Concluding Thoughts
OUTLINE

• A few last thoughts on Data Assimilation
• The rest of the Space/Time review
• Where to go from here
Ensemble Analysis

• Allows us to account for multiple sources of uncertainty by sampling from them
  – Model state variables / initial conditions
  – Parameter uncertainty
  – Covariates / drivers / scenarios
  – Different models

• For computational reasons, the number of ensemble members, m, is often not large
  – Some error propagation is better than none!

• For large m, analogous to model CI / PI
Model Averaging

- Any forecasting/prediction (not just DA)
- Often have multiple candidate models that are not significantly different (AIC, DIC, etc)
- Predict using a weighted average of the predictions by each model

\[
P(X|D) = \sum_k P(X|M_k, D) P(M_k|D)
\]
Figure 3: BMA Predictive PDF (thick curve) and its Five Components (thin curves) for the 48-Hour Sea-Level Pressure Forecast at Power River, B.C., Initialized at 0000 UTC on February 23, 2000. Also shown are the ensemble member forecasts and range (solid horizontal line and bullets), the BMA 90% prediction interval (dotted lines), and the verifying observation (solid vertical line).
Bayesian Model Averaging

\[ P(X|D) = \sum P(X|M_k, D) P(M_k|D) \]

\[ P(M_k|D) = \frac{P(D|M_k) P(M_k)}{\sum P(D|M_k) P(M_k)} \]

\[ P(D|M_k) = \int P(D|\theta_k, M_k) P(\theta_k|M_k) \, d\theta_k \]

FIGURE 11.1. An extension of the graph of Figure 1.7 to highlight elements of the prediction process.
Predictions and Decision Making

• Predicting the behavior of some natural phenomenon
  – Global change, endangered spp, disease spread
• Fully specified uncertainties
  – Information content inversely proportional to forecast uncertainty
  – Falsely overconfident prediction leads to poor decisions
• Contingent on explicit scenarios
  – Indicate possibilities, not definitive probabilities
• What is forecastable?

Clark et al 2001 “Ecological Forecasting”
State Space Model Revisited

• Previously have considered the State Space model in an “offline” mode
  – Likelihood of $X_t$ depends on $X_{t-1}$ AND $X_{t+1}$

• Easy to implement in an “online” mode where $X_t$ depends only on $X_{t-1}$

• Flexible framework for either state estimation or both state and parameter estimation

• At each Analysis step, treat parameter posteriors from previous step as priors for this step
Bayesian State Space Model

\[ X_t = f(X_{t-1}) + \epsilon_t \]
\[ Y_t = g(X_t) + \omega_t \]

- \( Y \) = observed data
- \( X \) = latent time series
- \( \epsilon \) = process error
- \( \omega \) = observation error

Process Model
Data Model
Random Walk State Space Model

Data Model

Process Model

Parameter Model
Generality of the State Space Model

- Neither X nor Y need be Normal
- X and Y don't need to be the same type of data
- X and Y don't need to have the same time scale
- Easily handles missing data (gaps) and irregularly spaced data
- Easily handles multiple data sources (Y's), which don't need to be the same type or synchronous
- Easily handles time-integrated observations
Basic Mark-Recapture State Space

• Process model

\[
P(X_t = 1 | X_{t-1} = 1) = s_t
\]
\[
P(X_t = 1 | X_{t-1} = 0) = 0
\]
\[
P(X_t = 0 | X_{t-1} = 1) = 1 - s_t
\]
\[
P(X_t = 0 | X_{t-1} = 0) = 1
\]

Bernoulli Survival Probability

• Observation model

\[
P(Y_t = 1 | X_t = 1) = p_t
\]
\[
P(Y_t = 1 | X_t = 0) = 0
\]
\[
P(Y_t = 0 | X_t = 0) = 1
\]
\[
P(Y_t = 0 | X_t = 1) = 1 - p_t
\]

Bernoulli Detection Probability

• Priors on p and s (e.g. Beta)
Longitudinal Data (aka Repeated Measures)

- The same observational unit (plot, individual, etc) is often measured repeatedly over time
- Usually have many such observation units
- Observations on the same unit over time are not independent
Alternatives for Repeated Measures

- Random effects
  - By time: Assumes all observational units move up or down in sync
    - probably won't solve lack of independence
  - By unit: Assumes a unit is offset from “average” by some constant amount

- Autoregressive: AR(1)
  - Assumes each unit is similar from one time step to the next but not that units are synchronized

- State Space

- With short t.s. almost impossible to distinguish AR vs individual effects
Intervention Analysis

• Treatment effects in TIME
  – Pretreatment data establishes unit differences

• Hypotheses are usually that one or more parameters changed with treatment

• Alternate model (NULL) is no change in parameters with treatment

• Time can be modeled as
  – Covariate (explicit)
  – Time varying treatment (implicit)
  – Autocorrelation
FIGURE 9.21. Four examples of ways in which an intervention might occur. The trend in the solid line “Treatment A” might not match that of the response to the treatment. For example, a step change in a treatment $x$ (a) might elicit a trajectory of change in a response variable $y$. 
The Big Picture

• Quantify states & relationships
  - What is Y?
  - How is Y related to X?
• Test Hypotheses
• Prediction
• Decision making
The Big Picture

• “Confronting Models with Data”

• Probability theory as a coherent framework for making inference and prediction

• Set of flexible but powerful tools
  – “It's all the same model”

• Build your statistical model around your problem!
Where to go from here...

- Ecological Models and Data in R (Bolker 2008)
  - Mostly Likelihood based
  - Lots of great R tricks
  - Well written / easy to follow
  - PDF's of draft version still on Ben's webpage

- Bayesian Methods for Ecology (McCarthy 2007)
  - Lots of good / simple BUGS examples
  - DANGEROUS for the untrained
    - Devoid of theory, caveats, assumptions, and any discussion of numerical methods
• Hierarchical Modeling for the Environmental Sciences (Clark and Gelfand 2006)
  – Mostly a collection of case studies

• Bayesian Data Analysis (Gelman et al 2004)
  – The standard reference for Bayes and Hierarchical Bayes
  – Is a stats book, not biostats, so it's even harder to read than Clark

• A number of other books listed on the course website
Advanced Topics...

• Spatial:
  Banerjee et al 2004 Hierarchical modeling and analysis for spatial data

• Time-series
  Diggle 1990 Time Series: A biostatistical introduction

• Data assimilation
  Lewis et al 2006 Dynamics Data Assimilation
Additional Resources

• Email lists:
  – BUGS
    • In addition to an endless stream of people asking questions, will have occasional postings about workshops
    • Often useful nuggets of wisdom
  – R-sig-ecology
    • Mostly frequentist, but many useful R tricks for ecological data
    • Equivalent lists for many other subjects: e.g. genetics, phylogenetics, geostatistics, dynamic models
Where am I going from here...

- Course will next be offered Spring 2012
- I already know the labs have to be simplified
- I already know Clark book is difficult
- Looking for feedback beyond just ICES
  - How can I make lecture more interactive?
  - What topics should be covered more/less?
  - What labs/lectures were most/least useful?
  - Should we do project presentations?
  - Exposure vs expertise?
  - What would increase proficiency / independence?
It usually takes multiple exposures to this material before it really sinks in