Inference based model reduction for complex dynamical systems

Fei Lu

Department of Mathematics, Johns Hopkins

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

April 3, 2019
KI-Net: Dimension reduction in physical and data sciences
Department of Mathematics, Duke University

Inference based stochastic model reduction

1. Motivation and objective
2. An inference approach
3. Data assimilation with reduced models
4. Model reduction for PDEs
    - Kuramoto-Sivashinsky
    - **Stochastic Burger equation**

Motivation: data assimilation in weather/climate prediction

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

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

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

- HighD multiscale full chaotic/ergodic systems:
  - can only afford to resolve $x' = f(x)$ online
  - y: unresolved variables (subgrid-scales)
- Discrete noisy observations: missing i.c.
- Ensemble prediction: need many simulations

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

Goal: Develop a closed reduced model of x that

- captures key statistical + dynamical properties
- use it for online state estimation and prediction

reduce spatial dimension + increase time step-size

Various efforts in closure model reduction:

- Direct constructions:
  - Mori-Zwanzig formalism
  - relaxiation approximations
  - Galerkin methods: non-linear Galerkin, Petrov - Galerkin
  - linear response (Fluctuation-dissipation theory)
    (filtering / feedback control /)

- Inference
  - hypoellitpic SDEs, GLE and SDDEs
  - manifold learning
  - neural network learning
  - discrete-time (time series) models

(system identification / parametrization / error quantification)

Inference-based model reduction

SDEs and time series – dynamical models

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 but non-intrusive
Inference[1]	likelihood
Discretization[2]	error correction by data

_ _ _ _ _ _ _ _ _

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

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

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
- NARMAX in system identification $Z_n = \Phi(Z, X) + \xi_n$,

Tasks:

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

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

Structure derivation: derive the terms in Φ_n

$$X_n = X_{n-1} + R_h(X_{n-1}) + Z_n,$$
$$Z_n = \Phi_n + \xi_n,$$
$$\Phi_n = \sum_{j=1}^{p} a_j X_{n-j} + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(X_{n-j}) + \sum_{j=1}^{q} c_j \xi_{n-j}.$$

The nonlinear structure is problem dependent.

Techniques: drive terms from

- the reduced differential systems (IDEs/SDEs)
- numerical schemes
- inertial manifolds

and estimate their coefficients from data.

Parameter estimation: conditional maximum likelihood[1]

$$\begin{aligned} X_n &= X_{n-1} + R_h(X_{n-1}) + Z_n, \\ Z_n &= \Phi_n + \xi_n, \\ \Phi_n &= \sum_{j=1}^{p} a_j X_{n-j} + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(X_{n-j}) + \sum_{j=1}^{q} c_j \xi_{n-j}; \end{aligned}$$

Conditioned on $\xi_1, \ldots, \xi_q$, the log-likelihood of $X_{q+1:N}$ is

$$l(\theta|\xi_1, \ldots, \xi_q) = -\sum_{n=q+1}^{N} \frac{|Z_n - \Phi_n|^2}{2\sigma^2} - \frac{N-q}{2} \ln \sigma^2.$$

- compute Φ_n and ξ_n recursively for $n > q$
- maximum likelihood estimator (MLE) by optimization
- $\{\xi_n\}$ can be non-Gaussian to get other lost fucntions
- when $q = 0$: a least squares regression

[1] e.g. Hamilton 94: Time Series Analysis

Overview:

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

Discrete-time stochastic parametrization

NARMA

$$
\begin{aligned}
X_n &= X_{n-1} + R_h(X_{n-1}) + Z_n, \\
Z_n &= \Phi_n + \xi_n, \\
\Phi_n &= \sum_{j=1}^{p} a_j X_{n-j} + \sum_{j=1}^{q} c_j \xi_{n-j} \\
&\quad + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(X_{n-j}).
\end{aligned}
$$

1. compute $R_h(x)$
2. derive structure
3. estimate parameters

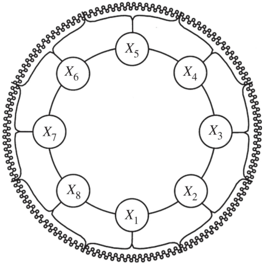
Data assimilation with model reduction

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

Data assimilation:

- estimate the state of a forward model (FM):
  - (x, y) for the full model
  - x for a reduced model of x
- predict x (by ensembles of solutions of FM)
  - Widely used method: Ensemble Kalman filters (EnKF)

