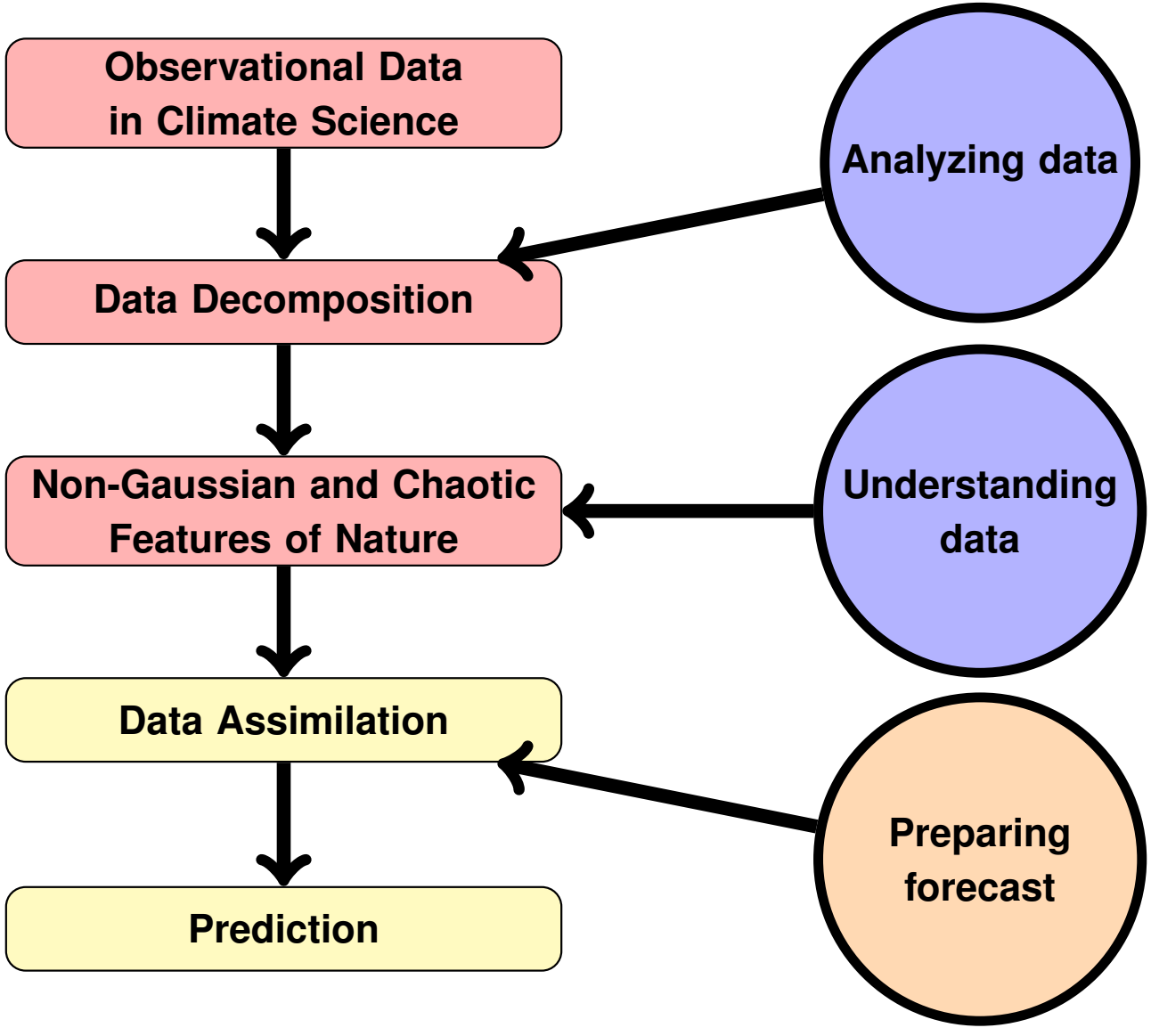
Nonlinear Data Assimilation Methods

Nan Chen

Department of Mathematics
University of Wisconsin-Madison

Math 717, Spring 2024


Observational Data
in Climate Science
Data Decomposition
Non-Gaussian and Chaotic
Features of Nature
Data Assimilation
Prediction
Analyzing data
Understanding
data
Preparing
forecast

