UNBC UNIVERSITY OF NORTHERN BRITISH COLUMBIA

Advanced Data Assimilation

in strongly nonlinear systems

via Sigma-point Kalman Filters

Jaison Thomas Ambadan

and

Youmin Tang

Environmental Science & Engineering Program
Natural Resources & Environmental Science

Outline

Introduction

Sigma-point Kalman filters

Simulations with Lorenz model

A Reduced SPKF

Summary & Conclusion

- Introduction
- Sigma-point Kalman filters
- Simulations with Lorenz ’63 & ’95 models
- A Reduced SPKF
- Summary & Conclusion.

Introduction

Data assimilation: A dynamical state-space estimation problem

$$\boldsymbol{\theta}_k = \boldsymbol{f}\left(\boldsymbol{\theta}_{k-1}, \boldsymbol{q}_{k-1}\right)$$

$$\psi_k = \boldsymbol{h}\left(\boldsymbol{\theta}_k, \boldsymbol{r}_k\right)$$

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

Sigma-point Kalman filter

- **The SPKF algorithm has been successfully implemented in many areas like robotics, artificial intelligence, natural language processing, and global positioning systems navigation.**

- **The SPKF makes use of a reformulated Kalman gain K and "chooses" the ensemble deterministically in such a way that it can estimate the statistical moments of the nonlinear model accurately; in other words, the forecast error covariance equation is computed using deterministically chosen samples, called "sigma-points".**

Ref: (Nørgad Magnus et al. 2000; Ito and Xiong 2000; Lefebvre et al. 2002;Wan and Van Der Merve 2000; Haykin 2001; Van der Merwe 2004, Julier et al. 1995; Van der Merwe and Wan4 April 2001,M;).

Kalman gain & Covariance update equations

Standard equations:

$$K_k = P_{\theta_k}^{-} H^{\mathrm{T}} \left[H P_{\theta_k}^{-} H^{\mathrm{T}} + R \right]^{-1}$$

$$P_{\theta_k} = \left(I - K_k H \right) P_{\theta_k}^{-}$$

Alternative forms:

$$K_k = P_{\theta_k \tilde{\psi}_k} P_{\tilde{\psi}_k}^{-1}$$

$$P_{\theta_k} = P_{\theta_k}^{-} - K_k P_{\tilde{\psi}_k} K_k^{\mathrm{T}}$$

$P_{\theta_k \tilde{\psi}_k}$ ➔ cross-covariance between the state and observation errors

$P_{\tilde{\psi}_k}$ ➔ error covariance of the difference between model and observations

No linear assumption for measurement function.

Ref: (Simon, D. 2006, Gelb, A. 1974)

SP-Unscented Kalman filter

Unscented transformation (Julier et al. 1995; Julier 1998; Wan and Van Der Merve 2000; Julier 2002).

$$\mathcal{X}_{k,0} = \theta_k \qquad w_0^{(m)} = \frac{\lambda}{(L+\lambda)} \tag{2.42}$$

$$\mathcal{X}_{k,i}^{+} = \theta_k + \left(\sqrt{(L+\lambda)\,\boldsymbol{P}_{\boldsymbol{\theta}_k}}\right)_i \quad i=1,\ldots,L \qquad w_0^{(c)} = \frac{\lambda}{(L+\lambda)} + \left(1-\alpha^2+\beta\right) \tag{2.43}$$

$$\mathcal{X}_{k,i}^{-} = \theta_k - \left(\sqrt{(L+\lambda)\,\boldsymbol{P}_{\boldsymbol{\theta}_k}}\right)_i \quad i=(L+1),\ldots,2L \qquad w_i^{(m)} = w_i^{(c)} = \frac{1}{2\,(L+\lambda)} \quad i=1,\ldots,2L \tag{2.44}$$

$\lambda = \alpha^2\,(L+\kappa) - L$ is a scaling parameter.

