# A filter algorithm combining ensemble transform Kalman filter and imortance sampling

S. Nakano

The Institute of Statistical Mathematics
Graduate University for Advance Studies

23 October 2013

# About me

- Brief my history:
  - BS, MS, DS, in Earth and Planetary Science, Kyoto University, 1999, 2001, 2004
    (Geomagnetism, magnetospheric physics)
  - Researcher, in Kyoto University, 2004–2005
  - Researcher, Assistant Professor, in the Institute of Statistical Mathemattics 2005–
- Current research interests:
  - Sequential (ensemble-based) data assimilation algorithms
  - Application of data assimilation techniques to space physics
  - Application of nonlinear filtering methods (e.g., particle filter)
  - Sampling methods in Monte Carlo models

# The Institute of Statistical Mathematics

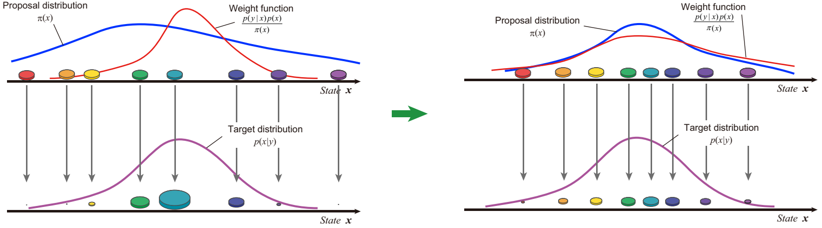
- Established in 1944.
- Moved to Tachikawa (from the metropolitan area of Tokyo) in 2009
- 3 departments:
  - Department of Statistical Modeling
  - Department of Data Science
  - Department of Mathematical Analysis and Statistical Inference
- 5 project-oriented units:
  - Risk Analysis Research Center
  - Research and Development Center for Data Assimilation
  - Survey Science Center
  - Research Center for Statistical Machine Learning
  - Service Science Research Center

# Motivation of this work

- Observations in geosciences are sometimes nonlinear or non-Gaussian.
- Many filtering algorithms such as the ensemble Kalman filter and the ensemble transform Kalman filter are based on a linear Gaussian observation model.
- Even if such algorithms are applied to a case with nonlinear or non-Gaussian observation, they would provide some errors or biased estimates.
- In order to allow nonlinear or non-Gaussian observation, the importance sampling approach is employed.

# Importance sampling

- The particle filter algorithm is based on the importance sampling method, which represents the posterior PDF by weighted sample.
- In the importance sampling, the proposal distribution should be chosen as similar to the target distribution.



# Particle filter

Particle filters are based on the importance sampling approach.
There are two ways to derive the most basic particle filter (so-called the bootstrap filter).

- From the joint distribution of the whole sequence $x_{0:k}$

$$p(x_{0:k}|y_{1:k}) \approx \sum_{i=1}^{N} \frac{p(x_{0:k}^{q,(i)}|y_{1:k})}{q(x_{0:k}^{q,(i)}|y_{1:k})} \delta(x_{0:k} - x_{0:k}^{q,(i)}),$$

- From the marginal

$$p(x_k|y_{1:k}) \approx \sum_{i=1}^{N} \frac{p(x_k^{\pi,(i)}|y_{1:k})}{\pi(x_k^{\pi,(i)}|y_{1:k})} \delta(x_k - x_k^{\pi,(i)}).$$

# Particle filter

If we start with the joint distribution:

$$p(x_{0:k}|y_{1:k}) \approx \sum_{i=1}^{N} \frac{p(x_{0:k}^{q,(i)}|y_{1:k})}{q(x_{0:k}^{q,(i)}|y_{1:k})} \delta(x_{0:k} - x_{0:k}^{q,(i)}),$$

Assuming $p(x_k|x_{0:k-1}, y_{1:k-1}) = p(x_k|x_{k-1})$ and $p(y_k|x_{0:k}, y_{1:k-1}) = p(y_k|x_k)$,

$$p(x_{0:k}|y_{1:k}) \approx \frac{1}{Z'} \sum_{i=1}^{N} \frac{p(y_k|x_k^{q,(i)})\, p(x_k^{q,(i)}|x_{k-1}^{q,(i)})\, p(x_{0:k-1}^{q,(i)}|y_{1:k-1})}{q(x_k^{q,(i)}|x_{k-1}^{q,(i)}, y_{1:k})\, q(x_{0:k-1}^{q,(i)}|y_{1:k-1})} \delta(x_{0:k} - x_{0:k}^{q,(i)}),$$

If $p(x_{0:k-1}|y_{1:k-1})$ is represented by equally weighted particles $\{x_{0:k-1}^{(i)}\}_{i=1}^{N}$,

$$p(x_{0:k}|y_{1:k}) \approx \frac{1}{Z'} \sum_{i=1}^{N} \frac{p(y_k|x_k^{q,(i)})\, p(x_k^{q,(i)}|x_{k-1}^{(i)})}{q(x_k^{q,(i)}|x_{k-1}^{(i)}, y_{1:k})} \delta(x_{0:k} - x_{0:k}^{q,(i)}).$$

If $p(x_k|x_{k-1}^{(i)})$ is chosen as $q(x_k|x_{0:k-1}^{(i)}, y_{1:k})$,

$$p(x_{0:k}|y_{1:k}) \approx \frac{1}{Z'} \sum_{i=1}^{N} p(y_k|x_k^{q,(i)})\delta(x_{0:k} - x_{0:k}^{q,(i)}).$$

# Particle filter

If we consider the marginal:

$$p(x_k|y_{1:k}) \approx \sum_{i=1}^{N} \frac{p(x_k^{\pi,(i)}|y_{1:k})}{\pi(x_k^{\pi,(i)}|y_{1:k})} \delta(x_k - x_k^{\pi,(i)}).$$

Assuming that $p(y_k|x_{0:k}, y_{1:k-1}) = p(y_k|x_k)$,

$$p(x_k|y_{1:k}) \approx \frac{1}{Z} \sum_{i=1}^{N} \frac{p(y_k|x_k^{\pi,(i)})\, p(x_k^{\pi,(i)}|y_{1:k-1})}{\pi(x_k^{\pi,(i)}|y_{1:k})} \delta(x_k - x_k^{\pi,(i)}).$$

If $p(x_k|y_{1:k-1})$ is chosen as $\pi(x_k|y_{1:k})$,

$$p(x_k|y_{1:k}) \approx \frac{1}{Z} \sum_{i=1}^{N} p(y_k|x_k^{\pi,(i)}) \delta(x_k - x_k^{\pi,(i)}).$$

Sampling from

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})\, p(x_{k-1}|y_{1:k-1})\, dx_{k-1}$$

is done via a Markov chain sampling.

