Data Assimilation with Stochastic Model Reduction of Chaotic 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)

Snowbird, SIAM-DS, May 23, 2019
Model error from sub-grid scales

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

From: Wikipedia, ECWMF

The Lorenz 96 system

Wilks 2005
Model error from sub-grid scales

ECMWF: 16 km horizontal grid → $10^9$ freedoms

The Lorenz 96 system
Wilks 2005

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

Observe only $\{x(nh)\}_{n=1}^N$.
Forecast $x(t), t \geq Nh$.

- HighD multiscale full systems:
  - can only afford to resolve $x' = f(x)$
  - $y$: unresolved variables (subgrid-scales)

- Discrete noisy observations: missing i.c.
- Ensemble prediction: need many simulations
Model error from sub-grid scales

ECMWF: 16 km horizontal grid → $10^9$ freedoms

The Lorenz 96 system
Wilks 2005

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

Observe only \( \{x(nh)\}_{n=1}^N \).
Forecast \( x(t), t \geq Nh \).

HighD multiscale full systems:

- can only afford to resolve \( x' = f(x) \)
- \( y \): unresolved variables (subgrid-scales)

Discrete noisy observations: missing i.c.

Ensemble prediction: need many simulations

→ How to account for the model error \( U(x, y) \)?
Given a highD multiscale full system:

\[ x' = f(x) + U(x, y), \quad y' = g(x, y). \]

Ensemble prediction: can afford to resolve \( x' = f(x) \) online.

**Accounting model error** \( U(x, y) \) **from subgrid scales**

- **Indirect approaches:** correct the forecast ensemble
  - e.g. Inflation and Localization;\(^1\) relaxation/bias correction\(^2\),
  - in Assimilation step; deficiency in forecast model remains
- **Direct approach:** improve the forecast model
  - parametrization methods \(^3\), non-Markovian \(^4\),
  - random perturbation, averaging+homogenization \(^5\)

---

\(^1\) Mitchell-Houtekamer(00), Hamill-Whitaker(05), Anderson(07)
\(^2\) Zhang-etc (04), Dee-Da Silva(98)
\(^3\) Palmer, Arnold+(01,13), Wilks(05), Meng-Zhang(07), Danforth-Kalnay-Li+(08,09), Berry-Harlim(14), Mitchell-Carrissi(15), Van Leeuwen etc(18)
\(^4\) Chorin+(00-15), Marjda-Timofeyev-Harlim+(03-13), Chekroun-Kondrashov-Gil+(11,15), Cromellin+Vanden-Eijinden(08), Gottwald+(15)
\(^5\) Hamill-Whitaker(05), Houtekamer+(09), Pavliotis-Stuart(08), Gottwald+(12-13)
1. Stochastic model reduction
   (reduction from simulated data)
   ▶ Discrete-time stochastic parametrization (NARMA)

2. Data assimilation with the reduced model
   (Noisy data + reduced model $\rightarrow$ state estimation and prediction)
Stochastic model reduction

\[ x' = f(x) + U(x, y), \quad y' = g(x, y). \]

Why stochastic reduced models?

- The system is “ergodic”: \( \frac{1}{N} \sum_{n=1}^{N} F(x(nh)) \xrightarrow{N \to \infty} \int F(x) \mu(dx) \)
- \( U(x, y) \) acts like a stochastic force
Stochastic model reduction

Why stochastic reduced models?
- The system is “ergodic”: \( \frac{1}{N} \sum_{n=1}^{N} F(x(nh)) \xrightarrow{N \to \infty} \int F(x) \mu(dx) \)
- \( U(x, y) \) acts like a stochastic force

Memory effects (Mori, Zwanzig, Chorin, Kubo, Majda, Wilks, Ghil, . . .)
- Mori-Zwanzig formalism → generalized Langevin equ.
  \[
  \frac{dx}{dt} = \underbrace{\mathbb{E}[\text{RHS}|x]}_{\text{Markov term}} + \int_{0}^{t} K(x(s), t - s)ds + \underbrace{W_t}_{\text{noise}},
  \]
- Fluctuation-dissipation theory → Hypoelliptic SDEs
  \[
  dX = a(X, Y)dt + Y; \quad dY = b(X, Y)dt + c(X, Y)dW,
  \]
- Parametrization: multi-layer stochastic models

Goal: develop a non-Markovian stochastic reduced system for \( x \)
Discrete-time stochastic parametrization

