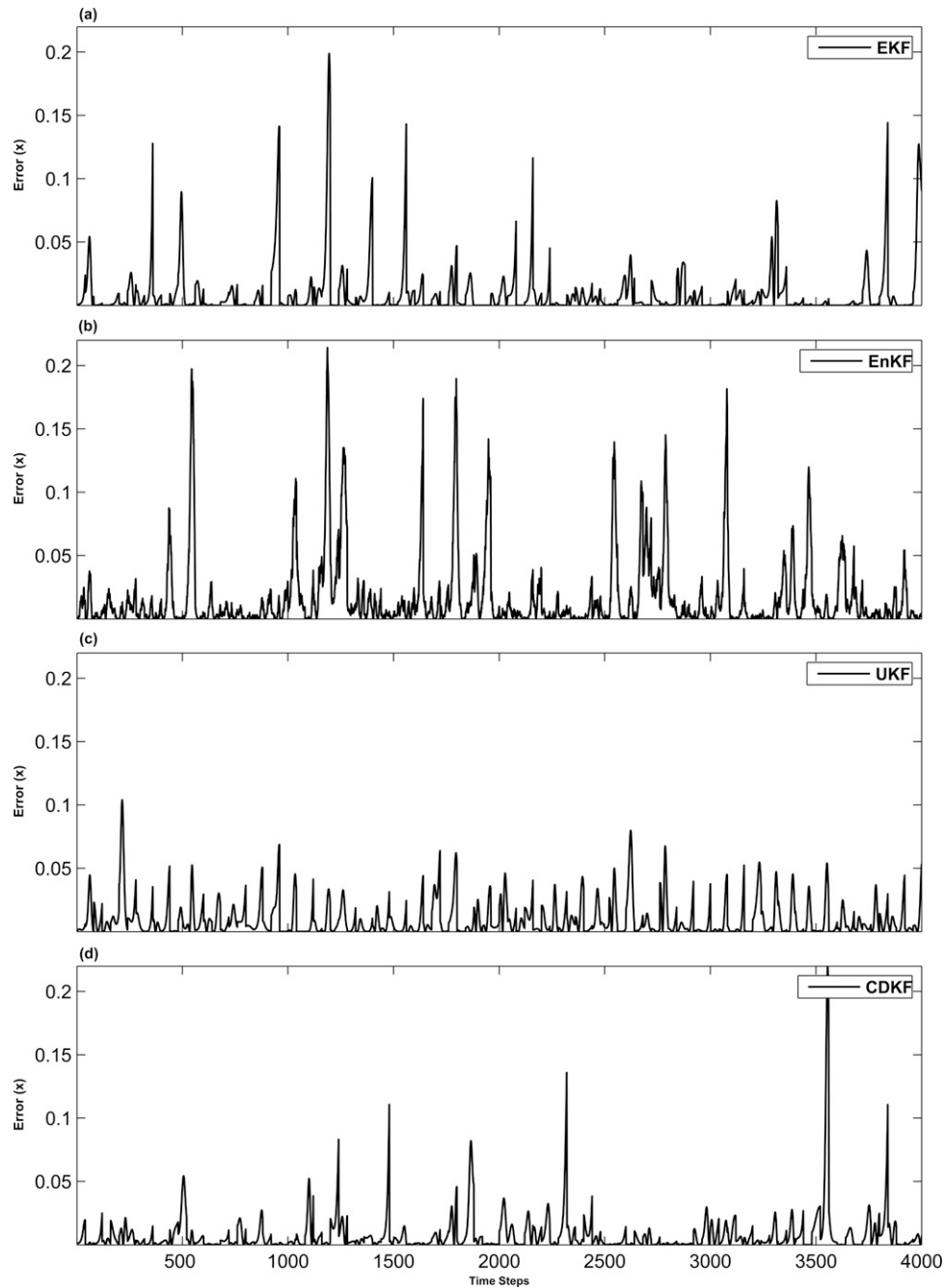
FIG. 7. Case 3—Assimilation errors for state estimation (corresponding to Fig. 6): (a) EKF (b) EnKF (c) SP-UKF and (d) SP-CDKF.

the estimation of one parameter can result in inaccurate estimation of the other. Initial parameters were generated using the same method as in the previous case: adding normal distributed noise of covariance 100 to the true parameters. Figures 10a and 10b show the results of the simultaneous estimation of ρ and β using SP-UKF and SP-CDKF, respectively. In the case of single parameter estimation, SPKF assimilation is able to approximate the true parameter much faster compared to the two-parameter case. This suggests that more frequent observations might be needed to accurately estimate both parameters.

c. Joint estimation of model states and parameters

Data assimilation problems involving inaccurate model states and parameters arise in many situations in
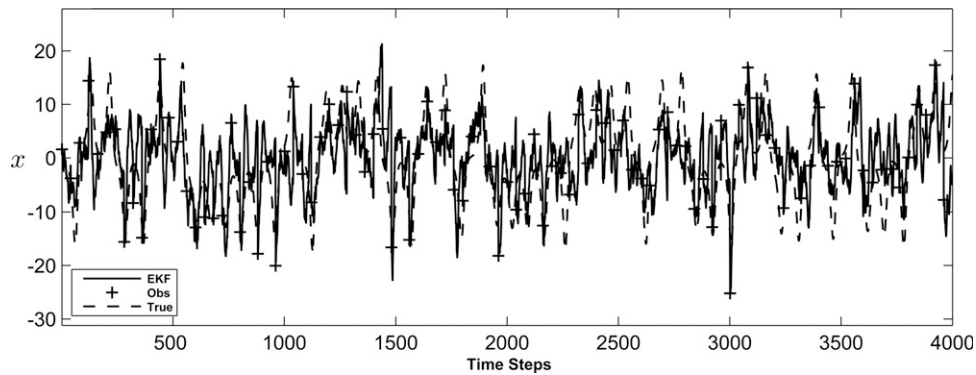
FIG. 8. Case 3—EnKF assimilation solutions with 1000 ensembles, for the Lorenz model with fewer observations and a tenfold increase in the observation noise levels (symbols as in Fig. 1).

meteorology and physical oceanography. In this situation our task is to estimate the model states and parameters simultaneously from a set of noisy observations. In this experiment we used SPKF data assimilation schemes for the joint estimation of parameters and states simultaneously. The experimental setup is identical to that of the state estimation discussed in the above section where the interval between noisy observations is set to 25 and the noise covariance is 2. In the joint estimation approach the model states and parameters evolve in time simultaneously; model states are estimated at each assimilation step using the estimated parameters, which are estimated from the prior states. In this simulation we estimated the model state x and dynamical parameter σ simultaneously. Figures 11 and 12 show the joint estimation results and the corresponding error square for SP-UKF assimilation, respectively; Figs. 13 and 14 do the same for SP-CDKF data assimilation.

