

# Introduction to multivariate state-space models

using multi-site population data as the example

Eli Holmes

*FISH 507 – Applied Time Series Analysis*

19 January 2017

# Topics

---

## Lecture

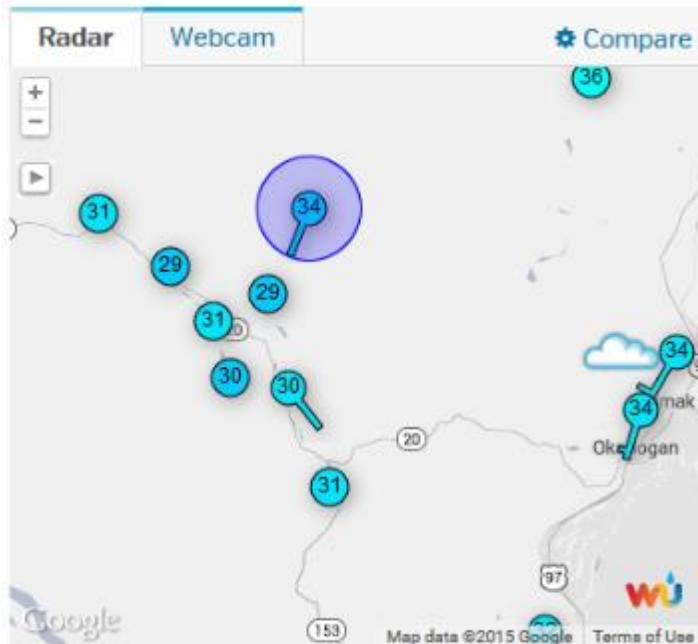
- Short example of multivariate observations
- Examples of multivariate structure in population data
- How to express these structures mathematically
- Adding a multivariate observation process

## Computer Labs

- Analysis of population structure using multi-site data
- Combining diverse data sources to estimate an underlying model

# Other examples

Combine multiple station data into a single metric

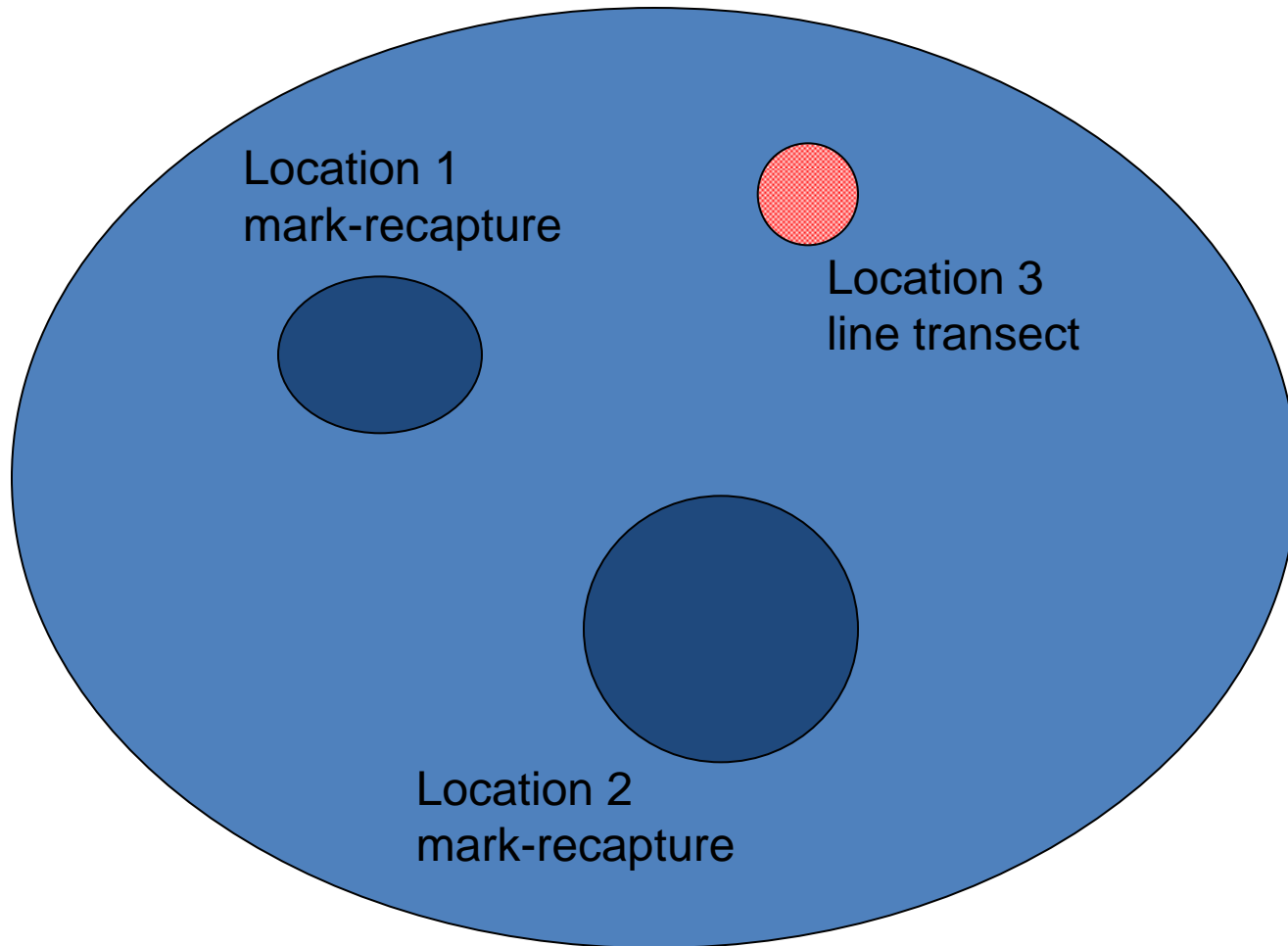


Multiple individuals measured over time



# Imagine we have 3 sampling locations for a population

---



# Mathematically we can write

---

$$x_t = x_{t-1} + u + w_t, w_t \sim N(0, q)$$

$$y_{1,t} = x_t + v_{1,t}, v_{1,t} \sim N(a_1, r_1)$$

$$y_{2,t} = x_t + v_{2,t}, v_{2,t} \sim N(a_2, r_2)$$

$$y_{3,t} = x_t + v_{3,t}, v_{3,t} \sim N(a_3, r_3)$$

observations

population  
size

noise

# The observation part can be rewritten

---

We need to fix one of the  $a$ 's.  
Traditionally we fix to the first to 0.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_t = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x_t + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_t$$

observations

Z matrix

population  
size

bias

noise

# The model with one $a$ fixed to zero

---

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_t = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_t$$

observations

Z matrix

population  
size

bias

noise

# The observation errors are multivariate. For now, let's assume Normality

---

The variance-covariance matrix tells you how the observation errors are related. Are they independent? Or do they covary? Do they have the same variance or different variances?

