Announcements

- Today’s reading Dietze Chapter 2 “From Models to Data”
- Hands-on Activity 1 due 2/4
- Wednesday 2/6 Discussion “Predictability”
  - Petchey et al. 2015
- Friday 2/8 reading: Ch 3 - Data
How Theory is Taught
Logistic Growth
Logistic Growth

Unstable

Stable
Fit to data?

Log Transformed Moose Pop

Log of Moose Pop

years


Exp
Ricker
95% CI
**Fit to data?**

**Observed vs. Modeled Moose Density**

**Linear Scale**

- Data
- Mean Ricker Model
- Process Model Minus Error

![Graph showing observed vs. modeled moose density over time. The graph illustrates different fits as FUNCTION and PROCESS.]
Fit to data?

Focuses on equilibrium

Real world problems are transients

Focuses on determinism

Real world is heterogeneous and noisy

Uncertainty?
Observation Error

\[ N_{t+1} = N_t + rN_t \left(1 - \frac{N_t}{K}\right) \]

\[ N_t^{(O)} \sim N(N_t, \tau_{obs}) \]
Parameter Error

\[ N_{t+1} = N_t + rN_t \left(1 - \frac{N_t}{K}\right) \]

\[ \begin{bmatrix} r \\ K \end{bmatrix} \sim N_2 \left(\begin{bmatrix} r_0 \\ K_0 \end{bmatrix}, \Sigma_{\text{param}}\right) \]
Initial Conditions

$N_{t=0} \sim logN(N_0, \tau_{IC})$
Additive Variability

\[ N_{t+1} = N_t + r_N t \left( 1 - \frac{N_t}{K} \right) + \varepsilon_t \]
\[ \varepsilon_t \sim N(0, \tau_{add}) \]

Parameter Variability

\[ N_{t+1} = N_t + r_t N_t \left( 1 - \frac{N_t}{K_t} \right) \]
\[ \begin{bmatrix} r_t \\ K_t \end{bmatrix} \sim N_2 \left( \begin{bmatrix} r_0 \\ K_0 \end{bmatrix}, \Sigma_{process} \right) \]
1 state variable
3 parameters
11 uncertainties
Think Distributions !!

- What is a random variable
- What is a probability distribution
- Common distributions and their representation in R
Probability distributions
What is a random variable?

“a variable that can take on more than one value, in which the values are determined by probabilities”

Does not have a single, fixed value
Coin Flips

Flip $n$ coins. Count the number of heads.
What is a probability distribution?

A function that assigns a probability to a random variable

\[ P(X = x_k) = p_k \]

given:

\[ 0 \leq p_k \leq 1 \]

\[ \sum p_k = 1 \]

\[ P(X = 3) = \frac{1}{6} \]
Discrete distributions

probability of a given number of events

Poisson($X \mid \lambda$)
Continuous distributions

- X is a continuous random variable
- $f(X) > 0$
- $\int f(X) \, dx = 1$

$Pr(z \leq Z \leq z + \delta z) = \int_{z}^{z+\delta z} f(z) \, dz$
Absent: beta, binomial, gamma, exponential, Laplace, Pareto, Bernoulli, geometric, hypergeometric, Wishart
Drawing random numbers
Monte Carlo Simulation

![Graph showing a line and data points labeled n = 1.](image-url)
Can we forecast ecology like we forecast weather?
HOW DO WE MEASURE PREDICTABILITY?

$Y_{t+1} = f(Y_t, X_t | \bar{\theta} + \alpha) + \varepsilon$
What causes var to increase with time?

\[
\text{Var}[Y_{t+1}] \approx \left( \frac{\partial f}{\partial Y} \right)^2 \text{Var}[Y_t] + \left( \frac{\partial f}{\partial X} \right)^2 \text{Var}[X] + \left( \frac{\partial f}{\partial \theta} \right)^2 \left( \text{Var}[\theta] + \text{Var}[\alpha] \right) + \text{Var}[\varepsilon]
\]

= INTERNAL + EXTERNAL + PARAMETERS + RANDOM EFFECTS + PROCESS ERROR
INTERNAL STABILITY

\[ \text{Var}[Y_{t+1}] \approx \left( \frac{\partial f}{\partial Y} \right)^2 \text{Var} \left[ \frac{Y_t}{\text{stability}} \right] + \text{Var} \left[ \frac{\text{IC}}{\text{uncert}} \right] \]
WEATHER FORECASTING: AN INITIAL CONDITIONS PROBLEM

\[ \text{Var}[Y_{t+1}] \approx \left(\frac{\partial f}{\partial Y}\right)^2 \text{Var}[Y_t]^{\text{stability}} \text{Var}[Y_t]^{\text{ic uncertain}} \]

Slingo & Palmer. 2011. Phil. Trans. R. Soc. A
INTERNAL STABILITY

All other terms grow linearly
**EXOGENOUS STABILITY**

+ \( \left( \frac{\partial f}{\partial X} \right)^2 \frac{\text{Var}[X]}{\text{driver}_{\text{sens}}} \)
+ \( \frac{\text{Var}[X]}{\text{driver}_{\text{uncert}}} \)

- Predictable if low sensitivity or low uncertainty

- Anova vs Regression design: *How much does X affect Y?*

- \( \text{Var}[x] \) also needs to be forecast
  - Different X for forecast than explain?
  - Not in model select, over complex
  - Rel. importance increases with time

- Endogenous (DD) vs Exogenous (DI) continuum
PARAMETER ERROR

\[ + \left( \frac{\partial f}{\partial \theta} \right)^2 \left( \frac{\text{Var}[\theta]}{\text{param uncert}} + \frac{\text{Var}[\alpha]}{\text{param variability}} \right) + \frac{\text{Var}[\varepsilon]}{\text{process error}} \]
PROCESS ERROR

\[ + \left( \frac{\partial f}{\partial \theta} \right)^2 \left( \frac{\text{Var}[\tilde{\theta}]}{\text{param uncertain}} + \frac{\text{Var}[\alpha]}{\text{param variability}} \right) + \text{Var}[\varepsilon] \]

• Inherent stochasticity (irreducible)
• Structural uncertainty
• Heterogeneity & variability
  - need to accommodate, even if can’t explain
COV & SCALING

- At large scales, average over drivers \((X)\), heterogeneity \((\alpha)\), & variability \((\varepsilon)\)

- Internal stability \((Y)\) increases in importance

- Scaling very dependent on spatial and temporal auto- & cross-correlation

\[
\sum \sum \frac{\partial f}{\partial X_i} \frac{\partial f}{\partial X_j} COV[X_i, X_j]
\]
If we added CI, which is most important?

Confounds structure, driver and parameter error

Data-free IC

No process error or variability

Friedlingstein et al 2006

Annual land flux (PgCyr⁻¹)

Year
NATURE OF THE PREDICTION PROBLEM...

• Theory
  • What drives dynamics?
  • Generality across processes and locations

• Practice
  • What can we predict?
  • How to tackle new systems

• Methods
  • What to measure
  • How we build models
  • How we assimilate data

\[ \text{Var}[Y_{t+1}] \approx \left( \frac{\partial f}{\partial Y} \right)^2 \text{Var}[Y_t] + \left( \frac{\partial f}{\partial X} \right)^2 \text{Var}[X] + \left( \frac{\partial f}{\partial \theta} \right)^2 (\text{Var}[\theta] + \text{Var}[\alpha]) + \text{Var}[\epsilon] \]

\[ = \text{INTERNAL} + \text{EXTERNAL} + \text{ERROR} \]
DISCOVER WHETHER NATURE IS PREDICTABLE