Outline

- Notations and review of Kalman filter
- Ensemble smoother (ES)
- Ensemble Kalman smoother and filter (ENKS/ENKF)
- Particle filter

Notations and review of Kalman filter

Notations:

state variable: ψ, measurement: d, error cov from obs: $C_{\epsilon\epsilon}$,

ensemble forecast error cov: $C_{\psi\psi}$, observational operator: $\mathcal{M}$, $r = \mathcal{M}[C^f_{\psi\psi}]$

Kalman update formula:

$$\psi^a = \psi^f + r^T(\mathcal{M}^T[r] + C_{\epsilon\epsilon})^{-1}(d - \mathcal{M}[\psi^f])$$
$$C^a_{\psi\psi} = C^f_{\psi\psi} - r^T(\mathcal{M}^T[r] + C_{\epsilon\epsilon})^{-1} r$$

with Kalman gain

$$K = \mathcal{M}^T[C^f_{\psi\psi}](\mathcal{M}^T[\mathcal{M}[C^f_{\psi\psi}]] + C_{\epsilon\epsilon})^{-1} = r^T(M^T[r] + C_{\epsilon\epsilon})^{-1}$$

In 1D linear Gaussian case:

Model: $u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1}$,

Obs: $v_{m+1} = gu_{m+1} + \sigma^o_{m+1}$.

Posterior update:

$$\bar{u}_{m+1|m+1} = (1 - K_{m+1}g)\bar{u}_{m+1|m} + K_{m+1}v_{m+1}.$$
$$r_{m+1|m+1} = (1 - K_{m+1}g)r_{m+1|m}.$$

with: $K_{m+1} = \dfrac{g r_{m+1|m}}{r^o + g^2 r_{m+1|m}}$

Notations:

state variable: ψ, measurement: d, error cov from obs: $C_{\epsilon\epsilon}$,

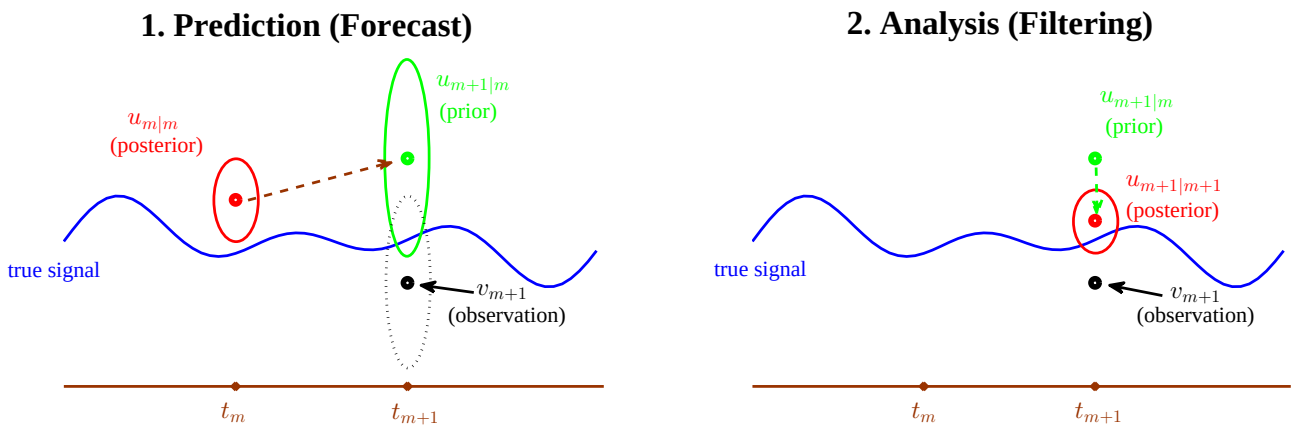
ensemble forecast error cov: $C_{\psi\psi}$, observational operator: $\mathcal{M}$, $r = \mathcal{M}[C^f_{\psi\psi}]$

