Data assimilation issues in weather and climate

Eric J. Kostelich

ARIZONA STATE UNIVERSITY
SCHOOL OF MATHEMATICS AND STATISTICS

MSRI Climate Change Summer School
July 18, 2008

Co-workers:

Istvan Szunyogh, Brian Hunt, Edward Ott,

Eugenia Kalnay, Jim Yorke

and many others!

Thanks to: Dave Kuhl

Papers, preprints, and codes:

http://www.weatherchaos.umd.edu
http://math.asu.edu/~eric

Principal papers

Preprints: www.weatherchaos.umd.edu

Initial papers:

- E. Ott *et al.*, *Tellus A* **56** (2004), 415–428.
- I. Szunyogh *et al.*, *Tellus A* **57** (2005), 528–545.

Refined mathematical implementation: B. R. Hunt, E. K., I. Szunyogh, *Physica D* **230** (2007) 112–126.

Results with real data: I. Szunyogh, E.K. *et al.*, *Tellus A* **60** (2008) 113–130.

Some big questions

- Why is it so hard to predict the weather?
- If a 7-day weather forecast is hard, what confidence can we have in a 70-year forecast?
- All models have errors; can we trust them?
- All measurements have errors; can we use them?

The mathematics of uncertainty

- Statistical tools: least squares, ANOVA, ARMA, Kalman Filter, etc.
- Where is the crucial information in a noisy time series?
  - the last few measurements? (time domain)
  - the last few "cycles"? (frequency domain)
  - the last few "patterns"? (some other domain)
- Atmospheric flows (to excellent approximations) are governed by deterministic equations
  - Navier-Stokes equations, Bernoulli's Law, barotropic equation, hydrostatic law, ...
  - Can bigger computers improve forecasts?

Picard's existence theorem

Suppose $\mathbf{f}(\mathbf{x}, t)$ is Lipschitz continuous in a neighborhood N of $(\mathbf{x}_0, t_0)$, i.e.,

$$\|\mathbf{f}(\mathbf{x}, t) - \mathbf{f}(\mathbf{y}, t)\| \leq L\,\|\mathbf{x} - \mathbf{y}\|$$

for some constant L whenever $\mathbf{x}, \mathbf{y} \in N$. Then the initial value problem

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}, t) \quad \text{with} \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

has a unique solution in an interval around t_0 (the size of which depends on N and $\mathbf{f}$).

Perfect initial data $\Longrightarrow$ Perfect predictability

Gronwall's inequality

Given $\mathbf{f}$, N, and L as before, and suppose that $\hat{\mathbf{x}}(t_0) = \hat{\mathbf{x}}_0$ approximates $\mathbf{x}(t_0) = \mathbf{x}_0$. Then

$$\|\mathbf{x}(t) - \hat{\mathbf{x}}(t)\| \leq \|\mathbf{x}_0 - \hat{\mathbf{x}}_0\| \, e^{L(t-t_0)}.$$

This is the best estimate that we can expect in general.
Example:

$$x' = Lx \quad \text{with} \quad x(0) = x_0 \quad \text{and} \quad \hat{x}(0) = \hat{x}_0.$$

Then

$$|x(t) - \hat{x}(t)| = |x_0 - \hat{x}_0| \, e^{Lt}.$$

A hint of the difficulties

- Uncertainties in initial conditions may amplify exponentially in time!
- The details are highly equation dependent
- Example: $x' = -Lx$ has the same Lipschitz constant, but

$$|x(t) - \hat{x}(t)| = |x_0 - \hat{x}_0|\, e^{-Lt} \to 0 \quad \text{as} \quad t \to 0$$

- Under what circumstances do uncertainties grow?

Simple case: Linear systems with constant coefficients

Suppose $\mathbf{x}' = \mathbf{A}\mathbf{x}$ where $\mathbf{A} \in \mathbf{R}^{n\times n}$ has n distinct real eigenvalues. The initial condition $\mathbf{x}(0) = \mathbf{x}_0$ yields the solution

$$\mathbf{x}(t) = \mathbf{x}_0 e^{\mathbf{A}t} = c_1 e^{\lambda_1 t}\mathbf{v}_1 + \cdots + c_n e^{\lambda_n t}\mathbf{v}_n$$

where $[\mathbf{x}_0]_V = (c_1, \ldots, c_n)^{\mathrm{T}}$ in the basis of eigenvectors. (Analogous results for repeated and complex eigenvalues.)

Net result: Linear systems with constant coefficients

- Errors in initial conditions in the system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ grow exponentially with time whenever $\mathbf{A}$ has a positive eigenvalue (or an eigenvalue with positive real part).
- This is a global result.

