Particle Filters with Nudging in Multiscale Chaotic Systems: with Application to the Lorenz-96 Atmospheric Model

H. C. Yeong, R. T. Beeson and N. Sri Namachchivaya
Department of Aerospace Engineering, University of Illinois, Urbana, Illinois, USA

Summary. Particle methods in nonlinear filtering suffer from the "curse of dimensionality", the issue of particles degeneracy within a sample, which increases exponentially with system dimension. As a result, particle filtering for high dimensional systems can benefit from some form of dimensional reduction. This paper presents lower dimensional particle-based filters that are developed by combining stochastic dimensional reduction and nonlinear filtering. The particle filter is further adapted to the complexities of multi-scale signals that are inherently chaotic, by combining importance sampling and stochastic optimal control techniques. Specifically, a control is superimposed on particle dynamics to drive particles to locations most representative of observations. The control is chosen to minimize a cost functional, such that the prior distribution of the particles matches as closely as possible the posterior distribution of the true state given observations, while still staying true to the signal dynamics.

Introduction

Atmospheric models used in numerical weather forecasting are examples of multi-scale complex systems. It is well known from the study of non-linear dynamical systems that even simple deterministic non-linear systems can give rise to complex behavior which is statistically indistinguishable from that of completely random processes. One consequence of this is that it may be possible to describe apparently complex signals using simple chaotic systems. In this paper, we consider a simple heuristic atmospheric model, which nonetheless exhibits many of the difficulties arising in realistic models, to gain insight into predictability and data assimilation.

An ongoing challenge for inference and prediction of large-scale complex systems is to efficiently *analyze and assimilate* high dimensional data produced by vast numbers of engineered and natural systems. This work focuses on the challenges in data assimilation, which arise from the interactions between uncertainties, nonlinearities, and observations. Lower dimensional ensemble Kalman filter (enKF) and sequential importance sampling particle filter (SIS PF) are developed based on the theoretical framework of stochastic dimensional reduction and nonlinear filtering. The SIS PF is further adapted to the complexities of data assimilation in inherently chaotic models, when observations are sparse in time.

The Lorenz '96 Model

The Lorenz '96 [1] atmospheric model that mimics midlatitude atmospheric dynamics with microscopic convective processes, is a useful tool for testing data assimilation methods for use in numerical weather simulations, owing to its transparency and low computational cost. A stochastic version of the two-timescales Lorenz-96 model is used here – the full multi-scale system is defined as the "truth", whereas a truncated, homogenized version is used as testbed for the data assimilation schemes.

The Lorenz '96 model was originally introduced in [1] to mimic an unspecified scalar meteorological quantity at midlatitude. A latitude circle is divided into $K = 36$ sectors, and each sector is subdivided into $J = 10$ subsectors:

$$dX_t^k = \left(-X_t^{k-1}(X_t^{k-2} - X_t^{k+1}) - X_t^k + F + \frac{h_x}{J}\sum_{j=1}^{J} Z_t^{k,j} \right) dt + \sigma_x dW_t, \quad k = 1, \ldots, K, \tag{1a}$$

$$dZ_t^{k,j} = \frac{1}{\varepsilon}\left(-Z_t^{k,j+1}(Z_t^{k,j+2} - Z_t^{k,j-1}) - Z_t^{k,j} + h_z X_t^k \right) dt + \frac{1}{\sqrt{\varepsilon}}\sigma_z dV_t, \quad j = 1, \ldots, J. \tag{1b}$$

X_t^k represents a slow-scale atmospheric variable at time t in the k^{th} sector. $(W, V) \in \mathbb{R}^{k\times kj}$ are independent Gaussian noises to represent unresolved uncertainties and $\sigma_x \in \mathbb{R}^{k\times k}$, $\sigma_z \in \mathbb{R}^{kj\times kj}$. (*Note:* We use superscripts k and j to conform with the typical spatial indexing notation used for the Lorenz '96 model. In sections that follow, subscripts k and j will be used as discrete time indices, not to be confused with the spatial indices of the Lorenz model). Each X_t^k is coupled to its neighbors X_t^{k+1}, X_t^{k-1}, and X_t^{k-2} to mimic the westerly wind pattern in midlatitude.

The influence of multiple spatio-temporal scales is incorporated by dividing each sector k into J subsectors, and introducing $Z_t^{k,j}$ in each subsector. Coupling between neighbors model advection between sectors and subsectors, while coupling between each sector and its subsectors models damping. The dynamics of the X-components is also subjected to linear external forcing F.

$\varepsilon << 1$ in (1b) is a small timescale separation parameter. Thus, in (1), each X_t^k represents a slowly-varying, large amplitude atmospheric quantity, with J fast-varying, low amplitude quantities, $Z_t^{k,j}$, associated with it. In the context of climate modeling, the slow component is also known as the resolved climate modes while the fast-varying component is known as the unresolved non-climate modes. In the fast scale dynamics in [1], the nonlinear effects of order $\frac{1}{\varepsilon}$ are dominant while the linear and slow-scale forcing are of orders $\frac{1}{\sqrt{\varepsilon}}$ and 1, respectively. Here, we use a version of the Lorenz '96 model used in [2] and [3], in which the nonlinear, linear and slow scale effects in the fast dynamics are all of order $\frac{1}{\varepsilon}$. In this setting, [2] showed that (for a lower order version of the Lorenz '96 model) the fast scale dynamics display ergodic properties such that the averaging technique described in the lower dimensional filtering section can be

utilized to average out the fast dynamics when we are only interested in the slow dynamics (coarse-grained process). This is taken advantage of to reduced the dimension of the filtering problem in the work presented here.
Considering (1), in which only quadratic nonlinearity is present, the motivation behind adding stochastic forcing is to represent higher order unresolved effects, hence the stochastic effects are made small compared to the linear and quadratic effects via σ_x and σ_z.