Simultaneously estimating both state and parameter values increases the nonlinearity of the assimilation problem, thereby increasing the assimilation time needed to retrieve them. From Figs. 12 and 14, we can see that the ES of the parameter estimate decreases with time. One interesting feature in Figs. 11–14 is that when the estimated parameters are far away the true values, the model states still have good estimation. This is because the initial model errors for the states are much higher than those for the parameter; thus, the analysis weights the observations much more than the model simulation associated with inaccurate parameters. When the estimated parameter gradually approaches the true value, the ES of model state estimate seems not to decrease significantly. This is because as the model error decreases with assimilation time (i.e., as the model becomes more and more accurate), the model state becomes more sensitive to slight changes in the estimated parameter.

In summary, all the above experiments—including state, parameter, and joint estimation with different observation frequencies and noise levels—show that sigma-point Kalman filters are efficient and good assimilation algorithms for a highly nonlinear Lorenz system. If the observation density is high and noise level is
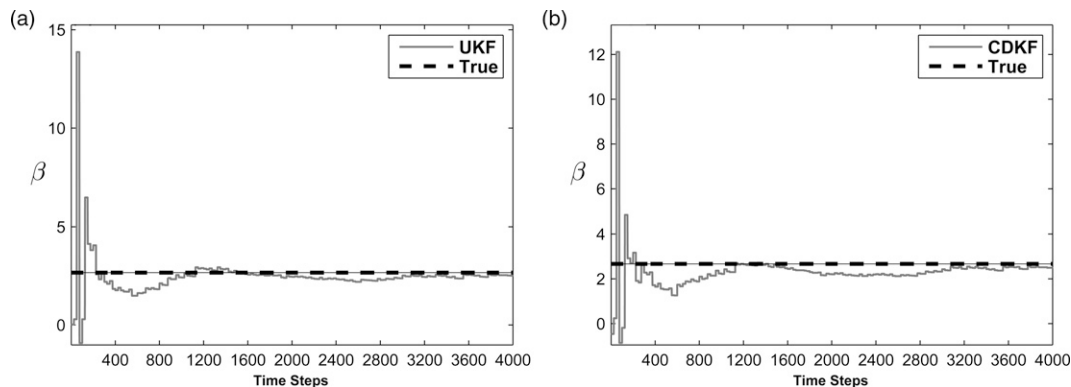

FIG. 9. Parameter estimation: (a) SP-UKF assimilation; (b) SP-CDKF assimilation. True β = dashed line; estimated β = solid line.
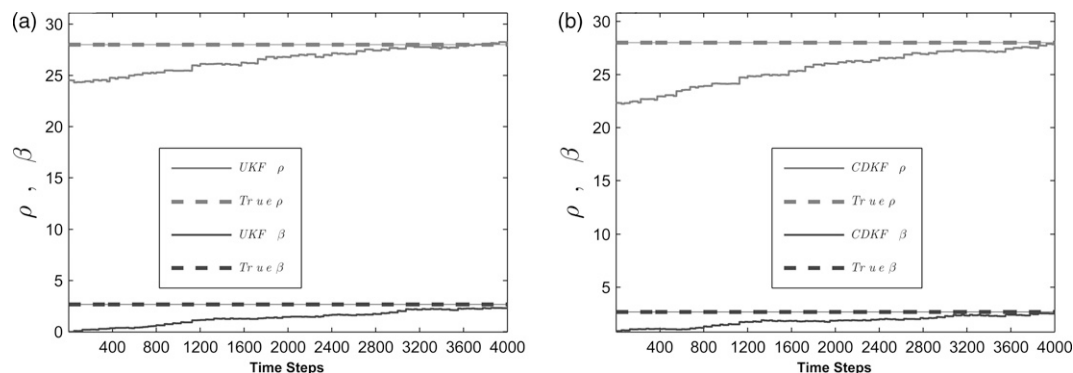
FIG. 10. Simultaneous estimation of two parameters: (a) SP-UKF assimilation; (b) SP-CDKF assimilation. True β = solid thick line, true ρ = dashed line; estimated β = dashed–dotted solid line, estimated ρ = solid thin line.

small, all the data assimilation methods discussed above would estimate the model state accurately, but at the cost of additional computational expense and the requirement of TLM for EnKF and EKF. But even if the noise level is high and the observations are less frequent, SPKF can estimate the model states and parameters with good accuracy, without the requirements of TLM and costly computation.

5. SPKF data assimilation in higher-dimensional systems

In the preceding sections, we have demonstrated the power and merits of SPKF, as well as its advantages over EKF and EnKF, by the low-dimensional Lorenz model. One of the crucial issues in evaluating a data assimilation algorithm is its computational expense when applied to realistic models that have a large dimensionality. In this section we will further explore the SPKF using higher-dimensional Lorenz models.

For an L-dimensional system, the number of sigma points required to estimate the true mean and covariance is $2L + 1$. As described in the previous sections, this procedure works well for low-dimensional models like the Lorenz 1963 model, but $2L + 1$ sigma-point integration is computationally unfeasible if the dimension system is of the order of tens of millions, as in global GCMs. Julier (Julier and Uhlmann 2002; Julier 2003; Julier and Uhlmann 2004) has shown that by us-
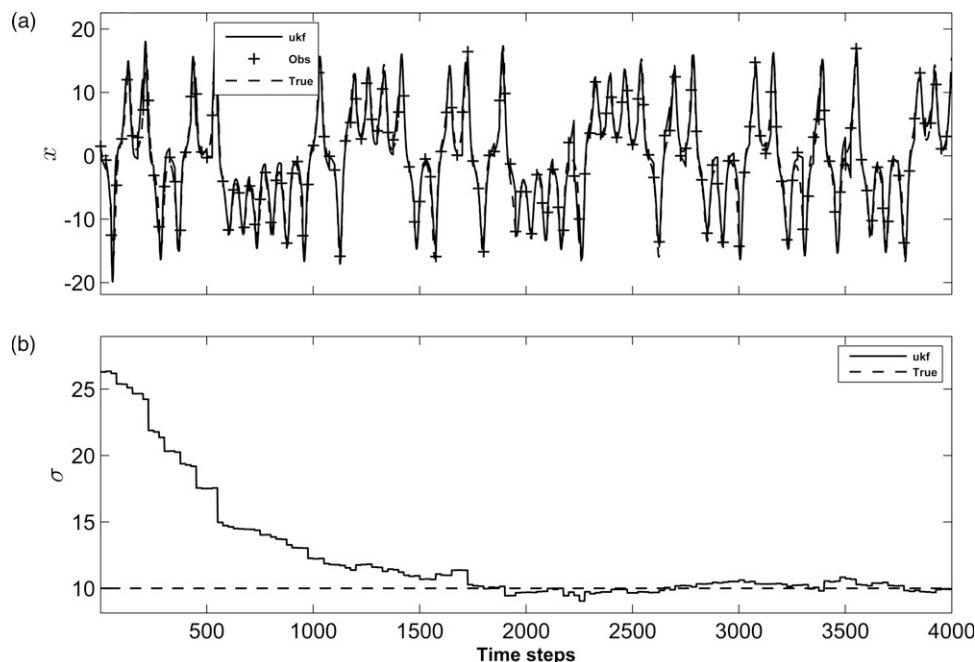

