Particle Filters and Finite Ensemble Kalman Filters in Large Dimensions: Theory, Applied Practice, and New Phenomena

Andrew J. Majda

Morse Professor of Arts and Science
Department of Mathematics
Founder of Center for Atmosphere Ocean Science (CAOS)
Courant Institute, New York University
and
Principal Investigator, Center for Prototype Climate Modeling (CPCM)
NYU – Abu Dhabi (NYUAD)

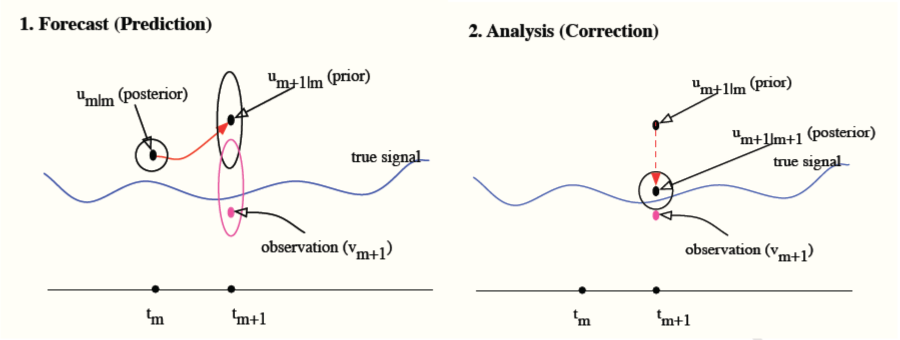
CAOS colloquium
September 12, 2018

Outline

- Introduction: Kalman Filters and Finite Ensemble Kalman Filters for Complex Systems
- Applied Practice, New Phenomena, and Math Theory for Finite Ensemble Kalman Filters
- Preventing Catastrophic Filter Divergence using Adaptive Additive Inflation
- State Estimation and Prediction using Clustered Particle Filters
- High Dimensional Challenges for Analyzing Finite Ensemble Kalman Filters
- Robustness and Accuracy of Finite Ensemble Kalman Filters in Large Dimensions

Outline

Filtering the Turbulent Signals[1] [2]



Filtering is a two-step process involving statistical prediction of the state variables through a forward operator followed by an analysis step at the next observation time which corrects this prediction on the basis of the statistical input of noisy observations of the system.

[1] Majda & Harlim, book *Cambridge University Press*, 2012;
[2] Majda, Harlim, & Gershgorin, review article, *Discrete Contin. Dyn. Syst*, 2010.

Filtering

- Signal-observation system with random coefficients

$$\text{Signal:} \quad X_{n+1} = A_n X_n + B_n + \xi_{n+1}, \quad \xi_{n+1} \sim \mathcal{N}(0, \Sigma_n)$$
$$\text{Observation:} \quad Y_{n+1} = H_n X_{n+1} + \zeta_{n+1}, \quad \zeta_{n+1} \sim \mathcal{N}(0, I_q)$$

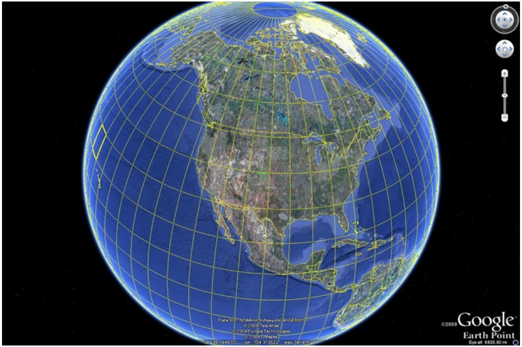
  Goal: estimate X_n based on $Y_1, \ldots, Y_n$

- A_n, B_n, H_n (stationary) sequence of random matrices and vectors.

- A_n can be unstable sometimes. H_n can be on and off.

Weather forecast

- Signal: $X_{n+1} = A_n X_n + B_n + \xi_{n+1}$,
  Observation: $Y_{n+1} = H X_{n+1} + \zeta_{n+1}$.
- Weather forecast:
  Signal: atmosphere and ocean “follow" a PDE.
  Obs: weather station, satellite, sensors, ...
- Main challenge: high dimension, $d \sim 10^6 - 10^8$.



