Clustered Particle Filtering for High-Dimensional Non-Gaussian Systems

Yoonsang Lee

Center for Computational Sciences and Engineering
Lawrence Berkeley National Laboratory

Joint work with
Andrew J Majda, Courant Institute

Data assimilation and Non-Gaussian statistics

- Non-Gaussian features in Geophysical fluids
- Ensemble based methods : use Gaussian assumption
- Particle filters

$$p(x) = \sum_{k}^{K} w_k \delta(x - x_k)$$

where $w_k \geq 0$ and $\sum_k w_k = 1$.

Limitation of particle filters

- Not applicable to high-dimensional systems
- Particle collapse : A small fraction of particles have the most weights
- Number of particles increases exponentially with the dimension of the system
- No localization : observation affects all state variables even if they are not correlated

Clustered Particle Filters (CPF), L. and Majda, PNAS

A new class of particle filters to address the issues of ensemble-based filters and standard particle filters

Key features

- Capture non-Gaussian statistics
- Use a relatively few particles
- Implements coarse-grained localization through the clustering of state variables
- Particle adjustment
- Simple but robust even with sparse and high-quality observations
- No adjustable parameter

For simplicity of the description of the algorithm, we assume that the observation is linear, which observes partial components of the state variable

Schematics of several particle filters

Total dimension is 6 and two observations at x_2 and x_5

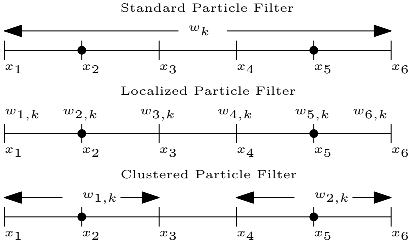


Figure: Schematics of particle weight, w_k, for the k-th particle.

- Standard particle filter uses the same particle weight at different locations
- Localized particle filter uses different weights at different locations
- In CPF, sparse observation network determines the clustering of state variables; two clusters for CPF
- Weights are the same in the same cluster

Another problem of particle filter

- The mean of $p(x) = \sum_k^K w_k \delta(x - x_k), \quad w_k \leq 0, \sum_k w_k = 1$ is a convex combination of $x_k, w_k x_k$
- If the posterior mean is not in the convex hull of the prior samples, the particle filter cannot represent the correct posterior distribution ($\because$ particle filtering updates only the particle weights)

Particle Adjustment

Adjust the prior particles to match the Kalman posterior mean and covariance (i.e., update the prior under Gaussian assumption) **if** the prior particles cannot represent the observation

$$y_j \notin \{\sum_k^K q_k[\mathbf{x}^f_{C_j,k}]|, \forall q_k \geq 0 \text{ such that } \sum_k q_k = 1\}$$

y_j : j-th observation component, $\mathbf{x}^f_{C_j}$: prior particles in the j-th cluster C_j
Note several adjustment or transformation methods of ensemble-based methods can be applied to the particle adjustment. In this study, we use EAKF by Anderson.

Hard Threshold Clustered Particle Filter Algorithm - one step assimilation

Given :
1) N_{obs} observations $\{y_1, y_2, ..., y_{N_{obs}}\}$
2) prior K particles $\{\mathbf{x}^f_{C_j,k}, k = 1, 2, ..., K\}$ and weight vectors $\{\omega^f_{l,k}, k = 1, 2, ..., K\}$ for each cluster $C_l, l = 1, 2, .., N_{obs}$

For y_j from $j = 1$ to N_{obs}

If $y_j \notin \{\sum_k^K q_k \mathbf{H}[\mathbf{x}^f_{C_j,k}]|, \forall q_k \geq 0$ such that $\sum_k q_k = 1\}$

Do particle adjustment

Else Use particle filtering

Update $\{\omega^f_{j,k}\}$ using standard PF update

If $K_{eff} = \frac{1}{\sum_k (\omega^a_{l,k})^2} < \frac{K}{2}$

Do resampling

Add additional noise to the resampled particles

$$\mathbf{x}_{C_l,Resample(k)} \leftarrow \mathbf{x}_{C_l,Resample(k)} + \epsilon \tag{1}$$

where ϵ is IID Gaussian noise with zero mean and variance r_{noise}

End If

End If

End For

Multiscale Clustered Particle Filtering

- Multiscale data assimilation (particle filter, ensemble filter) Lee and Majda, **Multiscale Methods for Data Assimilation in Turbulent Systems**, SIAM MMS, 2015
- Probability distribution : conditional Gaussian mixture

$$p(u) = \sum_{k}^{K} w_k \delta(\overline{u} - \overline{u}_k) \mathcal{N}(u'(\overline{u}_k), R'(\overline{u}))$$

- Particle filtering for the large scales and Kalman update for the small scales
- Particle adjustment : Accounts for representative error, the error due to the contribution of unresolved scales

Multiscale Clustered Particle Filtering

- For the j-th observation component v_j, the posterior particle weights are given by

$$w^a_{l,k} = \begin{cases} \frac{w^f_{l,k} I_k}{\sum_k w^f_{l,k} I_k} & l = j, \\ w^f_{l,k} & l \neq j \end{cases} \quad (2)$$

  where $I_k = \int p(v_j | \mathbf{x}_{C_l,k}, \mathbf{y}_{C_l}) p(\mathbf{y}_{C_l} | \mathbf{x}_k) d\mathbf{y}_{C_l}$.
- For the particle adjustment step, use the standard Ensemble update formula using an increased observation error variance (i.e., representation error)

$$\text{Kalman gain } G = R^f H^T (H R^f H^T + r_o I + R')^{-1} \quad (3)$$

Two-layer coupled Lorenz-96 system

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) + \lambda_1 \sum_{j=1}^{J} y_{i,j} - d_1 x_i + F, \quad i = 1, 2, ..., I$$
$$\frac{dy_{i,j}}{dt} = \frac{a_L x_i + a_S y_{i,j+1}}{\epsilon}(y_{i,j-1} - y_{i,j+2}) - \lambda_2 x_i - d_2 y_{i,j}, \quad j = 1, 2, ..., J \tag{4}$$

