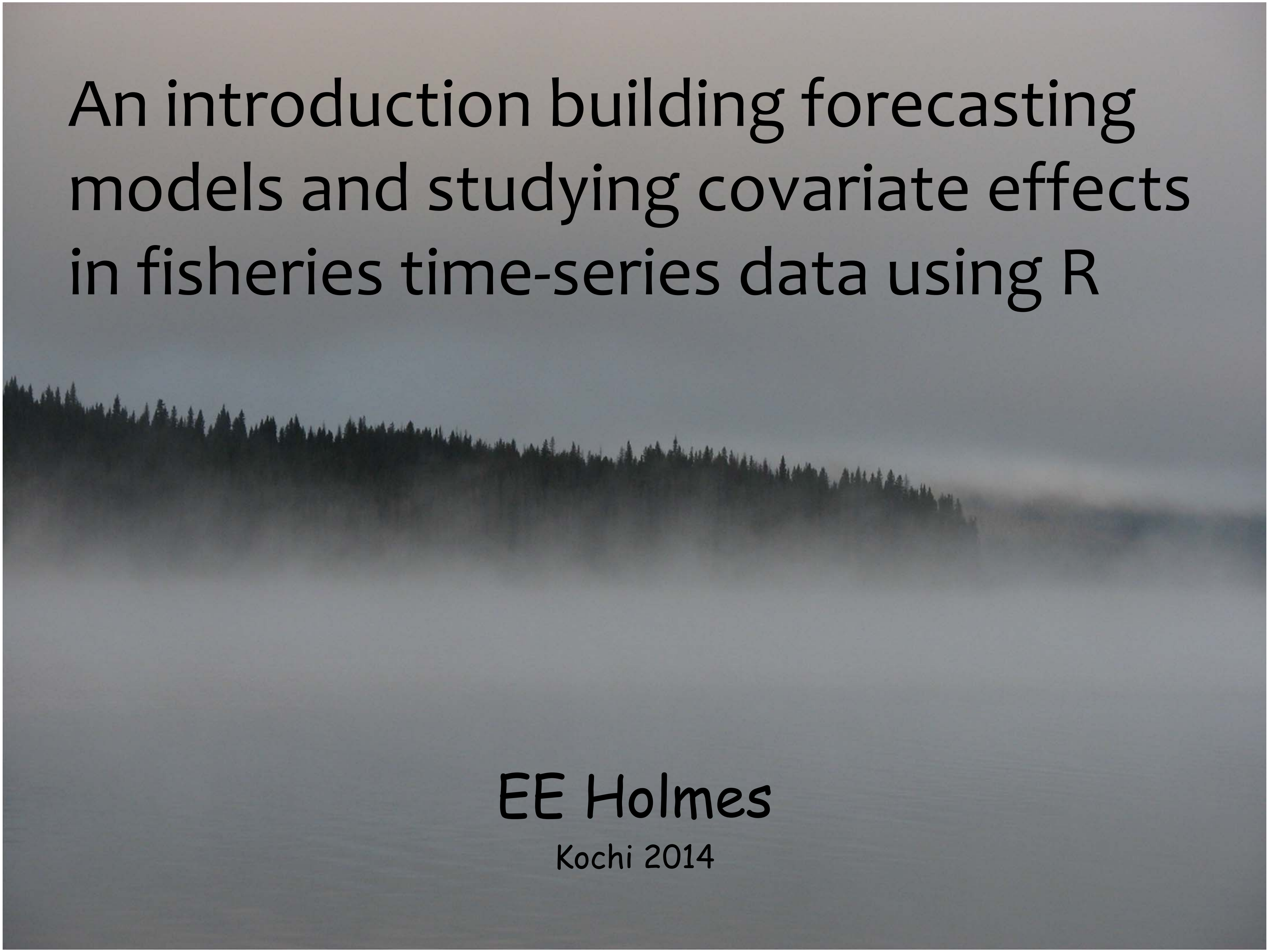


An introduction building forecasting models and studying covariate effects in fisheries time-series data using R

EE Holmes

Kochi 2014

A little bit about me

- I am a quantitative ecologist at a federal research lab

Northwest Fisheries Science Center

National Marine Fisheries Service

National Oceanic & Atmospheric Administration

Seattle, WA USA

- I am also an affiliate faculty member

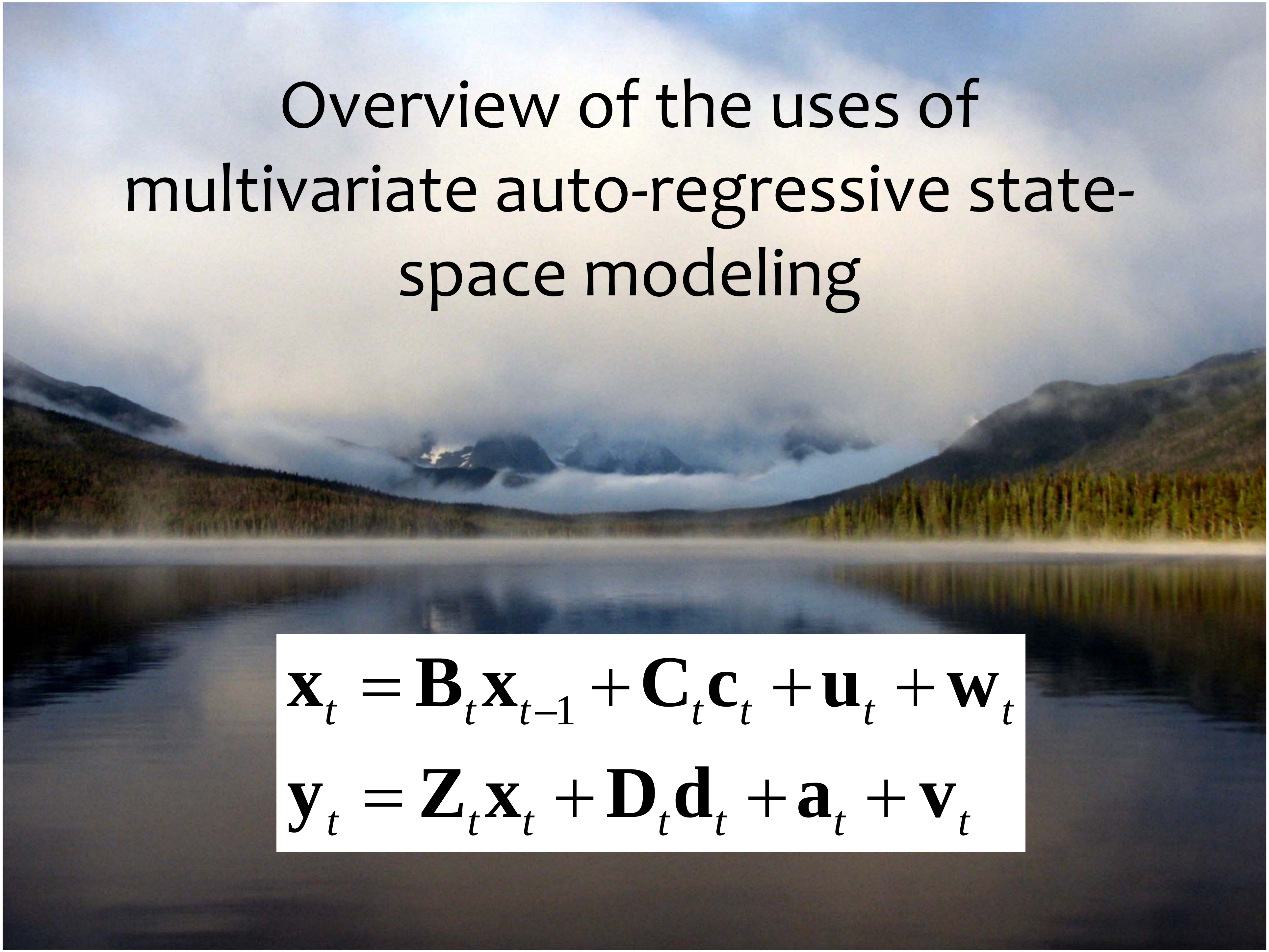
School of Aquatic and Fishery Sciences

University of Washington

- I work in a research group that develops statistical methods for the analysis of multivariate time series data

- In addition to teaching and publishing, I serve on federal scientific teams charged with analyses relating to status and risk assessment and management decisions
  - Endangered and threatened species
  - Pacific NW salmon
  - Puget Sound herring
  - Puget Sound rockfish
  - Southern Resident killer whales
  - Steller sea lions
  - Lake and marine plankton
  - Plus many other marine mammals

http://faculty.washington.edu/eeholmes

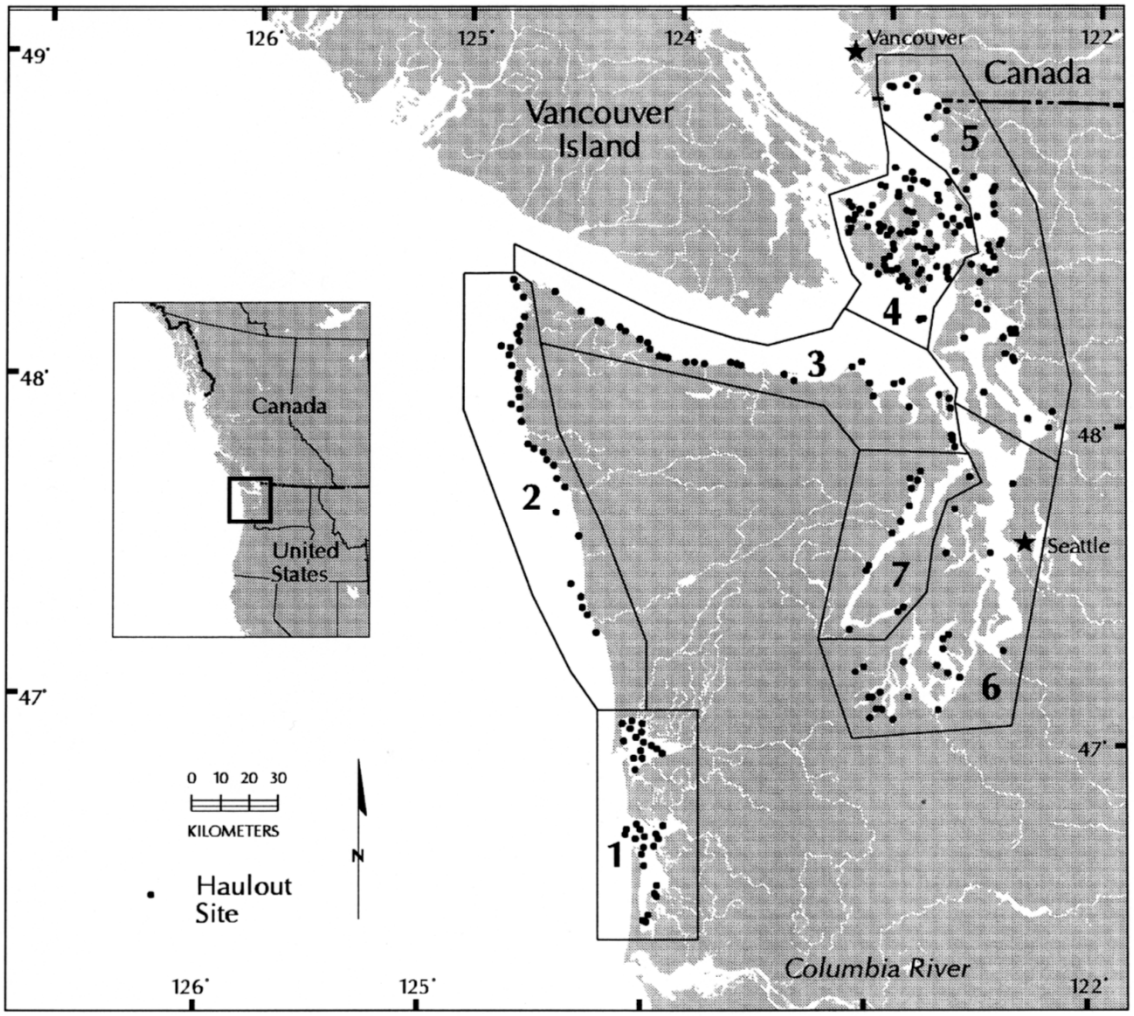


Overview of the uses of multivariate auto-regressive state-space modeling

$$\mathbf{x}_t = \mathbf{B}_t \mathbf{x}_{t-1} + \mathbf{C}_t \mathbf{c}_t + \mathbf{u}_t + \mathbf{w}_t$$