The sigma-point vector is then propagated through the nonlinear model

SP-Central Difference KF

- **In SP-CDKF the analytical derivatives in EKF are replaced by numerically evaluated central divided differences, based on Sterling's polynomial interpolation.**

$$\mathcal{X}_{k,0} = \bar{\boldsymbol{\theta}}_k \qquad w_0^{(m)} = \frac{\delta^2 - L}{\delta^2}$$

$$\mathcal{X}_{k,i}^{+} = \bar{\boldsymbol{\theta}}_k + \left(\sqrt{\delta^2 \boldsymbol{P}_{\theta_k}}\right)_i \quad i = 1, \ldots, L \qquad w_i^{(m)} = \frac{1}{2\delta^2} \quad i = 1, \ldots, 2L$$

$$\mathcal{X}_{k,i}^{-} = \bar{\boldsymbol{\theta}}_k - \left(\sqrt{\delta^2 \boldsymbol{P}_{\theta_k}}\right)_i \quad i = (L+1), \ldots, 2L \qquad w_i^{(c_1)} = \frac{1}{4\delta^2} \quad i = 1, \ldots, 2L$$

$$w_i^{(c_2)} = \frac{\delta^2 - 1}{4\delta^4} \quad i = 1, \ldots, 2L$$

where δ is the central difference step size

Ref: (**Ito and Xiong 2000; Nørga°d Magnus et al. 2000; Lefebvre et al. 2002).**

SPKF and EnKF

- We can extend the sigma-point approach to ensemble Kalman filter and formulate the ensemble Kalman filter as a **"general sigma-point Kalman filter"** without a specific selection scheme.

we can consider each perturbed initial state as a EnKF *sigma-point.*

$$\mathcal{X}_{k,m} = \boldsymbol{\theta}_{k,m} + \epsilon_{k,m} \qquad w = \frac{1}{(M-1)}$$

Now the ensemble forecast error covariance given by

$$\boldsymbol{P}^f = \boldsymbol{P}^-_{\theta_k} \approx \frac{1}{M-1}\sum_{m=1}^{M}\left(\mathcal{X}^f_{k,m} - \overline{\mathcal{X}^f_k}\right)\left(\mathcal{X}^f_{k,m} - \overline{\mathcal{X}^f_k}\right)^{\mathrm{T}}$$

$$= w\sum_{m=1}^{M}\left(\mathcal{X}^f_{k,m} - \hat{\boldsymbol{\theta}}^-_k\right)\left(\mathcal{X}^f_{k,m} - \hat{\boldsymbol{\theta}}^-_k\right)^{\mathrm{T}}$$

where

$$\hat{\boldsymbol{\theta}}^-_k \approx \overline{\mathcal{X}^f_k} = w\sum_{m=1}^{M}\mathcal{X}^f_{k,m}$$

- In other words the forecast sigma-points in SP-UKF and SP-CDKF and other Sigma-point Kalman filter algorithms are actually specific ensembles conditioned on the specific selection schemes

Lorenz '63 model

$$\frac{dx}{dt} = \sigma\left(y - x\right) + q^x$$

$$\frac{dy}{dt} = \rho x - y - xz + q^y$$

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

True value: integrate the model over 4000 time steps using prescribed parameters and initial conditions.

