Data-Driven Nonparametric Likelihood Functions: Continue

John Harlim
Department of Mathematics
Department of Meteorology & Atmospheric Science
Institute of CyberScience.
The Pennsylvania State University

March 20, 2019

Plan of the talk:

- Reproducing kernel with diffusion maps basis.
- A data assimilation application: An online estimation of observation model error.
- Recover missing dynamics.

Nonparametric likelihood functions

Recall that given $\varphi_j(\theta) \in L^2(\mathcal{M}, q)$ and $\psi_k(y) \in L^2(\mathcal{N}, \tilde{q})$ be the, respective, orthonormal bases. Then,

$$p(y|\theta) = \sum_{k,j} \varphi_j(\theta)[C_{Y\Theta} C_{\Theta\Theta}^{-1}]_{kj} \psi_k(y) \tilde{q}(y)$$

where,

$$\left[C_{Y\Theta}\right]_{jk} \approx \frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_j(y_i) \varphi_k(\theta_i),$$

$$\left[C_{\Theta\Theta}\right]_{jk} \approx \frac{1}{N} \sum_{i=1}^{N} \varphi_j(\theta_i) \varphi_k(\theta_i).$$

Remarks: In the following application, given training data $\{\theta_i, y_i\}_{i=1,\ldots,N}$, we will employ diffusion maps to construct basis functions $\varphi_j(\theta_i)$ and $\psi_k(y_i)$.

Biased observation model error problems in DA

The Kalman based DA formulation assumes unbiased observation model error, e.g.,

$$y_i = h(x_i) + \eta_i, \quad \eta_i \sim \mathcal{N}(0, R).$$

Suppose the operator h is un known. Instead, we are only given $\tilde{h}$, then

$$y_i = \tilde{h}(x_i) + b_i,$$

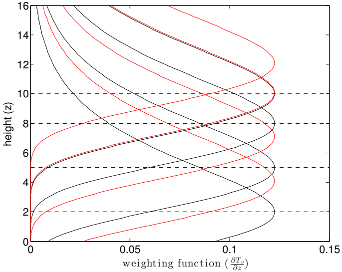
where we introduce a biased model error, $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i$.

Example: Basic radiative transfer model

Consider solutions of the stochastic cloud model[1], $\{T(z), \theta_{eb}, q, f_d, f_s, f_c\}$.
Based on this solutions, define a basic radiative transfer model as follows,

$$h_\nu(x) = \theta_{eb} T_\nu(0) + \int_0^\infty T(z) \frac{\partial T_\nu}{\partial z}(z)\, dz,$$

where T_ν is the transmission between heights z to ∞ that is defined to depend on q.
The weighting function, $\frac{\partial T_\nu}{\partial z}$ are defined as follows:



[1]Khouider, Biello, Majda 2010

Example: Basic radiative transfer model

Suppose the deep and stratiform cloud top height is $z_d = 12$km, while the cumulus cloud top height is $z_c = 3$km. Define $f = \{f_d, f_c, f_s\}$ and $x = \{T(z), \theta_{eb}, q\}$. Then the cloudy RTM is given by,

$$\begin{aligned} h_\nu(x,f) &= (1 - f_d - f_s)\left[\theta_{eb} T_\nu(0) + \int_0^{z_d} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\right] \\ &\quad + (f_d + f_s)T(z_t)T_\nu(z_d) + \int_{z_d}^{\infty} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz \end{aligned}$$

Example: Basic radiative transfer model

Suppose the deep and stratiform cloud top height is $z_d = 12$km, while the cumulus cloud top height is $z_c = 3$km. Define $f = \{f_d, f_c, f_s\}$ and $x = \{T(z), \theta_{eb}, q\}$. Then the cloudy RTM is given by,

$$
\begin{aligned}
h_\nu(x, f) &= (1 - f_d - f_s)\Big[\theta_{eb} T_\nu(0) + \int_0^{z_d} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\Big] \\
&\quad +(f_d + f_s)T(z_t)T_\nu(z_d) + \int_{z_d}^{\infty} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz \\
&= (1 - f_d - f_s)\Big[(1 - f_c)\Big(\theta_{eb} T_\nu(0) + \int_0^{z_c} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\Big) \\
&\quad +f_c T(z_c)T_\nu(z_c) + \int_{z_c}^{z_d} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\Big] \\
&\quad +(f_d + f_s)T(z_d)T_\nu(z_t) + \int_{z_d}^{\infty} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz
\end{aligned}
$$

