Data → Models

NCAR Summer School, Mathematics & Climate
July 14, 2010

Christopher M. Danforth

Department of Mathematics & Statistics
Complex Systems Center
Vermont Advanced Computing Center

University of Vermont
Outline

Lyapunov Exponents
- Atmosphere
- Solar System

Data Assimilation
- Toy Climate Models

Model Error
- Another Toy
- Global Models
“If we should observe a hurricane, we might ask ourselves, ‘Why did this hurricane form?’ If we could determine the exact initial conditions at an earlier time, and if we should feed these conditions, together with a program for integrating the exact equations, into an electronic computer, we should in due time receive a forecast from the computer, which would show the presence of a hurricane.
Lorenz, Tellus, 1960

“If we should observe a hurricane, we might ask ourselves, ‘Why did this hurricane form?’ If we could determine the exact initial conditions at an earlier time, and if we should feed these conditions, together with a program for integrating the exact equations, into an electronic computer, we should in due time receive a forecast from the computer, which would show the presence of a hurricane.

We then might still be justified in asking why the hurricane formed. The answer that the physical laws required a hurricane to form from the given antecedent conditions might not satisfy us, since we were aware of that fact even before integrating the equations.”
Outline

Lyapunov Exponents
  Atmosphere
  Solar System

Data Assimilation
  Toy Climate
  Models

Model Error
  Another Toy
  Global Models
Lorenz (1963) System

\[
\begin{align*}
\frac{dx}{dt} &= \sigma(y - x) \\
\frac{dy}{dt} &= \rho x - y - xz \\
\frac{dz}{dt} &= xy - \beta z
\end{align*}
\]
Lorenz (1963) System
Lorenz (1963) System
FIG. 3 (color). Average locations of regions with low BV dimensions are shown through the pointwise time average of the BV dimension calculated from ensemble forecasts every 12 h from 10 February 2000 to 30 July 2000. Red (blue) depicts regions in which the BV dimension tends to be low (high).
Outline

Lyapunov Exponents
  Atmosphere
  Solar System

Data Assimilation
  Toy Climate
  Models

Model Error
  Another Toy
  Global Models
Question: Is our Solar System Chaotic?
Answer: Chaos/Order Separatrix Passes Directly Through the Current Observational Error Ball

Two Sibling Pairs Differing by $10^{-7}$

- $10^0$ - $10^{-2}$
- $10^{-4}$ - $10^{-6}$

Two Sibling Pairs Differing by $10^{-7}$ Million Years

2-norm of position vector difference (15−D)
Stability of the Solar System

Hayes et. al. 2010 MNRAS in press

Lyapunov Time (My)

Fractional Change to Jupiter Orbital Radius

Fractional Change to Neptune Orbital Radius

Hayes et. al. 2010 MNRAS in press
Stability of the Solar System

Hayes et al. 2010 MNRAS in press

Lyapunov Exponents
Atmosphere
Solar System
Data Assimilation
Toy Climate
Models
Model Error
Another Toy
Global Models
Fractal Basin Boundaries

Hayes et. al. 2010 MNRAS in press
Fractal Basin Boundaries
Hayes et. al. 2010 MNRAS in press
Fractal Basin Boundaries

Hayes et. al. 2010 MNRAS in press
Fractal Basin Boundaries

Hayes et al. 2010 MNRAS in press

Lyapunov Time (My)
Fractal Basin Boundaries

Hayes et. al. 2010 MNRAS in press

Lyapunov Time (My)

Frame 13/39
Outline

Lyapunov Exponents
  Atmosphere
  Solar System

Data Assimilation
  Toy Climate
  Models

Model Error
  Another Toy
  Global Models
Data Assimilation

Image: Kameron Harris
A Toy Climate Experiment

A

B

heat sink

heat source

gravity

fluid flow
A Toy Climate Experiment
A Toy Climate Experiment

Credit: Glenn Russell
Typical Observations of Delta Temp (A-B)
Forcing: Small
Stable Conduction
Typical Observations of Delta Temp (A-B)
Forcing: Medium
Stable Convection
Typical Observations of Delta Temp (A-B)
Forcing: Large
Chaotic Convection
Outline

Lyapunov Exponents
Atmosphere
Solar System

Data Assimilation
Toy Climate
Models

Model Error
Another Toy
Global Models
CFD Simulation

Data Assimilation

Image: Kameron Harris
Figure 4: Two views of the 2D thermosyphon's attractor. 