Kalman update formula:

$$\psi^a = \psi^f + r^T(\mathcal{M}^T[r] + C_{\epsilon\epsilon})^{-1}(d - \mathcal{M}[\psi^f])$$
$$C^a_{\psi\psi} = C^f_{\psi\psi} - r^T(\mathcal{M}^T[r] + C_{\epsilon\epsilon})^{-1} r$$

with Kalman gain

$$K = \mathcal{M}^T[C^f_{\psi\psi}](\mathcal{M}^T[\mathcal{M}[C^f_{\psi\psi}]] + C_{\epsilon\epsilon})^{-1} = r^T(M^T[r] + C_{\epsilon\epsilon})^{-1}$$



Now instead of solving the mean and covariance analytically, we use Monte Carlo simulation (known as ensembles) to approximate these statistics.

Ensemble representation of error statistics

Notations:

state variable: ψ, measurement: d, error cov from obs: $C_{\epsilon\epsilon}$,

ensemble forecast error cov: $C_{\psi\psi}$, observational operator: $\mathcal{M}$, $r = \mathcal{M}[C^f_{\psi\psi}]$

The ensemble covariance is defined as

$$C_{\psi\psi} = \overline{(\psi - \overline{\psi})(\psi - \overline{\psi})^T}$$

Define a matrix

$$A(x, t_i) \in \mathbb{R}^{n\times N}$$

where n is the dimension of the state variables while N is the number of ensembles. We denote

$$A_i = A(x, t_i) = (\psi^1(x, t_i), \psi^2(x, t_i), \ldots, \psi^N(x, t_i))$$

The ensemble mean is stored in each column of $\overline{A}(x, t_i)$,

$$\overline{A}(x, t_i) = A(x, t_i)1_N, \qquad 1_N = \begin{pmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \ddots & \vdots \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

Define the ensemble perturbation matrix

$$A'(x, t_i) = A(x, t_i) - \overline{A}(x, t_i) = A(x, t_i)(I - 1_N)$$

The ensemble covariances $C^e_{\psi\psi}(x_1, x_2, t_i) \in \mathbb{R}^{n \times n}$ can be defined as

$$C^e_{\psi\psi}(x_1, x_2, t_i) = \frac{A'(x, t_i)(A'(x, t_i))^T}{N - 1}.$$

Define the ensemble matrix for the joint state from t_0 to t_i,

$$\widetilde{A}_i = \begin{pmatrix} A(x, t_0) \\ \vdots \\ A(x, t_i) \end{pmatrix}$$

The space-time ensemble covariance at time instants t_1 and t_2 then becomes

$$\widetilde{C}^e_{\psi\psi}(x_1, t_1, x_2, t_2) = \frac{A'(x_1, t_2)(A'(x_2, t_2))^T}{N - 1}.$$

Ensemble representation for measurements

At the data time $t_{i(j)}$, we have given a vector of measurements $d_j \in \mathbb{R}^{m_j}$, with m_j being the number of measurements at this time. We can define the N vectors of perturbed measurements as

$$d_j^l = d_j + \epsilon_j^l, \qquad l = 1, \ldots, N$$

which can be stored in the columns of a matrix

$$D_j = (d_j^1, d_j^2, \ldots, d_j^N) \in \mathbb{R}^{m_j \times N}$$

The ensemble of measurement perturbations is given by

$$E_j = (\epsilon_j^1, \epsilon_j^2, \ldots, \epsilon_j^N) \in \mathbb{R}^{m_j \times N}$$

from which we can construct the ensemble representation of the measurement error covariance matrix

$$C_{\epsilon\epsilon}^e(t_{i(j)}) = \frac{E_j E_j^T}{N-1}$$

Ensemble smoother (ES)

Given the preparation above, the ensemble smoother (ES) is given by

$$\widetilde{A}_k^a = \widetilde{A}_k + \mathcal{M}^T[\widetilde{C}_{\psi\psi}^e](\mathcal{M}^T[\mathcal{M}[\widetilde{C}_{\psi\psi}^e]] + C_{\epsilon\epsilon}^e)^{-1}(D - \mathcal{M}[\widetilde{A}_k])$$

where

$$D = \begin{pmatrix} D_1 \\ \vdots \\ D_m \end{pmatrix}, \qquad \mathcal{M} = \begin{pmatrix} \mathcal{M}_1 \\ \vdots \\ \mathcal{M}_m \end{pmatrix}, \qquad C_{\epsilon\epsilon}^e = \begin{pmatrix} C_{\epsilon\epsilon}^e(t_{i(1)}) & & \\ & & \\ & & C_{\epsilon\epsilon}^e(t_{i(m)}) \end{pmatrix}$$