# Particle filter

- According to the joint distribution formula

$$p(x_{0:k}|y_{1:k}) \approx \frac{1}{Z'} \sum_{i=1}^{N} \frac{p(y_k|x_k^{q,(i)})\, p(x_k^{q,(i)}|x_{k-1}^{q,(i)})\, p(x_{0:k-1}^{q,(i)}|y_{1:k-1})}{q(x_k^{q,(i)}|x_{k-1}^{q,(i)}, y_{1:k})\, q(x_{0:k-1}^{q,(i)}|y_{1:k-1})} \delta(x_{0:k} - x_{0:k}^{q,(i)}),$$

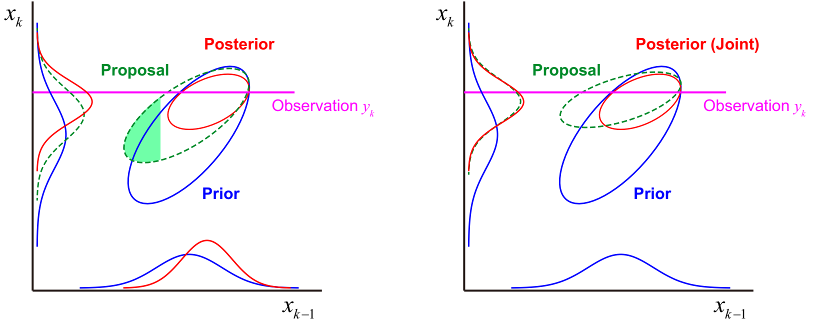
  the transition proposal $q(x_k|x_{k-1}, y_{1:k})$ should be designed such that $q(x_k|x_{k-1}, y_{1:k})\, q(x_{0:k-1}|y_{1:k-1})$ is similar to $p(x_{0:k}|y_{1:k})$.
- According to the marginal distribution formula

$$p(x_k|y_{1:k}) \approx \frac{1}{Z} \sum_{i=1}^{N} \frac{p(y_k|x_k^{\pi,(i)})\, p(x_k^{\pi,(i)}|y_{1:k-1})}{\pi(x_k^{\pi,(i)}|y_{1:k})} \delta(x_k - x_k^{\pi,(i)}),$$

  $\pi(x_k^{\pi,(i)}|y_{1:k})$ should be designed as similar to $p(x_k|y_k)$.

# Particle filter

- Joint distribution formula:
  - The transition proposal $q(x_k|x_{k-1}, y_{1:k})$ is designed such that $q(x_k|x_{k-1}, y_{1:k})\, q(x_{0:k-1}|y_{1:k-1})$ becomes similar to $p(x_{0:k}|y_{1:k})$.
- Marginal distribution formula:
  - The proposal $\pi(x_k^{\pi,(i)}|y_{1:k})$ is designed as similar to $p(x_k|y_k)$.



# Particle filter

- Joint distribution formula:
  - It is inevitable that a part of particles become ineffective. (It is normally impossible to satisfy $q(x_k|x_{k-1}, y_{1:k})\, p(x_{0:k-1}|y_{1:k-1}) = p(x_{0:k}|y_{1:k})$. )
  - But, the assumption for the transition density $p(x_k|x_{k-1})$ is usually more acceptable than the assumption for the forecast density $p(x_k|y_{1:k-1})$.
- Marginal distribution formula
  - Sampling from $p(x_k|y_{1:k-1})$ does not matter.
  - But, it is required to evaluate $p(x_k|y_{1:k-1})$ for all the particles. It is usually impossible to evaluate the forecast density $p(x_k|y_{1:k-1})$.
  - In this work, $p(x_k|y_{1:k-1})$ is assumed to be Gaussian.

# Forecast

- Suppose that the posterior PDF at the time step $t_{k-1}$, $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$, is given, the forecast PDF $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$ can be obtained by the following integral:

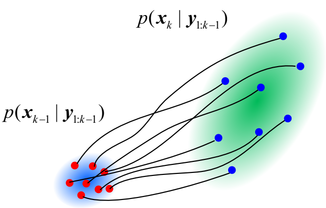
$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})\, p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}) d\boldsymbol{x}_{k-1}.$$

- We assume that the system dynamics is deterministic and fully known; that is, the system model is written in the following form:

$$\boldsymbol{x}_k = \mathcal{M}(\boldsymbol{x}_{k-1}),$$

which corresponds to the assumption that $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ is a delta function.

# Prediction with the Monte Carlo method



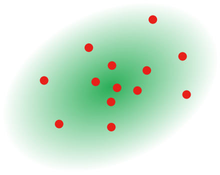

- Suppose that the particles $\{\boldsymbol{x}^{(i)}_{k-1|k-1}\}$ are Monte Carlo samples drawn from the posterior at the previous time step $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$.
- If we generate each member of the forecast ensemble by applying the dynamical system model for each particle:

$$\boldsymbol{x}^{(i)}_{k|k-1} = \mathcal{M}(\boldsymbol{x}^{(i)}_{k-1|k-1}),$$

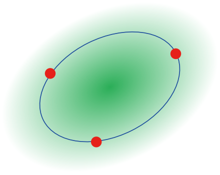
  it provides a random sample from the forecast distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$.
- Non-Gaussian features of the forecast distribution can be represented with a sufficiently large number of particles.
- The exact mean and covariance can be obtained for $N \to \infty$.

# Forecast

- This Monte Carlo approache can deal with non-Gaussian PDF. But, it assumes that the number of particles is sufficiently large.
- To obtain the forecast, a model run is required for each of the particles.
- In practical applications, the ensemble size is typically limited to smaller than the state dimension.
- If the ensemble size $N$ is smaller than the rank of the state covariance matrix, the ensemble would form a simplex in an $(N-1)$-dimensional subspace.
- It would be difficult to represent the non-Gaussianity by the simplex.

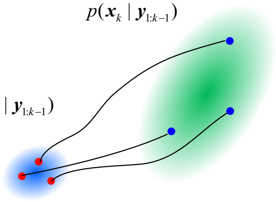


Monte Carlo representation



Simplex representation

# Simplex representation




- If an ensemble is obtained by applying the system model to each particle:

$$x^{(i)}_{k|k-1} = \mathcal{M}(x^{(i)}_{k-1|k-1}).$$

  the moments are considered up to the second order (Wang et al, 2004).

- The first and second order moments are approximated with second order accuracy. (The Taylor expansion of the dynamical model up to the second-order is considered except for the uncertainties in the subspace complimentary to the ensemble subspace. )
- We then use only the first and second order moments of the forecast ensemble and assume that the forecast PDF is Gaussian.