FIG. 11. SP-UKF assimilation for joint estimation of state and parameter: (a) dashed line = true state x, observations = +, solid line = estimated x; (b) dashed line = true σ, solid line = estimated σ.
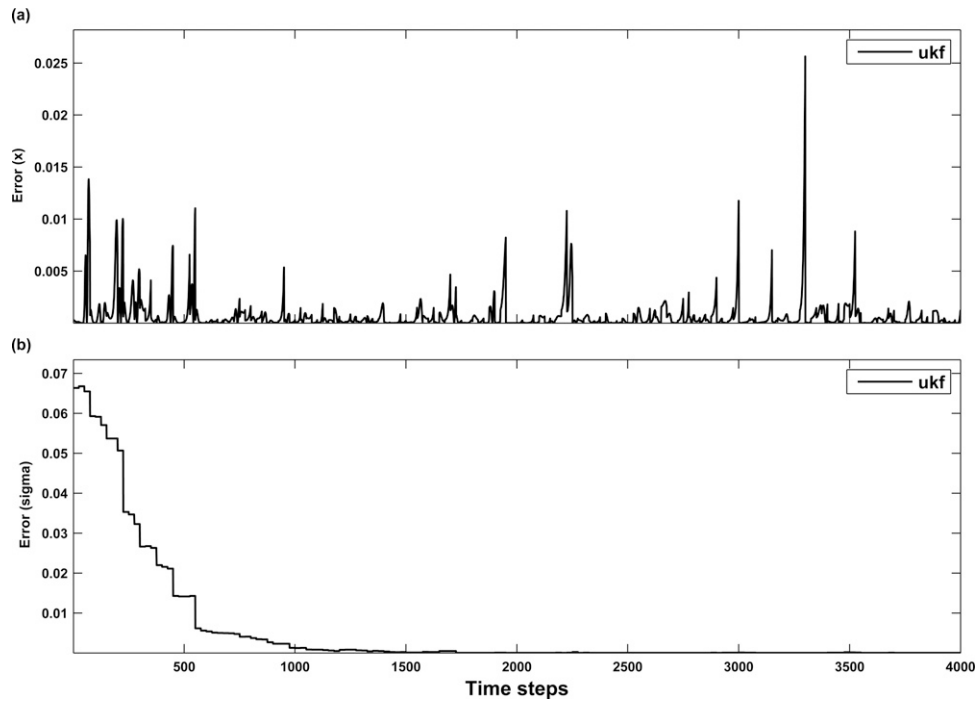
FIG. 12. Assimilation errors for SP-UKF joint estimation: (a) estimation error for x (b) estimation error for σ.

ing *simplex unscented transformation* the minimum number of sigma points that gives same estimation accuracy as SP-UKF can be reduced to $L + 1$. These sigma points are called simplex sigma points, but for higher-dimensional systems this $L + 1$ simplex sigma-point integration is still computationally intractable. A possible solution to this problem is to reduce the number of sigma points by selecting a particular subset of
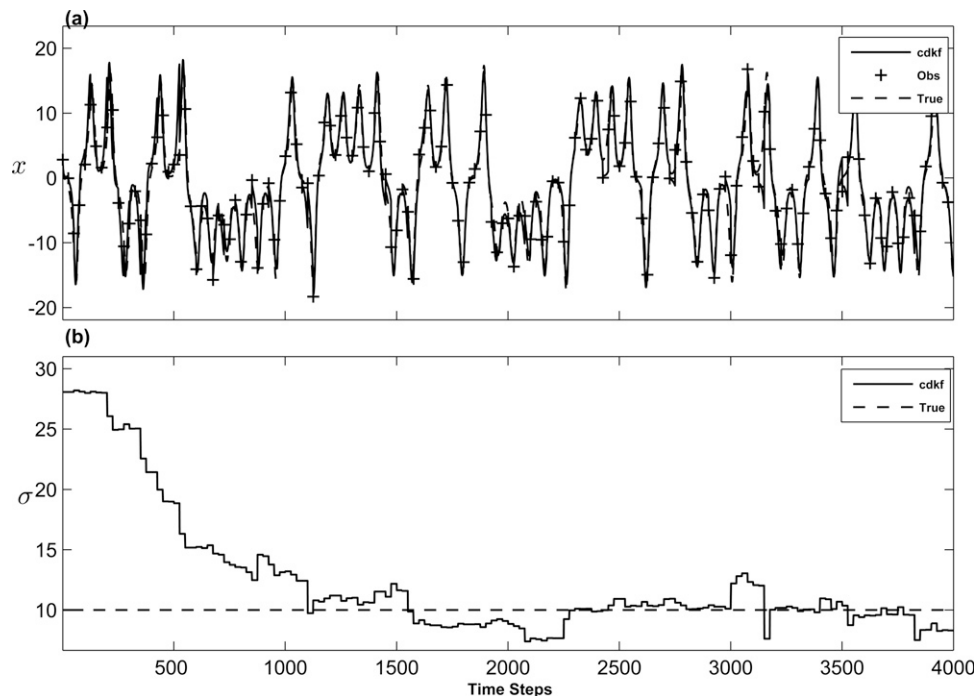

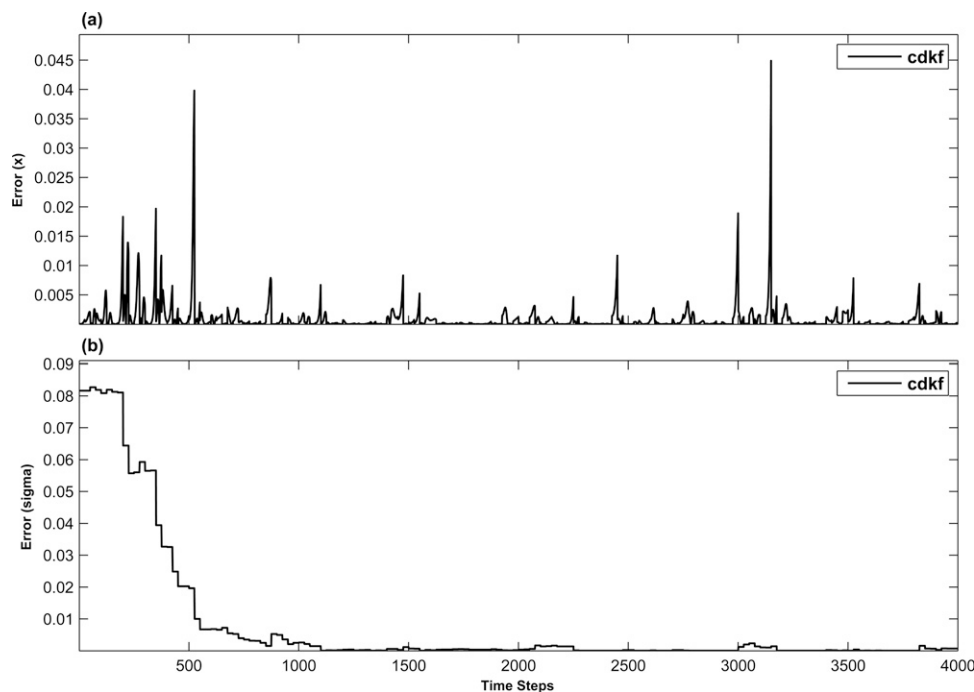FIG. 13. As in Fig. 11, but for SP-CDKF.

