A Statistical Approach to Modeling Nonlinear Systems

Barbara A. Bailey
Department of Mathematics and Statistics
San Diego State University

Outline

• Nonlinear Models and Systems
• Application: Lorenz System
• Inference for Parameters of Nonlinear Models
• Application: Modeling Calanus Abundance
• Application: Modeling Cloud Cover
• Future Work
• A general class of nonlinear models is from data of observed univariate responses $Y_i$, dependent on corresponding $d$-dimensional inputs $\mathbf{x}_i$:

$$ Y_i = f(\mathbf{x}_i; \theta) + e_i $$

where $\theta$ is a $p$-dimensional vector of unknown parameters and $\{e_i\}$ is a sequence of i.i.d random variables.

• A nonlinear autoregressive process is a univariate time series:

$$ x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-d}; \theta) + e_t $$

where $\{e_t\}$ is a sequence of i.i.d random variables.
Example: Lorenz System

• Coupled, nonlinear system of three first order differential equations:

\[
\begin{align*}
\frac{dx}{dt} &= -s(x - y) \\
\frac{dy}{dt} &= -xz + rx - y \\
\frac{dz}{dt} &= xy - bz
\end{align*}
\]

for \( s = 10, r = 28 \) and \( b = 8/3 \) get famous “butterfly”

• Data: numerically integrate and add noise at every integration time-step.

• State-Space System:

\[
\begin{align*}
X_t &= F_1(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{1,t} \\
Y_t &= F_2(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{2,t} \\
Z_t &= F_3(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{3,t}
\end{align*}
\]
Lorenz Time Series, $X(t), Z(t), Y(t)$
Phase Space for Noisy Lorenz System
Lyapunov Exponents: Measuring Sensitivity to Initial Conditions

Taylor Series expansion to approximate the action of the map $F$ on two initial state vectors $X_1,Y_1$

$$X_2 - Y_2 = F(X_1) - F(Y_1)$$
$$\approx DF(X_1)(X_1 - Y_1)$$

Let $J_t = DF(X_t)$, the Jacobian matrix of $F$. By the chain rule for differentiation

$$X_n - Y_n \approx J_n \cdot J_{n-1} \cdots J_1(X_1 - Y_1)$$

Global Lyapunov Exponent:

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \ln \| J_n J_{n-1} \cdots J_1 \|$$

Local Lyapunov Exponent:

$$\lambda_n(t) = \frac{1}{n} \ln \| J_{n+t-1} J_{n+t-2} \cdots J_t \|$$
Chaos Facts

• If $X_t$ is ergodic, stationary and bounded. Then $\lambda$ exists and is independent of the trajectory (Multiplicative Ergodic Theorem of Oseledec)

• A system with $\lambda > 0$ has the property of “sensitive dependence on initial conditions” and is chaotic.
Neural Networks

- Class of nonlinear models:
  \[ Y = f(x; \theta) + e \]

- The form of the model:
  \[ f(x) = \beta_0 + \sum_{i=1}^{k} \beta_i \varphi(x^T \gamma_i + \mu_i) \]
  where \( \varphi(u) = e^u/(1 + e^u) \)

  - Net parameters are estimated by nonlinear least squares.
  - Total number of parameters is \( p = 1 + k(d + 2) \) where \( d \) is dimension of \( x \).
  - Complexity of the model chosen by Cross Validation:
    \[ V_c = \frac{\frac{1}{n} RSS}{(1 - p_{\frac{c}{n}})^2} \]
Neural Network Fits to Noisy Lorenz System, Y(t)
Lorenz LLEs

(a) Number of Steps vs. LLEs

(b) LLEs vs. time steps
Confidence Intervals for Parameters of Nonlinear Models

Based on asymptotics of MLE

Assumptions:

A1 \( X_t \) is stationary and ergodic.
A2 \( e_t \) is i.i.d. \( N(0, \sigma^2) \).
A3 \( \theta \in \Theta \), and \( \Theta \) is compact.
A4 \( F \), first, second and third partials of \( F \) exist and are continuous and uniformly bounded for all \( \theta \in \Theta \).

Theorem: Under Assumptions A1-A4, there exists a ML estimator \( \hat{\theta}_n \) of \( \theta \) for which
\[
\hat{\theta}_n \xrightarrow{a.s.} \theta \quad \text{and} \quad I_n^{1/2}(\theta)(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, 1) \quad \text{as} \quad n \to \infty.
\]

Corollary:
\[
-2 \ln \left( \frac{L(\theta)}{L(\hat{\theta})} \right) \xrightarrow{d} \chi^2(p)
\]
Confidence Intervals for Parameters of Nonlinear Models

- Approximate Confidence set for $\theta$:

$$\mathcal{A}_\theta = \{ \theta : -2\ln(L(\theta)/L(\hat{\theta})) \leq c \}$$

- Applications:

$$\mathcal{A}_\theta = \left\{ \theta : S(\theta) \leq S(\hat{\theta}) \left[ 1 + \frac{p}{n-p} F(p, n-p, \alpha) \right] \right\}$$

where $S(\theta)$ is the residual sum of squares and $\hat{\theta}$ is the least-squares estimate of $\theta$.