Example: Basic radiative transfer model

Suppose the deep and stratiform cloud top height is $z_d = 12$km, while the cumulus cloud top height is $z_c = 3$km. Define $f = \{f_d, f_c, f_s\}$ and $x = \{T(z), \theta_{eb}, q\}$. Then the cloudy RTM is given by,

$$
\begin{aligned}
h_\nu(x,f) &= (1-f_d-f_s)\Big[\theta_{eb}T_\nu(0) + \int_0^{z_d} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\Big] \\
&\quad + (f_d+f_s)T(z_t)T_\nu(z_d) + \int_{z_d}^{\infty} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz \\
&= (1-f_d-f_s)\Big[(1-f_c)\Big(\theta_{eb}T_\nu(0) + \int_0^{z_c} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\Big) \\
&\quad + f_c T(z_c)T_\nu(z_c) + \int_{z_c}^{z_d} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\Big] \\
&\quad + (f_d+f_s)T(z_d)T_\nu(z_t) + \int_{z_d}^{\infty} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz
\end{aligned}
$$

One can check that $h_\nu(x,0)$ corresponds to cloud-free RTM.

Observation model error in data assimilation

Suppose the observations are generated with

$$y_\nu = h_\nu(x, f) + \eta, \qquad \eta \sim \mathcal{N}(0, R)$$

The difficulty in estimating the cloud fractions, cloud top heights and (in reality we don't know precisely how many clouds under a column) induces model error.

Observation model error in data assimilation

Suppose the observations are generated with

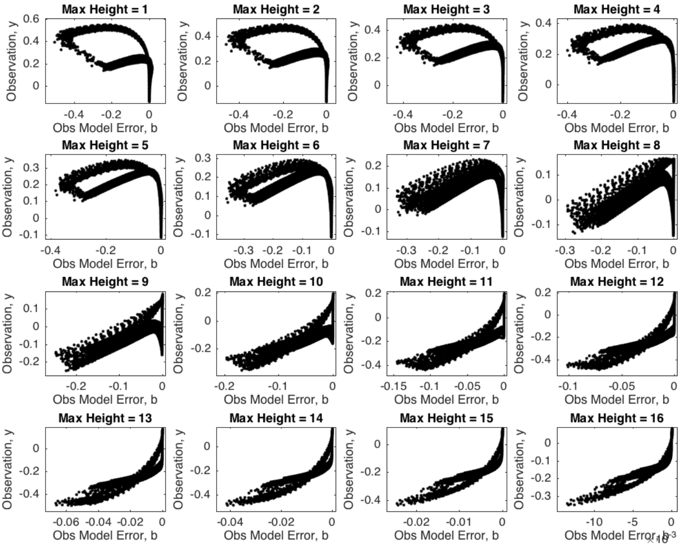
$$y_\nu = h_\nu(x, f) + \eta, \qquad \eta \sim \mathcal{N}(0, R)$$

The difficulty in estimating the cloud fractions, cloud top heights and (in reality we don't know precisely how many clouds under a column) induces model error.
In an extreme case, we consider filtering with a cloud-free RTM:

$$y_\nu = h_\nu(x, 0) + b_\nu$$

where $b_\nu = h_\nu(x, f) - h_\nu(x, 0) + \eta$ is model error with bias.

