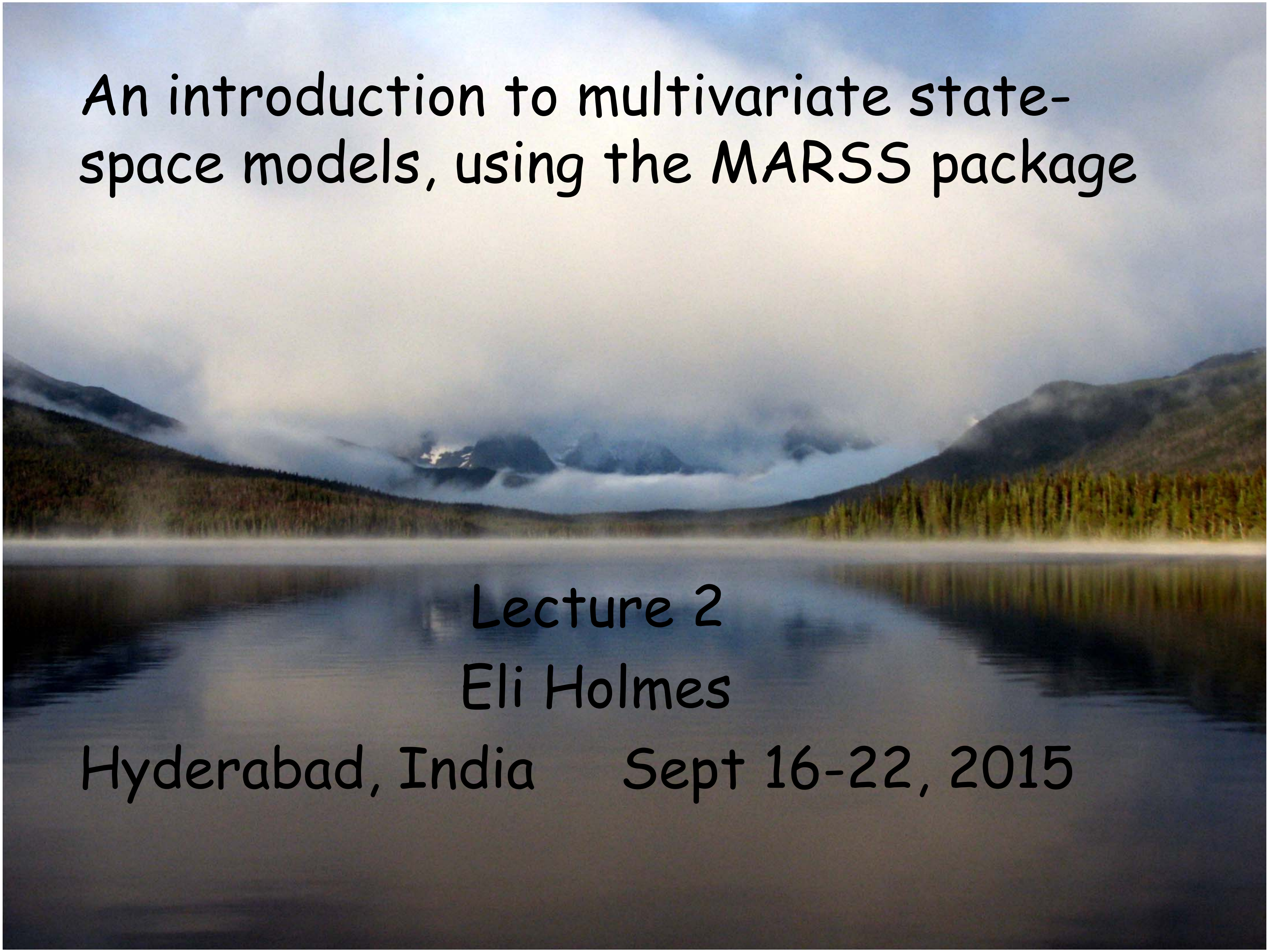


An introduction to multivariate state-space models, using the MARSS package

Lecture 2

Eli Holmes

Hyderabad, India Sept 16-22, 2015

Topics: How to combine time series

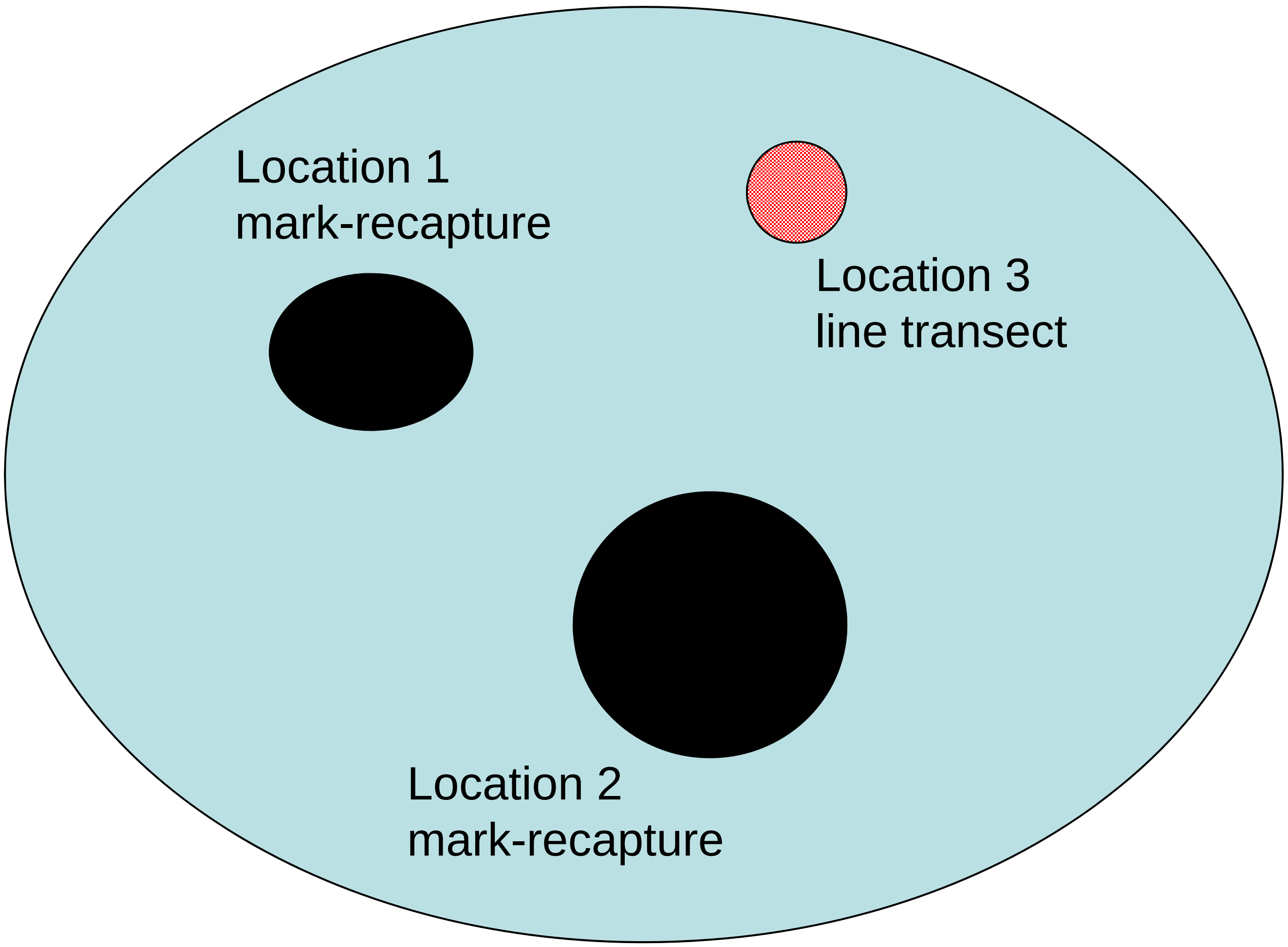
Lecture

- Short example of multivariate observations
- Hand-ons R examples. Starting fitting some data.
- How to explore population structure with MARSS models
- Adding a multivariate observation process
- Model comparison using AIC and AIC weights

More advanced computer Labs

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

Imagine we have 3 sampling locations for a population (denoted x).



Mathematically we can express this like

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

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

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

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

observations

population size

noise

The observation part can be rewritten as a matrix

We need to fix one of the a's.
Traditionally we fix 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 normal.

