J. Adv. Model. Earth Syst., **Vol. 3**, M12005, 16 pp.

# Sigma-point particle filter for parameter estimation in a multiplicative noise environment

*Jaison Thomas Ambadan[1,2] and Youmin Tang[1,3]*

[1] Environmental Science and Engineering, University of Northern British Columbia, Prince George, British Columbia, Canada.
[2] International Max-Planck Research School on Earth System Modelling, Max-Planck Institute for Meteorology, Hamburg, Germany.
[3] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, State Oceanic Administration, Hangzhou, China.

A pre-requisite for the ''optimal estimate'' by the ensemble-based Kalman filter (EnKF) is the Gaussian assumption for background and observation errors, which is often violated when the errors are multiplicative, even for a linear system. This study first explores the challenge of the multiplicative noise to the current EnKF schemes. Then, a Sigma Point Kalman Filter based Particle Filter (SPPF) is presented as an alternative to solve the issues associated with multiplicative noise. The classic Lorenz '63 model and a higher dimensional Lorenz '96 model are used as test beds for the data assimilation experiments. Performance of the SPPF algorithm is compared against a standard EnKF as well as an advanced square-root Sigma-Point Kalman Filters (SPKF). The results show that the SPPF outperforms the EnKF and the square-root SPKF in the presence of multiplicative noise. The super ensemble structure of the SPPF makes it computationally attractive compared to the standard Particle Filter (PF).

DOI:10.1029/2011MS000065

## 1. Introduction

The Ensemble Kalman Filter (EnKF) data assimilation method has attracted broad attention in the atmosphere and ocean modelling community because of its simplicity as well as its ease of implementation [e.g., *Evensen*, 2003; *Zhang and Snyder*, 2007]. The major strengths of the EnKF include: i) There is no need to calculate the tangent linear or adjoint of forecast models, which is quite difficult for General Circulation Models (GCMs). ii) The background error covariance matrix is propagated in time via the full non-linear model (no linear approximation). iii) It suits modern parallel computing [*Keppenne*, 2000]. The issues related to the standard EnKF, such as the ensemble member perturbation, have become less important with the introduction of Ensemble Square Root Kalman Filter (EnSRKF), Local Ensemble Transform Kalman Filter (LETKF), Sigma-point Kalman Filter (SPKF), and their variants [e.g., *Anderson*, 2001, 2002; *Tippett et al.*, 2003; *Hunt et al.*, 2007; *Hamill*, 2006; *Ambadan and Tang*, 2009].

Important assumptions involved in the above mentioned methods are: (i) the background (or process) and observation (or measurement) noises are additive, and (ii) associated probability density functions are Gaussian (under those assumptions, the estimate will be ''globally optimal''). However, those assumptions may not hold in reality. For example, the probability density function of daily or weekly averages of many atmospheric variables is non-Gaussian, even though long-term averages tend to follow Gaussian distribution. Recently, several studies have also shown that persistent nonlinear circulation regimes in the atmosphere and associated deviations from the Gaussian probability distributions can be modeled with multiplicative noise [*Sura and Sardeshmukh*, 2008; *Sardeshmukh and Sura*, 2009; *Sardeshmukh*, 2010]. In this case, the multiplicative noise corresponds to the state dependent variations of stochastic feedback from unresolved system components [*Sura et al.*, 2005]. Another interesting example is the stochastic parameterization method such as the stochastic kinetic energy backscatter (SKEB) scheme used in many ensemble prediction systems [*Shutts*, 2005; *Shutts et al.*, 2008; *Berner et al.*, 2008, 2009;

**To whom correspondence should be addressed.**
Youmin Tang, Environmental Science and Engineering, University of Northern British Columbia, Prince George, BC V2N 4Z9, Canada.
(ytang@unbc.ca)

*Charron et al.*, 2010]. The SKEB schemes, which are designed to account for the dissipations in the forecast model, introduce perturbations, which are state-dependent directly or indirectly. These perturbations introduce stochasticity into the model, and are expected to increase the spread of the forecast ensemble. In a broad sense, one may consider these schemes as multiplicative noises models. These stochastic schemes can create non-Gaussian statistics, and may cause the forecast model to deviate from Gaussianity. The representativeness errors due to the unresolved scales may also be considered as multiplicative noise since they are state dependent and correlated in time [*Janjic and Cohn*, 2006]. In general, the multiplicative process noise is attributed to the internally evolving dynamical and numerical errors and the observation or measurement noise corresponds to external noise. If the noise is multiplicative (state dependent) and the model is nonlinear, both the internal and external noises play an important role in the estimation statistics.

Non-Gaussianity and controlling noise have been recently an extensive research topic in data assimilation community [e.g., *Peña et al.*, 2010]. Data assimilation methods based on conditional mean estimate such as the iterated Kalman filter [*Jazwinsky*, 1970; *Cohn*, 1997] have yielded limited success in non-Gaussian scenarios. In variational assimilation methods such as 4D-var, an asymmetric cost function might be useful for assimilating non-Gaussian variables as shown by some studies [*Tsuyuki et al.*, 2003; *Koizumi et al.*, 2005; *Honda et al.*, 2005]. *Fletcher and Zupanski* [2006a, 2006b, 2007] proposed two different approaches to deal with non-Gaussian variables in a 3D variational data assimilation framework. The first approach uses a transform to make the log-normal random variable into a normal random variable and the second one uses the correct distribution for a collection of normal and log-normal random variables through a hybrid distribution which gives a different cost function to minimize. However, this approach may not be effective in all cases as shown by *Fletcher and Zupanski* [2007]. *Zupanski* [2005] developed the Maximum-Likelihood Ensemble Filter (MLEF), a hybrid filter based on variational assimilation method and the EnKF. The MLEF uses a nonlinear costfunction similar to the 3D-var and could be useful in some cases where observations are log-normally distributed. *Jardak et al.* [2010] give a comparison of the assimilation performances of MLEF, EnKF and PF under additive noise and Gaussian assumptions. Theoretically the MLEF method could be used for assimilating non-Gaussian variables. However, its performance may be different if the random variable cannot be transformed to GRV and the corresponding noises are multiplicative. The motivation of this work is to explore the EnKF based methods in the presence of multiplicative noise, and in particular, the effects of multiplicative noise on them. A Sigma-point particle filter (SPPF) will be presented and its applicability to multiplicative noise models and non-Gaussian systems will be explored.

This paper is structured as follows: Section (2) gives a general overview of parameter estimation using ensemble based Kalman filters. Section (3) introduces the Sigma-Point Particle Filter approach, while Section (4) describes experimental and implementation details of the schemes in the highly nonlinear Lorenz '63 and Lorenz '96 models. Section (5) summarizes the conclusion.

## 2. Overview of EnKF Parameter Estimation

One of the main objectives of data assimilation is to tune the parameters of a dynamical model by deterministically using observations such that they can perform more accurate simulations or predictions. Recursive parameter estimation using EnKF has garnered modelers attention and made considerable progress [*Annan and Hargreaves*, 2004; *Annan*, 2005; *Annan et al.*, 2005a, 2005b; *Hacker and Snyder*, 2005; *Aksoy et al.*, 2005, 2006a, 2006b; *Tong and Xue*, 2008a, 2008b]. *Annan* [2005] and *Annan and Hargreaves* [2004] estimated the parameters of various models using EnKF, where they introduced a preconditioning procedure and scaling to improve the error covariance matrix, which may introduce additional computational burdens. *Aksoy et al.* [2005, 2006a, 2006b] and *Tong and Xue* [2008a, 2008b] used the EnSRKF formulation where they estimated the model parameters from noisy observations. In their approach, the Kalman gain term is replaced by a scaling parameter in the state update equation, which acts as an alternative to perturbing observations in the analysis step of standard EnKF. However, there are reports that the standard EnKF generates poor parameter estimates, especially for high nonlinear systems [e.g., *Kivman*, 2003]. Recently, *Ambadan and Tang* [2009] estimated the parameters of the Lorenz '63 model using Sigma-point Kalman filters (SPKF), which use deterministic sampling of ensemble for calculating the error statistics [*Julier et al.*, 1995; *Nørgåd Magnus et al.*, 2000; *Ito and Xiong*, 2000; *Lefebvre et al.*, 2002; *Wan and Van Der Merve*, 2000; *Haykin*, 2001; *Van der Merwe et al.*, 2004]. All the above mentioned experiments were performed under the assumption that the state and observation noises are additive, and follow Gaussian distribution. In the following sections, we will introduce a recently developed hybrid particle filter data assimilation method, called Sigma-Point Particle Filter (SPPF), which use existing SPKF technique for resampling [*Van der Merwe et al.*, 2000]. We will also show that the SPPF scheme is more suitable in such situations where multiplicative noise is inherent in the model.

## 3. The Sigma-Point Particle Filter

In this section, we will briefly review the SPPF algorithm, which will serve as a theoretical background for the experiments presented in the later sections. The theory and derivations presented in this section are mainly based on the works by *Doucet et al.* [2000], *Van der Merwe et al.* [2000], *Haykin* [2001], *Arulampalam et al.* [2002], *Van der Merwe and Wan* [2001a, 2001b], *Schon* [2006], and *Simon* [2006].

Consider a stochastic process defined by a nonlinear differential equation of first order in time:

$$\dot{\theta}_k = \mathbf{f}(\theta_k) + \mathbf{g}(\theta_k)\mathbf{q}_k \tag{1}$$

where $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are in general nonlinear functions of the state $\theta_k$, and $\mathbf{q}_\mathrm{k}$ is the random force. The random force is generally considered as a zero-mean Gaussian process or white noise. In the case of additive noise $\mathbf{g}(\cdot)$ is a constant (e.g., 1.0), i.e., independent of the state $\theta_k$, and the stochastic process given by (1) is Markovian. (In recursive estimation, the states evolve in time according to a Markov process. The Markovian property implies that given the present state, the future states are independent of the past states, which is one of the primary properties of recursive Bayesian estimators such as KF.) On the other hand, in the case of multiplicative process noise, $\mathbf{g}(\cdot)$ is a linear or nonlinear function of $\theta_k$, and the process is no longer Markovian.