Lower dimensional filtering

The filtering problem here is a 396-dimensional problem ($K = 36$, $J = 10$). Observations are taken as the slow states with additive Gaussian noise, collected at discrete numerical integration timesteps, at intervals equivalent to 1.5 days in real time – approximately the error doubling time for the slow-scale system when external forcing $F = 8$. We are only interested in estimating the slow-scale processes, which enables us to effectively reduce the dimension of the filtering problem. The lower-dimensional filters presented here are based on the theoretical results in [4]. The idea is that, as the timescale separation grows large ($\varepsilon \searrow 0$), the slow-scale process converges in distribution to a process that can be represented by a system with the same dimension as the slow states, in which the effects of the fast-scale has been averaged appropriately. Specifically, for (1), $X \xrightarrow{\text{d.}} \bar{X}$ as $\varepsilon \searrow 0$, where $\bar{X}$ satisfies a 36-dimensional equation similar to (1a), with the fast-scale forcing effect appropriately averaged. From a numerical simulation point of view, instead of simulating the 396-dimension system (1), we can simulate the 36-dimensional equation for $\bar{X}$ to obtain probabilistically similar results. Hence, $\bar{X}$ can be used as the states for the enKF and SIS PF to construct posterior filtering densities of the slow-scale process of (1). We call these the homogenized enKF and homogenized hybrid PF (henKF and HHPF), respectively. The henKF and HHPF are introduced in [3] and [5], respectively. In the following, we briefly describe the main theoretical result of [4] that supports the henKF and HHPF.
A system of the type (1) can be represented by a general two-timescale stochastic differential equation (SDE) along with observation:

$$dX_t^\varepsilon = b(X_t^\varepsilon, Z_t^\varepsilon)dt + \sigma(X_t^\varepsilon, Z_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x \in \mathbb{R}^m, \tag{2a}$$

$$dZ_t^\varepsilon = \frac{1}{\varepsilon} f(X_t^\varepsilon, Z_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}} g(X_t^\varepsilon, Z_t^\varepsilon)dV_t \quad Z_0^\varepsilon = z \in \mathbb{R}^n, \tag{2b}$$

$$dY_t^\varepsilon = h(X_t^\varepsilon, Z_t^\varepsilon)dt + dB_t, \quad Y_0^\varepsilon = \mathbf{0}_{d\times 1}. \tag{2c}$$

Here X^ε and Z^ε are, respectively, the slow and fast components of the state, or signal, and Y^ε is the observation. $0 < \varepsilon << 1$ is the timescale separation parameter and $(W, V, B) \in \mathbb{R}^{k+l+d}$ are independent standard Brownian motions. For conditions on the drift and diffusion coefficients b, f, h and σ, g, we refer the reader to [4].
In the fast timescale that is of order ε, variations of the fast component are observed, but the slow component will be approximately constant. Assume that for every fixed $x \in \mathbb{R}^m$, the process Z^x of (2b) with $X^\varepsilon = x$ fixed is ergodic and converges exponentially fast to its unique stationary distribution. Then, stochastic averaging theory (see, for example, [6]) tells us that under suitable conditions, X^ε converges in law to X^0 as $\varepsilon \searrow 0$, where X^0 is the solution of an SDE

$$dX_t^0 = \bar{b}(X_t^0)dt + \bar{\sigma}(X_t^0)dW_t, \quad X_0^0 = x \in \mathbb{R}^m. \tag{3}$$

Denote by $p_\infty(z; x)$ the stationary density of Z^x in the slow timescale. The coefficients in (3) are

$$\bar{b}(x) \overset{\text{def}}{=} \int_{\mathbb{R}^n} b(x, z)p_\infty(z; x)dz, \qquad \bar{\sigma}\bar{\sigma}^*(x) \overset{\text{def}}{=} \int_{\mathbb{R}^n} \sigma\sigma^*(x, z)p_\infty(z; x)dz.$$

X^0 can be interpreted as a representation of the slow component of (2) with the fast scale effects averaged out. X^0 is used to construct an averaged lower-dimensional filter π^0.
Let $\mathbb{Q}$ denote the probability measure on the filtration generated by the standard Brownian motion (W, V, B). More precisely, for any bounded measurable function φ on $\mathbb{R}^m$, the best estimate of $\varphi(X^\varepsilon)$ given information from observations is the measure-valued process $(\pi_t^\epsilon, t \geq 0)$ that is the conditional expectation $\pi_t^\varepsilon(\varphi) = \mathbb{E}_\mathbb{Q}\left[\varphi(X_t^\varepsilon, Z_t^\varepsilon) \middle| \mathcal{Y}_t^\epsilon\right]$, where $\mathbb{E}_\mathbb{Q}$ is the expectation with respect to $\mathbb{Q}$ and $\mathcal{Y}_t^\varepsilon = \sigma\left\{Y_s^\varepsilon : 0 \leq s \leq t\right\} \vee \mathcal{N}$ is the sigma-algebra (information) generated by the observations up to time t, where $\mathcal{N}$ are the $\mathbb{Q}$-negligible sets. π^ε is the optimal filter. The Kallianpur-Striebel formula tells us that π^ε can be written in terms of the unnormalized filter ρ^ε:

$$\pi_t^\varepsilon(\varphi) = \frac{\rho_t^\varepsilon(\varphi)}{\rho_t^\varepsilon(1)}, \qquad \rho_t^\varepsilon(\varphi) \overset{\text{def}}{=} \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\varphi(X_t^\varepsilon)(D_t^\varepsilon)^{-1} \middle| \mathcal{Y}_t^\varepsilon\right],$$

where $\mathbb{P}^\varepsilon$ is given by the measure change

$$D_t^\varepsilon \overset{\text{def}}{=} \left.\frac{d\mathbb{P}^\varepsilon}{d\mathbb{Q}}\right|_{\mathcal{F}_t} = \exp\left(-\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)^* dB_s - \frac{1}{2}\int_0^t \|h(X_s^\varepsilon, Z_s^\varepsilon)\|^2 ds\right).$$

