Data Assimilation in Multiscale Complex Systems: Lorenz 96

Nishanth Lingala, N. Sri Namachchivaya and Hoong Chieh Yeong

Departments of Aerospace Engineering & Information Trust Institute
University of Illinois at Urbana-Champaign, IL, USA

July 23, 2015
8th International Workshop on the
Analysis of Multitemporal Remote Sensing Images
Annecy, France.

Supported by: CMMI -1030144, and AFOSR FA9550-12-1-0390

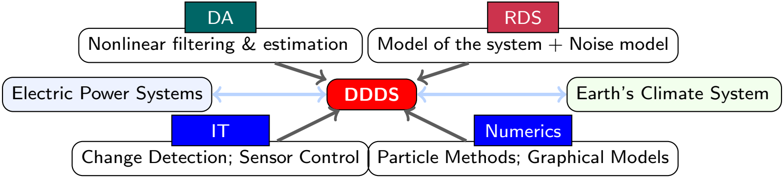
Data Driven Dynamical Systems (DDDS)

The goals of DDDS are to develop novel mathematical and statistical methods for (i) discovering fundamental structures and extracting useful information contained in high dimensional data; and (ii) assimilating this information into evolving dynamical models in order to understand large-scale data-centric problems.

Optimize the tradeoff between the detection delay and the frequency of false alarms.

Data Driven Dynamical Systems (DDDS)

The goals of DDDS are to develop novel mathematical and statistical methods for (i) discovering fundamental structures and extracting useful information contained in high dimensional data; and (ii) assimilating this information into evolving dynamical models in order to understand large-scale data-centric problems.



Optimize the tradeoff between the detection delay and the frequency of false alarms.

Motivation: Simple or "conceptual" coupled atmosphere - ocean model

Signal Process : $\varepsilon \dot{Z}_t^\varepsilon = g(X_t^\varepsilon, Z_t^\varepsilon, \xi_t^\varepsilon), \; Z_0^\varepsilon = z \in \mathbb{R}^m,$ core atmosphere model

$\dot{X}_t^\varepsilon = f(X_t^\varepsilon, Z_t^\varepsilon, \xi_t^\varepsilon), \; X_0^\varepsilon = x \in \mathbb{R}^n,$ core ocean model

where, Z_t^ε represent the general circulation of the atmosphere; X_t^ε are the ocean components such as the density gradients, angular momentum, etc.. ξ^ε represents the unmodeled dynamics of the system or an additive noise, and the initial conditions (z, x) are random variables. $n + m$-dimensional stochastic p

Observation process: Current Meters, CTDs (salinity and temperature every hour or so for a period of up to two years), buoys, acoustic releases. For example, Meridional Overturning Circulation and Heatflux Array (MOCHA) deployed along 26.5° N in the Atlantic.

The observation process is a function of the signal process corrupted by noise

$$Y_t^\varepsilon = \int_0^t h^\varepsilon(X_s^\varepsilon, Z_s^\varepsilon) ds + V_t,$$

where $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is called the sensor function.

Motivation: Simple or "conceptual" coupled atmosphere - ocean model

Signal Process : $\varepsilon \dot{Z}_t^\varepsilon = g(X_t^\varepsilon, Z_t^\varepsilon, \xi_t^\varepsilon),\ Z_0^\varepsilon = z \in \mathbb{R}^m,$ core atmosphere model

$$\dot{X}_t^\varepsilon = f(X_t^\varepsilon, Z_t^\varepsilon, \xi_t^\varepsilon),\ X_0^\varepsilon = x \in \mathbb{R}^n, \text{ core ocean model}$$

where, Z_t^ε represent the general circulation of the atmosphere; X_t^ε are the ocean components such as the density gradients, angular momentum, etc.. ξ^ε represents the unmodeled dynamics of the system or an additive noise, and the initial conditions (z, x) are random variables. $n+m$-dimensional stochastic p

Observation process: Current Meters, CTDs (salinity and temperature every hour or so for a period of up to two years), buoys, acoustic releases. For example, Meridional Overturning Circulation and Heatflux Array (MOCHA) deployed along 26.5° N in the Atlantic.

The observation process is a function of the signal process corrupted by noise

$$Y_t^\varepsilon = \int_0^t h^\varepsilon(X_s^\varepsilon, Z_s^\varepsilon)ds + V_t,$$

where $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is called the sensor function.

Main goals: Develop efficient methods to

- estimate the initial conditions / current states / parameters
- obtain reliable forecast of the system evolution – prediction.

Our interest is to estimate the slowly varying signal (coarse-grained signal) X_t^ε at time t on the basis of the sigma-algebra $\sigma\{Y_s^\varepsilon : 0 \le s \le t\}$.

- More precisely for each $t \ge 0$, we want to find the conditional law of the slowly varying signal (coarse-grained signal)

$$\pi_t^{\varepsilon,x}(A) \stackrel{\text{def}}{=} \mathbb{P}\{X_t^\varepsilon \in A | Y_s^\varepsilon : 0 \le s \le t\}, \quad \text{for all } A \in \mathscr{B}(\mathbb{R}^n).$$

- $\pi_t^\varepsilon(A)$ is governed by high dimensional SPDEs: "Curse-of-dimensionality".

- Show the conditional law $\{\pi_t^{\varepsilon,x}\}$ for the coarse - grained dynamics converge to a process $\{\pi_t^0\}$ that is governed by a lower dimensional linear SPDE.

