Hierarchical Bayesian Statistical Modeling and Prediction

Mark Berliner

Ohio State University

MBI Workshop: Uncertainty in Ecological Analysis

April 3, 2006

Outline

I Issues and Motivations

II Bayesian Hierarchical Modeling

III Brief Example

IV Discussion

I. Selected Issues

- **Complexity: Many processes at various space-time scales**
- **Advances in computing enable numerical modeling (biophysical, biogeochemical, biogeographic,...), but**
  - **coupling, parameterization, ...**
  - **dynamics and change**
  - **enlarges the list of stuff we don't know.**
- **Complex, diverse, large-to-massive, yet <u>incomplete</u> datasets.**

I. Goals

- "Model-data fusion", while maintaining uncertainty management
- Linking processes across space-time
- Quantifying information about change
- Decision making and decision support

II. Framework: Bayesian Hierarchical Modeling

- Bayesian principles
  - combine information
  - uncertainty quantification & management: inputs and outputs are probability distributions
  - decision making
- Practice
  - seek effective data-model compromises
  - hierarchical thinking!!

II. Bayesian Hierarchical Modeling (BHM)

- **HM: sequence of conditional probability distributions**

$$[\mathbf{x}, \mathbf{y}, \mathbf{z}] = [\mathbf{x} \mid \mathbf{y}, \mathbf{z}]\,[\mathbf{y} \mid \mathbf{z}]\,[\mathbf{z}]$$

- **BHM: Model the component distributions for <u>all</u> unknowns: observations, processes, parameters**

II. Meta-BHM

1. Data Model $[\ \mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}\ ]$

2. Process Model $[\ \mathbf{X} \mid \boldsymbol{\theta}\ ]$

3. Parameter Model $[\ \boldsymbol{\theta}\ ]$

Bayes' Theorem: $[\ \mathbf{X}, \boldsymbol{\theta} \mid \mathbf{Y}\ ]$

Compare To "Strawmen"

- "Statistics": $[\ \mathbf{Y} \mid \boldsymbol{\theta}\ ]$
- "Bayesian Statistics": add $[\ \boldsymbol{\theta}\ ]$
- "Scientific Tradition": $[\ \mathbf{X} \mid \tilde{\boldsymbol{\theta}}(\mathbf{Y})\ ]$

II. What Does This Buy Us?

Combining Information

A Diverse datasets

- <u>Sat</u>ellite data and <u>Sta</u>tion data on X of interest
  Complicated dependence structure, but (?)

$$[\mathbf{Sat}, \mathbf{Sta} | \mathbf{X}, \boldsymbol{\theta}] = [\mathbf{Sat} | \mathbf{X}, \boldsymbol{\theta}][\mathbf{Sta} | \mathbf{X}, \boldsymbol{\theta}]$$

- <u>Cou</u>nt data; <u>Env</u>ironment data on X

$$[\mathbf{Cou}, \mathbf{Env} | \mathbf{Population}, \mathbf{X}, \boldsymbol{\theta}] = [\mathbf{Cou} | \mathbf{Population}, \boldsymbol{\theta}][\mathbf{Env} | \mathbf{X}, \boldsymbol{\theta}]$$

$$[\mathbf{Population}, \mathbf{X} | \boldsymbol{\theta}] = [\mathbf{Population} | \mathbf{X}, \boldsymbol{\theta}][\mathbf{X} | \boldsymbol{\theta}]$$

B Physical-statistical modeling: $[\ \mathbf{X}\ | \boldsymbol{\theta}]$ from a model

- From PDE's to stochastic models
- Model output as <u>data</u>

C Various combinations

Treatment of Complexity

A "Unwrap" some complexities

B Conditioning of X at different space-time scales

$$[\textbf{local rain}|\textbf{regional rain}]$$

"up/down-scaling"

C Space-time and dynamics

$$\mathbf{X_{t+1}} = \mathbf{M}(\mathbf{X_t}, \boldsymbol{\theta}) + \mathbf{e_{t+1}}$$

versus

$$\mathbf{X_{t+1}} = \mathbf{M}(\mathbf{X_t}, \boldsymbol{\theta}_\mathbf{t}) + \mathbf{e_{t+1}}$$

$$[\{\boldsymbol{\theta}_t\}]$$

Problem Description: Meta-HBM

- **Climate and climate change:**
  Forcings → Climate → "Weather" → Impacts

$$[\mathbf{I}|\mathbf{W}][\mathbf{W}|\mathbf{C}][\mathbf{C}|\mathbf{F}][\mathbf{F}]$$

  Messsage: Regional/Local Env forecasting does not rely on forecasts of future "weather", but rather its prob. distribution

- **Evolution:**
  Forcings → Climate → "Weather" → Impacts
  → Phenotype → Genotype

$$[\mathbf{G_{t+1}}|\mathbf{P_t},\mathbf{G_t}][\mathbf{P_t}|\mathbf{G_t},\mathbf{I}][\mathbf{G_t}][\mathbf{I}|\mathbf{W}][\mathbf{W}|\mathbf{C}][\mathbf{C}|\mathbf{F}][\mathbf{F}]$$