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_t \sim MVN \left( \mathbf{0}, \begin{bmatrix} \eta_1^2 & \eta_{1,2} & \eta_{1,3} \\ \eta_{1,2} & \eta_2^2 & \eta_{3,2} \\ \eta_{1,3} & \eta_{3,2} & \eta_3^2 \end{bmatrix} \right)$$



# Example observation error var-cov matrices

$$\begin{bmatrix} \eta_1^2 & \eta_{1,2} & \eta_{1,3} \\ \eta_{1,2} & \eta_2^2 & \eta_{3,2} \\ \eta_{1,3} & \eta_{3,2} & \eta_3^2 \end{bmatrix}$$

unconstrained

$$\begin{bmatrix} \eta^2 & \alpha & \alpha \\ \alpha & \eta^2 & \alpha \\ \alpha & \alpha & \eta^2 \end{bmatrix}$$

“equal varcov”

$$\begin{bmatrix} \eta_1^2 & 0 & 0 \\ 0 & \eta_2^2 & 0 \\ 0 & 0 & \eta_3^2 \end{bmatrix}$$

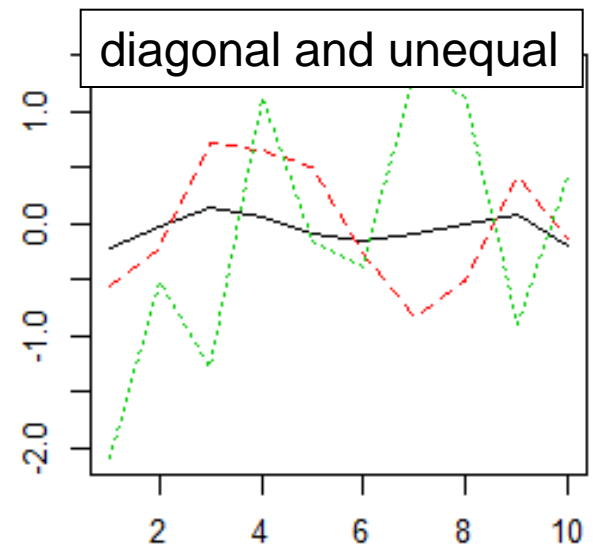
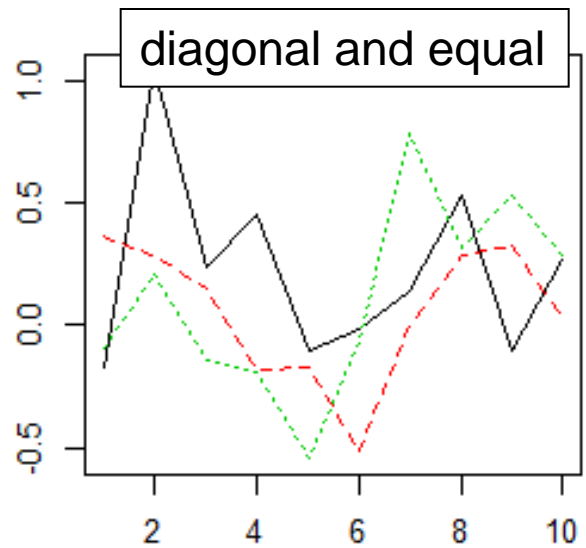
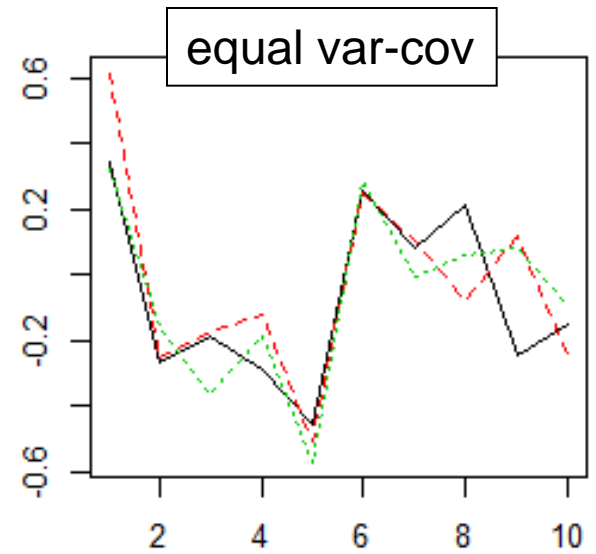
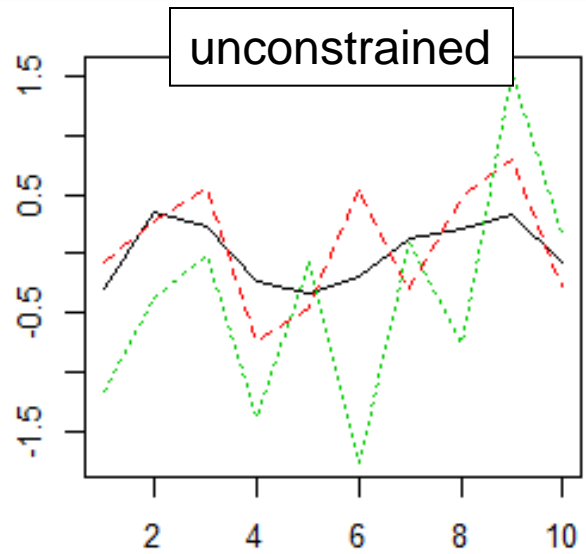
unique variances and uncorrelated errors

diagonal

$$\begin{bmatrix} \eta^2 & 0 & 0 \\ 0 & \eta^2 & 0 \\ 0 & 0 & \eta^2 \end{bmatrix}$$

identical variances and uncorrelated errors

# Example of errors coming from these variance-covariance matrices



error  
= how  
much  
the pop.  
growth  
rate is  
above or  
below  
average

`mvrnorm(10,rep(0,3),diag(.1,3))`

# Fitting MARSS models using the MARSS R Package

---

- Fits MARSS models
- *Model specification is 1-to-1 with the equation for the model*
- General, fits any MARSS model with Gaussian errors.
- **BUT**
- Maximum likelihood
- Slow for large data sets. Students working with large data sets have gotten huge speed improvements by coding their models in TMB (or ADMB).

# MARSS R Package Model Syntax

---

$$X(t) = \mathbf{B} X(t-1) + \mathbf{U} + w(t), w(t) \sim N(0, \mathbf{Q})$$

$$Y(t) = \mathbf{Z} X(t) + \mathbf{A} + v(t), v(t) \sim N(0, \mathbf{R})$$

- `fit2=MARSS(y,model=mod.list)`
- `y` is data; `model` tells MARSS what the parameters are
- The parameters are MATRICES
- You write matrices just like they appear in your model on paper. Matrices must be MATRICES (not scalars, not vectors)
- You pass `model` to MARSS as a list

$$X(t) = \mathbf{B} X(t-1) + \mathbf{U} + w(t), w(t) \sim N(0, \mathbf{Q})$$

$$Y(t) = \mathbf{Z} X(t) + \mathbf{A} + v(t), v(t) \sim N(0, \mathbf{R})$$

```
mod.list=list(  
  U=matrix("u"),  
  x0=matrix(0),  
  B=matrix(1),  
  Q=matrix(0.1),  
  Z=matrix(1),  
  A=matrix(0),  
  R=matrix("r"),  
  tinitx=0)
```

Let's say we want to fit this model:

$$x_t = x_{t-1} + u + w_t, w_t \sim N(0, 0.1)$$

$$y_t = x_t + v_t, v_t \sim N(0, r)$$

$$x_0 = 0$$

Write in matrix form:

$$[x]_t = [1][x]_{t-1} + [u] + [w]_t, [w]_t \sim MVN(0, [0.1])$$

$$[y]_t = [1][x]_t + [v]_t, [v]_t \sim MVN(0, [r])$$

$$x_0 = [0]$$

```
mod.list=list(  
  Q=matrix(0.1)  
)
```

$$X(t) = \mathbf{B} X(t-1) + \mathbf{U} + w(t), w(t) \sim N(0, \mathbf{Q})$$

$$Y(t) = \mathbf{Z} X(t) + \mathbf{A} + v(t), v(t) \sim N(0, \mathbf{R})$$

```

mod.list=list(
  U=matrix("u"),
  x0=matrix(0),
  B=matrix(1),
  Q=matrix(0.1),
  Z=matrix(1,2,1),
  A= matrix(list(0,"a2"),2,1),
  R= matrix(list("r",0,0,"r"),
  tinitx=0)