In (a), a time-delay reconstruction, with $\tau = 60$ s, is used to plot the FLUENT system attractor using the monitored mass flow rate. In (b), data points show $x_1$ and $x_2$ of the EM analyses generated by EKF with an assimilation window of 30 s. Each is colored by the bred vector growth rate over the prior window. Note how trajectories that move through the far edge of either lobe create distinctive loops near the center of the opposite lobe after a regime change.

Figure 5: Results of 3D-Var assimilating the same data are shown for three different assimilation windows. In (a), observations are made frequently enough to keep the forecast close to the truth. In (b), the filter has satisfactory overall performance ($\text{RMSE} \approx 35\% \text{ of } \langle q^2 \rangle$); note the error spike around $8 \times 10^3$ s when the forecast and truth end up in different regimes. With the largest window (c), DA is incapable of keeping the forecast in the correct regime. The largest errors tend to occur at regime changes.

Ground forecasts. The unobserved model state variables are still poor predictions.

Model errors further complicate the implementation of DA in realistic forecasting scenarios. For instance, to test the effect of model errors on different DA schemes, Kalnay et al. [24] created a Lorenz truth with $r = 28$...
Figure 4: Two views of the 2D thermosyphon's attractor. In (a), a time-delay reconstruction, with $\tau = 60$ s, is used to plot the FLUENT system attractor using the monitored mass flow rate. In (b), data points show $x_1$ and $x_2$ of the EM analyses generated by EKF with an assimilation window of 30 s. Each is colored by the bred vector growth rate over the prior window. Note how trajectories that move through the far edge of either lobe create distinctive loops near the center of the opposite lobe after a regime change.

Figure 5: Results of 3D-Var assimilating the same data are shown for three different assimilation windows. In (a), observations are made frequently enough to keep the forecast close to the truth. In (b), the filter has satisfactory overall performance ($\text{RMSE} \approx 35\%$ of $\sqrt{\langle q^2 \rangle}$); note the error spike around $8 \times 10^3$ s when the forecast and truth end up in different regimes. With the largest window (c), DA is incapable of keeping the forecast in the correct regime. The largest errors tend to occur at regime changes.

Ground forecasts. The unobserved model state variables are still poor predictions. Model errors further complicate the implementation of DA in realistic forecasting scenarios. For instance, to test the effect of model errors on different DA schemes, Kalnay et al. [24] created a Lorenz truth with $r = 28$. A

Image: Kameron Harris
Figure 8: The EKF-assimilated trajectory during the largest-oscillation regime change, corresponding to the bottom right-most point of Fig. 7. The system starts in the negative (counterclockwise) regime and stalls near the convecting state for a long time, causing fluid in the bottom to heat up, manifesting in a highly negative $x_3$. The fluid turns over one more time, the wide swing in the trajectory to the left of the figure, then begins to rotate clockwise before quickly returning to the counterclockwise regime. Color indicates the 30 s assimilation window BV growth rate.

Figure 9: FLUENT simulation showing the temperature profile, units of K, of steady counterclockwise convecting flow. The loop parameters are $Ra = 1.2 \times 10^4$ and $R/r = 3$ (for visualization). In the chaotic case, opposite anomalous regions of warmer and cooler fluid are superimposed on this temperature profile. As these pass through the loop, the “tongues” of warm and cool fluid extending into the top and bottom halves of the loop will grow and shrink simultaneously until the hot tongue visible near 2 o’clock reaches the opposite side of the loop. The flow then stalls and reverses direction.

This study will use DA and breeding to qualitatively predict regime changes and new regime duration. Besides informing when regime changes will occur, the evident gradient of bred vector growth rates up the “steps” in Fig. 7 should be useful in determining the duration of the upcoming regime. In the more distant future, the imperfect model experiment we have devised could be used to compare the relative performance of other DA algorithms (4D-Var, ETKF) or synchronization approaches (adaptive nudging, see [2]).

A laboratory thermosyphon device is in construction; in the next stage of this research, we will attempt to forecast the regime changes of the physical experiment.