- Utilize the reduced-dimension linear SPDE for $\{\pi_t^0\}$ to develop efficient particle filtering methods - Homogenized Hybrid Particle Filter (HHPF).

MAIN GOALS: Develop efficient methods to

- estimate the initial conditions / current states / parameters
- obtain reliable forecast of the system evolution – prediction.

Our interest is to estimate the slowly varying signal (coarse-grained signal) X_t^ε at time t on the basis of the sigma-algebra $\sigma\{Y_s^\varepsilon : 0 \le s \le t\}$.

- More precisely for each $t \ge 0$, we want to find the conditional law of the slowly varying signal (coarse-grained signal)

$$\pi_t^{\varepsilon,x}(A) \stackrel{\text{def}}{=} \mathbb{P}\{X_t^\varepsilon \in A | Y_s^\varepsilon : 0 \le s \le t\}, \quad \text{for all } A \in \mathscr{B}(\mathbb{R}^n).$$

- $\pi_t^\varepsilon(A)$ is governed by high dimensional SPDEs: "Curse-of-dimensionality".
- Show the conditional law $\{\pi_t^{\varepsilon,x}\}$ for the coarse - grained dynamics converge to a process $\{\pi_t^0\}$ that is governed by a lower dimensional linear SPDE.
- Utilize the reduced-dimension linear SPDE for $\{\pi_t^0\}$ to develop efficient particle filtering methods - Homogenized Hybrid Particle Filter (HHPF).

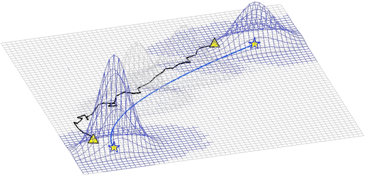


Figure: Evolution of density (FPE)

Zakai equation:

$$\underbrace{du^{\varepsilon}(t, x, z) = \mathscr{L}_{\varepsilon}^{*} u^{\varepsilon}(t, x, z)dt}_{\text{Fokker-Planck equation}} + u^{\varepsilon}(t, x, z)h(x, z)dY_t^{\varepsilon}$$

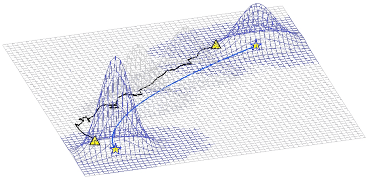


Figure: Evolution of density (FPE)

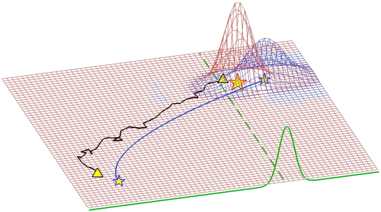


Figure: Evolution of conditional density

Zakai equation:

$$\underbrace{du^{\varepsilon}(t,x,z) = \mathscr{L}_{\varepsilon}^{*} u^{\varepsilon}(t,x,z)dt}_{\text{Fokker-Planck equation}} + u^{\varepsilon}(t,x,z)h(x,z)dY_t^{\varepsilon}$$

Technical Goals: how information interacts with complex scales

The homogenization of multiscale SDEs is a standard technique. However the convergence of the corresponding filtering problem is not trivial, that is, under the condition that

$$(X^\varepsilon, Y^\varepsilon) \Rightarrow (X, Y),$$

showing that

$$\mathbb{E}[f(X^\varepsilon)|\mathcal{F}^{Y^\varepsilon}] \to \mathbb{E}[f(X)|\mathcal{F}^{Y}]$$

for all bounded continuous functions f is not trivial.

– the convergence of $(X^\varepsilon, Y^\varepsilon)$ to (X, Y) itself does not guarantee the convergence of filters.
– while estimation, information is lost when conditioning on Y instead of Y^ε.

To come out of this conundrum, we recall that the filter is a "map from observations to distributions" and is sufficiently continuous. This filter works well enough when the limiting observations are replaced by the pre-limit ones:

$$\boxed{\pi_t^0(\cdot, \vec{\mathbf{Y}}^\varepsilon_{[0,t]}) \quad \text{is close to} \quad \pi_t^{\varepsilon,x}(\cdot, \vec{\mathbf{Y}}^\varepsilon_{[0,t]})}$$

In other words, the filter of the coarse-grained dynamics conditioned to the real observation is close to x-marginal of the original filter!!

Multi-dimensional Case: Homogenization in multiscale filtering

Construction of a Homogenized filter

Find ρ^0:

$$d\rho_t^0(\varphi) = \rho_t^0\left(\bar{\mathcal{L}}\varphi\right)dt + \rho_t^0\left(\bar{h}\varphi\right)dY_t^\varepsilon, \quad \rho_0^0(\varphi) = \mathbb{E}\left[\varphi\left(X_0^0\right)\right]$$

$\pi_t^0(\varphi) \stackrel{\text{def}}{=} \frac{\rho_t^0(\varphi)}{\rho_t^0(1)}$ and $\bar{\mathcal{L}} = \sum_{i=1}^m \bar{b}_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m \bar{a}_{ij}(x,z)\frac{\partial^2}{\partial x_i \partial x_j}$

Goal:

1. Find a suitable version of π^0 associated with (X^0, Y^ε) (homogenized process, actual observation)
2. Show that $\pi^{\varepsilon,x}$ is close to π^0 in L^p-sense as $\varepsilon \to 0$[a]

$$\lim_{\varepsilon\to 0}\sup_{t\le T}\mathbb{E}_{\mathbb{Q}}\left[d\left(\pi_t^{\varepsilon,x}, \pi_t^0\right)^p\right] = 0, \quad \forall T > 0$$

[a] P. Imkeller, N. Sri Namachchivaya, N. Perkowski, and H. Yeong, Dimensional reduction in nonlinear filtering: A homogenization approach. *Annals of Applied Probability*, Vol. 23, No. 6, 2290-2326, 2013 (extension of J. H. Park, N. Sri Namachchivaya, and R. B. Sowers, Dimensional reduction in nonlinear filtering. *Nonlinearity*, Vol. 23, 2010, *Stochastics and Dynamics*, Vol. 8(3) 543-560, 2008)

Multi-dimensional Case: Homogenization in multiscale filtering

Construction of a Homogenized filter

Find ρ^0:

$$d\rho_t^0(\varphi) = \rho_t^0\left(\bar{\mathcal{L}}\varphi\right)dt + \rho_t^0\left(\bar{h}\varphi\right)dY_t^\varepsilon, \quad \rho_0^0(\varphi) = \mathbb{E}\left[\varphi\left(X_0^0\right)\right]$$

$\pi_t^0(\varphi) \stackrel{\text{def}}{=} \frac{\rho_t^0(\varphi)}{\rho_t^0(1)}$ and $\bar{\mathcal{L}} = \sum_{i=1}^m \bar{b}_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m \bar{a}_{ij}(x,z)\frac{\partial^2}{\partial x_i \partial x_j}$

Goal:

1. Find a suitable version of π^0 associated with (X^0, Y^ε) (homogenized process, actual observation)
2. Show that $\pi^{\varepsilon,x}$ is close to π^0 in L^p-sense as $\varepsilon \to 0$[a]

$$\lim_{\varepsilon\to 0}\sup_{t\le T}\mathbb{E}_{\mathbb{Q}}\left[d\left(\pi_t^{\varepsilon,x},\pi_t^0\right)^p\right] = 0, \quad \forall T > 0$$

[a] P. Imkeller, N. Sri Namachchivaya, N. Perkowski, and H. Yeong, Dimensional reduction in nonlinear filtering: A homogenization approach. *Annals of Applied Probability*, Vol. 23, No. 6, 2290-2326, 2013 (extension of J. H. Park, N. Sri Namachchivaya, and R. B. Sowers, Dimensional reduction in nonlinear filtering. *Nonlinearity*, Vol. 23, 2010, *Stochastics and Dynamics*, Vol. 8(3) 543-560, 2008)

Homogenized Hybrid Particle Filter (HHPF)

"curse of dimensionality" is partially resolved by the Homogenized Filtering Equations

$$d\rho_t^0(\varphi) = \rho_t^0\left(\bar{\mathcal{L}}\varphi\right)dt + \rho_t^0\left(\bar{h}\varphi\right)dY_t^\varepsilon, \quad \rho_0^0(\varphi) = \mathbb{E}\left[\varphi\left(X_0^0\right)\right] \tag{1}$$

Particle method

- Numerical approximation of π^0 or ρ^0
- (weighted) particles to represent (conditional) density

Main Idea: The solution of the nonlinear filtering equation is approximated by a system of N particles with varying weights

$$U_N^\varepsilon(t) = \sum_{j=1}^{N} w^{x_t^{\varepsilon,j}} \delta_{x_t^{\varepsilon,j}},$$

where $\{x_t^{\varepsilon,j}; 1 \leq j \leq N\}$ is a set of random particles which evolve according to system dynamics and $w^{x_t^{\varepsilon,j}}$s are the corresponding weights.

Homogenized Hybrid Particle Filter (HHPF)

"curse of dimensionality" is partially resolved by the Homogenized Filtering Equations

$$d\rho_t^0(\varphi) = \rho_t^0\left(\bar{\mathcal{L}}\varphi\right)dt + \rho_t^0\left(\bar{h}\varphi\right)dY_t^{\varepsilon}, \quad \rho_0^0(\varphi) = \mathbb{E}\left[\varphi\left(X_0^0\right)\right] \tag{1}$$

Particle method

- Numerical approximation of π^0 or ρ^0
- (weighted) particles to represent (conditional) density

Main Idea: The solution of the nonlinear filtering equation is approximated by a system of N particles with varying weights

$$U_N^{\varepsilon}(t) = \sum_{j=1}^{N} w^{x_t^{\varepsilon,j}} \delta_{x_t^{\varepsilon,j}},$$

where $\{x_t^{\varepsilon,j}; 1 \leq j \leq N\}$ is a set of random particles which evolve according to system dynamics and $w^{x_t^{\varepsilon,j}}$s are the corresponding weights.