The total number of measurements is $M = \sum_{j=1}^m m_j$. We have $D \in \mathbb{R}^{M\times N}$, $\mathcal{M} \in \mathbb{R}^M$ and $C_{\epsilon\epsilon}^e \in \mathbb{R}^{M\times M}$.
Define the ensemble innovation vector

$$D' = D - \mathcal{M}[\widetilde{A}_k].$$

Define the measurements of the ensemble perturbations

$$S = \mathcal{M}[\widetilde{A}_k'] \in \mathbb{R}^{M\times N}.$$

Define a matrix C as

$$C = SS^T + (N-1)C_{\epsilon\epsilon}^e$$

With

$$D' = D - \mathcal{M}[\widetilde{A}_k], \qquad S = \mathcal{M}[\widetilde{A}'_k] \in \mathbb{R}^{M\times N}, \qquad C = SS^T + (N-1)C^e_{\epsilon\epsilon}$$

and

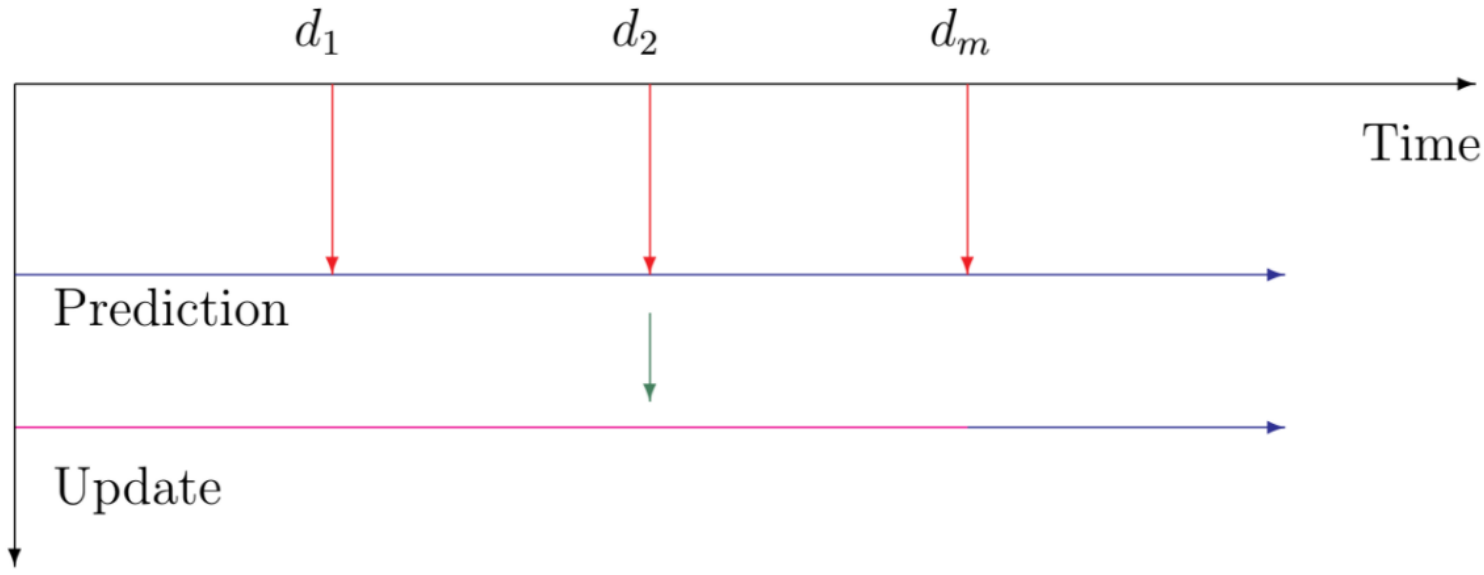
$$\widetilde{C}^e_{\psi\psi}(x_1,t_1,x_2,t_2) = \frac{A'(x_1,t_2)(A'(x_2,t_2))^T}{N-1}, \qquad C^e_{\epsilon\epsilon}(t_{i(j)}) = \frac{E_j E_j^T}{N-1}$$

