Nonparametric modeling of dynamical systems

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

March 20, 2019

Plan of the talk

- Diffusion Forecast: Probabilistic prediction of deterministic dynamics and Itô diffusion processes.
- Data assimilation for specifying initial densities from observed data for diffusion forecast.

Probabilistic forecasting problem:

Suppose the variables of interest $x(t) \in \mathcal{M} \subset \mathbb{R}^n$ satisfy,

$$dx = a(x)dt + b(x)\, dW_t,$$

with distribution characterized by a density function $p(x,t)$ that satisfies a PDE called the Fokker-Planck equation,

$$\partial_t p = -\nabla \cdot (ap) + \frac{1}{2}\nabla \cdot \nabla \cdot (bb^\top p) \equiv \mathcal{L}^* p.$$

Probabilistic forecasting problem:

Suppose the variables of interest $x(t) \in \mathcal{M} \subset \mathbb{R}^n$ satisfy,

$$dx = a(x)dt + b(x)\,dW_t,$$

with distribution characterized by a density function $p(x,t)$ that satisfies a PDE called the Fokker-Planck equation,

$$\partial_t p = -\nabla \cdot (ap) + \frac{1}{2}\nabla \cdot \nabla \cdot (bb^\top p) \equiv \mathcal{L}^* p.$$

Probabilistic forecasting problem:
Given initial distribution $p(x,0) = p_0(x)$, one is interested to find

$$p(x,t) = e^{t\mathcal{L}^*} p_0(x)$$

and the corresponding statistics,

$$\mathbb{E}[f](t) \equiv \int_{\mathbb{R}^n} f(x)p(x,t)\,dx.$$

of some function f.

Probabilistic forecasting problem:

Classical solutions:

- If one knows $a(x), b(x), \mathcal{M}$, BC's, and IC's, then solve the Fokker-Planck equation with appropriate PDE solvers.

Probabilistic forecasting problem:

Classical solutions:

- If one knows $a(x), b(x), \mathcal{M}$, BC's, and IC's, then solve the Fokker-Planck equation with appropriate PDE solvers.
- For high dimensional application, apply Monte-Carlo (ensemble forecasting, see Epstein 1969, Leith 1974), i.e., Sample initial conditions $x^k \sim p_0(x)$ and solve an ensemble of initial value problems,

$$\begin{aligned} dx &= a(x)dt + b(x)\,dW_t, \\ x(0) &= x^k, \quad k = 1, \ldots, K. \end{aligned}$$

Suppose the ensemble solutions at time $t_i > 0$ is denoted by x_i^k, then one can compute the statistics via Monte-Carlo,

$$\mathbb{E}[f](t_i) \equiv \int_{\mathbb{R}^n} f(x)p(x,t)\,dx \approx \frac{1}{N}\sum_{k=1}^{N} f(x_i^k)$$

Nonparametric forecasting

Problem: Given only a time series of of $x_i = x(t_i), i = 1, \ldots, N$, solve the Fokker-Planck equation for $p(x, t)$. In other words, $\mathcal{L}^*$, $a(x)$, $b(x)$, $\mathcal{M}$, and BC's are all unknown.

Nonparametric forecasting

Problem: Given only a time series of of $x_i = x(t_i), i = 1, \ldots, N$, solve the Fokker-Planck equation for $p(x,t)$. In other words, $\mathcal{L}^*$, $a(x)$, $b(x)$, $\mathcal{M}$, and BC's are all unknown.

Basically, we want to solve a partial differential equation without knowing the equation!

Nonparametric forecasting

Problem: Given only a time series of of $x_i = x(t_i), i = 1, \ldots, N$, solve the Fokker-Planck equation for $p(x,t)$. In other words, $\mathcal{L}^*$, $a(x)$, $b(x)$, $\mathcal{M}$, and BC's are all unknown.

Basically, we want to solve a partial differential equation without knowing the equation!

Assumption: The dynamics are ergodic so $\mathcal{M}$ is the attractor of the system and the sampling measure is the same as the invariant measure. That is, $x_i \sim p_{eq}(x)$, where $\mathcal{L}^* p_{eq} = 0$.

