State Estimation and Prediction using clustered Particle Filters

Center for Atmosphere Ocean Science (CAOS)
Courant Institute of Mathematical Sciences (CIMS)
New York University

Lecture on Filtering Turbulent Signals in Complex Systems
Main reference


Multiscale data assimilation

Filtering is the process of obtaining the best statistical estimate of a natural system from partial observations of the true signal. With Bayes formula,

\[
p_{m+1,+}(u) \equiv p_{m+1}(u \mid v_{m+1}) = \frac{p_{m+1}(v_{m+1} \mid u) p_{m+1,-}(u)}{\int p_{m+1}(v_{m+1} \mid u) p_{m+1,-}(u) \, du}.
\]
Examples

- Kalman Filter

  Suppose Gaussian prior and linear observations

  \[ p_-(u) \sim \mathcal{N}(\bar{u}^-, R^-), \quad v = Gu + \sigma, \quad \sigma \sim \mathcal{N}(0, R_0), \]

  then the posterior becomes \( p_+(u) = \mathcal{N}(\bar{u}^+, R^+) \) with

  \[
  \bar{u}^+ = \bar{u}^- + K(v - G\bar{u}^-),
  \]

  \[
  R^+ = (I - KG)R^-,
  \]

  \[
  K = R^-G^T(GR^-G^T + R_0)^{-1}.
  \]
Examples

- Kalman Filter
  Suppose Gaussian prior and linear observations

\[ p_-(u) \sim \mathcal{N}(\bar{u}^-, R^-), \quad v = Gu + \sigma, \quad \sigma \sim \mathcal{N}(0, R_0), \]

then the posterior becomes \( p_+(u) = \mathcal{N}(\bar{u}^+, R^+) \) with

\[ \bar{u}^+ = \bar{u}^- + K (v - G\bar{u}^-), \]
\[ R^+ = (I - KG) R^- , \]
\[ K = R^- G^T (GR^- G^T + R_0)^{-1}. \]

- Particle Filter
  Suppose particle representation of the prior

\[ p_-(u) = \sum_j p_{j,-} \delta (u - u_j), \]

then get the posterior weights and distribution by

\[ p_{j,+} \propto p(v \mid u_j) p_{j,-}, \quad p_+(u) = \sum_j p_{j,+} \delta (u - u_j). \]
Strategies for filtering turbulent dynamical systems

Particle Filter v.s. Kalman Filter
- Particle filter based on Monte-Carlo approaches with various resampling strategies provides better estimates of low-dimensional systems than Kalman filter in the presence of strong nonlinearity and highly non-Gaussian distributions;
- These particle filtering strategies become less feasible for high-dimensional turbulent systems considering computational constraints and particle collapses (Curse of ensemble/dimensionality).

Bayesian Modeling and Reduced Order Filtering Strategies
- Ensemble adjustment Kalman filter (EAKF), Anderson;
- Rank histogram particle filter (RHPF), Anderson;
- Maximum entropy particle filter (MEPF), Majda & Harlim;
- Implicit particle filter/smoother, van Leeuwen, Miller, Chorin, etc.
Ensemble Kalman Filter

What is ensemble Kalman filter? Approximate the prior mean and covariance using an ensemble

- Computationally cheap
- Low dimensional ensemble state approximation for extremely high dimensional turbulent dynamical systems
- Sampling errors and model errors
- Covariance and localization
Catastrophic Filter Divergence

- Observations are typically sparse and infrequent as in oceanography.
- Ensemble filtering methods can suffer from catastrophic filter divergence with sparse and infrequent observations and small observation errors.
- Catastrophic filter divergence drives the filter prediction to machine infinity although the underlying system remains in a bounded set.

Occurrence of catastrophic filter divergence

- EAKF for two-layer QG equation
- Snapshots of posterior upper layer stream function by low-latitude ocean code
- Observation points are marked with circles
- Catastrophic filter divergence is invoked after the 600-th cycle
Data assimilation and Non-Gaussian statistics

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

\[ p(x) = \sum_{k=1}^{K} w_k \delta(x - x_k) \]

where \( w_k \geq 0 \) and \( \sum_{k=1}^{K} w_k = 1 \).
Direct particle filter algorithm

Prior distribution is given by

\[ p^-(x) = \sum_{k=1}^{K} w_k^- \delta(x - x_k) \]

where \( w_k^- \geq 0 \) and \( \sum_{k=1}^{K} w_k^- = 1 \).