For the purpose of presentation, the standard state space equations for an $L$ dimensional model are given by,

$$\theta_k = \mathbf{f}(\theta_{k-1}, \mathbf{q}_{k-1}) \tag{2}$$

$$\psi_k = \mathbf{h}(\theta_k, \mathbf{r}_k) \tag{3}$$

Here $\theta_k$ is the state vector at time $k$, $\mathbf{f}(\cdot)$ is the forecast model, $\psi_k$ is the observed state, $\mathbf{h}(\cdot)$ is the observation function, and $\mathbf{q}_k$ and $\mathbf{r}_k$ are the zero-mean random noises corresponding to the background and observations respectively. Given the imperfection of model states and observations, the recursive Bayesian estimation of the state space model given by equations (2) and (3) is actually the Kalman Filter (KF), Extended Kalman Filter (EKF), EnKF, SPKF etc., under Gaussian assumption. Appendix A summarizes the least square formulation of Kalman gain, which is the core of the SPKF approach. A pre-requisite for KF is the Gaussian distribution of background and observation errors, under which the KF provides the globally optimal estimate for state-space equations. The Gaussian assumption reflects the fact that the KF is designed based on the minimization of the analysis error variance (i.e., the trace of error covariance), which ignores the higher order moments. For a non-Gaussian system, the solution by KF may not be optimal, and it could be even erroneous. In his seminal paper, *Kalman* [1960] confined the filter to linear systems and linear measurement functions. In fact, it has been shown that the standard Kalman gain used in KF, EKF and EnKF is the special case of equation (A19) when the measurement function is linear or locally linearized, and the noise is additive [*Ambadan and Tang*, 2009]. The EnKF and SPKF algorithms use the same optimality criterion in their algorithms. In the following sections, we will show that the EnKF and the SPKF failed to estimate the model parameters accurately in the presence of multiplicative noise and underlying non-Gaussian probability distribution, and in such case the SPPF assimilation scheme is found to be more accurate.

The basic idea behind the particle filter is to represent the underlying probability distribution by a set of samples known as particles, and associated weights. *Van Leeuwen* [2009] provided a clear overview of generic particle filters and of their role in geophysical estimation problems. In a broad sense the particles are similar to the ensembles in the EnKF. In a particle filter the probability density function is fully propagated in time whereas in the Kalman filter only the first and second moments are propagated in time. The probability density is approximated using an empirical function given by,

$$\begin{gathered} p(\theta_k|\psi_{1:k}) \approx \sum_{m=1}^{M} \tilde{q}_k^{(m)} \delta\left(\theta_k - \theta_k^{(m)}\right), \\ \sum_{m=1}^{M} \tilde{q}_k^{(m)} = 1, \qquad \tilde{q}_k^{(m)} \geq 0, \forall\ m \end{gathered} \tag{4}$$

where $\theta_k^{(m)}; m=1,\ldots,M$ are the independent and identically distributed (i.i.d.) particles, at time step $k$, with corresponding weights $\tilde{q}_k^{(m)}$, and $\delta(\cdot)$ is the Dirac-delta function. Here $m$ represents the particle index. Practically, it is almost impossible to get i.i.d. samples at any time $k$ from the posterior density function (4), but this limitation can be circumvented by using importance sampling from a *proposal distribution*. The choice of the *proposal distribution* is one of the most important factors in importance sampling schemes. Several strategies for choosing *proposal distribution* have been proposed in the literature. The most popular schemes include the Sampling Importance Re-sampling (SIR), the Residual sampling, and the minimum variance sampling. For further details and references, see *Gordon et al.* [1993], *Kitagawa* [1996], *Isard and Blake* [1998], *Liu and Chen* [1998], *Doucet et al.* [2000], *Doucet et al.* [2001], *Haykin* [2001], *Arulampalam et al.* [2002], and *Schon* [2006].

The SPPF, first introduced by *Van der Merwe et al.* [2000], has wide applications in robotics, and artificial intelligence. *Van der Merwe et al.* [2000] suggested that significant improvement on the particle resampling can be accomplished by using a Kalman filter for the proposal distribution. By using more advanced Kalman filters such as the square-root EnKF, or the SPKFs one can generate a better proposal distribution for the particle filter thereby propagating the statistics more accurately. The family of SPKF algorithms includes the Sigma-Point Unscented Kalman Filter (SP-UKF) [*Julier et al.*, 1995; *Wan and Van Der Merve*, 2000], Sigma-Point Central Difference Kalman Filter (SP-CDKF) [*Nørgåd Magnus et al.*, 2000;

*Ito and Xiong*, 2000] and their square root versions [*Haykin*, 2001; *Van der Merwe and Wan*, 2001a, 2001b]. *Julier et al.* [1995] has shown that for the nonlinear model given by (2), the number of sigma-points needed to compute precisely the mean and covariance of the model state at time $k$ is $2L$+1, where $L$ is the number of degrees of freedom. The selection scheme for the sigma-points for SP-UKF is based on the scaled unscented transformation, and that for the SP-CDKF is based on the *sterling's interpolation* formula [*Press et al.*, 1992; *Ito and Xiong*, 2000; *Nørgåd Magnus et al.*, 2000]. In our experiments we have used the square-root SP-CDKF for generating the proposal distribution because of its well known numerical stability [*Van der Merwe*, 2004].

In SP-CDKF the analytical derivatives in EKF are replaced by numerically evaluated *central divided differences*. For implementing the SP-CDKF, augmented state vectors are constructed by concatenating the original model state, and the background and observation error vectors. The augmented *sigma-point state vectors* are calculated using the following selection scheme:

$$\mathbf{X}_{k,0} = \bar{\theta}_k \qquad \mathbf{w}_0^{(m)} = \frac{\delta^2 - L}{\delta^2} \tag{5}$$

$$\begin{aligned} \mathbf{X}_{k,i}^{+} &= \bar{\theta}_k + \left(\sqrt{\delta^2 \mathbf{P}_{\theta_k}}\right)_i \qquad i = 1,\ldots,L \\ \mathbf{w}_i^{(m)} &= \frac{1}{2\delta^2} \qquad i = 1,\ldots,2L \end{aligned} \tag{6}$$

$$\begin{aligned} \mathbf{X}_{k,i}^{-} &= \bar{\theta}_k - \left(\sqrt{\delta^2 \mathbf{P}_{\theta_k}}\right)_i \qquad i = (L+1),\ldots,2L \\ \mathbf{w}_i^{(c_1)} &= \frac{1}{4\delta^2} \qquad i = 1,\ldots,2L \end{aligned} \tag{7}$$

$$\mathbf{w}_i^{(c_2)} = \frac{\delta^2 - 1}{4\delta^4} \qquad i = 1,\ldots,2L \tag{8}$$

where $\delta$ is the central difference step size, and $\mathbf{w}_i^{(m)}$ is the weighting term corresponding to the $i^{\text{th}}$ sigma-point for computing the mean, and $\mathbf{w}_i^{(c)}$ that for the covariance. The sigma-points are then propagated through the forecast model, and the approximated mean, covariance and cross-covariance for the calculation of Kalman gain are computed as follows:

$$\hat{\theta}_k^{-} \approx \sum_{i=0}^{2L} \mathbf{w}_i^{(m)} \mathbf{X}_{k,i}^{\theta} \tag{9}$$

$$\hat{\psi}_k^{-} \approx \sum_{i=0}^{2L} \mathbf{w}_i^{(m)} \mathbf{Y}_{k,i}^{\theta} \tag{10}$$

$$\mathbf{P}_{\theta_k}^{-} \approx \sum_{i=1}^{L} \left[ \mathbf{w}_i^{(c_1)} \left( \mathbf{X}_{k,i}^{\theta} - \mathbf{X}_{k,L+i}^{\theta} \right)^2 + \mathbf{w}_i^{(c_2)} \left( \mathbf{X}_{k,i}^{\theta} + \mathbf{X}_{k,L+i}^{\theta} - 2\mathbf{X}_{k,0}^{\theta} \right)^2 \right] \tag{11}$$

$$\mathbf{P}_{\tilde{\psi}_k}^{-} \approx \sum_{i=1}^{L} \left[ \mathbf{w}_i^{(c_1)} \left( \mathbf{Y}_{k,i}^{\theta} - \mathbf{Y}_{k,L+i}^{\theta} \right)^2 + \mathbf{w}_i^{(c_2)} \left( \mathbf{Y}_{k,i}^{\theta} + \mathbf{Y}_{k,L+i}^{\theta} - 2\mathbf{Y}_{k,0}^{\theta} \right)^2 \right] \tag{12}$$

$$\mathbf{P}_{\theta_k \tilde{\psi}_k} \approx \sum_{i=0}^{L} \mathbf{w}_i^{(m)} \left( \mathbf{X}_{k,i}^{\theta_k} - \hat{\theta}_k^{-} \right) \left( \mathbf{Y}_{k,i} - \hat{\psi}_k^{-} \right)^{\mathrm{T}} \tag{13}$$

Equations (9)–(13) form the core part for generating the proposal distribution (Sigma-point particles) for SPPF. The SP-CDKF generated proposal distribution in SPPF may be Gaussian approximate. However, it has been shown that as long as the Kalman filter generated distribution overlaps with the proposal distribution, this approximation results in a better particle filter implementation [*Van der Merwe et al.*, 2004]. One of the advantages of using the SP-CDKF for generating the proposal distribution is that it uses only one "control parameter" ($\delta$) compared to three in SP-UKF. The SPPF algorithm is summarized as follows (here we repeat the SPPF algorithm derived by *Van der Merwe* [2004]):

I. Initialization: k = 0
   For $i$ = 1 … $N$ draw particles $\theta_0^i$ from the prior $p$ ($\theta_0$)
II. For time $k$ = 1, 2…