FIG. 14. As in Fig. 12, but for SP-CDKF.

sigma points from the original sigma-point space, which can approximate the error statistics of the model. In the following subsections we will examine this possibility.

a. A subspace approach with sigma points: Design and implementation

It has been shown that the number of degrees of freedom necessary to describe most large-scale geophysical systems is finite, and their dominant variability can be described by a limited number of modes or functions (Teman 1991; Lermusiaux and Robinson 1999; Lermusiaux 1997). These functions evolve in time and space in accordance with the system. The techniques commonly used for deriving such functions include dynamical normal modes, dynamical singular vectors and values, empirical orthogonal functions (EOFs; Lorenz 1965; Davis 1977; Weare and Nasstrom 1982; Wallace et al. 1992; von Storch and Frankignoul 1997), principal interaction and oscillation patterns (PIPs and POPs, respectively; Hasselmann 1988; von Storch et al. 1995; Penland 1989; Schnur et al. 1993), and radial functions and wavelets (Meyers et al. 1993; Gamage and Blumen 1993). Lermusiaux (Lermusiaux and Robinson 1999; Lermusiaux 1997) proposed a method to reduce error space, called error subspace statistical estimation (ESSE). In the ESSE approach, a reduced rank approximation P_k^p to the error covariance P_k should be defined by minimizing the norm of the difference between P_k and P_k^p; i.e.,

$$\text{rank}\left(P_k^p\right) = p_{\min\|P_k - P_k^p\|}. \tag{58}$$

According to the minimum criterion (58), the error subspace is characterized by the singular vectors and values of P_k.

We follow an idea similar to ESSE to form a sigma-point subspace that approximates the mean and error covariance of system. In our approach, it is assumed that when the estimate of a system's full errors requires all sigma points, its dominant errors can be estimated using the most important sigma points. Theoretically these most important sigma points should be chosen based on (58). However, this will introduce huge complexity and be difficult to implement. For simplicity, as a good start toward a complete solution to the problem, we have used principal component analysis (PCA) to identify the most important sigma points that influence the evolution of error covariance. The main idea behind using PCA is to represent the multidimensional sigma-point space by a fewer number of sigma points while still retaining the main features of the original sigma-point space; that is, sigma points in the principal component space are used to calculate the error propagation. The selection of sigma points is based on the proportion of variances. Specifically, instead of using the full sigma-point space, we use some leading principal components, thereby reducing the number of sigma points required to approximate forecast error covariance. In the following subsections, we will see the po-

tential of this approach in the assimilation of higher-dimensional systems.

b. Experiments with the Lorenz 1995 model

The Lorenz 1995 (Lorenz 2006) model is a one-dimensional atmospheric model introduced by E. Lorenz in 1995 to explain the dynamics of weather at a fixed latitude. It has error growth characteristics similar to those of full NWP models. The model contains K variables $X_1, \ldots, X_k$, which may be thought of as atmospheric variables in K sectors of a latitude circle, governed by

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F, \qquad (59)$$

where the constant F, called the forcing term, is independent of k. By using the cyclic boundary conditions, the definition of X_k is extended to all values of k; that is, X_{k-K} and X_{k+K} equal X_k. It is assumed that a unit time $\Delta t = 1$ is associated with 5 days.

The experimental setup is similar to that of Lorenz (Lorenz 2006), where $K = 36$ and the magnitude of the forcing is set to 8, for which the system is chaotic. The system is integrated using the fourth-order Runge–Kutta method, with a time step of $\Delta t = 0.05$ (i.e., 6 h). The experiments were carried out with random initial conditions, and the observations were generated by adding normally distributed noise $N(0, \sqrt{2})$ to the true states. Also, the interval of observation is set to 10; i.e., the observed states are assimilated to the nonlinear model at every 10 time steps. A more detailed discussion of the model and its characteristics can be found in Lorenz (2006, 2005) and Lorenz and Emmanuel (1998).

c. Performance and evaluation

For all the cases to be discussed, we assume that the model and observation errors are uncorrelated in both space and time. In the first case we use "full" sigma-point space for the calculation of error covariance. Thus we have a total of 217 sigma points, hence 217 ensemble members. Figure 15a shows the state estimate using the SP-UKF. As can be seen, SP-UKF can estimate the model states similar to true values, indicating good capability of the original SPKF methods in estimating model states.

In the second case we use the *reduced sigma-point subspace* to calculate the error covariance. In this case we select 40 sigma points, which account for more than 90% of the total variance. The result of this experiment is shown in Fig. 15b. As can be seen, the model states can be fairly well estimated by the reduced SPKF, although its estimate accuracy is not as good as the origi-

nal SPKF. This suggests a possible solution to applying SPKF for high-dimensional systems. For the sake of completeness, we performed an EnKF assimilation experiment with 40 ensembles. The ensemble is generated using the same approach as the previous experiment with the Lorenz 1963 model, in which we used 19 ensembles (Fig. 3a). The result of this experiment is shown in Fig. 15c. Comparing Figs. 15b and 15c reveals that the reduced SPKF is better than the EnKF for the state estimate, especially for the magnitude estimate. It is apparent that the EnKF underestimates the magnitude of model states during the transition phase period.

We also performed the SPKF assimilation experiment for the 960-variable Lorenz '95 model. The experimental setup is identical to that in the previous cases except that $K = 960$. Two cases are studied with the model. In the first case we use all sigma points (a total of 5761 sigma points), and in the second case we use 200 important sigma points for the calculation of error covariances. The results of these experiments are shown in Fig. 16. For comparison, we also performed an EnKF assimilation experiment with 200 ensembles, and the result of this experiment is shown in Fig. 16c. Apparently the reduced SPKF leads to a better estimate than the EnKF in both phase and magnitude simulation. As can be seen in the EnKF, the estimated state is often out of the phase of "true" trajectory, which is absent in the reduced SPKF. The correlation between the estimated trajectory and true trajectory is 0.59 for the reduced SPKF and 0.10 for the EnKF.

A great deal of additional research is needed for better design and implementation of these techniques applied to atmosphere or ocean GCMs for state, parameter, and joint estimation problems. However, the above experimental results are promising, and a variety of possible extensions to these techniques could be developed to deal with more complicated situations.

6. Discussion and conclusions

The EKF and EnKF, two important Kalman-type filters, have been widely applied for atmospheric and oceanic data assimilation because of their efficient and simple algorithms. The major weaknesses of the EKF and EnKF are that the former needs to calculate the tangent linear model or Jacobian for linearization of nonlinear forecast models and that the EnKF performance is greatly dependent on ensemble size, which is often an intractable burden for computation. The EKF and EnKF cannot deal with the systems directly if observed data are a nonlinear transformation of states.