Observations (y_ν) v Model error (b_ν)



State estimation of the model error

We propose a secondary filter to estimate the statistics for b_i as follows:

$$\begin{tikzcd}
\begin{array}{c}\text{Prior}\\ p(x_i)\end{array} \arrow[r, "\text{Primary Filter}"] \arrow[d] & \begin{array}{c}\text{Posterior}\\ p(x_i \mid y_i)\end{array} \\
\begin{array}{c}\text{Error Prior}\\ p(b)\end{array} \arrow[r, "\text{Secondary Filter}"] & \begin{array}{c}\text{Error Posterior}\\ p(b \mid y_i)\end{array} \arrow[u] \\
\begin{array}{c}\text{Observation}\\ y_i\end{array} \arrow[r, "\text{RKHS+Training Data}"] \arrow[u] & \begin{array}{c}\text{Likelihood}\\ p(y_i \mid b)\end{array} \arrow[u]
\end{tikzcd}$$

We employ the RKHS theory to train a nonparametric likelihood function $p(y_i|b)$[2].

[2] Berry and H, Mon. Wea. Rev. 2017.

Recall that $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i = y_i - \tilde{x}_i$.

Given the prior ensemble $\{x_i^{b,k}\}_{k=1\ldots,K}$:

Recall that $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i = y_i - \tilde{x}_i$.

Given the prior ensemble $\{x_i^{b,k}\}_{k=1\dots,K}$:

- Compute $\bar{y}^b = \frac{1}{K}\sum_{k=1}^{K} \tilde{h}(x_i^{b,k})$.

Recall that $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i = y_i - \tilde{x}_i$.

Given the prior ensemble $\{x_i^{b,k}\}_{k=1\ldots,K}$:

- Compute $\bar{y}^b = \frac{1}{K}\sum_{k=1}^{K} \tilde{h}(x_i^{b,k})$.
- Define $Y_i = [\tilde{h}(x_i^{b,K}), \ldots, \tilde{h}(x_i^{b,K})]$ and $P_{yy,i} = \frac{1}{K-1} Y_i Y_i^\top$.

Recall that $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i = y_i - \tilde{x}_i$.

Given the prior ensemble $\{x_i^{b,k}\}_{k=1\ldots,K}$:

- Compute $\bar{y}^b = \frac{1}{K}\sum_{k=1}^{K} \tilde{h}(x_i^{b,k})$.
- Define $Y_i = [\tilde{h}(x_i^{b,K}), \ldots, \tilde{h}(x_i^{b,K})]$ and $P_{yy,i} = \frac{1}{K-1} Y_i Y_i^\top$.
- Assume that

$$p(b) = \mathcal{N}(y_i - \bar{y}_i^b, P_{yy,i} + R)$$

Recall that $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i = y_i - \tilde{x}_i$.

Given the prior ensemble $\{x_i^{b,k}\}_{k=1\ldots,K}$:

- Compute $\bar{y}^b = \frac{1}{K}\sum_{k=1}^{K} \tilde{h}(x_i^{b,k})$.
- Define $Y_i = [\tilde{h}(x_i^{b,K}), \ldots, \tilde{h}(x_i^{b,K})]$ and $P_{yy,i} = \frac{1}{K-1} Y_i Y_i^\top$.
- Assume that

$$p(b) = \mathcal{N}(y_i - \bar{y}_i^b, P_{yy,i} + R)$$

- Then apply the secondary Bayesian on training data set b_ℓ with nonparametric likelihood from RKHS.

$$p(b|y_i) \propto p(b)p(y_i|b)$$

Recall that $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i = y_i - \tilde{x}_i$.

