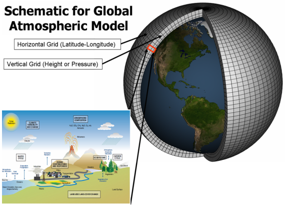
Data assimilation with stochastic model reduction

Fei Lu

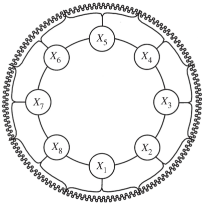
Department of Mathematics, Johns Hopkins

Joint with: Alexandre J. Chorin (UC Berkeley)
Kevin K. Lin (U. of Arizona)
Xuemin Tu (U. of Kansas)

Nov. 21, 2017
Banff International Research Station



ECMWF: 16 km horizontal grid $\to 10^9$ freedoms



The Lorenz 96 system

Wilks 2005

High-dimensional Full system + **Discrete partial data** → **Prediction**

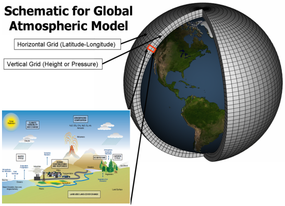
$x' = f(x) + U(x, y),$
$y' = g(x, y).$

Observe only $\{x(nh)\}_{n=1}^{N}$.

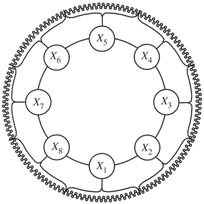
Forecast $x(t), t \geq Nh$.

- Complex full systems:
  - can only afford to resolve $x' = f(x)$
  - y: unresolved variables (subgrid-scales)
- Discrete partial observations: missing i.c.
- Ensemble prediction: need many simulations

Motivation: weather/climate prediction



ECMWF: 16 km horizontal grid $\to 10^9$ freedoms



The Lorenz 96 system

Wilks 2005

High-dimensional Full system + **Discrete partial data** - - → **Prediction**

$x' = f(x) + U(x, y),$
$y' = g(x, y).$

Observe only $\{x(nh)\}_{n=1}^{N}$.

Forecast $x(t), t \geq Nh$.

- Complex full systems:
  - can only afford to resolve $x' = f(x)$
  - y: unresolved variables (subgrid-scales)
- Discrete partial observations: missing i.c.
- Ensemble prediction: need many simulations

$\to$ Develop a reduced model that

- quantifies the model error $U(x, y)$
- captures key statistical + dynamical properties

Outline

- Stochastic model reduction
  (A first step: reduction from simulated data)
  - Discrete-time stochastic parametrization (NARMA)
  - Application to chaotic dynamical systems
- Data assimilation with the reduced model
  (An intermediate step: NARMA + noisy data $\rightarrow$ state estimation and prediction)

Stochastic model reduction

$x' = f(x) + U(x,y), \; y' = g(x,y).$
Data $\{x(nh)\}_{n=1}^{N}$

Memory effects (Mori, Zwanzig, Chorin, Ghil, Majda, Wilks, . . .)

- Takens Theorem: delay embedding
- Mori-Zwanzig formalism: “generalized Langevin equation”

$$\frac{dx}{dt} = \underbrace{f(x)}_{\text{Markov term}} + \underbrace{\int_0^t K(x(t-s),s)ds}_{\text{memory}} + \underbrace{\dot{W}_t}_{\text{noise}},$$

Goal: a non-Markovian stochastic reduced system for x

Differential system or discrete-time system?

$X' = f(X) + Z(t,\omega)$ | $X_{n+1} = X_n + R_h(X_n) + Z_n$

informative, neat | messy

Inference[1] | likelihood

Discretization[2] | error correction by data

---

[1] Talay, Mattingly, Stuart, Higham, Milstein,Tretyakov, . . .

[2] Brockwell, Sørensen, Pokern, Wiberg, Samson,. . .