```

Let's say we want to fit a model where two sites are sampling temperature  $x$  in a lake:

Our temperature model:

$$[x]_t = [1][x]_{t-1} + [u] + [w]_t, [w]_t \sim MVN(0, [0.1])$$

$$[x]_0 = [0]$$

Our two temperature sensors:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} [x]_t + \begin{bmatrix} 0 \\ a_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_t$$

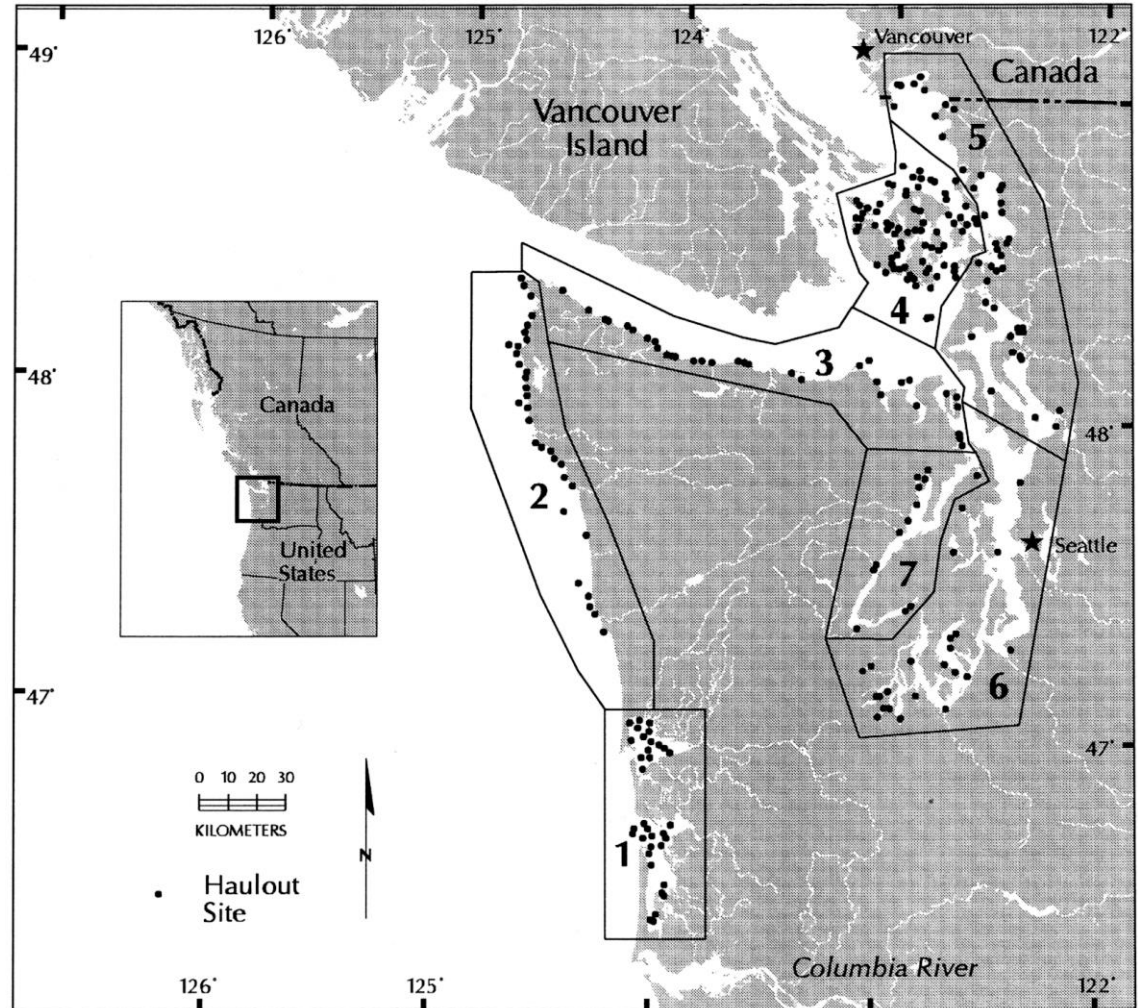
$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_t \sim MVN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \right)$$

# Some short examples

---

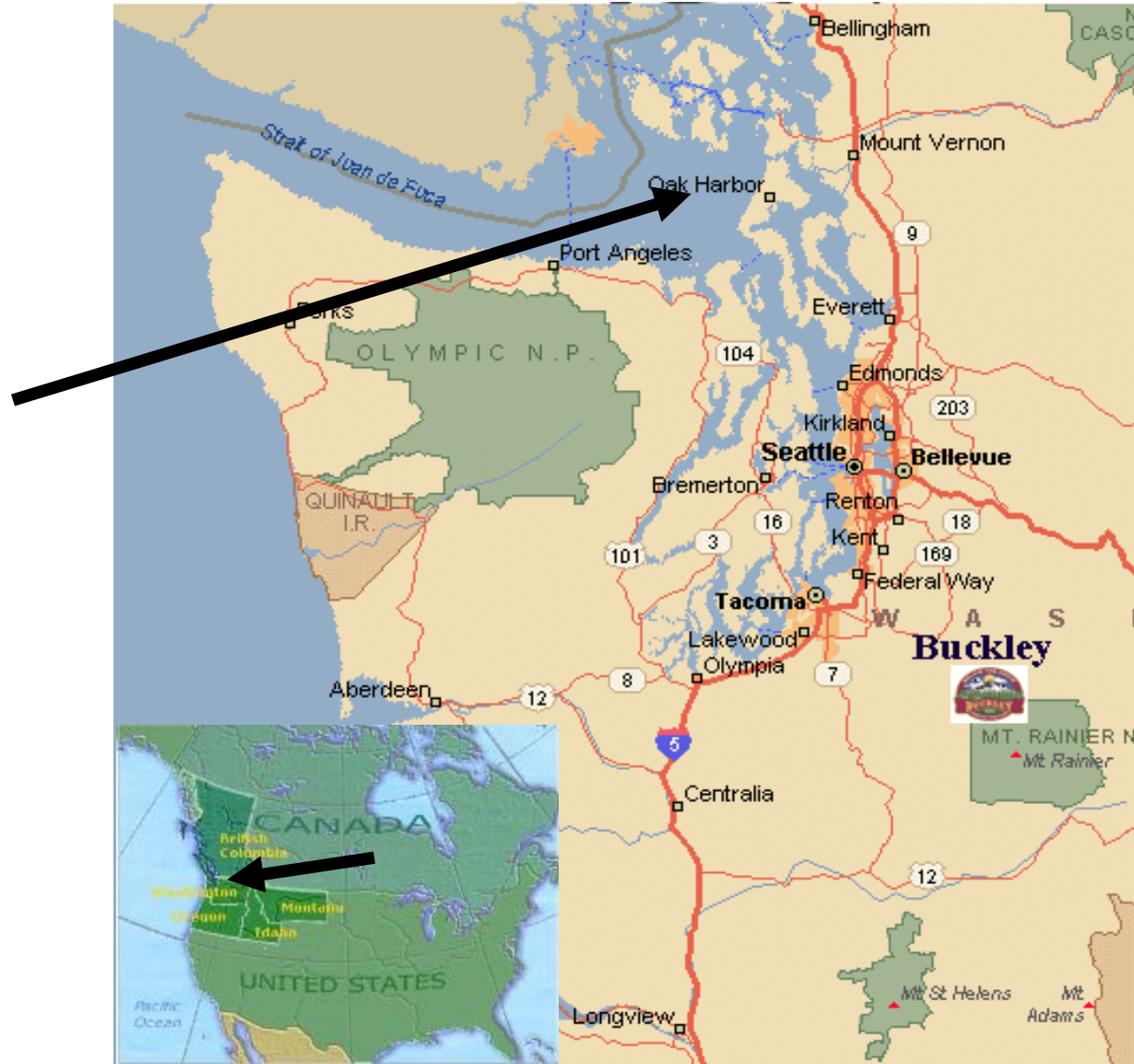
- MARSS\_example\_1.R
- MARSS\_example\_2.R
- MARSS\_example\_3.R