  Message: Conjectures about how this works translate to testable assumptions on forms of conditional distributions

III. Pacific SST: Berliner, Wikle & Cressie *J. Climate 2000*

- **Driven, quasi-periodic dynamic behavior of emergent structures.**
- **Data model: regress monthly SST data on spatial EOF (Prin. Comp.)**

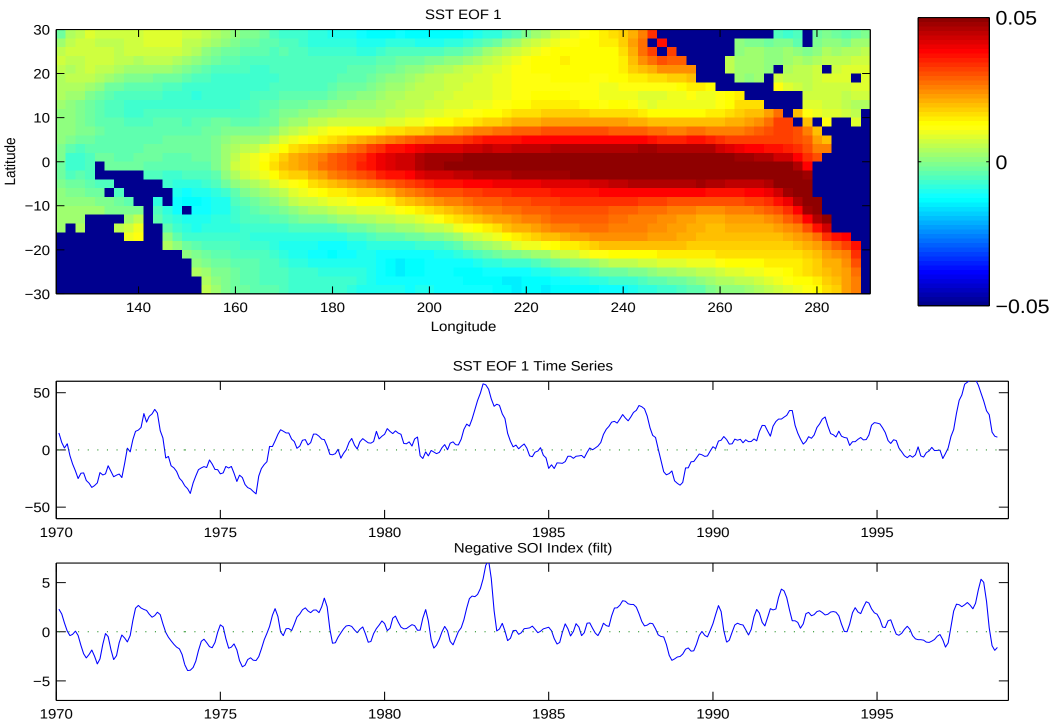
$$\mathbf{Y_t = \sum_{i=1}^{k} a_{i,t} u_i + e_t}$$

Length of Y = 2,520, but k = 9.
Message: Dimension reduction for emergent structures

- **Process model:**
  - **Dynamic model for parameters, with time-varying parameters**

$$\mathbf{a}_{t+7} = \mathbf{H}_t \, \mathbf{a}_t + \boldsymbol{\epsilon}_{t+7}$$

  - **Markov mixture model: cool, normal, warm states (9 $\mathbf{H}_t$'s)**
    Message: More of emergent structures
  - **mixture prob's depend on atmosphere:**
    SOI (pressure), & Winds
    Message: Build in the "drivers", or surrogates, of the dynamics


SST EOF 1
Latitude
Longitude
0.05
0
−0.05
SST EOF 1 Time Series
Negative SOI Index (filt)
30
20
10
0
−10
−20
−30
140
160
180
200
220
240
260
280
50
0
−50
5
−5
1970
1975
1980
1985
1990
1995

IV. Discussion

- Using Science Models: Review Berliner (2003) JGR
  - Scipione & Berliner (1992) Proc. Bayesian Sect. ASA "Lorenz-3" example.
  - Royle et al. (1998) Winds over Labrador Sea.
  - Wikle et al. (2001) JASA Spatio-temporal winds over tropical ocean.
  - Hoar et al. (2003) JCGS More of the previous.
  - Berliner et al. (2003) JGR Air-sea interaction.

$$(\nabla^2 - \frac{1}{r^2})\frac{\partial\psi}{\partial t} = -J(\psi, \nabla^2\psi) - \beta\frac{\partial\psi}{\partial x} + \frac{1}{\rho H}\,\mathrm{curl}_z\boldsymbol{\tau}(W) - \gamma\nabla^2\psi + a_h\nabla^4\psi$$

- Computation: MCMC and ISMC Hybrids
- Challenge: Linking models over space-time
- Meta-BHM to real BHM: science, statistics, & dimension reduction