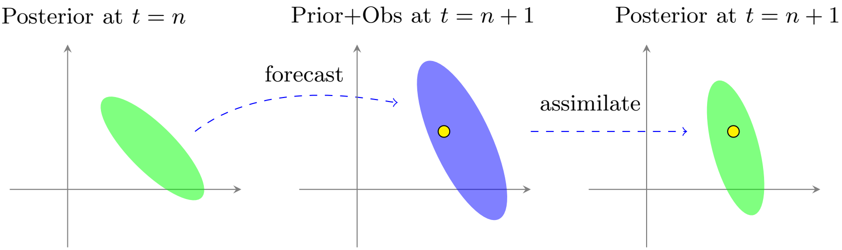
Kalman filter

- Use Gaussian: $X_n|_{Y_{1\dots n}} \sim \mathcal{N}(m_n, R_n)$
- Forecast step: $\hat{m}_{n+1} = A_n m_n + B_n, \widehat{R}_{n+1} = A_n R_n A_n^T + \Sigma_n$.
- Assimilation step: apply the Kalman update rule

$$m_{n+1} = \hat{m}_{n+1} + \mathcal{G}(\widehat{R}_{n+1})(Y_{n+1} - H_n \hat{m}_{n+1}), \quad R_{n+1} = \mathcal{K}(\widehat{R}_{n+1})$$

$$\mathcal{G}(C) = CH_n^T(I_q + H_n C H_n^T)^{-1}, \quad \mathcal{K}(C) = C - \mathcal{G}(C)H_n C$$

- Complexity: $O(d^3)$.



Random Kalman filters

Random Kalman Filters are very useful as computationally cheap filtering for forecast and filtering models (*Majda & Harlim book, 2012*). SPEKF, DSS (Branicki & Majda, 2014, 2018)

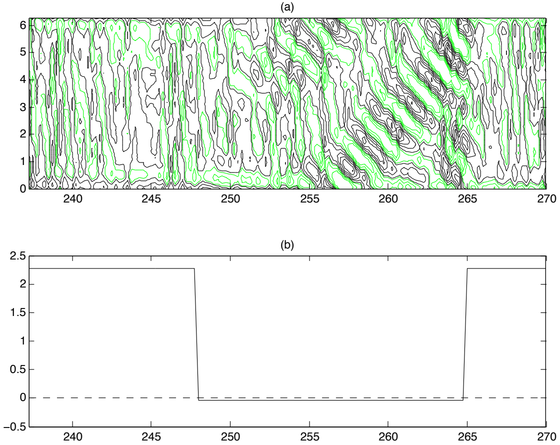


Figure: Spatial pattern for a turbulent system of the externally forced barotropic Rossby wave equation with instability through intermittent negative damping. Note the coherent wave train that emerges during the unstable regime.

Sampling+Gaussian

- Monte Carlo: use samples to represent a distribution:

$$X^{(1)}, \dots, X^{(K)} \sim p, \quad \frac{1}{K}\sum_{k=1}^{K} \delta_{X^{(k)}} \approx p.$$

- Ensemble $\{X_n^{(k)}\}_{k=1}^{K}$ to represent $\mathcal{N}(\overline{X}_n, C_n)$

$$\overline{X}_n = \frac{\sum X_n^{(k)}}{K}, \quad S_n = [\Delta X_n^{(1)}, \cdots, \Delta X_n^{(K)}], \quad C_n = \frac{S_n S_n^T}{K-1}.$$

or through ensemble spread $C_n = \frac{S_n S_n^T}{K-1}$

$$S_n = [X_n^{(1)} - \overline{X}_n, \cdots, X_n^{(K)} - \overline{X}_n],$$

Ensemble Kalman filter (EnKF)

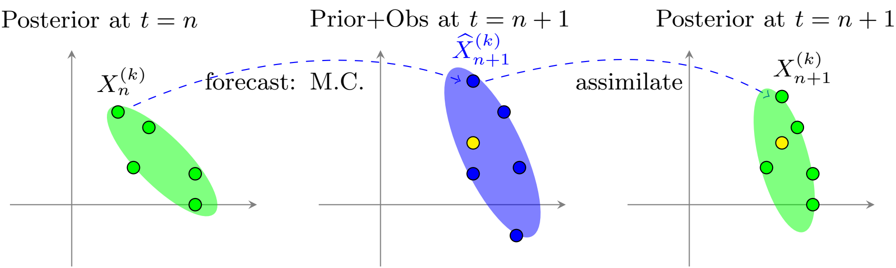
- Forecast step

$$\widehat{X}^{(k)}_{n+1} = A_n X^{(k)}_n + B_n + \zeta^{(k)}_{n+1}, \quad \widehat{C}_{n+1} = \frac{\widehat{S}_{n+1}\widehat{S}^T_{n+1}}{K-1}$$

- EAKF assimilation step, find $S_{n+1} = \mathbf{A}_{n+1}\widehat{S}_{n+1}$

$$\overline{X}_{n+1} = \overline{\widehat{X}}_{n+1} + \mathcal{G}(\widehat{C}_{n+1})(Y_{n+1} - H_n\overline{\widehat{X}}_{n+1}), \quad C_{n+1} = \mathcal{K}(\widehat{C}_{n+1})$$

- Complexity: $O(dK^2)$.