Nonparametric forecasting

Problem: Given only a time series of of $x_i = x(t_i), i = 1, \ldots, N$, solve the Fokker-Planck equation for $p(x,t)$. In other words, $\mathcal{L}^*$, $a(x)$, $b(x)$, $\mathcal{M}$, and BC's are all unknown.

Basically, we want to solve a partial differential equation without knowing the equation!

Assumption: The dynamics are ergodic so $\mathcal{M}$ is the attractor of the system and the sampling measure is the same as the invariant measure. That is, $x_i \sim p_{eq}(x)$, where $\mathcal{L}^* p_{eq} = 0$.

Remark: Our approach is nonparametric in the sense that we do not impose any parametric form in modeling $a(x)$ and $b(x)$. Surely, the method has parameters.

Review of Galerkin method

Had we know the PDE,

$$p(x, t+\tau) = e^{\tau\mathcal{L}^*} p(x,t),$$

one can solve this problem with a Galerkin method. That is, pick a basis function $\psi_j(x)$ depending on the geometry and represent the solutions of the PDE as linear combinations of these basis functions,

$$p(x, t+\tau) = \sum_j c_j(t+\tau)\psi_j(x),$$

then solve the system of ODE's,

$$c_k(t+\tau) = \sum_j \langle \psi_k, e^{\tau\mathcal{L}^*}\psi_j\rangle c_j(t),$$

under the appropriate inner product.

Review of Galerkin method

Had we know the PDE,

$$p(x, t+\tau) = e^{\tau\mathcal{L}^*} p(x, t),$$

one can solve this problem with a Galerkin method. That is, pick a basis function $\psi_j(x)$ depending on the geometry and represent the solutions of the PDE as linear combinations of these basis functions,

$$p(x, t+\tau) = \sum_j c_j(t+\tau)\psi_j(x),$$

then solve the system of ODE's,

$$c_k(t+\tau) = \sum_j \langle \psi_k, e^{\tau\mathcal{L}^*}\psi_j\rangle c_j(t),$$

under the appropriate inner product.

Problem: We don't have $\psi_j(x)$ and we don't know $\mathcal{L}^*$

Recall of the diffusion maps

- We use the diffusion maps[1],[2] to generate a data driven basis function φ_j. This kernel based method uses the data $x_i \sim p_{eq}(x)$ to approximate eigenfunctions of a weighted Laplacian,

$$\hat{\mathcal{L}} = p_{eq}^{-1}\mathrm{div}(p_{eq}\nabla \quad) = \nabla \log(p_{eq}) \cdot \nabla + \Delta$$

which is the generator of a gradient flow with isotropic diffusion,

$$dx = \nabla \log(p_{eq})\, dt + \sqrt{2}dW_t.$$

- One can check that the eigenfunctions of the adjoint operator $\hat{\mathcal{L}}^*$ are $\varphi_j(x)p_{eq}(x)$, which forms an orthonormal basis for $L(\mathcal{M}, p_{eq}^{-1})$ assuming that $\hat{\mathcal{L}}$ has pure point spectra.

[1] Coifman & Lafon, Appl. Comput. Harmon. Anal., 2006

[2] Berry & H, Appl. Comput. Harmon. Anal., 2016.

Diffusion Forecast

With these data-driven basis functions, we represent the solutions of the Fokker-Planck equation as,

$$p(x, t+\tau) = \sum_k c_k(t+\tau)\varphi_k(x)p_{eq}(x),$$

where

$$\begin{aligned} c_k(t+\tau) &= \sum_j \langle \varphi_k p_{eq}, e^{\tau\mathcal{L}^*}\varphi_j p_{eq}\rangle_{p_{eq}^{-1}} c_j(t), \\ &= \sum_j \langle e^{\tau\mathcal{L}}\varphi_k, \varphi_j\rangle_{p_{eq}} c_j(t), \end{aligned}$$

Q: How do you estimate $\langle e^{\tau\mathcal{L}}\varphi_k, \varphi_j\rangle_{p_{eq}}$?