Discrete-time stochastic parametrization

NARMA(p, q)

$$X_n = X_{n-1} + R_h(X_{n-1}) + Z_n,$$

$$Z_n = \Phi_n + \xi_n,$$

$$\Phi_n = \underbrace{\sum_{j=1}^{p} a_j X_{n-j} + \sum_{j=1}^{r}\sum_{i=1}^{s} b_{i,j} P_i(X_{n-j})}_{\text{Auto-Regression}} + \underbrace{\sum_{j=1}^{q} c_j \xi_{n-j}}_{\text{Moving Average}}$$

- $R_h(X_{n-1})$ from a numerical scheme for $x' \approx f(x)$
- Φ_n depends on the past

Tasks:

Structure derivation: terms and orders (p, r, s, q) in Φ_n

– physical laws, asymptotic behavior, discretization

Parameter estimation: a_j, $b_{i,j}$, c_j, and σ

– conditional likelihood methods

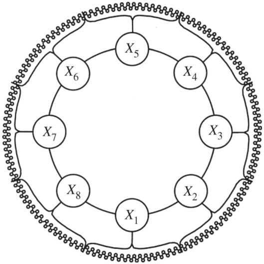
Example I: the Lorenz 96 system

A chaotic dynamical system (a simplified atmospheric model)

$$\frac{d}{dt}x_k = x_{k-1}\left(x_{k+1} - x_{k-2}\right) - x_k + 10 - \frac{1}{J}\sum_j y_{k,j},$$

$$\frac{d}{dt}y_{k,j} = \frac{1}{\varepsilon}[y_{k,j+1}(y_{k,j-1} - y_{k,j+2}) - y_{k,j} + x_k],$$

where $x \in \mathbb{R}^{18}$, $y \in \mathbb{R}^{360}$. $\epsilon = 0.5 \to$ no scale separation.



Wilks 2005

Find a reduced system for $x \in \mathbb{R}^{18}$ based on

- Data $\{x(nh)\}_{n=1}^{N}$
- $\frac{d}{dt}x_k \approx x_{k-1}\left(x_{k+1} - x_{k-2}\right) - x_k + 10.$

- NARMA:

$$
\begin{aligned}
x^n &= x^{n-1} + R_h(x^{n-1}) + z^n; \; z^n = \Phi^n + \xi^n, \\
\Phi^n &= a + \sum_{j=1}^{p} \sum_{l=1}^{d_x} b_{j,l} (x^{n-j})^l + \sum_{j=1}^{p} c_j R_h(x^{n-j}) + \sum_{j=1}^{q} d_j \xi^{n-j}.
\end{aligned}
$$

$$
p = 2, d_x = 3; \; q = \begin{cases} 1, & h = 0.01; \\ 0, & h = 0.05. \end{cases}
$$

- **NARMA:**

$$
\begin{aligned}
x^n &= x^{n-1} + R_h(x^{n-1}) + z^n; \; z^n = \Phi^n + \xi^n, \\
\Phi^n &= a + \sum_{j=1}^{p} \sum_{l=1}^{d_x} b_{j,l} (x^{n-j})^l + \sum_{j=1}^{p} c_j R_h(x^{n-j}) + \sum_{j=1}^{q} d_j \xi^{n-j}.
\end{aligned}
$$

$$
p = 2, d_x = 3; \; q = \begin{cases} 1, & h = 0.01; \\ 0, & h = 0.05. \end{cases}
$$

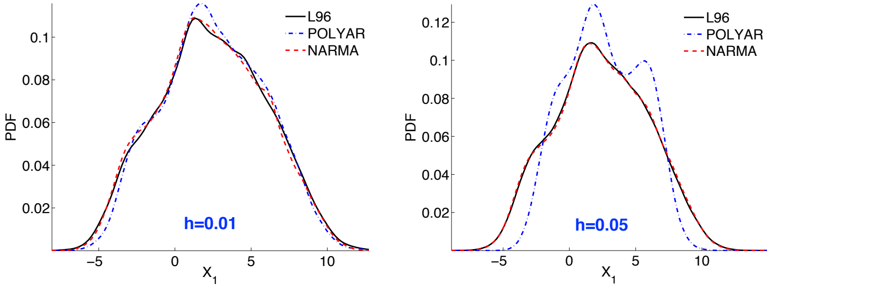
- **Polynomial autoregression (POLYAR)**[3]

$$
\begin{aligned}
\frac{d}{dt} x_k &= x_{k-1}\left(x_{k+1} - x_{k-2}\right) - x_k + 10 + \boxed{U}, \\
U &= P(x_k) + \eta_k, \; \text{with } d\eta_k(t) = \phi \eta_k(t) + dB_k(t).
\end{aligned}
$$

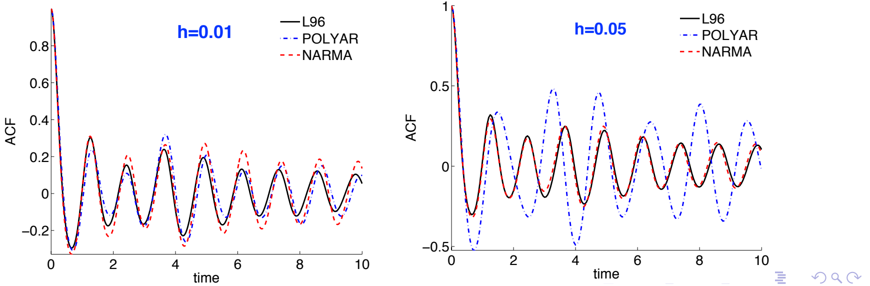
where $P(x) = \sum_{j=0}^{d_x} a_j x^j$. Optimal $d_x = 5$.