Outline

Application of Finite Ensemble KFs

Application:

- Successful weather forecast and oil reservoir management.
- Recently been applied to deep neural networks.
- $K = 50$ ensembles can forecast $d = 10^6$ dimensional systems.
- Extreme saving: $10^{10} = dK^2 \ll d^3 = 10^{18}$.
- Localization & inflation for successful filtering: needed practically to avoid filter divergence

Major contributions (introduced by geoscientists):

- G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics.* J Geophys Res Oceans, 1994.
- P. L. Houtekamer, and H. L. Mitchell. *Data assimilation using an ensemble Kalman filter technique.* Mon. Weather Rev., 1998.
- Gaspari, G., and S. Cohn. *Construction of correlation functions in two and three dimensions.* Quart. J. Roy. Meteor. Soc., 1999. Quart. J. Roy. Meteor. Soc.,
- J. L. Anderson. *An ensemble adjustment Kalman filter for data assimilation.* Mon. Weather Rev., 2001.
- C. H. Bishop, B. J. Etherton, and S. J. Majumdar. *Adaptive sampling with the ensemble transform kalman filter.* Mon. Weather Rev., 2001.
- E. Kalnay. Atmospheric modeling, data assimilation, and predictability. Cambridge university press, 2003.

New Phenomena: Catastrophic Filter Divergence

Surprising pathological discovery in EnKF

For filtering forced dissipative system with absorbing ball property (such as L-96 model), EnKF can explode to machine infinity in finite time! (Harlim and Majda 2008; Gottwald and Majda, *NPG* 2013)

- 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

Rigorous math contirbutions

Well-posedness of EnKF: D Kelly, KJ Law, and A Stuart. *Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time.* Nonlinearity, 2014.

• D Kelly, A J Majda, X Tong, *Concrete Ensemble Kalman Filters with Rigorous Catastrophic Filter Divergence*, PNAS 112 (34), pp. 10589–10594, 2015.

Rigorous nonlinear stability for finite ensemble Kalman filter (EnKF)

(Xin Tong, Majda, Kelly, *Nonlinearity* 2015)

Filter divergence – a potential flaw for EnKF:

- Catastrophic filter divergence: the ensemble members diverging to infinity,
- Lack of stability: the ensemble members being trapped in locations far from the true process.

Finding practical conditions and modifications to rule out filter divergence with rigorous analysis:

- Ruling out catastrophic filter divergence by establishing an energy principle for the filter ensemble.
- Looking for energy principles inherited by the Kalman filtering scheme.
- Looking for modification schemes of EnKF that ensures an energy principle and preserving the original EnKF performance (Xin Tong, Majda, Kelly, *Comm. Math. Sci.*, 2015).
- Verifying the nonlinear stability of EnKF through geometric ergodicity.

Rigorous example of catastrophic divergence:

- For filtering a nonlinear map with absorbing ball property (Kelly, Majda, Xin Tong, *PNAS* 2015).

Outstanding problem: Why and when is there accuracy in mean for $M \leq N$?

Outline

Preventing catastrophic filter divergence using adaptive additive inflation

Main reference

- Y. Lee, A.J. Majda, and D. Qi, (2016) **Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence**, Monthly Weather Review, 145(2), pp. 669-682.

Catastrophic Filter Divergence and Adaptive Inflation

- D Kelly, A J Majda, X Tong, (2015) **Concrete Ensemble Kalman Filters with Rigorous Catastrophic Filter Divergence**, PNAS 112 (34), pp. 10589–10594
- X. Tong, A J Majda, D Kelly, (2016) **Nonlinear Stability of the Ensemble Kalman Filter with Adaptive Covariance Inflation**, Commun. Math. Sci., 14 (5), pp. 1283–1313

Goal :

- Demonstrate catastrophic filter divergence
- Idealized baroclinic turbulence model for the ocean: two-layer quasigeostrophic equation
- Adaptive inflation to prevent catastrophic filter divergence
- Comparison between different reduced-order forecast models; importance of accurate parameterization of small scales

Ensemble Kalman Filter and Catastrophic Filter Divergence (review)

Finite ensemble Kalman filter: Approximate the prior mean and covariance using an ensemble