Although the thermosyphon is far from representing anything as complex and vast as Earth’s weather and climate, there are many characteristics our toy climate shares with global atmospheric models. Both are, at best, only an approximate representation of the numerous processes that govern the Earth’s climate. Global models and the EM model both parameterize fine-scale processes that determine large-scale behavior.
Outline

Lyapunov Exponents
  Atmosphere
  Solar System

Data Assimilation
  Toy Climate
  Models

Model Error
  Another Toy
  Global Models
Lorenz and Emanuel, 1996

\[ \frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{iJ} y_j \]

\[ \frac{dy_j}{dt} = -cby_{j+1}(y_{j+2} - y_{j-1}) - cy_j + \frac{hc}{b} x_{\text{floor}[(j-1)/J]+1} \]

Schematic for \( I = 8 \) and \( J = 4 \):
FIG. 2.2. System perturbation time series. The effect of a five unit perturbation is observed over a 5 day time series. After 5 days, the perturbation effects half of the locations, indicating advection to the east and more slowly to the west ($I = 40$, $J = 16$).

FIG. 2.3. Extended perturbation time series. The initial state of Fig. 2.2 is observed over a 5 day time series beginning at day 50. Particular perturbations can be traced throughout the time series ($I = 40$, $J = 16$).

Adjusting the parameters of the system and examining the time series at a single site illustrates the dependence of the slow variables upon coupling to the fast variables. The relative significance of the fast modes can be observed by varying $h$ as shown in Fig. 2.4 ($I = 6$, $J = 16$). For $h = 1$, a regular pattern emerges as an energy equilibrium is achieved between external forcing and dissipation. However, as the significance of the fast modes is reduced, the time series becomes more complex.
Dynamics

System

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{ij} y_j
\]

\[
\frac{dy_j}{dt} = -cby_{j+1}(y_{j+2} - y_{j-1}) - cy_j + \frac{hc}{b} x_{\text{floor}[(j-1)/J]+1}
\]
Dynamics

▶ System

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{ij} y_j
\]

\[
\frac{dy_j}{dt} = -cby_{j+1}(y_{j+2} - y_{j-1}) - cy_j + \frac{hc}{b} x_{\text{floor}[(j-1)/J]+1}
\]

▶ Model

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F
\]
III. METHODS

From a given initial condition, the trajectory of the truth ($z_a$) is created by integration of the system (i.e. using both the slow and fast variables in Equations (1) and (2)). A two and three dimensional view of this attractor is shown in Fig. 3.8.

The forecast for each ensemble member is then completed by setting $h = 0.5$ in the governing equations (hereafter referred to as the 'model'). In other words, the model is rendered imperfect by dampening the effect of the fast modes in Equation (2) by 50%. The same two and three dimensional slices can be seen in Fig. 3.9, now for the model. This particular experimental design was chosen as it is typical for global atmospheric models to attempt to parameterize sub-grid scale behavior, e.g. for phenomena occurring on a finer temporal/spatial scale. For both the truth and model forecasts, integration of the differential equations is completed using the fourth order Runge-Kutta method with a time step of 0.01. Rigorous shadowing attempts would be made using far more advanced methods of integration, with much smaller time steps. However, for the purpose of this study of short forecasts, the difference is negligible.

FIG. 3.8. System Attractor. The left panel shows a two-dimensional view of the system attractor looking at $x_1$ vs $x_3$. The right panel shows a three-dimensional view using $x_1$, $x_2$, and $x_4$ ($I = 4$, $J = 16$).

FIG. 3.9. Model attractor. The left frame shows a two-dimensional view of a typical forecast looking at $x_1$ vs $x_3$. The right frame shows a three-dimensional view of the forecast using $x_1$, $x_2$, and $x_4$ ($I = 4$, $J = 16$).

Ensemble Creation

Long integrations of the system on randomized initial values were performed to establish the shape of the attractor. A set of 500 different $I$-dimensional points were then chosen at 250-day intervals. This spacing was chosen to ensure that the initial states sampled different regions of the system attractor, and neighboring states were uncorrelated. A 'true' trajectory from each of these states was determined using a 50-day integration of the system. The goal is then to use the model to shadow these trajectories with an ensemble of 20 members.

At each of the 500 initial states, an $I$-dimensional hypersphere was constructed encompassing 100 neighboring states from the system attractor. A neighboring state is defined to be one within 5% of the climatological span of $x_i$ in the $i$th dimension.

