Diffusion Forecast: A nonparametric modeling approach

John Harlim[1]

Department of Mathematics
Department of Meteorology and Atmospheric Science
The Pennsylvania State University

Collaborators: Tyrus Berry (PSU), Dimitrios Giannakis (Courant Institute)

May 24, 2016

[1]Supports: Office of Naval Research and the National Science Foundation

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.

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...)

Review of Galerkin method

Had we know the PDE,

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

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

$$p(x,t) = \sum_j c_j(t)\varphi_j(x),$$

then solve the system of ODE's,

$$c_k(t) = \sum_j \langle \varphi_k, e^{t\mathcal{L}^*}\varphi_j \rangle c_j(0),$$

under the appropriate inner product.

Review of Galerkin method

Had we know the PDE,

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

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

$$p(x,t) = \sum_j c_j(t)\varphi_j(x),$$

then solve the system of ODE's,

$$c_k(t) = \sum_j \langle \varphi_k, e^{t\mathcal{L}^*}\varphi_j \rangle c_j(0),$$

under the appropriate inner product.

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

Diffusion Forecast

- We use the diffusion maps[2],[3] to generate a data driven basis function φ_j. This is a kernel based method!
- We[4] approximate $e^{\tau\mathcal{L}}$ with a shift operator S_τ defined as follows: $S_\tau f(x_i) = f(x_{i+1})$, where $t_{i+1} = t_1 + \tau$. In this coordinate basis, we can approximate

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

where $\{x_i\}_{i=1}^N \sim p_{eq}(x)$ is the training data time series.

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

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

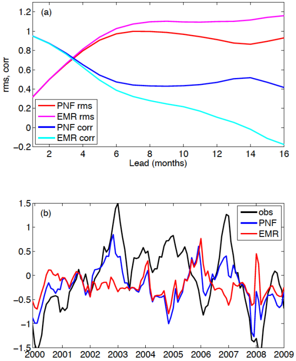
[4] Berry, Giannakis, and H, Phys. Rev. E 2015.

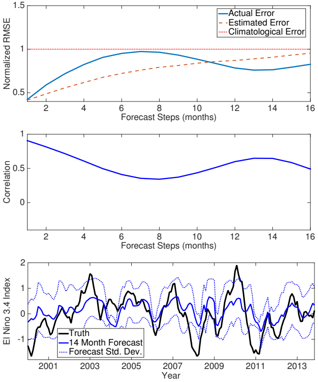
Application: Forecasting the El Nino Index 3.4

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.





Application: Modeling missing dynamics [5]

Consider forecasting problem,

$$\dot{x} = f(x, \theta),$$

with unknown time evolving $\dot{\theta} = g(\theta, \dot{W})$. We assume that besides missing the dynamics of $\theta \in \mathcal{M}$, we only measure noisy observations at discrete time,

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

For illustration, we consider coupling L96 with an L63 as follows,

$$\begin{aligned}
\dot{x}_i &= \theta x_{i-1}x_{i+1} - x_{i-1}x_{i-2} - x_i + 8, \\
\theta &= (a_1/40 + 1), \\
\dot{a}_1 &= 10(a_2 - a_1)/\epsilon, \\
\dot{a}_2 &= (28a_1 - a_2 - a_1 a_3)/\epsilon, \\
\dot{a}_3 &= (a_1 a_2 - 8a_3/3)/\epsilon,
\end{aligned}$$

[5] Berry & H, J Comput. Phys., 2016.