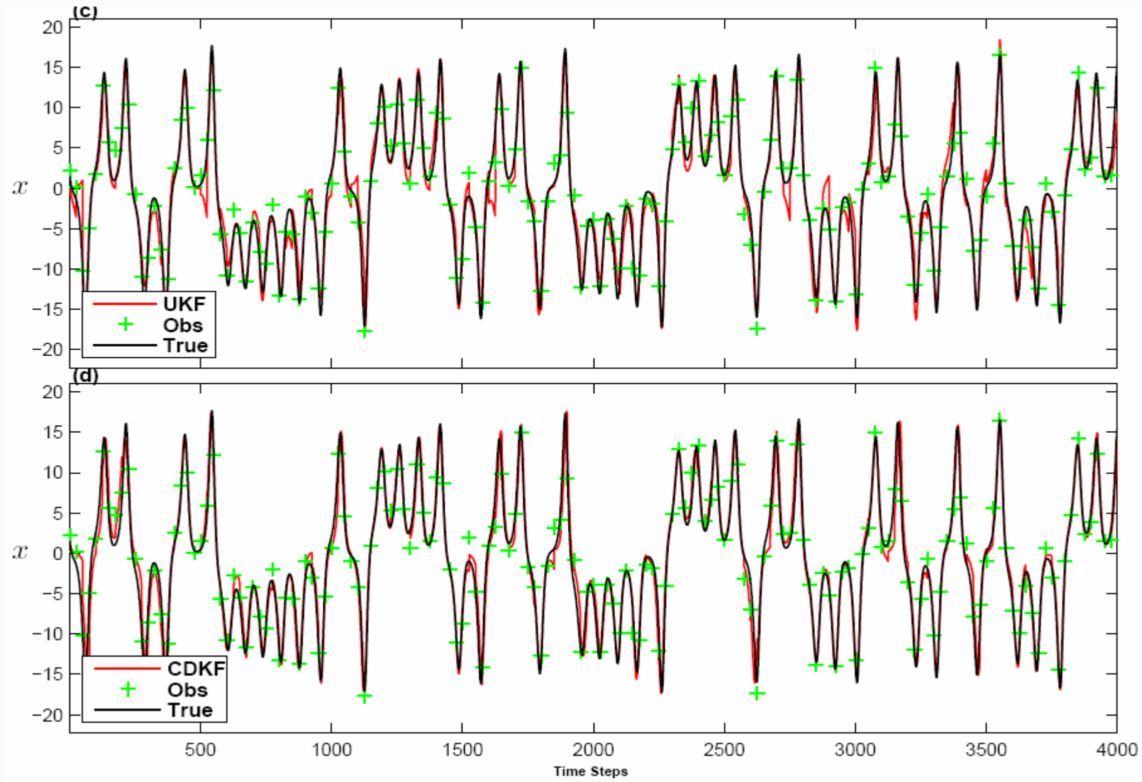
Observation: true value plus normal distribute noise;

Experimental conditions are the same as those used by Miller (Miller, 1994) and Evensen (Evensen 1997)

Lorenz model: state estimation

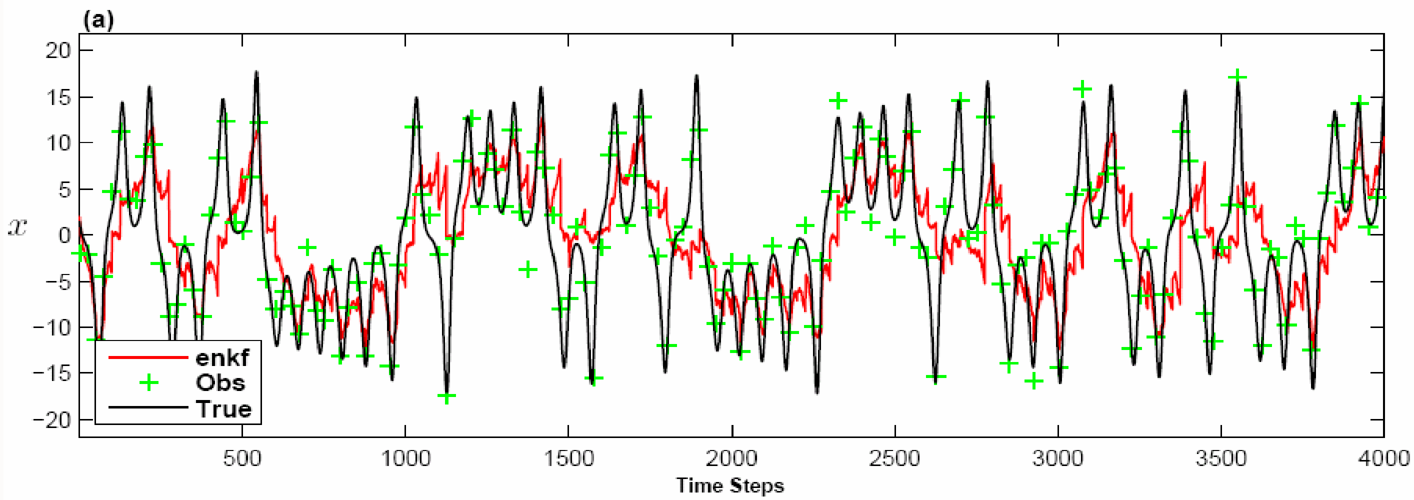
Observation and initial conditions: True values plus normal distributed noise. The assimilation interval is 25 time steps.