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

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

The observation errors have a var-cov matrix

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

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

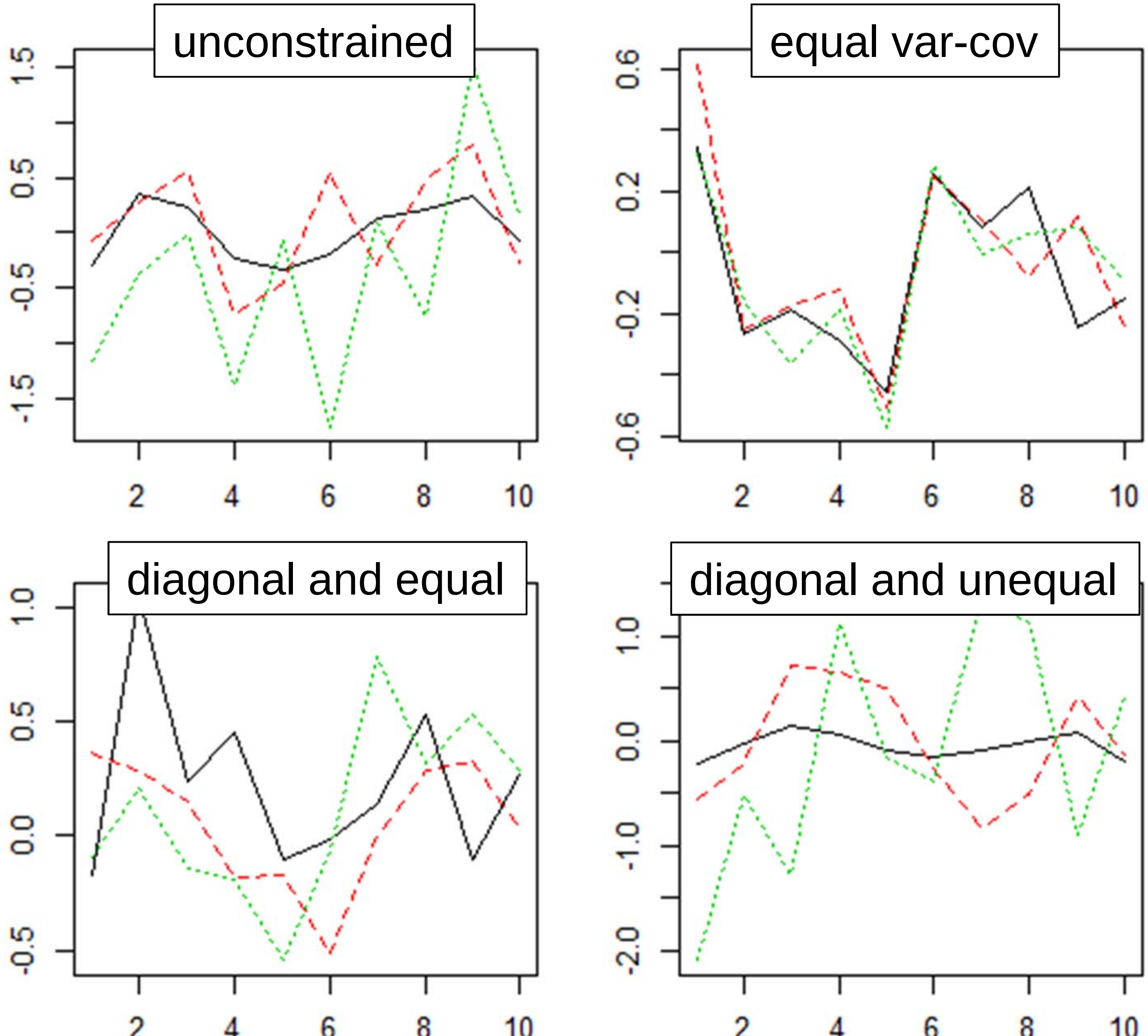
diagonal

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

identical variances and uncorrelated errors

Example of errors coming from these variance-covariance matrices

error



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

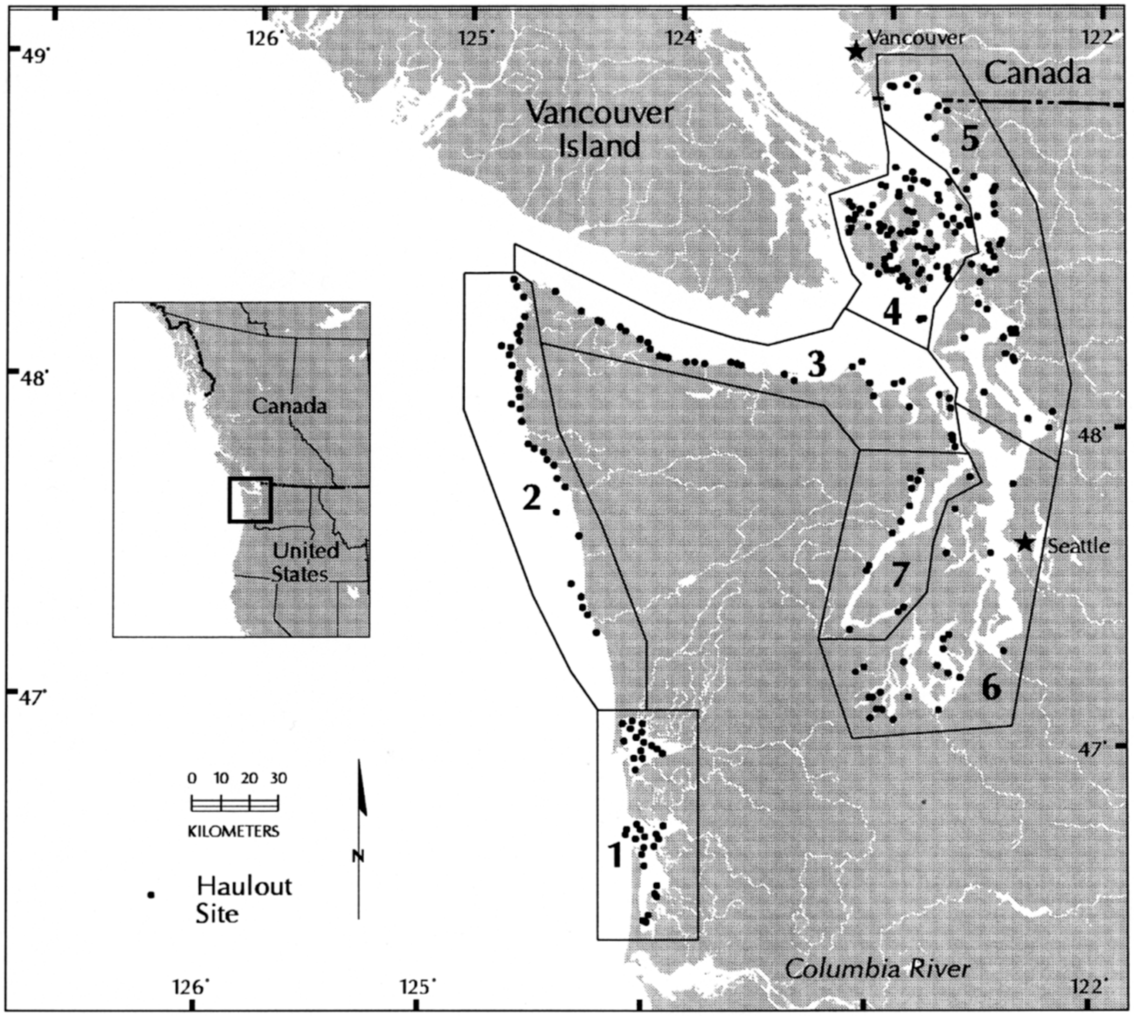
Some short examples. Estimates x with 3 observation time series