# Multi-site data (Pacific harbor seals)

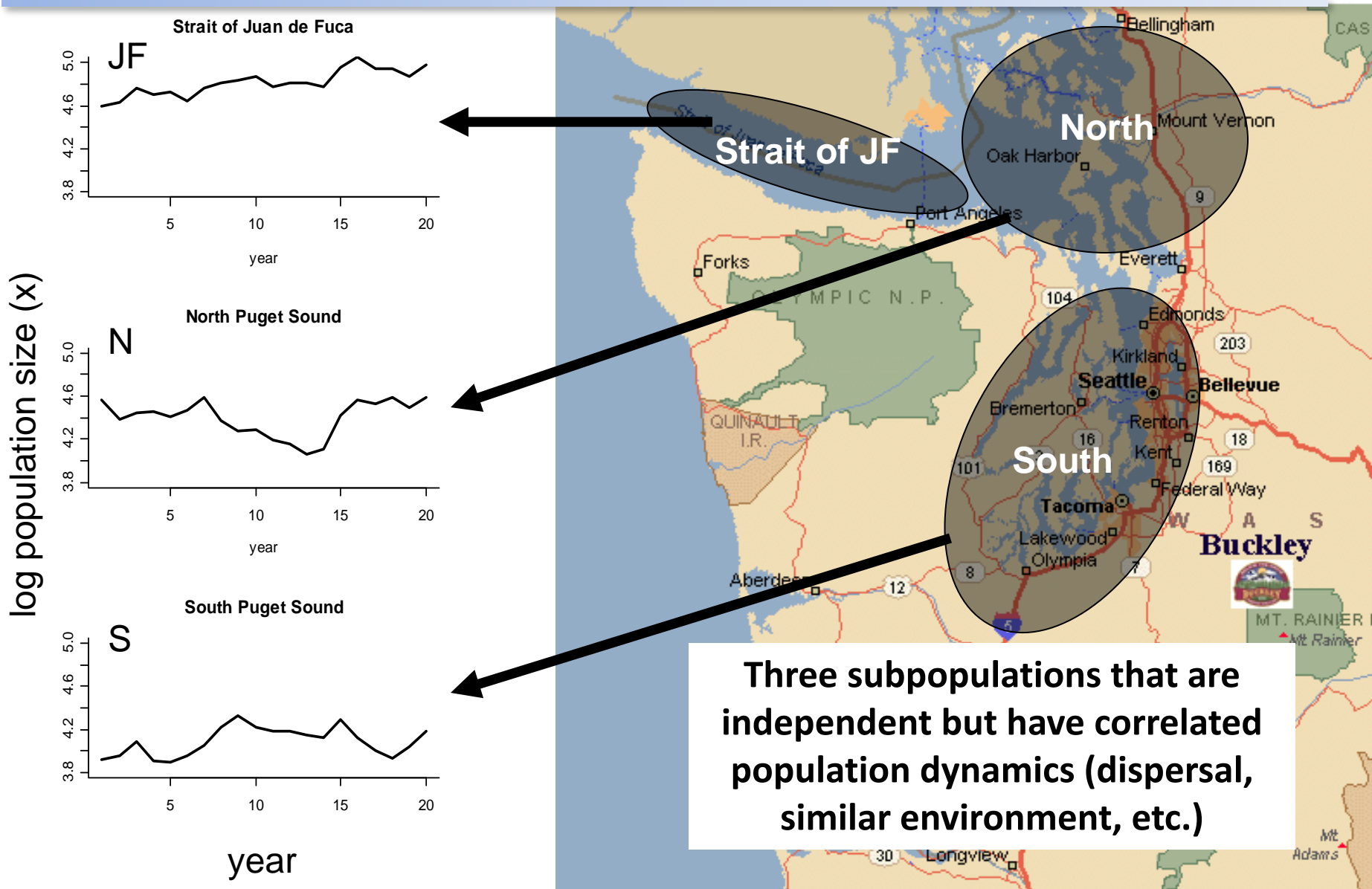




# An example: modeling the population dynamics of harbor seals in Puget Sound, WA



# Let's hypothesize (and model) that the population has 3 subpopulations



# A multivariate model for the population (not the observations but the actual population)

## Multivariate stochastic exponential growth

$$\begin{bmatrix} x_{JF,t} \\ x_{N,t} \\ x_{S,t} \end{bmatrix} = \begin{bmatrix} x_{JF,t-1} \\ x_{N,t-1} \\ x_{S,t-1} \end{bmatrix} + \begin{bmatrix} u_{JF} \\ u_N \\ u_S \end{bmatrix} + \begin{bmatrix} w_{JF,t} \\ w_{N,t} \\ w_{S,t} \end{bmatrix}$$

3 different  $x$ 's, one for each subpopulation

3 mean population growth rate terms

3 different process errors

$$e \sim \text{MVN}(0, \mathbf{Q})$$

# The population model in matrix form

---

Exponential population growth with drift (tendency to increase or decline)

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

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

Each parameter has “structure”. Different structures imply different population structure.

# The mean population growth rates ( $u$ ) can have spatial structure

---

$$\begin{bmatrix} u_{JF} \\ u_N \\ u_S \end{bmatrix}$$

unconstrained (all different)

$$\begin{bmatrix} u \\ u \\ u \end{bmatrix}$$

all the same

$$\begin{bmatrix} u_{JF} \\ u_{N\&S} \\ u_{N\&S} \end{bmatrix}$$

Strait of Juan de Fuca different  
North and South same

# The process error var-cov matrix can have structure: $\mathbf{w}_t \sim \text{MVN}(0, \mathbf{Q})$

$$\begin{bmatrix} \sigma_{JF}^2 & \sigma_{JF,N} & \sigma_{JF,S} \\ \sigma_{JF,N} & \sigma_N^2 & \sigma_{N,S} \\ \sigma_{JF,S} & \sigma_{N,S} & \sigma_S^2 \end{bmatrix}$$

unconstrained

variances all different and year-to-year population changes covary

$$\begin{bmatrix} \sigma_{JF}^2 & 0 & 0 \\ 0 & \sigma_N^2 & 0 \\ 0 & 0 & \sigma_S^2 \end{bmatrix}$$