$N(0,\sqrt{2})$



An “unfair” comparison!

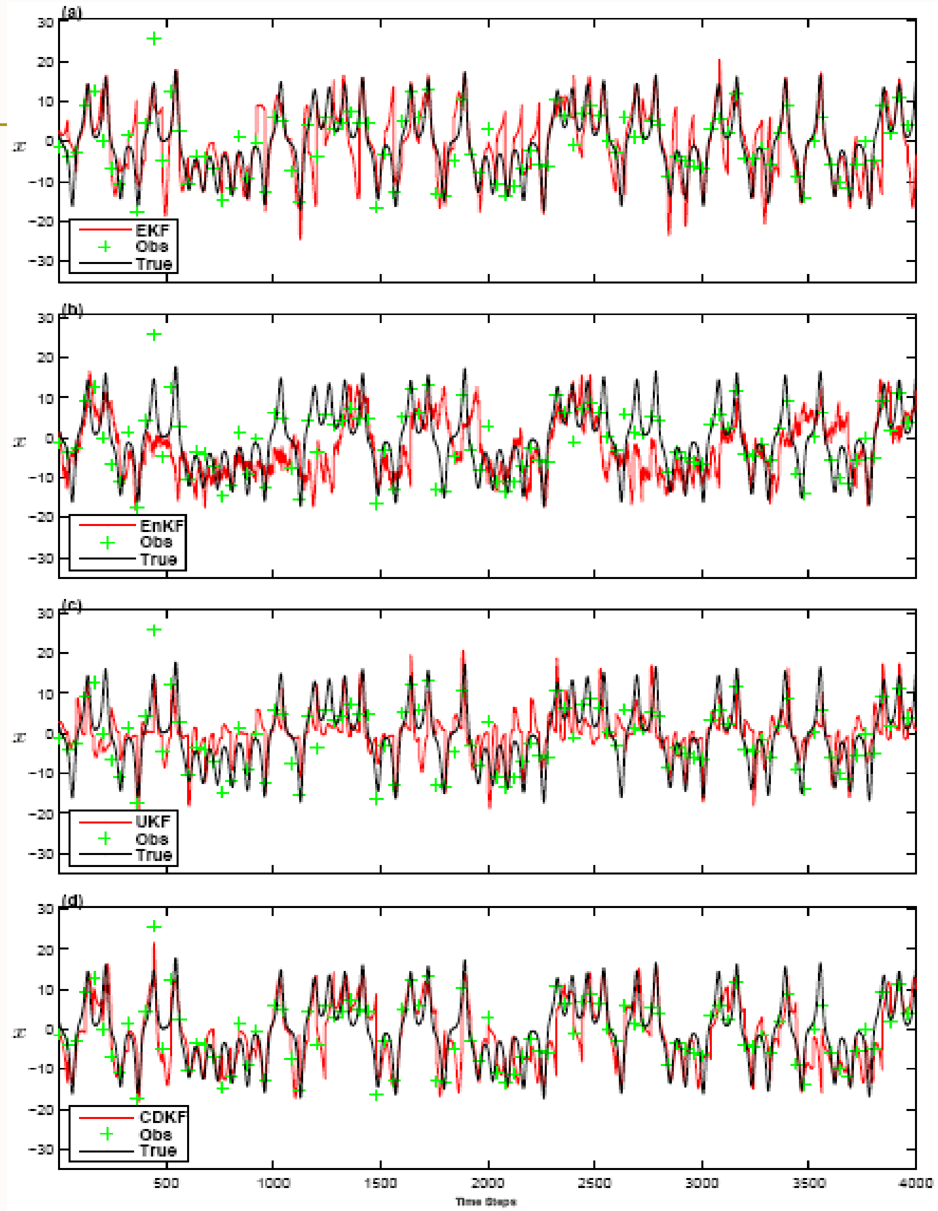
EnKF with 19 members



Assimilation Method	Computation Time (in Seconds)	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

noise $N(0, \sqrt{20})$.

Assimilation interval is 40 time steps.



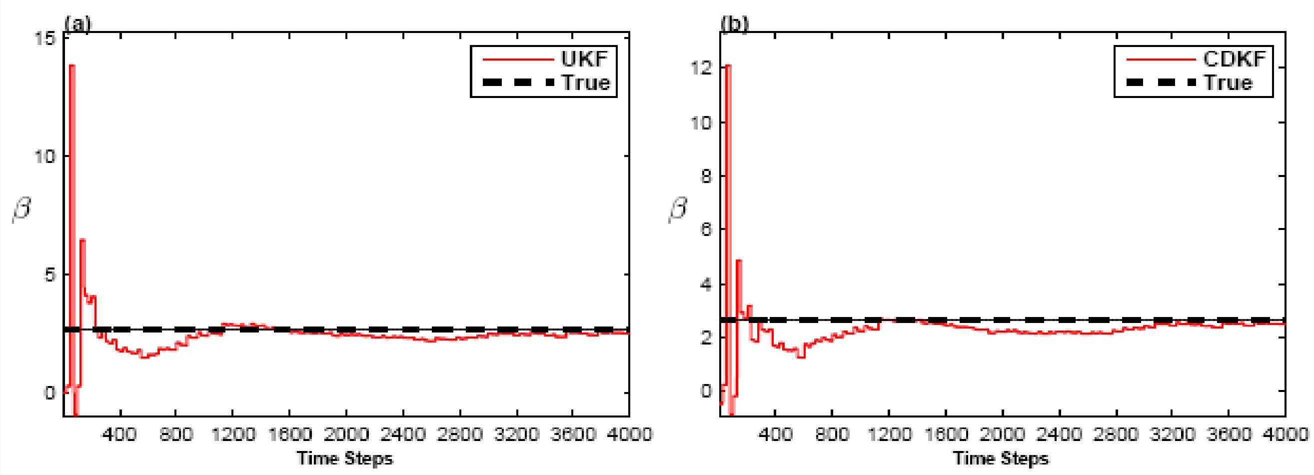
Parameter estimation

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

$$\psi_k = \boldsymbol{f}\left(\boldsymbol{\theta}_k, \boldsymbol{\Lambda}_k\right) + \boldsymbol{r}_k^{\Lambda}$$

where $\boldsymbol{f}(\cdot)$ is the nonlinear measurement model given by the Lorenz equations