Under $\mathbb{P}^\varepsilon$, the mean of B_t is shifted by $\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)ds$, so the observation process Y^ε is a standard Brownian motion and is independent of $(X^\varepsilon, Z^\varepsilon)$. The unnormalized filter ρ^ϵ satisfies the Zakai equation:

$$d\rho_t^\varepsilon(\varphi) = \rho_t^\varepsilon(\mathcal{L}^\varepsilon\varphi)dt + \rho_t^\varepsilon(h\varphi)dY_t^\varepsilon, \quad \rho_0^\varepsilon(\varphi) = \mathbb{E}_{\mathbb{Q}}\left[\varphi(X_0^\varepsilon, Z_0^\varepsilon)\right], \tag{4}$$

where $\mathcal{L}^\varepsilon = \mathcal{L}_S + \frac{1}{\varepsilon}\mathcal{L}_F$, is the differential operator (generator) associated with the process $(X^\varepsilon, Z^\varepsilon)$. $\mathcal{L}_S = b(x,z)^*\nabla_x + \frac{1}{2}\text{tr}\left((\sigma\sigma^*)(x,z)\nabla_x^2\right)$ and $\mathcal{L}_F = f(x,z)^*\nabla_z + \frac{1}{2}\text{tr}\left((gg^*)(x,z)\nabla_z^2\right)$ are the generators of the slow and fast components, respectively, and * denotes the transpose. The solution to (4) can be approximated numerically by a sample of finite number of particles with error proportional to inverse of the sample size. This representation is the particle filter. For details on nonlinear filtering theory and particle approximation, the text [7] is a good reference.
In [4], a homogenized (unnormalized) filter ρ^0 is constructed to be the solution of

$$d\rho_t^0(\varphi) = \rho_t^0(\bar{\mathcal{L}}\varphi)dt + \rho_t^0(\bar{h}\varphi)dY_t^\varepsilon, \quad \rho_0^0(\varphi) = \mathbb{E}_{\mathbb{Q}}\left[\varphi(\bar{X}_0)\right], \tag{5}$$

where $\bar{h}$ is a suitably averaged version of h and $\bar{\mathcal{L}} \overset{\text{def}}{=} \bar{b}(x)^*\nabla_x + \frac{1}{2}\text{tr}\left((\bar{\sigma}\bar{\sigma}^*)(x)\nabla_x^2\right)$. In other words, ρ^0 is the unnormalized homogenized filter

$$\rho_t^0(\varphi) = \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\varphi(X_t^0)\tilde{D}_t^0\middle|\,\mathcal{Y}_t^\varepsilon\right], \quad \text{where} \quad \tilde{D}_t^0 \overset{\text{def}}{=} \exp\left(\int_0^t \bar{h}(X_s^0)^* dY_s^\epsilon - \frac{1}{2}\int_0^t \|\bar{h}(X_s^0)\|^2 ds\right),$$

constructed using the homogenized process X^0, but driven by the real observation Y^ε instead of a "homogenized observation" $\bar{Y}$. This is practical for implementation of the homogenized filter in applications since such homogenized observation is usually not available. Should it be available, using it may lead to loss of information for estimating the true signal X^ε. The corresponding normalized filter is $\pi_t^0(\varphi) = \frac{\rho_t^0(\varphi)}{\rho_t^0(1)}$.
In [4] it is proved that for any $T > 0$, the difference between the original filter and the filter for the coarse-grained dynamics constructed using (5) goes to zero as $\varepsilon \searrow 0$, i.e.,

$$\lim_{\epsilon\to 0}\mathbb{E}\left[d(\pi_T^{\epsilon,x}, \pi_T^0)\right] = 0, \tag{6}$$

where d denotes a suitable distance on the space of probability measures that generates the topology of weak convergence. In particle filtering approximation of the nonlinear filter, the particles live in the state space of the signal process and each particle represents a stochastic realization of the signal process. [8] showed that the number of particles required to suitably represent the signal's distribution scales exponentially with the dimension of the state. For the signal (2a), (2b), the ensemble size N_s needs to be greater than $\exp\{(m+n)^{\frac{1}{3}}\}$ to appropriately represent the density of $(X^\varepsilon, Z^\varepsilon)$ on $\mathbb{R}^{m+n}$, which is an issue when $m+n$ is large. This is the "curse of dimensionality". However, if we are only interested in estimating X^ε, the result (6) of [4] states that ρ^0 can be used, which requires $N_s > \exp\{m^{\frac{1}{3}}\}$, which may be more manageable. This is the basis for the henKF and HHPF of [3] and [5], respectively. The homoginized filters require computation of averaged coefficients $\bar{b}$, $\bar{\sigma}$ and $\bar{h}$, which can be accomplished either analytically when possible or using a multicale numerical integration technique (for example the technique of [9] used in [3] and [5]).

An optimal prior density

Due to the chaotic behavior of the Lorenz '96 system, small errors in estimation grows over time. This presents an issue when constructing the prior filtering density (density based on signal dynamics, prior to observations update) using particles, when observations are sparse in time, as estimation error from the last assimilation step grows at an exponential rate in between observations. In order to construct a better prior density for the HHPF, we introduce an additive "control" for each particle that steers it towards a location most representative of the next observation. Specifically, for each particle i, we solve a stochastic optimal control problem over each inter-observations time interval, where the control is determined by minimizing a cost function that is quadratic in the control energy and the distance between the particle's location and the next available observation.
We describe the stochastic optimal control problem in the single timescale setting to simplify notations. Consider the continuous-time signal with discrete-time observation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad X_0 = x \in \mathbb{R}^m, \tag{7a}$$

$$Y_{t_k} = h(X_{t_k}) + B_{t_k}, \qquad Y_0 = 0_{d\times 1}, \tag{7b}$$

where W is a standard k-dimensional $\mathbb{Q}$-Brownian motion as before B_{t_k} is a d-dimensional mean-zero Gaussian random variable with covariance R for observation times $\{t_k,\, t_k \geq 0\}$. In order to ensure particle trajectories do not deviate too far from the truth in between observations Y_{t_k} and $Y_{t_{k+1}}$, we introduce an additive "control" in the dynamics of each particle in $[t_k, t_{k+1}]$. This "control" steers it towards a location most representative of the observation $Y_{t_{k+1}}$. The dynamics of particle i is modified from (7) to be

$$d\hat{X}_t^i = \left(b(\hat{X}_t^i) + u_t^i\right)dt + \sigma(\hat{X}_t^i)dW_t, \qquad t \in (t_k, t_{k+1}). \tag{8}$$