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{x}_t + \mathbf{D}_t \mathbf{d}_t + \mathbf{a}_t + \mathbf{v}_t$$

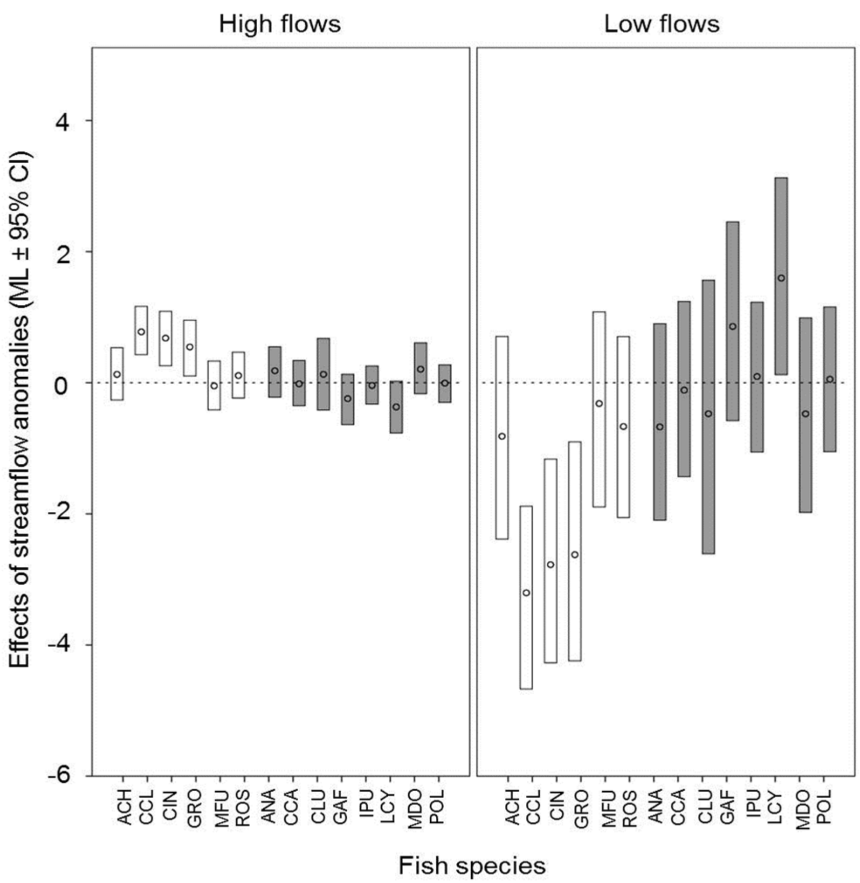
Combining multi-site data to estimate trends and identify spatial structure



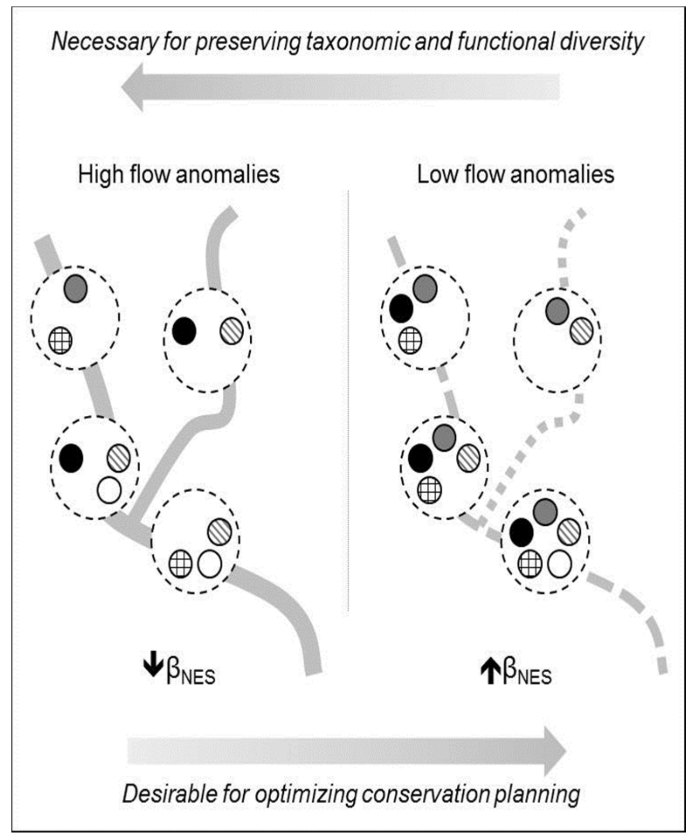
ISSUES

- Observation error
- Missing years
- Different types of sampling (boat vs plane vs land)
- Some sites more independent than others

Estimating the effects of covariates and seasonal drivers on population growth

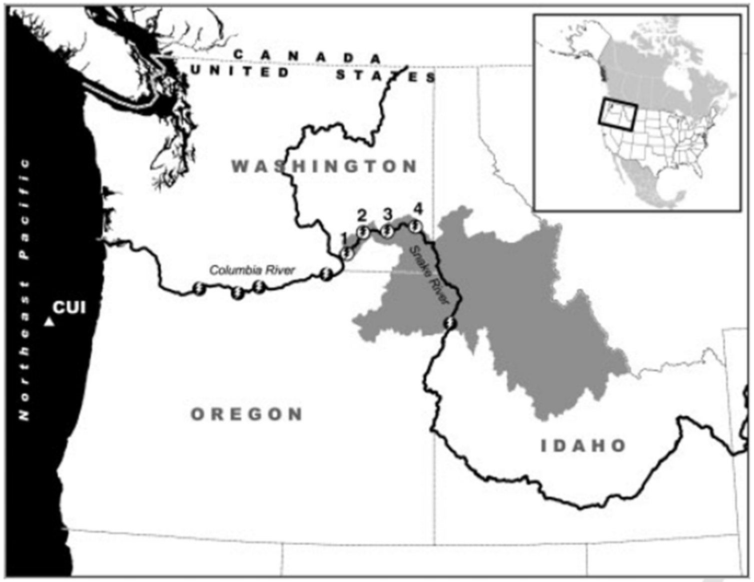


Desert fish native and non-native response to flash flood events in dry and wet season

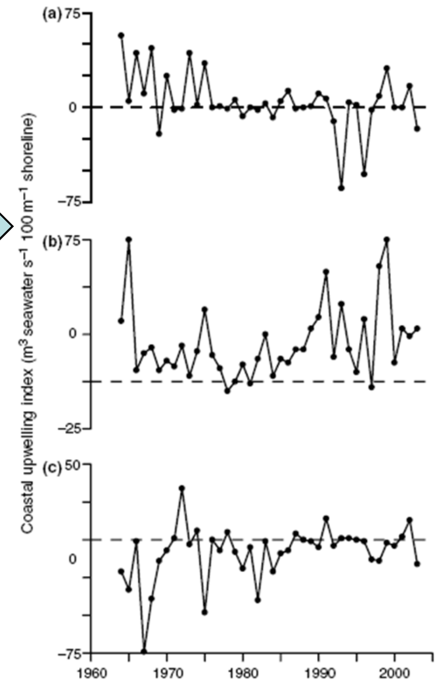


Recommendations about flow management

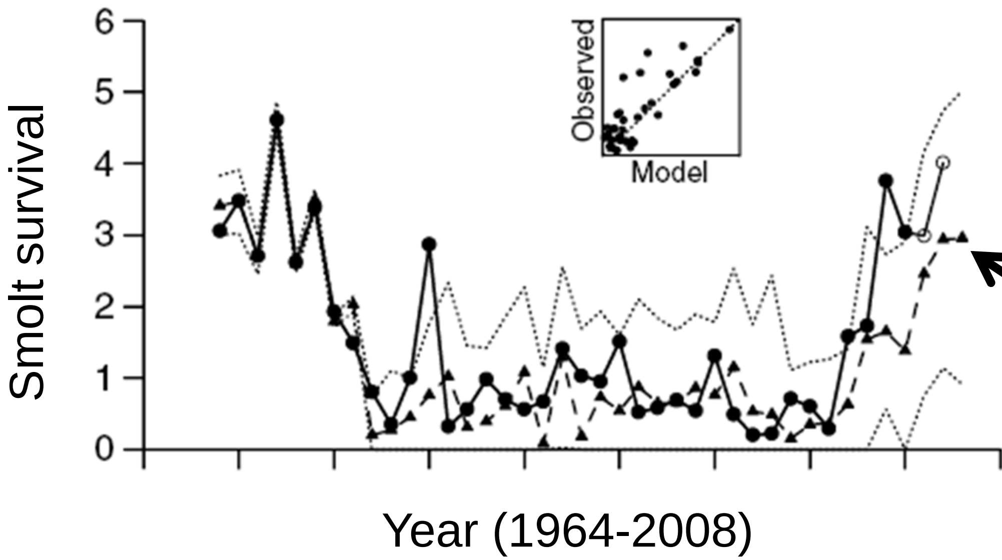
Developing forecasting models based on the relationship between fish abundance and covariates