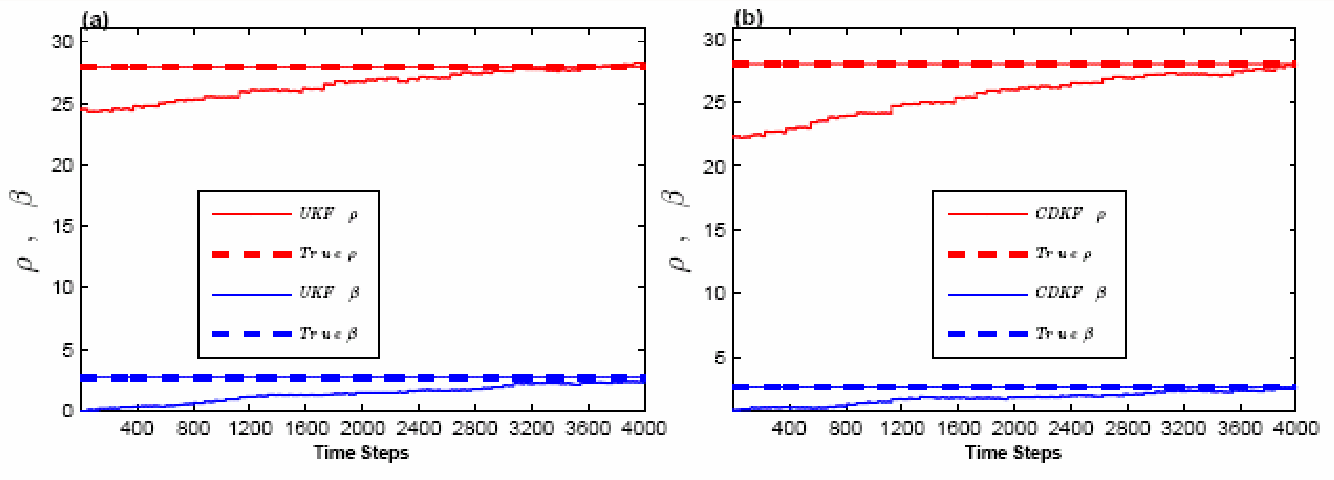
$\boldsymbol{\Lambda}$ is the parameter vector which constitutes the dynamical parameters

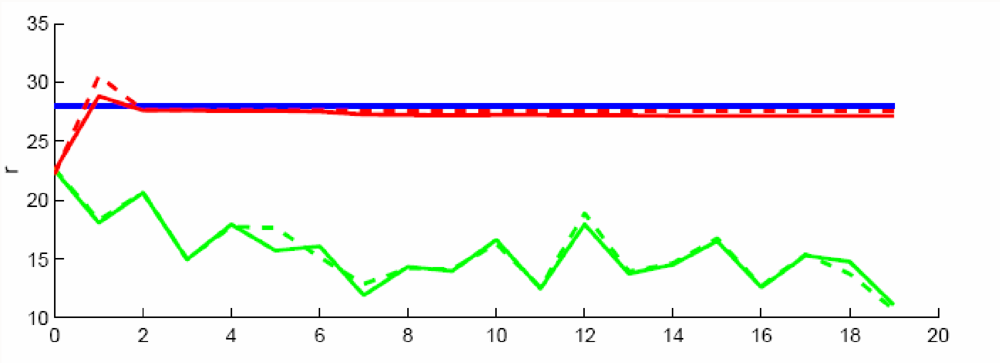


Parameter estimation

Simultaneous estimation of two parameters:

Initial parameters = true parameter plus normal distributed noise of covariance 100.





Kivman, G.A. 2003: Sequential parameter estimation for stochastic system, Nonlinear. Process. in Geophysics, 10, 253-259.

A “big” drawback of SPKF

- For an L-dimensional system, the number of sigma-points required to estimate the true statistics is 2L+1
- 2L+1 sigma-point integration is impossible, when the dimension system is of the order of millions as happens often in GCMs

A Possible solution – reducing sigma-points

$$\mathcal{X}_{k,0} = \theta_k \qquad w_0^{(m)} = \frac{\lambda}{(L+\lambda)} \qquad (2.42)$$

$$\mathcal{X}_{k,i}^{+} = \theta_k + \left(\sqrt{(L+\lambda)\,\boldsymbol{P}_{\boldsymbol{\theta}_k}}\right)_i \quad i=1,\ldots,L \qquad w_0^{(c)} = \frac{\lambda}{(L+\lambda)} + \left(1-\alpha^2+\beta\right) \qquad (2.43)$$

$$\mathcal{X}_{k,i}^{-} = \theta_k - \left(\sqrt{(L+\lambda)\,\boldsymbol{P}_{\boldsymbol{\theta}_k}}\right)_i \quad i=(L+1),\ldots,2L \qquad w_i^{(m)} = w_i^{(c)} = \frac{1}{2(L+\lambda)} \quad i=1,\ldots,2L \qquad (2.44)$$

A subspace approach with sigma-points:

Lermusiaux et. al. [Lermusiaux, 1997; Lermusiaux and Robinson, 1999] proposed a method to reduce error space, called the Error Subspace Statistical Estimation (ESSE).