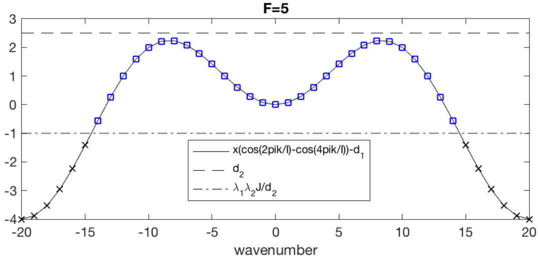
where x_i is periodic in i and $y_{i,j}$ is periodic in both i and j.

- $\mathbf{x} = \{x_i\}$: slow-climate variable of size I
- $\mathbf{y} = \{y_{i,j}\}$: fast-weather variable of size IJ
- $\epsilon > 0$: an explicit time-scale separation parameter
- F : an external slow forcing
- Resolved variable $Y_i = \frac{1}{J}\sum_j y_{ij}$

Weakly Chaotic Regime

- $I = 40$, $J = 10$, there are total 440 variables
- $a_L = 1$, $a_S = 0$, $F = 5$, $\lambda_1 = 1/4$, $\lambda_2 = -1$, $d_1 = 1.5$, $d_2 = 2.5$, $\epsilon = 1$

Linearly stability



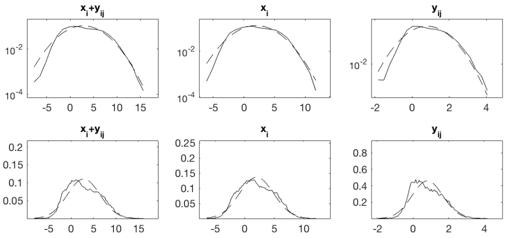
Blue square : linearly unstable modes

Weakly Chaotic Regime

Climatological properties

	x_i	y_{ij}
mean	2.01	0.80
variance	8.51	0.75
skewness	0.18	0.38
kurtosis	2.40	2.68
corr length	$\lesssim 1$	$\lesssim 1$
corr time	2.93	3.52

Table: Climatological properties of the weakly chaotic regime



Experiment setup

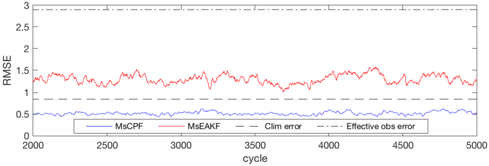
- Each observation component directly observes the sum of x_j and $y_{j,5}$ (observation has contributions from both the large- and small-scale components)
- Raw observation error variance is only 1% of the total variance (and thus the most observation error comes from the unresolved scale components)
- Run 5000 assimilation cycles and use the last 3000 cycles to measure the filter performance
- 50 samples for both MsCPF and MsEAKF

Experiment setup

- We compare two multiscale filtering methods, MsCPF and MsEAKF
  - MsCPF : Multiscale Clustered Particle Filter
  - MsEAKF : Multiscale EAKF (which uses EAKF for the large-scale estimation)
- To see the effect of the filtering method, we use the perfect forecast model (i.e., the forecast use the same numerical method as the method to generate the true signal)

Experiment results

Time series of forecast RMS errors



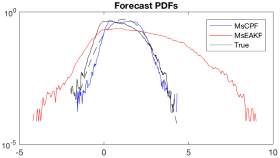
Time averaged RMS errors and pattern correlation in parenthesis.

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.05	0.52 (0.90)	1.30 (0.64)	0.55 (0.83)	1.46 (0.52)
0.10	0.54 (0.87)	1.32 (0.63)	0.61 (0.81)	1.53 (0.50)

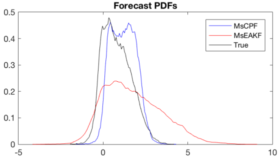
Table: Climatological error is 0.844. Effective observation error is 2.900.

Experiment results

Forecast PDFs by MsCPF (blue) and MsEAKF (red) along with the true value (black)



(a) in log-scale



(b) without scaling

Figure: Dash-line is the Gaussian fit to the true PDF.

Forecast relative entropy using the forecast estimate PDFs.

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.05	0.1631	0.3024	0.1787	0.4328
0.10	0.1791	0.3234	0.1891	0.4523

Summary

Multiscale data assimilation method using the clustered particle filter for resolved scales

- Captures non-Gaussian statistics
- Efficient - requires only a small number of particles
- Robust under sparse and high-quality observations
- Clustering of state variables
- Particle adjustment to prevent particle collapse

Future works

- Dense and vector observations
- Two- and three-dimensional spaces
- Bayesian parameter estimation

References

- Y. Lee and A.J. Majda, **Multiscale Data Assimilation and Prediction using Clustered Particle Filters**, submitted
- Y. Lee and A.J. Majda, **State estimation and prediction using clustered particle filters**, PNAS, 113(51), 14609–14614, 2016

Thank you for your attention