Predictors (coastal upwelling) for smolt survival

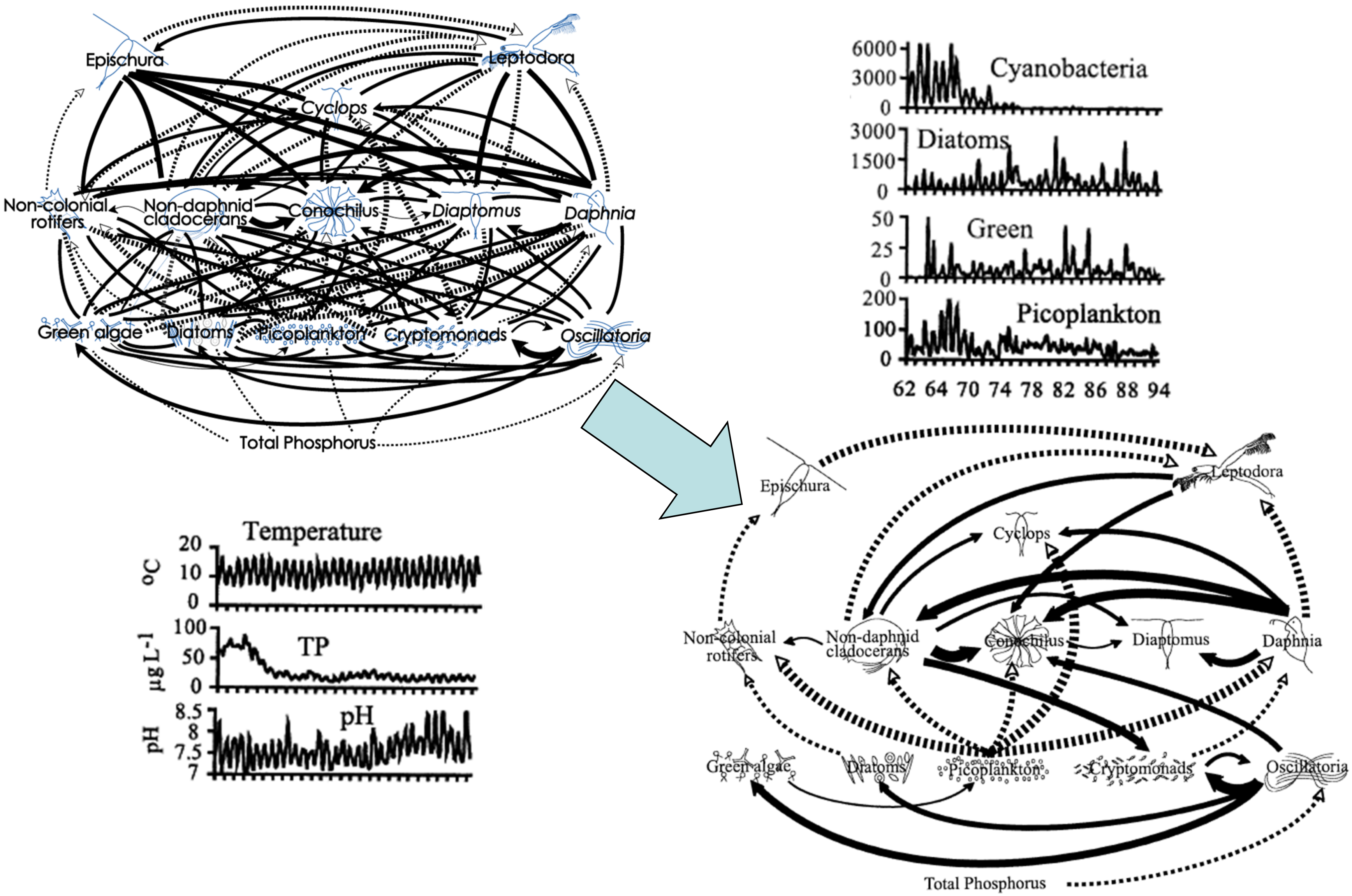


Smolt survival is used to set allowable harvest



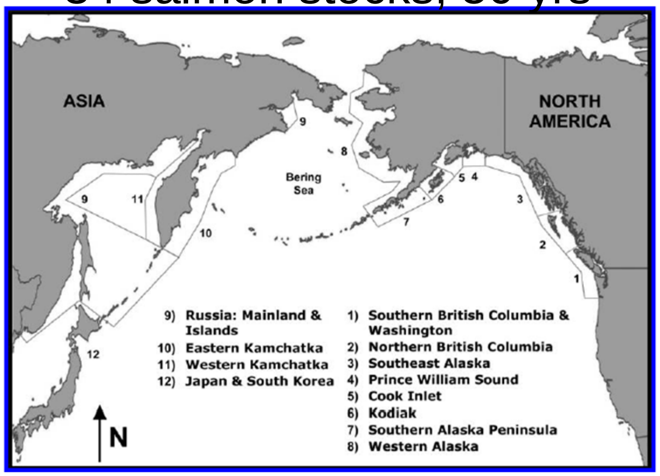
1 yr ahead Forecast that uses predictors but allows the relationship between predictor and survival to change

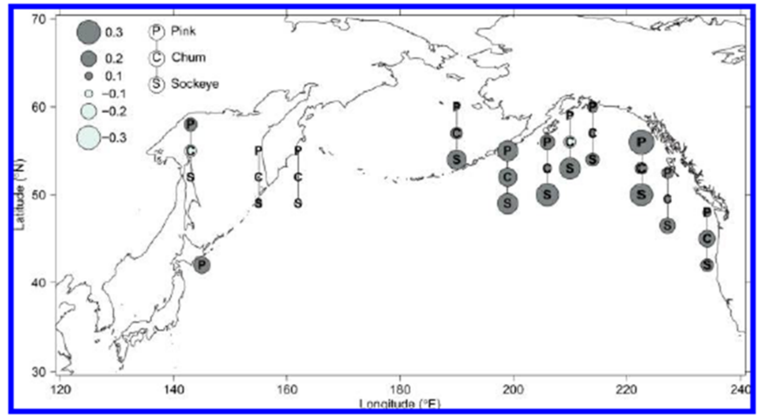
Estimating species interactions to understand community function and to forecast



Dynamic Factor Analysis (DFA)

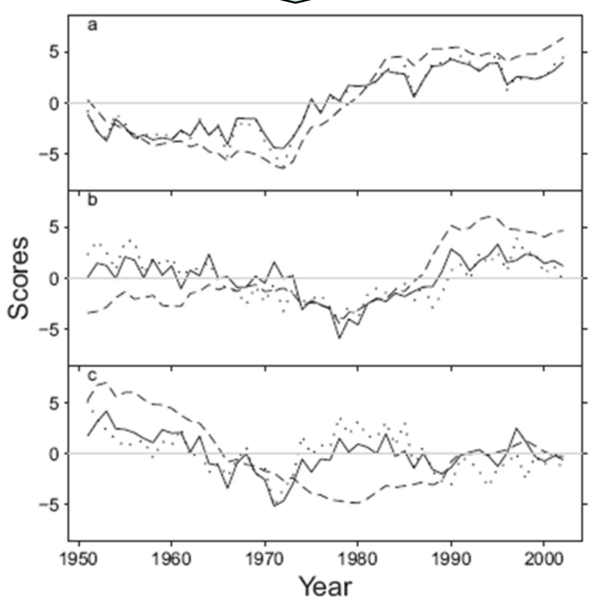
34 salmon stocks, 50 yrs

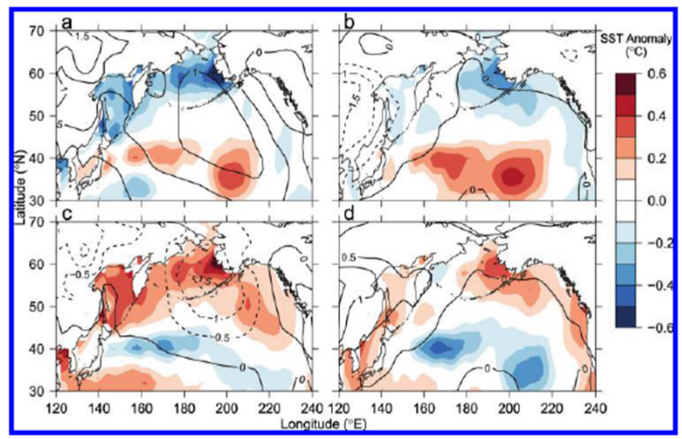




Look for patterns in which stocks are similarly correlated with drivers

Find the trends that explain this complex data set





Try to understand the environmental features that drive those trends

The multivariate AR state-space model

$$\mathbf{x}_t = \mathbf{B}_t \mathbf{x}_{t-1} + \mathbf{C}_t \mathbf{c}_t + \mathbf{u}_t + \mathbf{w}_t$$

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{x}_t + \mathbf{D}_t \mathbf{d}_t + \mathbf{a}_t + \mathbf{v}_t$$

$$\mathbf{w}_t \sim MVN(0, \mathbf{Q}_t)$$

$$\mathbf{v}_t \sim MVN(0, \mathbf{R}_t)$$

All the material I will cover is here:

http://tinyurl.com/Kochi2014

E. E. Holmes, E. J. Ward, and M. D. Scheuerell

Analysis of multivariate time-series using the MARSS package

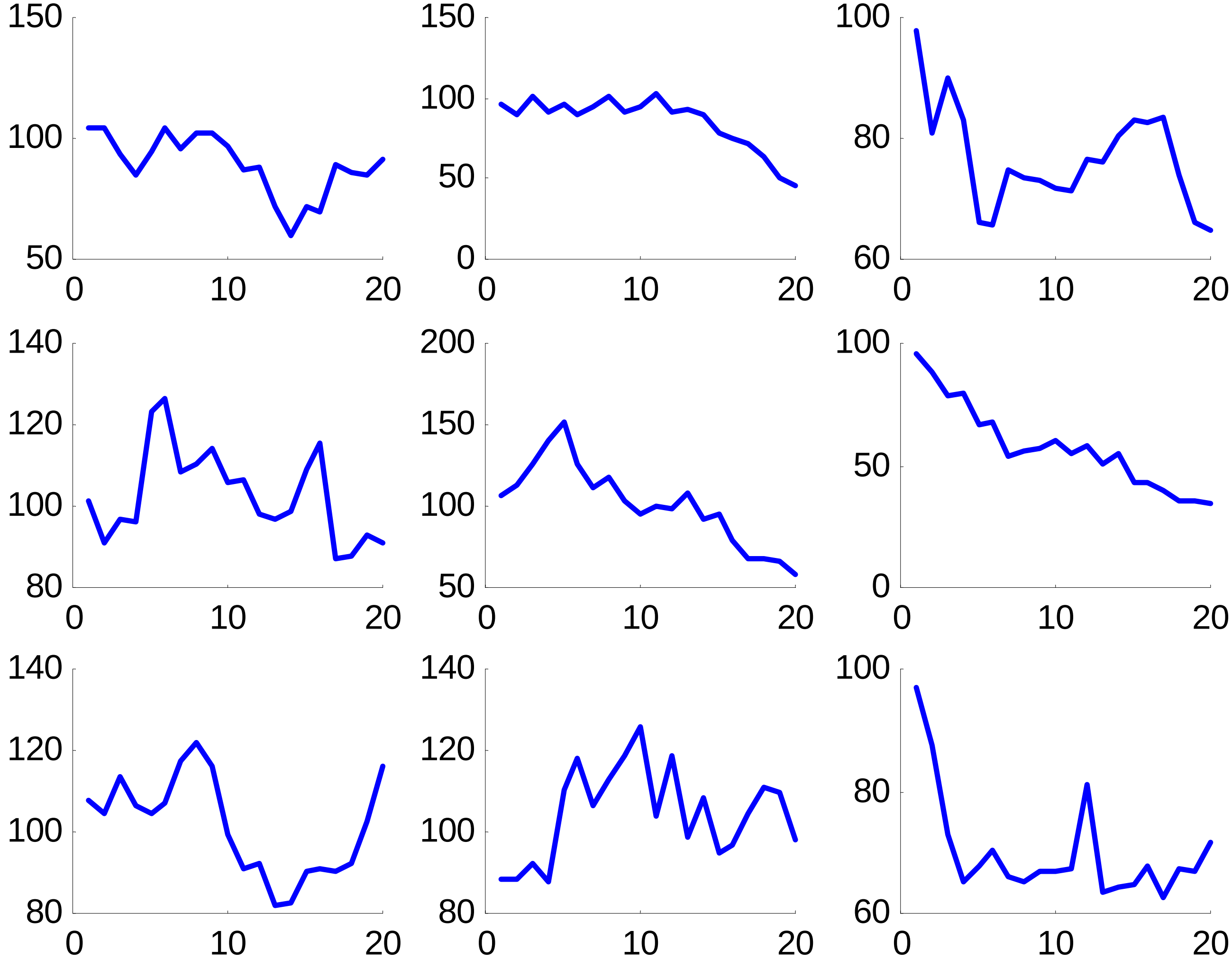
version 3.10.1

May 30, 2014

Northwest Fisheries Science Center, NOAA
Seattle, WA, USA

MARSS User Guide

Analysis of RANDOM WALKS



Random walks, aka AR processes, arise often in the analysis of fisheries data

- Many biological processes depend on the past

Today = f(Yesterday) + noise

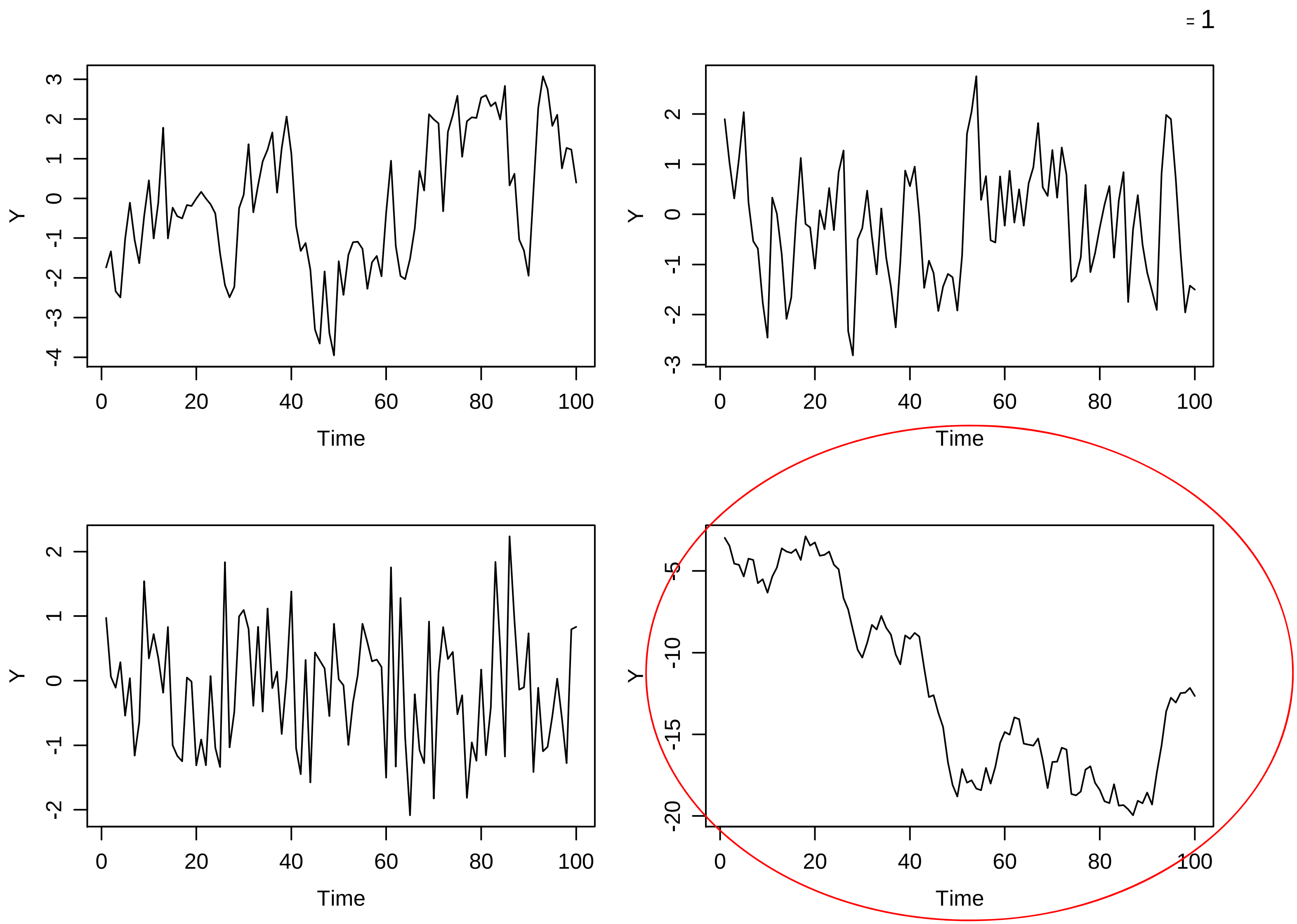
* animal movement

* gene frequency

* population growth

- fish
- algae
- birds

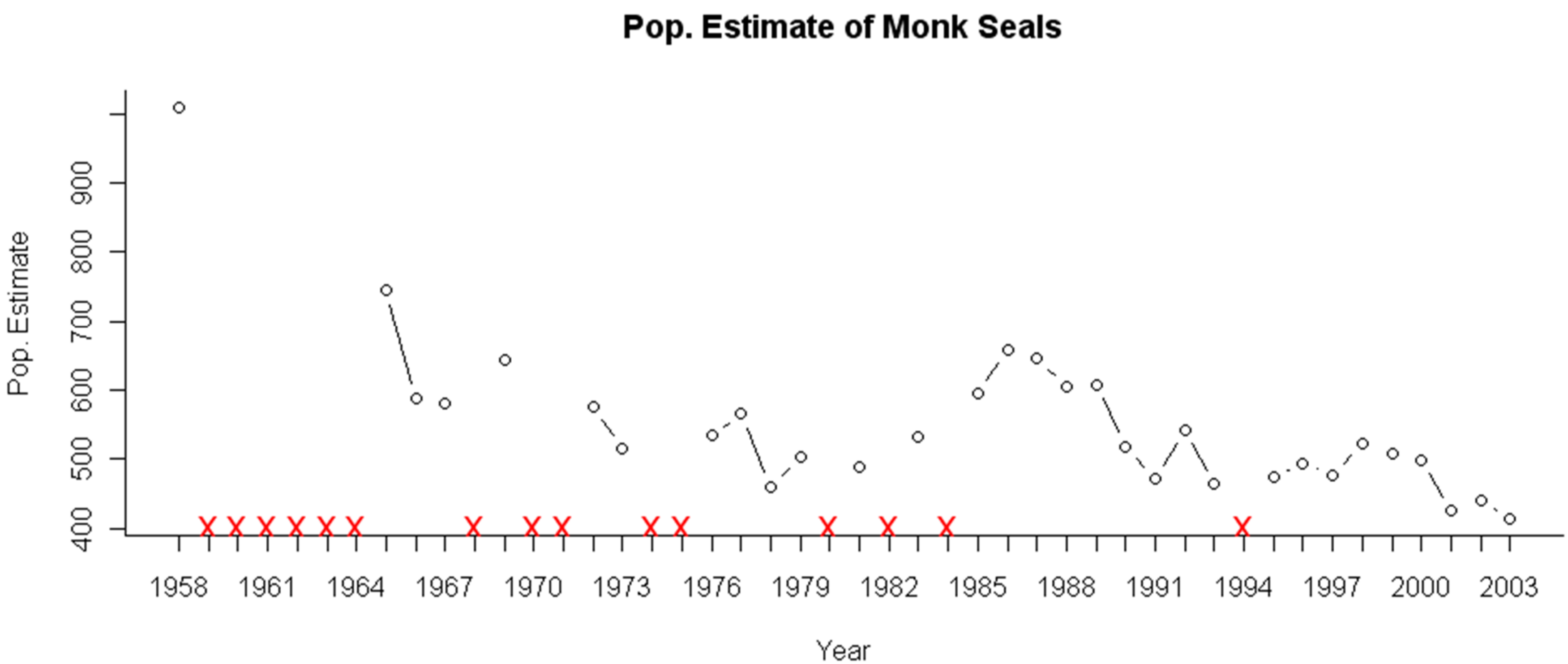
Classical time-series analysis focuses on stationary data. Population data is often non-stationary.



Topics

Introduction to univariate AR state-space models

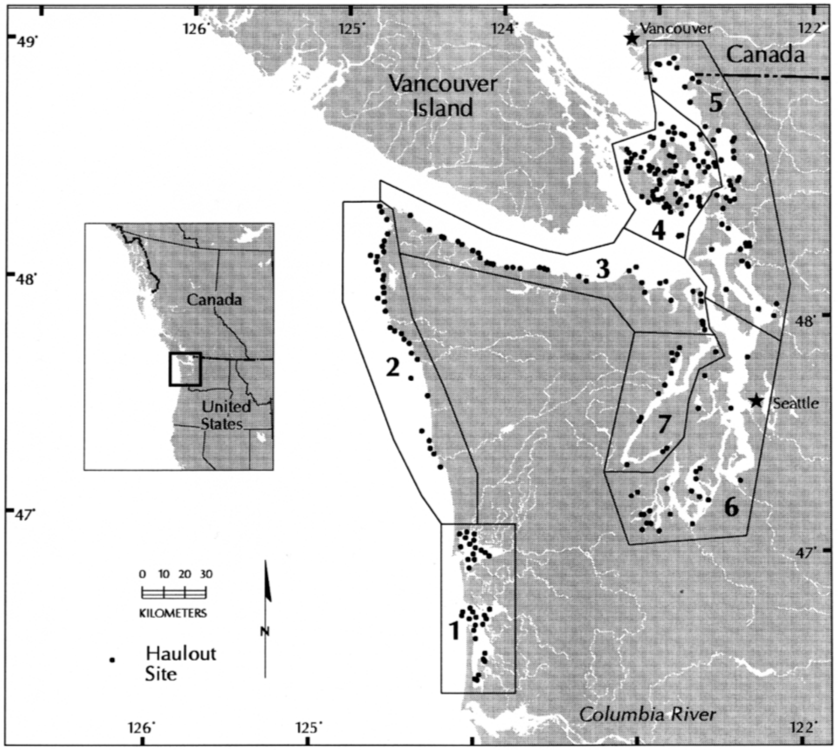
- Definition of process versus observation error
- **Hands on** with some R code and simulations.
- Adding density dependence (feedbacks)



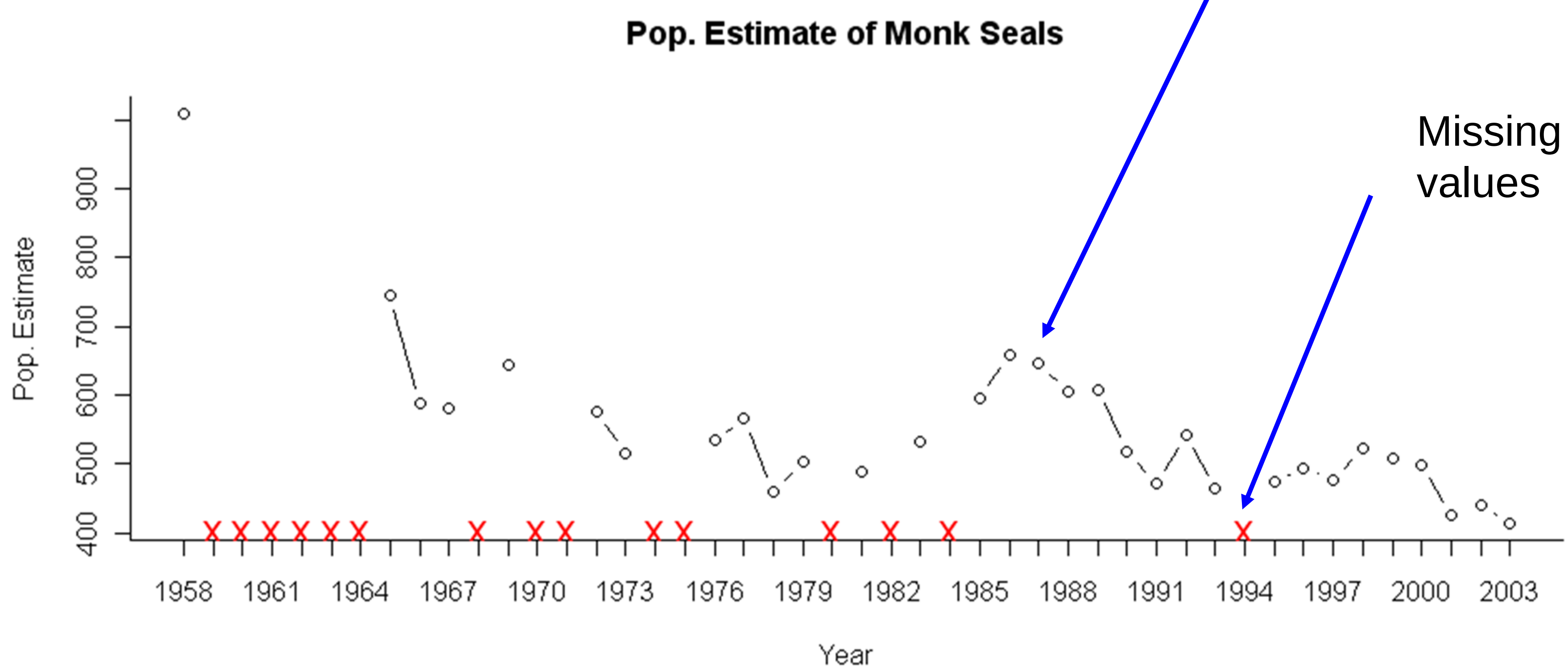
Topics

Introduction to multivariate AR lag-1 state-space models

- Integrating multi-site data into a single population estimate
- Creating forecasting models using time-series data and covariates