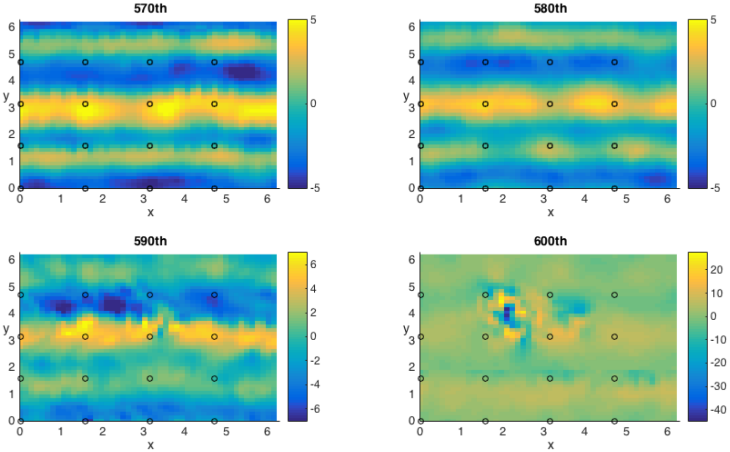
- Computationally cheap
- Low dimensional ensemble state approximation for extremely high dimensional turbulent dynamical systems
- Sampling errors and model errors
- Covariance and localization

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



Covariance Inflation

Inflate covariance, C^f, to overcome problems caused by

- sampling errors due to insufficient ensemble numbers
- model errors from an imperfect model

Covariance inflation gives more weight on the observation by introducing more uncertainty in the forecast.

Several inflation methods

$$\text{Additive inflation : } C^f \leftarrow C^f + \lambda I$$

$$\text{Multiplicative inflation : } C^f \leftarrow (I + \lambda) C^f$$

- Constant inflation improves filter skill in many applications
- However, it does not prevent catastrophic filter divergence with sparse and high quality observations

Adaptive Covariance Inflation

A simple remedy (Tong. et al, 2016) inflates the covariance adaptively

$$\lambda = c_a \Theta (1 + \Xi) \mathbf{1}_{\{\Theta > M_1 \text{ or } \Xi > M_2\}}$$

- Θ measures the innovation process

$$\Theta := \frac{1}{K} \sum_{k=1}^{K} \|H\tilde{v}^{(k)} - z\|^2, \quad \tilde{v}^k : \text{prior ensemble}, z : \text{observation}$$

- Ξ measures the cross-covariance between the observed and unobserved variables

$$\Xi = \left\| \frac{1}{K-1} \sum_{k=1}^{K} \left(\tilde{x}^{(k)} - \overline{x}\right) \left(\tilde{y}^{(k)} - \overline{y}\right)^T \right\|, \tilde{v}^{(k)} = (\tilde{x}^{(k)}, \tilde{y}^{(k)}), \tilde{x}^{(k)} = H\tilde{v}^{(k)}$$

  $\tilde{x}$: observed variable, $\tilde{y}$: unobserved variable

- M_1 and M_2 are fixed positive thresholds to decide whether the filter is performing well or not

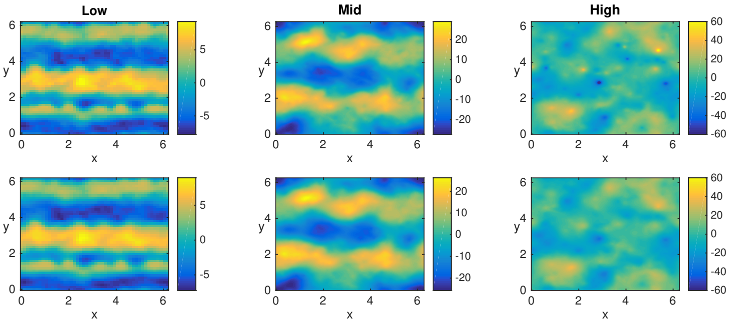
Two-layer quasigeostrophic equation

$$
\begin{aligned}
\partial_t q_1 &= -\mathbf{v}_1 \cdot \nabla q_1 - \partial_x q_1 - (k_\beta^2 + k_d^2) v_1 - \nu \Delta^4 q_1, \\
\partial_t q_2 &= -\mathbf{v}_2 \cdot \nabla q_2 + \partial_x q_2 - (k_\beta^2 - k_d^2) v_2 - r \Delta \psi_2 - \nu \Delta^4 q_2, \\
q_1 &= \Delta \psi_1 + \frac{k_d^2}{2}(\psi_2 - \psi_1), \quad q_2 = \Delta \psi_2 - \frac{k_d^2}{2}(\psi_2 - \psi_1)
\end{aligned}
$$

- q_j : potential vorticity in the upper ($j = 1$) and lower ($j = 2$) layers
- r : linear Ekman drag coefficient at the bottom layer
- ν : hyperviscosity
- k_d : deformation wavenumber
- k_β : nondimensionalized variation of the vertical projection of Coriolis frequency with latitude
- **There is net transfer of kinetic energy from small to large scales!**