diagonal

unique variances and year-to-year population changes are uncorrelated

$$\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

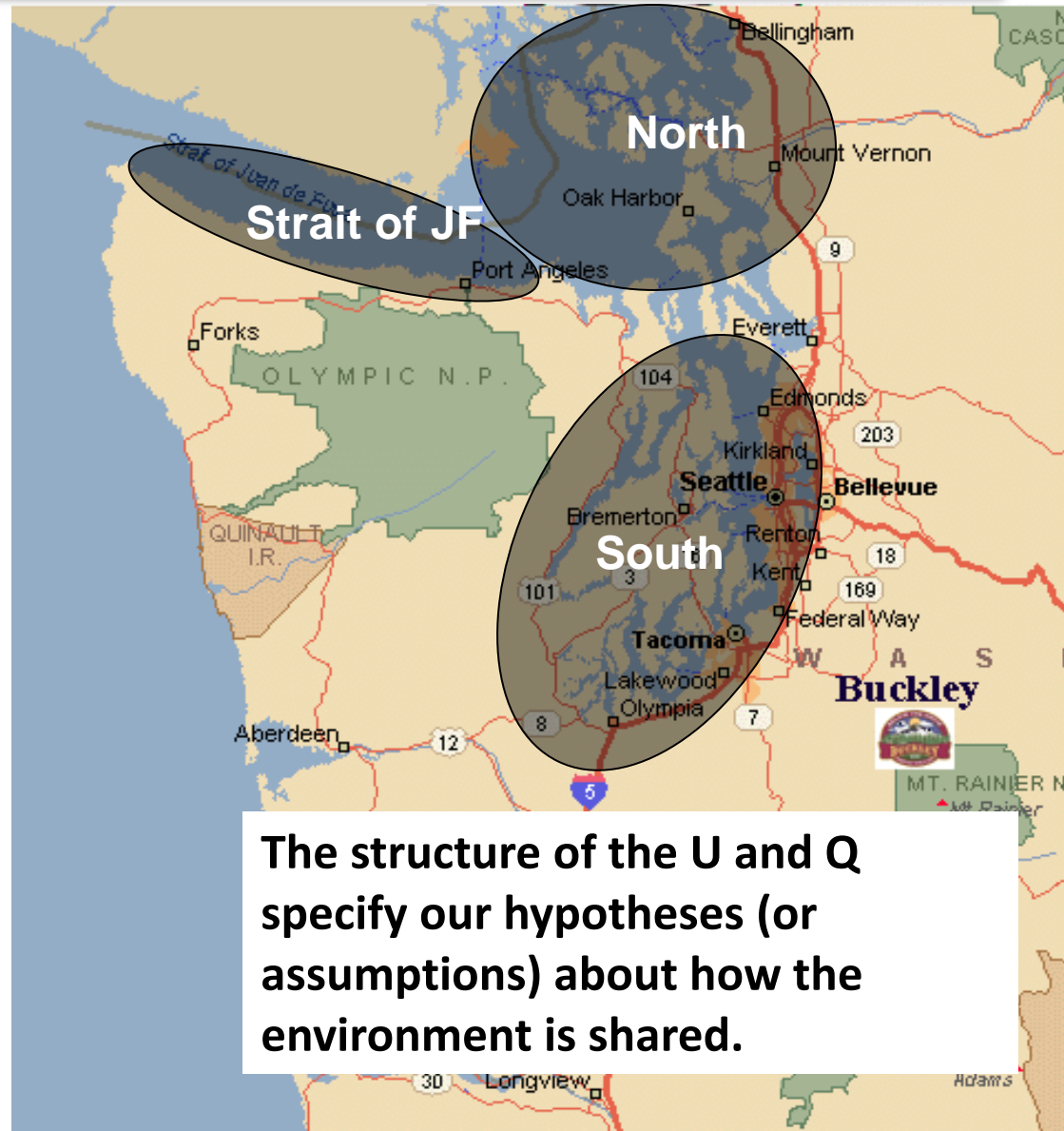
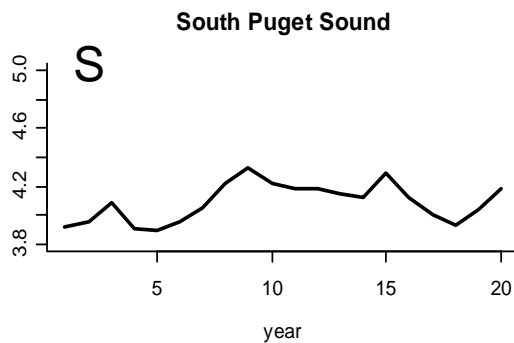
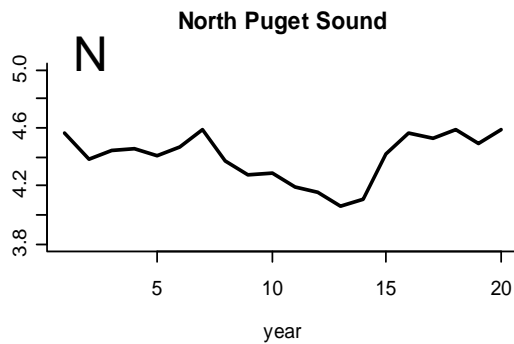
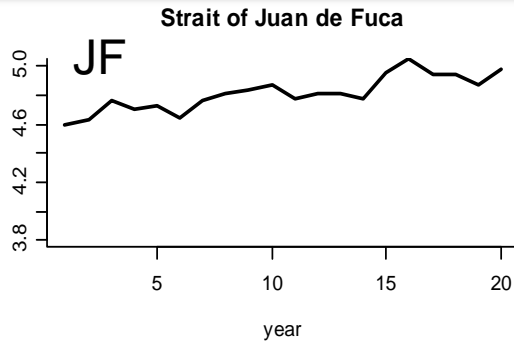
diagonal

same variances and year-to-year population changes are uncorrelated

$$\begin{bmatrix} \sigma^2 & \alpha & \alpha \\ \alpha & \sigma^2 & \alpha \\ \alpha & \alpha & \sigma^2 \end{bmatrix}$$

JF has unique variance;  
N & S share the same variance  
yr-to-yr changes have equal covariance

$$X_t = X_{t-1} + u + w_t$$



The structure of the  $U$  and  $Q$  specify our hypotheses (or assumptions) about how the environment is shared.