The control u^i is chosen as the control u that minimizes the cost functional:

$$J(t_k, x, u) \stackrel{\text{def}}{=} \mathbb{E}_{t_k,x}\left[\frac{1}{2}\int_{t_k}^{t_{k+1}} u(s)^* Q(\hat{X}_s^i)^{-1} u(s)\, ds + g(Y_{t_{k+1}}, \hat{X}_{t_{k+1}}^i)\right], \tag{9}$$

where $Q(x) \stackrel{\text{def}}{=} \sigma\sigma^*(x)$ and $g(y,x) \stackrel{\text{def}}{=} \frac{1}{2}\left(y - h(x)\right)^* R^{-1}\left(y - h(x)\right)$. $E_{t_k,x}[\cdot]$ is expectation with respect to the probability measure of the process $\hat{X}^i$ that starts at x at time t_k. In other words, the control drives $\{\hat{X}_t^i, t \in [t_k, t_{k+1}]\}$ as close as possible to the location indicated by observation at time t_{k+1}, but with minimal effort in order to not overpower the actual signal dynamics. The latter is achieved by minimization of the running cost $\int_{t_k}^{t_{k+1}} u(s)^* Q(\hat{X}_s^i)^{-1} u(s)\, ds$ that is inversely proportional to the signal noise variance.

By standard optimal control technique, the value function $V(t,x) \stackrel{\text{def}}{=} \inf_u J(t,x,u)$ satisfies the Hamilton-Jacobi-Bellman (HJB) equation and the optimal control is $u(t,x) = -Q\,\nabla_x V(t,x)$. The HJB equation is

$$\frac{\partial V}{\partial t} + b(x)^* \nabla_x V + \frac{1}{2}\mathrm{tr}\left(Q\nabla_x^2 V\right) - \frac{1}{2}\nabla_x^* V Q \nabla_x V = 0, \qquad t \in [t_k, t_{k+1}], \qquad V(t_{k+1}, x) = g(Y_{t_{k+1}}, x). \tag{10}$$

The nonlinearity $\frac{1}{2}\nabla_x^* V Q \nabla_x V$ can be removed by employing a Cole-Hopf transformation as in [10, 11]: $V(t,x) = -\log\Phi(t,x)$. The optimal control becomes

$$u(t,x) = \frac{1}{\Phi(t,x)} Q \nabla_x \Phi(t,x), \tag{11}$$

where $\Phi(t,x)$ and $\nabla_x\Phi(t,x)$ satisfy linear second order PDEs respectively, which are obtained from successive differentiation of (10). The solutions $\Phi(t,x)$ and $\nabla_x\Phi(t,x)$ to the linear second order PDEs can be obtained by the Feynman-Kac formula, which results in the optimal control solution:

$$u(t,x) = -Q\mathbb{E}_{t,x}\left[\hat{w}_{t_{k+1}}(Y_{t_{k+1}}, \eta_{t_{k+1}}^{t,x})\, e^{\int_t^{t_{k+1}} \left(\nabla_x b(\eta_s^{t,x})\right)^* ds}\, \nabla_x g(Y_{t_{k+1}}, \eta_{t_{k+1}}^{t,x})\right], \tag{12}$$

$$\text{where} \qquad \hat{w}_{t_{k+1}}(Y_{t_{k+1}}, \eta_{t_{k+1}}^{t,x}) \stackrel{\text{def}}{=} \frac{e^{-g(Y_{t_{k+1}}, \eta_{t_{k+1}}^{t,x})}}{\mathbb{E}_{t,x}\left[e^{-g(Y_{t_{k+1}}, \eta_{t_{k+1}}^{t,x})}\right]}$$

and $\eta^{t,x}$ is the process that evolves according to

$$d\eta_s^{t,x} = b(\eta_s^{t,x})ds + \sigma(\eta_s^{t,x})d\tilde{W}_s, \qquad s \in [t, t_{k+1}], \quad \eta_t^{t,x} = x, \tag{13}$$

where $\tilde{W}$ is a standard Brownian motion. Given observation $Y_{t_{k+1}}$, the optimal control (12) steers a particle's trajectory in $[t_k, t_{k+1}]$ based on the expected deviation from $Y_{t_{k+1}}$ at time t_{k+1}. $\hat{w}_t(Y_{t_{k+1}}, \eta_{t_{k+1}}^{t,x})$ can be interpreted as the weight of the path $\{\eta_s^{t,x}; s \in [t, t_{k+1}]\}$ based on how well its t_{k+1} location agrees with observation $Y_{t_{k+1}}$. This result is the same as the optimal control obtained by applying the Clark-Ocone formula to the solution of the HJB equation (10) as in [12], when signal noise is additive, and the path integral solution of [13]. The prior density constructed using particles $\hat{X}_{t_{k+1}}^i$ is optimal in the sense that particles starting from the same location at time t_k have minimum weight variance.