Reference simulations

- Snapshots of upper layer stream functions
- 256×256 grid points for both layers
- Zonal jets under the β-plane effect



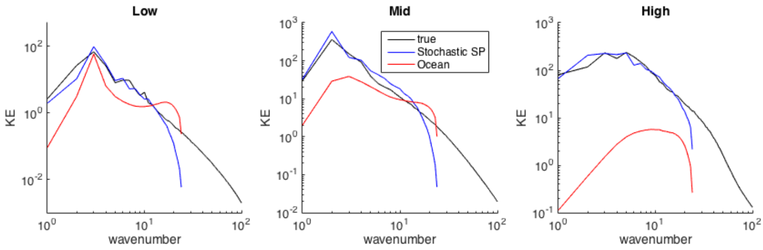
Forecast models

48×48 grid points for both layers (200 times cheaper than the full resolution model)

- Ocean code : uses only a coarse grid without parameterizing the small scales
- Stochastic superparameterization : parameterizes the effect of the small scales by modeling the small scales as randomly oriented plane waves; captures the inverse cascade of kinetic energy

  Majda and Grooms, JCP, 2014, Grooms, Lee and Majda, SIAM MMS, 2015

Time averaged kinetic energy spectra by direct numerical simulation (true), stochastic superparameterization (blue) and ocean code (red)



Numerical Experiments

- True signal : Full resolution; 256×256 grid points for each layer.
- Forecast : 48×48 grid points; 200 times cheaper than the full resolution
- Ensemble Adjustment Filter with 17 ensemble members
- Observes only the upper layer stream function
- 4×4 uniform observation network with an error variance less than 1% of the total variance;
- Infrequent observation interval 0.008 longer than the eddy turnover time 0.006.

Several inflation methods

We will consider the following four different inflation approaches

$$\lambda = c_c + c_a\Theta(1+\Xi)\mathbf{1}_{\{\Theta>M_1 \text{ or } \Xi>M_2\}}$$

- No inflation (noI) : $c_c = c_a = 0$
- Constant Inflation (CI) : $c_a = 0$
- Adaptive Inflation (AI) : $c_c = 0$
- Constant+Adaptive Inflation (CAI) : $c_c \neq 0, c_a \neq 0$

We also test covariance localization using the smooth localization function by Gaspari and Cohn and a localization radius 8 grid points.

Catastrophic filter divergence occurrence percentage out of 100 different runs

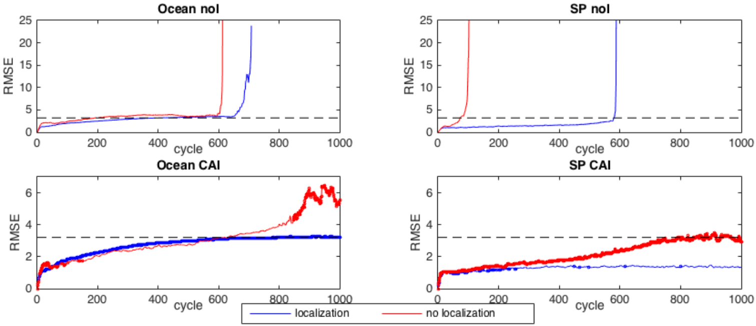
	Low		Mid		High	
no localization	Ocean	SP	Ocean	SP	Ocean	SP
noI	78%	84%	98%	97%	90%	85%
CI	63%	87%	80%	76%	45%	57%
AI	3%	0%	2%	0%	5%	0%
CAI	0%	0%	0%	0%	0%	0%

	Low		Mid		High	
with localization	Ocean	SP	Ocean	SP	Ocean	SP
noI	40%	24%	19%	38%	44%	64%
CI	15%	11%	9%	12%	22%	8%
AI	1%	0%	0%	0%	0%	0%
CAI	0%	0%	0%	0%	0%	0%

- No inflation and constant inflation do not prevent divergene
- Localization does not prevent divergence
- Adaptive inflation depends on the forecast model; Adaptive inflation with SP prevents divergence
- Constant inflation in addition to adaptive inflation is required to account for other errors, such as approximation in thresholding

Time series of the upper layer RMS errors (low latitude case)

