**Forecasting**

**Projection**

“**Probabilistic statement that it is possible that something will happen in the future**” given boundary condition scenarios.

**Prediction**

“**Probabilistic statement that something will happen in the future based on what is known today**”

---

MacCracken 2001
WHY FORECAST?
CLIMATE CHANGE

Stationarity Is Dead: Whither Water Management?

P. C. D. Milly, Julio Betancourt, Malin Falkenmark, Robert M. Hirsch, Zbigniew W. Kundzewicz, Dennis P. Lettenmaier, Ronald J. Stouffer

Climate change undermines a basic assumption that historically has facilitated management of water supplies, demands, and risks.

Science 2008

DECISIONS ARE ABOUT THE FUTURE
Forecasts should be updated when new data becomes available.
EcoCast is a dynamic ocean management tool that aims to minimize fisheries/yield and maximize fisheries/yielding catch in real-time. It maps the daily relative yield/ target catch probabilities. Species weightings reflect management priorities and recent catch events. Environmental data are used to predict areas where species are likely to be found.

Mean Particulate Domoic Acid Probability: Jun-01-2018 to Jun-30-2018

California Harmful Algae Risk Mapping
(C-HARM)
California Harmful Algae Risk Mapping (C-HARM)
REAL-TIME MONITORING

= REAL-TIME SCIENCE
NETWORK SCIENCE

TERN, LTER, CZO, FluxNET, GLEON, NutNet, etc.
FORECASTS ARE QUANTITATIVE, SPECIFIC, & FALSIFIABLE

NEAR-TERM ITERATIVE FORECASTING IS WIN-WIN
Forecasts are a priori and out-of-sample.
Can we forecast ecology like we forecast weather?

- Theory
- Methods
- Think Probabilistic
HOW DO WE MEASURE PREDICTABILITY?

\[ Y_{t+1} = f(Y_t, X_t | \bar{\theta} + \alpha) + \varepsilon \]

- Background/null uncertainty
- Forecast limit
- Decision relevant scales
- IC

Time, Space
WHAT CAUSES VAR TO INCREASE WITH TIME?

\[
\text{Var}[Y_{t+1}] \approx \left( \frac{\partial f}{\partial Y} \right)^2 \frac{\text{Var}[Y_t]}{\text{IC uncert}} + \left( \frac{\partial f}{\partial X} \right)^2 \frac{\text{Var}[X]}{\text{driver uncert}} + \left( \frac{\partial f}{\partial \theta} \right)^2 \left( \frac{\text{Var}[\theta]}{\text{param uncert}} + \frac{\text{Var}[\alpha]}{\text{param variability}} \right) + \text{Var}[\epsilon] \text{ process error}
\]

\[
= \text{INTERNAL} + \text{EXTERNAL} + \text{PARAMETERS} + \text{RANDOM EFFECTS} + \text{PROCESS ERROR}
\]
Logistic Growth

Moose Density

Year

- Data
- FIT AS PROCESS
- FIT AS FUNCTION

INTERNAL STABILITY

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

All other terms grow linearly
$N_{t=0} \sim \log N(N_0, \tau_{IC})$
WEATHER FORECASTING: AN INITIAL CONDITIONS PROBLEM

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

\[ \text{IC uncertainty} \]
EXOGENOUS STABILITY

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

- Anova vs Regression design: \textit{How much} does X affect Y?
- Endogenous (DD) vs Exogenous (DI) continuum
PARAMETER ERROR

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

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

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

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

• Inherent stochasticity (irreducible)

• Structural uncertainty

• Heterogeneity & variability

  • need to accommodate, even if can’t explain
Additive Variability

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

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 !!
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?

Data-free IC
No process error or variability
Confounds structure, driver and parameter error

Friedlingstein et al 2006
How do the drivers of forecast uncertainty vary across ecological system?

Thomas *unpublished*

**UNCERTAINTY:**
- Green: driver: meteorology
- Dark Green: driver: meteorology downscaling
- Black: initial conditions (IC)
- Red: parameter
- Light Blue: process

**Surface water temperature**

**Forest net ecosystem exchange**
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

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

\[ = \text{INTERNAL} + \text{EXTERNAL} + \text{ERROR} \]
DISCOVER WHETHER NATURE IS PREDICTABLE
Willow Creek Forecast for 2019-07-10

To: Michael Dietze

Below is today's 16 day forecast for NEE, LE, and Soil Moisture Fraction at Willow Creek, Wisconsin.

Net Ecosystem Exchange Forecast for 2019-07-10
PREDICTABILITY INCREASES WITH SCALE

...both for functional groups

Out-of-sample validation at NEON sites
Why Forecast ➔ Theory ➔ Vision / Challenges

**NCEP Operational Forecast Skill**

36 and 72 Hour Forecasts @ 500 MB over North America

[100 * (1-SI/70) Method]

- 36 Hour Forecast
- 72 Hour Forecast

**NCEP Central Operations January 2015**
Forecasts should be updated when new data becomes available.

**Initial Conditions (IC)**

**Drivers**

**Parameter (Param)**

**Forecast (Fx)**

**Analysis (Obs)**

**Model (MODEL)**

**Data (DATA)**
Precision controls influence

Less Precise Data

Less Precise Model
State-Variable Data Assimilation

\[ P(\theta|y) \propto P(y|\theta) P(\theta) \]

Updated State  Data  Model
ECOLOGICAL FORECASTING

• Is more than forward simulation
• Requires a fusion of models and data
• Must address multiple sources of uncertainty and variability
• Think Probabilistically!!