The Lorenz 96 system



Wilks 2005

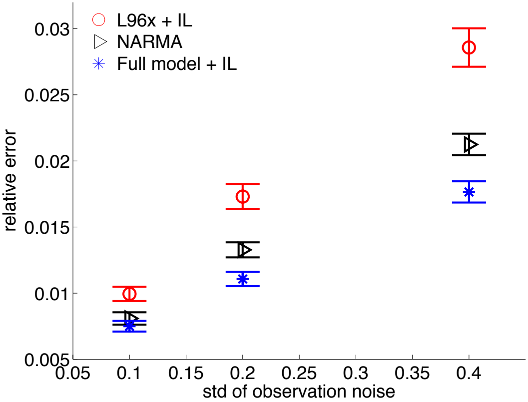
Estimate and predict x based on

➢ Noisy Data $z(n) = x(nh) + \mathbf{W(n)}$

➢ **Forward models**

- **L96x**: the truncated model
  $\frac{d}{dt}x_k \approx x_{k-1}\left(x_{k+1} - x_{k-2}\right) - x_k + 10$
  (account for the model error by IL in EnKF)
- **NARMA** (account for the model error by parametrization in the forward model)

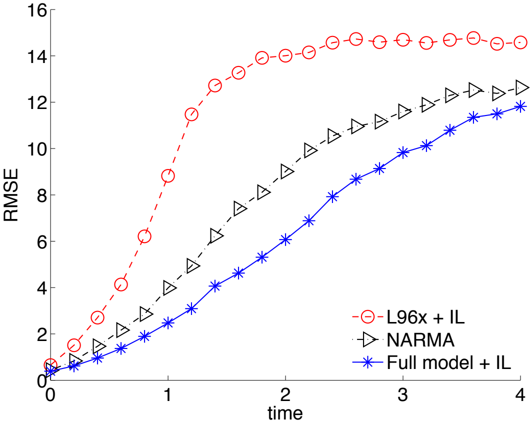
Relative error of state estimation



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: NARMA improves performance of DA.

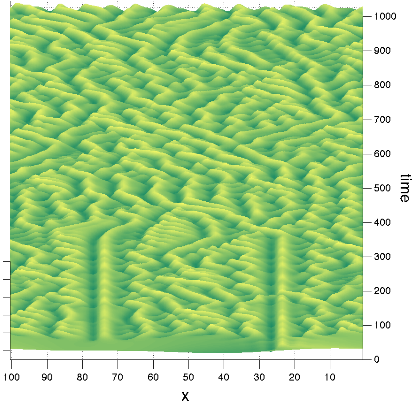
Model reduction for PDEs

(chaotic/stochastic PDEs)

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

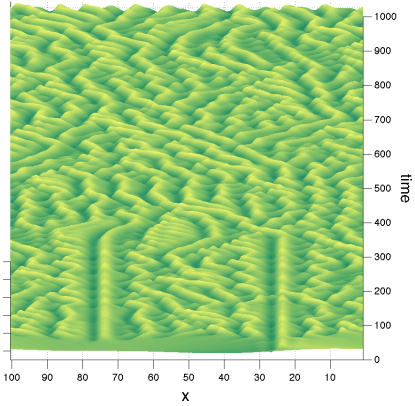
Problem setting: $\nu = 3.43$

- Observing only 5 Fourier modes every 10 time steps
- to predict their evolution
- 100-fold reduction

The Kuramoto-Sivashinsky equation

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

Spatio-temporally chaotic



solved with 128 Fourier modes

Problem setting: $\nu = 3.43$

- Observing only 5 Fourier modes every 10 time steps
- to predict their evolution
- 100-fold reduction

Reduced models:

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

---

[a] Stinis 12
[b] Lu-Lin-Chorin17

Key point 1: long-term statistics $\leftrightarrow$ Large time behavior of PDE[2]

Inertial manifolds $\mathcal{M}$: - finite-dimensional, positively invariant manifolds

- exponentially attracts all trajectories

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

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

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

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

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

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

---

[2] Foias, Constantin, Temam, Sell, Jolly, Kevrekidis, Titi et al (88-94)

Key point 1: long-term statistics $\leftrightarrow$ Large time behavior of PDE[2]

Inertial manifolds $\mathcal{M}$: - finite-dimensional, positively invariant manifolds

- exponentially attracts all trajectories

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

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

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

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

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

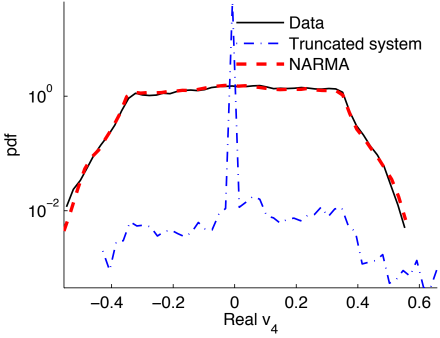
- $\frac{dw}{dt} \approx 0 \Rightarrow w \approx A^{-1}QB(u + w)$,
- Fixed point: $\psi_0 = 0$; $\psi_{n+1} = A^{-1}QB(u + \psi_n)$.