Given the prior ensemble $\{x_i^{b,k}\}_{k=1\ldots,K}$:

- Compute $\bar{y}^b = \frac{1}{K}\sum_{k=1}^K \tilde{h}(x_i^{b,k})$.
- Define $Y_i = [\tilde{h}(x_i^{b,K}), \ldots, \tilde{h}(x_i^{b,K})]$ and $P_{yy,i} = \frac{1}{K-1} Y_i Y_i^\top$.
- Assume that

$$p(b) = \mathcal{N}(y_i - \bar{y}_i^b, P_{yy,i} + R)$$

- Then apply the secondary Bayesian on training data set b_ℓ with nonparametric likelihood from RKHS.

$$p(b|y_i) \propto p(b)p(y_i|b)$$

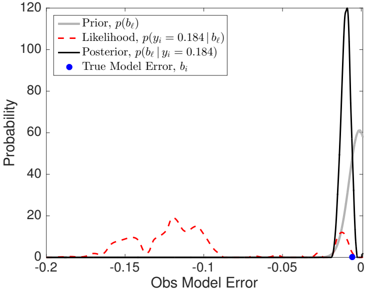
- Compute the mean and variance of the observation model error,

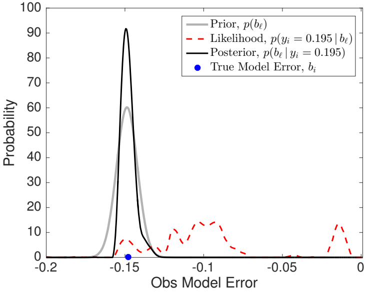
$$\begin{aligned} \hat{\mu}_{b_i} &= \mathbb{E}[b|y_i] \\ \hat{R}_{b_i} &= Var[b|y_i] \end{aligned}$$

  and use these terms to compensate for bias and variance of the observation model error in the primary filter.

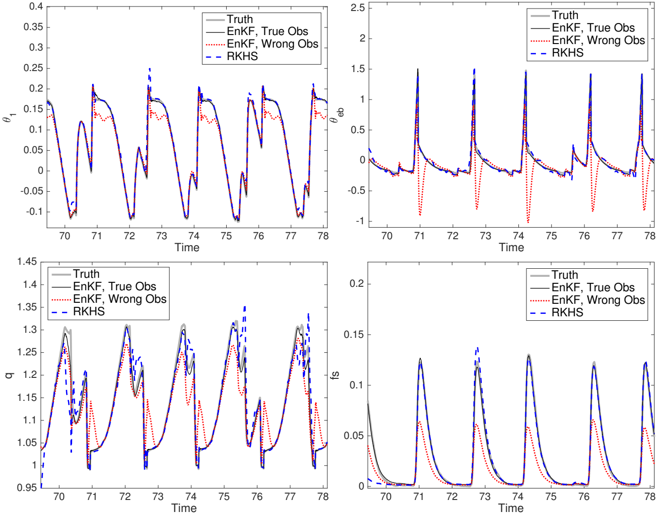
Secondary Bayesian filter

$$p(b|y_i) \propto p(b)p(y_i|b)$$



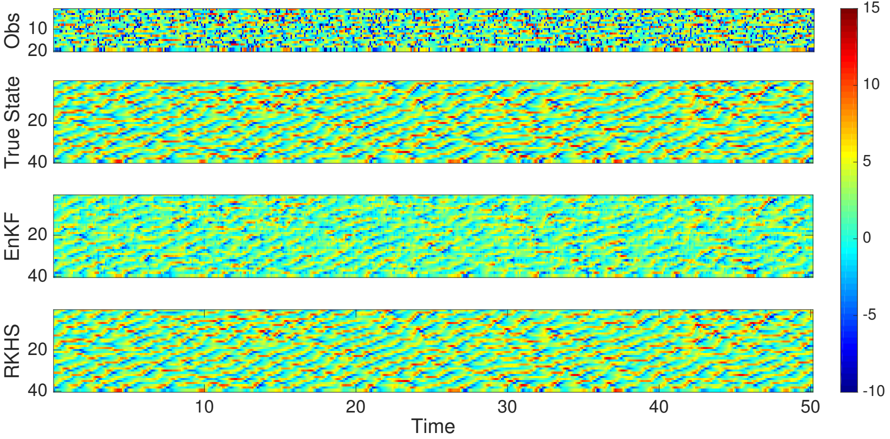


Filter estimates (with adaptive tuning of R and Q[3]).