- lecture_2_example_1.R
- lecture_2_example_2.R
- lecture_2_example_3.R

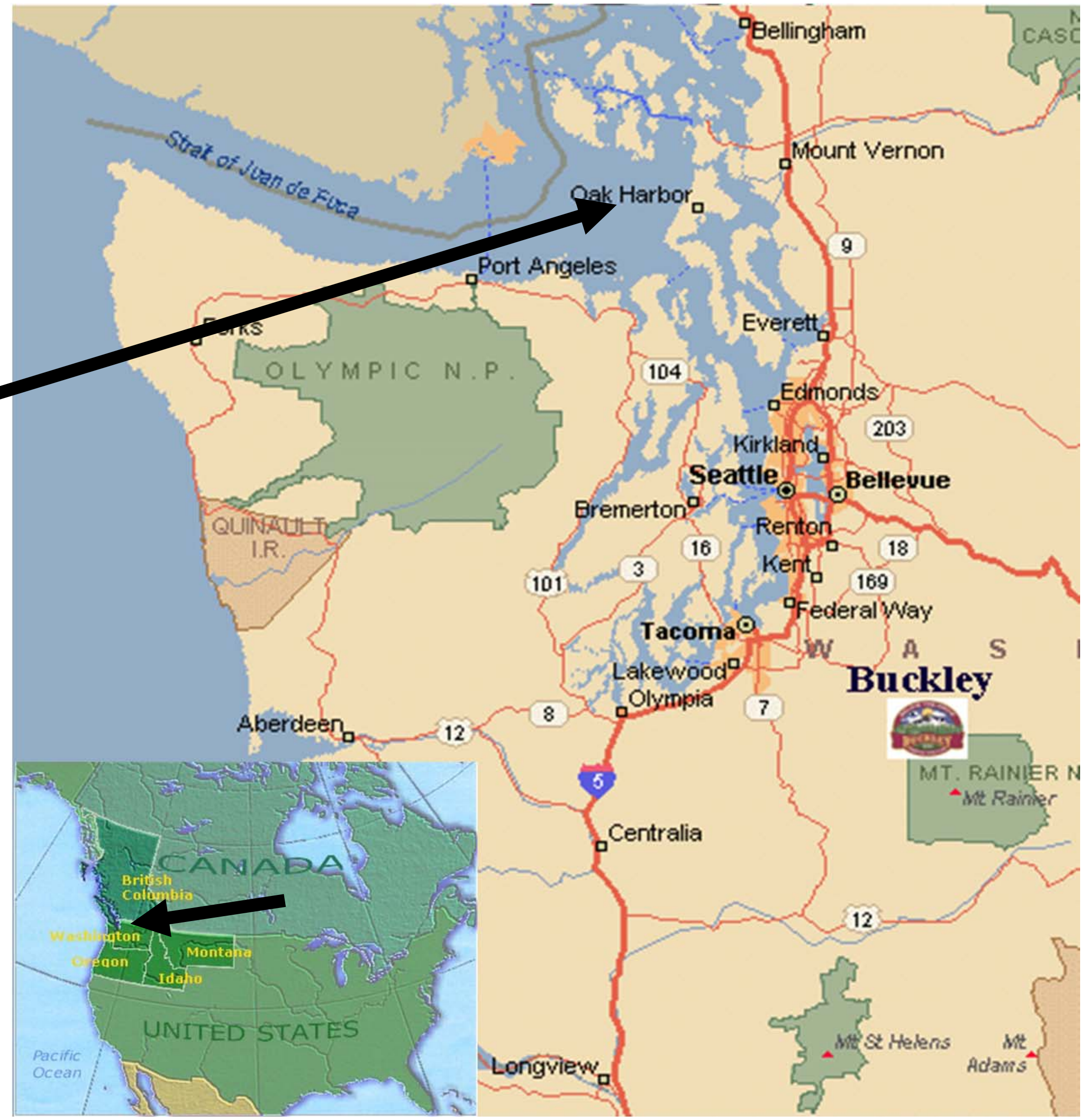
A more realistic example:

Chapter_CombiningTrendData.R

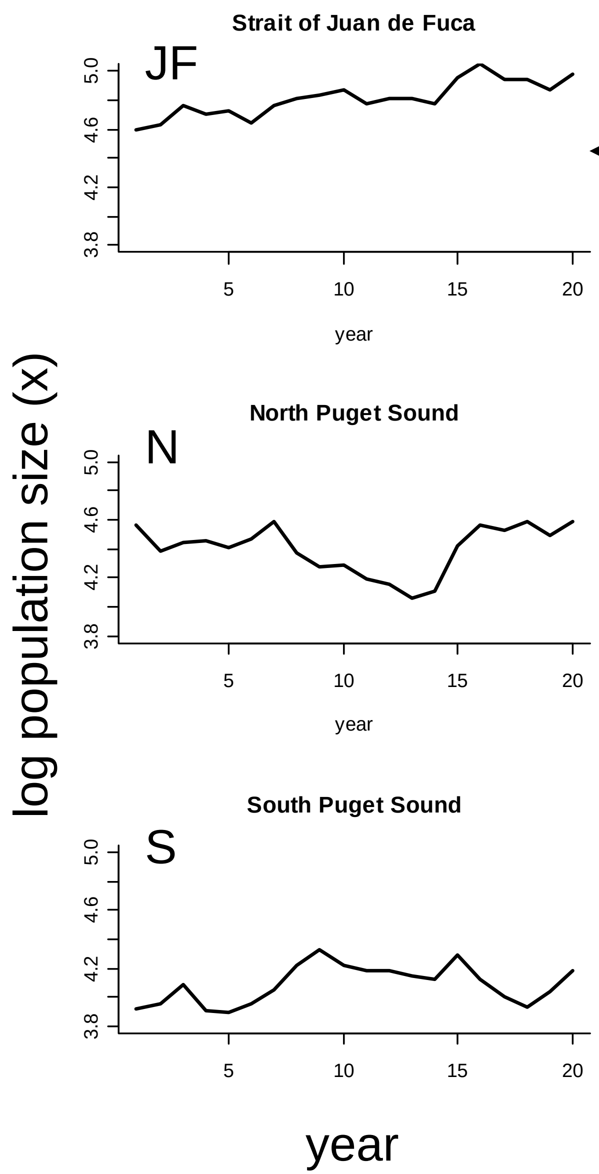
Multi-site data (Pacific harbor seals)