# Exact mean vector

The mean vector of the forecast distribution can be calculated via the following integral:

$$\boldsymbol{x}_{k|k-1} = \int \mathcal{M}(\boldsymbol{x}_{k-1}) \cdot p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})\, d\boldsymbol{x}_{k-1}.$$

Since we can represent the PDF $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$ by the generative model

$$\boldsymbol{x}_k|\boldsymbol{y}_{1:k} \sim \boldsymbol{x}_{k|k} + \mathsf{X}_{k|k}\boldsymbol{z}_k, \quad (\boldsymbol{z}_k \sim \mathcal{N}(\boldsymbol{z}_k; \boldsymbol{0}, \mathrm{I})),$$

the above equation can be rewritten as

$$\boldsymbol{x}_{k|k-1} = \int \mathcal{M}\left(\boldsymbol{x}_{k-1|k-1} + \mathsf{X}_{k-1|k-1}\boldsymbol{z}_k\right) \cdot p\left(\boldsymbol{z}_{k-1}\right)\, d\boldsymbol{z}_{k-1}.$$

# Exact mean vector

We can approximate $\mathcal{M}(\boldsymbol{x}_{k-1})$ by the following Taylor expansion:

$$\begin{aligned}\mathcal{M}(\boldsymbol{x}_{k-1}) &= \mathcal{M}(\boldsymbol{x}_{k-1|k-1} + \mathsf{X}_{k-1|k-1}\boldsymbol{z}_{k-1}) \\ &= \mathcal{M}(\boldsymbol{x}_{k-1|k-1}) + \nabla\mathcal{M}(\boldsymbol{x}_{k-1|k-1}) \cdot \mathsf{X}_{k-1|k-1}\boldsymbol{z}_{k-1} \\ &\quad + \frac{[\mathsf{X}_{k-1|k-1}\boldsymbol{z}_{k-1}]^T \nabla^2\mathcal{M}(\boldsymbol{x}_{k-1|k-1}) [\mathsf{X}_{k-1|k-1}\boldsymbol{z}_{k-1}]}{2} + \cdots .\end{aligned}$$

Using this Taylor expansion, we obtain

$$\begin{aligned}&\boldsymbol{x}_{k|k-1} \\ &= \int \mathcal{M}\left(\boldsymbol{x}_{k-1|k-1} + \mathsf{X}_{k-1|k-1}\boldsymbol{z}_{k}\right) \cdot p\left(\boldsymbol{z}_{k-1}\right)\, d\boldsymbol{z}_{k-1} \\ &= \mathcal{M}(\boldsymbol{x}_{k-1|k-1}) + \frac{1}{2}\mathrm{tr}\left(\mathsf{X}_{k-1|k-1}^T \left[\nabla^2\mathcal{M}(\boldsymbol{x}_{k-1|k-1})\right] \mathsf{X}_{k-1|k-1}\right) + O(\delta x^4).\end{aligned}$$

# Mean with a simplex

The ensemble mean can also be reduced as follows:

$$
\begin{aligned}
&\bar{\boldsymbol{x}}_{k|k-1} \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathcal{M}(\boldsymbol{x}^{(i)}_{k-1|k-1}) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{M}(\boldsymbol{x}_{k-1|k-1} + \mathsf{X}_{k-1|k-1}\boldsymbol{z}^{(i)}_{k-1}) \\
&= \mathcal{M}(\boldsymbol{x}_{k-1|k-1}) + \frac{1}{2}\mathrm{tr}\left(\mathsf{X}^{T}_{k-1|k-1}\left[\nabla^2\mathcal{M}(\boldsymbol{x}_{k-1|k-1})\right]\mathsf{X}_{k-1|k-1}\right) \\
&\quad + \frac{1}{6}\sum_{i,j,k}\sum_{l}\frac{\partial^3\mathcal{M}}{\partial x_i \partial x_j \partial x_k}\mathsf{X}^{il}_{k-1|k-1}\mathsf{X}^{jl}_{k-1|k-1}\mathsf{X}^{kl}_{k-1|k-1} + O(\delta x^4).
\end{aligned}
$$

Therefore, the ensemble mean $\bar{\boldsymbol{x}}_{k|k-1}$ approximates the forecast mean with accuracy up to the second order term.

However, the third order term does not agree with that of the analytical mean.

# Covariance matrix

The covariance matrix of the forecast distribution can be calculated via the following integral:

$$
\begin{aligned}
&\mathsf{V}_{k|k-1} \\
&= \int [\mathcal{M}(\boldsymbol{x}_{k-1}) - \boldsymbol{x}_{k-1|k-1}][\mathcal{M}(\boldsymbol{x}_{k-1}) - \boldsymbol{x}_{k-1|k-1}]^T \cdot p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})\, d\boldsymbol{x}_{k-1}.
\end{aligned}
$$

It can be reduced as follows:

$$
\begin{aligned}
&\mathsf{V}_{k|k-1} \\
&\approx \int [\nabla\mathcal{M}(\boldsymbol{x}_{k-1|k-1})]\, \mathsf{X}_{k-1|k-1} \boldsymbol{z}_{k-1} \boldsymbol{z}_{k-1}^T \mathsf{X}_{k-1|k-1}^T\, [\nabla\mathcal{M}(\boldsymbol{x}_{k-1|k-1})]^T \\
&\qquad \times p\,(\boldsymbol{z}_{k-1})\; d\boldsymbol{z}_{k-1} \\
&= [\nabla\mathcal{M}(\boldsymbol{x}_{k-1|k-1})]\, \mathsf{X}_{k-1|k-1} \mathsf{X}_{k-1|k-1}^T\, [\nabla\mathcal{M}(\boldsymbol{x}_{k-1|k-1})]^T .
\end{aligned}
$$

# Covariance matrix with a simplex

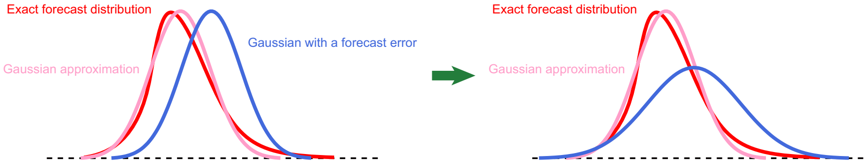
The covariance matrix of the ensemble can be reduced as follows:

$$\bar{\mathsf{V}}_{k|k-1} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathcal{M}(\boldsymbol{x}_{k-1|k-1}^{(i)}) - \bar{\boldsymbol{x}}_{k|k-1}\right]\left[\mathcal{M}(\boldsymbol{x}_{k-1|k-1}^{(i)}) - \bar{\boldsymbol{x}}_{k|k-1}\right]^T$$
$$\approx \left[\nabla\mathcal{M}(\boldsymbol{x}_{k-1|k-1})\right]\mathsf{X}_{k-1|k-1}\mathsf{X}_{k-1|k-1}^T\left[\nabla\mathcal{M}(\boldsymbol{x}_{k-1|k-1})\right]^T.$$

Therefore, the ensemble covariance $\bar{\mathsf{V}}_{k|k-1}$ approximates the covariance matrix of the forecast distribution with accuracy up to the second order term. (The third order term does not agree with the analytical value again.)

# Covariance inflation

- If the number of particles is limited, the ensemble mean can deviate from the exact forecast mean.
- This means that the forecast PDF produced by a limited-sized ensemble may not cover the probable region of the state space.
- The inflation of the covariance matrix could be a good way to avoid missing the probable value.



# Covariance inflation

The following rough argument would suggest that the discrepancy from the exact forecast can be reduced by inflating the covariance.