[3] Berry and Sauer, Tellus A 2013

Corrupted observations in Lorenz-96 model



Sensitivity to observation noise variance and time

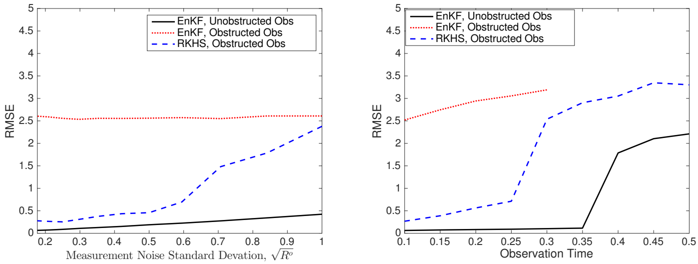


Figure : Left: RMSE as a function of the measurement error standard deviation, $\sqrt{R^o}$ for a fixed observation time $\Delta t = 0.1$. Right: RMSE as a function of observation time for $R^o = 2^{-5}$. Adaptive Q and R estimation method is implemented.

Averaging Theory[4]

Consider fast-slow systems with $\epsilon \ll 1$,

$$\frac{dx}{dt} = f(x,y), \quad x(0) = x_0 \tag{1}$$

$$\frac{dy}{dt} = \frac{1}{\epsilon} g(x,y), \quad y(0) = y_0. \tag{2}$$

[4] see e.g. Kurtz 1975, Papanicolaou 1976, Pavliotis and Stuart 2000.

Averaging Theory[4]

Consider fast-slow systems with $\epsilon \ll 1$,

$$\frac{dx}{dt} = f(x,y), \quad x(0) = x_0 \tag{1}$$

$$\frac{dy}{dt} = \frac{1}{\epsilon} g(x,y), \quad y(0) = y_0. \tag{2}$$

Assume that for fixed x, (2) is ergodic with invariant meas $\mu_x(dy)$.

[4] see e.g. Kurtz 1975, Papanicolaou 1976, Pavliotis and Stuart 2000.

Averaging Theory[4]

Consider fast-slow systems with $\epsilon \ll 1$,

$$\frac{dx}{dt} = f(x,y), \quad x(0) = x_0 \tag{1}$$

$$\frac{dy}{dt} = \frac{1}{\epsilon} g(x,y), \quad y(0) = y_0. \tag{2}$$

Assume that for fixed x, (2) is ergodic with invariant meas $\mu_x(dy)$. Then, as $\epsilon \to 0$, up to finite time $\mathcal{O}(1)$, $x(t)$ can be approximated by $X(t)$ that solves,

$$\frac{dX}{dt} = \bar{f}(X) = \int f(x,y)\mu_x(dy), \quad X(0) = x_0. \tag{3}$$

[4] see e.g. Kurtz 1975, Papanicolaou 1976, Pavliotis and Stuart 2000.

Averaging Theory[4]

Consider fast-slow systems with $\epsilon \ll 1$,

$$\begin{aligned} \frac{dx}{dt} &= f(x,y), \quad x(0) = x_0 &(1)\\ \frac{dy}{dt} &= \frac{1}{\epsilon} g(x,y), \quad y(0) = y_0. &(2) \end{aligned}$$

Assume that for fixed x, (2) is ergodic with invariant meas $\mu_x(dy)$. Then, as $\epsilon \to 0$, up to finite time $\mathcal{O}(1)$, $x(t)$ can be approximated by $X(t)$ that solves,

$$\frac{dX}{dt} = \bar{f}(X) = \int f(x,y)\mu_x(dy), \quad X(0) = x_0. \qquad (3)$$

Via ergodic theory, $\bar{f}(X) = \lim_{t\to\infty} \int_0^t f(X, z_x(s))\, ds$, where $z_x(s)$ is solution of (2) for fixed x.

[4] see e.g. Kurtz 1975, Papanicolaou 1976, Pavliotis and Stuart 2000.