in hand, the analysis can be expressed as

$$\begin{aligned}
\widetilde{A}^a_k &= \widetilde{A}_k + \widetilde{A}'_k\mathcal{M}^T[\widetilde{A}'_k]\left(\mathcal{M}[\widetilde{A}'_k]\mathcal{M}^T[\widetilde{A}'_k] + (N-1)C^e_{\epsilon\epsilon}\right)^{-1} D' \\
&= \widetilde{A}_k + \widetilde{A}_k(I - 1_N)S^T C^{-1} D' \\
&= \widetilde{A}_k(I + (I - 1_N)S^T C^{-1} D') \\
&= \widetilde{A}_k(I + S^T C^{-1} D') \\
&= \widetilde{A}_k X
\end{aligned}$$

where we have used the fact that $1_N S^T = 0$.

The above converges towards the exact solution of the Bayesian formulation with increasing ensemble size if the assumption of Gaussian statistics is true. This requires that all priors are Gaussian and that a linear model is used. In this linear case it will also converge towards the representer solution.
It can be illustrated that the ES may have problems with nonlinear dynamical models.


d_1
d_2
d_m
Time
Prediction
Update

Ensemble Kalman smoother and filter (ENKS/ENKF)

The joint pdf for the model prediction until $t_{i(1)}$ is

$$f(\psi_0, \psi_1, \ldots, \psi_{i(1)}) \propto f(\psi_0) \prod_{i=1}^{i(1)} f(\psi_i|\psi_{i-1})$$

The analysis based on the first observation d_1 is

$$f(\psi_0, \psi_1, \ldots, \psi_{i(1)}|d_1) \propto f(\psi_0, \psi_1, \ldots, \psi_{i(1)}) f(d_1|\psi_{i(1)})$$

Starting from A_0, a forward stochastic integration until the first available observational data gives the ensemble prediction

$$\widetilde{A}^f_{i(1)} = \begin{pmatrix} A_0 \\ A^f_1 \\ \vdots \\ A^f_{i(1)} \end{pmatrix}$$

Using the ES, we have

$$\begin{aligned} \widetilde{A}^a_{i(1)} &= \widetilde{A}^f_{i(1)} + \widetilde{A}^{f\prime}_{i(1)} \mathcal{M}_1^T [\widetilde{A}^{f\prime}_{i(1)}] \left(\mathcal{M}_1 [\widetilde{A}^{f\prime}_{i(1)} \mathcal{M}_1^T [\widetilde{A}^{f\prime}_{i(1)}] + (N-1) C^e_{\epsilon\epsilon}(t_{i(1)})] \right)^{-1} D'_1 \\ &= \widetilde{A}^f_{i(1)} + \widetilde{A}^f_{i(1)} (I - 1_N) S_1^T C_1^{-1} D'_1 \\ &= \widetilde{A}^f_{i(1)} \left(I + (I - 1_N) S_1^T C_1^{-1} D'_1 \right) = \widetilde{A}^f_{i(1)} \left(I + S_1^T C_1^{-1} D'_1 \right) = \widetilde{A}^f_{i(1)} X_1 \end{aligned}$$

where we have used the following definitions

$$D'_j = D_j - M_j[\widetilde{A}^f_{i(1)}], \qquad S_j = \widetilde{A}^{f\prime}_{i(1)}, \qquad C_j = S_j S_j^T + (N-1) C_{\epsilon\epsilon}(t_{i(j)})$$

In general:

$$f(\psi_0, \psi_1, \ldots, \psi_{i(j)} | d_1, \ldots, d_j) \propto f(\psi_0, \psi_1, \ldots, \psi_{i(j)} | d_1, \ldots, d_{j-1}) f(d_j | \psi_{i(j)})$$