- We want to attain the Gaussian forecast distribution, $\mathcal{N}(\boldsymbol{x}_{k|k-1}, \mathsf{V}_{k|k-1})$, where $\boldsymbol{x}_{k|k-1}$ and $\mathsf{V}_{k|k-1}$ are the exact mean vector and the exact covariance matrix, respectively.
- We estimate the forecast distribution as a Gaussian $\mathcal{N}(\bar{\boldsymbol{x}}_{k|k-1}, \alpha^2 \bar{\mathsf{V}}_{k|k-1})$, where $\bar{\boldsymbol{x}}_{k|k-1}$ and $\bar{\mathsf{V}}_{k|k-1}$ are the sample mean and the sample covariance of the forecast ensemble.
- The cross entropy

$$-\int \mathcal{N}(\boldsymbol{x}_{k|k-1}, \mathsf{V}_{k|k-1}) \log \mathcal{N}(\bar{\boldsymbol{x}}_{k|k-1}, \alpha^2 \bar{\mathsf{V}}_{k|k-1})\, d\boldsymbol{x}$$

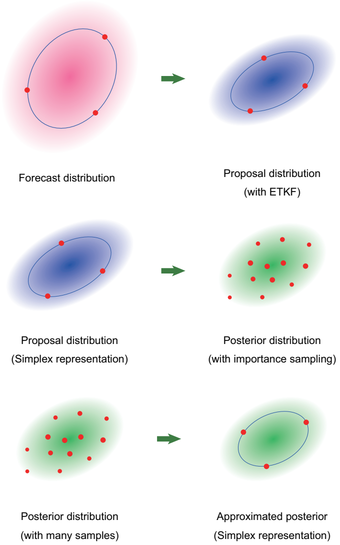
  is minimized when

$$\alpha^2 = \frac{1}{\dim \boldsymbol{x}_k} \mathrm{tr}\left[\bar{\mathsf{V}}_{k|k-1}^{-1} \mathsf{V}_{k|k-1} + \bar{\mathsf{V}}_{k|k-1}^{-1} (\bar{\boldsymbol{x}}_{k|k-1} - \boldsymbol{x}_{k|k-1})(\bar{\boldsymbol{x}}_{k|k-1} - \boldsymbol{x}_{k|k-1})^T\right].$$

- If $E\left[\bar{\mathsf{V}}_{k|k-1}^{-1}\right] = \mathsf{V}_{k|k-1}^{-1}$,

$$E[\alpha^2] = 1 + \frac{1}{\dim \boldsymbol{x}_k} \mathrm{tr}\left[\mathsf{V}_{k|k-1}^{-1} (\bar{\boldsymbol{x}}_{k|k-1} - \boldsymbol{x}_{k|k-1})(\bar{\boldsymbol{x}}_{k|k-1} - \boldsymbol{x}_{k|k-1})^T\right] \geq 1.$$

# Overview




- We consider cases in which the forecast PDF is represented by a simplex representation with a limited-size ensemble.
- To allow nonlinear or non-Gaussian observation models, the simplex representation is converted into a Monte Carlo representation. Then the importance sampling method is applied.
- Finally, the importance sampling result is converted into a simplex representation again.

# Some definitions

- Suppose that the forecast distribution is represented by an ensemble $\{\boldsymbol{x}_{k|k-1}^{(1)}, \ldots, \boldsymbol{x}_{k|k-1}^{(N)}\}$.
- The mean of the forecast distribution is obtained as:

$$\bar{\boldsymbol{x}}_{k|k-1} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_{k|k-1}^{(i)}.$$

- We define a matrix $\mathsf{X}_{k|k-1}$ and $\mathsf{Y}_{k|k-1}$ as

$$\mathsf{X}_{k|k-1} = \frac{1}{\sqrt{N}} \begin{pmatrix} \delta\boldsymbol{x}_{k|k-1}^{(1)} & \cdots & \delta\boldsymbol{x}_{k|k-1}^{(N)} \end{pmatrix}, \quad \mathsf{Y}_{k|k-1} = \frac{1}{\sqrt{N}} \begin{pmatrix} \delta\boldsymbol{y}_{k|k-1}^{(1)} & \cdots & \delta\boldsymbol{y}_{k|k-1}^{(N)} \end{pmatrix},$$

  where $\delta\boldsymbol{x}_{k|k-1}^{(i)} = \boldsymbol{x}_{k|k-1}^{(i)} - \bar{\boldsymbol{x}}_{k|k-1}$ and $\delta\boldsymbol{y}_{k|k-1}^{(i)} = H_k(\boldsymbol{x}_{k|k-1}^{(i)}) - \overline{H_k(\boldsymbol{x}_{k|k-1})}$, respectively, and we assumed the following observation model

$$\boldsymbol{y}_k = H_k(\boldsymbol{x}_k) + \boldsymbol{w}_k.$$

- The covariance matrix of the forecast (predictive) distribution is written as $\mathsf{V}_{k|k-1} = \mathsf{X}_{k|k-1}\mathsf{X}_{k|k-1}^T$.

# Ensemble transform Kalman filter (ETKF)

- The mean of the filtered distribution is obtained according to the Kalman filter algorithm:

$$\bar{\boldsymbol{x}}^{\dagger}_{k|k} = \bar{\boldsymbol{x}}_{k|k-1} + \mathsf{K}_k \left(\boldsymbol{y}_k - \mathsf{H}_k \boldsymbol{x}_{k|k-1}\right)$$

- The square root of the covariance matrix is also calculated as $\mathsf{X}^{\dagger}_{k|k} = \mathsf{X}_{k|k-1}\mathsf{T}_k$, where the matrix $\mathsf{T}_k$ is designed to satisfy $\mathsf{V}^{\dagger}_{k|k} = \mathsf{X}^{\dagger}_{k|k}\mathsf{X}^{\dagger T}_{k|k}$ and $\mathsf{X}^{\dagger}_{k|k}\mathbf{1} = \mathbf{0}$, where $\mathbf{1} = (1 \ \cdots \ 1)^T$. The latter condition is required to preserve the mean of the PDF.
- Using the following eigen-value decomposition

$$\mathsf{Y}_{k|k-1}\mathsf{R}_k^{-1}\mathsf{Y}_{k|k-1} = \mathsf{U}_k\Lambda_k\mathsf{U}_k^T,$$

the matrices $\mathsf{K}_k$ and $\mathsf{T}_k$ are obtained as follows:

$$\mathsf{K}_k = \mathsf{X}_{k|k-1}\mathsf{U}_k(\mathrm{I}_N + \Lambda_k)^{-1}\mathsf{U}_k^T\mathsf{Y}_{k|k-1}^T\mathsf{R}_k^{-1},$$
$$\mathsf{T}_k = \mathsf{U}_k(\mathrm{I}_N + \Lambda_k)^{-\frac{1}{2}}\mathsf{U}_k^T,$$

where $\mathsf{R}_k$ is the covariance matrix of the observation noise.