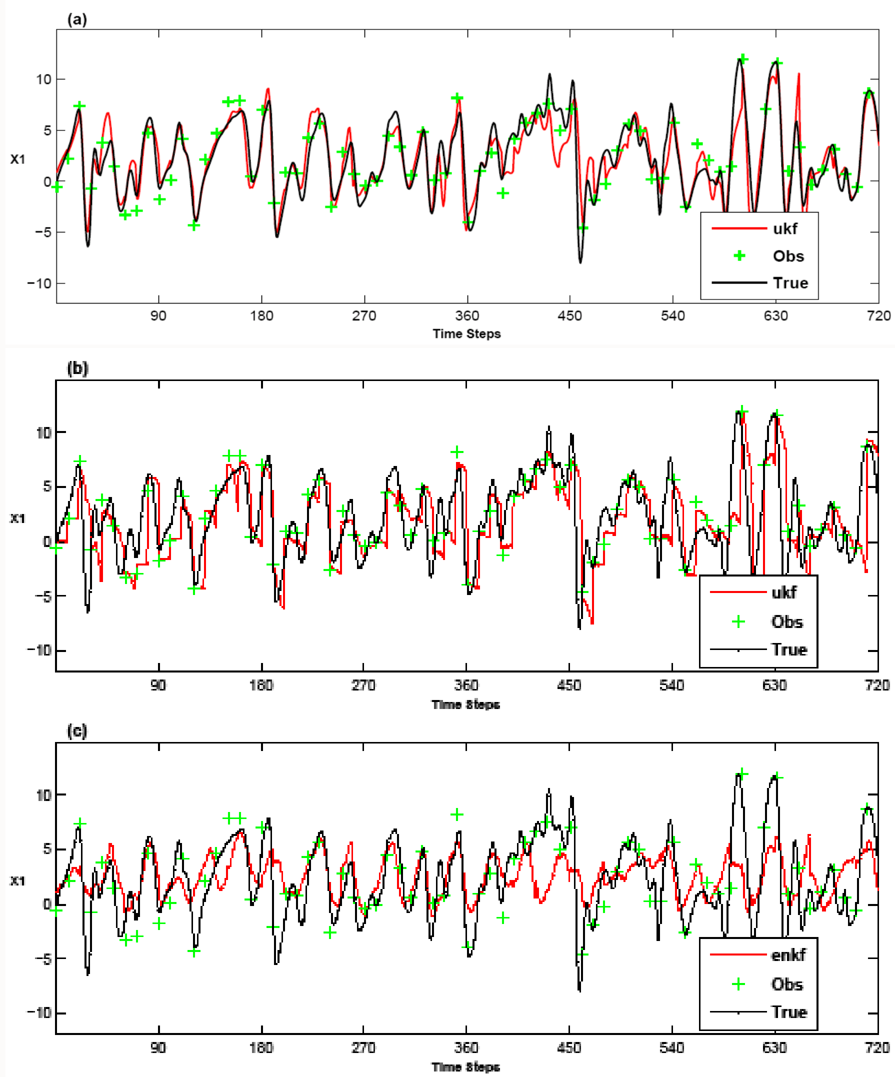
Theoretically the most important sigma-points should be chosen based on reduced rank approximation. We have used principal component analysis (PCA) to identify the most important sigma-points.