   1. Importance sampling step:
      For $i$ = 1 … $N$:

      (a) Update the prior distribution for each particle with the SPKF

         (i) Calculate the sigma points for the particle, $\mathbf{X}_{k,i} = \begin{bmatrix} \mathbf{X}_{k,0} & \mathbf{X}_{k,j}^{+} & \mathbf{X}_{k,j}^{-} \end{bmatrix}$ where $\mathbf{X}_{k,0}$; $\mathbf{X}_{k,j}^{+}$ and $\mathbf{X}_{k,j}^{-}$ are the sigma point vectors given by

$$\mathbf{X}_{k,0} = \theta_k \tag{14}$$

$$\mathbf{X}_{k,j}^{+} = \bar{\theta}_k + \left(\sqrt{\delta^2 \mathbf{P}_{\theta_k}}\right)_i \tag{15}$$

$$\mathbf{X}_{k,j}^{-} = \bar{\theta}_k - \left(\sqrt{\delta^2 \mathbf{P}_{\theta_k}}\right)_i \tag{16}$$

(ii) Propagate sigma points in time (forecast step of SPKF):

$$\mathbf{X}_k^f = \mathbf{f}\left(\mathbf{X}_{k-1}, \mathbf{X}_{k-1}^q\right) \tag{17}$$

$$\hat{\theta}_k^- = \mathbf{w}^{(m)} \sum_{m=1}^{M} \mathbf{X}_{k,m}^f \tag{18}$$

$$\mathbf{P}_{\theta_k}^- = \sum_{i=1}^{L} \left[ \mathbf{w}_i^{(c_1)} \left( \mathbf{X}_{k,i}^f - \mathbf{X}_{k,L+i}^f \right)^2 + \mathbf{w}_i^{(c_2)} \left( \mathbf{X}_{k,i}^f + \mathbf{X}_{k,L+i}^f - 2\mathbf{X}_{k,0}^f \right)^2 \right] \tag{19}$$

(iii) Measurement update (analysis step of SPKF):

$$\mathbf{Y}_k^f = \mathbf{h}\left(\mathbf{X}_k^f, \mathbf{X}_k^r\right) \tag{20}$$

$$\hat{\psi}_k^- = \mathbf{w}^{(m)} \sum_{m=1}^{M} \mathbf{Y}_{k,m}^f \tag{21}$$

$$\mathbf{P}_{\tilde{\psi}_k}^- = \sum_{i=1}^{L} \left[ \mathbf{w}_i^{(c_1)} \left( \mathbf{Y}_{k,i}^f - \mathbf{Y}_{k,L+i}^f \right)^2 + \mathbf{w}_i^{(c_2)} \left( \mathbf{Y}_{k,i}^f + \mathbf{Y}_{k,L+i}^f - 2\mathbf{Y}_{k,0}^f \right)^2 \right] \tag{22}$$

$$\mathbf{P}_{\theta_k \tilde{\psi}_k} = \sum_{i=0}^{L} \mathbf{w}_i^{(m)} \left( \mathbf{X}_{k,i}^f - \hat{\theta}_k^- \right) \left( \mathbf{Y}_{k,i} - \hat{\psi}_k^- \right)^{\mathbf{T}} \tag{23}$$

$$\mathbf{K}_k = \mathbf{P}_{\theta_k \tilde{\psi}_k} \mathbf{P}_{\tilde{\psi}_k}^{-1} \tag{24}$$

$$\hat{\theta}_{k,i} = \hat{\theta}_k^- + \mathbf{K}_k (\psi_k - \hat{\psi}_k^-) \tag{25}$$

$$\mathbf{P}_{\hat{\theta}_{k,i}} = \mathbf{P}_{\theta_k}^- - \mathbf{K}_k \mathbf{P}_{\tilde{\psi}_k} \mathbf{K}_k^{\mathrm{T}} \tag{26}$$

(b) Sample $\mathbf{X}_{k,i} \sim \mathbf{N}\left(\hat{\theta}_{k,i}; \mathbf{P}_{\hat{\theta}_{k,i}}\right)$ (SPKF analysis distribution

For $i = 1 \ldots N$; evaluate the important weights, and normalize the weights:

$$w_{k,i} = w_{k-1,i} \frac{\text{likelihood}_{k,i} \times \text{prior}_{k,i}}{\text{proposal}_{k,i}} \tag{27}$$

$$= w_{k-1,i} \frac{p(\psi_k | \theta_{k,i})\ p(\theta_{k,i} | \theta_{k-1,i})}{p\left(\theta_{k,i} | \psi_{1:k,i}\right)} \tag{28}$$

$$\tilde{w}_{k,i} = \frac{w_{k,i}}{\sum_{j=0}^{N} w_{k,i}} \tag{29}$$

2. Resample the particles using the above weights (by multiplying with important weights)
3. Approximate the posterior distribution; and the estimate

$$\hat{\theta}_k \approx \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_{k,i} \tag{30}$$

A more detailed interpretation and derivation of the above expression is given by *Van der Merwe* [2004].

## 4. Parameter Estimation Experiments With Multiplicative Noise Models

In general the parameter estimation involves a nonlinear mapping given by

$$\mathbf{Y}_\mathbf{k} = \mathbf{N}(\theta_k, \mathbf{\Lambda}) \tag{31}$$

where the nonlinear map $\mathbf{N}(\cdot)$ may be the dynamical model $\mathbf{f}(\cdot)$ or an empirical model such as a neural network, parameterized by the vector $\mathbf{\Lambda}$. (In general $Y_k$ refers to the mapped vector (e.g., temperature) from the state vector $\theta_k$, (e.g., radiance), and $\mathbf{N}$ is the nonlinear function which is the mapping function (e.g., radiative transfer model).)

The state space model for the parameter estimation problem can be written as,

$$\mathbf{\Lambda}_k = \mathbf{\Lambda}_{k-1} + \mathbf{q}(\mathbf{\Lambda}_k) \mathbf{W}_k^\theta \tag{32}$$

$$\psi_k = \mathbf{f}(\theta_k, \mathbf{\Lambda}_k) + \mathbf{r}(\psi_k) \mathbf{W}_k^\psi \tag{33}$$

where $\mathbf{f}(\cdot)$ is the nonlinear model, $\mathbf{\Lambda}$ is the parameter vector which constitutes the dynamical parameters (or empirical parameters in the case of empirical model), $\mathbf{q}(\cdot)$ and $\mathbf{r}(\cdot)$ represent the multiplicative noise models corresponding to the model states and observations, and $\mathbf{W}^\theta$, and $\mathbf{W}^\psi$ are random white noises corresponding to the respective noise models. The state space model for the parameter estimation is similar to the state estimation except for the fact that the state (here states are parameters) time evolution is linear (equation (32)) and the measurement function is nonlinear (equation (33)). In this particular situation equation (32) may be considered as a linear stochastic system with multiplicative forcing. In the following subsection we will use the *Lorenz* [1963, 1996] models as test beds for our parameter estimation experiments. In all the experiments the state observations are related to the model parameters through the nonlinear dynamical model.

### 4.1. Experiments With Lorenz '63 Model

In the data assimilation community, the *Lorenz* [1963] model has served as a test bed for examining the properties of various data assimilation methods as the model shares many common features with the atmospheric circulation and the climate system in terms of variability and predictability [*Gauthier*, 1992; *Palmer*, 1993; *Miller et al.*, 1994; *Evensen*, 1997]. The model can be used to simulate nearly-regular oscillations or irregular chaotic variations by adjusting the model parameters that control the non-linearity of the system. In our experiments, we used a modified version of the standard *Lorenz* [1963] model with additional noise terms, given by

$$\frac{dx}{dt} = \sigma(y - x) + q(x)w^x \tag{34}$$

$$\frac{dy}{dt} = \rho x - y - xz + q(y)w^y \tag{35}$$

$$\frac{dz}{dt} = xy - \beta z + q(z)w^z \tag{36}$$

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, and the parameters $\sigma$, $\rho$, and $\beta$ will be referred to as dynamical parameters. $q(\cdot)$ represents the state dependent (multiplicative) background errors, and $w$ is the Gaussian white noise. The true data are created by integrating the model using the fourth-order Runge-Kutta scheme [*Press et al.*, 1992], with parameters $\sigma$, $\rho$, and $\beta$ set to 10.0, 28.0, and 8/3 respectively, and initial conditions set to 1.508870, –1.531271, and 25.46091 as by *Miller et al.* [1994] and *Evensen* [1997].

To apply the assimilation algorithms, we discretize the nonlinear Lorenz model (34)–(36) using the fourth-order Runge-Kutta method and write it in the form of state space equations given by (32) and (33), where $\theta_k$ represents the system state vector (a column vector composed of $x$, $y$ and $z$), $\mathbf{f}(\cdot)$ is the Lorenz model and $\mathbf{q}_k$ is the random (white) process noise vector (column vector composed of $q^x$, $q^y$ and $q^z$). The measurement function $\psi_k$, required for the application of the EnKF parameter estimation, is the nonlinear model itself, connecting the state observations and model parameters.

For all the experimental cases (involving multiplicative noise) to be discussed, the observation data sets are generated by setting:

$$\mathbf{q}(\theta_k)\mathbf{w}_k^\theta = C_m \theta_k \mathbf{w}^\theta \tag{37}$$

where $C_m$ is a constant, called the multiplicity factor, which determines the strength of the state influence in the multiplicative noise. $\theta$ is the system state vector, and $w^\theta$ is the normally distributed white noise $N(0, \sqrt{2})$. (In our study we focus only on linear multiplicative noise model where $C_m$ is a constant. However in many real situations such as the stochastic kinetic energy backscatter (SKEB) schemes, the multiplicative noise models are nonlinear in general.) This white noise distribution is similar to that in the additive noise experiments of *Miller et al.* [1994] and *Evensen* [1997]. The observation interval is set to 25, i.e., the observations are assimilated to the model at every 25 steps.

Two particular cases were studied. Case 1: the background noise (or internal noise) is additive and the observation noise (or external noise) is multiplicative. Case 2: both the background and the observation noises are multiplicative.