Homogenized Hybrid Particle Filter (HHPF)[1]

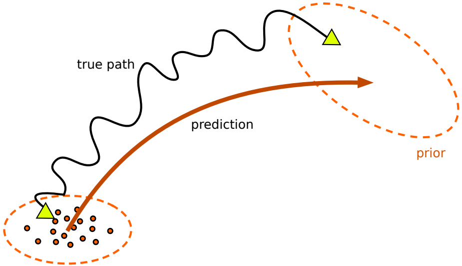


Figure: Initial condition

Zakai equation: $\underbrace{d\bar{u}^{\varepsilon}(x,t) = \bar{\mathcal{L}}\bar{u}^{\varepsilon}(x,t)dt}_{\text{Fokker-Planck equation}} + \bar{h}^{*}(x,t)\bar{u}^{\varepsilon}(x,t)dY_t^{\varepsilon}$

Weight update: $w_{t+\delta t}^{\varepsilon,i} = \exp\left\{\int_t^{t+\delta t} \bar{h}^{*}(x_s^i,s)dY_s^{\varepsilon} - \frac{1}{2}\int_t^{t+\delta t}\left\|\bar{h}(x_s^i,s)\right\|^2 ds\right\}$

$$\bar{w}_{t+\delta t}^{\varepsilon,i} = \frac{w_{t+\delta t}^{\varepsilon,i}}{\sum_{i=1}^{N_s} w_{t+\delta t}^{\varepsilon,i}}$$

[1] J. H. Park, N. Sri Namachchivaya, H. Yeong, Particle Filters in a multi-scale environment: Homogenized hybrid particle filter. J. Appl. Mech., 78(6), 2011

Homogenized Hybrid Particle Filter (HHPF)[1]

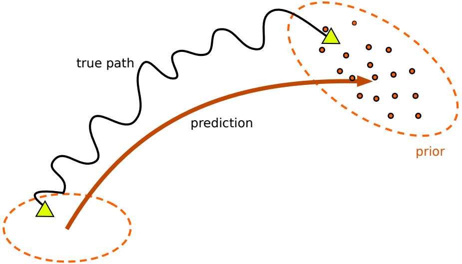


Figure: Particles propagation

Zakai equation: $\underbrace{d\bar{u}^{\varepsilon}(x,t) = \bar{\mathcal{L}}\bar{u}^{\varepsilon}(x,t)dt}_{\text{Fokker-Planck equation}} + \bar{h}^{*}(x,t)\bar{u}^{\varepsilon}(x,t)dY_t^{\varepsilon}$

Weight update: $w_{t+\delta t}^{\varepsilon,i} = \exp\left\{\int_t^{t+\delta t} \bar{h}^{*}(x_s^i,s)dY_s^{\varepsilon} - \frac{1}{2}\int_t^{t+\delta t}\left\|\bar{h}(x_s^i,s)\right\|^2 ds\right\}$

$\bar{w}_{t+\delta t}^{\varepsilon,i} = \frac{w_{t+\delta t}^{\varepsilon,i}}{\sum_{i=1}^{N_s} w_{t+\delta t}^{\varepsilon,i}}$

[1] J. H. Park, N. Sri Namachchivaya, H. Yeong, Particle Filters in a multi-scale environment: Homogenized hybrid particle filter. J. Appl. Mech., 78(6), 2011

Homogenized Hybrid Particle Filter (HHPF)[1]

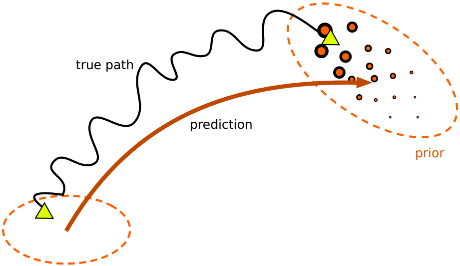


Figure: Weights update with observation Y_t^{ε}

Zakai equation: $\underbrace{d\bar{u}^{\varepsilon}(x,t) = \bar{\mathcal{L}}\bar{u}^{\varepsilon}(x,t)dt}_{\text{Fokker-Planck equation}} \quad + \bar{h}^{*}(x,t)\bar{u}^{\varepsilon}(x,t)dY_t^{\varepsilon}$

Weight update: $w_{t+\delta t}^{\varepsilon,i} = \exp\left\{ \int_t^{t+\delta t} \bar{h}^{*}(x_s^i, s) dY_s^{\varepsilon} - \frac{1}{2} \int_t^{t+\delta t} \left\| \bar{h}(x_s^i, s) \right\|^2 ds \right\}$

$$\bar{w}_{t+\delta t}^{\varepsilon,i} = \frac{w_{t+\delta t}^{\varepsilon,i}}{\sum_{i=1}^{N_s} w_{t+\delta t}^{\varepsilon,i}}$$

D

[1] J. H. Park, N. Sri Namachchivaya, H. Yeong, Particle Filters in a multi-scale environment: Homogenized hybrid particle filter. J. Appl. Mech., 78(6), 2011

Homogenized Hybrid Particle Filter (HHPF)[1]

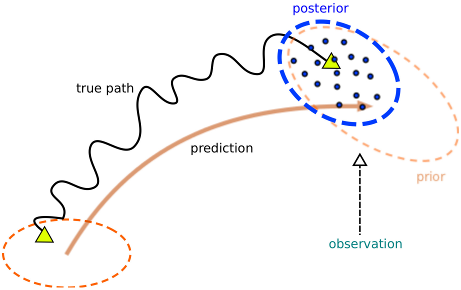


Figure: Resample → Conditional density

Zakai equation: $\underbrace{d\bar{u}^\varepsilon(x,t) = \bar{\mathcal{L}}\bar{u}^\varepsilon(x,t)dt}_{\text{Fokker-Planck equation}} + \bar{h}^*(x,t)\bar{u}^\varepsilon(x,t)dY^\varepsilon_t$