A Reduced SPKF - (Lorenz ’96 model: 36 variables)

with Full Sigma-points (241)

with reduced Sigma-points (20)

with EnKF of 20 members



Introduction

Sigma-point Kalman filters

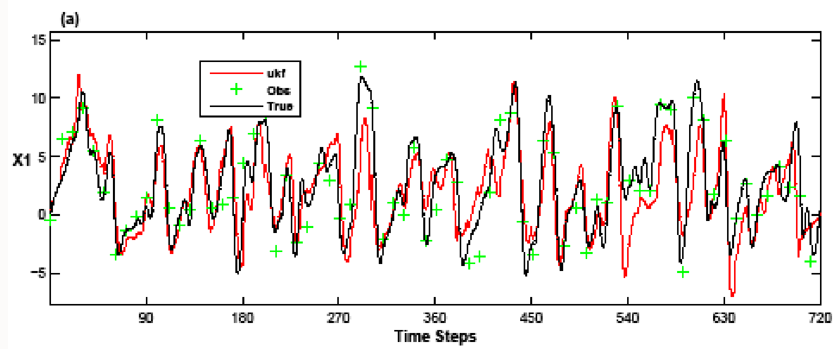
Simulations with Lorenz model

A Reduced SPKF

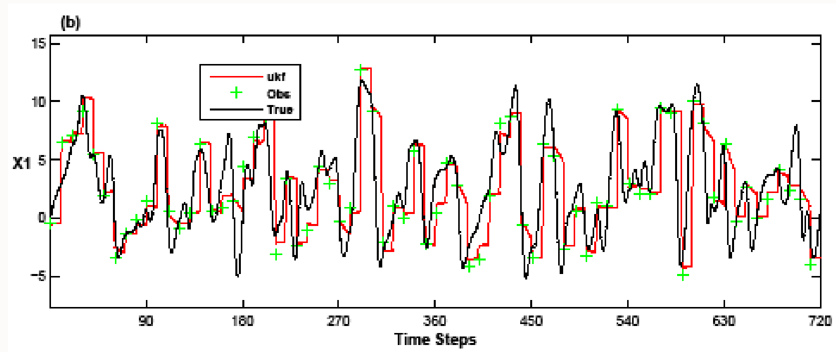
Summary & Conclusion

A Reduced SPKF - (Lorenz ’96 model: 960 variables)

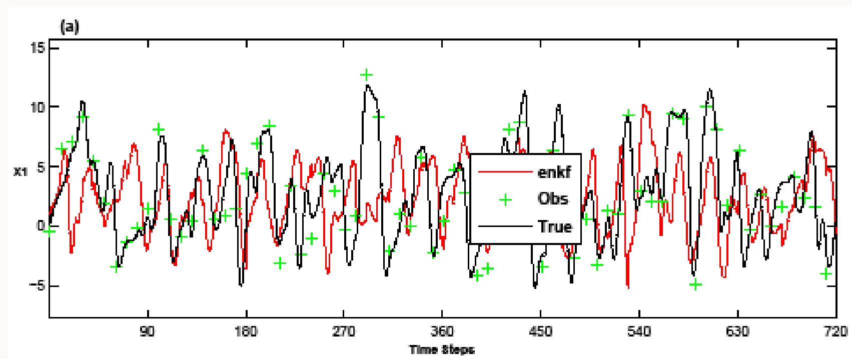
with full Sigma-points (5761)



with reduced Sigma-points (200)



with EnKF of 200 members



Conclusion

- The SPKF is a technique for a derivative-less optimal estimation using a deterministic sampling approach that ensures accurate estimation of error statistics.
- The SPKF is capable of estimating model state and parameters with better accuracy than EKF and EnKF for strong nonlinear systems.
- The SPKF is practically difficult for high dimensional systems. A possible solution is to reduce the number of sigma-points by "selecting a particular set of sigma-points".

Thank You.

Acknowledgements

This work was supported by NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant and CFCAS (Canadian foundation for Climate and Atmospheric Sciences) network project of "Global Ocean-Atmosphere Prediction and Predictability".

Canadian Foundation for Climate and Atmospheric Sciences (CFCAS)

Fondation canadienne pour les sciences du climat et de l'atmosphère (FCSCA)

NSERC CRSNG

GOAPP Global Ocean-Atmosphere Prediction and Predictability