Example of univariate data: Count data



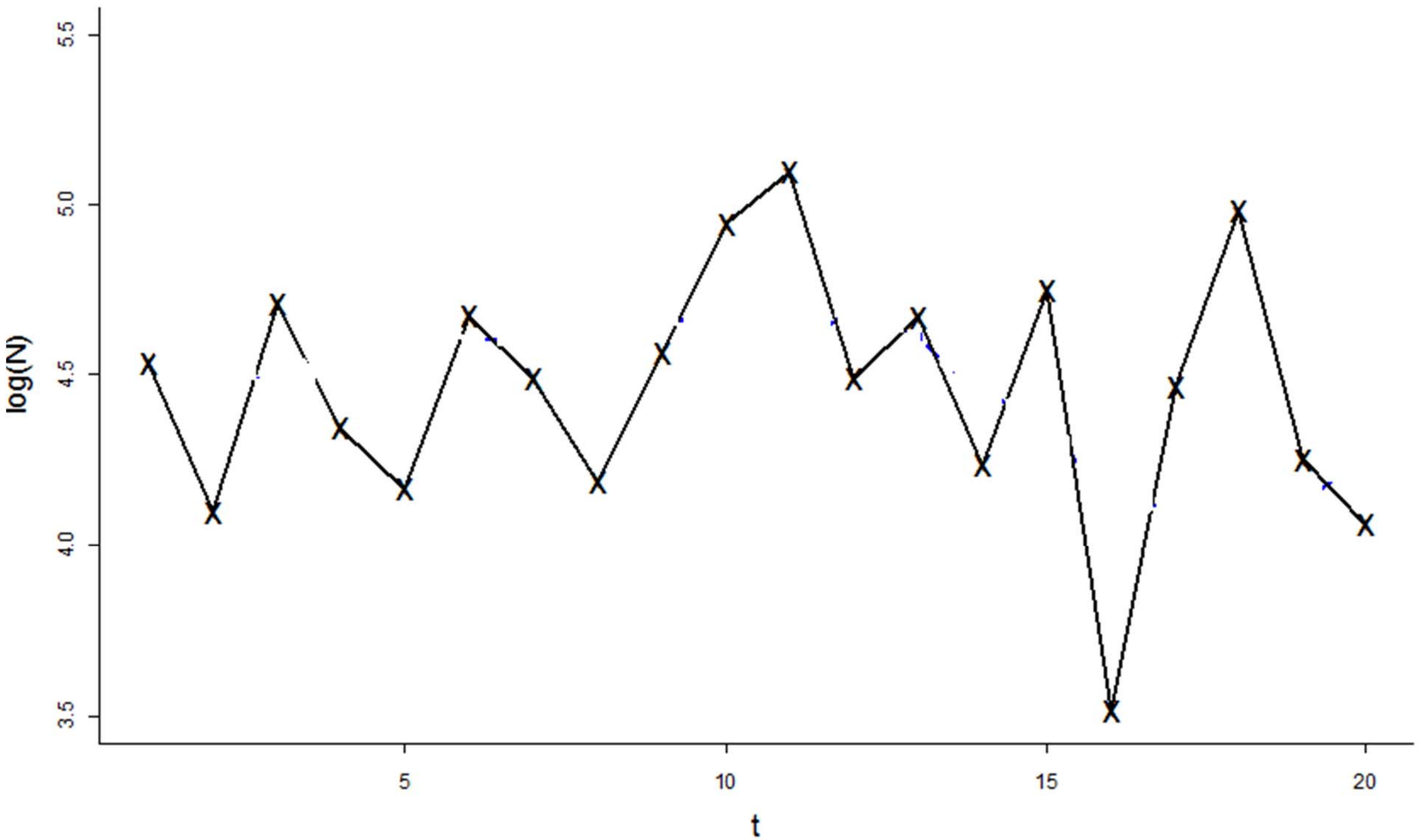
Yearly (usually) population (or subpopulation) counts

Missing values

Some of the variability is due to the population and some is due to observation error

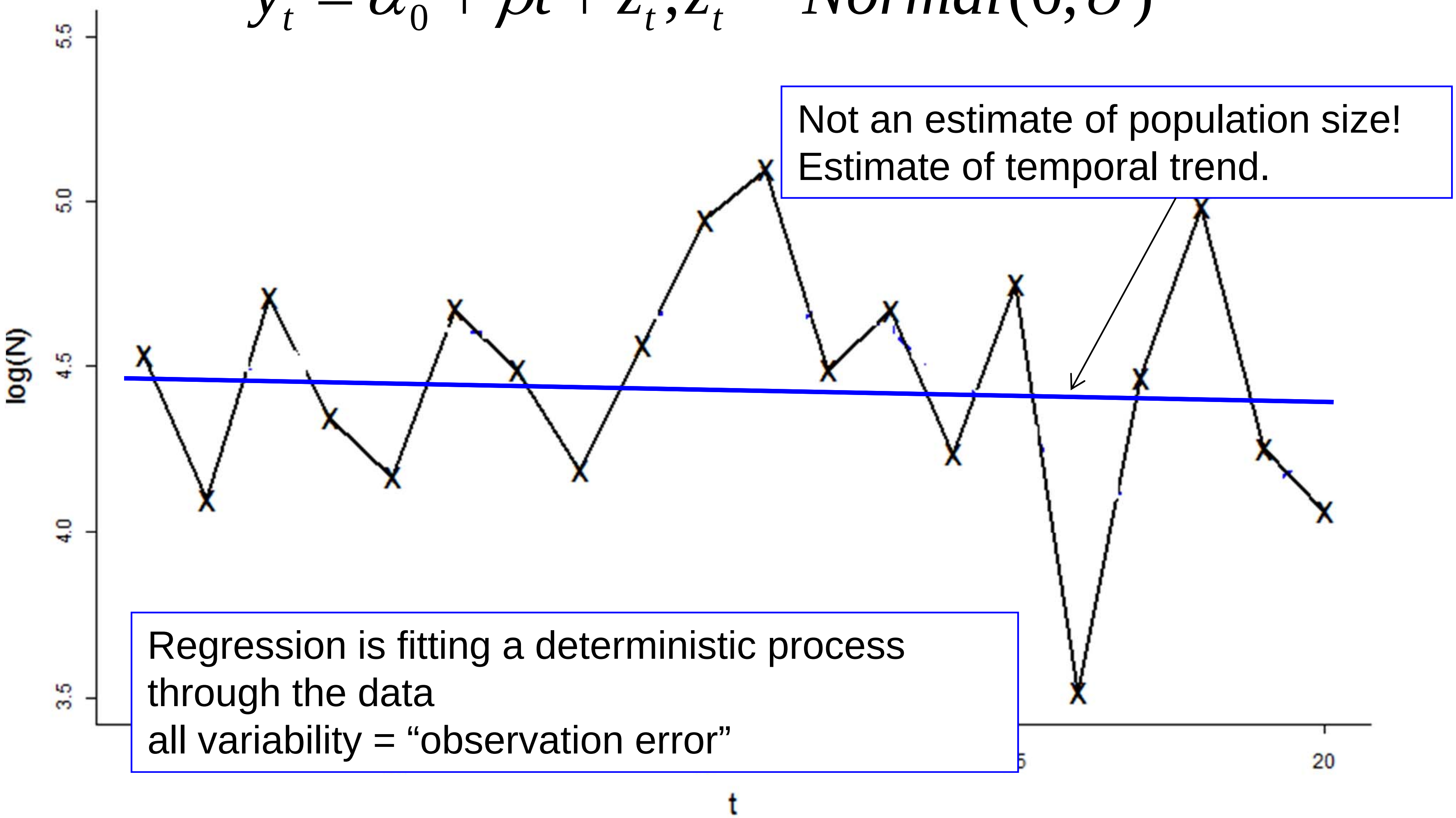
There IS some number of sea lions in our population in year x, but we don't know that number precisely. It is "hidden".

Suppose we have some count data (logged).

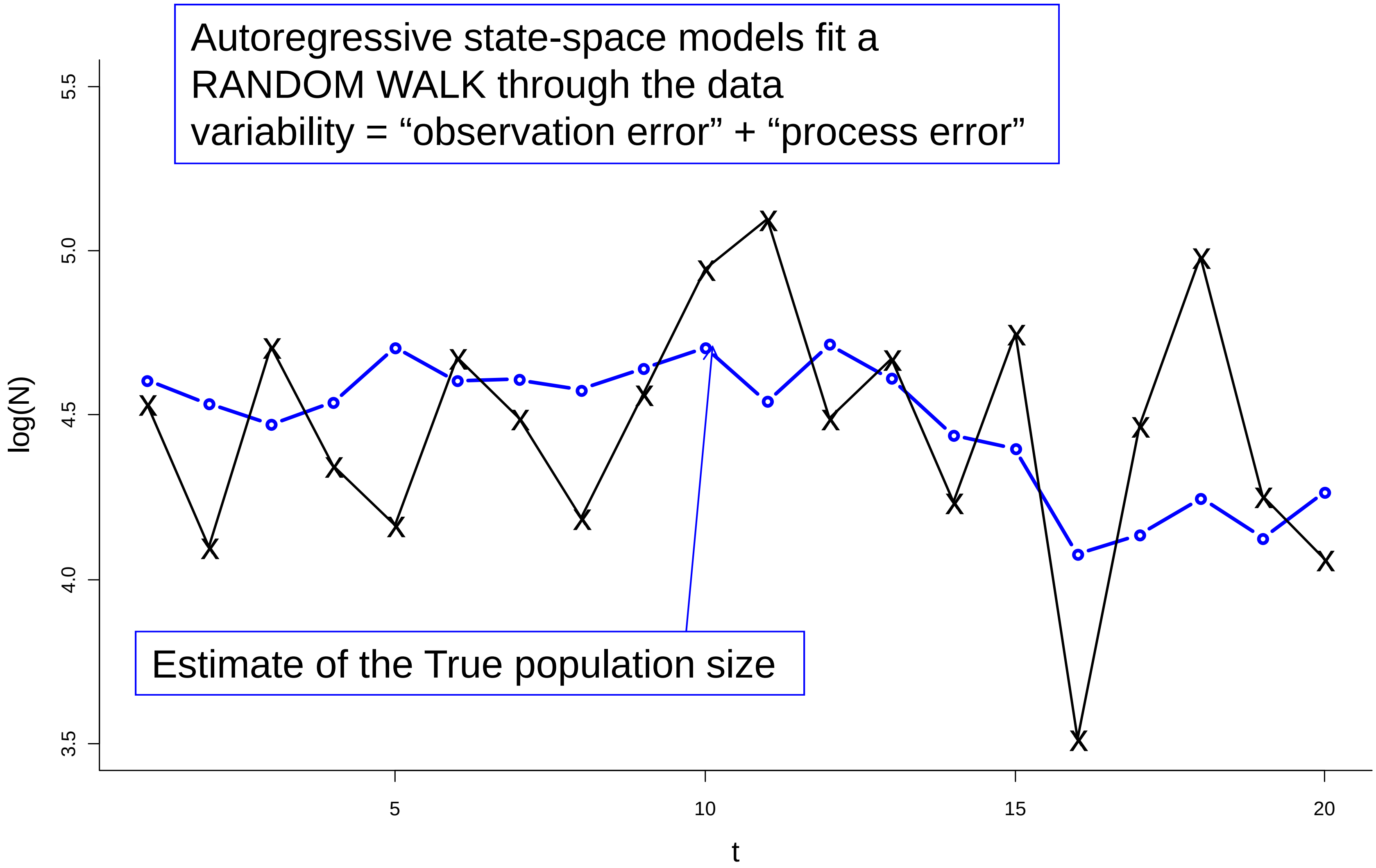


What about fitting a regression line through the data?