Long-term statistics

Empirical probability density function (PDF)

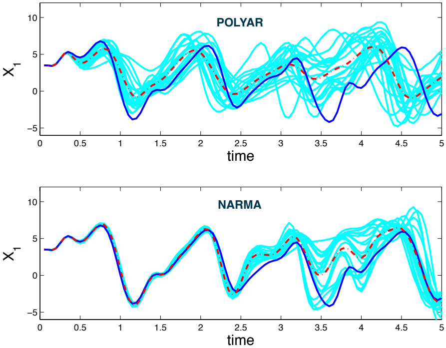


Empirical autocorrelation function (ACF)



Prediction ($h = 0.05$)

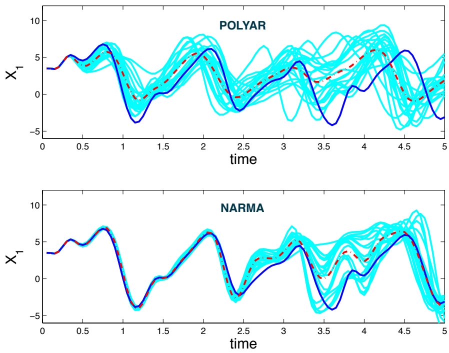
A typical ensemble forecast:



- forecast trajectories in cyan
- true trajectory in blue

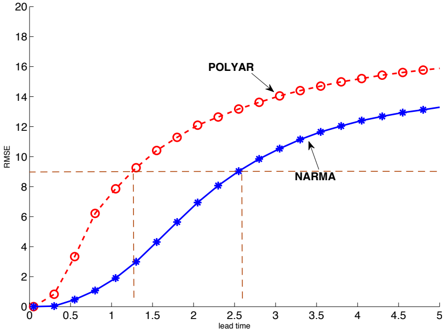
Prediction ($h = 0.05$)

A typical ensemble forecast:



- forecast trajectories in cyan
- true trajectory in blue

RMSE of many forecasts:



Forecast time:

POLYAR: $T \approx 1.25$

NARMA: $T \approx 2.5$

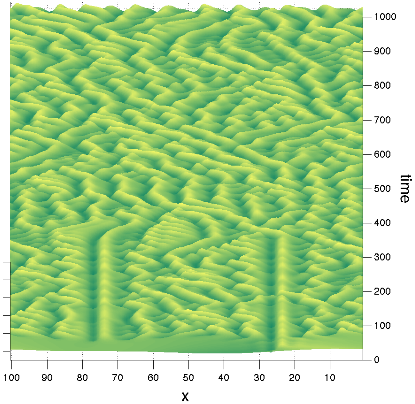
(Full model: $T \approx 2.5$

"Best" forecast time achieved!)

Example II: the Kuramoto-Sivashinsky equation

$$v_t + v_{xx} + v_{xxxx} + vv_x = 0\,,\, t > 0,\, x \in [0, 2\pi\nu],\, \textit{periodic}.$$

Spatio-temporally chaotic

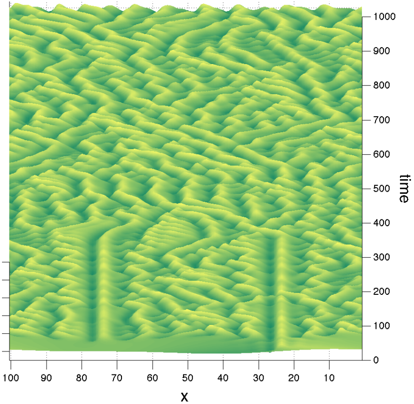


solved with 128 Fourier modes

Example II: the Kuramoto-Sivashinsky equation

$$v_t + v_{xx} + v_{xxxx} + vv_x = 0\,,\, t > 0, x \in [0, 2\pi\nu], \textit{periodic}.$$

Spatio-temporally chaotic



solved with 128 Fourier modes

Problem setting: $\nu = 3.43$

- Observing only 5 Fourier modes
- to predict their evolution

Reduced models:

- the truncated system not accurate
- Discrete-time sto. paramtrization[a]: derive structure from inertial manifold
  $\rightarrow$ an effective NARMA model

---

[a] Lu-Lin-Chorin17

Key point 1: long-term statistics $\leftrightarrow$ Large time behavior of PDE[4]

Inertial manifolds $\mathcal{M}$: - finite-dimensional, positively invariant manifolds
- exponentially attracts all trajectories

Let $v = u + w$. Rewrite the KSE:

$$\frac{du}{dt} = Au + Pf(u + w)$$
$$\frac{dw}{dt} = Aw + Qf(u + w)$$

On $\mathcal{M}$, $w = \psi(u)$

$$\frac{du}{dt} = PAu + Pf(u + \psi(u)).$$

Approximate inertial manifolds (AIMs): approximate $w = \psi(u)$

- $\frac{dw}{dt} \approx 0 \Rightarrow w \approx A^{-1}Qf(u + w)$,
- Fixed point: $\psi_0 = 0$; $\psi_{n+1} = A^{-1}Qf(u + \psi_n)$.

Key point 2: parametrize the AIM

- AIM with 5 modes: unstable
  (An accurate AIM requires $m = dim(u)$ to be large!)
- use the terms; estimate their coefficients from data
  $\rightarrow$ an effective NARMA model

NARMA with AIMs[5]

The AIMs hint at how the high modes depend on the low modes:

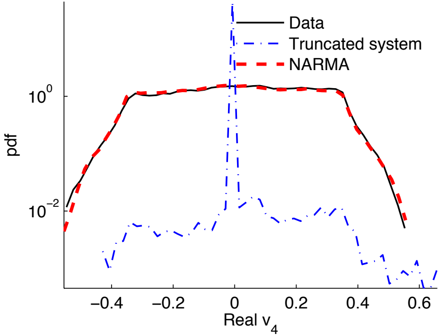
$$|k| > K: \quad \hat{v}_k \approx \psi_{1,k} = \left(A^{-1} Qf(u)\right)_k \Rightarrow \hat{v}_k \approx c_k \sum_{1 \le |l|, |k-l| \le K} \hat{v}_l \hat{v}_{k-l}.$$

$$\widetilde{u}_j^n = \begin{cases} u_j^n, & 1 \le j \le K; \\ i \sum_{l=j-K}^{K} u_l^n u_{j-l}^n, & K < j \le 2K. \end{cases}$$

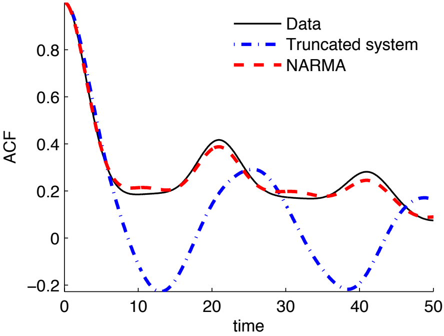
A discrete-time stochastic system: $(p = 2, q = 1)$

$$\begin{aligned} u_k^{n+1} &= u_k^n + hR_k^h(u^n) + hz_k^n, \\ z_k^n &= \Phi_k^n + \xi_k^n, \\ \Phi_k^n(\theta_k) &= \mu_k + \sum_{j=0}^{p} b_{k,j} u_k^{n-j} + \sum_{j=1}^{K} c_{k,j} \widetilde{u}_{j+K}^n \widetilde{u}_{j+K-k}^n + c_{k,(K+1)} R_k^h(u^n) + \sum_{j=1}^{q} d_{k,j} \xi_k^{n-j}. \end{aligned}$$

Long-term statistics:



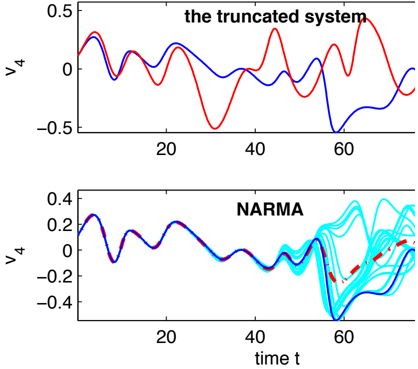
probability density function



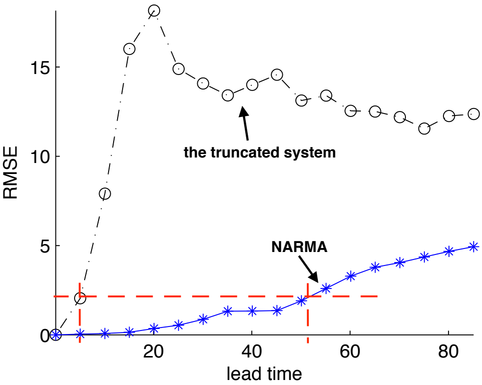
auto-correlation function

Prediction

A typical forecast:



RMSE of many forecasts:



Forecast time:

the truncated system: $T \approx 5$

the NARMA system: $T \approx$ **50**

Part II: Data assimilation with the reduced model

$x' = f(x) + U(x,y)$, $y' = g(x,y)$.
Noisy data: $x(nh) + W(n)$, $n = 1, 2, \ldots$

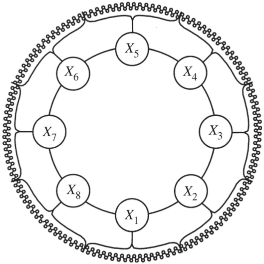
Data assimilation:

- state estimation and prediction
- EnKF: ensemble from forward model + Kalman update

Assume: we can simulate the full system offline
$\rightarrow$ use the solution to quantify model error $U(x,y)$ by

- tuning inflation and localization of EnKF
- deriving a NARMA reduced model

The Lorenz 96 system

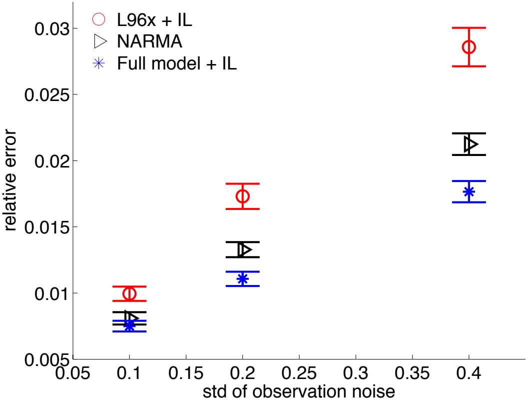


Wilks 2005

Estimate and predict x based on

- ➢ Noisy Data $z(n) = x(nh) + W(n)$
- ➢ Forward models
  - **L96x:** $\frac{d}{dt}x_k \approx x_{k-1}(x_{k+1} - x_{k-2}) - x_k + 10$ (account for the model error by IL in EnKF)
  - NARMA (account for the model error by parametrization in the forward model)

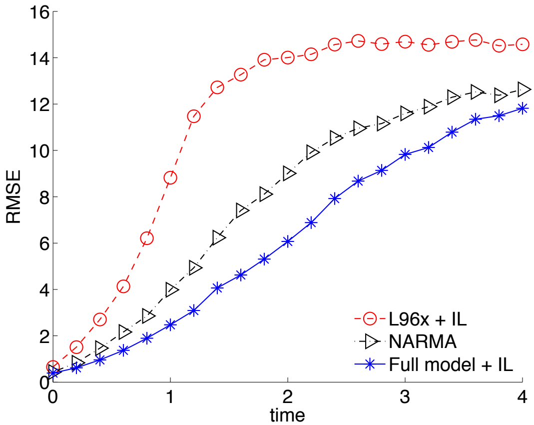
Relative error of state estimation



Relative error for different observation noises.

(ensemble size: =1000 for L96x and NARMA; =10 for the full model)

RMSE of state prediction

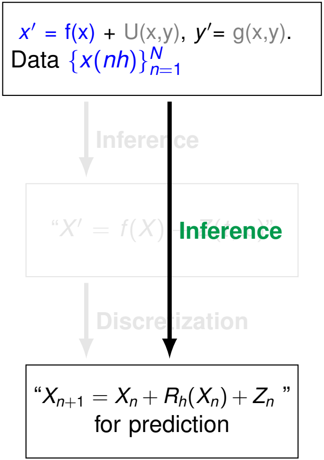


RMSE of 10^4 ensemble forecasts.

(ensemble size: =1000 for L96x and NARMA; =10 for the full model)

Summary: The NARMA modeling improves performance of DA.

Summary and ongoing work



Data-driven stochastic model reduction by
Discrete-time stochastic parametrization

- effective non-Markovian reduced model (**NARMA**)
  - captures key statistical-dynamical features
  - makes medium-range forecasting
- Improves performance of Data assimilation

Open and ongoing work: if noisy data only?

- DA with non-Markovian models
- inference for hidden non-Markovian models

Thank you!