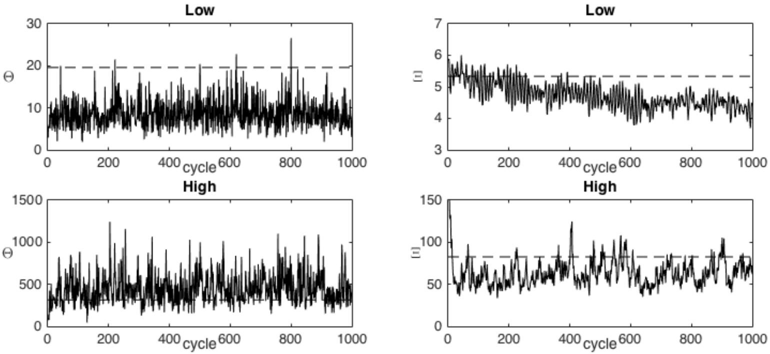
The cycles at which inflation is triggered are marked with filled circles. Dash-line : standard deviation of the stream function



- Constant + adaptive inflation stabilizes filters
- Filtering skill depends on the forecast model and localization
- High skill with Superparameterization

Time series of the two statistics, Θ and Ξ

- Θ : innovation
- Ξ : cross-covariance between the observed and unobserved variables).
- Dash-line : threshold values



- For the low latitude case, Ξ plays an important role in preventing catastrophic filter divergene

Summary of covariance inflation

- Catastrophic filter divergence can be invoked when observations are sparse, infrequent and of high quality
- Checked for baroclinic turbulence, nontrivial test model
- Constant inflation and localization do not prevent catastrophic filter divergence
- Adaptive inflation by Tong et al. can prevent catastrophic filter divergence; cross-correlation between observed and unobserved variables play important roles in addition to innovations
- Robust forecast models can affect the occurrence of catastrophic filter divergence in addition to accurate results
- High skill with Superparameterization

Outline

State estimation and prediction using clustered particle filters

Main reference

- Y. Lee and A.J. Majda, (2016) **State estimation and prediction using clustered particle filters**, PNAS, 113(51), 14609-14614.

Multiscale data assimilation

- Y. Lee and A.J. Majda, (2015) **Multiscale methods for data assimilation in turbulent systems**, SIAM Multiscale Modeling and Simulation, 13(2), 691–173.

Data assimilation and non-Gaussian statistics (review)

- 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 $\sum_k w_k = 1$.

Limitations of particle filters

- Not applicable to high-dimensional systems (Bengtsson T, Bickel P, Li B, 2008)
- 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

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

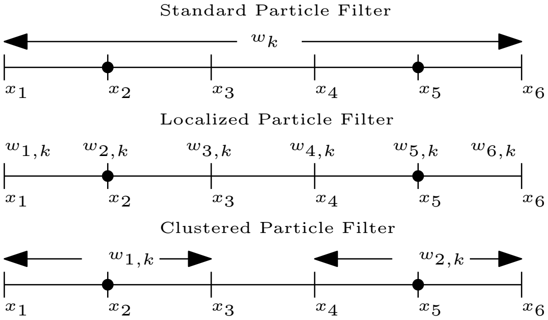


Figure: Schematics of particle weight, w_k, for the k-th particle.

- 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
- Weights are the same in the same cluster

Particle Adjustment

- 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 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

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

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 the method of 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

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

      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

MMT model : wave turbulence

$$i\partial_t\psi = |\partial_x|^{1/2}\psi - |\psi|^2\psi + iF + iD\psi \tag{2}$$

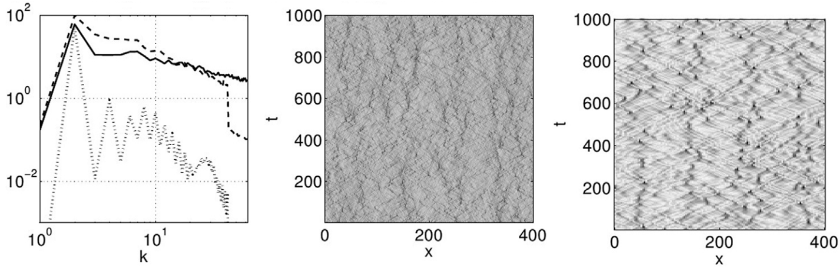
in a periodic domain of length L with large-scale forcing set to $F = 0.0163\sin(4\pi x/L)$ and dissipation D for both the large and small scales.