In this study we introduced and presented two recently proposed derivativeless sigma-point Kalman filters. The SPKF is a technique for implementing a de-
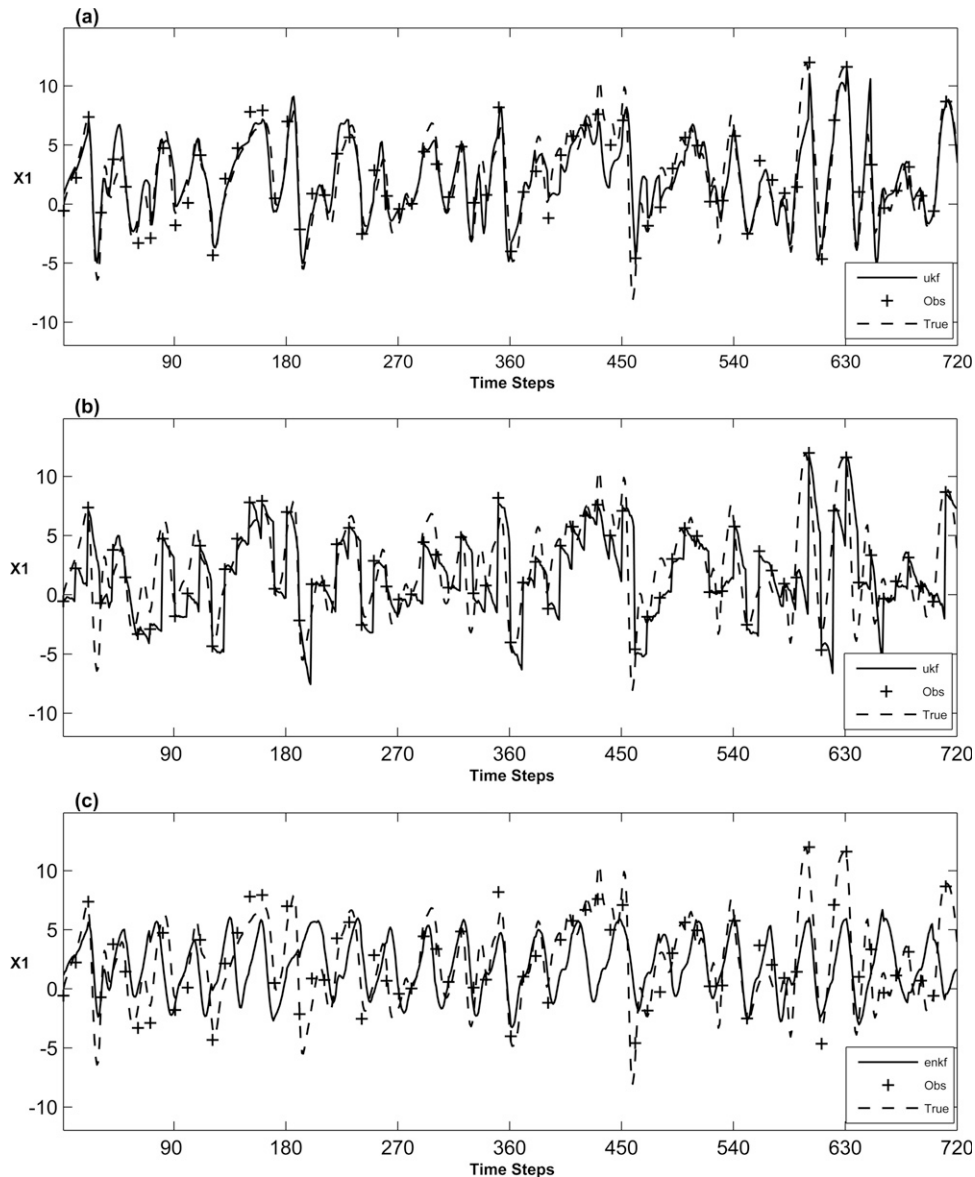
FIG. 15. The 36-variable Lorenz 1995 model assimilation solutions for X_1: (a) case 1, (b) case 2, and (c) EnKF with 40 ensembles. The dashed curve is the true model, + indicates the noisy observations, and the solid trajectory indicates the assimilated solutions.

rivativeless optimal estimation using a novel deterministic sampling approach that ensures a small size of samples to accurately estimate forecast error statistics. It is unlike EnKF, in which a random sampling strategy is used. The technique employed in SPKF is that it reinterprets the standard Kalman gain and covariance update equation in such a way that it does not need linearization of the nonlinear prediction model and nonlinear measurement operator, and it can capture the statistical moments of the nonlinear model accurately using deterministic sampling technique. Thus, in SPKF the forecast error covariance equation is computed using deterministically chosen samples, called sigma points. In a broad sense, the SPKF algorithm can be considered as a particular case of the ensemble Kalman filter with a specific sample selection scheme. In other words, the forecast sigma points in SPKF algorithms are actually specific ensembles conditioned on the specific selection schemes, which can represent error statistics accurately. Also, the ensemble forecast step in SPKF can be parallelized by running each ensemble member on a separate processor of a parallel computer (or cluster), resulting in huge computational savings.
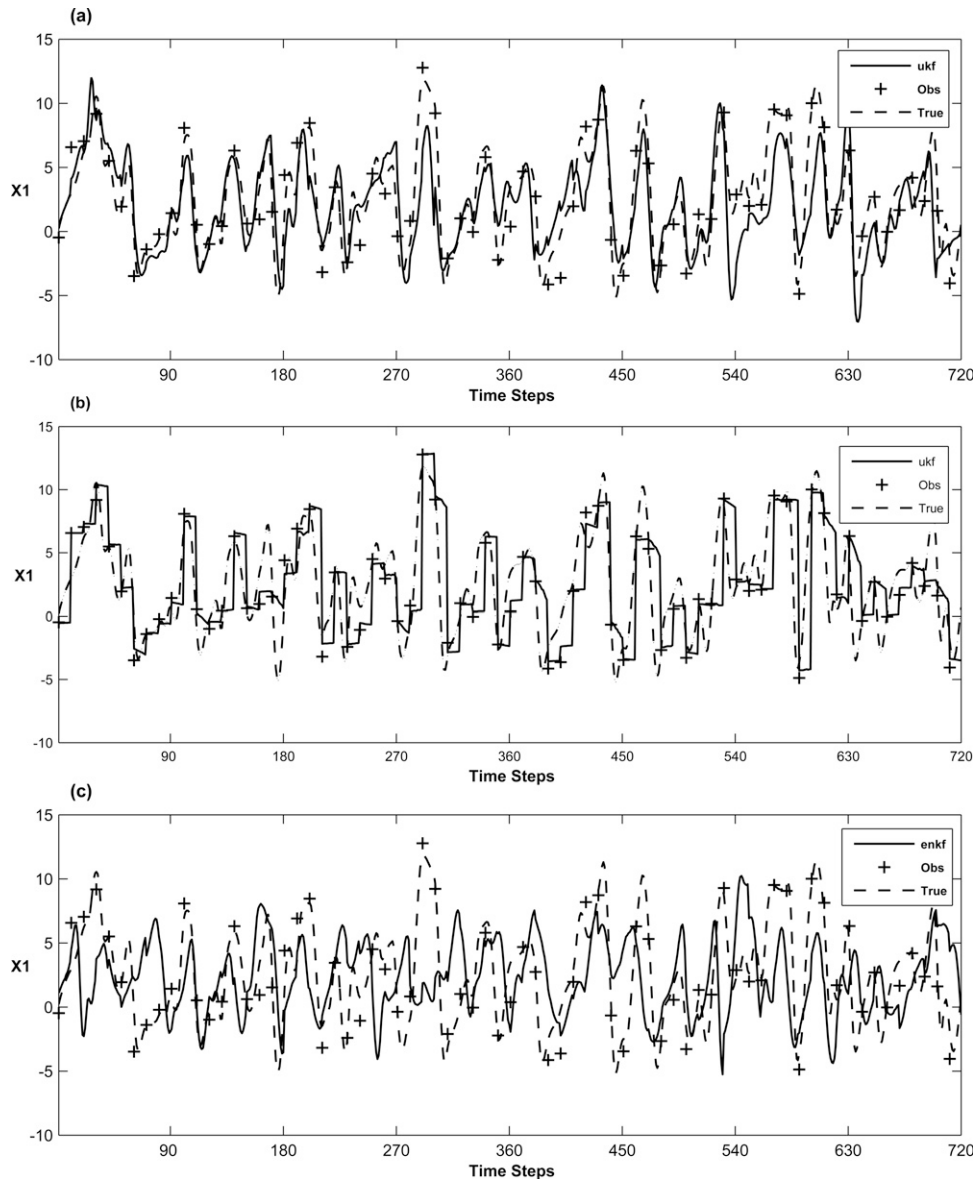
FIG. 16. The 960-variable Lorenz 1995 model assimilation solutions for X_1: (a) case 1, (b) case 2, and (c) EnKF with 200 ensembles. Colors and symbols as in Fig. 15.