Observation data:

\[ v_{m+1} = g(u_{m+1}) + \sigma_{m+1}^0 \]

- posterior weights updated by

\[ w_{k,+} = \frac{p(v \mid u_k) w_{k,-}}{\sum_k p(v \mid u_k) w_{k,-}}, \quad p^+(u) = \sum_k w_{k,+} \delta(u - u_k) \]

- resample according to the effective ensemble size

\[ N_{\text{eff}} = \frac{1}{\sum_{k=1}^{K} w_k^2} \]
Example: Lorenz 96 system

The Lorenz 96 model was introduced to mimic the large-scale behavior of the mid-latitude atmosphere around a circle of constant latitude.

\[
\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) - u_i + F_i, \quad i = 0, 1, \ldots, J - 1. \tag{1}
\]

with \(J = 40\) the number of grids and \(F_i\) the deterministic forcing. The quadratic part conserves energy, that is, \(B(\mathbf{u}, \mathbf{u}) \cdot \mathbf{u} = 0\).

<table>
<thead>
<tr>
<th>(F)</th>
<th>Linear Analysis</th>
<th>Mean Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(k_{\text{begin}})</td>
<td>(k_{\text{end}})</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>
Figure: Numerical solutions of L-96 model in space-time through MC simulations for weakly chaotic ($F = 5$), strongly chaotic ($F = 8$), and fully turbulent ($F = 16$) regime.
Filter Performance with Sparse Infrequent High Quality Observations $F = 8, r_0 = 0.01, \Delta t = 0.25, p = 4$

Figure from Majda & Harlim book: *Filtering complex turbulent systems.*

Figure 15.14: RMS errors as functions of time: $\Delta t = 0.25, r^o = 0.01, K = 100, P = 4, L = 3$. 
Limitation of particle filters

- Not applicable to high-dimensional systems: difficult to sample a wide range of spatiotemporal scales
- 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 uncorrelated
Localized particle filter (Poterjoy, MWR)

Implements localization for particle filters and uses vector valued particle weights.

- Successful results for frequent and short observations

However

- Very complicated algorithm mainly due to several additional steps to stabilize the filter
- Frequent resampling steps
- Not robust for sparse and infrequent observations, which are typical in oceanography
Clustered Particle Filters (CPF), Lee 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
Schematics of several particle filters

**Standard Particle Filter**

![Standard Particle Filter]

**Localized Particle Filter**

![Localized Particle Filter]

**Clustered Particle Filter**

![Clustered Particle Filter]

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

- Total dimension is 6 and two observations at $x_2$ and $x_5$
- 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
Particle Adjustment

- The mean of $p(x) = \sum_{k}^{K} w_k \delta(x - x_k)$, $w_k \leq 0, \sum_{k} w_k = 1$ is a convex combination of $x_k, w_k x_k$
- If the observation cannot be represented by a convex combination of the prior particles, the posterior mean is never close to the observation ($\because$ particle filtering updates only the particle weights)

Adjust the prior particles to match the Kalman posterior mean and covariance if the prior particles cannot represent the observation

$$y_j \notin \left\{ \sum_{k}^{K} q_k [x_{C_j,k}^f], \forall q_k \geq 0 \text{ such that } \sum_{k} q_k = 1 \right\}$$

$y_j$ : $j$-th observation component, $x_{C_j,k}^f$ : 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 the method of EAKF by Anderson.
Hard Threshold Clustered Particle Filter Algorithm - one step assimilation

Given:
1) $N_{obs}$ observations \{\(y_1, y_2, \ldots, y_{N_{obs}}\)\}
2) prior $K$ particles \(\{x_{C_j,k}^f, k = 1, 2, \ldots, K\}\) and weight vectors \(\{\omega_{l,k}, k = 1, 2, \ldots, K\}\) for each cluster $C_l, l = 1, 2, \ldots, N_{obs}$

For $y_j$ from $j = 1$ to $N_{obs}$
   If $y_j \notin \{\sum_k q_k H[x_{C_j,k}^f], \forall q_k \geq 0 \text{ such that } \sum_k q_k = 1\}$
      Do particle adjustment
   Else Use particle filtering
      Update \(\{\omega_{j,k}\}\) using standard PF update
      If $K_{eff} = \frac{1}{\sum_k (\omega_{l,k}^a)^2} < \frac{K}{2}$
         Do resampling
         Add additional noise to the resampled particles
         \[
         x_{C_l, \text{Resample}(k)} \leftarrow x_{C_l, \text{Resample}(k)} + \epsilon
         \] (2)
         where $\epsilon$ is IID Gaussian noise with zero mean and variance $r_{noise}$
      End If
   End If
End For
Test model - Lorenz-96