- 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

Majda, McLaughlin, Tabak, J. Nonlinear Sci., 1997
Cai, Majda, McLaughlin, Tabak, Physica D: Nonlinear Phenomena, 2001

Reference and stochastic superparameterization (SP) results

Reference uses 8192 grid points while stochastic SP uses only 128 grid points
(Majda and Grooms, JCP, 2014, Grooms and Majda, Commun. Math. Sci, 2014)
Left : Time-averaged kinetic energy by reference (solid line), stochastic superparameterization (dash line), unparameterized model (dots)

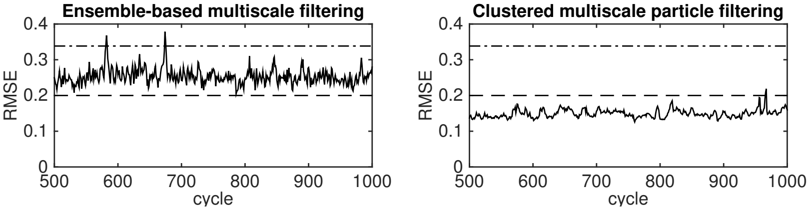


Middle and Right : Time series of $|\psi|$ of the reference (middle) and stochastic superparameterization (right)

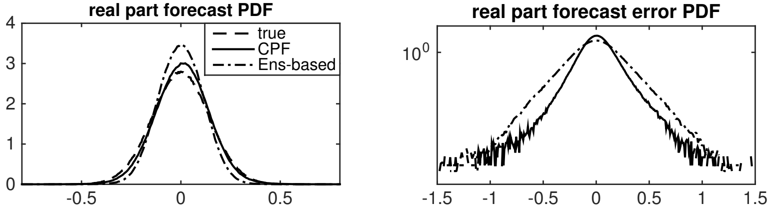
Filtering results of the MMT model

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

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



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



- Superior performance of CPF

Summary of clustered particle filters

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 work:

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

Outline

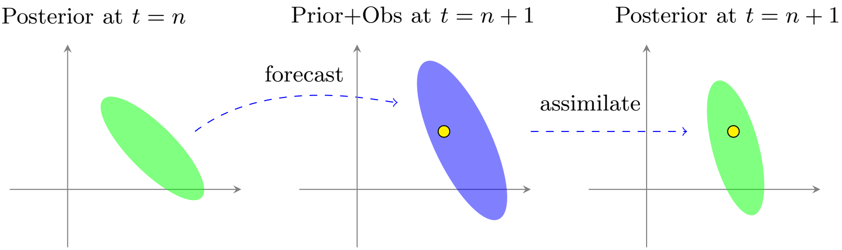
Review: Kalman filter

- Use Gaussian: $X_n|_{Y_{1\dots n}} \sim \mathcal{N}(m_n, R_n)$
- Forecast step: $\hat{m}_{n+1} = A_n m_n + B_n, \widehat{R}_{n+1} = A_n R_n A_n^T + \Sigma_n$.
- Assimilation step: apply the Kalman update rule

$$m_{n+1} = \hat{m}_{n+1} + \mathcal{G}(\widehat{R}_{n+1})(Y_{n+1} - H_n \hat{m}_{n+1}), \quad R_{n+1} = \mathcal{K}(\widehat{R}_{n+1})$$

$$\mathcal{G}(C) = CH_n^T (I_q + H_n C H_n^T)^{-1}, \quad \mathcal{K}(C) = C - \mathcal{G}(C) H_n C$$

- Complexity: $O(d^3)$.



Review: Gaussian Sampling and Ensemble Kalman filter (EnKF)

Sampling + Gaussian

- Monte Carlo: use samples to represent a distribution:

$$X^{(1)}, \dots, X^{(K)} \sim p, \quad \frac{1}{K}\sum_{k=1}^{K} \delta_{X^{(k)}} \approx p.$$

- Ensemble $\{X_n^{(k)}\}_{k=1}^{K}$ to represent $\mathcal{N}(\overline{X}_n, C_n)$

$$\overline{X}_n = \frac{\sum X_n^{(k)}}{K}, \quad S_n = [\Delta X_n^{(1)}, \cdots, \Delta X_n^{(K)}], \quad C_n = \frac{S_n S_n^T}{K-1}.$$

Finite EnKF

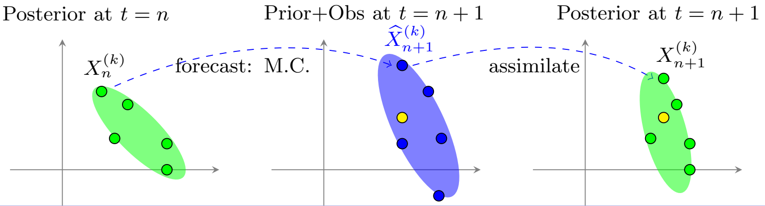
- Forecast step

$$\widehat{X}_{n+1}^{(k)} = A_n X_n^{(k)} + B_n + \zeta_{n+1}^{(k)}, \quad \widehat{C}_{n+1} = \frac{\widehat{S}_{n+1}\widehat{S}_{n+1}^T}{K-1}$$

- EAKF assimilation step, find $S_{n+1} = \mathbf{A}_{n+1}\widehat{S}_{n+1}$

$$\overline{X}_{n+1} = \overline{\widehat{X}}_{n+1} + \mathcal{G}(\widehat{C}_{n+1})(Y_{n+1} - H_n\overline{\widehat{X}}_{n+1}), \quad C_{n+1} = \mathcal{K}(\widehat{C}_{n+1})$$

- Complexity: $O(dK^2)$.