Gradient Flow System

If the underlying dynamical system is indeed a gradient flow with isotropic diffusion with invariant density,

$$p_{eq}(x) = \exp(-U(x)/D)$$

where $D > 0$ is a constant, then diffusion maps will approximate,

$$\hat{\mathcal{L}} = \nabla \log(p_{eq}) \cdot \nabla + \Delta = D^{-1} \nabla U \cdot \nabla + \Delta.$$

and our goal is to approximate the generator

$$\mathcal{L} = D\hat{\mathcal{L}} = \nabla U \cdot \nabla + D\Delta.$$

HW for students: Estimate D from timeseries?

Gradient Flow System[3]

If $\{\varphi_k\}$ denote a set of orthonormal basis of $\hat{\mathcal{L}}$ under $L^2(\mathcal{M}, p_{eq})$, it is then immediate to see that,

$$e^{\tau\mathcal{L}}\varphi_k = e^{\lambda_k D t}\varphi_k,$$

which means that,

$$\langle e^{\tau\mathcal{L}}\varphi_k, \varphi_j\rangle_{p_{eq}} = e^{\lambda_k D\tau}\delta_{k,j}.$$

The solution of the Fokker-Planck is given as,

$$\begin{aligned}
p(x, t+\tau) &= \sum_k c_k(t+\tau)\varphi_k(x)p_{eq}(x) \\
&= \sum_k \sum_j \langle e^{\tau\mathcal{L}}\varphi_k, \varphi_j\rangle_{p_{eq}} c_j(t)\varphi_k(x)p_{eq}(x) \\
&= \sum_k e^{\lambda_k D\tau} c_k(t)\varphi_k(x)p_{eq}(x)
\end{aligned}$$

[3]Berry and H, SIAM J. Uncertainty Quantification, 2015.

Example

Given data set of $x(t)$ that solves the following chaotic oscillator,

$$
\begin{aligned}
\dot{x} &= x - x^3 + \frac{\gamma}{\epsilon} y_2, \\
\dot{y}_1 &= \frac{10}{\epsilon^2}(y_2 - y_1), \\
\dot{y}_2 &= \frac{1}{\epsilon^2}(28 y_1 - y_2 - y_1 y_3), \\
\dot{y}_3 &= \frac{1}{\epsilon^2}(y_1 y_2 - \frac{8}{3} y_3).
\end{aligned}
$$

It is well known that for a sufficiently small γ, ϵ, the dynamics can be approximated by[4],

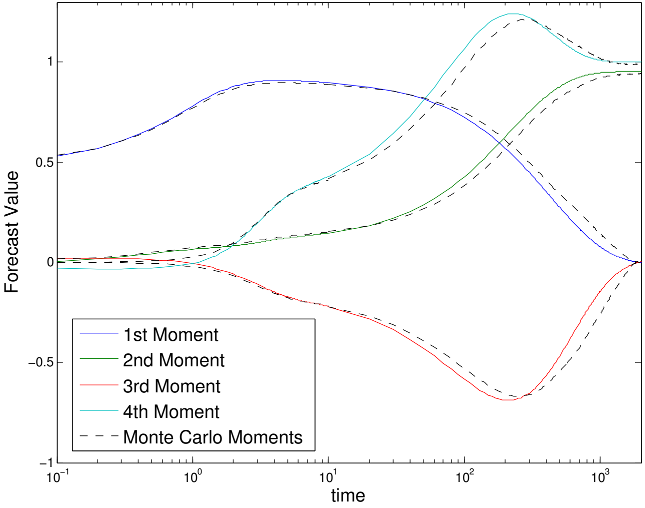
$$dX = X(1 - X^2)\, dt + \sigma\, dW_t.$$

[4] Givon, Kupferman, & Stuart 2006

Evolution of the density

(Loading Video...)