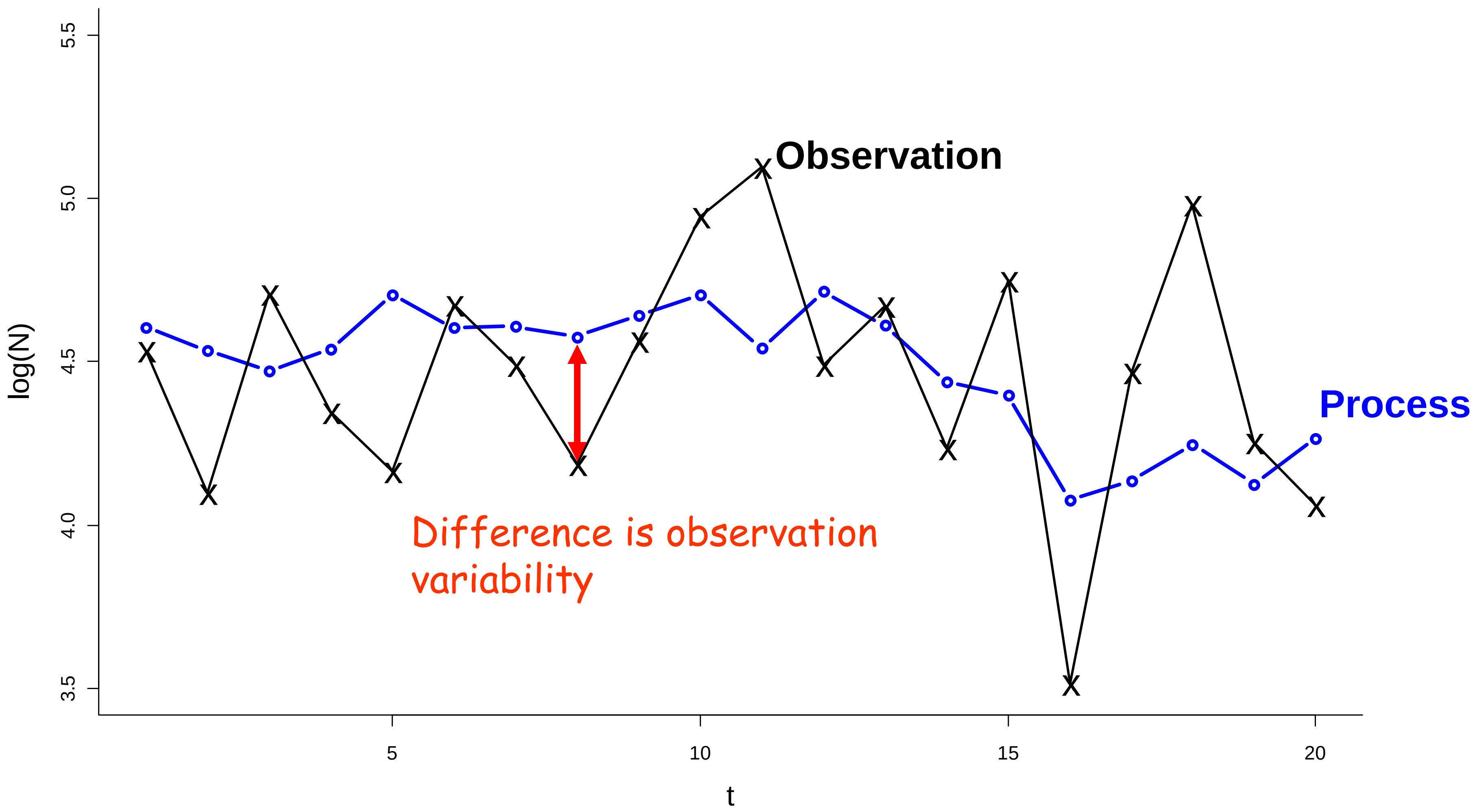
$$y_t = \alpha_0 + \beta t + z_t ; z_t \sim Normal(0, \sigma)$$



Versus fitting an autoregressive state-space model



Two types of variability
#1 observation or "non-process" variability

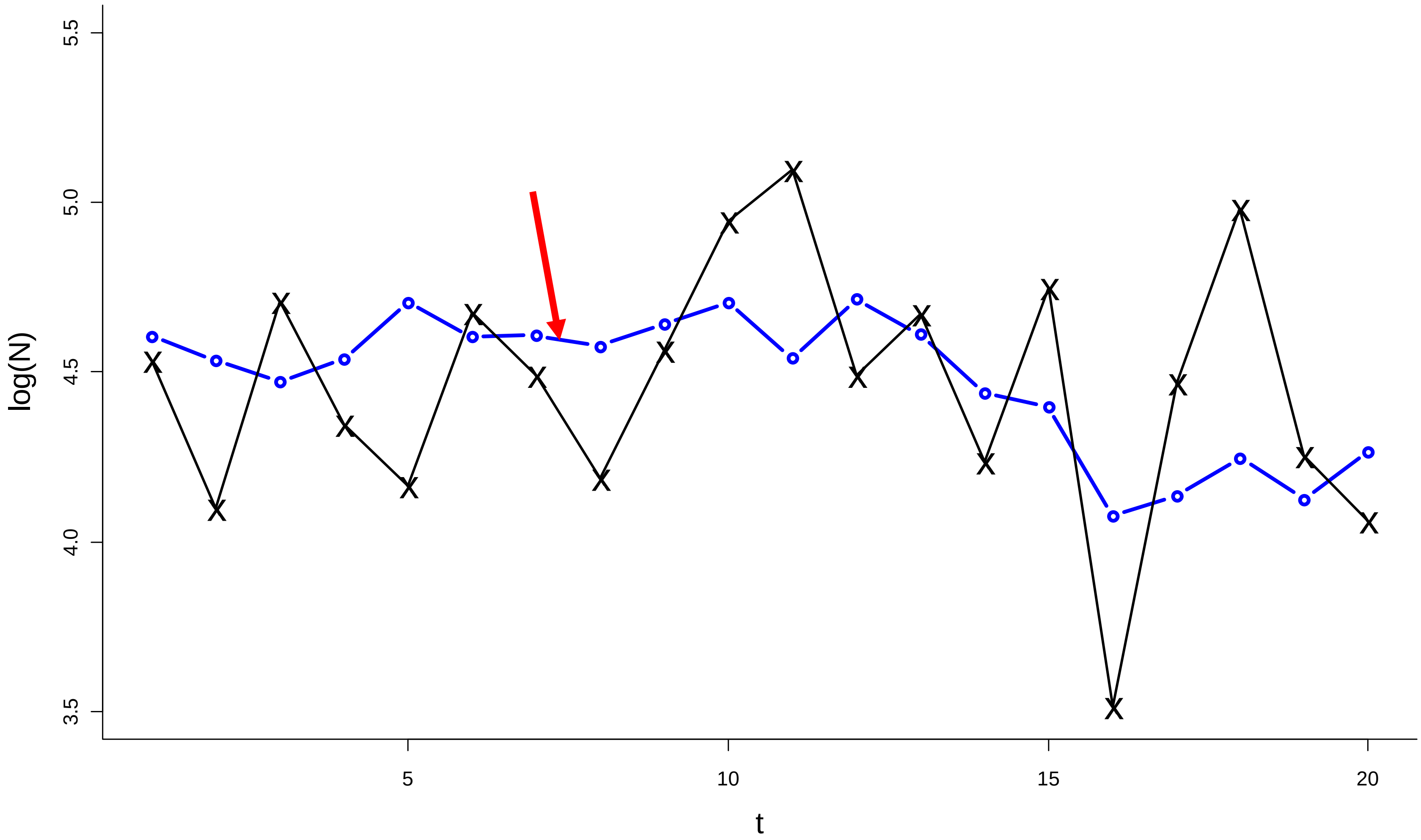


The observation variance (and bias) is often unknowable

- Sightability varies (year-to-year, day-to-day, etc.) due to a myriad of factors that may not be fully understood or measureable

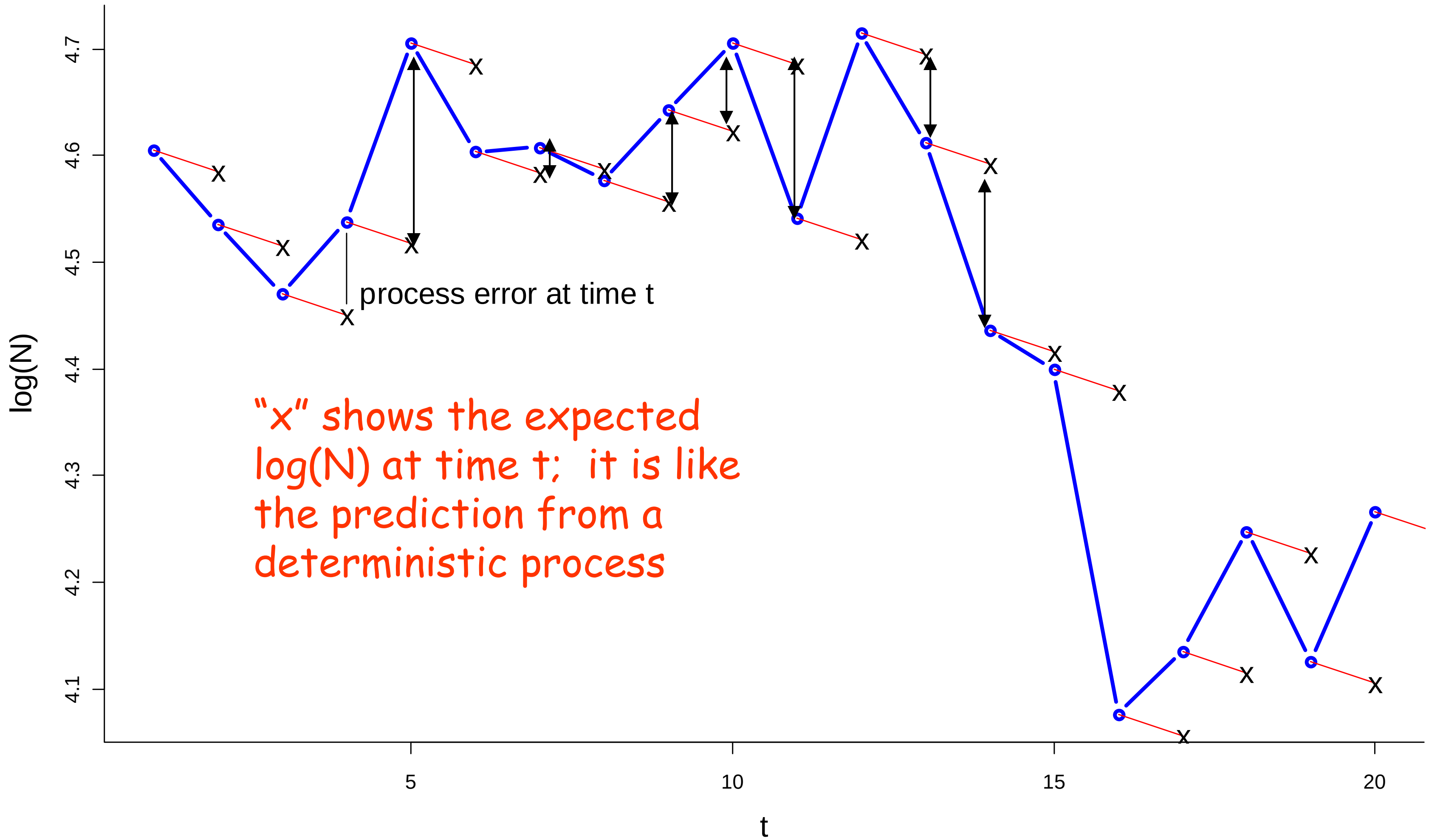
- Sampling variability--due to how you actually count animals--is just one component of observation variance

Two types of variability
#2 Process variability



Process error is the difference between the expected population size and the actual value