Define the ensemble prediction matrix

$$\widetilde{A}^f_{i(j)} = \begin{pmatrix} \widetilde{A}^a_{i(j-1)} \\ A^f_{i(j-1)+1} \\ \vdots \\ A^f_{i(j)} \end{pmatrix}$$

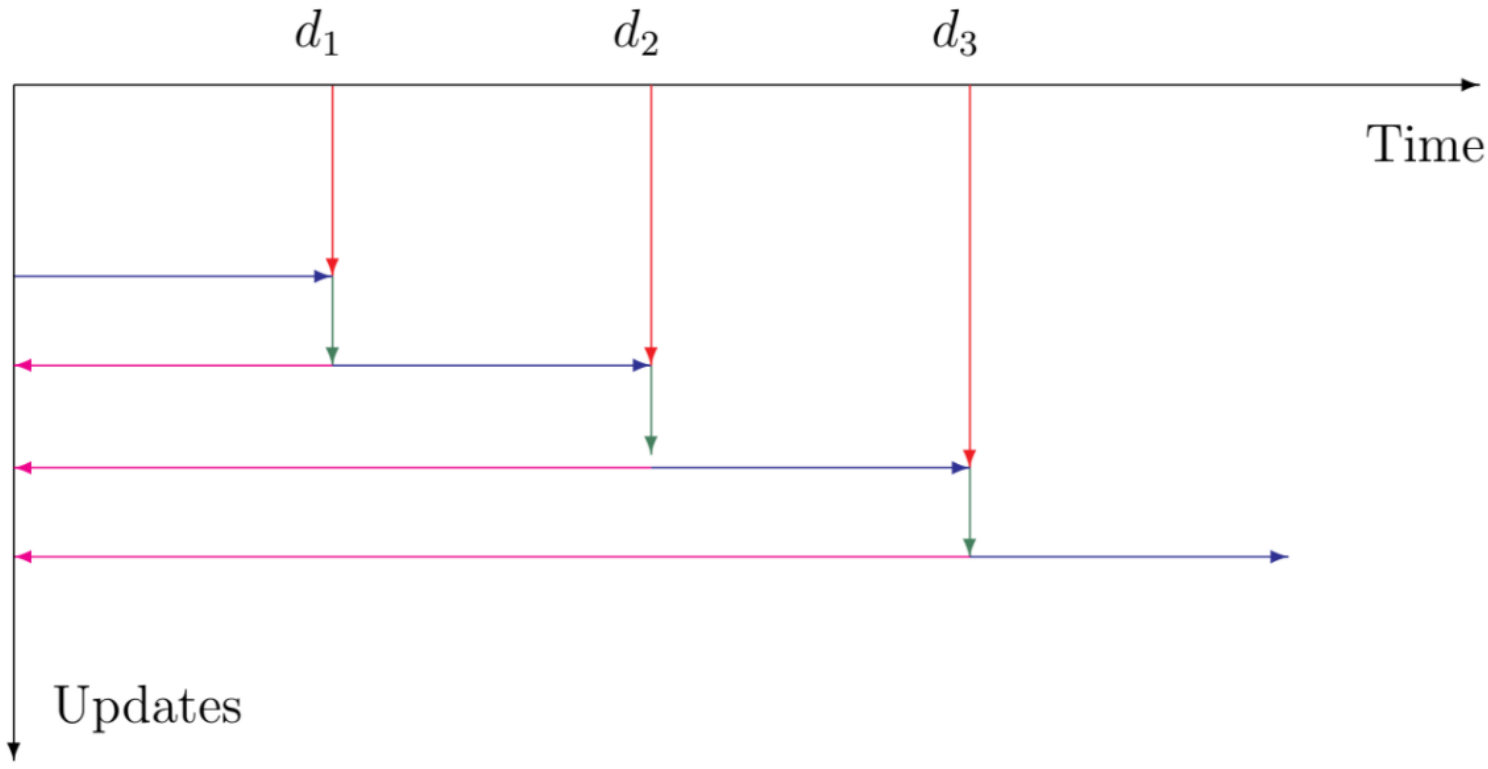
where the ensemble prediction $A^f_{i(j-1)}, \ldots, A^f_{i(j)}$ is obtained by ensemble integration starting from the final analyzed result in $\widetilde{A}^a_{i(j-1)}$. We can then compute the EnKS update based on using the measurements at time $t_{i(j)}$ as,