Forecasting moments



Bayesian filtering problem

Consider now a dynamical system with continuous-time noisy observations

$$
\begin{aligned}
dx &= -\nabla U(x)\, dt + \sqrt{2D}\, dW, \\
dz &= h(x)\, dt + \sqrt{R} dW_z.
\end{aligned}
$$

Bayesian filtering problem

Consider now a dynamical system with continuous-time noisy observations

$$\begin{aligned} dx &= -\nabla U(x)\, dt + \sqrt{2D}\, dW, \\ dz &= h(x)\, dt + \sqrt{R} dW_z. \end{aligned}$$

The filtering problem is to obtain the following conditional distribution,

$$p(x|z) \propto p(x)p(z|x)$$

and its statistics. Indeed, the un-normalized conditional density, $r(x,t) = p(x,t) \int_{\mathcal{M}} r(x,t) dV(x)$, solves a linear SPDE known as the Zakai equation,

$$dr = \mathcal{L}^* r\, dt + r h^\top R^{-1} dz.$$

Bayesian filtering problem

Consider now a dynamical system with continuous-time noisy observations

$$\begin{aligned} dx &= -\nabla U(x)\, dt + \sqrt{2D}\, dW, \\ dz &= h(x)\, dt + \sqrt{R} dW_z. \end{aligned}$$

The filtering problem is to obtain the following conditional distribution,

$$p(x|z) \propto p(x)p(z|x)$$

and its statistics. Indeed, the un-normalized conditional density, $r(x,t) = p(x,t) \int_{\mathcal{M}} r(x,t) dV(x)$, solves a linear SPDE known as the Zakai equation,

$$dr = \mathcal{L}^* r\, dt + r h^\top R^{-1} dz.$$

Key idea: project this SPDE into the diffusion coordinates.

Evolution of the posterior density

Example filtering the double-well potential with observation function $h(x) = (x - 0.05)^2$.

(Loading Video...)

Evolution of the posterior density

Example filtering the double-well potential with observation function $h(x) = x^2$.

(Loading Video...)

Response Problem (climate sensitivity)

Given data set $x_i \sim p_{eq}(x)$ of the current climate (or equilibrium state) and an external perturbation, δU, on the unknown potential U, compute the response of the statistics as functions of time,

$$\delta\mathbb{E}[A(x)](t) = \mathbb{E}_{p^\delta}[A(x)](t) - \mathbb{E}_{p_{eq}}[A(x)], \quad t \geq 0.$$

for functional A.

Response Problem (climate sensitivity)

Given data set $x_i \sim p_{eq}(x)$ of the current climate (or equilibrium state) and an external perturbation, δU, on the unknown potential U, compute the response of the statistics as functions of time,

$$\delta\mathbb{E}[A(x)](t) = \mathbb{E}_{p^\delta}[A(x)](t) - \mathbb{E}_{p_{eq}}[A(x)], \quad t \geq 0.$$

for functional A.

Remark: If the dynamical models for the unperturbed system are known, given the functional of the perturbation, one can just apply Monte-Carlo.

Response Problem (climate sensitivity)

Recall the variable bandwidth diffusion kernel[5], we prove that our discrete diffusion maps with variable bandwidth kernel $K_\epsilon(x,y) = \exp(-\frac{||x-y||^2}{\epsilon\rho(x)\rho(y)})$ produces

$$L_\epsilon f = \Delta f + 2(1-\alpha)\nabla f \cdot \frac{\nabla q}{q} + (d+2)\nabla f \cdot \frac{\nabla \rho}{\rho} + \mathcal{O}(\epsilon)$$

The key idea is to apply diffusion maps on the data set $x_i \sim p_{eq}(x)$ with the variable bandwith kernel with bandwidth function $\rho(x) = p_{eq}(x)^{-1/2} e^{-\frac{\delta U(x)}{D(d+2)}}$. This allows us to learn the generator corresponding to the perturbed system,

$$dx = -\nabla(U(x) + \delta U(x))\, dt + \sqrt{2D} dW.$$

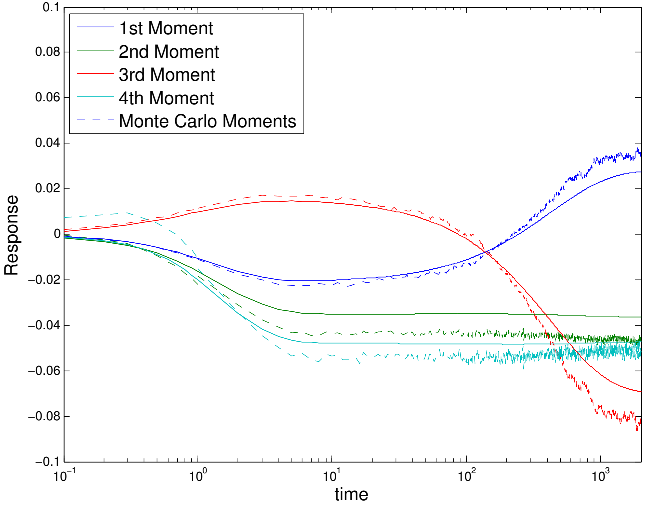
[5] Berry & H, Appl. Harmon. Comput. Anal., 2016

Response of the density due to an external forcing

(Loading Video...)