We assume that the parameter $\rho$ is uncertain. Our task is to estimate the correct value of $\rho$ from infrequent noisy observations using a noisy model. In Case 1 the observations are generated using the multiplicative noise model given by equation (37), where the multiplicity factor $C_m$ is set to 0.02. Figure 1a shows the distribution of the variables $X$ and $Z$, and Figure 1b shows the distribution of the corresponding additive and multiplicative noises used in the experiments. It is clear from the probability plot that the observations (Figure 1a) are non-Gaussian. The multiplicative noise shown in Figure 1b also shows non-Gaussian features, and is symmetric. However, the symmetric nature may not be the case for real observations. In all our experiments, we set the initial parameter to zero.

Figure 2a shows the parameter estimation results using the EnKF scheme. The number of ensemble members used in the experiments is 100. Similarly, Figure 2b shows the results using the square-root SP-CDKF, which uses $2L+1$ sigma-points for the estimation. As can be seen in Figure 2a the EnKF scheme failed to estimate the parameter. On the other hand the performance of the square-root SP-CDKF is better but the parameter is still slightly overestimated as shown in Figure 2b. It should be noted that the estimate might be sensitive to the initial guess. In fact the performance of the SP-CDKF can be adjusted by tuning the central difference parameter. However in all our trial experiments the SP-CDKF either underestimates or overestimates the true parameter even though it converges very fast compared to other Kalman filter schemes.

We repeated the experiment with the SPPF scheme, which is a hybrid Particle filter-Kalman filter. The results are shown in Figure 2c, and they are remarkably better compared to any of the Kalman filter based assimilation schemes including the advanced square-root filters.

In Case 2, the situation is much more realistic and may give rise to complex non-Gaussian distribution. Here we focus only on the performances of the advanced square root SP-CDKF and the SPPF since the role of generic EnKF methods are in this case irrelevant. Figures 3a and 3b show the results of the experiments, which are similar to Case 1 where the square root SP-CDKF overestimates/underestimates the model parameter, and the SPPF scheme estimates the parameter with better accuracy.

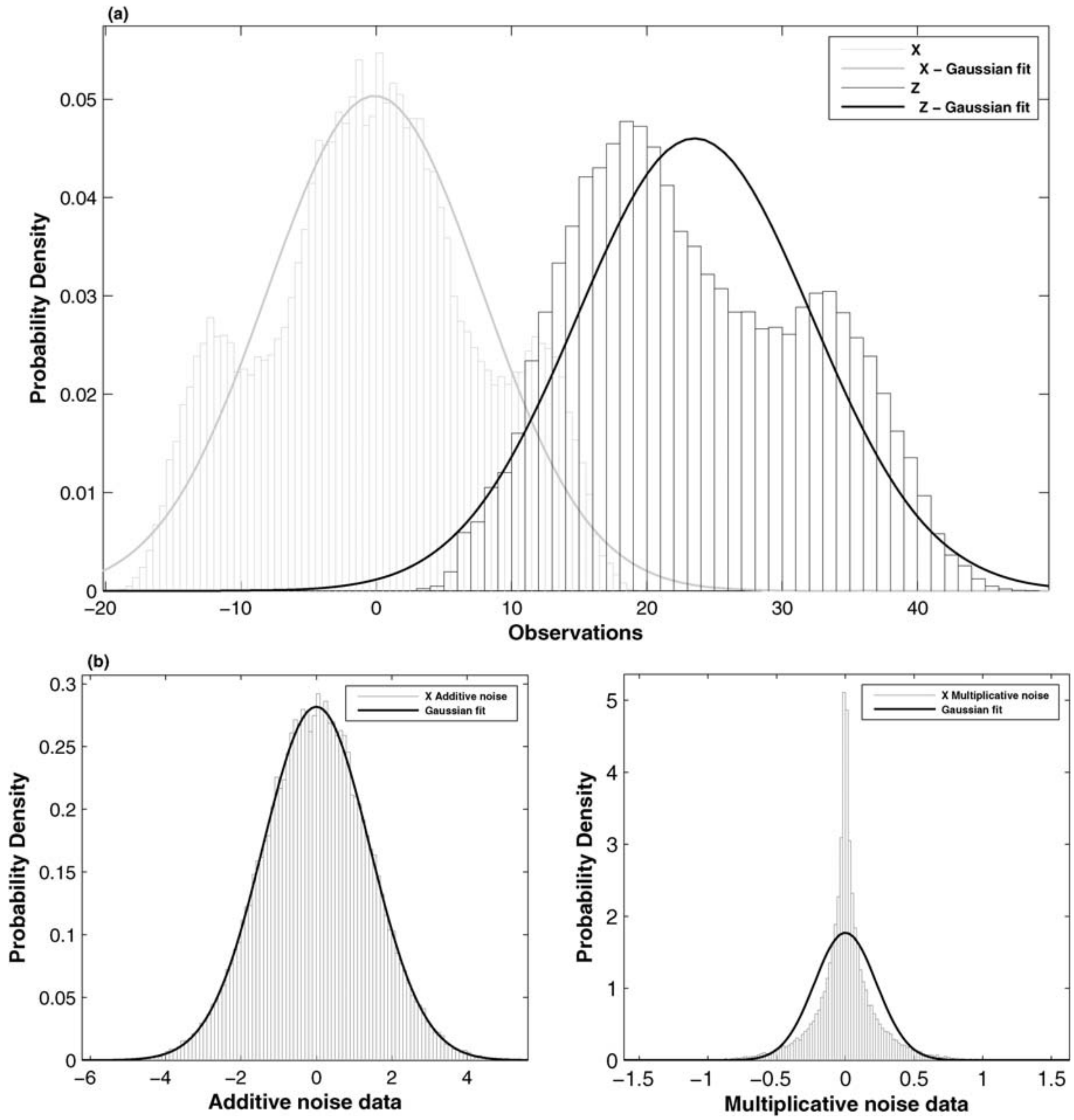


**Figure 1.** (a) Observation distribution: gray - X, and black - variable Z. (b) Noise distributions for X: (left) additive noise, (right) multiplicative noise; solid curves represent corresponding Gaussian fits.

To study the effect of the multiplicity factor on the SPPF assimilation scheme, we have increased $C_m$ from 0.02 to 0.2. The results of the experiments are shown in Figure 3c. From the figure it can be concluded that irrespective of the strength of the multiplicative noise, the SPPF scheme was able to estimate the parameter accurately. Table 1 gives the Root Mean Squared Error (RMSE) values of all the above experiments, which in general confirms the results from the figures.

In summary, we have investigated the merits and de-merits of different Kalman filter based ensemble data assimilation schemes in a multiplicative model noise environment, using the low-dimensional Lorenz '63 model. Important features in evaluating the performance of a data assimilation algorithm are its robustness and computational expense as they can become issues when it is applied to higher dimensional models. In the following section we will further explore the above mentioned schemes using the higher dimensional Lorenz '96 model.

## 4.2. Experiments With Lorenz '96 Model

To explain the dynamics of weather at a fixed latitude, *Lorenz* [1996] introduced a one dimensional atmosphere

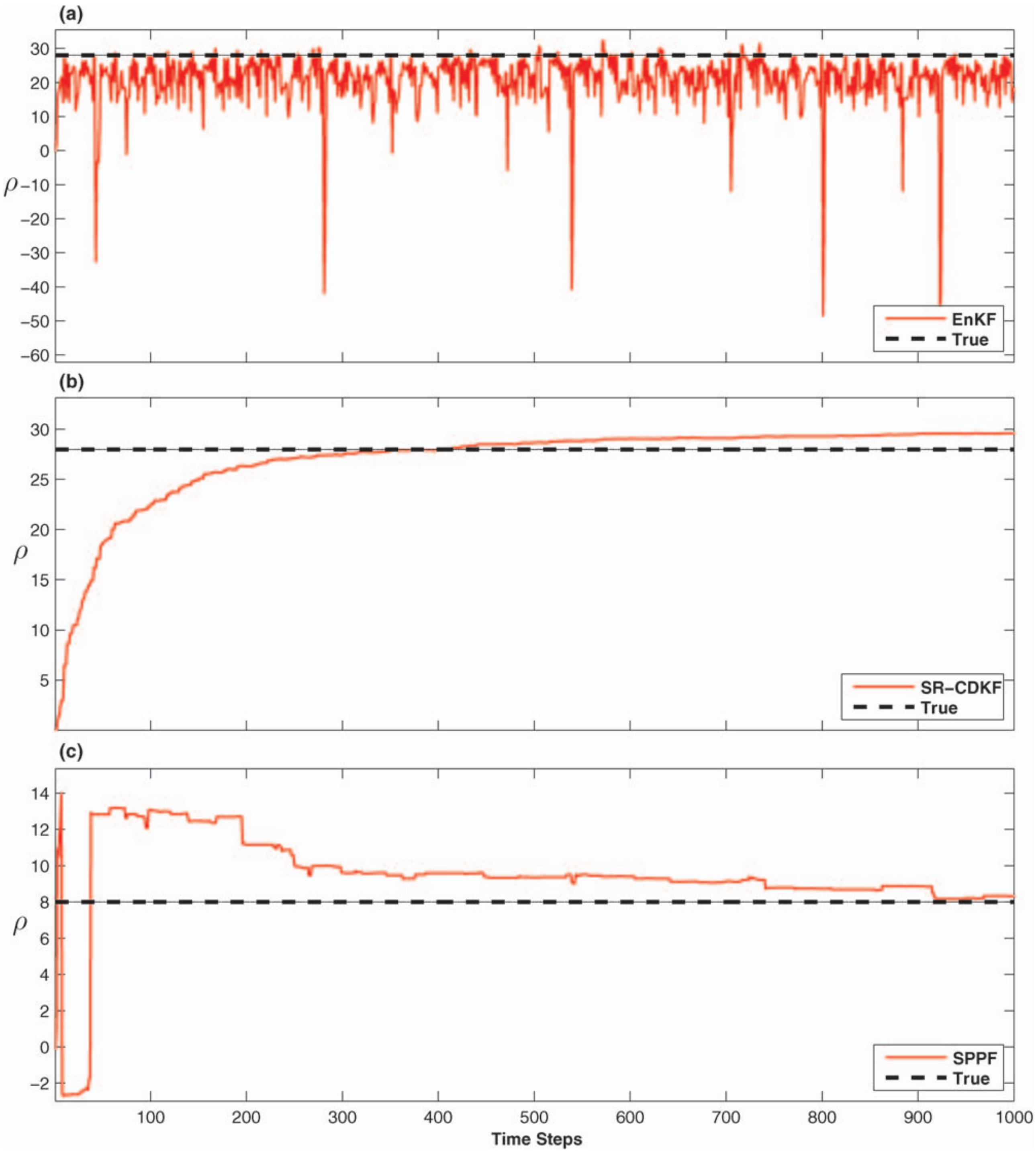


**Figure 2.** Lorenz '63 model: (a) EnKF and (b) SR-CDKF, and (c) SPPF with 100 particles. True $\rho$ - dashed line, estimated $\rho$ - solid red line.