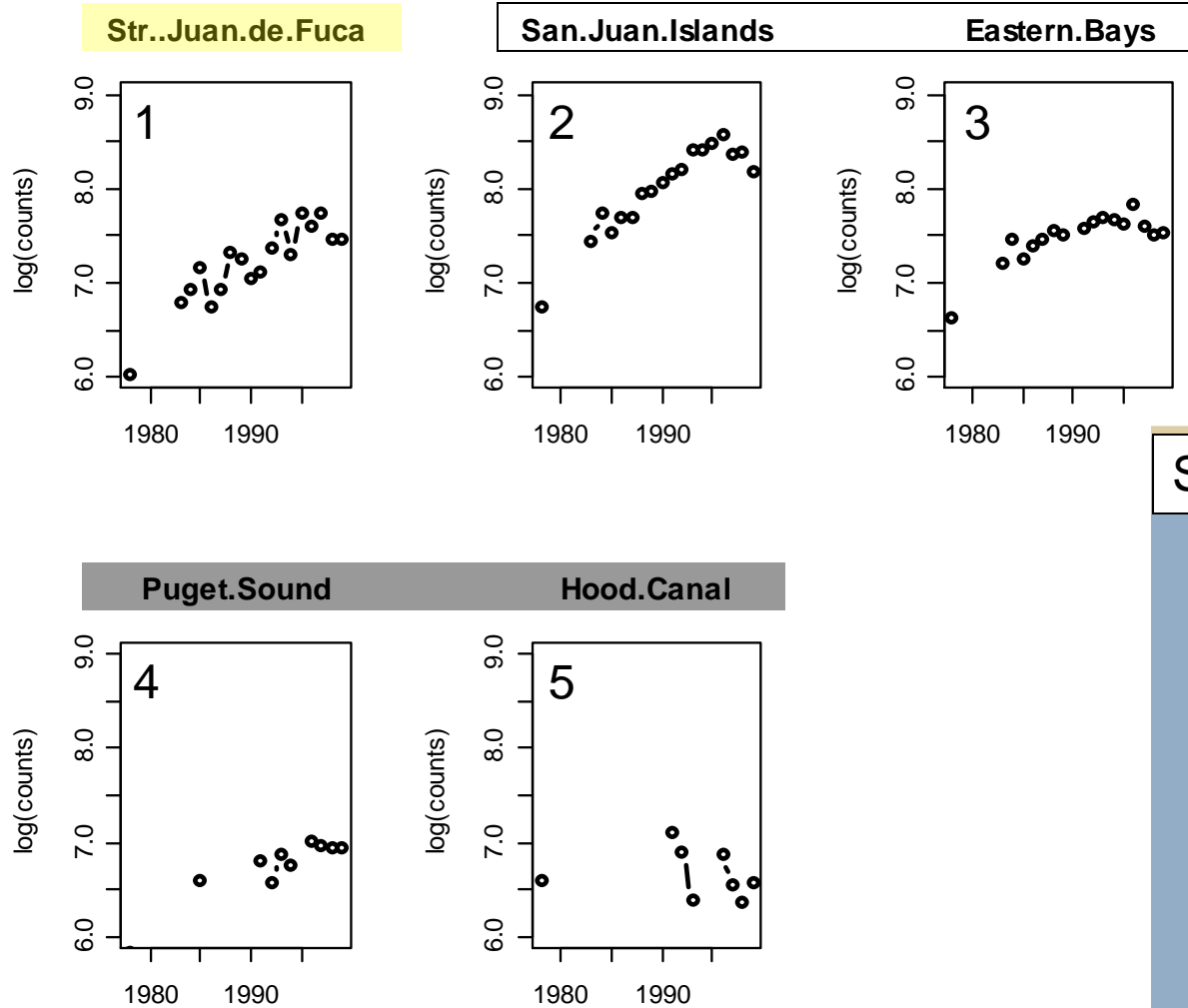
We observe our subpopulations and those observations have error



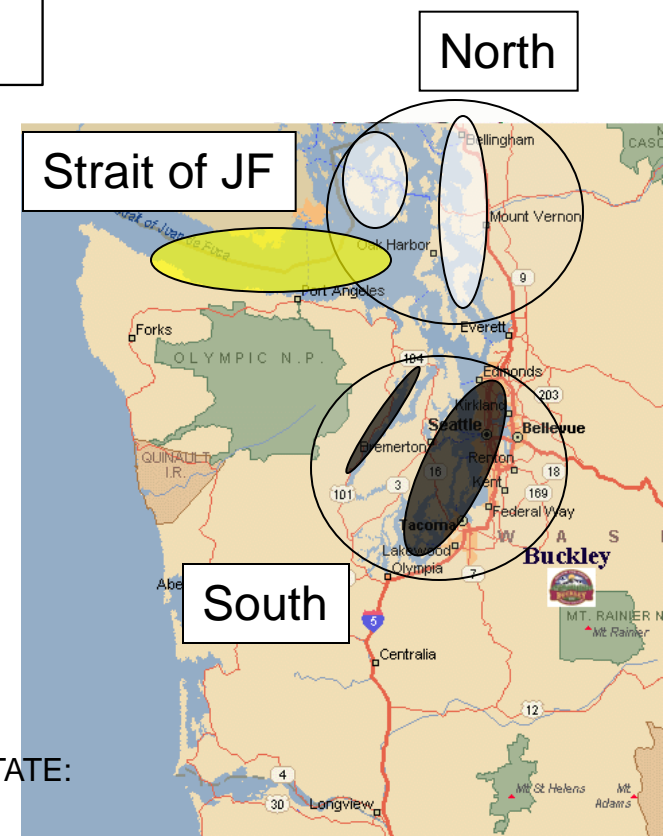
sea lion



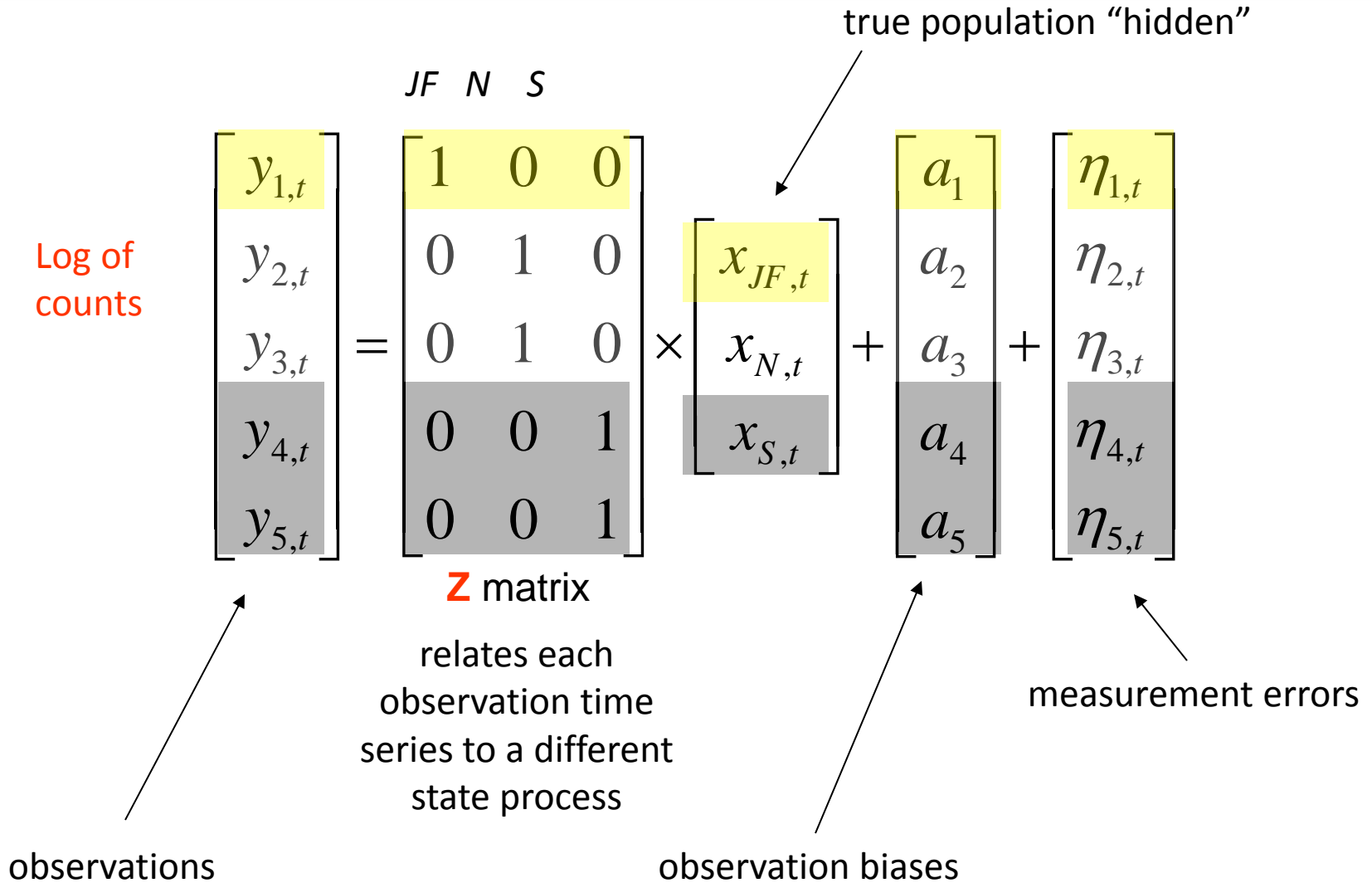
# The obs. err. model specifies how the observed time series are related to the true subpopulation sizes