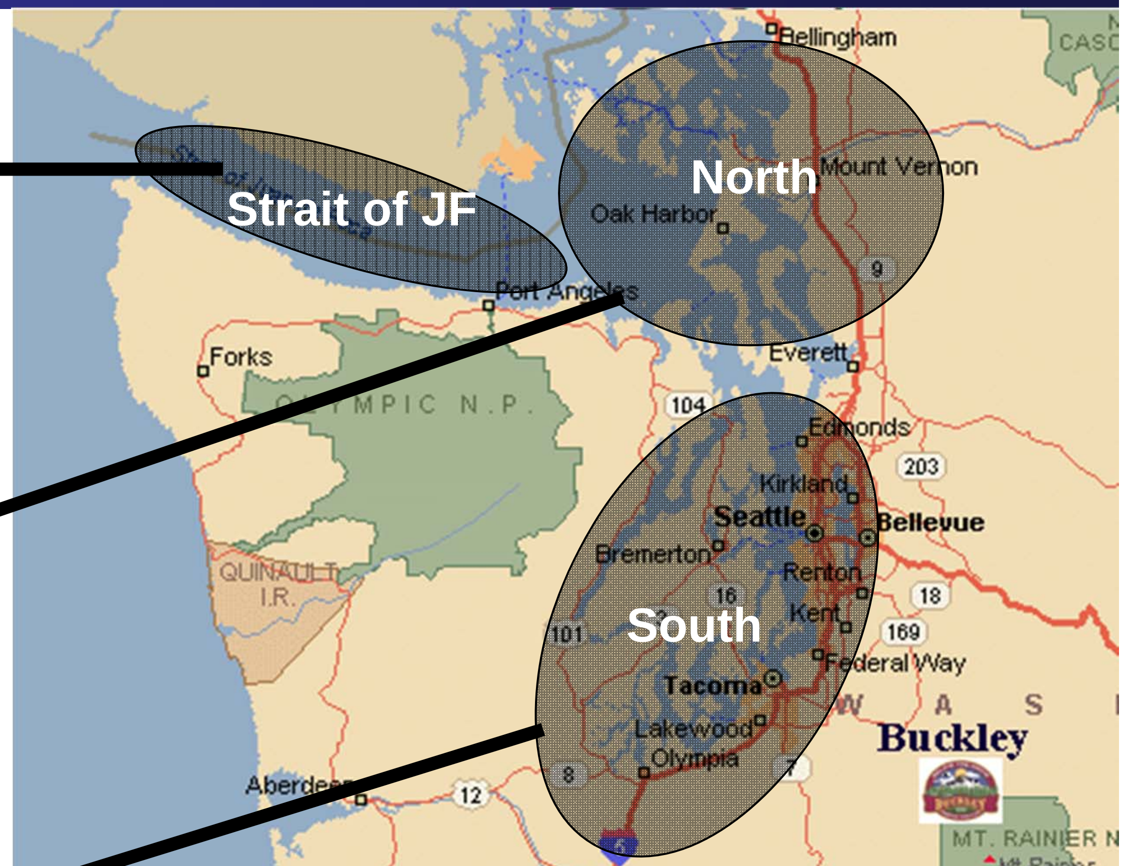


An example: modeling the population dynamics of harbor seals in Puget Sound, WA



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





Three subpopulations that are independent but have correlated population dynamics (dispersal, similar environment, etc.)

A multivariate model for 3 subpopulations

Multivariate stochastic exponential growth

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

3 different x's, one for each subpopulation

3 mean population growth rate terms

3 different process errors

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

The population model in matrix notation

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

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

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: e_t ~MVN(0,Q)

$$\begin{bmatrix} \sigma^2_{JF} & \sigma_{JF,N} & \sigma_{JF,S} \\ \sigma_{JF,N} & \sigma^2_{N} & \sigma_{N,S} \\ \sigma_{JF,S} & \sigma_{N,S} & \sigma^2_{S} \end{bmatrix}$$

unconstrained

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

$$\begin{bmatrix} \sigma^2_{JF} & 0 & 0 \\ 0 & \sigma^2_{N} & 0 \\ 0 & 0 & \sigma^2_{S} \end{bmatrix}$$

