Uncertainty Quantification, Data Assimilation and Prediction of Complex Nonlinear Turbulent Dynamical Systems

Instructor: Nan Chen

Department of Mathematics
University of Wisconsin-Madison

Graduate Course, MATH 801, Spring 2020
MWF 8:50AM – 9:40AM, B239 Van Vleck Hall
Uncertainty Quantification, Data Assimilation and Prediction of Complex Nonlinear Turbulent Dynamical Systems

The course should be interesting for undergraduate students, graduate students, and postdocs in pure and applied mathematics, physics, engineering, and climate, atmosphere, ocean science interested in modeling and simulating complex turbulent dynamical systems.

This course is a self-contained introduction to these topics from nonlinear dynamics, stochastic modeling, numerical algorithms and real applications. Many lectures involve hot research topics. Students and young researchers can certainly follow the lectures with elementary background.

**Prerequisites:** undergraduate courses in ODEs, PDEs and numerics.
Office hours: Monday and Wednesday 4:20pm – 5:20pm

Instructor’s email: chennan@math.wisc.edu

Lecture notes will be uploaded to Canvas.

There will be no final exam.

You are welcome to give a lecture based on one of the topics listed in the next three pages.
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 22,24</td>
<td>Overview</td>
</tr>
<tr>
<td>2</td>
<td>Jan 27,29,31</td>
<td>Introduction to Information Theory and UQ</td>
</tr>
<tr>
<td>3</td>
<td>Feb 3,5,7</td>
<td>Stochastic Toolkits</td>
</tr>
<tr>
<td>4</td>
<td>Feb 10,12,14</td>
<td>Filtering, data assimilation and state estimation</td>
</tr>
<tr>
<td>5</td>
<td>Feb 17,19,21</td>
<td>Introduction to optimal control theory</td>
</tr>
<tr>
<td>Week</td>
<td>Date</td>
<td>Topic</td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>6</td>
<td>Feb 24, 26,28</td>
<td><strong>Conditional Gaussian nonlinear modeling framework</strong></td>
</tr>
<tr>
<td>7</td>
<td>Mar 2,4,6</td>
<td><strong>Lagrangian data assimilation</strong></td>
</tr>
<tr>
<td>8</td>
<td>Mar 9,11,(13)</td>
<td><strong>Nonlinear smoothing and state estimation</strong></td>
</tr>
<tr>
<td>9</td>
<td>Mar 16,18,20</td>
<td><strong>Spring break</strong></td>
</tr>
<tr>
<td>10</td>
<td>(Mar 23,25,27)</td>
<td><strong>(Optimal design of observational network)</strong></td>
</tr>
<tr>
<td>11</td>
<td>Mar 30, Apr 1,3</td>
<td><strong>Parameter estimation</strong></td>
</tr>
<tr>
<td>Week</td>
<td>Date</td>
<td>Topic</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>12</td>
<td>Apr 6,8,10</td>
<td>Efficient solver of high-dimensional Fokker-Planck equations</td>
</tr>
<tr>
<td>13</td>
<td>Apr 13,15,17</td>
<td>Developing stochastic low-order models for predicting intermittent time series and extreme events</td>
</tr>
<tr>
<td>14</td>
<td>Apr 20,22,24</td>
<td>Fluctuation-Dissipation and Linear Response Theory</td>
</tr>
<tr>
<td>14</td>
<td>Apr 27,29, May 1</td>
<td>Application of UQ and stochastic models to studying El Niño-Southern Oscillation</td>
</tr>
</tbody>
</table>

The instructor will be out of town on Mar 13, 23, 25 and 27.
Overview
Example 1: Prediction of Hurricane Sandy (Oct 30, 2012)
NEW YORK UNIVERSITY

TO: THE NYU COMMUNITY

FROM: Jules Martin, Vice President for Global Security and Crisis Management

RE: Cancellation of Classes and Normal University Operations on Monday, Oct 29.

The safety and well-being of the NYU community is of foremost concern to us. Due to the hazards presented by Hurricane Sandy and the announced shut-down this evening at 7:00 pm of the New York’s transit systems, all classes, activities, and events will be cancelled tomorrow, Monday, Oct 29. University offices will be closed.

This communication does not apply to NYU Langone Medical Center, including the School of Medicine, which will separately be making determinations and communicating directly with the Med Ctr. community.

Note: On Oct 1, the location and the situation of the Joaquin was almost the same as that of the Sandy!
The Butterfly Effect.

flap flap

Yesss... Yesss...

by J.L. Westover

www.mrlovenstein.com
Why we need to quantify the uncertainty?

- Nature is chaotic/turbulent. Small errors can be enlarged significantly shortly.
- Stochastic models are used to describe the small-scale features due to the incomplete knowledge of the underlying physics.
- Important for improving the prediction skill.