that shares similar error growth characteristics as full Numerical Weather Prediction (NWP) models. In our experiments, we used a modified version of the model containing $K$ variables $X_1, \cdots, X_k$, which may be thought of as atmospheric variables in $K$ sectors of a latitude circle, governed by,

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F + \underbrace{q(X_k)w^{X_k}} \qquad (38)$$

where the constant $F$ is called the forcing term. The last term (under-bracketed expression) in equation (38) forms the noise model, which is given by equation (37). By using cyclic

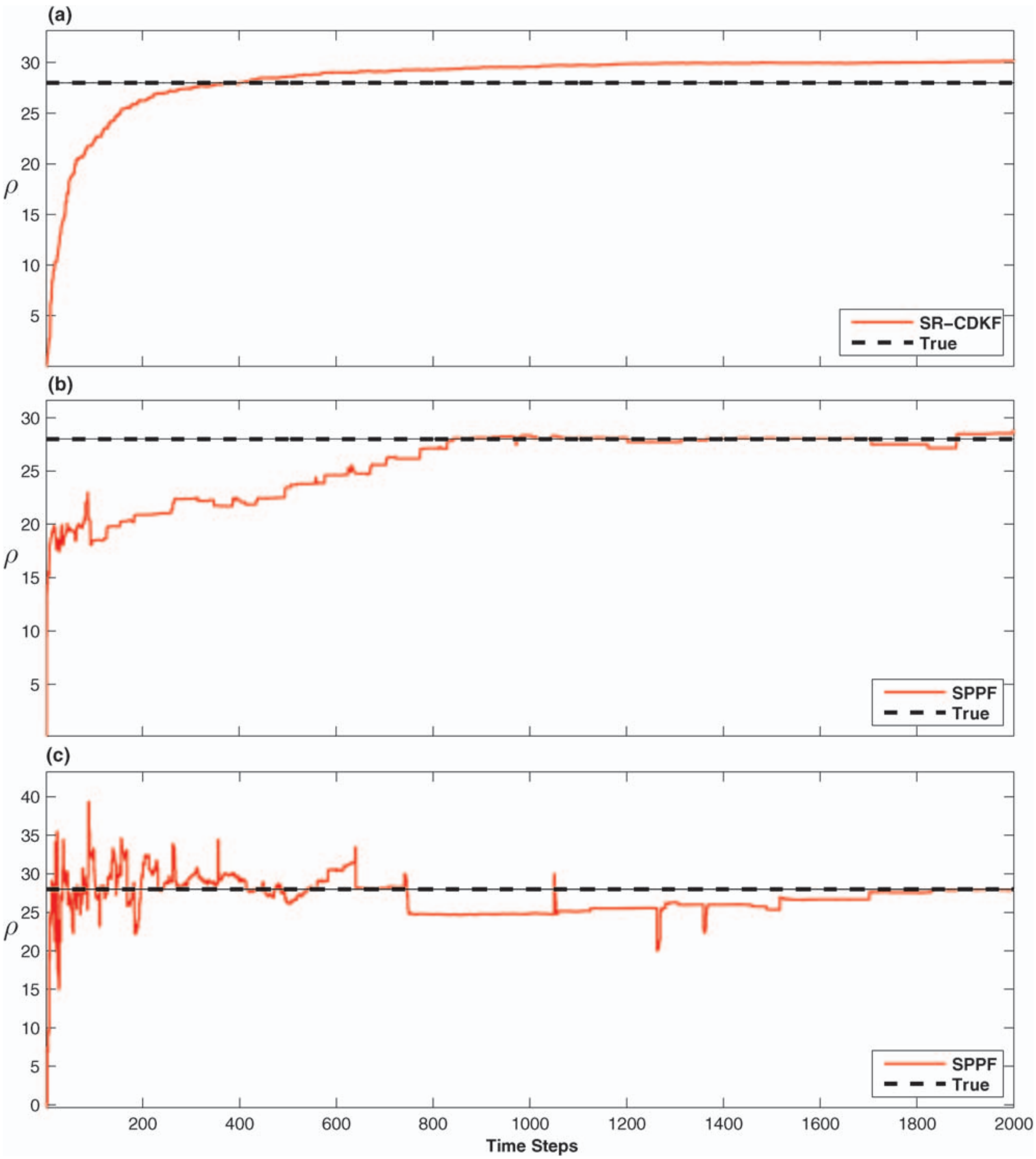


**Figure 3.** Lorenz '63 model: (a) SR-CDKF and (b) SPPF using 100 particles. (c) SPPF with a higher multiplicity factor of 0.2. True $\rho$ - dashed line, estimated $\rho$ - solid red line.

boundary conditions, the definition of $X_k$ is extended to all values of $k$; i.e., $X_{k-K}$ and $X_{k+K}$ equal to $X_k$. It is assumed that a unit time $\Delta t=1$ corresponds to 5 days.

The experimental setup is similar to that of *Lorenz and Emmanuel* [1998], where $K$=40 and the magnitude of the forcing term is set to 8 for which the system is chaotic. The system was integrated using fourth-order Runge-Kutta scheme, with integration step $\Delta t=0.05$ (i.e., 6 hours). The experiments were carried out with random initial conditions, and the observations were generated by applying the noise model to the true model. For different case studies the strength of the multiplicative noise was controlled by setting the multiplicity factor either to 0.02 (weak case) or 0.2 (strong case). Also the observation interval was set to 5, i.e.,

**Table 1.** Parameter Estimation: Root Mean Squared Error

| Assimilation Method | Observation Error | Background Error | Multiplicity Factor, $C_m$ | RMSE |
|---|---|---|---|---|
| EnKF (L63, Figure 2a) | Multiplicative | Additive | 0.02 | 10.2340 |
| SR-CDKF (L63, Figure 2b) | Multiplicative | Additive | 0.02 | 4.5692 |
| SPPF (L63, Figure 2c) | Multiplicative | Additive | 0.02 | 3.4762 |
| SR-CDKF (L63, Figure 3a) | Multiplicative | Multiplicative | 0.02 | 3.5363 |
| SPPF (L63, Figure 3b) | Multiplicative | Multiplicative | 0.02 | 3.8475 |
| SPPF (L63, Figure 3c) | Multiplicative | Multiplicative | 0.2 | 2.6334 |
| SR-CDKF (L96, Figure 4a) | Multiplicative | Multiplicative | 0.02 | 1.4193 |
| SPPF (L96, Figure 4b) | Multiplicative | Multiplicative | 0.02 | 2.9593 |
| SR-CDKF (L96, Figure 4c) | Multiplicative | Multiplicative | 0.2 | 1.5403 |
| SPPF (L96, Figure 4d) | Multiplicative | Multiplicative | 0.2 | 3.0108 |

the observed states are assimilated to the nonlinear model at every 5 steps. A complete discussion of the Lorenz '96 model is give by *Lorenz* [1996], *Lorenz and Emmanuel* [1998], and *Lorenz* [2005, 2006a, 2006b].

Here we focus on a case study where both the model and measurement noises are multiplicative, which is similar to the second case study using the Lorenz '63 model described in the previous subsection. In all experiments in this section, we assume that observations of all the states are available, and the forcing term $F$ is uncertain. Initially we set the forcing term $F$ to zero, and our aim is to estimate the actual forcing term $F$ accurately from the observed state variables so that we will be able to tune the dynamical model for a more accurate prediction.

Estimation results are shown in Figures 4a and 4b respectively. These results imply that the pure Kalman filter based methods either underestimate or over-estimate the parameter. It is due to the fact that pure Kalman filter based optimal estimation methods rely only on the first two moments, which are insufficient for estimating non-Gaussian statistics. In all the cases described above the SPPF scheme is very successful in estimating the parameters with reasonable accuracy.

The results of experiments using a higher multiplicity factor are shown in Figures 4c and 4d. The results once again re-iterate the fact that pure Kalman filter methods fail in non-Gaussian scenarios whereas the hybrid SPPF scheme estimates the parameter accurately. However, the RMSE values corresponding to the Lorenz '96 model experiments are relatively higher than those of the square-root SP-CDKF. This is due to an initial fluctuation in SPPF estimate. The RMSE values may get smaller for SPPF if one takes a longer assimilation period, since the SP-CDKF converges to an under-estimated value (almost constant) after a certain assimilation steps.

## 5. Discussions and Conclusion

Over the last decade, the data assimilation community made significant progresses towards the development and application of ensemble based Kalman filter data assimilation schemes. The EnKF and its derivatives have been widely applied to various fields, in particular atmosphere and ocean sciences. However, a preliminary limit imposed in carrying out all the above mentioned Kalman filters is that the states, observations and associated noise models should follow a Gaussian distribution. On the other hand, the multiplicative noise typically introduced in some systems may cause non-Gaussianity, which is a major concern for the Kalman filter based ensemble data assimilation, and has not been well addressed in the literature. Recently, *Anderson* [2010] introduced the Ranked Histogram Filter (RHF), which is a promising workaround to deal with non-Gaussian observation space. Notwithstanding those improvements, the Kalman filter based methods still lacks the ability to handle non-Gaussian statistics. In such cases, hybrid methods may be more useful.