Numerical solutions

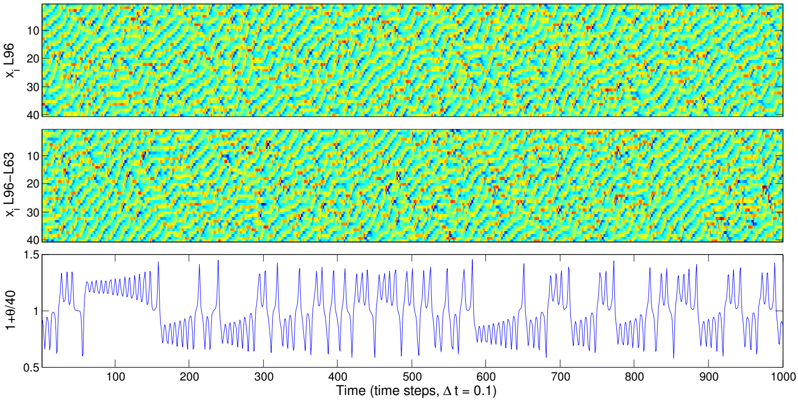


Figure : Comparison of the uncoupled Lorenz-96 (top) model dynamics with the coupled L96-L63 system (middle) and the coefficient $(\theta/40+1)$ (bottom).

Extracting training data

Extract training data set for θ from observations y.

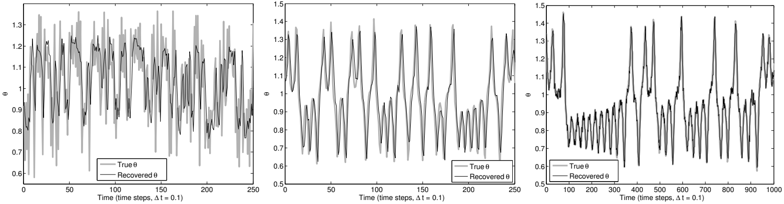


Figure : Comparison the true time series of θ and the recovered time series for $\epsilon = 0.25$ (left), $\epsilon = 1$ (middle), and $\epsilon = 4$ (right).

Here we used the adaptive method[6] to solve

$$\begin{aligned} \dot{x} &= f(x, \theta) \\ \dot{\theta} &= \sqrt{Q_{\theta\theta}}\,\dot{W} \end{aligned}$$

[6] Berry & Sauer, Tellus A 2013

Numerical results

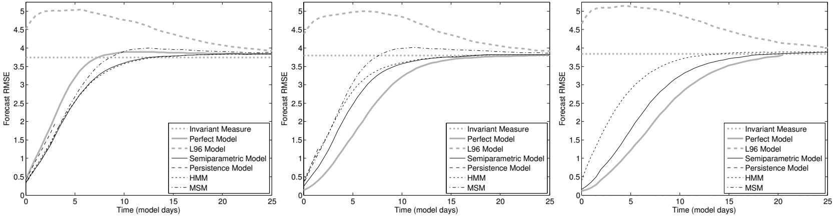


Figure : Comparison of forecasting methods with recovered training data and filtered initial condition for $\epsilon = 0.25$ (left), $\epsilon = 1$ (middle), and $\epsilon = 4$ (right).

We consider: standard L96 model, persistence model, fitting MSM model, HMM (sampling from recovered θ), perfect model.

References:

- Berry & H, “Variable Bandwidth Diffusion Kernels”, Appl. Comput. Harmon. Anal. 40, 68-96, 2016.
- Berry & H, “Nonparametric Uncertainty Quantification for Stochastic Gradient Flows”, SIAM J. Uncertainty Quantification 3(1), 484-508, 2015.
- Berry, Giannakis, & H, “Nonparametric Forecasting of Low-dimensional Dynamical Systems”, Phys. Rev. E 91 032915, 2015.
- Berry & H, “Forecasting Turbulent Modes with Nonparametric Diffusion Models: Learning from Noisy Data”, Physica D, 57-76, 2016.
- Berry & H, “Semiparametric modeling: Correcting low-dimensional model error in parametric models”, J. Comput. Phys., 308, 305-321, 2016.
- Berry, Giannakis, & H, Reply to “Comments on ‘Nonparametric Forecasting of Low-dimensional Dynamical Systems’ ”, Phys. Rev. E 93 036202, 2016.