Numerical experiments

For numerical experiments on the Lorenz '96 model, a timescale separation of $\varepsilon = 0.01$ is used. The fast- and slow-scale couplings are taken as $(h_x, h_z) = (-1, 10)$, while the slow-scale external forcing is set as $F_x = 10$. The true signal is taken as the $\mathbb{R}^{36+360}$ vector

$$\begin{aligned}&\left[[\mathbf{X}^\varepsilon]^* \quad [\mathbf{Z}^\varepsilon]^*\right]^* \\ &= \left[\begin{bmatrix}X^1 & \dots & X^{36}\end{bmatrix}^* \quad \begin{bmatrix}Z^{1,1} & \dots & Z^{1,10} & Z^{2,1} & \dots & Z^{2,10} & Z^{3,10} & \dots & Z^{36,10}\end{bmatrix}^*\right]^*,\end{aligned}$$

simulated according to (1). We have a multiscale system with additive noise of the form

$$\begin{aligned} d\mathbf{X}_t^\varepsilon &= b(\mathbf{X}_t^\varepsilon, \mathbf{Z}_t^\varepsilon)dt + \sigma_x d\mathbf{W}_t, \quad \mathbf{X}^\varepsilon \in \mathbb{R}^{36}, \\ d\mathbf{Z}_t^\varepsilon &= \frac{1}{\varepsilon} f(\mathbf{X}_t^\varepsilon, \mathbf{Z}_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}}\sigma_z d\mathbf{V}_t, \quad \mathbf{Z}^\varepsilon \in \mathbb{R}^{360}, \end{aligned}$$

where $\mathbf{W} \in \mathbb{R}^{36}$ and $\mathbf{V} \in \mathbb{R}^{360}$ are independent standard Brownian motions. The slow-scale signal noise covariance σ_x is taken as a sparse square matrix each with 1 on the diagonal and 0.05 on the first two sub- and super-diagonals. The fast-scale noise covariance σ_z is set similarly. Observation is taken as the complete slow-scale component $\mathbf{X}^\varepsilon$, with standard Gaussian noise, recorded at discrete timesteps:

$$Y_{t_k}^\varepsilon = H\begin{bmatrix}\mathbf{X}_{t_k}^\varepsilon \\ \mathbf{Z}_{t_k}^\varepsilon\end{bmatrix} + B_{t_k},$$

where $B_{t_k} \sim \mathcal{N}(0_{36\times 1}, I_{36\times 36})$ for all $k = 1, 2, \ldots$ and $H = I_{36\times 36}$. The time interval in between observations is equivalent to 1.5 days in real time. The deterministic version of (1) with $F_x = 10$ has error doubling time $\tau_d \approx 1.6$ days ([1]). Hence, observations intervals are approximately equal to the error doubling time.
We are interested only in the coarse-grained dynamics $\mathbf{X}$, hence we can apply the homogenization result described in the lower dimensional filtering section and use a reduced dimension (36-dimensional) filter in place of the full (36+360)-dimensional filter. The henKF and HHPF are implemented. For the homogenized filters, the signal is given by

$$d\mathbf{X}_t^0 = \bar{b}(\mathbf{X}_t^0)dt + \sigma_x d\mathbf{W}_t, \quad \mathbf{X}^0 \in \mathbb{R}^{36}, \tag{14}$$

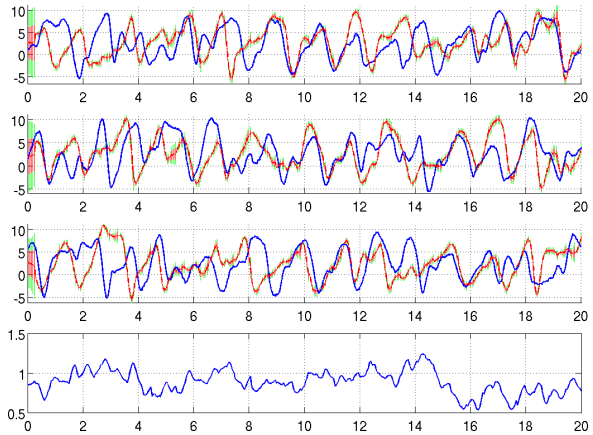
where the fast-scale effects in $\bar{b}$ have been averaged w.r.t. the stationary distribution of the fast component. Particles are sampled to represent the state of (14). Since there is only one time scale with no small parameter ε, numerical integration of the homogenized system can be performed using timestep of order 10^2 larger than that for the multiscale system. Numerical experiments are performed for an interval equivalent to 100 days in real time.
The following filters are implemented:

- sequential importance sampling particle filter (PF)
- homogenized hybrid particle filter (HHPF)
- homogenized hybrid particle filter with optimal nonlinear particle control (HHPF$_c$): nonlinear control using (12); if, for each particle, a sample of size n_s is used to approximate the expectation in (12), then we denote the corresponding filter by HHPF$_{c\,(n_s)}$.
- ensemble Kalman filter (enKF) with no homogenization
- homogenized ensemble Kalman filter (henKF); this is the same as the scheme in the wide timescale separation setting in [3]

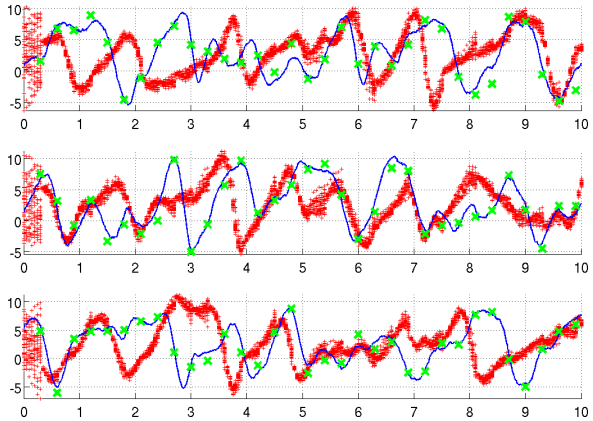
Each filter is implemented with 30 particles. For each particle in the HHPF$_c$, a sample of 30 particles is used to approximate the expectation in (12) for the optimal control, indicated in the subscript in HHPF$_{c\,(30)}$.
The filtering results of the HHPF, HHPF$_{c,(30)}$ and henKF for one experiment are shown in Figures 1, 2 and 3. Solid blue plots are the true states. In Figures (a), broken red plots with error bars are filter means with $\pm 1\times$sample standard deviations. Green error bars are $\pm 2\times$sample standard deviatons. The error plot is the time varying normalized RMSE

$$e_t \stackrel{\text{def}}{=} \frac{\|X_t^{\text{true}} - X_t^{\text{filter}}\|}{\|X_t^{\text{true}}\|}, \tag{15}$$

where X^{filter} is computed using the sample mean of the respective filters. In Figures (b), red crosses are particle trajectories. Solid green crosses in Figures (b) indicate observations. Observation noise is of the same order as signal noise, but noise amplitudes are small compared to the signal, hence observations are quite accurate.