Weight update: $w^{\varepsilon,i}_{t+\delta t} = \exp\left\{\int_t^{t+\delta t} \bar{h}^*(x^i_s,s)dY^\varepsilon_s - \frac{1}{2}\int_t^{t+\delta t}\left\|\bar{h}(x^i_s,s)\right\|^2 ds\right\}$

$$\bar{w}^{\varepsilon,i}_{t+\delta t} = \frac{w^{\varepsilon,i}_{t+\delta t}}{\sum_{i=1}^{N_s} w^{\varepsilon,i}_{t+\delta t}}$$

[1] J. H. Park, N. Sri Namachchivaya, H. Yeong, Particle Filters in a multi-scale environment: Homogenized hybrid particle filter. J. Appl. Mech., 78(6), 2011

Lorenz '63 model - A toy model of atmospheric convection

$$\begin{aligned} \dot{X}_t &= -\sigma X_t + \sigma Y_t && + \xi_x(t) \\ \dot{Y}_t &= \rho X_t - Y_t - X_t Z_t && + \xi_y(t) \\ \dot{Z}_t &= -\beta Z_t - X_t Y_t && + \xi_z(t) \end{aligned}$$

- Typical values for parameters are $\sigma = 10$ (Prandtl number), $\rho = 8/3$ and β can vary; model exhibits chaotic behavior at $\beta = 28$ ($\lambda = 0.9056$ and $\tau_d = 0.77$ units)
- "True" signal generated and observations taken every 50 timesteps (0.2 units) which represent approximately one-fourth of doubling time (sparse data).
- signal noise ξ added is vector of Gaussian random numbers premultiplied by correlation matrix $\begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix}$.
- observation is taken as signal plus Gaussian noise with correlation matrix $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.
- Particle filters implemented with 20 particles, resample when effective number of particles falls below 5

Lorenz '63 model - A toy model of atmospheric convection

$$\dot{X}_t = -\sigma X_t + \sigma Y_t \qquad + \xi_x(t)$$
$$\dot{Y}_t = \rho X_t - Y_t - X_t Z_t \qquad + \xi_y(t)$$
$$\dot{Z}_t = -\beta Z_t - X_t Y_t \qquad + \xi_z(t)$$

- Typical values for parameters are $\sigma = 10$ (Prandtl number), $\rho = 8/3$ and β can vary; model exhibits chaotic behavior at $\beta = 28$ ($\lambda = 0.9056$ and $\tau_d = 0.77$ units)
- "True" signal generated and observations taken every 50 timesteps (0.2 units) which represent approximately one-fourth of doubling time (sparse data).
- signal noise ξ added is vector of Gaussian random numbers premultiplied by correlation matrix $\begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix}$.
- observation is taken as signal plus Gaussian noise with correlation matrix $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.
- Particle filters implemented with 20 particles, resample when effective number of particles falls below 5

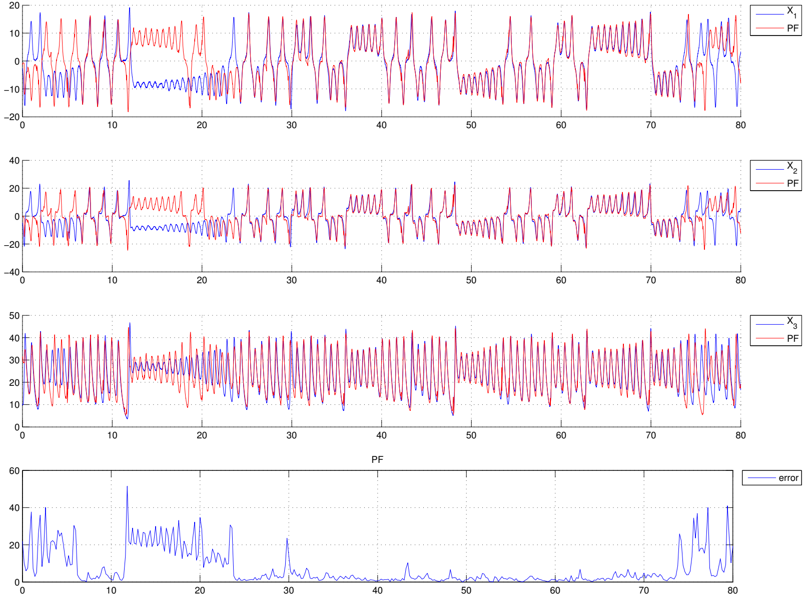


Figure: Lorenz '63 model: Standard Particle filter; $N_s = 3$

"Close, but no cigar" when it comes to data assimilation in chaotic systems with sparse data – observations taken over discrete times close to "error doubling time".

Difficult Issues:

- Systems with positive Lyapunov exponents – small errors in the estimate of the current state can grow to have a major impact on the forecast.
- Probability space in large-dimensional systems is "empty" – very few of the particles end up close to the actual location, and hence receive large fraction of the weight – the curse of dimensionality and particle collapse.

Solutions:

- Use control methods and importance sampling as a basic and flexible tool for the construction of the proposal density.
- Superimpose a control on the particle dynamics to drive the particles to locations most representative of the observations without losing the sample diversity.