Standard test model for data assimilation: 40-dimensional Lorenz-96

\[
\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i = 1, 2, \ldots, J = 40. \tag{3}
\]

- mimics baroclinic turbulence in the midlatitude atmosphere
- energy conserving nonlinear advection
Experiment setup for L96 with F=8, standard test regime

- Test three different methods - CPF, EAKF and Localized PF
- Use 50 samples
- Linear observation,
  \[ y = x + \xi, \quad \xi : \text{observation error} \]
- 20 regularly spaced observations
- Observation is of high quality, observation error is less than 1% of the total variance
Filter test results for L96 with \( F=8 \)

Time series of the RMS errors of EAKF, CPF, and Localized PF

Figure: Dash-dot line: observation error 0.22. Dash line: the climatological error 3.64.

- Localized PF has no filtering skill
- EAKF and CPF show accurate filter skill
- Note: EAKF requires tuning parameters to achieve robust results; no tuning for CPF
Time series of the RMS errors of CPF with and without particle adjustment

- On average, the particle adjustment is triggered only 30% of all assimilation steps
- The effect of the particle adjustment is significant
Experiment setup for L96 with F=5, strongly non-Gaussian test regime

- Use 50 samples
- **Linear** and **Nonlinear** observation,

\[ y = h(x) + \xi, \quad \xi : \text{observation error} \]

\[ h(x) = x \text{ or } \log(|x|) \]

- Observation error variance 0.5
- Randomly chosen **irregular** observation network

**Figure:** Dots : observation points, Vertical lines : cluster boundaries
Filter test results for L96 with F=5

Time series of the RMS errors of EAKF and CPF

- **Linear obs (top)**: CPF has smaller RMS errors than EAKF
- **Nonlinear obs (bottom)**: CPF shows filtering skill while EAKF has no skill

**Note** No tuning for CPF while EAKF uses tuning of inflation and localization
Forecast and forecast error PDFs

- $x^f$: forecast mean of $x$
- $\hat{x}_7$ and $\hat{x}_8$: the two most energetic Fourier modes

- Forecast PDF (top): the PDFs of CPF are on top of the true signal. EAKF does not capture the true signal PDFs
- Forecast error PDF (bottom): CPF has sharp peaks than EAKF
Multiscale Clustered Particle Filtering

- Multiscale data assimilation (particle filter, ensemble filter)
  

- Probability distribution: conditional Gaussian mixture

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

- 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
MMT model: wave turbulence

\[ i \partial_t \psi = |\partial_x|^{1/2} \psi - |\psi|^2 \psi + iF + iD\psi \]  (4)

in a periodic domain of length $L$ with large-scale forcing set to $F = 0.0163 \sin(4\pi x/L)$ and dissipation $D$ for both the large and small scales.

- shallow energy spectrum $k^{-5/6}$
- inverse cascade of energy from small to large scales
- non-Gaussian extreme event statistics caused by intermittent instability and breaking of solitons
- small scales carry more than two-thirds of the total variance

Visualization of $|\psi(x, t)|$ from simulation with $F_0 = 0.0163$; darker colors indicate higher amplitudes. Here the number of Fourier modes are $64^2 \approx 4000$.

Reference and stochastic superparameterization (SP) results

Reference uses 8192 grid points while stochastic SP uses only 128 grid points

Left: Time-averaged kinetic energy by reference (solid line), stochastic superparameterization (dash line), unparameterized model (dots)

Middle and Right: Time series of $|\psi|$ of the reference (middle) and stochastic superparameterization (right)
Filtering results of the MMT model

Time series of the large-scale RMS errors; 64 observations

Dash line: climatological error 0.20, Dash-dot line: effective observation error 0.34

Ensemble-based multiscale filtering

Clustered multiscale particle filtering

Forecast PDF and forecast error PDF of the large-scale real part

- Superior performance of CPF
Summary

We proposed the clustered particle filter

- 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
- Applied to Lorenz 96 and wave turbulence (multiscale data assimilation)
- Accurate filter performance

Future works

- Dense and vector observations
- Two- and three-dimensional spaces