Harder case: Nonlinear systems

- Local result: Suppose $\mathbf{x}_0$ is a fixed point for $\mathbf{x}' = \mathbf{f}(\mathbf{x})$.
- $\mathbf{x}_0$ is hyperbolic if the eigenvalues of the Jacobian matrix $\mathbf{A} = \mathbf{Df}(\mathbf{x}_0)$ are all nonzero (or have nonzero real part).
- Hartman-Grobman theorem: There exists a change of coordinates that maps solutions of $\mathbf{x}' = \mathbf{f}(\mathbf{x})$ onto solutions of the linear system $\mathbf{x}' = \mathbf{A}\mathbf{x}$ in a neighborhood of $\mathbf{x}_0$ whenever $\mathbf{x}_0$ is hyperbolic.

Basic classification of hyperbolic fixed points

Sink: All eigenvalues negative (or negative real part).

Saddle: Some eigenvalues negative and some positive (or some negative and some positive real parts).

Source: All eigenvalues positive (or positive real part).

- Sinks are stable, i.e., insensitive to small initial errors.
- Saddles and sources are unstable (sensitive).

Example: The damped nonlinear pendulum

Assume linear friction:

$$x'' + kx' + \sin x = 0 \quad \text{with} \quad k > 0$$

Define $x_1 = \text{position}(= x)$ and $x_2 = \text{velocity}(= x')$. The equivalent first-order system is

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= -kx_2 - \sin x_1 \end{aligned}$$

There are two fixed points:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \pi \\ 0 \end{pmatrix}$$

Fixed point analysis

The linearized equation about each fixed point $\mathbf{p}$ is

$$\mathbf{x}' = \mathbf{Df}(\mathbf{p})\,\mathbf{x} = \begin{pmatrix} 0 & 1 \\ -\cos x_1 & -k \end{pmatrix} \mathbf{x}.$$

At $\mathbf{p} = (0,0)$: $\mathbf{Df}(0,0) = \begin{pmatrix} 0 & 1 \\ -1 & -k \end{pmatrix}$ with eigenvalues

$$\lambda_\pm = \frac{-k \pm \sqrt{k^2 - 4}}{2}$$

so $\lambda_\pm < 0$ (or $\operatorname{Re}\lambda_\pm < 0$). So $(0,0)$ is a sink (stable).

Fixed point analysis II

At $\mathbf{p} = (\pi, 0)$: $\mathbf{Df}(\pi, 0) = \begin{pmatrix} 0 & 1 \\ +1 & -k \end{pmatrix}$ with eigenvalues

$$\lambda_{\pm} = \frac{-k \pm \sqrt{k^2 + 4}}{2}$$

so $\lambda_{-} < 0 < \lambda_{+}$. Hence $(\pi, 0)$ is a saddle (sensitive). Initial perturbations grow exponentially (at least initially).

The bottom line

- Small changes to initial conditions at the saddle point lead to large short-term changes in the solution.
- On the other hand, the long-term evolution is perfectly predictable.

Forced, damped, nonlinear systems

When damped nonlinear systems are forced strongly enough, they often become *chaotic*.

In a chaotic process, every point is a sensitive point.

Uncertainties in the initial condition of a chaotic process make it hard to predict—even if the process is deterministic.

The Hénon map

Introduced by

M. Hénon, Comm. Math. Phys. **50** *(1976) 69–77.*

It can be written as

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} a - x_n^2 + b y_n \\ x_n \end{pmatrix}$$

Take $a = 2.12$ and $b = -0.3$. Almost every initial condition sufficiently close to the origin yields a chaotic attractor.

The Lorenz ’96 model

Introduced by Edward Lorenz and Kerry Emanuel, *J. Atmos. Sci.* **55** (1998), 399–414. Simple model of generalized “weather” at N points on a latitude circle:

$$x_j' = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F, \quad x_{N+1} \equiv x_1$$

- The nonlinear terms simulate advection and conserve the total energy, defined as $\frac{1}{2}(x_1^2 + \cdots + x_N^2)$
- The linear terms dissipate the total energy
- F represents external forcing ($F = 8$)
- $x_1 = \cdots = x_N = F$ is a fixed point ($N = 40$)

The geometry of uncertainty

- Suppose our knowledge of the initial condition $\mathbf{x}_0$ is a "circle" of uncertainty (i.e., the underlying pdf is circularly symmetric and centered about $\mathbf{x}_0$).
- How does a dynamical system propagate the uncertainty?