$$\widetilde{A}^a_{i(j)} = \widetilde{A}^f_{i(j)} X_j$$

where

$$X_j = I + S_j^T C_j^{-1} D_j'$$

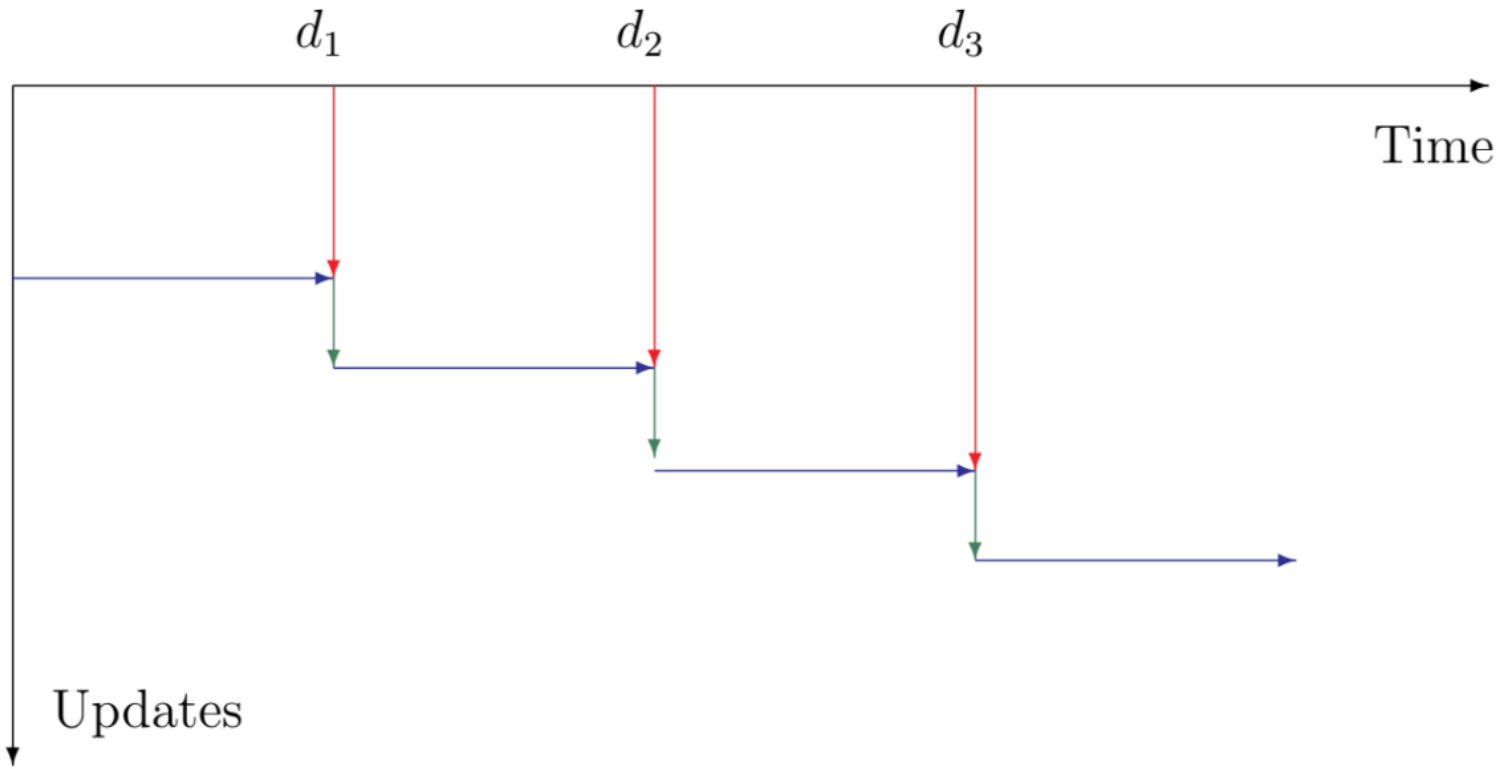
and $\widetilde{A}^f_{i(j)}$ has been updated from all previous measurements $d_1, \ldots, d_{j-1}$.


d_1
d_2
d_3
Time
Updates

The EnKF can be most easily characterized as a simplification of the EnKS where the analysis acts on the ensemble only at the measurement times. Thus, there is no information propagated backward in time like in the EnKS. We now only consider the analysis step at time $t_{i(j)}$, and the analysis equation is rewritten as

$$A^a_{i(j)} = A^f_{i(j)} X_j$$

where the ensembles at all prior times are discarded in the analysis.