# Sampling from the ETKF estimate

- The ETKF estimates the filtered (posterior) distribution as a Gaussian distribution $\mathcal{N}(\bar{\boldsymbol{x}}^{\dagger}_{k|k}, \mathsf{V}^{\dagger}_{k|k})$.
- However, it does not actually calculate the covariance matrix $\mathsf{V}^{\dagger}_{k|k}$ itself. Instead, a square root of the covariance matrix $\mathsf{X}^{\dagger}_{k|k}$ is calculated.
- Using the matrix $\mathsf{X}^{\dagger}_{k|k}$, we can easily generate a large number of random numbers obeying $\mathcal{N}(\bar{\boldsymbol{x}}^{\dagger}_{k|k}, \mathsf{V}^{\dagger}_{k|k})$ using the following generative model:

$$\boldsymbol{x}_k = \bar{\boldsymbol{x}}^{\dagger}_{k|k} + \mathsf{X}^{\dagger}_{k|k} \boldsymbol{z}_k, \quad \text{where } \boldsymbol{z}_k \sim \mathcal{N}(\boldsymbol{0}, \mathrm{I}_N).$$

# Importance sampling

- Since the posterior distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ is written as

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \frac{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}{\pi(\boldsymbol{x}_k)}\pi(\boldsymbol{x}_k) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})\pi(\boldsymbol{x}_k)}\pi(\boldsymbol{x}_k),$$

  the posterior $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ can be represented by the importance sampling using the sample drawn from $\pi(\boldsymbol{x}_k)$:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{j=1}^{M} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{\pi(j)})p(\boldsymbol{x}_k^{\pi(j)}|\boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})\pi(\boldsymbol{x}_k^{\pi(j)})}\delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{\pi(j)}).$$

- In the normal particle filter, the forecast $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$ is used as $\pi(\boldsymbol{x}_k)$.
- On the other hand, we use the estimate of $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ obtained by the ETKF as the proposal $\pi(\boldsymbol{x}_k)$.

# Importance sampling

- If we obtain the proposal $\pi(\boldsymbol{x}_k)$ by the ETKF, we can generate a large number of particles from $\pi(\boldsymbol{x}_k)$ according to the following generative model:

$$\boldsymbol{x}_k^{\pi(j)} = \bar{\boldsymbol{x}}_{k|k}^{\dagger} + \mathsf{X}_{k|k}^{\dagger}\boldsymbol{z}_k^{(j)} \quad \left(\boldsymbol{z}_k^{(j)} \sim \mathcal{N}(\mathbf{0}, \mathrm{I}_N)\right).$$

- In order to approximate the posterior $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ using the importance sampling method as follows:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{j=1}^{M} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{\pi(j)})p(\boldsymbol{x}_k^{\pi(j)}|\boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})\pi(\boldsymbol{x}_k^{\pi(j)})}\delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{\pi(j)}),$$

  we need to calculate

$$\frac{p(\boldsymbol{x}_k^{\pi(j)}|\boldsymbol{y}_{1:k-1})}{\pi(\boldsymbol{x}_k^{\pi(j)})}$$

  for each particle $\boldsymbol{x}_k^{\pi(j)}$. (We can obtain $p(\boldsymbol{y}_k|\boldsymbol{x}_k^{\pi(j)})$ from the observation model. )

# Importance weight

- According to the generative model

$$\boldsymbol{x}_k^{\pi(j)} = \bar{\boldsymbol{x}}_{k|k}^{\dagger} + \mathsf{X}_{k|k}^{\dagger}\boldsymbol{z}_k^{(j)} \quad \left(\boldsymbol{z}_k^{(j)} \sim \mathcal{N}(\mathbf{0}, \mathrm{I}_N)\right),$$

$\pi(\boldsymbol{x}_{k|k}^{\pi,(j)})$ can be associated with the probability density for $\boldsymbol{z}_k^{(j)}$, $p(\boldsymbol{z}_k^{(j)})$.

- The probability density $p(\boldsymbol{z}_k^{(j)})$ is proportional to $\exp\left(-\|\boldsymbol{z}_k^{(j)}\|^2/2\right)$.
- Considering that $\mathsf{X}_{k|k}^{\dagger}$ satisfies the mean-preserving condition $\mathsf{X}_{k|k}^{\dagger}\mathbf{1} = \mathbf{0}$, the component parallel to $\mathbf{1}$ is projected onto a null space. We therefore obtain

$$\pi(\boldsymbol{x}_{k|k}^{\pi,(j)}) \propto \exp\left[-\frac{1}{2}\left(\left\|\boldsymbol{z}_k^{(j)}\right\|^2 - \frac{(\mathbf{1}^T\boldsymbol{z}_k^{(j)})^2}{N}\right)\right].$$

We consider that a sample from the forecast $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$ is generated according to the following model:

$$\boldsymbol{x}_k = \bar{\boldsymbol{x}}_{k|k-1} + \mathsf{X}_{k|k-1}\boldsymbol{z}_k \quad \left(\boldsymbol{z}_k \sim \mathcal{N}(\mathbf{0}, I_N)\right).$$

We can then evaluate the probability density that $\boldsymbol{x}_{k|k}^{\pi,(j)}$ is drawn from the forecast distribution as follows:

$$\begin{aligned}
\boldsymbol{x}_{k|k}^{\pi,(j)} &= \bar{\boldsymbol{x}}_{k|k}^{\dagger} + \mathsf{X}_{k|k}^{\dagger}\boldsymbol{z}_k^{(j)} = \bar{\boldsymbol{x}}_{k|k-1} + \mathsf{K}_k\left(\boldsymbol{y}_k - \overline{\boldsymbol{h}_k(\boldsymbol{x}_{k|k-1})}\right) + \mathsf{X}_{k|k-1}\mathsf{T}_k\boldsymbol{z}_k^{(j)} \\
&= \bar{\boldsymbol{x}}_{k|k-1} + \mathsf{X}_{k|k-1}\left[\mathsf{U}_k(\mathrm{I}_N + \Lambda_k)^{-1}\mathsf{U}_k^T\mathsf{Y}_{k|k-1}^T\mathsf{R}^{-1}\left(\boldsymbol{y}_k - \overline{\boldsymbol{h}_k(\boldsymbol{x}_{k|k-1})}\right) + \mathsf{T}_k\boldsymbol{z}^{(j)}\right] \\
&= \bar{\boldsymbol{x}}_{k|k-1} + \mathsf{X}_{k|k-1}\boldsymbol{\zeta}_k^{(j)}
\end{aligned}$$

where

$$\boldsymbol{\zeta}_k^{(j)} = \mathsf{U}_k(\mathrm{I}_N + \Lambda_k)^{-1}\mathsf{U}_k^T\mathsf{Y}_{k|k-1}^T\mathsf{R}^{-1}\left(\boldsymbol{y}_k - \overline{\boldsymbol{h}_k(\boldsymbol{x}_{k|k-1})}\right) + \mathsf{T}_k\boldsymbol{z}^{(j)}.$$