The problem of missing dynamics

Suppose we have no knowledge of the fast equation

$$\frac{dx}{dt} = f(x,y), \quad x(0) = x_0$$
$$\frac{dy}{dt} = \frac{1}{\epsilon} g(x,y), \quad y(0) = y_0.$$

This means we have no access to samples $z_x(t)$. At the same time if we don't have explicit expression for $\mu_x(dy)$, then we can't compute $\bar{f}(x)$.

The problem of missing dynamics

Suppose we have no knowledge of the fast equation

$$\frac{dx}{dt} = f(x, y), \quad x(0) = x_0$$
$$\frac{dy}{dt} = \frac{1}{\epsilon} g(x, y), \quad y(0) = y_0.$$

This means we have no access to samples $z_x(t)$. At the same time if we don't have explicit expression for $\mu_x(dy)$, then we can't compute $\bar{f}(x)$.

Given these constraints, we propose to use data set $\{x_i, y_i\}$ to construct an RKHS conditional density, $\mu_x(dy) = p(y|x)\,dy$.

Preliminary results[6]

Consider Lorenz-96 model,

$$
\begin{aligned}
\frac{dx_k}{dt} &= x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F + B_k, \\
\frac{dy_{j,k}}{dt} &= \frac{1}{\epsilon}\left(y_{j+1,k}(y_{j-1,k} - y_{j+2,k}) - y_{j,k} + h_y x_k\right),
\end{aligned}
$$

where $B_k = \frac{h_x}{J}\sum Y_{j,k}$, $k = 1, \ldots, K, j = 1, \ldots, J$.
The variables, x_k and $y_{j,k}$ are periodic.

In our simulation, we fix the parameters
$K = 18, J = 20, F = 10, h_x = -1, h_y = 1$ [5].

In this case, we are missing the $KJ = 360$ dimensional of dynamical components y.

---

[5] Crommelin & Vanden-Eijnden, Mon. Wea. Rev. 2008

[6] joint work with Shixiao Jiang

Preliminary results[8]

- For this example, we apply RKHS theory on timeseries of $\{x_i, B_i\}$ to construct $p(B|x)$.
- The reduced model consists of solving the slow dynamics with $B_k(t)$ replaced by $\mathbb{E}_{B_k|x_k(t)}[B_k](t)$,

$$\frac{dx_k}{dt} = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F + \mathbb{E}_{B_k|x_k}[B_k].$$

- We will use VBDM and Hermite basis.
- We compare this to a parametric technique of Wilks[7].