diagonal

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

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

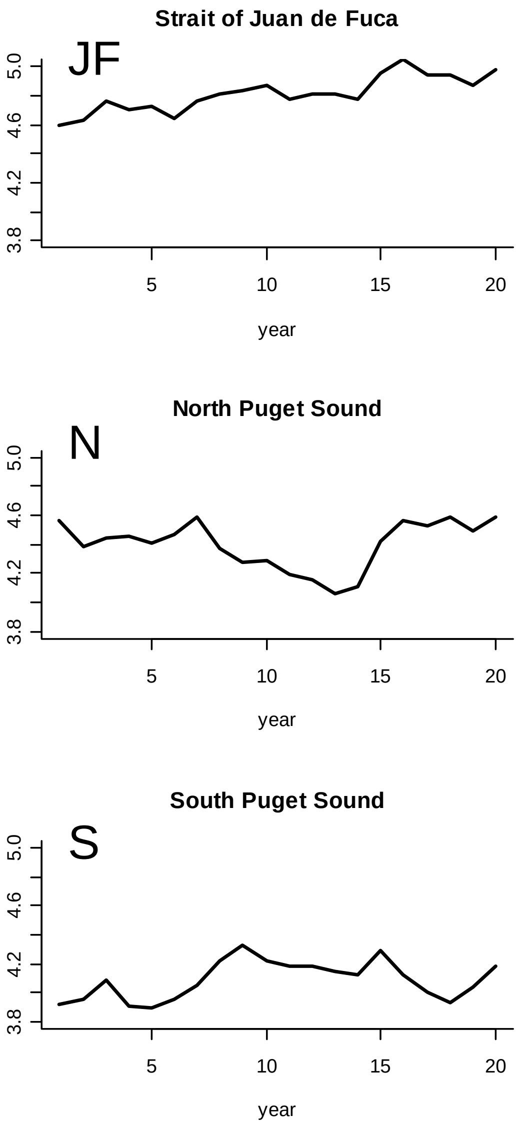
diagonal

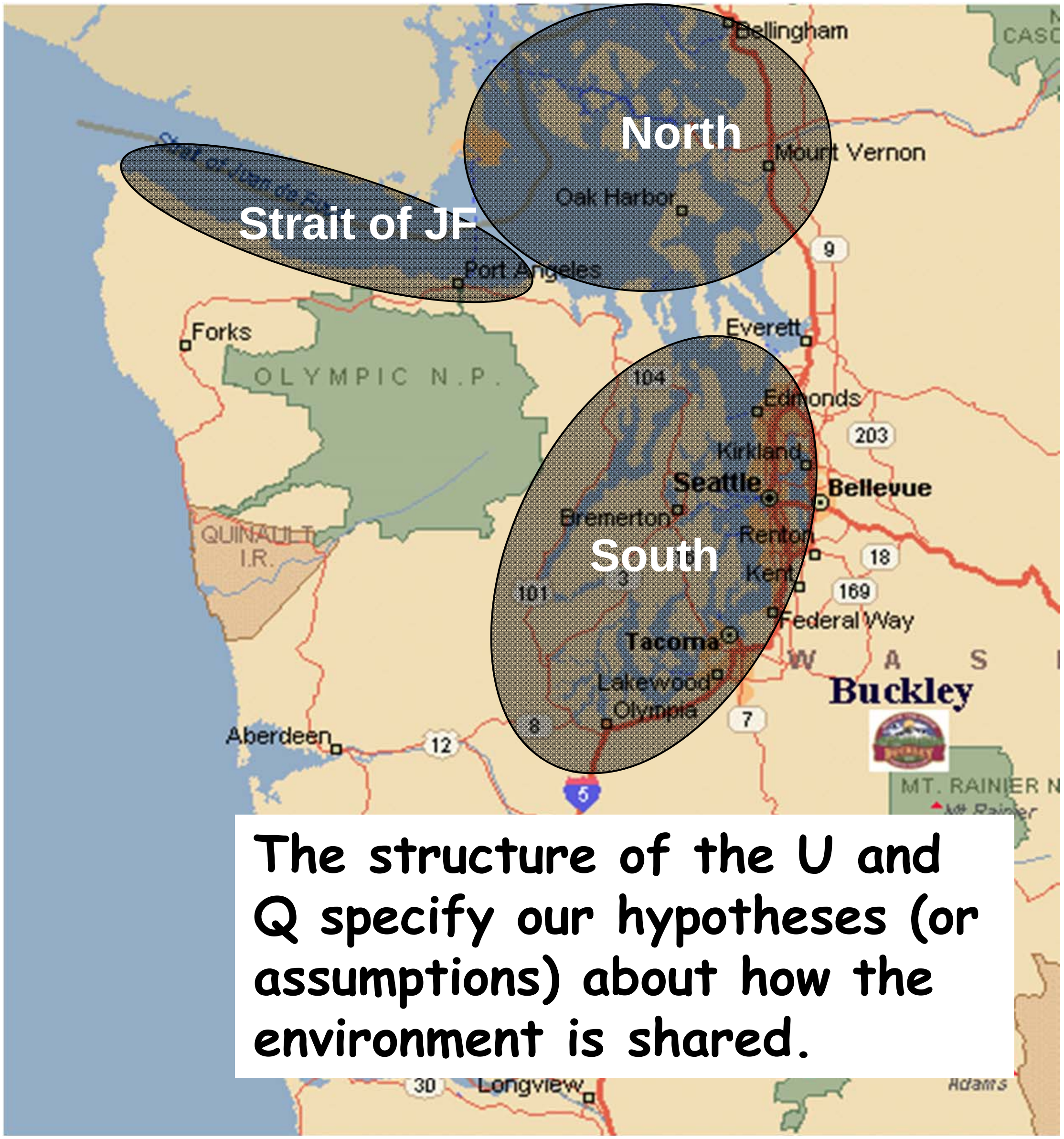
same variances and year-to-year population changes are uncorrelated