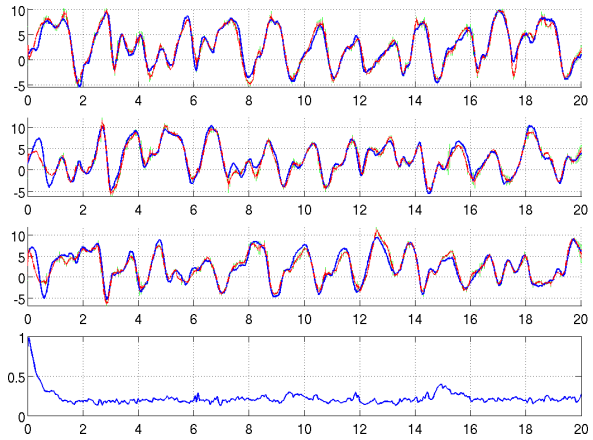


(a) (X^1, X^2, X^3)

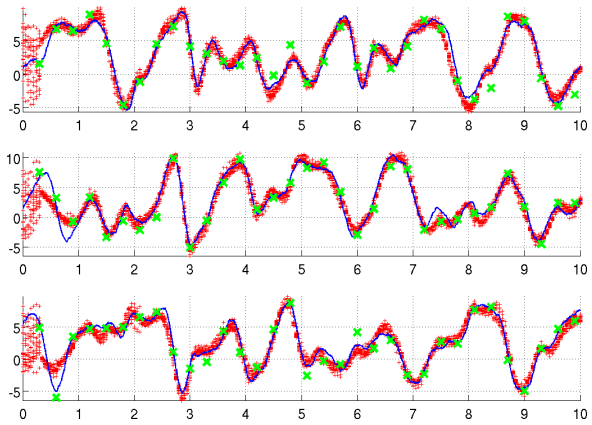


(b) (X^1, X^2, X^3) particles

Figure 1: HHPF

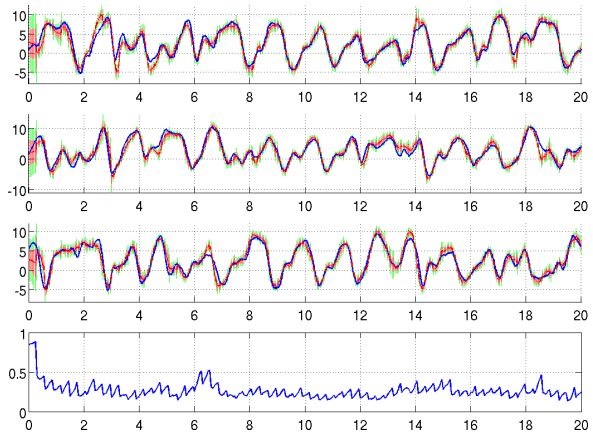


(a) (X^1, X^2, X^3)

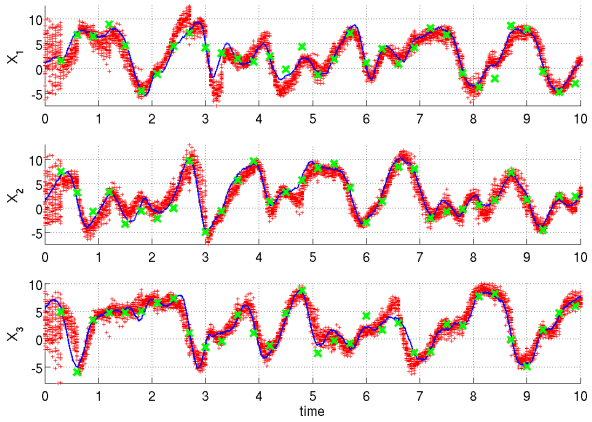


(b) (X^1, X^2, X^3) particles

Figure 2: $\text{HHPF}_{\text{c},(30)}$



(a) (X^1, X^2, X^3)



(b) (X^1, X^2, X^3) particles

Figure 3: henKF

The HHPF has reduced computation time compared to the PF. However, with sample size of 30, the PF and HHPF are found to be unable to track the true signal over the 100-day interval (Figure 1). Even with sample size increased to $\mathcal{O}(10^3)$, the HHPF does not do well. Applying the nonlinear optimal control to particles significantly improves the performance of the HHPF, as seen in Figure 2(a). The henKF and enKF are both able to track the true signal well, but the henKF is faster by taking advantage of homogenization. The advantage of the henKF and enKF over the standard particle filter is that, even though all particles are weighted equally, particle locations are corrected at each observation time based on observations (in the standard particle filter, particle locations are determined only by signal dynamics, only particle weights are changed by observations). The error plots of the henKF and enKF display more pronounced peaks, due to error growth in between observation times, which drop at observation times when new observations are assimilated. Error growth in between observation times are shown in the particle trajectories of the henKF in Figure 3(b). These error growths are more significant when observations frequency is decreased (Figure 5, where observations intervals is doubled to 3 days real time).

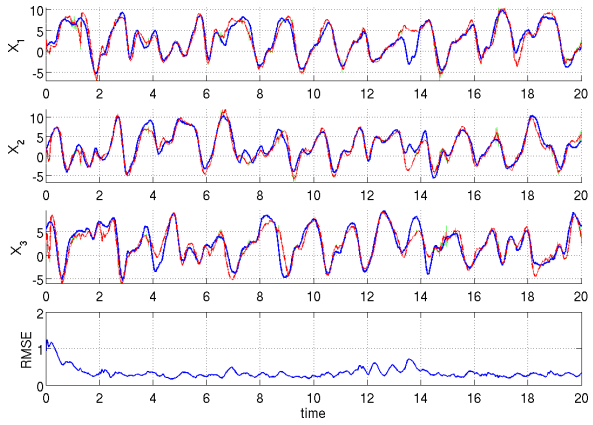
Table 1 shows the comparison of the filters over 40 independent experiments. The normalized RMSE (15) is integrated over time for each experiment and the average over 40 experiments are shown, along with averages of the RMSE at observation times and computational time.