"Close, but no cigar" when it comes to data assimilation in chaotic systems with sparse data – observations taken over discrete times close to "error doubling time".

Difficult Issues:

- Systems with positive Lyapunov exponents – small errors in the estimate of the current state can grow to have a major impact on the forecast.
- Probability space in large-dimensional systems is "empty" – very few of the particles end up close to the actual location, and hence receive large fraction of the weight – the curse of dimensionality and particle collapse.

Solutions:

- Use control methods and importance sampling as a basic and flexible tool for the construction of the proposal density.
- Superimpose a control on the particle dynamics to drive the particles to locations most representative of the observations without losing the sample diversity.

Optimal Nudging in Particle Filtering

Very few of the particles end up close to the actual location, and hence receive large fraction of the weight: the curse of dimensionality and particle collapse.

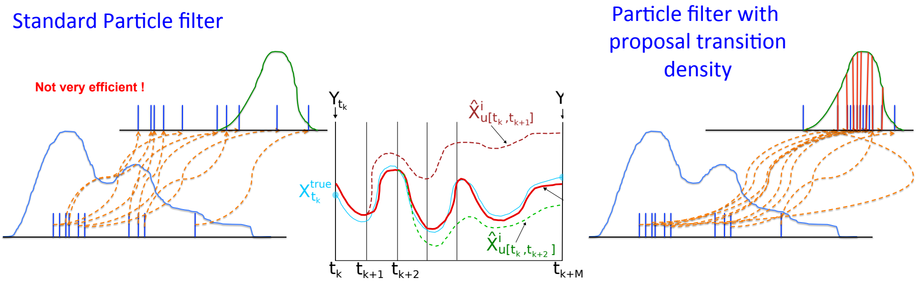


Figure: Standard particle filters, Optimal nudging and Nudged particle filters

Nudging consists of adding forcing terms to the "prognostic" equations, that steer the particles toward the observations – at the same time, not to over control so that the sample diversity is lost.

Control the particles to construct a new proposal density

Find the optimal control u which minimizes the cost:

$$J(t_k, x_k; u) = \hat{\mathbb{E}}_{t_k, x_k}\left[\frac{1}{2}\int_{t_k}^{t_{k+1}} \underbrace{u(s)^T Q^{-1} u(s)}_{\text{input energy}} ds + \underbrace{g(\hat{X}(t_{k+1}))}_{\text{terminal cost}}\right], \quad (2)$$

where $g(x) = \frac{1}{2}(Y_{k+1} - h(x))^T R^{-1}(Y_{k+1} - h(x))$ - penalty for being away from observation and $\hat{\mathbb{E}}_{t_k, x_k}$ is the probability measure generated by the controlled process $\hat{X}$ staring at x_k at time t_k:

$$d\hat{X}(s) = \bar{b}(\hat{X}(s))ds + u(s)ds + \sigma dW, \quad t_k \leq t \leq t_{k+1}, \quad \hat{X}(t_k) = x_k.$$

- less penalty on the size of the control in the directions with large noise amplitude (allow for more correction!).
- in directions where the quality of the observation is poor, we allow the particles to be further away from the observations.

Control the particles to construct a new proposal density

Find the optimal control u which minimizes the cost:

$$J(t_k, x_k; u) = \hat{\mathbb{E}}_{t_k, x_k} \left[\frac{1}{2} \int_{t_k}^{t_{k+1}} \underbrace{u(s)^T Q^{-1} u(s)}_{\text{input energy}} ds + \underbrace{g(\hat{X}(t_{k+1}))}_{\text{terminal cost}} \right], \qquad (2)$$

where $g(x) = \frac{1}{2}(Y_{k+1} - h(x))^T R^{-1}(Y_{k+1} - h(x))$ - penalty for being away from observation and $\hat{\mathbb{E}}_{t_k, x_k}$ is the probability measure generated by the controlled process $\hat{X}$ staring at x_k at time t_k:

$$d\hat{X}(s) = \bar{b}(\hat{X}(s))ds + u(s)ds + \sigma dW, \quad t_k \le t \le t_{k+1}, \quad \hat{X}(t_k) = x_k.$$

- less penalty on the size of the control in the directions with large noise amplitude (allow for more correction!).
- in directions where the quality of the observation is poor, we allow the particles to be further away from the observations.

Optimal Nudging Embedded on HHPF

The optimal control is $u(t) = -\sigma\sigma^T D_x V(t, \hat{X}(t))$, and taking $u(t) = \sigma v(t)$ for simplicity, we construct the proposal measure from the controlled process

$$\boxed{d\hat{X}(t) = \bar{b}(\hat{X}(t))dt + \sigma dW + \sigma v(t)dt,\ t_k \leq t \leq t_{k+1},\ \hat{X}(t_k) = x_k,} \quad (3)$$

where

$$v(t) = -\sigma^T D_x V(t, \hat{X}(t)).$$

We evolve the particles according to (3), using the principle of importance sampling, the weights are updated according to

$$w_i^{k+1} \propto \exp\left(-g(Y_{k+1}, \hat{X}_i(t_{k+1}))\right) \frac{d\mu_{\tilde{X}}{}^i}{d\mu_{\hat{X}}{}^i} w_i^k, \quad (4)$$

where $\mu_{\tilde{X}}$ and $\mu_{\hat{X}}$ are the measures generated by the original (**??**) and controlled processes (3), respectively, evolving for $t_k \leq t \leq t_{k+1}$ with starting point x at t_k.