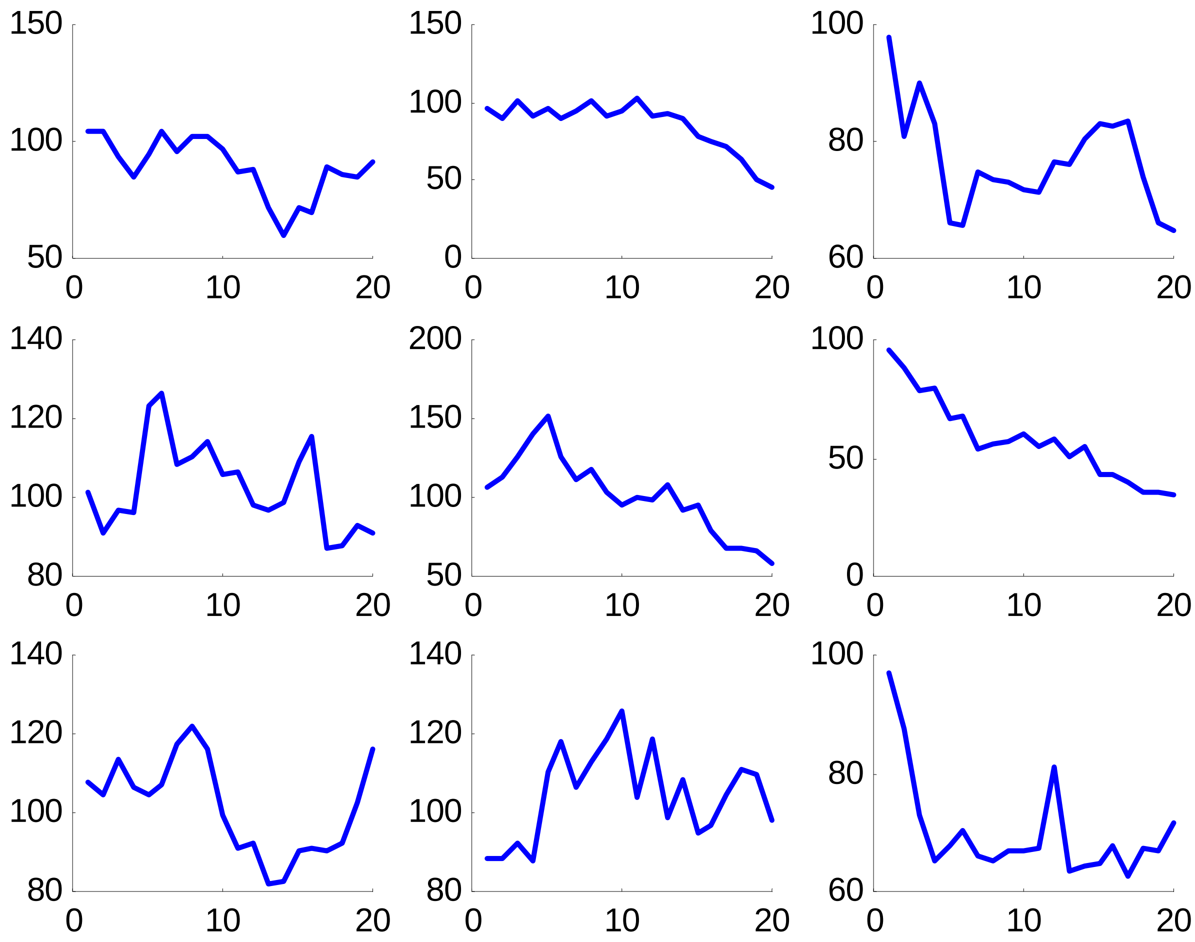
Let's say that the mean rate of decline is 2% per year*…



*u = -0.02

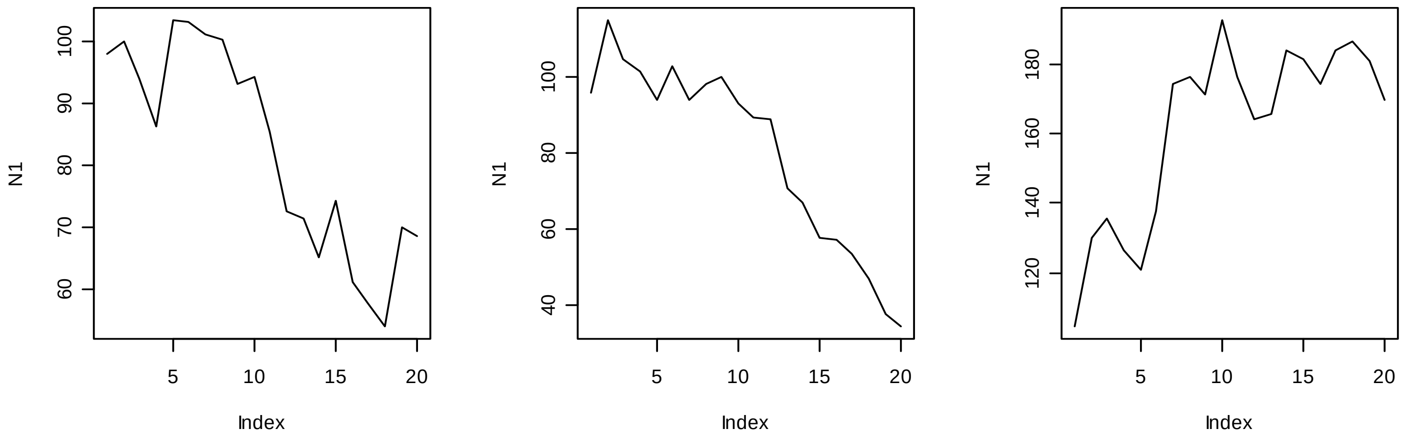
The process error leads to characteristic random walks: AR lag-1 with drift

All trajectories came from the same model:
$N_t = N_{t-1} \exp(-0.02 + e_t)$, e_t was Normal(mean=0.0, var=0.01)

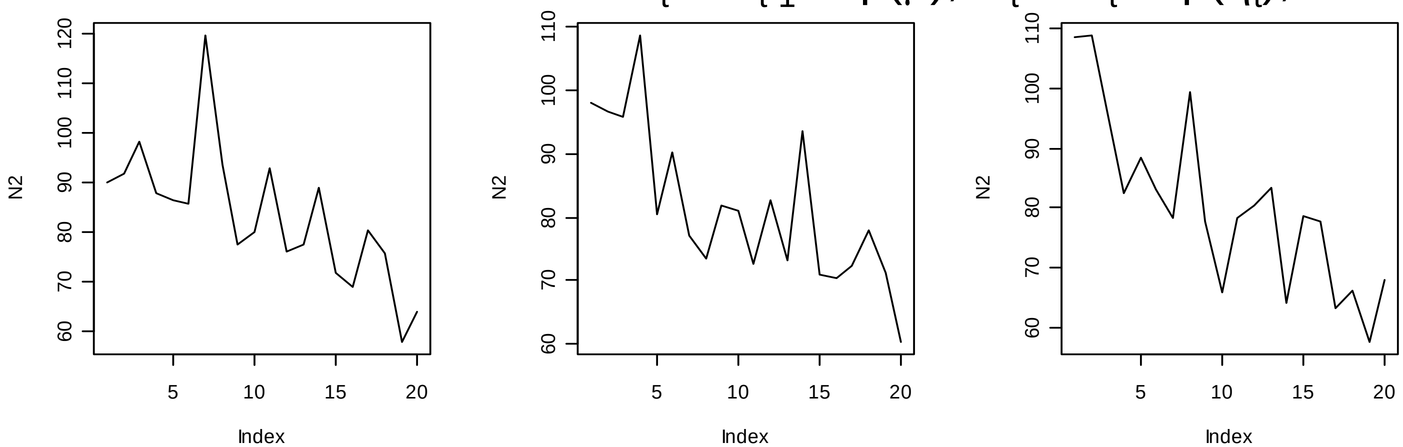


How can we separate process and observation variance? They affect a time series differently.

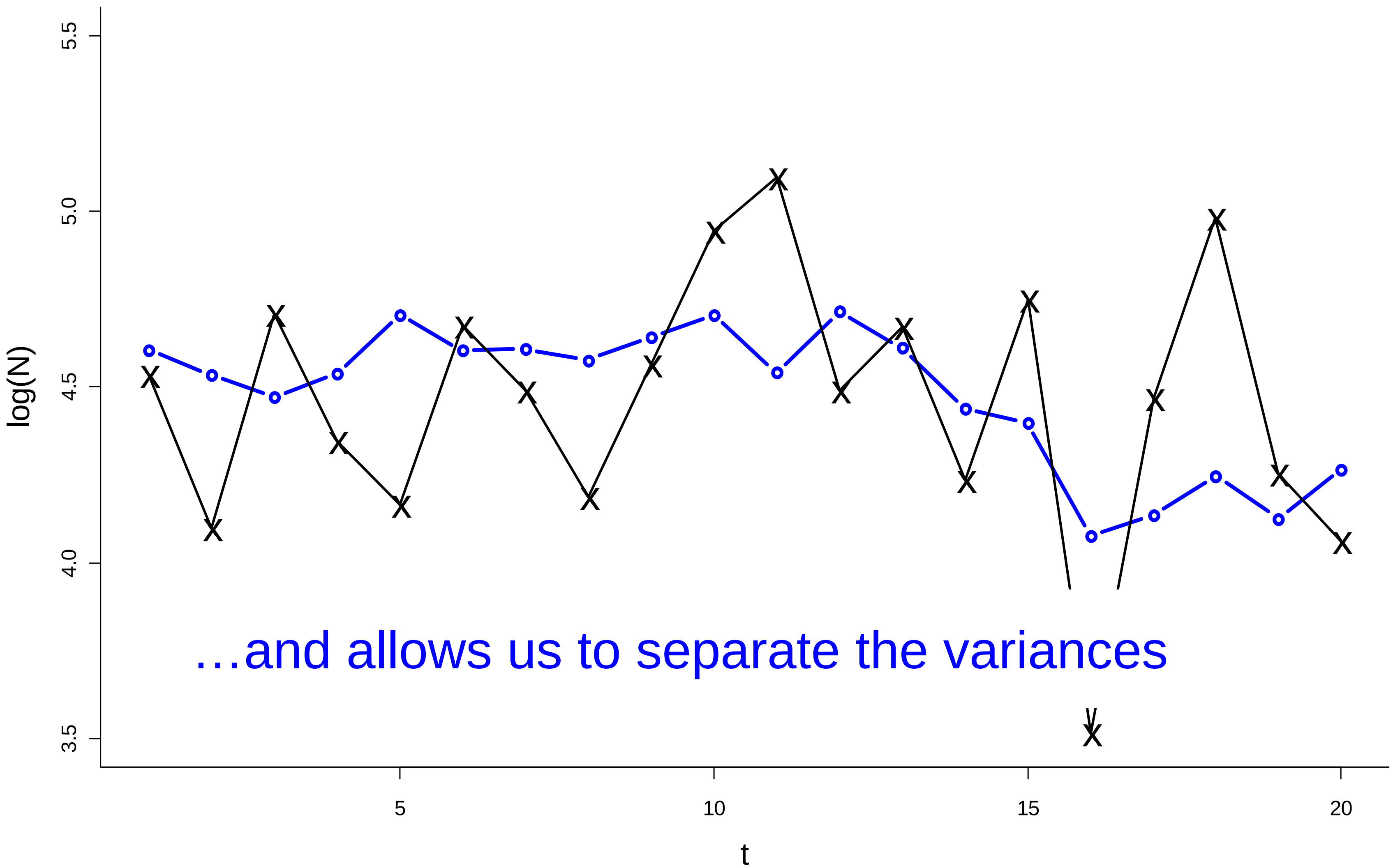
Process error: $N_t = N_{t-1}\exp(\mu + e_t)$,