The optimal control on particles enables the homogenized particle filter to track the true signal, and based on Table 1, the estimation is comparable to the homogenized and full ensemble Kalman filters. The optimal control of particles has a theoretically sound basis for application in nonlinear problems, as presented in the preceeding section on an optimal prior density. However, computation of the expectation in (12) incurs a high cost. By increasing N_s for the henKF, estimation error can be lowered with lower computational cost than the $\text{HHPF}_{\text{c}\,(30)}$ and enKF. It remains to be studied whether more efficient schemes can be found for computing (12). In the Lorenz '96 experiments, the HHPF_{c} is found to work well when the expectation in (12) is approximated with just one realization of the process η in (13), i.e. when the filter $\text{HHPF}_{\text{c}\,(1)}$ is used (see Table 1). In this case, the $\text{HHPF}_{\text{c}\,(1)}$ requires computational time comparable to the enKF, but the henKF is still the most efficient. Figures 4 and 5 show the $\text{HHPF}_{\text{c}\,(1)}$ and henKF when observations interval is doubled to 3 days. Error growth of the henKF particles within an inter-observations interval is more pronounced.

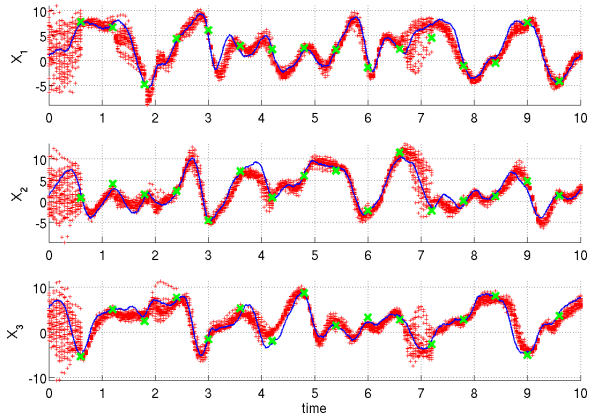
Filter	PF	HHPF	enKF	henKF
RMSE	18.974	19.140	5.252	5.432
$\text{RMSE}_{\text{obs times}}$	18.373	18.703	4.164	4.463
time	323 s	50 s	384 s	45 s

Filter	$\text{HHPF}^{2N_s}_{\text{c},(1)}$	$\text{HHPF}_{\text{c},(1)}$	$\text{HHPF}_{\text{c},(30)}$	Obs
RMSE	4.711	4.559	4.655	
$\text{RMSE}_{\text{obs times}}$	4,877	4.484	4.417	4.833
time	49 s	241 s	2705 s	

Table 1: RMSE integrated over time, RMSE at observations times integrated over time, and filter computation time, averaged over 40 experiments

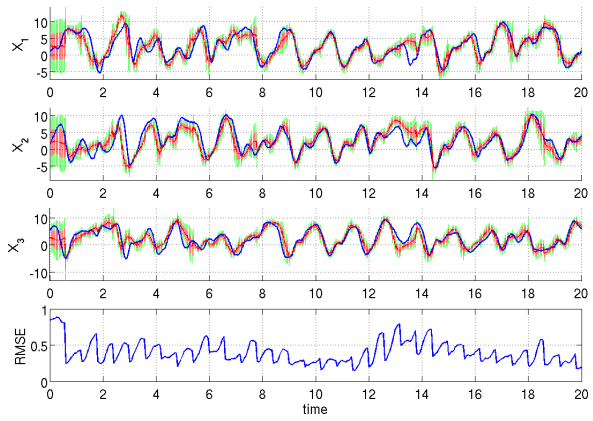


(a) (X^1, X^2, X^3)

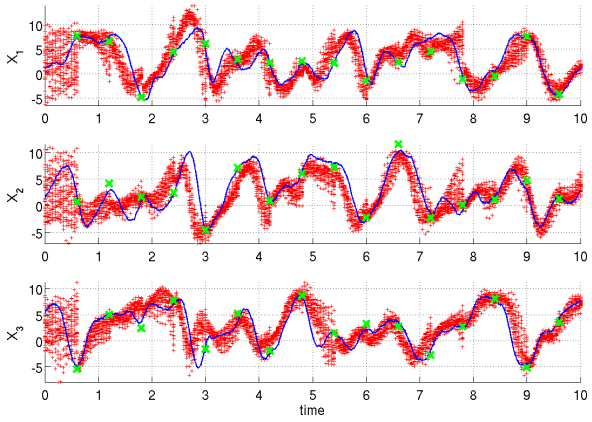


(b) (X^1, X^2, X^3) particles

Figure 4: $\text{HHPF}_{\text{c},(1)}$, observations every 3 days real time. Particle trajectories still remain close to true signal trajectory due to control



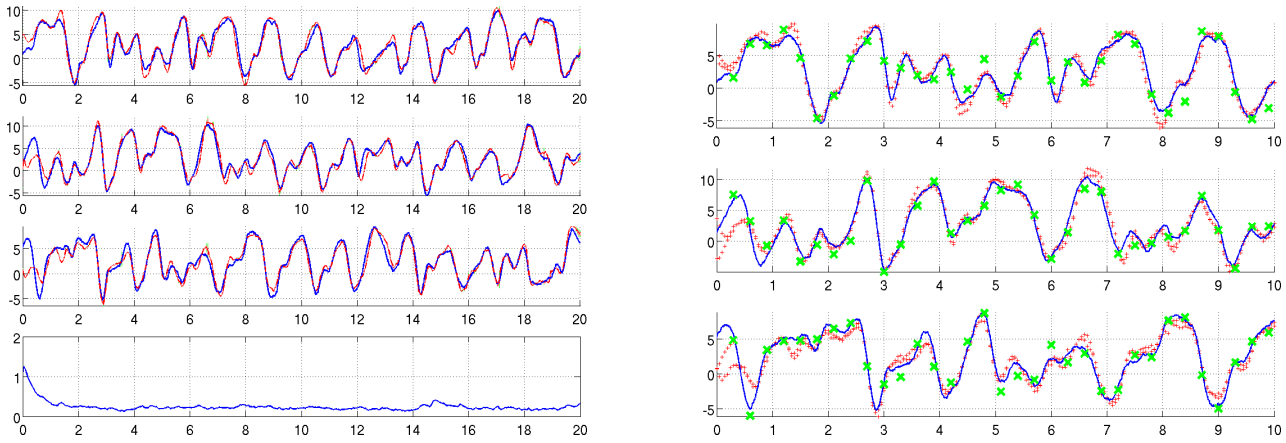
(a) (X^1, X^2, X^3)