$$\frac{d\mu_{\tilde{X}}}{d\mu_{\hat{X}}} = \exp\left(-\int_{t_k}^{t_{k+1}} v^T(s)\, dW(s) - \frac{1}{2}\int_{t_k}^{t_{k+1}} v(s)^T v(s)\, ds\right)$$

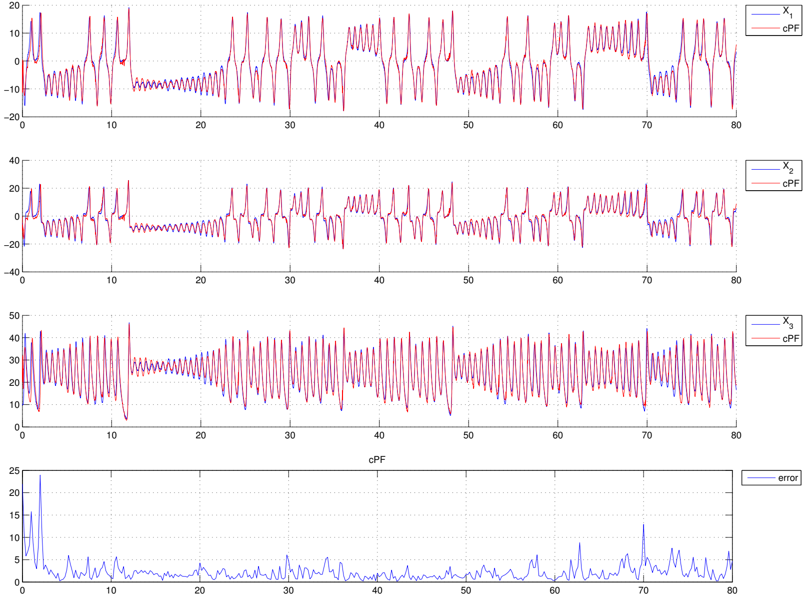


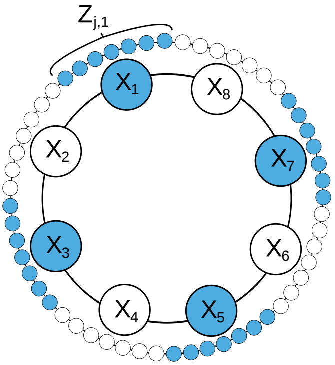
Figure: Lorenz '63 model: Particle filter with optimal control; $N_s = 3$

Lorenz-96[2]: Heuristic midlatitude atmospheric model

The large scale, low frequency variables X represent some scaler meteorological quantity, for example, vorticity or temperature at J equally spaced sites around the latitude circle. Each X is coupled to K small scale, high frequency Z variables.

$$\dot{X}_k = \overbrace{-X_{k-1}(X_{k-2} - X_{k+1})}^{\text{advection}} - \overbrace{X_k}^{\text{dissipation}} + \underbrace{F_x}_{\text{forcing}} + \frac{h_x}{J}\sum_{j=1}^{J} Z_{j,k}$$

$$\dot{Z}_{j,k} = \frac{1}{\varepsilon}(-Z_{j+1,k}(Z_{j+2,k} - Z_{j-1,k}) - Z_{j,k} + h_z X_k) + \frac{1}{\sqrt{\varepsilon}}\zeta(t)$$



- $J = 36$, $K = 10$, $\varepsilon = 1/128$
- sensitive to initial conditions: a large number of positive LEs (36% or 13/36 at F=10).
- filtering methods require modifications

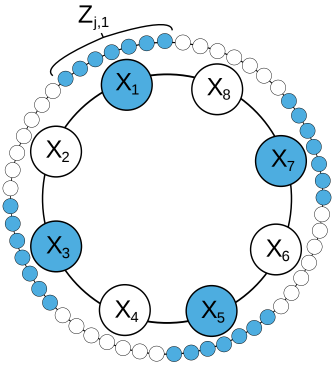
[2] E. N. Lorenz, *Proc. Seminar on Predictability*, 1(1), 1996

Lorenz-96[2]: Heuristic midlatitude atmospheric model

The large scale, low frequency variables X represent some scaler meteorological quantity, for example, vorticity or temperature at J equally spaced sites around the latitude circle. Each X is coupled to K small scale, high frequency Z variables.

$$\dot{X}_k = \overbrace{-X_{k-1}(X_{k-2} - X_{k+1})}^{\text{advection}} - \overbrace{X_k}^{\text{dissipation}} + \underbrace{F_x}_{\text{forcing}} + \frac{h_x}{J}\sum_{j=1}^{J} Z_{j,k}$$

$$\dot{Z}_{j,k} = \frac{1}{\varepsilon}(-Z_{j+1,k}(Z_{j+2,k} - Z_{j-1,k}) - Z_{j,k} + h_z X_k) + \frac{1}{\sqrt{\varepsilon}}\zeta(t)$$



- $J = 36$, $K = 10$, $\varepsilon = 1/128$

- sensitive to initial conditions: a large number of positive LEs (36% or 13/36 at F=10).
- filtering methods require modifications