**NARMA**(\(p, q\))

\[
X_n = X_{n-1} + R_h(X_{n-1}) + Z_n, \\
Z_n = \Phi_n + \xi_n,
\]

\[
\Phi_n = \sum_{j=1}^{p} a_j X_{n-j} + \sum_{j=1}^{r} \sum_{i=1}^{s} b_{i,j} P_i(X_{n-j}) + \sum_{j=1}^{q} c_j \xi_{n-j}
\]

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

**Tasks:**
- **Structure derivation:** terms and orders \((p, r, s, q)\) in \(\Phi_n\);
- **Parameter estimation:** \(a_j, b_{i,j}, c_j, \text{ and } \sigma\).
Overview:

\[ x' = f(x) + U(x, y), \ y' = g(x, y). \]

Data \( \{x(nh)\}_{n=1}^{N} \)

**Discrete-time stochastic parametrization**

### NARMA

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

1. compute \( R_h(x) \)
2. derive structure
3. estimate parameters
Application to the chaotic Lorenz 96 system

A chaotic dynamical system (a simplified atmospheric model)

\[
\frac{dx_k}{dt} = x_{k-1} (x_{k+1} - x_{k-2}) - x_k + 10 - \frac{1}{J} \sum_j y_{k,j},
\]

\[
\frac{dy_{k,j}}{dt} = \frac{1}{\epsilon} \left[ y_{k,j+1} (y_{k,j-1} - y_{k,j+2}) - y_{k,j} + x_k \right],
\]

where \( x \in \mathbb{R}^{18}, y \in \mathbb{R}^{360} \).

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

\( \triangleright \) Data \( \{ x(nh) \}_{n=1}^N \)

\( \triangleright \frac{dx_k}{dt} \approx x_{k-1} (x_{k+1} - x_{k-2}) - x_k + 10. \)

Wilks 2005
NARMA:

\[ x^n = x^{n-1} + R_h(x^{n-1}) + z^n; \quad z^n = \Phi^n + \xi^n, \]

\[ \Phi^n = a + \sum_{j=1}^{p} \sum_{l=1}^{d_x} b_{j,l}(x^{n-j})^l + \sum_{j=1}^{p} c_j R_h(x^{n-j}) + \sum_{j=1}^{q} d_j \xi^{n-j}. \]

\[ p = 2, d_x = 3; \quad q = \begin{cases} 1, & h = 0.01; \\ 0, & h = 0.05. \end{cases} \]
NARMA:

\[ x^n = x^{n-1} + R_h(x^{n-1}) + z^n; \quad z^n = \Phi^n + \xi^n, \]

\[ \Phi^n = a + \sum_{j=1}^{p} \sum_{l=1}^{d_x} b_{j,l}(x^{n-j})^l + \sum_{j=1}^{p} c_j R_h(x^{n-j}) + \sum_{j=1}^{q} d_j \xi^{n-j}. \]

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

Polynomial autoregression (POLYAR)\(^6\)

\[ \frac{d}{dt} x_k = x_{k-1} (x_{k+1} - x_{k-2}) - x_k + 10 + U, \]

\[ U = P(x_k) + \eta_k, \quad \text{with} \quad d\eta_k(t) = \phi \eta_k(t) + dB_k(t). \]

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

---

\(^6\)Wilks 05: an MLR model in atmosphere science
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! )
Outline

1. Stochastic model reduction
   (reduction from simulated data)
   - Discrete-time stochastic parametrization (NARMA)

2. Data assimilation with the reduced model
   (Noisy data + reduced model → state estimation and prediction)
Data assimilation with the reduced model

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

Data assimilation:
- estimate the state of a forward model
- make prediction (by ensembles of solutions)

The widely used method: Ensemble Kalman filters (EnKF)
The Lorenz 96 system

Estimate and predict $x$ based on

- Noisy Data $z(n) = x(nh) + W(n)$

- **Forward models**
  - **L96x**: the truncated model
    \[
    \frac{d}{dt} x_k \approx x_{k-1} (x_{k+1} - x_{k-2}) - x_k + 10
    \]
    (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 stochastic model improves performance of DA.
Summary and ongoing work

Accounting model error \( U(x, y) \) from subgrid scales

\[
\begin{align*}
x' &= f(x) + U(x, y), \\
y' &= g(x, y).
\end{align*}
\]

Data \( \{x(nh)\}_{n=1}^{N} \)

- **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

- **Improve the forecast model**
  - Improve performance of DA
Ongoing work:

- **noisy data: state estimation and model inference**
  - data assimilation with non-Markovian models
  - inference for hidden non-Markovian models

- **model reduction for (stochastic) PDEs**
  - stochastic Burgers equation, N-S equation
References

- **Data-driven stochastic model reduction**

- **Data assimilation**

Thank you!