We therefore obtain

$$p(\boldsymbol{x}_{k|k}^{\pi,(j)}|\boldsymbol{y}_{1:k-1}) \propto \exp\left[-\frac{1}{2}\left(\left\|\boldsymbol{\zeta}_k^{(j)}\right\|^2 - \frac{(\mathbf{1}^T\boldsymbol{\zeta}_k^{(j)})^2}{N}\right)\right].$$

As seen previously, the posterior distribution is approximated as

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{j=1}^{M} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{\pi(j)})p(\boldsymbol{x}_k^{\pi(j)}|\boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})\pi(\boldsymbol{x}_k^{\pi(j)})}\delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{\pi(j)}).$$

If we generate the proposal sample according to the following model:

$$\boldsymbol{x}_{k|k}^{\pi,(j)} = \bar{\boldsymbol{x}}_{k|k-1} + \mathsf{X}_{k|k-1}\boldsymbol{z}_k^{(j)} \quad \left(\boldsymbol{z}_k^{(j)} \sim \mathcal{N}(\boldsymbol{0}, \mathrm{I}_N)\right),$$

the weight for each particle can be given as follows:

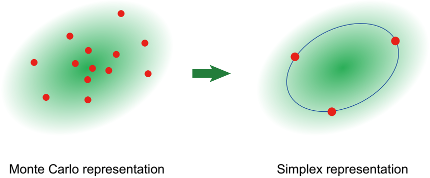
$$\beta_k^{(j)} \propto \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_{k|k}^{\pi,(j)})\ \exp\left[-\frac{1}{2}\left(\left\|\boldsymbol{\zeta}_k^{(j)}\right\|^2 - \frac{(\boldsymbol{1}^T\boldsymbol{\zeta}_k^{(j)})^2}{N}\right)\right]}{\exp\left[-\frac{1}{2}\left(\left\|\boldsymbol{z}_k^{(j)}\right\|^2 - \frac{(\boldsymbol{1}^T\boldsymbol{z}_k^{(j)})^2}{N}\right)\right]}.$$

We then obtain a new approximation of the posterior PDF:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{j=1}^{M} \beta_k^{(j)}\delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{\pi(j)}).$$

# Ensemble reconstruction

- Using the weight $\beta_k^{(j)}$, we can obtain a random sample from the posterior $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ with the rejection sampling method or the independent chain Metropolis-Hastings method.
- However, we consider the case in which a large ensemble size is not allowed. A small-size ensemble generated randomly would not give a good approximation of $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$.
- To avoid the errors due to the randomness, we construct a simplex approximation that represents the first and second order moments of the posterior.



# Moments on the $z$-space

If we calculate the mean and the covariance on the $\boldsymbol{z}$-space:

$$\bar{\boldsymbol{z}}_k = \sum_{i=1}^{M} \beta_k^{(j)} \boldsymbol{z}_k^{(j)}, \qquad \mathsf{V}_{z,k|k} = \sum_{i=1}^{M} \beta_k^{(j)} (\boldsymbol{z}_k^{(j)} - \bar{\boldsymbol{z}}_k)(\boldsymbol{z}_k^{(j)} - \bar{\boldsymbol{z}}_k)^T,$$

the mean and the covariance of the filtered distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ are given as follows:

$$\bar{\boldsymbol{x}}_{k|k} = \bar{\boldsymbol{x}}_{k|k}^{\dagger} + \mathsf{X}_{k|k}^{\dagger} \bar{\boldsymbol{z}}_k, \qquad \mathsf{V}_{k|k} = \mathsf{X}_{k|k} \mathsf{X}_{k|k}^T = \mathsf{X}_{k|k}^{\dagger} \mathsf{V}_{z,k|k} \mathsf{X}_{k|k}^{\dagger T}$$

where $\bar{\boldsymbol{x}}_{k|k}^{\dagger}$ and $\mathsf{X}_{k|k}^{\dagger}$ provide the estimate by the ETKF.

To avoid the bias of the ensemble mean, the new $\mathsf{X}_{k|k}$ should also satisfy

$$\mathsf{X}_{k|k} \mathbf{1} = \mathbf{0}.$$

We define the following matrix

$$\mathsf{A} = \mathrm{I}_N - \frac{1}{N}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix},$$

which obviously satisfies

$$\mathsf{A}\mathbf{1} = 0.$$

The covariance matrix $\mathsf{V}_{k|k}$ can then be written as follows:

$$\begin{aligned} \mathsf{V}_{k|k} &= \mathsf{X}^{\dagger}_{k|k}\mathsf{V}_{z,k|k}\mathsf{X}^{\dagger T}_{k|k} \\ &= \mathsf{X}^{\dagger}_{k|k}\mathsf{A}\,\mathsf{V}_{z,k|k}\mathsf{A}^T\mathsf{X}^{\dagger T}_{k|k} \end{aligned}$$

because obviously $\mathsf{X}^{\dagger}_{k|k} = \mathsf{X}^{\dagger}_{k|k}\mathsf{A}$.

When we calculate the eigen-value decomposition of the matrix $\mathsf{A}\,\mathsf{V}_{z,k|k}\mathsf{A}^T$ as

$$\mathsf{A}\,\mathsf{V}_{z,k|k}\mathsf{A}^T = \mathsf{U}_{z,k}\Gamma_k\mathsf{U}_{z,k}^T,$$

the matrix $\mathsf{U}_{z,k}$ contains an eigen-vector which is parallel to $\mathbf{1}$ and corresponds to zero eigen-value. Therefore, if we define $\mathsf{X}_{k|k}$ as

$$\mathsf{X}_{k|k} = \mathsf{X}_{k|k}^{\dagger}\mathsf{U}_{z,k}\Gamma_k^{\frac{1}{2}}\mathsf{U}_{z,k}^T,$$

it satisfies both of the necessary conditions:

$$\mathsf{X}_{k|k}\mathsf{X}_{k|k}^T = \mathsf{X}_{k|k}^{\dagger}\mathsf{V}_{z,k|k}\mathsf{X}_{k|k}^{\dagger T},$$
$$\mathsf{X}_{k|k}\mathbf{1} = \mathbf{0}.$$

# Ensemble reconstruction

Finally, we obtain ensemble perturbations:

$$\begin{pmatrix} \delta \boldsymbol{x}_{k|k-1}^{(1)} & \cdots & \delta \boldsymbol{x}_{k|k-1}^{(N)} \end{pmatrix} = \sqrt{N}\, \mathsf{X}_{k|k-1}.$$

We then obtain the filtered ensemble:

$$\boldsymbol{x}_{k|k}^{(i)} = \bar{\boldsymbol{x}}_{k|k} + \delta \boldsymbol{x}_{k|k}^{(i)}.$$

# Remark

- Using the generative model, $\boldsymbol{x}_k^{\pi(j)} = \bar{\boldsymbol{x}}_{k|k}^{\dagger} + \mathsf{X}_{k|k}^{\dagger}\boldsymbol{z}_k^{(j)}$, the ensemble members are generated in the subspace spanned by the ensemble members.
- We could consider a small uncertainty in the complement space as follows

$$\boldsymbol{x}_k^{\pi,(j)} = \bar{\boldsymbol{x}}_{k|k}^{\dagger} + \mathsf{X}_{k|k}^{\dagger}\boldsymbol{z}_k^{(j)} + \boldsymbol{\varepsilon}_k^{(j)},$$

  where $\boldsymbol{\varepsilon}_k^{(j)}$ is a random sample representing the uncertainty of the orthogonal complement space. But, this may invoke 'the curse of dimensionality'.
- As far as we ignore the complement space, we can convert between the importance sampling result and a spherical simplex representation through the calculation in the small subspace spanned by the forecast ensemble members. This would help reduce the computational cost.