[2] E. N. Lorenz, *Proc. Seminar on Predictability*, 1(1), 1996

Model parameters for sparse observations

- # of slow, large amplitude modes $K = 36$;
- # of fast, low amplitude modes $J = 10$ implies 360 total fast modes;
- constant external forcing $F_x = 10$;
- coupling coefficients $h_x = -0.8$, $h_z = 1$;
- $\varepsilon = 0.075$

Numerical integration

- $T = 20$ (time) units $= 100$ days, micro timestep $\delta t = 0.0005 = 3.6$ min;
- macro timestep $\Delta t = 0.05 = 6$ hrs;
- $N_s = 20$, Error doubling time $= 36$ hrs, observations at every 36 hrs
- Processor: Intel Xeon DP Hexa-core X5675s (dual, 3.07GHz)
- Integration scheme: Euler-Maruyama (stochastic), Runge-Kutta (deterministic) in MATLAB (R2010b)

HHPFs comparison

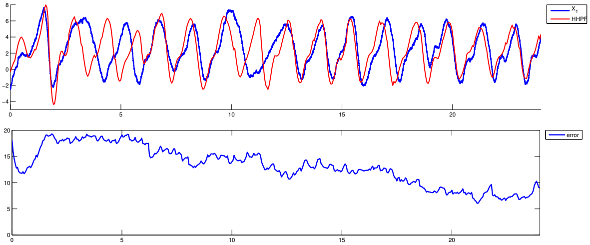


Figure: Direct HHPF

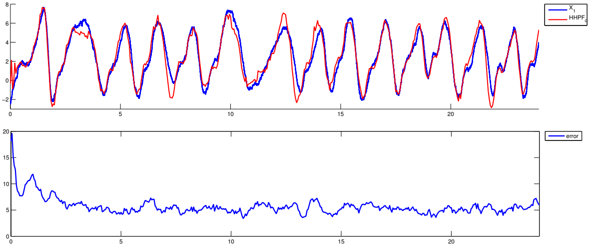


Figure: Optimized HHPF

Upper figure: true solution in blue. Lower figure: absolute error vs time.

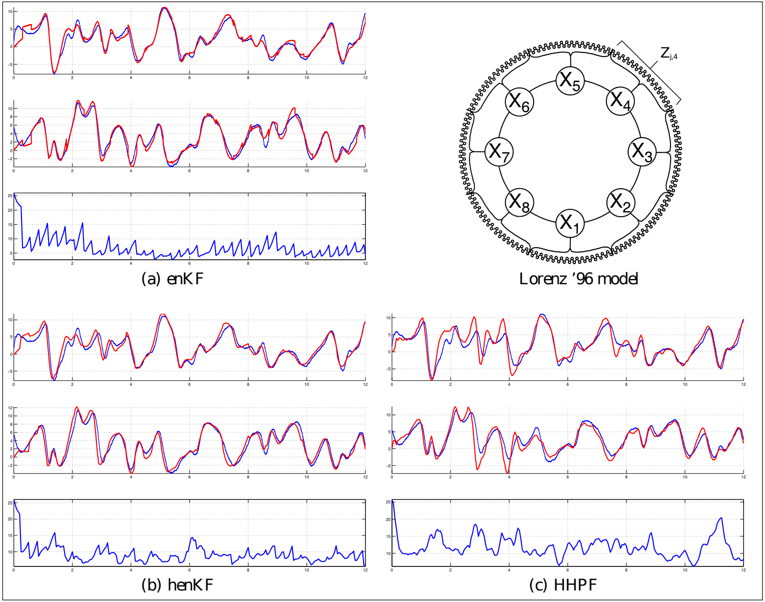


Figure: Estimation performance of the different filters; counterclockwise from the top left, the enKF, the henKF and the HHPF. The blue curve in the top two plots of each figure represents the true state trajectory while the red curve represents the estimated trajectory.

N_s	Opt. HHPF		Direct HHPF		Homog. EnKF		EnKF	
	time	*RMSE*	*time*	*RMSE*	*time*	*RMSE*	*time*	*RMSE*
10	82	26.04	26	69.27	26	32.53	785	25.96
20	97	24.52	27	58.09	27	26.92	587	20.05
50	241	24.78	68	42.71	70	24.53	703	17.32
100	493	24.92	139	36.77	134	23.75	861	16.21
200	1153	24.71	375	35.37	369	23.58	1539	16.46

Table: Average computation times (in sec.) and RMSEs for different sample sizes. Data recorded every 0.3 time units (6 $\Delta t \approx$ error doubling time of 36 hrs real time).

The computational time of the schemes for the same level of RMSE ($\approx$ 25):

EnKF (485 sec.) $>>$ **Opt.HHPF (97 sec.)** $\geq$ **Homog.EnKF (70 sec.)**

Conclusions: lower dimensional filters

We showed the efficient utilization of the low-dimensional models of the signal to develop a low-dimensional nonlinear filtering equations (Zakai-type equation) that determine the conditional law of the coarse-grained signal, X_t, of complex systems, in multi-scale environments.

Developed the *Homogenized Hybrid Particle Filter (HHPF)* which combined the results on reduced order nonlinear filtering with sequential Monte Carlo methods. HHPF reduces the effective number of variables in the evaluation of the conditional distribution needed in the Bayesian filter for data assimilation.

Numerical studies were presented to illustrate that in settings where the signal and observation dynamics are nonlinear a suitably chosen HHPF scheme can drastically outperform the regular particle filters.