At each hypersphere, the covariance for the 100 neighboring states is calculated, yielding a $I \times I$ matrix $C$. This distribution is then scaled to ensure that the average standard deviation is 5% of the climatological span of the system attractor. Thus, for each of the 500 hyperspheres, 20 initial ensemble members are chosen based on the distribution:

$$C_{init} = 0.052 \sigma^2_{clim} \lambda C$$ (3)
FIG. 2.7. A 10 day time series for $x_1$ and $y_1, 2, 3, 4$ using $I = 8, J = 4$. For the top frame, the solid line represents the time series for $x_1$ with fast mode coupling (the system). The dashed line has fast mode coupling turned off, and thus is integrating Equation (1) with $h = 0$. The bottom four frames are the fast mode time series coupled to $x_1$. That the fast variables all vary differently throughout the forecast, with amplitudes on the order of 10% those of the slow variables. Disturbances tend to propagate in about 4 days.
Model Error Estimation

'Truth' or Analysis

\[ \mathbf{x}^{\text{true}}(t) \rightarrow \mathbf{x}^{\text{true}}(t+1) \rightarrow \mathbf{x}^{\text{true}}(t+2) \]

6-hour forecasts

\[ \mathbf{x}^f(t+1) \rightarrow \mathbf{x}^f(t+2) \]

errors

\[ \mathbf{x}^e(t+1) \rightarrow \mathbf{x}^e(t+2) \]

Danforth and Kalnay, JAS, 2008
Model Error Estimation

Danforth and Kalnay, JAS, 2008
Dynamics

System

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{iJ} y_j
\]

\[
\frac{dy_j}{dt} = -cby_{j+1}(y_{j+2} - y_{j-1}) - cy_j + \frac{hc}{b} x_{\text{floor}[(j-1)/J]+1}
\]

Model

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F
\]
Dynamics

- **System**

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{iJ} y_j
\]

\[
\frac{dy_j}{dt} = -cby_{j+1}(y_{j+2} - y_{j-1}) - cy_j + \frac{hc}{b} x_{\text{floor}[(j-1)/J]+1}
\]

- **Model**

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F
\]

- **Model Error Correction**

\[
\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + G(\vec{x})
\]
Figure 4. Typical 10-day ensemble forecasts of $x_1$ using model (30), $F_{11005}^{14}$, with empirical correction terms described by (31). The dashed curve is a true solution of system (27), (28). The solid curves are a 20-member ensemble forecast of model (30), initialized according to Eq. (34). Forecasts empirically corrected by the observed bias of model (29), $D^{(2)}$, perform slightly better than forecasts not corrected at all, $D^{(1)}$. Ensemble divergence is typically significant by day 5 for both $D^{(1)}$ and $D^{(2)}$. Ensemble spread is weak for both Leith's empirical correction $D^{(3)}$ and the SVD correction $D^{(4)}$ with mode truncation $K_{11005}^5$. However, small spread is seen for perfect model forecasts $D^{(5)}$, and the effect is less evident for $F_{11005}^{14}$ and $F_{11005}^{18}$.
Model Error Correction

Danforth and Kalnay, JAS, 2008
Outline

Lyapunov Exponents
Atmosphere
Solar System

Data Assimilation
Toy Climate
Models

Model Error
Another Toy
Global Models
National Center for Environmental Prediction
Global Forecast Model

Credit: Nicholas Allgaier
Forecast Error

Credit: Nicholas Allgaier
Forecast Error

u wind (m/s) at 200 hPa
3/15/2009 GMT: 12:00
Lead Time: 000 hours

Credit: Nicholas Allgaier
Forecast Error

Credit: Nicholas Allgaier
Estimating Model Error

'Truth' or Analysis

\[
x_{\text{true}}(t) \quad x_{\text{true}}(t+1) \quad x_{\text{true}}(t+2)
\]

6-hour forecasts

\[
x_6^f(t+1) \quad x_6^f(t+2)
\]

\[
x_6^e(t+1) \quad x_6^e(t+2)
\]

\[x_{\text{true}}(t+1) - x_6^f(t+1) = x_6^e(t+1)\]

\[x_{\text{true}}(t+2) - x_6^f(t+2) = x_6^e(t+2)\]
SVD Modes for Regression

Danforth, Kalnay, Miyoshi, MWR 2007
Acknowledgements

Thanks to ...

University of Maryland
Eugenia Kalnay, James A. Yorke
Robert F. Cahalan, NASA GFSC
Takemasa Miyoshi, JMO, Hong Li,
Chaos group (Hunt, Kalnay, Kostelich,
Ott, Patil, Sauer, Szunyogh, Yorke)

Bates College
Chip Ross & Bonnie Shulman, Math
Mark Semon & George Ruff, Physics

University of Vermont
Darren Hitt, Mechanical Engineering
Floyd Vilmont, Lab

Students
El Hassan Ridouane, Postdoc
Nicholas Allgaier, MS
Ross Lieb–Lappen, MS
Kameron D. Harris, MS