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

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

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

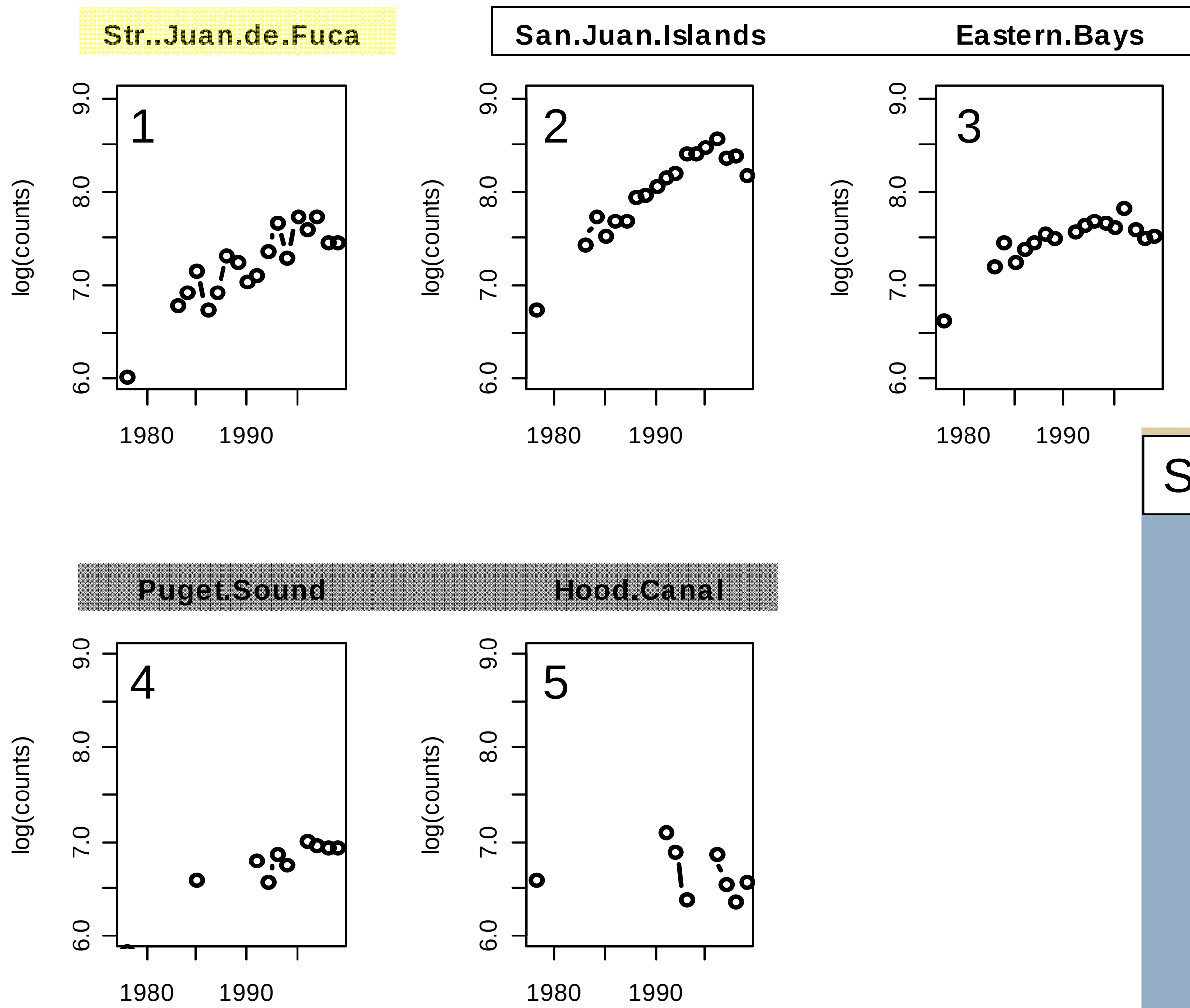




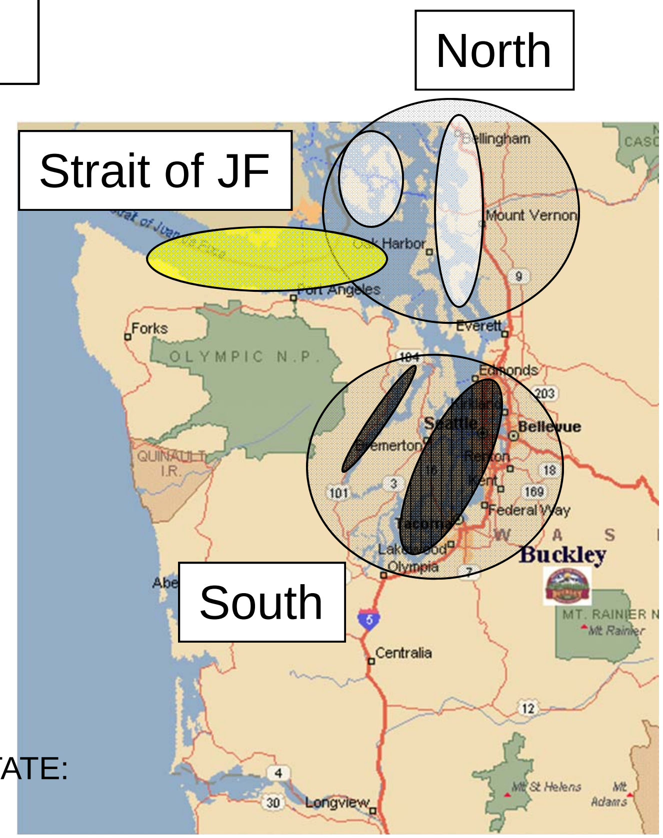
The structure of the U and Q specify our hypotheses (or assumptions) about how the environment is shared.