Observation error: $N_t = N_{t-1}\exp(\mu)$; $O_t = N_t \exp(\eta_t)$;



A state-space model combines a model for the hidden AR-1 process with a model for the observation process



To fit this model, we have to write it mathematically

Population growth

$$x_t = \log(N_t)$$

N_t is population size

$$x_t = x_{t-1} + u + w_t$$

Exponential growth model

$$w_t \sim Normal(0, q)$$

Normally distributed process errors

Observations

$$y_t = \log(O_t)$$

$$y_t = x_t + v_t$$

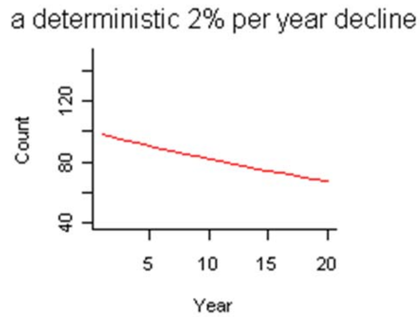
$$v_t \sim Normal(0, r)$$

Log-normally distributed observation errors

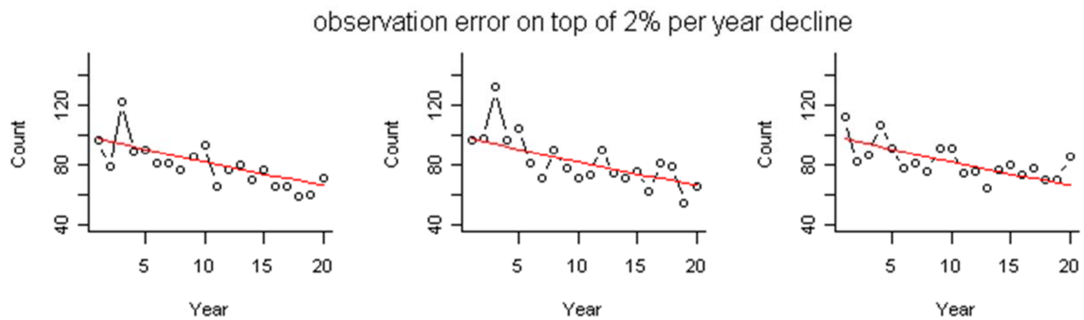
Let's simulate and try fitting some models

- Open up R and follow after me
- Lecture_2_univariate_example_1.R
- Lecture_2_univariate_example_2.R
- Lecture_2_univariate_example_3.R

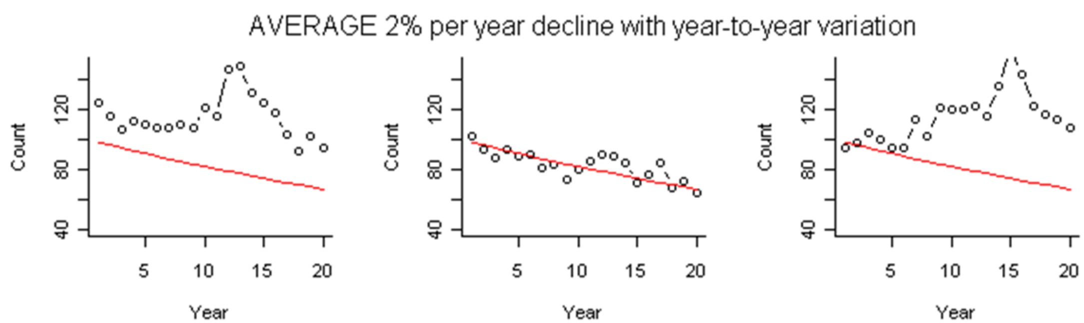
Deterministic, vs. obs. error, vs. proc. error
An example using population decline



Every year, decline 2%

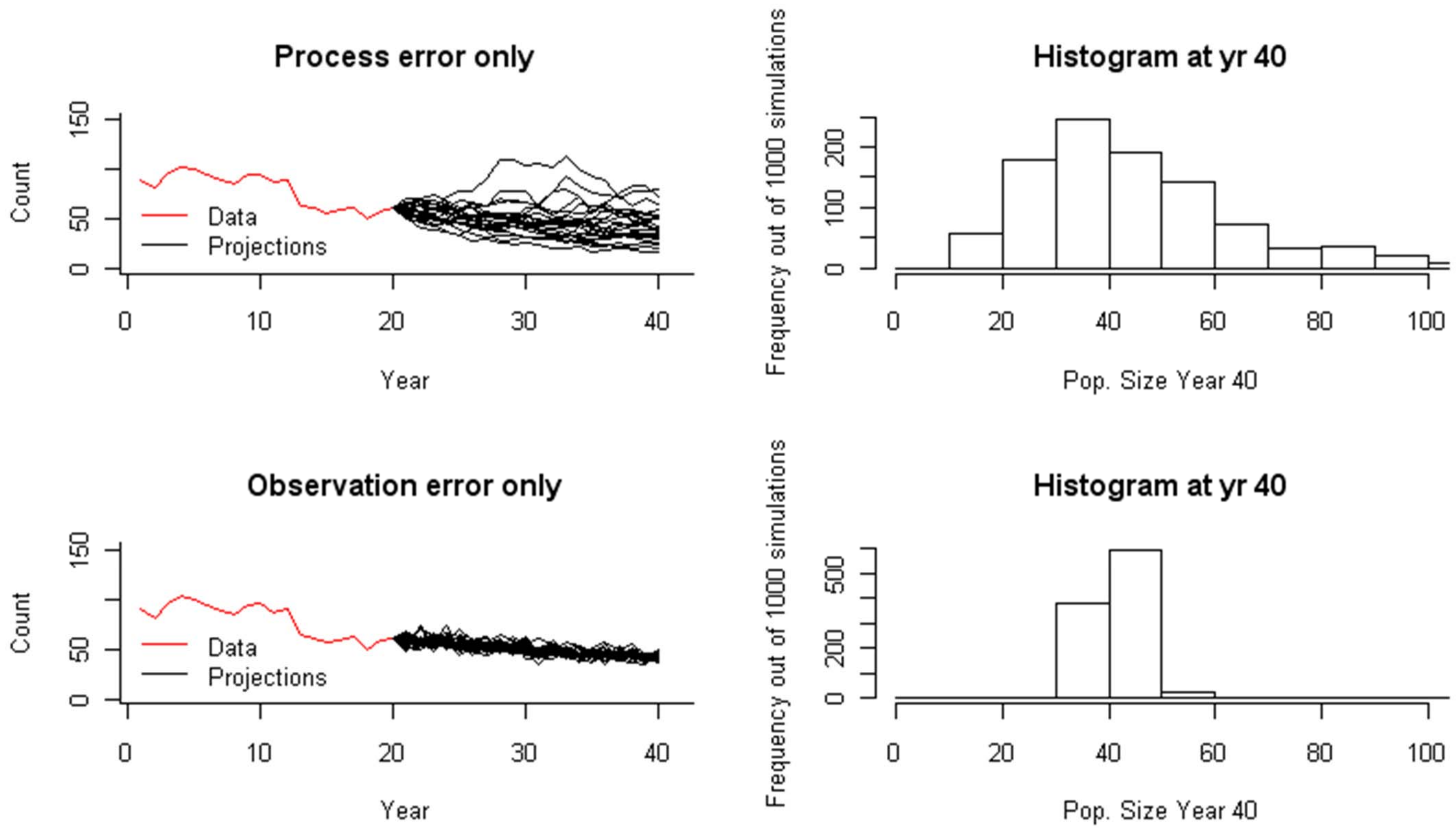


Every year, decline 2% but there is observation error



Average yearly decline is 2%, but actual declines vary from year to year

How you model your population data has a large impact on projection of the process



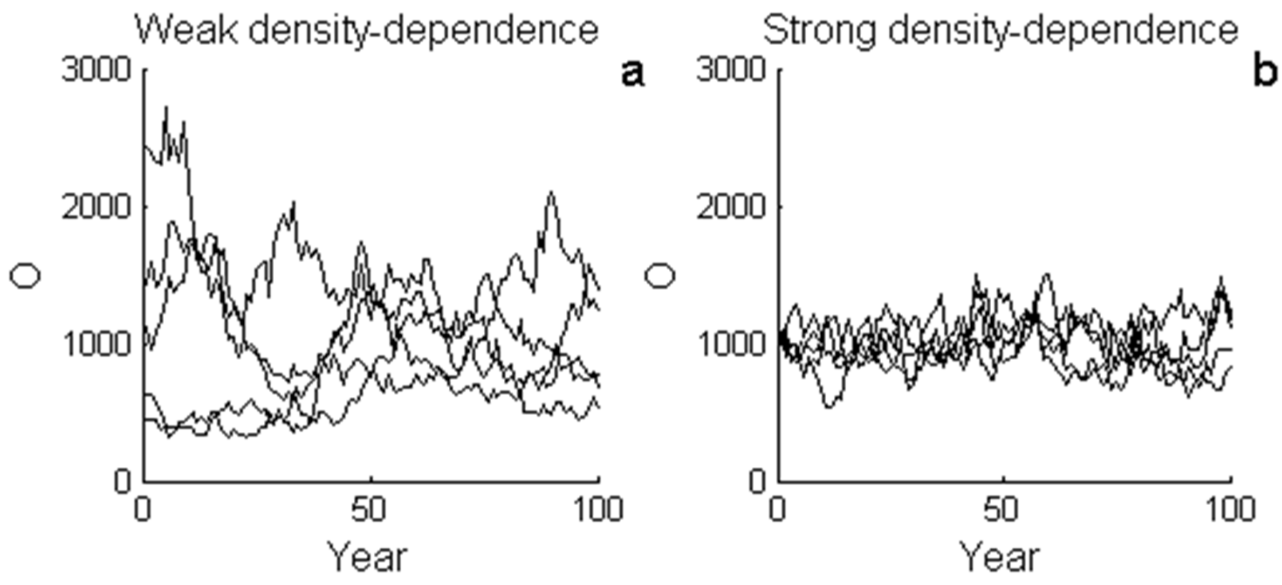
State-space model with density-dependence termed ‘mean-reverting‘. ---Day 3---

$$N_t = \exp(u + e_t) N_{t-1}^b$$

$$x_t = b x_{t-1} + u + e_t$$

Log-space

$$e_t \sim Normal(0, q)$$



b<1: Gompertz density-dependent process

Computer labs

from the MARSS User Guide

Chapter 6: Count-based population viability analysis (PVA) using corrupted data

```
library(MARSS)
RShowDoc("Chapter_PVA.R", package="MARSS")
```

Chapter 10: Analyzing noisy animal tracking data

```
RShowDoc("Chapter_AnimalTracking.R", package="MARSS")
```

1991

ESTIMATION OF GROWTH AND EXTINCTION PARAMETERS FOR ENDANGERED SPECIES[1]

BRIAN DENNIS
Department of Forest Resources and Department of Mathematics and Statistics,

Estimating risks in declining populations with poor data

2001

Elizabeth E. Holmes*

ESTIMATION OF POPULATION GROWTH AND EXTINCTION PARAMETERS FROM NOISY DATA

2003

STEVEN T. LINDLEY[1]

BEYOND THEORY TO APPLICATION AND EVALUATION: DIFFUSION APPROXIMATIONS FOR POPULATION VIABILITY ANALYSIS

2004

Nation

Ecology Letters, (2008) 11: E1–E5 doi: 10.1111/j.1461-0248.2008.01211.x

TECHNICAL COMMENT

Commentary on Holmes *et al.* (2007): resolving the debate on when extinction risk is predictable

2008

Stephen P. Ellner[1]* and Elizabeth E. Holmes[2]
[1]*Department of Ecology*

Abstract

We reconcile the findings of Holmes *et al.* (*Ecology Letters*, **10**, 2007, 1182) that 95% confidence intervals for quasi-extinction risk were narrow for many vertebrates of conservation concern, with previous theory predicting wide confidence intervals. We