Example: Lorenz 63 model

\[
\begin{align*}
    dx &= \sigma (y - x) \, dt, \\
    dy &= (x(\rho - z) - y) \, dt, \\
    dz &= (xy - \beta z) \, dt,
\end{align*}
\]
What is data assimilation and why it is necessary?

- Data assimilation (or filtering) is the process of combining data and a physical model to accurately determine the state of the system and quantify inherent uncertainties.

- Models have uncertainty. Observational data has noise (uncertainty). Good initial conditions are extremely important for predicting turbulent systems.

- Not all the model variables have observations.
Major Challenges

1. turbulent, chaotic, intermittent instability ...
2. nonlinear, non-Gaussian, extreme events ...
3. model error, no perfect knowledge of the truth
4. large dimensionality (typically $O(10^6) - O(10^9)$), computational issues
5. limited, sparse and noisy observations
6. simplification, approximations, discretization, parameterization of the unresolved processes.
7. ...
Important topics

1. How to measure the skill (i.e., the statistical accuracy) of a given imperfect model in reproducing the present states and predicting the future states in an unbiased fashion?
Important topics

1. How to measure the skill (i.e., the statistical accuracy) of a given imperfect model in reproducing the present states and predicting the future states in an unbiased fashion?

2. How to make the best possible estimate of model sensitivity to changes in external or internal parameters by utilizing the imperfect knowledge available of the present status? What are the most sensitive parameters for the change of the model status given uncertain knowledge of the present status?
Important topics

1. How to measure the skill (i.e., the statistical accuracy) of a given imperfect model in reproducing the present states and predicting the future states in an unbiased fashion?

2. How to make the best possible estimate of model sensitivity to changes in external or internal parameters by utilizing the imperfect knowledge available of the present status? What are the most sensitive parameters for the change of the model status given uncertain knowledge of the present status?

3. How to design cheap and practical reduced models that are nevertheless able to capture both the main statistical features of nature and the correct response to external/internal perturbations?
Important topics

1. How to measure the skill (i.e., the statistical accuracy) of a given imperfect model in reproducing the present states and predicting the future states in an unbiased fashion?

2. How to make the best possible estimate of model sensitivity to changes in external or internal parameters by utilizing the imperfect knowledge available of the present status? What are the most sensitive parameters for the change of the model status given uncertain knowledge of the present status?

3. How to design cheap and practical reduced models that are nevertheless able to capture both the main statistical features of nature and the correct response to external/internal perturbations?

4. How to develop a systematic data-driven nonlinear modeling and prediction framework that provides skillful forecasts and allows accurate quantifications of the forecast uncertainty?
Important topics

1. How to measure the skill (i.e., the statistical accuracy) of a given imperfect model in reproducing the present states and predicting the future states in an unbiased fashion?

2. How to make the best possible estimate of model sensitivity to changes in external or internal parameters by utilizing the imperfect knowledge available of the present status? What are the most sensitive parameters for the change of the model status given uncertain knowledge of the present status?

3. How to design cheap and practical reduced models that are nevertheless able to capture both the main statistical features of nature and the correct response to external/internal perturbations?

4. How to develop a systematic data-driven nonlinear modeling and prediction framework that provides skillful forecasts and allows accurate quantifications of the forecast uncertainty?

5. How to build effective models, efficient algorithms and unbiased quantification criteria for on-line data assimilation (state estimation or filtering) and prediction especially in the presence of model error?
Some more concrete examples

- How will the mean temperature change if the heating from the sun increases?
- How will the variance of the temperature respond to the changes of CO2 concentration?
- How will the mean velocity profile in the ocean behave if the salinity starts changing?
- How will the mean temperature in April change if the heating in January decreases?
A bit more details for data assimilation/filtering

1. Prediction (Forecast)

\[ u_{m+1|m} \] (prior)
\[ v_{m+1} \] (observation)
\[ \text{true signal} \]

\[ t_m \quad t_{m+1} \]

2. Analysis (Filtering)

\[ u_{m+1|m+1} \] (posterior)
\[ v_{m+1} \] (observation)
\[ \text{true signal} \]

▶ Basic tool: Bayesian formula

\[ p(u_{m+1}|v_{m+1}) \sim p(u_{m+1})p(v_{m+1}|u_{m+1}). \]

▶ In the situation with linear model and Gaussian noise, \( p(u_{m+1}|v_{m+1}) \) can be written down explicitly, which is known as “Kalman filter”.

▶ In the nonlinear or non-Gaussian case, there is often no closed form for \( p(u_{m+1}|v_{m+1}) \). Numerical methods can deal with the situation with dimension 1, 2 or 3. But the system may have a dimension \( \geq O(10^6) \).