- CI for a functional of the parameters, $\varphi(\theta)$ is the min and max of the set:

$$\mathcal{A}_{\varphi(\theta)} = \{ \varphi(\theta) : \theta \in \mathcal{A}_\theta \}$$
RSS Surface, $S(\theta)$
Recent Application

Nonlinear Time Series Analysis of Gulf of Maine Calanus Abundance

Collaborator: Andrew Pershing (Cornell University)

CAFE: Climate-based Assessment of Forecasting for Ecosystems in the Gulf of Maine

Goal: Understand the connection between variability and marine populations and apply this knowledge to develop ecological forecasts.

- *Calanus finmarchicus* (C) - log Calanus abundance as measured by the Continuous Plankton Recorder (CPR) surveys by NOAA Fisheries. (The major prey species of the endangered right whale.)

- Coupled Slope Water System (CSWS) - describes the system of currents and water masses between the NW Atlantic Shelf and the Gulf Stream.

- Regional Slope Water Index (RSWT) is a measure of the state of the CSWS. It is the first component of the PCA of the yearly mean temperature in 8 regions of the Gulf of Maine, Scotian Shelf, and adjacent slope waters between 150-200m.
• Nonlinear Time Series Model:

\[ C_t = f(C_{t-1}, RSWT_{t-3}, \sin((2\pi t/6), \cos(2\pi t/6)) + e_t \]

• Fit model to data 1978-2001

• Forecast 2002-2003, starting with data at time \( t \) and iterating model.
Bimonthly Data: Calanus Abundance and RSWT Index 1978-1999

Calanus Time Series

RSWT Time Series
Nonlinear Time Series Neural Network Fit
Model: $C_t = f(C_{t-1}, RSWT_{t-3}, \sin((2\pi t/6), \cos(2\pi t/6)) + e_t$
Forecasting 2002-2003

![Graph showing Calanus Abundance from 1980 to 2001. The graph includes predictions from 1980 to 2001 and forecasts for 2002-2003. The data shows fluctuations in Calanus abundance over the years.]
Future Work

• Look for other important variables to include in the Calanus prediction model.
  – Variables that are suspected to interact with Calanus.
  – We now have 6 other taxa.
Modeling Cloud Cover

Collaborator: William Collins (Climate Modeling Section, NCAR)

Cloud Modeling Motivation:

- Clouds play a fundamental role in controlling the amount of solar and infrared radiation available to the climate system
- Most clouds are smaller in area than the typical grid resolution of climate models

Objectives of the Statistical Model:

- Model the spatial and temporal distribution of cloud cover
- Link large scale climate variables with cloud cover
Data

- Raw data: 3 months of hourly IR data from satellite (TOGA COARE experiment)
  - Individual clouds have been identified and classified by temperature:
    1. Mesoscale convective clouds
    2. Deep clouds
    3. Mixed clouds

- Convective Data
  1. D(GCAPE)/Dt (generalized convective available potential energy)
     - local vertical component of moist available energy
  2. Richardson number - measure of low level shear
  3. Column mean relative humidity
IR Satellite Image and Identifying Clouds

Individual Clouds

Cloud 970
Modeling Clouds as a Dynamical System

The model:
\[
C_{t+1} = f(C_t, S_t) + e_t
\]
where \(S_t\) are other state variables.

Nearest-Neighbor Model:

For a gridded region: \(r\) rows and \(c\) columns \((L = rc)\)
For each site \(l\),
\[
C_{l(t+1)} = f(C_{l(t)}, D(GCAPE)/Dt_{n(l)t}, RICH_{n(l)t}, RHMEAN_{n(l)t}) + e_{l(t)}
\]
where \(e_{l(t)}\) is iid \(N(0, \sigma^2)\),
\(n(l)\) is the “nearest-neighborhood” of each site

\(f\) is a neural network model
Mean Cloud Fraction over sites

RMSE over sites

Cloud Fraction vs RMSE
Conclusions and Future Work

- A first-order lagged nearest-neighbor model is reasonable for modeling cloud cover over time.
- Space-time neural networks new class of space-time models.
- Examine closely the “neighborhood” structure.
- LLEs of “nearest-neighbor” dynamical system - STLLEs