5 sampling locations



# The observation model



# The observation errors have a var-cov matrix

---

$$\begin{bmatrix} \eta_1^2 & \eta_{1,2} & \eta_{1,3} & \eta_{1,4} & \eta_{1,5} \\ \eta_{1,2} & \eta_2^2 & \eta_{3,2} & \eta_{2,4} & \eta_{2,5} \\ \eta_{1,3} & \eta_{3,2} & \eta_3^2 & \eta_{3,4} & \eta_{3,5} \\ \eta_{1,4} & \eta_{2,4} & \eta_{3,4} & \eta_4^2 & \eta_{4,5} \\ \eta_{1,5} & \eta_{2,5} & \eta_{3,5} & \eta_{4,5} & \eta_5^2 \end{bmatrix}$$

unconstrained

$$\begin{bmatrix} \eta_1^2 & 0 & 0 & 0 & 0 \\ 0 & \eta_2^2 & 0 & 0 & 0 \\ 0 & 0 & \eta_3^2 & 0 & 0 \\ 0 & 0 & 0 & \eta_4^2 & 0 \\ 0 & 0 & 0 & 0 & \eta_5^2 \end{bmatrix}$$

unique  
variances and  
uncorrelated  
errors

diagonal

$$\begin{bmatrix} \eta^2 & 0 & 0 & 0 & 0 \\ 0 & \eta^2 & 0 & 0 & 0 \\ 0 & 0 & \eta^2 & 0 & 0 \\ 0 & 0 & 0 & \eta^2 & 0 \\ 0 & 0 & 0 & 0 & \eta^2 \end{bmatrix}$$

identical  
variances and  
uncorrelated  
errors

diagonal



# Instead of N, S, Str. J subpopulations, we could have other combinations and numbers of subpopulations

$$\begin{array}{l}
 \text{Str. JF} \\
 \text{San Isl.} \\
 \text{E. Bays} \\
 \text{P.S.} \\
 \text{Hood C.}
 \end{array}
 \begin{bmatrix}
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 0 & 1 & 0 \\
 0 & 0 & 1 \\
 0 & 0 & 1
 \end{bmatrix}$$

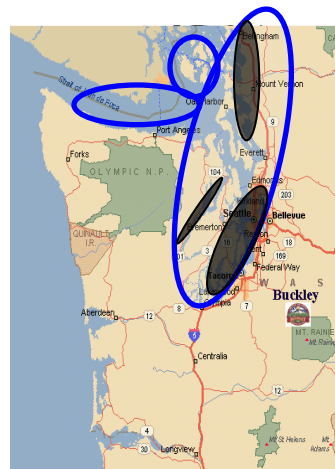
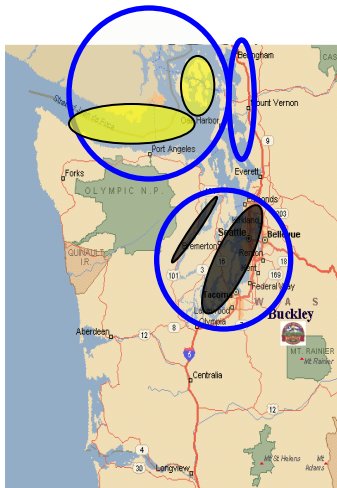
$$\begin{bmatrix}
 1 & 0 & 0 \\
 0 & 1 & 0 \\
 0 & 0 & 1 \\
 0 & 0 & 1 \\
 0 & 0 & 1
 \end{bmatrix}$$

$$\begin{bmatrix}
 1 \\
 1 \\
 1 \\
 1 \\
 1
 \end{bmatrix}$$

Str of Juan de Fuca & San Juan Is sites = 1<sup>st</sup> subpop  
 Eastern bays = 2<sup>nd</sup>  
 Hood C. & S. Puget S. = 3<sup>rd</sup>

Strait of Juan de Fuca = 1<sup>st</sup> sub pop  
 San Juan Is sites = 2<sup>nd</sup>  
 Eastern bays, Hood Canal & S. Puget Sound = 3<sup>rd</sup>

One Puget Sound population and all sites are sampling it  
**One population**



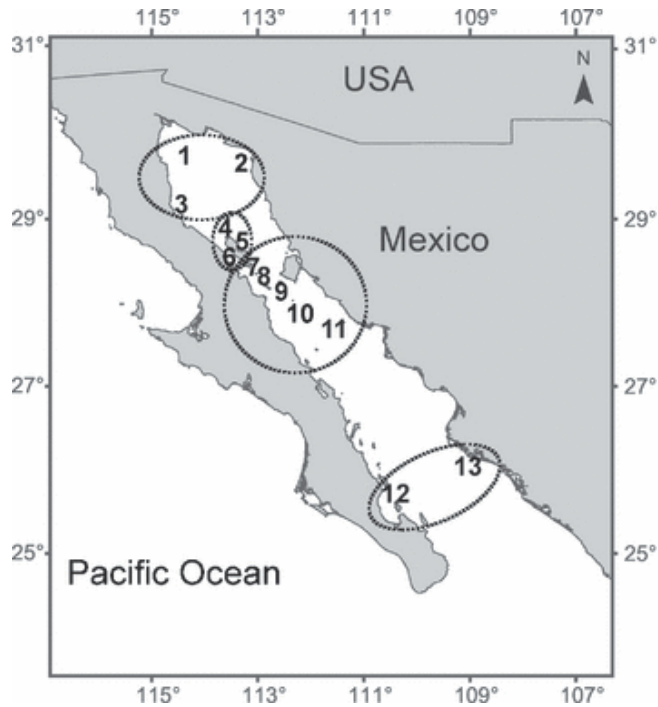
The same model can capture many different underlying population structures and observation structures

---

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{u} + \mathbf{w}_t \quad \text{where } \mathbf{w}_t \sim MVN(0, \mathbf{Q})$$

$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{v}_t \quad \text{where } \mathbf{v}_t \sim MVN(0, \mathbf{R})$$

***Inferring spatial structure from time-series data: using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico***



Tony Orr (NOAA, AFSC)

Hypotheses about the population structure:  
Diet, Disease, DNA, Distance  
(2 null models: no structure and fully structured)