Using the highly nonlinear low-dimensional Lorenz 1963 model and a higher-dimensional Lorenz '1995 model, we investigated the capability and performance of SPKF over standard KF-based data assimilation methods for three different classes of problems, namely state estimation, parameter estimation, and joint estimation. The results demonstrated that the SPKF has better estimate accuracy than EKF and EnKF for all experiments. SPKF experiments with a higher-dimensional model suggest that it is possible to reduce the number of sigma points, thereby reducing the computation time, by using a reduced sigma-point space approach. The results in this study are encouraging and suggest that the SPKF could become an effective method to assimilate observations into realistic models such as atmospheric or oceanic GCMs. The SPKF also has the advantage that it does not need tangent linear or Jacobian operators of the original models.

The SP-UKF and SP-CDKF data assimilation involves the calculation of the matrix square root of the state covariance matrix, which is a computationally intensive process. It has been shown that square root formulation of SP-UKF and SP-CDKF is numerically efficient and stable and has equal estimation accuracy

when compared to original SP-UKF and SP-CDKF (van der Merwe and Wan 2001a,b). Because the state space dimension of the model that we used in this study is relatively small, it is practically irrelevant to compare the numerical stability of the square root formulation with original SP-UKF and SP-CDKF implementation. Therefore, this issue is left for future study in GCMs.

In this study, we explored the SPKF using highly simplified nonlinear models. One might be concerned by the performance and efficiency of SPKF when a realistic GCM is used. Additional research is needed for better implementation of these techniques applied to data assimilation problems in atmospheric or ocean GCMs. Nonetheless, the present study represents a step in pursuing advanced data assimilation algorithms by using a simple nonlinear model, which shares some common features with complicated atmospheric and oceanic models. Future work will also study the parallelization of SPKF data assimilation in GCMs similar to EnKF parallelization because the propagation of each sigma point through the nonlinear model is independent. We are currently working on an implementation of the SPKF for a realistic ocean GCM that will investigate the estimation accuracy, numerical stability, and consistency, as well as the computational difficulties. These studies will be described in future work.

APPENDIX A

Reinterpretation of the Standard Kalman Gain

The optimal state update equation in the Kalman filter algorithm can be written as

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_k^- + \mathbf{K}_k(\boldsymbol{\psi}_k - \hat{\boldsymbol{\psi}}_k^-), \tag{A1}$$

where the superscript negative sign represents the prior states given by the following equations:

$$\hat{\boldsymbol{\theta}}_k^- = E[f(\boldsymbol{\theta}_{k-1}, \boldsymbol{q}_{k-1})] \quad \text{and} \tag{A2}$$

$$\hat{\boldsymbol{\psi}}_k^- = E[h(\boldsymbol{\theta}_k^-, \boldsymbol{r}_k)], \tag{A3}$$

$$= E[\boldsymbol{\psi}_k] = \mathbf{H}\hat{\boldsymbol{\theta}}_k^-, \tag{A4}$$

Here, $\mathbf{H}$ is the measurement operator and $E[\cdot]$ represents the mathematical expectation or the expected value. Thus, the state and covariance update equations can be rewritten as

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_k^- + \mathbf{K}_k(\boldsymbol{\psi}_k - \mathbf{H}\hat{\boldsymbol{\theta}}_k^-) \quad \text{and} \tag{A5}$$

$$\mathbf{P}_{\boldsymbol{\theta}_k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{\boldsymbol{\theta}_k}^-. \tag{A6}$$

Now the "standard" Kalman gain equation is given by

$$\mathbf{K}_k = \underbrace{\mathbf{P}_{\boldsymbol{\theta}_k}^-\mathbf{H}^T}(\underbrace{\mathbf{H}\mathbf{P}_{\boldsymbol{\theta}_k}^-\mathbf{H}^T + \mathbf{R})^{-1}}, \tag{A7}$$

where $\mathbf{P}_{\boldsymbol{\theta}_k}^-$ is the forecast error covariance matrix. The first underbracketed expression in the Kalman gain term can be interpreted as the cross-covariance between the state and observation errors (Gelb 1974; Simon 2006):

$$\mathbf{P}_{\boldsymbol{\theta}_k\tilde{\boldsymbol{\psi}}_k} = E[(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)(\boldsymbol{\psi}_k - \boldsymbol{\psi}_k^-)^T], \tag{A8}$$

$$= E[(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)(\mathbf{H}\boldsymbol{\theta}_k + \boldsymbol{r}_k - \mathbf{H}\boldsymbol{\theta}_k^-)^T], \tag{A9}$$

$$= E[(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)^T]\mathbf{H}^T$$
$$+ E[\boldsymbol{\theta}_k r_k^T] - E[\hat{\boldsymbol{\theta}}_k r_k^T], \tag{A10}$$