Particle filter

The main idea in particle filtering is to apply the Bayes theorem to update an ensemble of solutions (or particles) without assuming any Gaussianity on the prior or posterior distributions as well as on the observation likelihood function. Therefore, no explicit formulation like Kalman filter equations is used.

Consider an ensemble posterior state $\{u|^k_{m|m} \in \mathbb{R}^N\}_{k=1,\dots,K}$ of size K at time t_m with distribution function

$$p_{m|m}(u) = \sum_{k=1}^{K} w_m^k \delta(u - u^k_{m|m})$$

where the term w_m^k determines the weight (or importance) of each ensemble member $u^k_{m|m}$ with $\sum_k w_m^k = 1$ and $w_m^k \geq 0$.

Notice that with this density representation, we commit sampling errors in approximating the posterior distribution. Such errors are unavoidable in a high dimensional problem and this is the so-called “curse of dimensionality” issue in random sampling related problems.

The goal of each step in particle filtering is to produce a posterior particles with a distribution $p_{m+1|m+1}$ at the future time t_{m+1} that accounts for the statistical information about observations

$$v_{m+1} = g(u_{m+1}) + \sigma^o_{m+1} \qquad \sigma^o_{m+1} \sim \mathcal{N}(0, r^o I)$$

where the observation operator g can be nonlinear.

The first step in particle filtering is to produce a prior distribution at the observation time, $p_{m+1|m}$. Theoretically, for a dynamical system governed by the following stochastic differential equations,

$$du = f(u,t)dt + \sigma(u,t)dW(t)$$

The exact prior density is a solution of the Fokker-Planck equation at finite time $t_{m+1} > t_m$ with initial condition $p(u, t_m) = p_{m|m}(u)$.

In practical particle filtering, one does not solve this Fokker-Planck equation since it is unfeasible for high-dimensional problems and we only have a finite sample of the initial distribution pm|m through particles $\{u^k_{m|m}\}_{k=1,\dots,K}$. Typically, one propagates these particles forward in time by solving the SDE with initial conditions $\{u^k_{m|m}\}_{k=1,\dots,K}$ to obtain a prior ensemble $\{u^k_{m+1|m}\}_{k=1,\dots,K}$ and assumes that each prior particle is equally important, i.e.,

$$p_{m+1|m}(u) = \frac{1}{K}\sum_{k=1}^{K}\delta(u - u_{m+1|m})$$

Given an observation likelihoood function $p(v_{m+1}|u)$, the posterior density is produced through applying Bayes formula as follows,

$$p_{m+1|m+1}(u) = p(u|v_{m+1}) = \frac{p(v_{m+1}|u)p_{m+1|m}(u)}{\int p(v_{m+1}|u)p_{m+1|m}(u)du} = \sum_{k=1}^{K} w^k_{m+1}\delta(u - u_{m+1|m})$$

where

$$w^k_{m+1} = \frac{p(v_{m+1}|u^k_{m+1|m})}{\sum_{k=1}^{K} p(v_{m+1}|u^k_{m+1|m})}$$

- Notice here that the Bayes formula simply re-weights the particles.
- The main issue in practice is that many particles tend to have low weights in high dimensional systems or in systems with a moderate dimensional chaotic attractor.
- If this two-step process is repeated over and over, then typically only one ensemble member will remain to have a large weight and the remaining become negligible.
- One way to avoid this ensemble collapse is to resample the particles by duplicating those particles with large weights to replace those with small weights. The advantage of this resampling (known as sequential important resampling or SIR) is that we do not modify the particles so each ensemble member is dynamically balanced, which can be applied to stochastic systems.