In this paper, we have explored the impacts of multiplicative noise on ensemble based Kalman filter data assimilation methods in the context of parameter estimation problems. In parameter estimation in the presence of multiplicative noise, the nonlinearity of the measurement function also plays an important role. Our experiments show that all ensemble based Kalman filters, including EnKF, SPKF and square root SPKFs, either underestimate or overestimate the parameter, sometimes even diverging from the true value. The main reason for their poor performance is the fact that the multiplicative noise causes the system to deviate from Gaussianity. In such situations, it is difficult to approximate the statistical moments in a closed form, which is the necessary and sufficient condition for global optimality of the EnKFs.

Further, we introduced the recently developed SPPF scheme to the assimilation problem involving multiplicative noise. In the SPPF scheme, the particles are resampled using the SPKF scheme. Using a three-variable Lorenz '63, and a forty-variable Lorenz '96 model, we explored the merits and properties of SPPF. The results showed that the SPPF scheme can estimate the model parameters with reasonable accuracy and better than ensemble Kalman filters. The main advantages of using the hybrid method are that the number of particles is significantly reduced compared to the SIR particle filter, and the method works well in a multiplicative noise environment.

In our experiments, we assume that the dynamical parameters are stationary, and do not change with time. It is

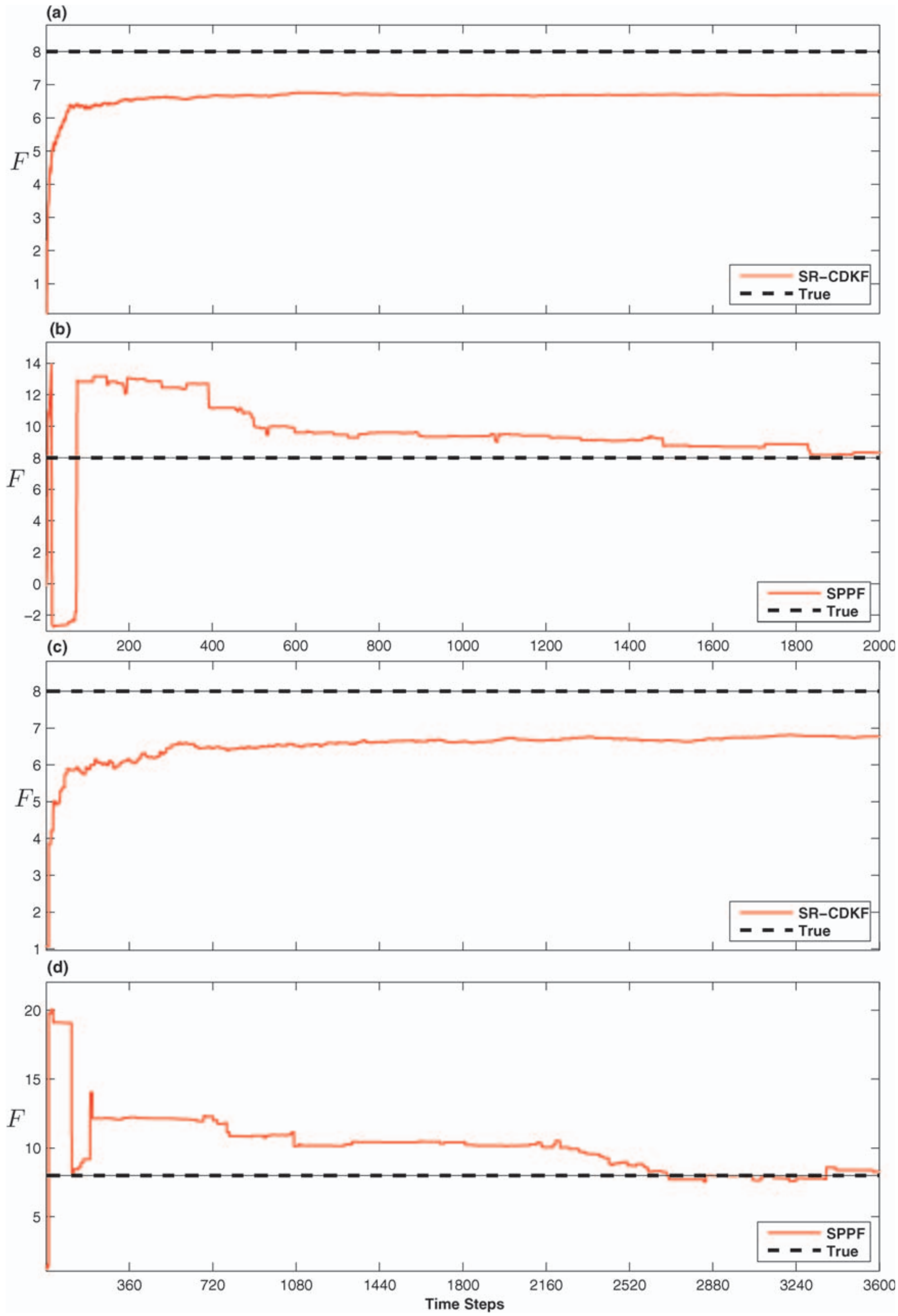


**Figure 4.** Lorenz '96 model: (a) SR-CDKF, (b) SPPF, and (c) SR-CDKF with a higher multiplicity factor of 0.2. (d) SPPF with a higher multiplicity factor of 0.2. True *F* - dashed line, estimated *F* - solid red line.

a common approach in parameter estimation using the Kalman filters [*Annan and Hargreaves*, 2004; *Annan*, 2005; *Annan et al.*, 2005a, 2005b; *Hacker and Snyder*, 2005; *Aksoy et al.*, 2005, 2006a, 2006b; *Tong and Xue*, 2008a, 2008b]. Other researchers have noted that the time series defined by equation (32) may not be stationary [*Dee*, 1995; *Evensen et al.*, 1998]. This lack of stationarity may add further complexity into the estimation problem. Besides, it may be possible that large differences in the initial parameter value may place the system in qualitatively different regimes. In fact, in such models, the original state has stable equilibriums (or stable limit cycles) while the true state was chaotic. We would imagine such cases would present a special challenge to any state-of-the art data assimilation technique.

Another interesting issue is the computational expense of the SPPF algorithm. In a broad sense, one may consider the SPPF scheme as a super-ensemble technique, where each sample is estimated through a subset of sigma-points and resampled accordingly. Compared to EnKFs and SPKFs, the computational requirement of SPPF is larger. However, the super-ensemble structure of the SPPF algorithm is highly parallelizable, and in theory, one can manage the computing time with the expense of more computing resources (more processors). On the other hand, the hybrid approach may help many researchers to use the existing EnKF based assimilation packages such as the Data Assimilation Research Test bed (DART) [*Anderson et al.*, 2009], which is optimized for many GCMs. It has no doubt that much additional research is required before applying the SPPF technique to highly dimensional systems like the GCMs.

In conclusion, in this study we have demonstrated that hybrid methods such as the SPPF can overcome the drawbacks of pure Kalman filter based ensemble data assimilation methods in the presence of multiplicative noise, and associated deviations from Gaussianity. Issues related to SPPF do not seem to impede their applications to high complexity models.

## Appendix A: Least Square Formulation of Kalman Gain

The state update equation for the state space model (2)–(3) is given by,

$$\hat{\theta}_k = \hat{\theta}_k^- + \mathbf{K}_k(\psi_k - \hat{\psi}_k^-) \quad \text{(A1)}$$

where $\mathbf{K}_k$ is the Kalman gain. The superscript "–" represents the prior states given by the following equations:

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

$$\hat{\psi}_k^- = \mathbf{E}\left[\mathbf{h}(\theta_k^-, \mathbf{r}_k)\right] \quad \text{(A3)}$$

where $\mathbf{E}[\cdot]$ represents the mathematical expectation or the expected value.

In general, the estimation error is defined as,

$$\tilde{\theta}_k = \theta_k - \hat{\theta}_k \quad \text{(A4)}$$

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

$$\tilde{\psi}_k = \psi_k - \hat{\psi}_k^- \quad \text{(A5)}$$

Substituting (A4) into the state update equation (A1), we can rewrite the estimation error as

$$\tilde{\theta}_k = \theta_k - \hat{\theta}_k^- - \mathbf{K}_k(\psi_k - \hat{\psi}_k^-) \quad \text{(A6)}$$

$$= \tilde{\theta}_k^- - \mathbf{K}_k\tilde{\psi}_k \quad \text{(A7)}$$

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

$$\mathbf{E}\left[\tilde{\psi}_k\right] = 0 \quad \text{(A8)}$$

$$\mathbf{E}\left[\tilde{\theta}_k\right] = 0 \quad \text{(A9)}$$

Now, the state error covariance, $\mathbf{P}_{\theta_k}$ and the cross covariance, $\mathbf{P}_{\theta_k\tilde{\psi}_k}$ between the state and observation error can be rewritten in terms of equations (A4) and (A5) and are given by

$$\mathbf{P}_{\theta_k} = \mathbf{E}\left[\tilde{\theta}_k\tilde{\theta}_k^{\mathrm{T}}\right] \quad \text{(A10)}$$

$$\mathbf{P}_{\theta_k\tilde{\psi}_k} = \mathbf{E}\left[\tilde{\theta}_k^- \tilde{\psi}_k^{\mathrm{T}}\right] \quad \text{(A11)}$$