Response statistics

Evolution of the moments compared to those of the stochastic Monte-Carlo simulations.



General Itô diffusion processes: $\hat{\mathcal{L}} \neq c\mathcal{L}$ for any $c \in \mathbb{R}$

The Dynkin formula states that for any fcn $f(x) \in C^2(\mathcal{M})$ on compact manifold, the solutions of the backward Kolmogorov of Ito diffusion,

$$u_t = \mathcal{L}u, \quad u(x, t_i) = f(x_i),$$

can be expressed as

$$e^{\tau\mathcal{L}} f(x_i) = \mathbb{E}_{x_i}[f(x_{i+1})],$$

where $t_{i+1} = t_i + \tau$.

General Itô diffusion processes: $\hat{\mathcal{L}} \neq c\mathcal{L}$ for any $c \in \mathbb{R}$

The Dynkin formula states that for any fcn $f(x) \in C^2(\mathcal{M})$ on compact manifold, the solutions of the backward Kolmogorov of Ito diffusion,

$$u_t = \mathcal{L}u, \quad u(x, t_i) = f(x_i),$$

can be expressed as

$$e^{\tau\mathcal{L}} f(x_i) = \mathbb{E}_{x_i}[f(x_{i+1})],$$

where $t_{i+1} = t_i + \tau$.
If we define a shift operator $S_\tau f(x_i) = f(x_{i+1})$, then

$$e^{\tau\mathcal{L}} f(x_i) = \mathbb{E}_{x_i}[S_\tau f(x_i)]$$

which gives a hint that we can approximate $e^{\tau\mathcal{L}}$ with S_τ.

Diffusion Forecast

We[6] approximate $e^{\tau\mathcal{L}}$ with a shift operator S_τ defined as follows: $S_\tau f(x_i) = f(x_{i+1})$, where $t_{i+1} = t_1 + \tau$. Numerically,

$$\begin{aligned}
\langle e^{\tau\mathcal{L}}\varphi_k, \varphi_j\rangle_{p_{eq}} &\approx \langle S_\tau\varphi_k, \varphi_j\rangle_{p_{eq}} \approx \frac{1}{N}\sum_{i=1}^{N} S_\tau\varphi_k(x_i)\varphi_j(x_i) \\
&= \frac{1}{N}\sum_{i=1}^{N} \varphi_k(x_{i+1})\varphi_j(x_i),
\end{aligned}$$

where $\{x_i\}_{i=1}^N \sim p_{eq}(x)$.

[6] Berry, Giannakis, and H, Phys. Rev. E 2015.

Diffusion Forecast

We[6] approximate $e^{\tau\mathcal{L}}$ with a shift operator S_τ defined as follows: $S_\tau f(x_i) = f(x_{i+1})$, where $t_{i+1} = t_1 + \tau$. Numerically,

$$\begin{aligned} \langle e^{\tau\mathcal{L}}\varphi_k, \varphi_j\rangle_{p_{eq}} &\approx \langle S_\tau\varphi_k, \varphi_j\rangle_{p_{eq}} \approx \frac{1}{N}\sum_{i=1}^{N} S_\tau\varphi_k(x_i)\varphi_j(x_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}\varphi_k(x_{i+1})\varphi_j(x_i), \end{aligned}$$

where $\{x_i\}_{i=1}^N \sim p_{eq}(x)$.
So all we need is the basis function φ_j which we learn from the diffusion maps and set

$$p(x,t+\tau) \approx \hat{p}(x,t+\tau) := \sum_{k=0}^{M}\sum_{j=0}^{M}\langle S_\tau\varphi_k, \varphi_j\rangle_{p_{eq}} c_j(t)\varphi_k(x)p_{eq}(x),$$

[6] Berry, Giannakis, and H, Phys. Rev. E 2015.