**Table 1.** Model performance, given by Akaike's Information Criterion (AIC) *b*-value, across the six hypotheses for the subpopulation configuration

| Parameters    |                   |             | Hypotheses ( <i>m</i> = no. subpopulations) |                         |                            |                             |                        |                                 |
|---------------|-------------------|-------------|---|-------------------------|----------------------------|-----------------------------|------------------------|---------------------------------|
| <i>u</i>      | <i>Q</i>          | <i>R</i>    | Panmictic<br>( <i>m</i> = 1)                | Diet<br>( <i>m</i> = 4) | Disease<br>( <i>m</i> = 4) | Distance<br>( <i>m</i> = 4) | DNA<br>( <i>m</i> = 2) | Independent<br>( <i>m</i> = 11) |
| Same          | Same              | Same        | 68.2  | 48.4                    | 49.8                       | 26.8                        | 38.9                   | 22.2                            |
| Unique        | Same              | Same        |   | 63.9                    | 72.8                       | 46.9                        | 46.6                   | 25.5                            |
| Same          | Unique            | Same        |   | 55.8                    | 57.6                       | 26.6                        | 34.2                   | 64.4                            |
| Same          | Same              | Unique      | 97.3  | 74.4                    | 73.2                       | 68.4                        | 67.3                   | 32.5                            |
| Unique        | Same              | Unique      |   | 87.1                    | 91.3                       | 71.8                        | 69.3                   | 65.6                            |
| Unique        | Unique            | Same        |   | 61.4                    | 84.1                       | 39.8                        | 38.3                   | 50.0                            |
| Same          | Unique            | Unique      |   | 102.8                   | 103.8                      | 202.1                       | 82.7                   | 114.7                           |
| Unique        | Unique            | Unique      |   | 111.8                   | 133.8                      | 167.8                       | 77.5                   | 169.2                           |
| Same          | Correlated        | Same        |   | 40.3                    | 63.1                       | 37.0                        | 38.3                   | 4804.7                          |
| <b>Unique</b> | <b>Correlated</b> | <b>Same</b> |   | 44.9                    | 87.2                       | <b>13.7</b>                 | 39.6                   | 989.4                           |
| Same          | Correlated        | Unique      |   | 110.3                   | 163.8                      | 321.4                       | 102.2                  | NA                              |
| Unique        | Correlated        | Unique      |   | 116.3                   | 176.5                      | 467.9                       | 94.5                   | NA                              |

Process errors (*Q*) may be independent (a diagonal matrix) with variances that are the same magnitude across subpopulations (same), independent with unequal variances across subpopulations (unique) or may be temporally correlated, meaning an unconstrained *Q* matrix (correlated). The growth rate (*u*) and observation error matrix (*R*) parameters may also be equal (same) or unique across subpopulations. The model best supported by the data is shown in bold; complex models that did not fully converge are not applicable.



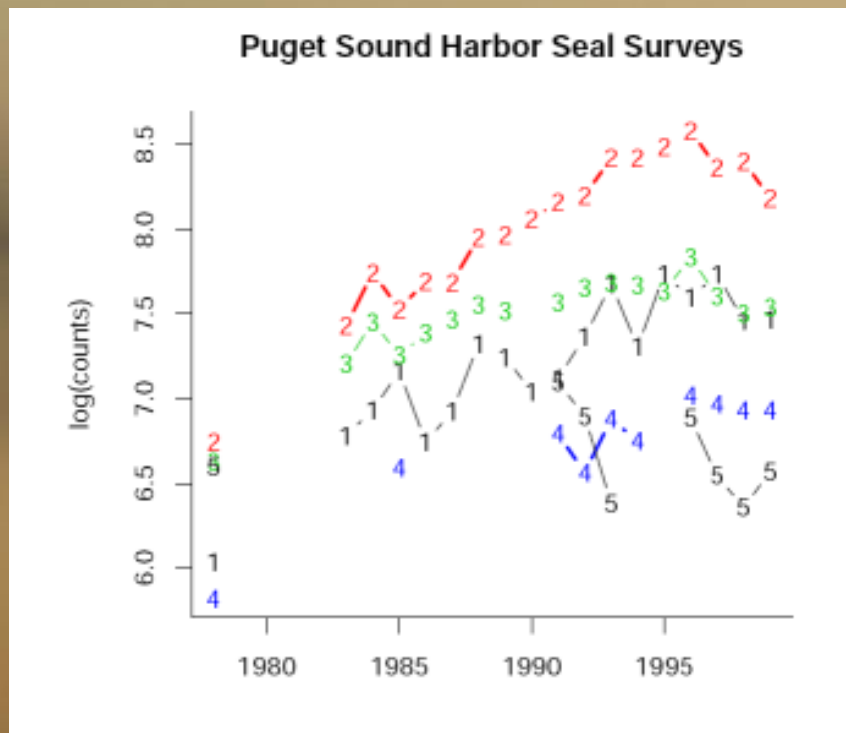
# The MARSS manual has two chapters using MARSS models to analyze spatial count data

harbor seal count data from the west coast of the USA



# Chapter 7 in HWS 2014

## Chapter 7: Combining multi-site data to estimate regional population trends



# Chapter 8 in HWS 2014

Identifying spatial structure and covariance in harbor seals on the west coast of the USA

2000km



# Shortcut for the Z matrix

|                   |   |  |                   |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
|-------------------|---|--|-------------------|---|---|---|-------------------|---|---|---|-------------------|---|---|---|------------------|---|---|---|--------------|---|---|---|-------------|---|---|---|-------------|---|---|---|-------------------|---|---|---|----------------|---|---|---|----------------|---|---|---|----------------|---|---|---|---|---------|
| count             | = | <table style="border-collapse: collapse;"> <tr><td>Coastal Estuaries</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>Olympic Peninsula</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>Str. Juan de Fuca</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>San Juan Islands</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>Eastern Bays</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>Puget Sound</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>CA.Mainland</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>CA.ChannelIslands</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>OR North Coast</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>OR South Coast</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>Georgia Strait</td><td>0</td><td>1</td><td>0</td></tr> </table> | Coastal Estuaries | 1 | 0 | 0 | Olympic Peninsula | 1 | 0 | 0 | Str. Juan de Fuca | 0 | 1 | 0 | San Juan Islands | 0 | 1 | 0 | Eastern Bays | 0 | 1 | 0 | Puget Sound | 0 | 1 | 0 | CA.Mainland | 0 | 0 | 1 | CA.ChannelIslands | 0 | 0 | 1 | OR North Coast | 1 | 0 | 0 | OR South Coast | 1 | 0 | 0 | Georgia Strait | 0 | 1 | 0 | $\begin{bmatrix} x_{wa,or,t} \\ x_{ps,t} \\ x_{ca,t} \end{bmatrix}$ | + a + v |
| Coastal Estuaries | 1 | 0  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| Olympic Peninsula | 1 | 0  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| Str. Juan de Fuca | 0 | 1  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| San Juan Islands  | 0 | 1  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| Eastern Bays      | 0 | 1  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| Puget Sound       | 0 | 1  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| CA.Mainland       | 0 | 0  | 1                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| CA.ChannelIslands | 0 | 0  | 1                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| OR North Coast    | 1 | 0  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| OR South Coast    | 1 | 0  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |
| Georgia Strait    | 0 | 1  | 0                 |   |   |   |                   |   |   |   |                   |   |   |   |                  |   |   |   |              |   |   |   |             |   |   |   |             |   |   |   |                   |   |   |   |                |   |   |   |                |   |   |   |                |   |   |   |   |         |

Z matrix

`factor(c("or.wa", "or.wa", "ps", "ps", "ps", "ps", "ca", "ca", "or.wa", "or.wa", "ps"))`