▶ The data assimilation needs to be run sequentially, namely at each time step.
Some information about the numerical weather prediction (NWP):

- Typical assimilation time: every 6 hours.
- Number of ensembles that is affordable: $< 100$.
- Dimension of the global circulation models (GCMs): $O(10^9)$.
- Kalman filter is not used in practice because of the large model error by linear approximation.
- The world most advanced high performance computers are used for NWP.
Some non-Gaussian examples

Monsoon time series

ENSO Nino 3 SST time series

El Niño

La Niña
The FitzHugh-Nagumo model — a prototype of an excitable system (e.g., a neuron).

\[ \epsilon du_i = \left( u_i - \frac{1}{3} u_i^3 - v_i + d_u(u_{i+1} - 2u_i + u_{i-1}) \right) dt + \sqrt{\epsilon \delta_1} dW_{u_i}, \]

\[ dv_i = (u_i + a) dt + \delta_2 dW_{v_i}, \quad \text{for } i = 1, \ldots, N. \]

The Lorenz 96 model — Model for baroclinic turbulence in the mid-latitude atmosphere.

\[ du_i = \left( u_{i-1}(u_{i+1} - u_{i-2}) - u_i + F \right) dt, \quad i = 1, \ldots, N. \]
Simple motivating examples for uncertainty quantification

Statistical description of the turbulence.

- How does the uncertainty change from (a) to (c)?
- What's the difference in the uncertainty in (a) and (d)?
- How to quantify the uncertainty in (e) and (f)?
Quantifying the model error: $q$ is the true PDF and $p$ is from imperfect model

Quantifying the lack of information: $q$ is the prior PDF and $p$ is the posterior PDF

Prior: from running the model. Posterior: combining model information with data.

Quantifying the prediction skill
A simple example with an intrinsic barrier for improving model sensitivity

- **Perfect model**

  
  \[
  \frac{du}{dt} = au + v + F, \\
  \frac{dv}{dt} = qu + Av + \sigma \dot{W},
  \]

- **Smooth Gaussian measure if**

  
  \[a + A < 0, \quad aA - q > 0.\]

- **Imperfect model**

  
  \[
  \frac{d u_M}{dt} = -\gamma M u_M + F_M + \sigma M \dot{W}_M.
  \]

A good imperfect model captures both the model fidelity and the model sensitivity of the truth.
Climate fidelity for imperfect model

\[
\frac{F_M}{\gamma_M} = -\frac{AF}{aA - q}, \quad \frac{\sigma^2_M}{2\gamma_M} = \frac{\sigma^2}{2(a + A)(aA - q)} \equiv E.
\]

Response to change in forcing

\[
\delta u = -\frac{A}{aA - q} \delta F, \quad \delta u_M = \frac{1}{\gamma_M} \delta F.
\]

With perfect fidelity

\[
\mathcal{P}(\pi_\delta, \pi_{\delta}^M) = \frac{1}{2} E^{-1} \left| -\frac{A}{aA - q} - \frac{1}{\gamma_M} \right| |\delta F|^2.
\]

With \(A > 0\) there is no finite minimum over \(\gamma_M\) and necessarily \(\gamma_M \to \infty\) in the approach to this minimum value. Thus, there is an intrinsic information barrier to skill in the mean response that cannot be overcome with the imperfect models even if they satisfy perfect model fidelity.

If \(A < 0\), then a unique minimum with \(\gamma^*_M = -A^{-1}(aA - q)\) is found, in which case both the model fidelity and model sensitivity are captured. For nonlinear and complex models, the Fluctuation-Dissipation Theorem is a good method.
An example of the so-called simple models in real world

Atmosphere

\[
\begin{align*}
(\partial_t + du)u' - yv' - \partial_x \theta' &= 0 \\
yu' - \partial_y \theta' &= 0 \\
(\partial_t + du)\theta' - (\partial_x u' + \partial_y v') &= Ha' \\
(\partial_t + dq)q' + Q(\partial_x u' + \partial_y v') &= -Ha' + \sigma_q \dot{W}_q \\
\partial_t a' &= \Gamma q' (\bar{a} + a') - \lambda a' + \sqrt{\lambda (\bar{a} + a')\bar{a}} \dot{W}_a
\end{align*}
\]

Ocean

\[
\begin{align*}
\partial_t U - \epsilon c_1 YV + \epsilon c_1 \partial_x H &= \epsilon c_1 \tau_x \\
YU + \partial_Y H &= 0 \\
\partial_t H + \epsilon c_1 (\partial_x U + \partial_Y V) &= 0
\end{align*}
\]

SST

\[
\partial_t T = -\epsilon c_1 \zeta E_q + \epsilon c_1 \eta H,
\]

- 9 equations of PDEs for the ENSO.
- With finite difference discretizations, the total dimension of the system is \(O(10^3)\).
- The only nonlinearity is in the equation of \(a'\).
- Multiplicative noise appears in the equation of \(a'\).
Thank you