Error estimate[7]

Theorem

Let $p(x,\tau) \in L^2(\mathcal{M}, p_{eq}^{-1})$ for all $t \geq 0$ be the solutions of FP eqn with bounded diffusion tensor, b, and assume that $\mathcal{M}$ is compact. Let $\hat{p}(x,\tau)$ be the approximate solutions from the diffusion forecast method. Then there exists sufficiently large $M > 0$ and constant $C > 0$ such that,

$$\mathbb{E}\Big[\|(p(\cdot,\tau) - \hat{p}(\cdot,\tau))^2\|^2_{p_{eq}^{-1}}\Big] \leq C\frac{\tau}{N}.$$

[7] H, Cambridge University Press 2018 (see Chapter 6).

Lorenz-63

Lorenz model:

$$\dot{x} = \sigma(y - x), \quad \dot{y} = x(\rho - z) - y, \quad \dot{z} = xy - \beta z.$$

(Loading Video...)

A 2D fast-slow stochastic system in a torus

Learning from a three-dimensional data $(x(\theta,\phi), y(\theta,\phi), z(\theta,\phi))$ whereas the intrinsic dynamical system is two dimensional (θ,ϕ).

(Loading Video...)

Specifying initial conditions given new observations

If the observation y is not noisy, $p(x) = \delta_y(x)$. Then the initial coefficients are given as,

$$c_k(0) = \langle p, \varphi_k \rangle = \varphi_k(y).$$

So all we need to do is to extend the diffusion maps basis on new data point y. This can be done using the Nystrom extension[8]

[8]Yang & H, J. Nonlinear Science, 2018.

[9]Berry & H, Physica D, 2016.

Specifying initial conditions given new observations

If the observation y is not noisy, $p(x) = \delta_y(x)$. Then the initial coefficients are given as,

$$c_k(0) = \langle p, \varphi_k \rangle = \varphi_k(y).$$

So all we need to do is to extend the diffusion maps basis on new data point y. This can be done using the Nystrom extension[8]

For discrete-time noisy observations y with likelihood function $p(y|x)$, we can apply the Bayes's theorem[9] to estimate initial density,

$$p(x|y) \propto p(x)p(y|x),$$

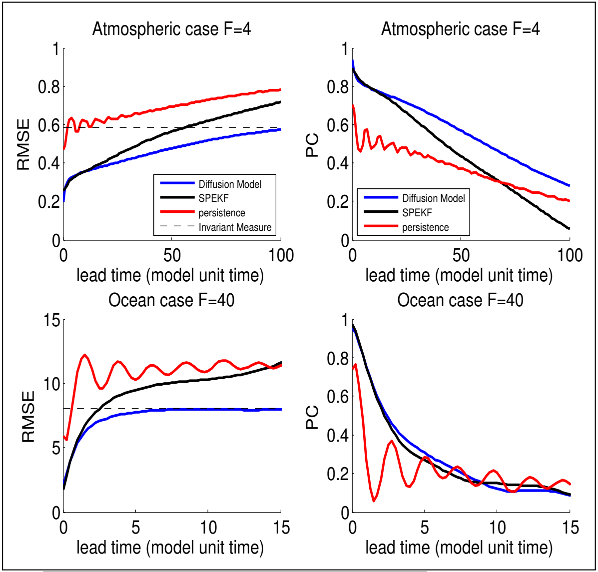
where the prior densities are from the diffusion forecast,

$$p(x) = \sum_j c_j(t)\varphi_j(x)p_{eq}(x)$$

[8]Yang & H, J. Nonlinear Science, 2018.

[9]Berry & H, Physica D, 2016.

Forecasting barotropic modes of QG turbulence[10]



Model: 2-layer QG with baroclinic instabilities. Parameter $F \propto L_D^{-2}$.

Diffusion forecasting is trained on noisy Fourier modes corresponding to 36 spatially uniform grid points. Training is performed on 5000 noisy data points (with time-delay embedding) while the forecast verification is done on separate 4000 data points.