$$= \mathbf{P}_{\boldsymbol{\theta}_k}^-\mathbf{H}^T. \tag{A11}$$

Similarly, the second underbracketed expression in Eq. (A7) can be interpreted as the error covariance of the difference between model and observation (Gelb 1974):

$$\mathbf{P}_{\tilde{\boldsymbol{\psi}}_k} = E[(\boldsymbol{\psi}_k - \hat{\boldsymbol{\psi}}_k^-)(\boldsymbol{\psi}_k - \hat{\boldsymbol{\psi}}_k^-)^T], \tag{A12}$$

$$= E[(\mathbf{H}\boldsymbol{\theta}_k + \boldsymbol{r}_k - \mathbf{H}\hat{\boldsymbol{\theta}}_k^-)(\mathbf{H}\boldsymbol{\theta}_k + \boldsymbol{r}_k - \mathbf{H}\hat{\boldsymbol{\theta}}_k^-)^T], \tag{A13}$$

$$= \mathbf{H}E[(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)^T]\mathbf{H}^T + E[\boldsymbol{r}_k r_k^T], \tag{A14}$$

$$+ \mathbf{H}E[(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)r_k^T] + E[\boldsymbol{r}_k(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)^T]\mathbf{H}^T, \tag{A15}$$

$$= \mathbf{H}\mathbf{P}_{\boldsymbol{\theta}_k}^-\mathbf{H}^T + \mathbf{R}. \tag{A16}$$

Therefore the Kalman gain can be rewritten as

$$\mathbf{K}_k = \mathbf{P}_{\boldsymbol{\theta}_k\tilde{\boldsymbol{\psi}}_k}\mathbf{P}_{\tilde{\boldsymbol{\psi}}_k}^{-1}. \tag{A17}$$

The main advantage of using this form of Kalman gain is that we can avoid the use of a measurement operator, especially when the measurement operator is a nonlinear function of the state. A complete statistical derivation of the above formulation can be found in Simon (2006).

APPENDIX B

An Alternate Formula for Updating the State Error Covariance Matrix

The estimation error is defined as

$$\tilde{\boldsymbol{\theta}}_k = \boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k. \tag{B1}$$

Similarly, the error between the noisy observation $\boldsymbol{\psi}_k$ and its prediction $\hat{\boldsymbol{\psi}}_k^-$ is given by

$$\tilde{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k - \hat{\boldsymbol{\psi}}_k^-. \tag{B2}$$

Substituting (B1) into the state-update Eq. (A1), we can rewrite the estimation error as

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_k &= \boldsymbol{\theta}_k - \boldsymbol{\theta}_k^- - \mathbf{K}_k\big(\boldsymbol{\psi}_k - \hat{\boldsymbol{\psi}}_k^-\big) \\ &= \tilde{\boldsymbol{\theta}}_k^- - \mathbf{K}_k\tilde{\boldsymbol{\psi}}_k. \end{aligned} \tag{B3}$$

Here we made use of the fact that the estimator is unbiased:

$$E\big[\tilde{\boldsymbol{\psi}}_k\big] = 0. \tag{B4}$$

Now, the state error covariance $\mathbf{P}_{\boldsymbol{\theta}_k}$ and the cross covariance $\mathbf{P}_{\boldsymbol{\theta}_k\tilde{\psi}_k}$ between the state and observation error given by Eqs. (A8) and (A12) can be rewritten in terms of Eqs. (B1) and (B2) and are given by

$$\mathbf{P}_{\boldsymbol{\theta}_k} = E[\tilde{\boldsymbol{\theta}}_k\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}] \quad \text{and} \tag{B5}$$

$$\mathbf{P}_{\boldsymbol{\theta}_k\tilde{\psi}_k} = E[\tilde{\boldsymbol{\theta}}_k^-\tilde{\boldsymbol{\psi}}_k^{\mathrm{T}}]. \tag{B6}$$

Taking the outer products and expectation of (B3) produces

$$\begin{aligned}E[\tilde{\boldsymbol{\theta}}_k\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}] &= E[(\tilde{\boldsymbol{\theta}}_k^- - \mathbf{K}_k\tilde{\boldsymbol{\psi}}_k)(\tilde{\boldsymbol{\theta}}_k^- - \mathbf{K}_k\tilde{\boldsymbol{\psi}}_k)^{\mathrm{T}}] \\ &= E[\tilde{\boldsymbol{\theta}}_k^-\tilde{\boldsymbol{\theta}}_{k^-}^{\mathrm{T}}] - E[\tilde{\boldsymbol{\theta}}_k^-\tilde{\boldsymbol{\psi}}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}] - E[\mathbf{K}_k\tilde{\boldsymbol{\psi}}_k\tilde{\boldsymbol{\theta}}_{k^-}^{\mathrm{T}}] \\ &\quad + E[\mathbf{K}_k\tilde{\boldsymbol{\psi}}_k\tilde{\boldsymbol{\psi}}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}]. \end{aligned} \tag{B7}$$

Using Eqs. (B5) and (B6), Eq. (B7) can be rewritten as

$$\mathbf{P}_{\boldsymbol{\theta}_k} = \mathbf{P}_{\boldsymbol{\theta}_k}^- - \mathbf{P}_{\boldsymbol{\theta}_k\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}} - \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k\boldsymbol{\theta}_k} + \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}}. \tag{B8}$$

Substituting the expression for Kalman gain, given by Eq. (A17), back into the above expression, the covariance update equation is given by

$$\mathbf{P}_{\boldsymbol{\theta}_k} = \mathbf{P}_{\boldsymbol{\theta}_k}^- - \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}}. \tag{B9}$$

A more detailed interpretation and derivation of the above expression can be found in Simon (2006).

REFERENCES

Advanced Micro Devices, cited 2007: AMD Opteron processor technical documents. Advanced Micro Devices, Inc. [Available online at http://www.amd.com/us-en/processors/technicalresources/0,,30_182_739_9003,00.html.]

Baheti, R. S., O. Halloran, and H. R. Itzkowitz, 1990: Mapping extended Kalman filters onto linear arrays. *IEEE Trans. Autom. Control,* **35,** 1310–1319.

Daum, F. E., and J. Fitzgerald, 1983: Decoupled Kalman filters for phased array radar tracking. *IEEE Trans. Autom. Control,* **28,** 269–283.

Davis, R. E., 1977: Techniques for statistical analysis and prediction of geophysical fluid systems. *Geophys. Astrophys. Fluid Dyn.,* **8,** 245–277.

Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.,* **97** (C11), 17 905–17 924.

——, 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.,* **125,** 1342–1354.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343–367.

Gamage, N., and W. Blumen, 1993: Comparative analysis of low-level cold fronts: Wavelet, Fourier, and empirical orthogonal function decompositions. *Mon. Wea. Rev.,* **121,** 2867–2878.

Gauthier, P., 1992: Chaos and quadric-dimensional data assimilation: A study based on the Lorenz model. *Tellus,* **44A,** 2–17.