(b) (X^1, X^2, X^3) particles

Figure 5: henKF, observations every 3 days real time. Particle trajectories deviate from true trajectory in between observation times

When observations are sufficiently accurate and signal noise is small, the optimal control ensures most of the particle filter sample members do not stray too far away from the truth at all times, so the sample mean provides a good estimate of the true signal. Since each particle is steered towards locations indicated by observations independently of the rest, we can expect a small sample size to be able to provide a good estimate of the true signal location. Indeed, Figure 6 shows the estimate from one experiment using a sample size of 2 ($\text{HHPF}^{2N_s}_{\text{c},(1)}$). The time-integrated RMSE and computation time are comparable to the henKF (see Table 1). The downside to steering particles is that the sample is not distributed well, tending to be clustered about a mean, albeit close to the truth. Accurate observations also result in most weight being concentrated on one particle that is closest to the truth (this is seen in all the particle filter variants). The number of particles with significant weight is small and often close to 1. As a result of these, the truth can fall outside $\pm 2\times$sample standard deviations from the mean, and higher order moments cannot be estimated. In further experiments, we find that the optimal control is still able to keep particles close to the truth when signal and observation noise amplitudes are large, but accuracy of the HHPF_{c} sample mean decreases and larger sample size is required, as one would expect.



(a) (X^1, X^2, X^3) (b) (X^1, X^2, X^3) particles

Figure 6: $\text{HHPF}_{\text{c},(1)}$ with sample size 2, observations every 36 hours

Although we have accurate estimate based on sample mean from a concentrated sample, we always desire a diverse sample in order to be able to properly capture the distribution of the true signal. A diferent form of optimal proposal density would be required, for example, similar to the scheme in [14, 15] that makes all particle weights almost equal at the observation step. However, when observations are accurate, the particles may again be concentrated. Different schemes remain to be studied, for example, schemes that utilize the existence of invariant manifolds.

Conclusions

The lower-dimensional filters are able to estimate the true slow-scale process using observations generated from the true signal, and are shown to have better computational efficiency compared to the multi-scale filter when applied on the Lorenz '96 model testbed. The lower-dimensional SIS PF is further adapted to address the chaotic nature of the model by constructing an optimal prior density that ensures particles do not stray too far from the locations indicated by observations and prevents particle weights from collapsing when observations are incorporated for constructing the posterior density.

Acknowledgements: The authors would like to acknowledge the support of the AFOSR under grant number FA9550-17-1-0001.

References

[1] Lorenz E.N. (1995) Predictability: A problem partly solved, *ECMWF Seminar Proceedings on Predictability*, **ECMWF**.

[2] Fatkullin I., Vanden-Eijnden E. (2004) A computational strategy for multiscale systems. *Journal of Computational Physics*, **Vol. 200** No. 2: 605–638.

[3] Harlim J., Kang, E.L. (2012) Filtering Partially Observed Multiscale Systems with Heterogeneous Multiscale Methods-Based Reduced Climate Models. *Monthly Weather Review*, **Vol. 140** No. 3: 860–873.

[4] Imkeller P., Sri Namachchivaya N., Perkowski N., Yeong H.C. (2013) Dimensional reduction in nonlinear filtering: A homogenization approach. *Annals of Applied Probability*, **Vol. 23** No. 5: 2290–2326.

[5] Park J.H., Sri Namachchivaya N., Yeong H. C. (2011) Particle Filters in a Multiscale Environment: Homogenized Hybrid Particle Filter. *Journal of Applied Mechanics*, **Vol. 78** No. 6: 1–10.

[6] Papanicolaou G.C., Stroock D., Varadhan S.R.S. (1977) Martingale approach to some limit theorems. *Papers from the Duke Turbulence Conference (Duke University, Durham, NC 1976), Paper No. 6*, Duke University.

[7] Bain A., Crisan D. (2009) Fundamentals of Stochastic Filtering. *Springer*, Berlin.

[8] Snyder C., Bengtsson T., Bickel P., Anderson J. (2008) Obstacles to high-dimensional particle filtering. **Monthly Weather Review**, **Vol. 136** No. 12: 4629–4640.

[9] E W., Liu D., Vanden-Eijnden E. (2005) Analysis of multiscale methods for stochastic differential equations. *Communications in Pure and Applied Mathematics*, **Vol. 58** No. 11: 1544–1585.

[10] Fleming W.H. (1978) Exit probabilities and optimal stochastic control. *Applied Mathematics and Optimization*, **Vol. 4** No. 1: 329–346.

[11] Fleming W.H. (1982) Logarithmic transformations and stochastic control. *Advances in Filtering and Optimal Stochastic Control: Proceedings of the IFIP-WG 7/1 Working Conference Cocoyoc, Mexico, February 1–6, 1982*: 131–141, Springer Berlin Heidelberg, Berlin, Heidelberg.

[12] Lingala N., Perkowski N., Yeong H.C., Sri Namachchivaya N., Rapti Z. (2014) Optimal nudging in particle filters. *Probabilistic Engineering Mechanics*, **Vol. 37** 160–169.

[13] Kappen H.J. (2005) Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics*, **Vol. 2005**: P11011.

[14] van Leeuwen P.J. (2010) Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, **Vol. 136** No. 653: 1991–1999.

[15] van Leeuwen P.J. (2011) Efficient nonlinear data-assimilation in geophysical fluid dynamics. *Computers and Fluids*, **Vol. 46** No. 1: 52–58.