Linear example: $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$

- Basic formula: matrix $\times$ circle $=$ ellipse
- Key idea: The singular value decomposition

$$\mathbf{A}_{m\times n} = \mathbf{U}_{m\times n}\mathbf{S}_{n\times n}\mathbf{V}^{\mathrm{T}}_{n\times n}$$

- $\mathbf{S} = \mathrm{diag}(s_1, s_2, \ldots, s_n)$ gives the singular values, which are the square roots of the eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.
- By convention, $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$.
- If C is the unit circle, then s_i is the length of the ith axis of $\mathbf{A}C$

Rank-r approximations

- The rank of $\mathbf{A}$ is the number of nonzero singular values
- The condition number is s_1/s_n
- Rank-r approximation of $\mathbf{A}$:

$$\hat{\mathbf{A}}_{m\times n} = \mathbf{U}_{m\times r}\hat{\mathbf{S}}_{r\times r}\mathbf{V}^{\mathrm{T}}_{r\times n}$$

  where $\hat{\mathbf{S}}$ consists of the first r nonzero singular values.
- $\hat{\mathbf{A}}$ is the best least-squares approximation of $\mathbf{A}$ insofar as $\hat{\mathbf{A}}$ is the (unique) rank-r matrix that minimizes

$$\|\hat{\mathbf{A}} - \mathbf{A}\|_F^2 = \sum_{i,j}(\hat{A}_{ij} - A_{ij})^2$$

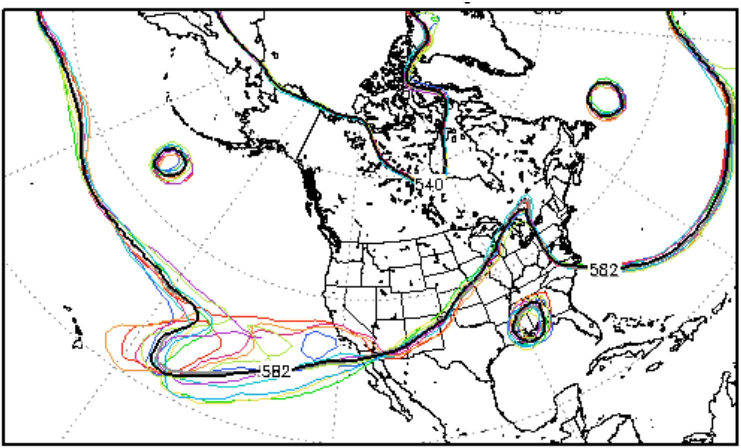
Ensemble forecasting

- How does a nonlinear model propagate a “circle” of uncertainty?
- One procedure: Given $\mathbf{x}' = \mathbf{f}(\mathbf{x})$, integrate the variational equations $\mathbf{U}' = \mathbf{Df}(\mathbf{x})\,\mathbf{U}$
- Not simple to do if $\mathbf{f}$ is big and complicated
- Simpler procedure: Integrate an ensemble of statistically equivalent initial conditions to approximate the uncertainty

The Global Forecast System

- The GFS is the operational global forecast model for the U. S. Weather Service
- Developed and maintained by the National Centers for Environmental Prediction (NCEP), a division of the National Oceanographic and Atmospheric Administration (NOAA)
- In the 1990's, NCEP began to generate ensemble forecasts to give meteorologists a quantitative estimate of forecast uncertainty

Spaghetti plot of a typical 72-hour forecast



Movie

- Movie: Time sequence of operational forecasts from 1 to 16 days starting from the same initial condition (noon on Oct. 16, 2007)
- 20 ensemble solutions (color)
- Variable shown: geopotential height at 500 mb

Key points

- The weather is a chaotic dynamical process
- Forecast uncertainty grows exponentially over short time scales…
- …and varies considerably in time and space
- Lorenz's estimate: The uncertainty in the global atmospheric state vector roughly doubles every 48 hours
- Places an upper bound on the predictability of the weather: 2 weeks

The data assimilation problem

- Without periodic corrections, the forecasts produced by a weather model would be no better than climatology
- Key question: Given a bunch of noisy observations and an imperfect model, find a “maximum likelihood” estimate of the global atmospheric state vector and its uncertainty

Some mathematical questions

- Find a useful representation of the forecast uncertainty
- Use the available observational data efficiently
- Estimate systematic errors (biases) in observations and forecast models
- Where should extra observations be targeted to best advantage?
- If weather isn't predictable after 2 weeks, to what extent can a climate prediction for 80 years from now be trusted?

Plan of the remaining lectures

- Today's lab: The singular value decomposition and rank-r approximations
- Monday's lectures: The Local Ensemble Transform Kalman Filter