# Experiment

We performed experiments using the Lorenz 96 model (Lorenz and Emanuel 1998):

$$\frac{dx_m}{dt} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + f$$

where $x_{-1} = x_{M-1}$, $x_0 = x_M$, and $x_{M+1} = x_1$. We take the dimension of a state vector $M$ to be 40 and the forcing term $f$ to be 8.

It was assumed that $x_m$ can be observed only if $m$ is an even number. This means that the half of the variables $x_m$ are observable.

First, the following nonlinear observation model is considered:

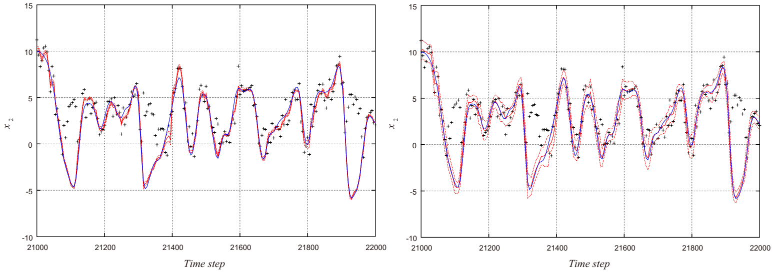
$$p(\boldsymbol{y}_k|\boldsymbol{x}_k) = \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \exp\left[-\frac{\left\|\boldsymbol{y}_k - \boldsymbol{h}(\boldsymbol{x}_k)\right\|^2}{2\sigma^2}\right],$$

Multiplicative covariance inflation (Anderson and Anderson, 1999) was applied. The inflation factor was set at 1.1.

# Result

An estimate of an observed variable

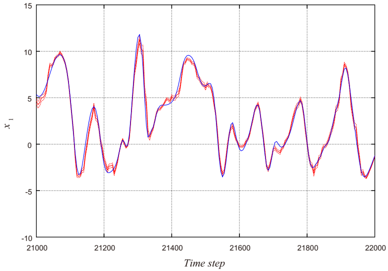
- With 32 ensemble members (and 2048 particles for importance sampling)
- RMSE: 0.43 (with the hybrid algorithm), 0.68 (with the ETKF)

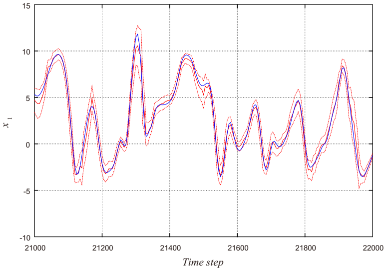


# Result

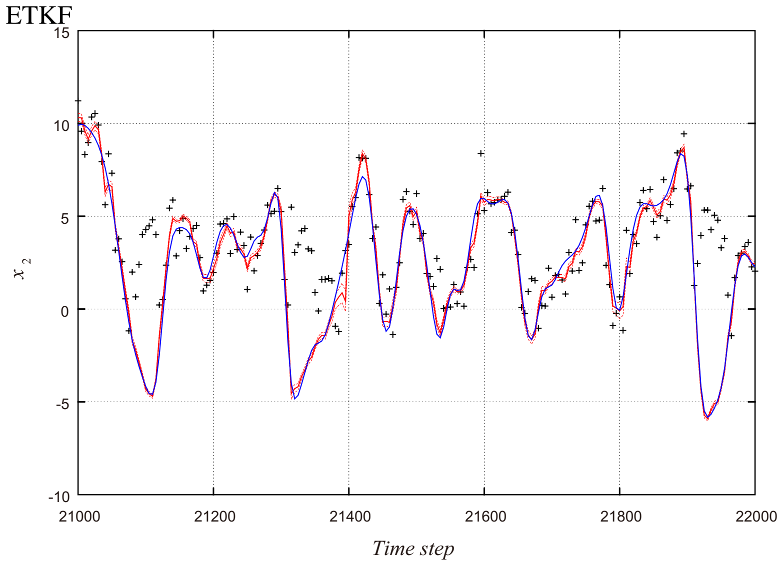
An estimate of an unobserved variable

- With 32 ensemble members (and 2048 particles for importance sampling)
- RMSE: 0.43 (with the hybrid algorithm), 0.68 (with the ETKF)

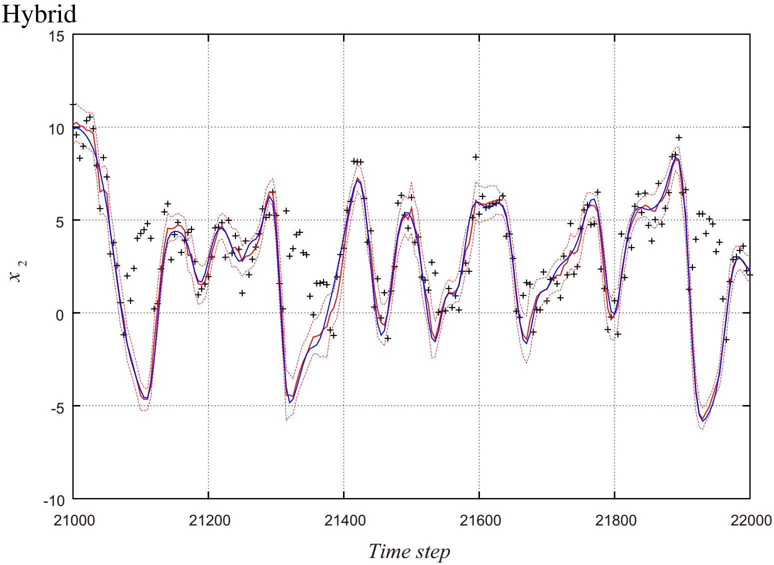


# Result


ETKF
15
10
5
0
-5
-10
21000
21200
21400
21600
21800
22000
$x_2$
Time step

# Result


Hybrid
15
10
5
0
-5
-10
21000
21200
21400
21600
21800
22000
$x_2$
Time step

Second, the following non-Gaussian observation model is considered:

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k) = \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \prod_{i=1}^{M} \exp\left[-\frac{\left\|\log \boldsymbol{y}_{i,k} - \log(\boldsymbol{x}_{2i,k}^2 + 1)\right\|^2}{2\sigma^2}\right],$$

where $\sigma = 0.4$.