[7] Wilks, Quart. J. of Roy. Meteor. Soc., 2005

[8] joint work with Shixiao Jiang

Case $\epsilon = 1/128$, we train $p(B_k(t)|x_k(t))$

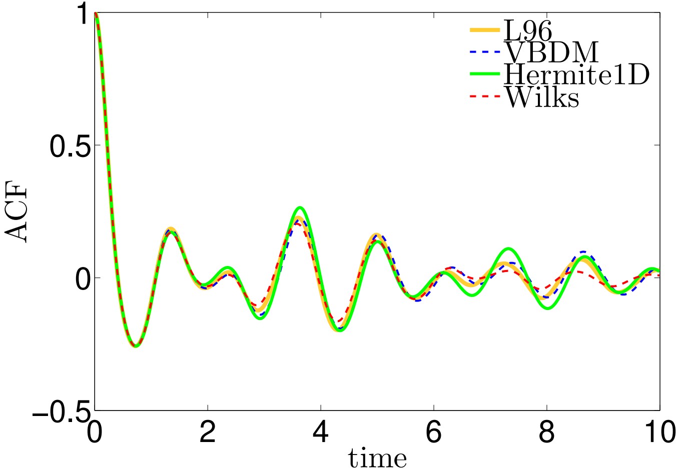


Figure : Autocorrelation function $\langle x_k(t)x_k(0)\rangle$

Case $\epsilon = 1/128$, we train $p(B_k(t)|x_k(t))$

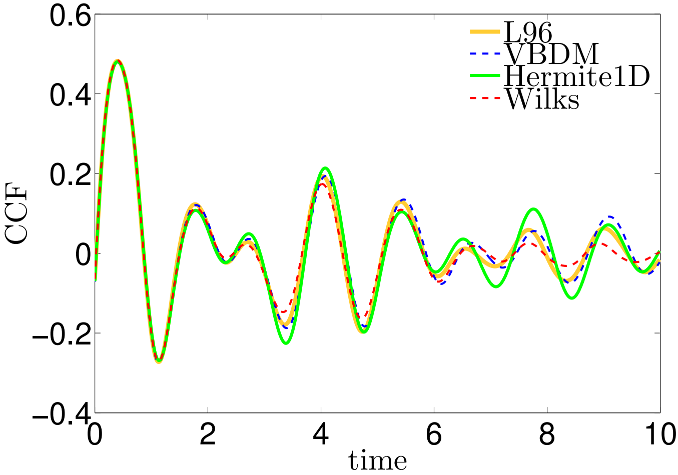


Figure : Cross-correlation function $\langle x_k(t)x_{k+1}(0)\rangle$

Case $\epsilon = 1/128$, we train $p(B_k(t)|x_k(t))$

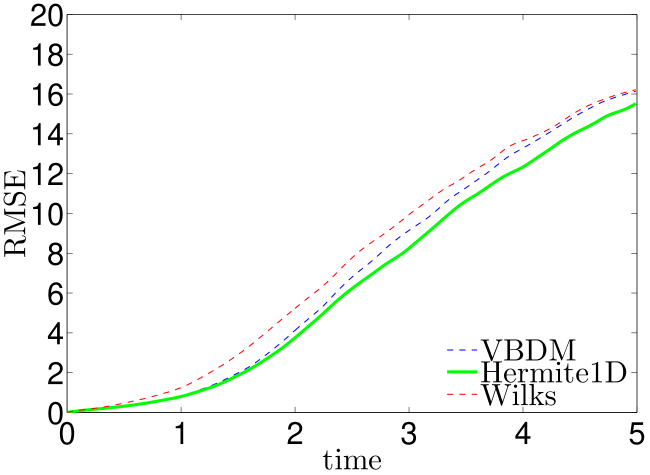


Figure : Lead forecast error (RMS)