Key point 2: parametrize the AIM

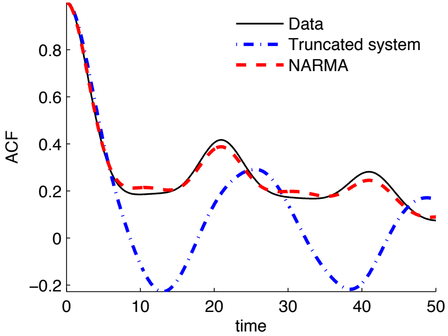
- AIM with 5 modes: unstable
  (An accurate AIM requires $m = dim(u)$ to be large!)
- use the terms; estimate their coefficients from data
  $\rightarrow$ an effective NARMA model

[2] Foias, Constantin, Temam, Sell, Jolly, Kevrekidis, Titi et al (88-94)

Long-term statistics:



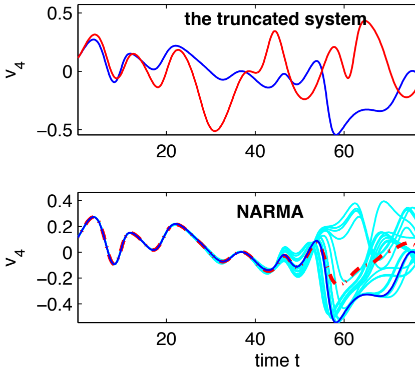
probability density function



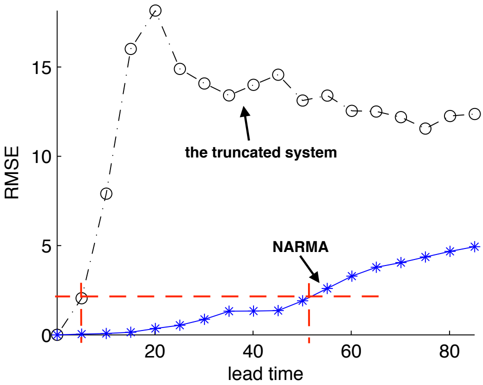
auto-correlation function

Prediction

A typical forecast:



RMSE of many forecasts:



Forecast time:

the truncated system: $T \approx 5$
the NARMA system: $T \approx \mathbf{50}$

Stochastic Burgers equation:

$$v_t = \nu v_{xx} - v v_x + f(x,t), \quad x \in [0, 2\pi], \text{ periodic}$$

where $f(x,t) = \sum_{k=1}^{K_0} \sin(kx) dW(t)$.

- $N \gtrsim 5/\nu$ Fourier modes are needed to resolve the Eq.

Goal: a closed model for $(\widehat{v}_{1:K})$, $K = 2K_0 \leq 1/\nu$.

- to capture energy spectrum and correlations (>PDF, ACF)
- Test setting:
  - Full model: $\nu = 0.05$, $K_0 = 4 \to$ random shocks
    $N = 128, dt \to$ CFL ≈ 0.2;
  - Reduced model: $K = 8$, $\delta = 20dt \to$ CFL ≈ 4;

Let $v = u + w$. In operator form:

$$\frac{du}{dt} = PAu + PB(u) + Pf + [PB(u+w) - PB(u)]$$

$$\frac{dw}{dt} = QAw + QB(u+w)$$

- No spectral gap $\rightarrow$ no inertial manifold (IM)
  No Approximate IM: ~~$\frac{dw}{dt} \approx 0 \Rightarrow w \approx A^{-1}QB(u+w)$.~~

Let $v = u + w$. In operator form:

$$\frac{du}{dt} = PAu + PB(u) + Pf + [PB(u+w) - PB(u)]$$
$$\frac{dw}{dt} = QAw + QB(u+w)$$

- No spectral gap $\rightarrow$ no inertial manifold (IM)
  No Approximate IM: ~~$\frac{dw}{dt} \approx 0 \Rightarrow w \approx A^{-1}QB(u+w)$.~~

Integration instead:

$$w(t) = e^{-QAt}w(0) + \int_0^t e^{-QA(t-s)}[QB(u(s)+w(s))]ds$$
$$w^n \approx c_0 QB(u^n) + c_1 QB(u^{n-1}) + \cdots + c_p QB(u^{n-p})$$