Gelb, A., 1974: *Applied Optimal Estimation.* MIT Press, 374 pp.

Hasselmann, K., 1988: PIPs and POPs: A general formalism for the reduction of dynamical systems in terms of principal interaction patterns and principal oscillation patterns. *J. Geophys. Res.,* **93,** 11 015–11 021.

Haykin, S., Ed., 2001: *Kalman Filtering and Neural Networks.* Wiley, 284 pp.

Houtekamer, P., and H. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811.

Ito, K., and K. Xiong, 2000: Gaussian filters for nonlinear filtering problems. *IEEE Trans. Autom. Control,* **45,** 910–927.

Julier, S., 1998: A skewed approach to filtering. *Proc. SPIE Conf. on Signal and Data Processing of Small Targets,* Orlando, FL, International Society for Optical Engineering, 271–282.

——, 2002: The scaled unscented transformation. *Proc. 2002 American Control Conf.,* Vol. 6, Anchorage, AK, IEEE, 4555–4559.

——, 2003: The spherical simplex unscented transformation. *Proc. 2003 American Control Conf.,* Vol. 3, Denver, CO, IEEE, 2430–2434.

——, and J. Uhlmann, 2002: Reduced sigma-point filters for the propagation of mean and covariances through nonlinear transformations. *Proc. 2002 American Control Conf.,* Vol. 2, Anchorage, AK, IEEE, 887–892.

——, and ——, 2004: Unscented filtering and nonlinear estimation. *Proc. IEEE,* **92,** 401–422.

——, ——, and H. Durrant-Whyte, 1995: A new approach for filtering nonlinear systems. *Proc. 1995 American Control Conf.,* Seattle, WA, IEEE, 1628–1632.

Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 1971–1981.

Kivman, G. A., 2003: Sequential parameter estimation for stochastic systems. *Nonlinear Processes Geophys.,* **10,** 253–259.

Lefebvre, T., H. Bruyninckx, and J. De Schutter, 2002: Comment on ''A new method for the nonlinear transformation of

means and covariances in filters and estimators.'' *IEEE Trans. Autom. Control,* **47,** 1406–1409.

Lermusiaux, P. F. J., 1997: Error subspace data assimilation methods for ocean field estimation: Theory, validation, and applications. Ph.D. thesis, Harvard University, 402 pp.

——, and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.,* **127,** 1385–1407.

Lorenz, E., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

——, 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus,* **17,** 321–333.

——, 2005: Designing chaotic models. *J. Atmos. Sci.,* **62,** 1574–1587.

——, 2006: Predictability—A problem partly solved. *Predictability of Weather and Climate,* T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 40–58.

——, and K. A. Emmanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.,* **55,** 399–414.

Meyers, S. D., B. G. Kelly, and J. J. O'Brien, 1993: An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of Yanai waves. *Mon. Wea. Rev.,* **121,** 2858–2866.

Miller, R., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

Nelson, A. T., 2000: Nonlinear estimation and modeling of noisy time-series by dual Kalman filtering methods. Ph.D. thesis, Oregon Graduate Institute of Science and Technology, 298 pp.

Nørgaard, M., N. K. Poulsen, and O. Ravn, 2000a: Advances in derivative-free state estimation for nonlinear systems. Tech. Rep. IMM-REP-1998-15, Dept. of Mathematical Modeling, Technical University of Denmark, 33 pp.

——, ——, and ——, 2000b: New developments in state estimation of nonlinear systems. *Automatica,* **36,** 1627–1638.

Ohmuro, T., 1984: A decoupled Kalman tracker using LOS coordinates. *Proc. Int. Symp. Noise and Clutter Rejection in Radars and Imaging Sensors,* Tokyo, Japan, IEEE, 451–455.

Palmer, T., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.,* **74,** 49–65.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C: The Art of Scientific Computing.* 2nd ed. Cambridge University Press, 994 pp.

Schei, T. S., 1997: A finite-difference method for linearization in nonlinear estimation algorithms. *Automatica,* **33,** 2053–2058.

Schnur, R., G. Schmitz, N. Grieger, and H. von Storch, 1993: Normal modes of the atmosphere as estimated by principal oscillation patterns and derived from quasi-geostrophic theory. *J. Atmos. Sci.,* **50,** 2386–2400.

Simon, D., 2006: *Optimal State Estimation, Kalman, H∞, and Nonlinear Approaches.* 1st ed. Wiley-Interscience, 526 pp.

Teman, R., 1991: Approximation of attractors, large eddy simulations and multiscale methods. *Proc. Roy. Soc. London,* **434A,** 23–29.

Van der Merwe, R., 2004: Sigma-point Kalman filters for probabilistic inference in dynamic state-space models. Ph.D. thesis, Oregon Health and Science University.

——, and E. A. Wan, 2001a: Efficient derivative-free Kalman filters for online learning. *Proc. 2001 European Symp. on Artificial Neural Networks (ESANN),* Bruges, Belgium, 6 pp. [Available online at http://www.dice.ucl.ac.be/Proceedings/esann/esannpdf/es2001-21.pdf.]

——, and ——, 2001b: The square-root unscented Kalman filter for state and parameter estimation. *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP),* Vol. 6, Salt Lake City, UT, IEEE, 3461–3464.

——, and ——, cited 2003: ReBEL: Recursive Bayesian estimation library. [Available online at http://choosh.csee.ogi.edu/rebel/index.html.]

——, A. Doucet, N. de Freitas, and E. Wan, 2000: The unscented particle filter. Tech. Rep. CUED/F-INFENG/TR 380, Cambridge University Engineering Department, 46 pp. [Available online at http://citeseer.ist.psu.edu/article/vandermerwe00unscented.html.]

——, E. A. Wan, and S. I. Julier, 2004: Sigma-point Kalman filters for nonlinear estimation and sensor fusion: Applications to integrated navigation. *AIAA Guidance, Navigation and Control Conf.,* Providence, RI, American Institute of Aeronautics and Astronautics, 5120–5122.

Von Storch, H., and C. Frankignoul, 1997: Empirical modal decomposition in coastal oceanography. *The Global Coastal Ocean,* K. Brink and A. R. Robinson, Eds., *The Sea,* Vol. 10, Wiley, 419–455.

——, G. Burger, R. Schnur, and J.-S. von Storch, 1995: Principal oscillation patterns: A review. *J. Climate,* **8,** 377–400.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Climate,* **5,** 561–576.

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, Canada, IEEE, 153–158.

——, and ——, 2001: Kalman filtering and neural networks. *The Unscented Kalman Filter,* S. Haykin, Ed., Wiley, 221–277.

Weare, B. C., and J. Nasstrom, 1982: Examples of empirical orthogonal function analyses. *Mon. Wea. Rev.,* **110,** 481–485.

Welch, G., and G. Bishop, 1995: An introduction to the Kalman filter. Tech. Rep. TR95-041, University of North Carolina, Chapel Hill, NC, 16 pp.