Case $\epsilon = 1/2$, we train $p(B_k(t)|x_k(t), B_k(t-\Delta t))$

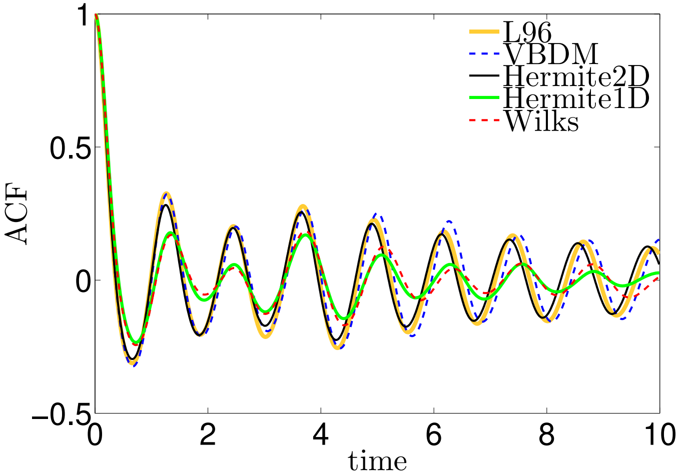


Figure : Autocorrelation function $\langle x_k(t) x_k(0) \rangle$

Case $\epsilon = 1/2$, we train $p(B_k(t)|x_k(t), B_k(t-\Delta t))$

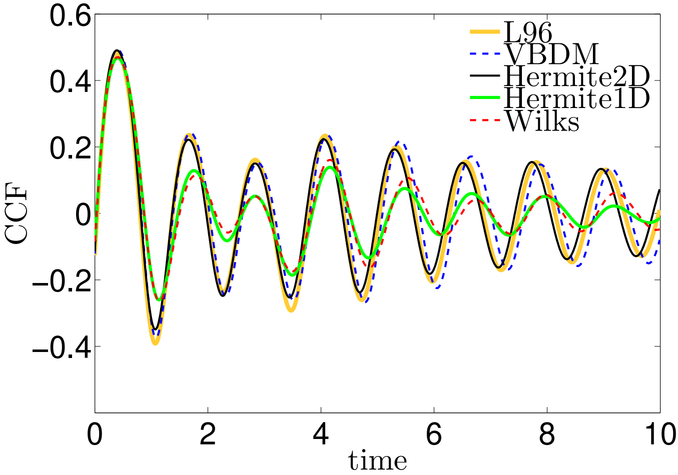


Figure : Cross-correlation function $\langle x_k(t) x_{k+1}(0) \rangle$

Case $\epsilon = 1/2$, we train $p(B_k(t)|x_k(t), B_k(t - \Delta t))$

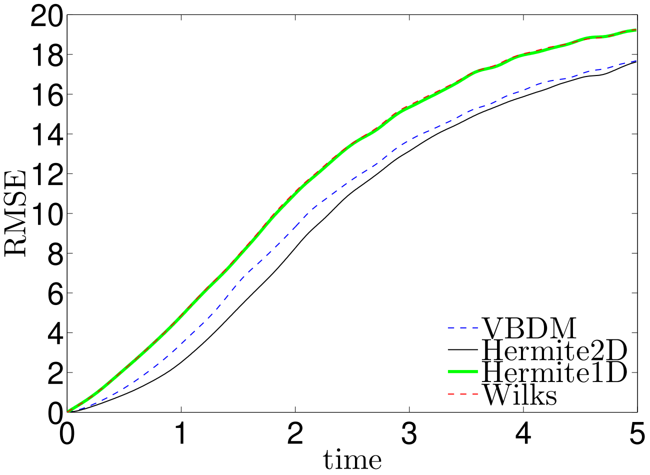
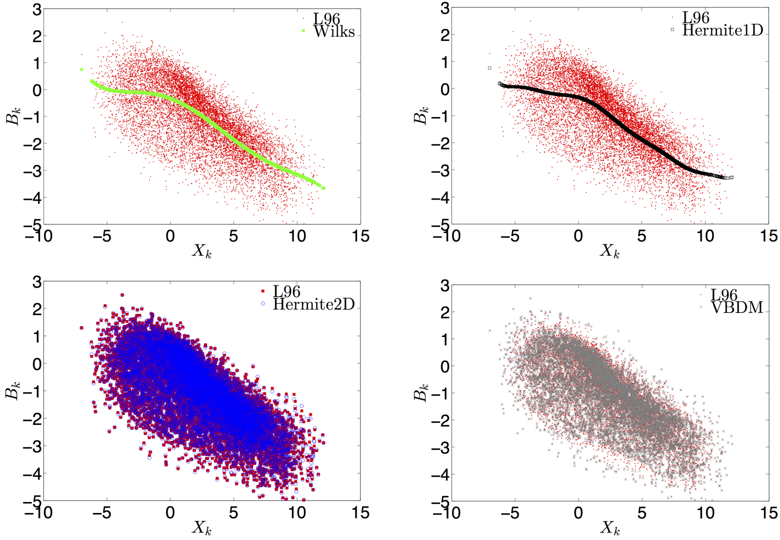


Figure : Lead forecast error (RMS)

Scatter plot of x_k vs B_k (and the estimated $\mathbb{E}[B_k]$)



References:

- S. Jiang and J. Harlim, *Parameter Estimation with Data-Driven Nonparametric Likelihood Functions*, arXiv:1804.03272.
- T. Berry and J. Harlim, *Correcting biased observation model error in data assimilation*, Mon Wea. Rev. 145(7), 2833-2853, 2017.

Collaborators:

- Shixiao Jiang, Postdoc at Department of Mathematics, The Pennsylvania State University.
- Tyrus Berry, Assistant Professor at Department of Mathematical Sciences, George Mason University.