Linear in parameter approximation:

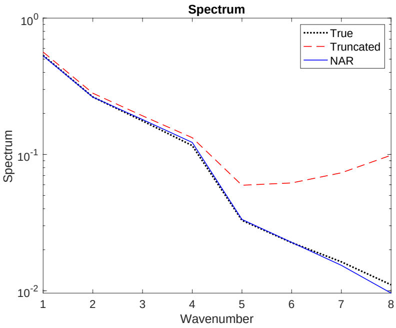
$$PB(u+w) - PB(u) = P[(uw)_x + (w^2)_x]/2 \approx P[(uw)_x]/2 + noise$$
$$\approx \sum_{j=0}^{p} c_j P[(u^n QB(u^{n-j}))_x] + noise$$

The NAR model has the form

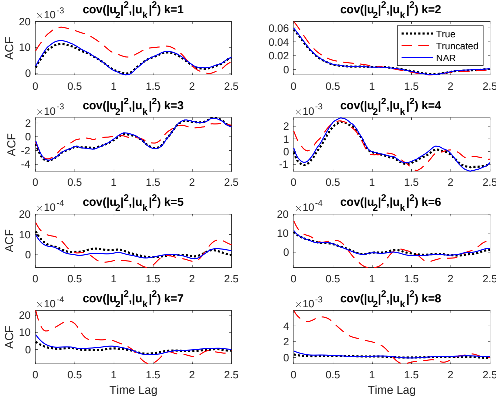
$$u_k^n = R^\delta(u_k^{n-1}) + f_k^n + g_k^n + \Phi_k^n,$$

with $\Phi_k^n := \Phi_k^n(u^{n-p:n-1}, f^{n-p:n-1})$ in form of

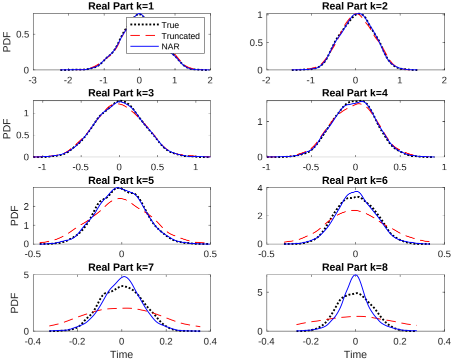
$$\Phi_k^n = \sum_{j=1}^{p} c_{k,j}^v u_k^{n-j} + c_{k,j}^R R^\delta(u_k^{n-j}) + c_{k,j}^w \sum_{\substack{|k-l|\le K, K<|l|\le 2K \\ \text{or } |l|\le K, K<|k-l|\le 2K}} \widetilde{u}_l^{n-1} \widetilde{u}_{k-l}^{n-j}$$



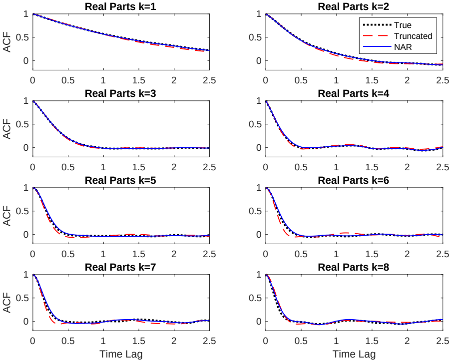
Energy spectrum



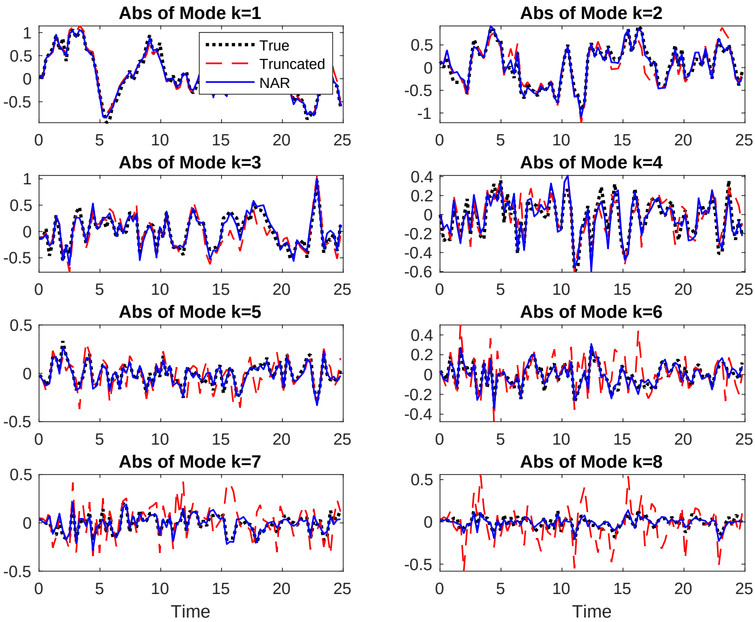
Cross-ACF of ernergy



probability density function



auto-correlation function



Prediction in response to force

Several questions:

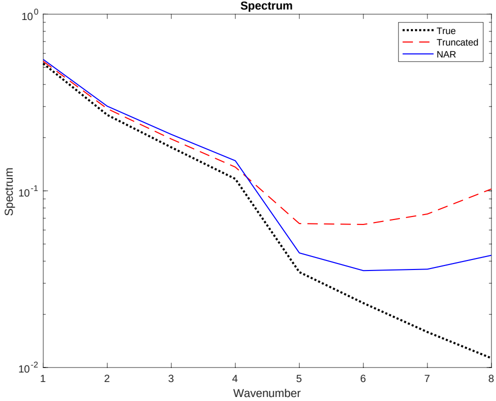
- How to determine Lag length? ($e^{-QA(t-s)}$ not good)

$$w(t) = e^{-QAt}w(0) + \int_0^t e^{-QA(t-s)}[QB(u(s) + w(s))]ds$$

$$w^n \approx c_0 QB(u^n) + c_1 QB(u^{n-1}) + \cdots + c_p QB(u^{n-p})$$

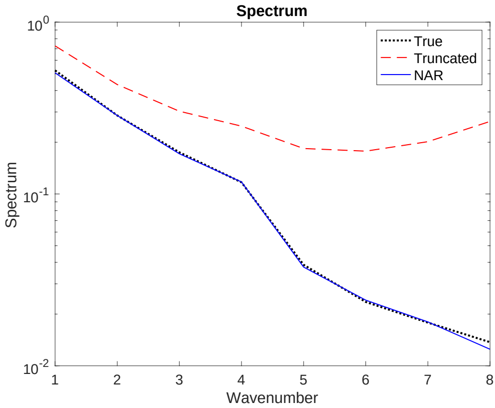
- Larger observation gap?
- Smaller ν?
- The role of $(w^2)_x$?
- Memory kernel? Correlated noise?
- Generalization to other dissipative systems?

When observe gap = 80 time steps:



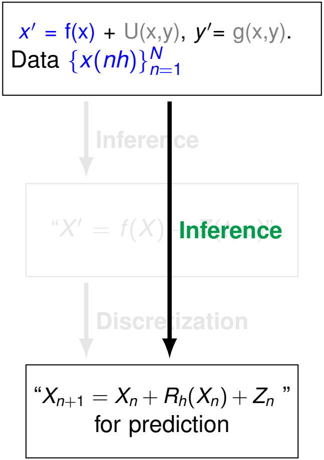
Energy spectrum

When viscosity becomes smaller $\nu = 0.02$ (more Burgulent)



Energy spectrum

Summary and ongoing work



- Stochastic model reduction by
  Discrete-time stochastic parametrization
  - simplifies the inference from data
  - incorporates memory flexibly
  - effective reduced model (**NARMA**)
    - capture key statistical-dynamical features
    - make medium-range forecasting
- Improves performance of Data assimilation

Open problems:

- model reduction
  - model selection
  - scaling of the discrete system? (With A. Chorin)
  - 2D N-S equation?
- noisy data: state estimation and model inference
  - data assimilation with non-Markovian models
  - inference for hidden non-Markovian models
- theoretical understanding of the approximation
  - distance between the two stochastic processes?
  - discrete-time Mori-Zwanzig formalism (With K. Lin)

    A Winer filter implementation (instead of the likelihood)

References

- Data-driven stochastic model reduction
  - Chorin-Lu: Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics. **PNAS 112 (2015)**, no. 32, 9804–9809.
  - Lu-Lin-Chorin: Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems. **CAMCoS, 11 (2016)**, no. 8, 4227–4246.
  - Lu-Lin-Chorin: Data-based stochastic model reduction for the Kuramoto – Sivashinsky equation. **Physica D, 340 (2017)**, 46–57.
  - Lin-Lu: Data-driven model reduction, Wiener projections, and the Mori-Zwanzig formalism. preprint (2019)
- Data assimilation
  - Lu-Tu-Chorin: Accounting for model error from unresolved scales in EnKFs: improving the forecast model. **MWR, 340 (2017)**.

Thank you!

FL acknowledges supported from LBL, NSF