We apply the Bayesian filter to initialize the forecast densities.

We compare it to:

- SPEKF: A stochastic parametric modeling approach with additive and multiplicative noises.
- persistent forecast: Tomorrow's forecast is exactly equal to today observation.

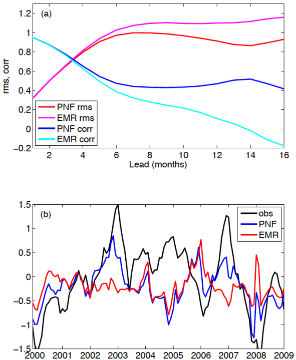
[10] Berry & H, Physica D, 2016.

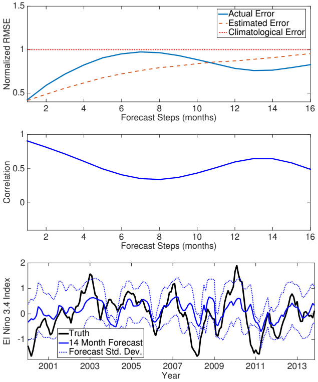
Application: Forecasting the El Nino Index 3.4

Left: Taken from comment to our paper by Kondrashov, Chekroun, & Ghil Phys. Rev. E, 2016. They published their modeling approach in PNAS 2011.

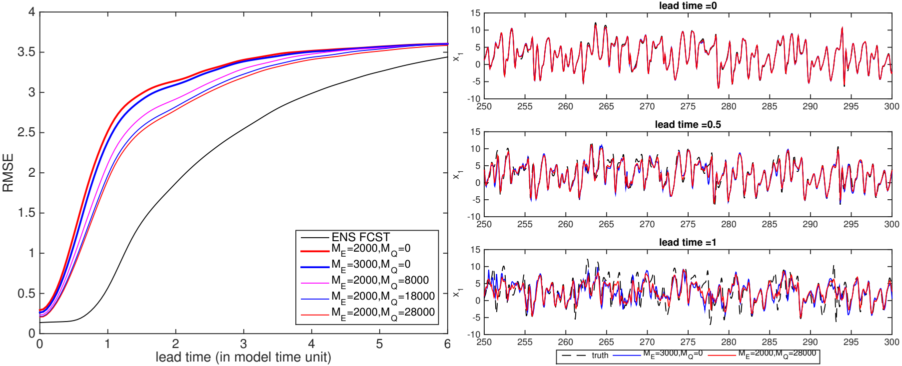
Right: Diffusion forecast is trained on only 600 data point (monthly between Jan 1950-Dec 1999. Forecast verification on Jan 2000-march 2014. Berry, Giannakis, & H Phys. Rev. E 2016.

14-month lead-time forecast skill:
PNF RMSE 0.86, PC 0.52
Diffusion Forecast RMSE 0.77, PC 0.64.





Example: Lorenz-96 6D [11]



Note: Training data $N = 10^6$. There are still plenty of room for improvement for high-dimensional problems.

[11]Yang and H, J. Nonlinear Science, 2018.

References:

- J. Harlim and H. Yang, *Diffusion forecasting model with basis functions from QR decomposition*, J. Nonlinear Sci 28, 847-872, 2018.
- T. Berry and J. Harlim, *Forecasting turbulent modes with nonparametric models: Learning from noisy data*, Physica D 320, 57-76, 2016.
- T. Berry and J. Harlim, *Nonparametric uncertainty quantification for stochastic gradient flows*, SIAM J Uncertainty Quantification 3(1), 484-508, 2015.
- T. Berry and J. Harlim, *Nonparametric forecasting of low-dimensional dynamical systems*, Physical Review E 91, 032915, 2015.
- J. Harlim, *Data-driven computational methods: Parameter and operator estimations*, Cambridge University Press, UK, 2018. (with supplementary MATLAB codes for diffusion forecast)

Collaborators:

- Tyrus Berry, Dept. of Mathematical Sciences, George Mason University.
- Dimitrios Giannakis, Courant Institute, New York University.
- Shixiao Jiang, Dept. of Mathematics, The Pennsylvania State University.
- Haizhao Yang, Dept. of Mathematics, National University of Singapore.