Taking the outer products and expectation of (A7) produces

$$\mathbf{E}\left[\tilde{\theta}_k\tilde{\theta}_k^{\mathrm{T}}\right] = \mathbf{E}\left[\left(\tilde{\theta}_k^- - \mathbf{K}_k\tilde{\psi}_k\right)\left(\tilde{\theta}_k^- - \mathbf{K}_k\tilde{\psi}_k\right)^{\mathrm{T}}\right] \quad \text{(A12)}$$

$$= \mathbf{E}\left[\tilde{\theta}_k^- \tilde{\theta}_{k-}^{\mathrm{T}}\right] - \mathbf{E}\left[\tilde{\theta}_k^- \tilde{\psi}_{k-}^{T}\mathbf{K}_k^{\mathrm{T}}\right] - \mathbf{E}\left[\mathbf{K}_k\tilde{\psi}_k\tilde{\theta}_{k-}^{\mathrm{T}}\right] + \mathbf{E}\left[\mathbf{K}_k\tilde{\psi}_k\tilde{\psi}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}\right] \quad \text{(A13)}$$

Using equations (A10) and (A11), equation (A13) can be rewritten as

$$\mathbf{P}_{\theta_k} = \mathbf{P}_{\theta_k}^- - \mathbf{P}_{\theta_k\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}} - \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k\theta_k} + \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}} \quad \text{(A14)}$$

Our aim is to minimize the trace of $\mathbf{P}_{\theta_k}$ for the unbiased estimator, i.e.,

$$\frac{\partial}{\partial \mathbf{K}_k}(\mathrm{Tr}(\mathbf{P}_{\theta_k})) = 0 \quad \text{(A15)}$$

We have

$$\mathrm{Tr}(\mathbf{P}_{\theta_k}) = \mathrm{Tr}\left(\mathbf{P}_{\theta_k}^{-} - \mathbf{P}_{\theta_k\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}} - \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k\theta_k} + \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}}\right) \quad (A16)$$

$$= \mathrm{Tr}\left[\left(\mathbf{K}_k - \mathbf{P}_{\theta_k\tilde{\psi}_k}\mathbf{P}_{\tilde{\psi}_k}^{-1}\right)\mathbf{P}_{\tilde{\psi}_k}\left(\mathbf{K}_k - \mathbf{P}_{\theta_k\tilde{\psi}_k}\mathbf{P}_{\tilde{\psi}_k}^{-1}\right)^{\mathrm{T}}\right] + \quad (A17)$$

$$\mathrm{Tr}\left(\mathbf{P}_{\theta_k}^{-} - \mathbf{P}_{\theta_k\tilde{\psi}_k}\mathbf{P}_{\tilde{\psi}_k}^{-1}\mathbf{P}_{\theta_k\tilde{\psi}_k}^{\mathrm{T}}\right) \quad (A18)$$

We want to choose $\mathbf{K}_k$ in order to minimize (A14). It can be easily verified that the above expression (here we have used the principle $\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$) (A14) is minimum when

$$\mathbf{K}_k = \mathbf{P}_{\theta_k\tilde{\psi}_k}\mathbf{P}_{\tilde{\psi}_k}^{-1} \quad (A19)$$

Here we have used the following identities,

$$\frac{\partial}{\partial \mathbf{A}}\left(\mathrm{Tr}\left(\mathbf{A}\mathbf{B}\mathbf{A}^{\mathrm{T}}\right)\right) = 2\mathbf{A}\mathbf{B} \quad (A20)$$

where $\mathbf{B}$ is symmetric, and

$$\frac{\partial}{\partial \mathbf{A}}\left(\mathrm{Tr}\left(\mathbf{A}\mathbf{C}^{\mathrm{T}}\right)\right) = \frac{\partial}{\partial \mathbf{A}}\left(\mathrm{Tr}\left(\mathbf{C}^{\mathrm{T}}\mathbf{A}\right)\right) = \mathbf{C} \quad (A21)$$

Substituting the expression for Kalman gain, given by equation (A19) back into the expression for the error covariance (A14), the covariance update equation is given by

$$\mathbf{P}_{\theta_k} = \mathbf{P}_{\theta_k}^{-} - \mathbf{K}_k\mathbf{P}_{\tilde{\psi}_k}\mathbf{K}_k^{\mathrm{T}} \quad (A22)$$

It has been shown that the standard Kalman gain used in KF, EKF and EnKF is a special case of equation (A19) when the measurement function is linear or locally linearized, and the noise is additive [*Ambadan and Tang*, 2009].

**Acknowledgments.** This work was supported by a NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant and an open grant of the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, P. R. China. The main work was completed during the period that J. T. Ambadan's visit Youmin Tang in summer 2010. J. T. Ambadan would like to acknowledge and thank the International Max-Planck Research School on Earth System Modelling (IMPRS-ESM), Max-Planck Institute for Meteorology (MPI-M), and Luis Kornblueh for making this visit possible, and for their continued support. The authors would also like to acknowledge and thank the Oregon Graduate Institute, Eric A. Wan and Rudolph Van der Merwe for providing the ReBEL (Recursive Bayesian Estimation library, 2003, available at http://choosh.csee.ogi.edu/rebel/index.html) tool kit, part of which has been used in this research work.

## References

Aksoy, A., F. Zhang, J. W. Nielsen-Gammon, and C. C. Epifanio (2005), Ensemble-based data assimilation for thermally forced circulations, *J. Geophys. Res.*, *110*, D16105, doi:10.1029/2004JD005718.

Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon (2006a), Ensemble-based simultaneous state and parameter estimation with MM5, *Geophys. Res. Lett.*, *33*, L12801, doi:10.1029/2006GL026186.

Aksoy, A., F. Zhang, and J. Neilsen-Gammon (2006b), Ensemble based simultaneous state and parameter estimation in a two-dimensional sea-breeze model, *Mon. Weather Rev.*, *134*, 2951–2970, doi:10.1175/MWR3224.1.

Ambadan, J., and Y. Tang (2009), Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems, *J. Atmos. Sci.*, *66*, 261–285, doi:10.1175/2008JAS2681.1.

Anderson, J. (2010), A non-Gaussian ensemble filter update for data assimilation, *Mon. Weather Rev.*, *138*, 4186–4198, doi:10.1175/2010MWR3253.1.

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano (2009), The data assimilation research testbed: A community facility, *Bull. Am. Meteorol. Soc.*, *90*(9), 1283–1296, doi:10.1175/2009BAMS2618.1.

Anderson, J. L. (2001), An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, *129*, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Anderson, J. L. (2002), A local least square framework for ensemble filtering, *Mon. Weather Rev.*, *131*, 634–642, doi:10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

Annan, J. (2005), Parameter estimation using chaotic time series, *Tellus, Ser. A*, *57*, 709–714, doi:10.1111/j.1600-0870.2005.00143.x.

Annan, J., and J. Hargreaves (2004), Efficient parameter estimation for a highly chaotic system, *Tellus, Ser. A*, *56*, 520–526, doi:10.1111/j.1600-0870.2004.00073.x.

Annan, J., J. Hargreaves, N. Edwards, and R. Marsh (2005a), Parameter estimation in an intermediate complexity Earth system model using an ensemble Kalman filter, *Ocean Modell.*, *8*, 135–154, doi:10.1016/j.ocemod.2003.12.004.

Annan, J., D. Lunt, J. Hargreaves, and P. Valdes (2005b), Parameter estimation in an atmospheric GCM using the ensemble Kalman filter, *Nonlinear Processes Geophys.*, *12*, 363–371, doi:10.5194/npg-12-363-2005.

Arulampalam, M., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filter for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.*, *50*(2), 188–197, doi:10.1109/78.978374.

Berner, J., F. Doblas-Reyes, T. Palmer, G. Shutts, and A. Weisheimer (2008), Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model, *Philos. Trans. R. Soc. A*, *366*, 2559–2577, doi:10.1098/rsta.2008.0033.

Berner, J., G. Shutts, M. Leutbecher, and T. Palmer (2009), A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system, *J. Atmos. Sci.*, *66*, 603–626, doi:10.1175/2008JAS2677.1.

Charron, M., L. Spacek, P. Houtekamer, H. Mitchell, L. Michelin, G. Pellerin, and N. Gagnon (2010), Towards random sampling of model error in the Canadian ensemble prediction system, *Mon. Weather Rev.*, *138*, 1877–1901, doi:10.1175/2009MWR3187.1.

Cohn, S. (1997), An introduction to estimation theory, *J. Meteorol. Soc. Jpn.*, *75*, 147–178.

Dee, D. P. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, *123*, 1128–1145, doi:10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2.

Doucet, A., S. Godsil, and C. Andrieu (2000), On sequential Monte-Carlo sampling methods for Bayesian filtering, *Stat. Comput.*, *10*, 197–208, doi:10.1023/A:1008935410038.

Doucet, A., N. De Freitas, and N. Gordon (2001), *Sequential Monte Carlo Methods in Practice*, 1st ed., 581 pp., Springer, New York.

Evensen, G. (1997), Advanced data assimilation for strongly nonlinear dynamics, *Mon. Weather Rev.*, *125*, 1342–1354, doi:10.1175/1520-0493(1997)125<1342:ADAFSN>2.0.CO;2.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367, doi:10.1007/s10236-003-0036-9.

Evensen, G., D. Dee, and J. Schröter (1998), Parameter estimation in dynamical models, in *Ocean Modeling and Parameterization, NATO ASI Ser., Ser. C,* vol. *516*, edited by E. P. Chassignet and J. Verron, pp. 373–398, Kluwer Acad., Dordrecht, Netherlands.

Fletcher, S., and M. Zupanski (2006a), A data assimilation method for log-normally distributed observational errors, *Q. J. R. Meteorol. Soc.*, *132*(621), 2505–2519, doi:10.1256/qj.05.222.

Fletcher, S., and M. Zupanski (2006b), A hybrid multivariate normal and lognormal distribution for data assimilation, *Atmos. Sci. Lett.*, *7*(2), 43–46, doi:10.1002/asl.128.

Fletcher, S., and M. Zupanski (2007), Implications and impacts of transforming lognormal variables into normal variables in VAR, *Meteorol. Z.*, *16*, 755–765, doi:10.1127/0941-2948/2007/0243.

Gauthier, P. (1992), Chaos and quadric-dimensional data assimilation: A study based on the Lorenz model, *Tellus, Ser. A*, *44*, 2–17, doi:10.1034/j.1600-0870.1992.00002.x.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993), Novel approach to nonlinear or non-Gaussian Bayesian filtering, *IEE Proc. F*, *140*, 107–113.