As $p(\boldsymbol{y}_k|\boldsymbol{x}_k)$ is maximized at $\boldsymbol{x}_{2i,k} = \pm\sqrt{|y_{i,k} - 1|}$, the following Gaussian approximation is employed to obtain a proposal distribution:

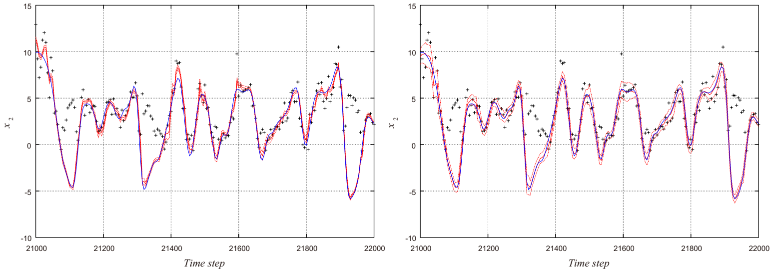
$$p'(\boldsymbol{y}_k|\boldsymbol{x}_k) = \frac{1}{(2\pi\sigma'^2)^{\frac{M}{2}}} \prod_{i=1}^{M} \exp\left[-\frac{\left\|\sqrt{|\boldsymbol{y}_{i,k} - 1|} - |\boldsymbol{x}_{2i,k}|\right\|^2}{2\sigma'^2}\right].$$

The inflation factor was set at 1.1 again.

# Result

An estimate of an observed variable

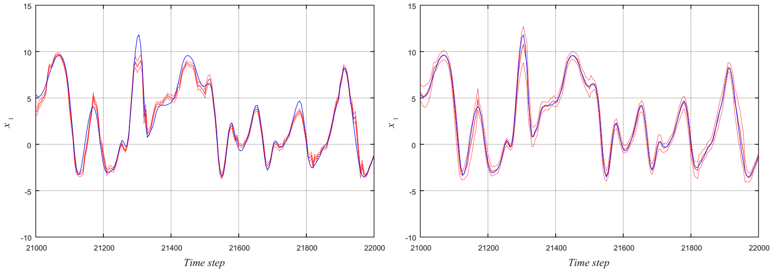
- With 32 ensemble members (and 2048 particles for importance sampling)
- RMSE: 0.28 (with the hybrid algorithm), 0.81 (with the ETKF)



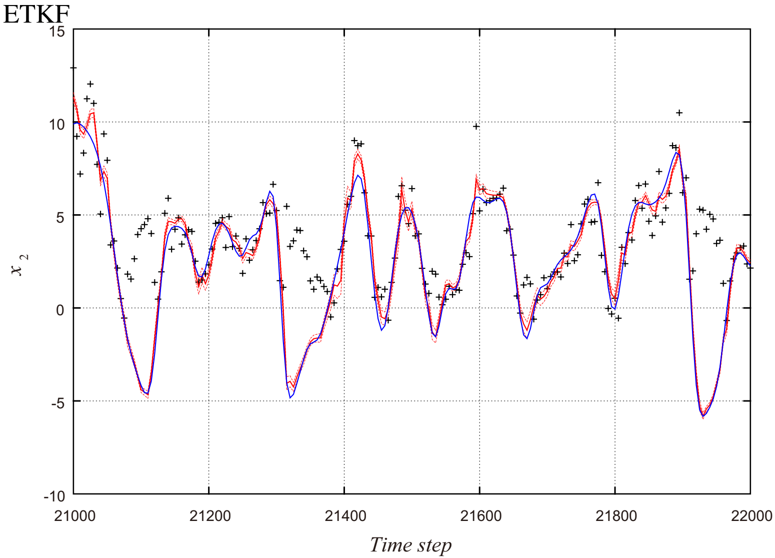
# Result

An estimate of an unobserved variable

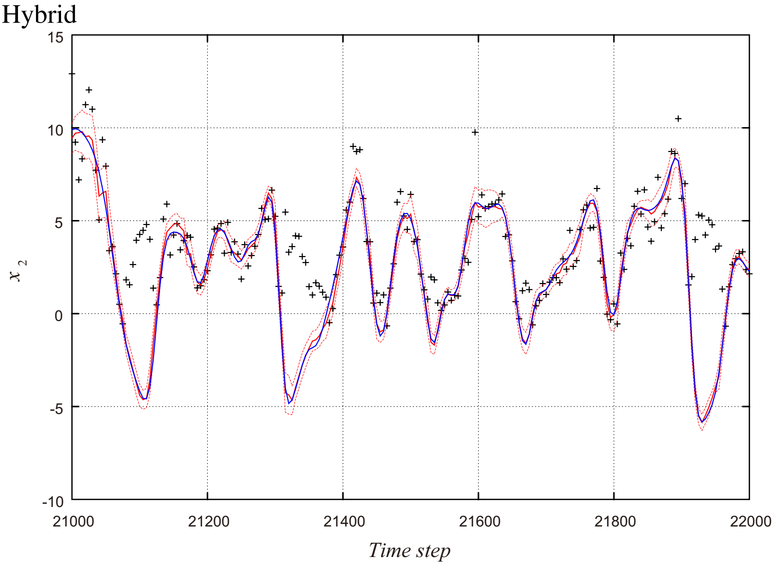
- With 32 ensemble members (and 2048 particles for importance sampling)
- RMSE: 0.28 (with the hybrid algorithm), 0.81 (with the ETKF)



# Result

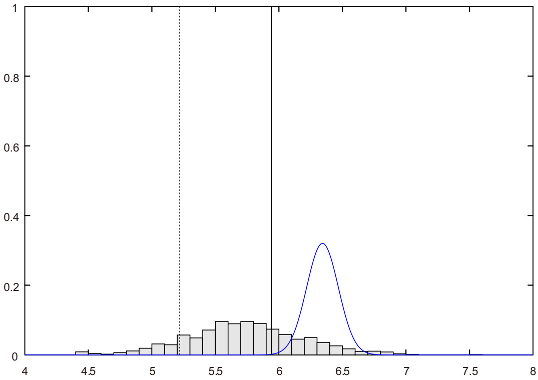

ETKF
15
10
5
0
-5
-10
21000
21200
21400
21600
21800
22000
$x_2$
Time step

# Result


Hybrid
15
10
5
0
-5
-10
21000
21200
21400
21600
21800
22000
x 2
Time step

# Result

- Blue line: the result with the ETKF
- Histogram: the result with the hybrid algorithm
- Solid vertical line: the true state
- Dashed vertical line: the observed value



# Summary

- We propose a hybrid algorithm which combines the ensemble transform Kalman filter (ETKF) and the importance sampling.
- While the importance sampling method requires a large number of particles, the ETKF is based on a spherical simplex representation which uses less particles than the state dimension. We thus need to make the conversion between a simplex representation and a Monte Carlo representation.
- In our approach, this conversion is performed in the low-dimensional subspace spanned by the forecast ensemble members.
- Even though the uncertainty is considered only in the subspace, the proposed approach seems to well work in the cases with nonlinear, non-Gaussian observation models in which the application of standard ensemble Kalman filters is not valid.