Challenges for analyzing finite ensemble Kalman filters

Ensemble size K to represent uncertainty of dimension d:

- Rank deficiency: $C_n = \frac{\sum_{k=1}^{K}(X_n^{(k)}-\overline{X}_n)(X_n^{(k)}-\overline{X}_n)^T}{K-1}$
  Has rank(C_n)$\leq K-1$, see as $\begin{bmatrix} C_n & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} \}\text{ K-1} \\ \}\text{ d-K+1} \end{matrix}$
- Instability of the dynamics: $\widehat{C}_{n+1} = A_n C_n A_n^T + \Sigma_n$
  What if span(C_n) does not cover expanding directions?
- Covariance decay by random sampling: $C_{n+1} = \mathcal{K}(\widehat{C}_{n+1})$
  $\mathcal{K}$: concave, monotone: $\mathbb{E}C_{n+1} = \mathbb{E}\mathcal{K}(\widehat{C}_{n+1}) \preceq \mathcal{K}(\mathbb{E}\widehat{C}_{n+1})$
- Spurious correlation in high dimension.
  Suppose $X_n^{(k)} \sim \mathcal{N}(0, I_d)$ i.i.d, by Bai-Yin's law

$$\|C_n - I_d\| \approx \sqrt{d/K} \quad \text{with large probability}$$

Outline

Robustness and Accuracy of Finite Ensemble Kalman Filters in Large Dimensions

Main references

- A. J. Majda and X. T. Tong, (2018). **Performance of Ensemble Kalman filters in large dimensions.** CPAM, 71(5), 892-937.
- A. J. Majda and X. T. Tong, (2017). **Rigorous accuracy and robustness analysis for two-scale reduced random Kalman filters in high dimensions.** accepted by CMS.

Main result

Theorem (Majda, Tong 16)

Suppose the system has a low effective filter dimension p, there is a variant of EnKF with a constant C, such that the EnKF reaches its proclaimed performance if $K > Cp$.

And the convergence rate applies for any finite ensemble with $K > Cp$ at the rate dK^2 which is the expected operational rate.

Next, we explain

- What variant of EnKF?
- How to define a low effective dimension?
- What does proclaimed performance mean?

EnKF variant for enhanced fidelity

Techniques to enhance fidelity

- Rank deficiency: additive inflation ρI_d

$$C_n^\rho = \rho I_d + \frac{\sum_{k=1}^K (X_n^{(k)} - \overline{X}_n)(X_n^{(k)} - \overline{X}_n)^T}{K-1} = \begin{bmatrix} C_n + \rho I_{K-1} & 0 \\ 0 & \rho I_{d-K+1} \end{bmatrix}$$

  The under represented direction: assume error strength is ρ.

- Instability of the dynamics. Increase noise strength

$$\widehat{X}_{n+1}^k = A_{n+1} X_n^{(k)} + \xi_{n+1}^{(k)}, \quad \xi_{n+1}^{(k)} \sim \Sigma_n^+$$
$$\Sigma_n \to \Sigma_n^+ = [\rho A_n A_n^T + \Sigma_n - \rho / r I_d],$$

  Σ_n^+ indicates the system instability.

- Covariance decay by random sampling.
  Multiplicative inflation: $\widehat{C}_{n+1} = r\widehat{C}_{n+1}$

- Spurious correlation in high dimension.
  Projecting to p principal directions of $\mathcal{K}(\widehat{C}_{n+1})$
  The leftover direction: assume error strength is ρ

Main theorem

Theorem (Majda, Tong 16)

Suppose the signal observation system is uniformly observable with m steps, and has a effective dimension p. Then for any c, there are C, F, D_F, M_n, so that if $K > Cp$

$$\mathbb{E}\|e_n\|_{C_n^\rho} \leq r^{-\frac{n}{6}}\mathbb{E}F(C_0)\sqrt{\|e_0\|^2_{C_0} + 2m} + M_n\sqrt{d}$$

With the constants bounded by

$$F(C) \leq D_F \exp(D_F \log^3 \|C\|), \quad \limsup_{n\to\infty} M_n \leq \frac{1+c}{1-r^{-\frac{m}{6}}}.$$

Cor1: exponential stability: the difference in mean converges to zero.
Cor2: ϵ scale noises lead to ϵ scale error for EnKF.

Thank you