Hacker, J., and C. Snyder (2005), Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL, *Mon. Weather Rev.*, *133*, 3260–3275, doi:10.1175/MWR3022.1.

Hamill, T. M. (2006), Ensemble-based atmospheric data assimilation, in *Predictability of Weather and Climate*, edited by T. Palmer and R. Hagedorn, pp. 124–156, Cambridge Univ. Press, Cambridge, U. K.

Haykin, E. S. (2001), *Kalman Filtering and Neural Networks*, 128 pp., Wiley, Hoboken, N. J.

Honda, Y., M. Nishijima, K. Koizumi, Y. Ohta, K. Tamiya, T. Kawabata, and T. Tsuyuki (2005), A pre-operational variational data assimilation system for a non-hydrostatic model at the Japan Meteorological Agency: Formulation and preliminary results, *Q. J. R. Meteorol. Soc.*, *131*(613), 3465–3475, doi:10.1256/qj.05.132.

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

Isard, M., and A. Blake (1998), in A smoothing filter for condensation, *Proceedings of 5th European Conference on Computer Vision, Lect. Notes Comput. Sci.*, vol. *1406*, pp. 767–781, Springer, New York.

Ito, K., and K. Xiong (2000), Gaussian filters for nonlinear filtering problems, *IEEE Trans. Autom. Control*, *45*(5), 910–927, doi:10.1109/9.855552.

Janjic, T., and S. Cohn (2006), Treatment of observation error due to unresolved scales in atmospheric data assimilation, *Mon. Weather Rev.*, *134*, 2900–2915, doi:10.1175/MWR3229.1.

Jardak, M., I. M. Navon, and M. Zupanski (2010), Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation, *Int. J. Numer. Methods Fluids*, *62*(4), 374–402.

Jazwinsky, A. (1970), *Stochastic Processes and Filtering Theory*, Academic, New York.

Julier, S., J. Uhlmann, and H. Durrant-Whyte (1995), A new approach for filtering nonlinear systems, in *Proceedings of the American Control Conference*, pp. 1628–1632, Inst. Electr. and Electron. Eng., Seattle, Wash.

Kalman, R. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME, Ser. D.*, *82*, 35–45.

Keppenne, C. L. (2000), Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter, *Mon. Weather Rev.*, *128*, 1971–1981, doi:10.1175/1520-0493(2000)128<1971:DAIAPE>2.0.CO;2.

Kitagawa, G. (1996), Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *J. Comput. Graphical Stat.*, *5*, 1–25, doi:10.2307/1390750.

Kivman, G. A. (2003), Sequential parameter estimation for stochastic systems, *Nonlinear Processes Geophys.*, *10*, 253–259, doi:10.5194/npg-10-253-2003.

Koizumi, K., Y. Ishikawa, and T. Tsuyuki (2005), Assimilation of precipitation data to the J. MA mesoscale model with a four-dimensional variational method and its impact on precipitation forecasts, *SOLA*, *1*, 45–48, doi:10.2151/sola.2005-013.

Lefebvre, T., H. Bruyninckx, and J. De Schutter (2002), Comment on ''New method for the nonlinear transformation of means and covariances in filters and estimators deterministic non-periodic flow,'' *IEEE Trans. Autom. Control*, *47*(8), 1406–1409, doi:10.1109/TAC.2002.800742.

Liu, J. S., and E. Chen (1998), Sequential Monte Carlo methods for dynamics systems, *J. Am. Stat. Assoc.*, *93*, 1032–1044, doi:10.2307/2669847.

Lorenz, E. (1963), Deterministic non-periodic flow, *J. Atmos. Sci.*, *20*, 13–141.

Lorenz, E. (1996), Predictability: A problem partly solved, in *Proceedings of the Seminar on Predictability,* vol. *1*, pp. 1–18, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.

Lorenz, E. (2005), Designing chaotic models, *J. Atmos. Sci.*, *62*, 1574–1587, doi:10.1175/JAS3430.1.

Lorenz, E. (2006a), Regimes in simple systems, *J. Atmos. Sci.*, *63*, 2056–2073, doi:10.1175/JAS3727.1.

Lorenz, E. N. (2006b), Predictability: A problem partly solved, in *Predictability of Weather and Climate*, edited by T. Palmer and R. Hagedorn, pp. 40–58, Cambridge Univ. Press, Cambridge, U. K.

Lorenz, E., and K. A. Emmanuel (1998), Optimal sites for supplementary weather observations: Simulations with a small model, *J. Atmos. Sci.*, *55*, 399–414, doi:10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

Miller, R., M. Ghil, and F. Gauthiez (1994), Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.*, *51*, 1037–1056, doi:10.1175/1520-0469(1994)051<1037:ADAISN>2.0.CO;2.

Nørgåd Magnus, P., N. K. Poulsen, and O. Ravn (2000), New developments in state estimation of nonlinear systems, *Automatica*, *36*, 627–1638, doi:10.1016/S0005-1098(99)00158-2.

Palmer, T. (1993), Extended-range atmospheric prediction and the Lorenz model, *Bull. Am. Meteorol. Soc.*, *74*, 49–65, doi:10.1175/1520-0477(1993)074<0049:ERAPAT>2.0.CO;2.

Peña, M., Z. Toth, and M. Wei (2010), Controlling noise in ensemble data assimilation schemes, *Mon. Weather Rev.*, *138*, 1502–1512, doi:10.1175/2009MWR2854.1.

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 Univ. Press, Cambridge, U. K.

Sardeshmukh, P. D. (2010), Global climate diagnosis in a linear stochastically forced framework, in *Seminar on Diagnosis of Forecasting and Data Assimilation Systems, 7–10 September 2009*, pp. 15–24, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.

Sardeshmukh, P. D., and P. Sura (2009), Reconciling non-Gaussian climate statistics with linear dynamics, *J. Clim.*, *22*, 1193–1207, doi:10.1175/2008JCLI2358.1.

Schon, T. B. (2006), Estimation of nonlinear dynamic systems–Theory and applications, Ph.D. thesis, Linkoping Stud. in Sci. and Technol., Linkoping Univ., Linkoping, Sweden.

Shutts, G. (2005), A kinetic energy backscatter algorithm for use in ensemble prediction systems, *Q. J. R. Meteorol. Soc.*, *612*, 3079–3102, doi:10.1256/qj.04.106.

Shutts, G., T. Allen, and J. Berner (2008), Stochastic parametrization of multiscale processes using a dual-grid approach, *Philos. Trans. R. Soc. A*, *366*, 2623–2639, doi:10.1098/rsta.2008.0035.

Simon, D. (2006), *Optimal State Estimation, Kalman, $H_\infty$, and Nonlinear Approaches,* 1st ed., pp. 318–320, Wiley Intersci., Hoboken, N. J.

Sura, P., and P. D. Sardeshmukh (2008), A global view of non-Gaussian SST variability, *J. Phys. Oceanogr.*, *38*, 639–647, doi:10.1175/2007JPO3761.1.

Sura, P., M. Newman, C. Penland, and P. D. Sardeshmukh (2005), Multiplicative noise and non-Gaussianity: A paradigm for atmospheric regimes? *J. Atmos. Sci.*, *62*, 1391–1409, doi:10.1175/JAS3408.1.

Tippett, M. K., J. Anderson, C. Bishop, T. Hamill, and J. Whitaker (2003), Ensemble square-root filters, *Mon. Weather Rev.*, *131*, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

Tong, M., and M. Xue (2008a), Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part I: Sensitivity analysis and parameter identifiability, *Mon. Weather Rev.*, *136*, 1630–1648, doi:10.1175/2007MWR2070.1.

Tong, M., and M. Xue (2008b), Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments, *Mon. Weather Rev.*, *136*, 1630–1648, doi:10.1175/2007MWR2070.1.

Tsuyuki, T., K. Koizumi, and Y. Ishikawa (2003), The JMA mesoscale 4D-VAR system and assimilation of precipitation and moisture data, in *Proceedings of the ECMWF/GEWEX Workshop on Humidity Analysis*, pp. 59–68, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.

Van der Merwe, R. (2004), Sigma-Point Kalman filters for probabilistic inference in dynamic state-space models, Ph.D. thesis, Oreg. Health and Sci. Univ., Portland.

Van der Merwe, R., and E. A. Wan (2001a), Efficient derivative-free Kalman filters for online learning, in *ESANN'2001 Proceedings: European Symposium on Artificial Neural Networks*, pp. 205–210, D-Facto, Brussels.

Van der Merwe, R., and E. A. Wan (2001b), The square-root unscented Kalman filter for state and parameter estimation, in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Inst. of Electr. and Electron. Eng., Salt Lake City, Utah.

Van der Merwe, R., A. Doucet, N. de Freitas, and E. Wan (2000), The unscented particle filter, technical report, Eng. Dep., Cambridge Univ., Cambridge, U. K.

Van der Merwe, R., E. A. Wan, and S. I. Julier (2004), Sigma-point Kalman filters for nonlinear estimation and sensor fusion: Applications to integrated navigation, in *AIAA Guidance, Navigation and Control Conference*, pp. 5120–5122, Am. Inst. of Aeronaut. and Astronaut., Providence, R. I.

Van Leeuwen, P. (2009), Particle filtering in geophysical systems, *Mon. Weather Rev.*, *137*, 4089–4114, doi:10.1175/2009MWR2835.1.

Wan, E., and R. Van Der Merve (2000), The unscented Kalman filter for nonlinear estimation, in *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pp. 153–158, IEEE Commun. Soc., Lake Louise, Alberta, Canada.

Zhang, F., and C. Snyder (2007), Ensemble-based data assimilation, *Bull. Am. Meteorol. Soc.*, *88*, 565–568, doi:10.1175/BAMS-88-4-565.

Zupanski, M. (2005), Maximum likelihood ensemble filter: Theoretical aspects, *Mon. Weather Rev.*, *133*, 1710–1726, doi:10.1175/MWR2946.1.