We observe x and those observations have error

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



5 sampling locations



Jeffries et al. 2003. TRENDS AND STATUS OF HARBOR SEALS IN WASHINGTON STATE: 1978–1999. J. Wildl. Manage. 67(1):208-219

The observation model

true population "hidden"

Log of counts

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} = \begin{array}{c} \begin{array}{ccc} JF & N & S \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \times \begin{bmatrix} x_{JF,t} \\ x_{N,t} \\ x_{S,t} \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \end{bmatrix}$$

Z matrix

relates each observation time series to a different state process

observations

observation biases

measurement errors

The observation errors have a var-cov matrix

$$\begin{bmatrix} \eta_1^2 & \eta_{1,2} & \eta_{1,3} & \eta_{1,4} & \eta_{1,5} \\ \eta_{1,2} & \eta_2^2 & \eta_{3,2} & \eta_{2,4} & \eta_{2,5} \\ \eta_{1,3} & \eta_{3,2} & \eta_3^2 & \eta_{3,4} & \eta_{3,5} \\ \eta_{1,4} & \eta_{2,4} & \eta_{3,4} & \eta_4^2 & \eta_{4,5} \\ \eta_{1,5} & \eta_{2,5} & \eta_{3,5} & \eta_{4,5} & \eta_5^2 \end{bmatrix} \quad \begin{bmatrix} \eta_1^2 & 0 & 0 & 0 & 0 \\ 0 & \eta_2^2 & 0 & 0 & 0 \\ 0 & 0 & \eta_3^2 & 0 & 0 \\ 0 & 0 & 0 & \eta_4^2 & 0 \\ 0 & 0 & 0 & 0 & \eta_5^2 \end{bmatrix} \quad \begin{bmatrix} \eta^2 & 0 & 0 & 0 & 0 \\ 0 & \eta^2 & 0 & 0 & 0 \\ 0 & 0 & \eta^2 & 0 & 0 \\ 0 & 0 & 0 & \eta^2 & 0 \\ 0 & 0 & 0 & 0 & \eta^2 \end{bmatrix}$$

unconstrained

unique variances and uncorrelated errors

diagonal

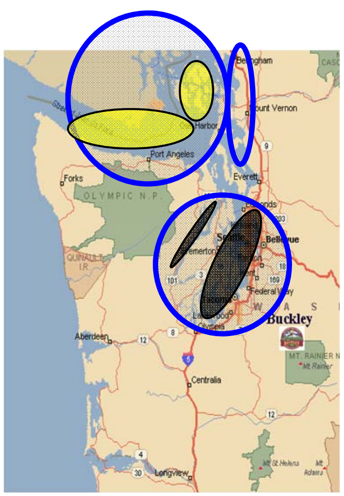
identical variances and uncorrelated errors

diagonal

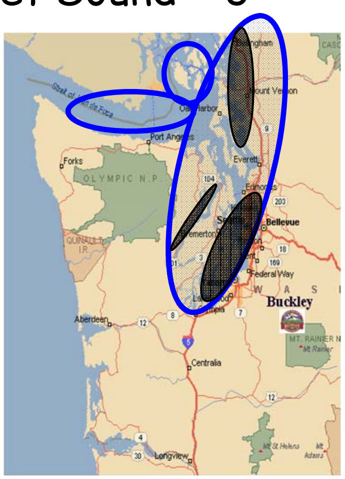
Instead of N, S, Str. J subpopulations, we could have other combinations and numbers of subpopulations

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

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



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



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

The harbor seal multivariate state-space model (MSSM) … in matrix form

3x1 vectors

3x3 matrix

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

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

5x5 matrix

5x1 vectors

* We can add density-dependence or interactions -- Day 4 --

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

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

Interactions with between the subpopulations

density dependence)

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{1,2} & b_{2,2} & b_{3,2} \\ b_{1,3} & b_{3,2} & b_{3,3} \end{bmatrix}$$

* We can add covariates that to explain some of the variability -- afternoon --

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

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

c and d are covariates (like temperature) that you are using to explain some of the variability

$$\mathbf{C}\mathbf{c}_t = \begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{1,2} & C_{2,2} \\ C_{1,3} & C_{3,2} \end{bmatrix} \times \begin{bmatrix} c_{1,t} \\ c_{2,t} \end{bmatrix}$$

* We can incorporate information on known observation variance

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

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

$$\mathbf{R} = \begin{bmatrix} \eta_1^2 & 0 & 0 & 0 & 0 \\ 0 & \eta_2^2 & 0 & 0 & 0 \\ 0 & 0 & \eta_3^2 & 0 & 0 \\ 0 & 0 & 0 & \eta_4^2 & 0 \\ 0 & 0 & 0 & 0 & \eta_5^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^2 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_{2,t}^2 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_{3,t}^2 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon_{4,t}^2 & 0 \\ 0 & 0 & 0 & 0 & \varepsilon_{5,t}^2 \end{bmatrix}$$

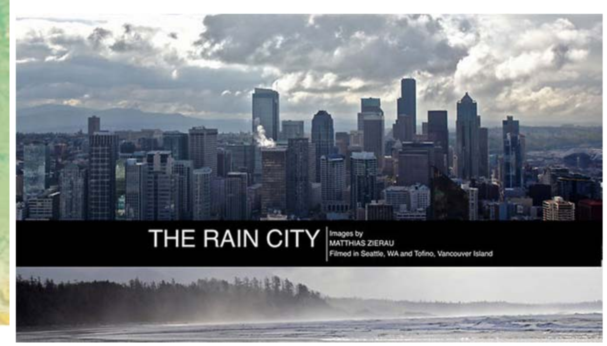
The unknown part which comes from the unknown way that the environment and individual attributes affects “measurability”

The known part which comes from the monitoring design (sample size)

Computer lab

Chapter 7 & 8: Identifying spatial structure and covariance in harbor seals on the west coast of the USA

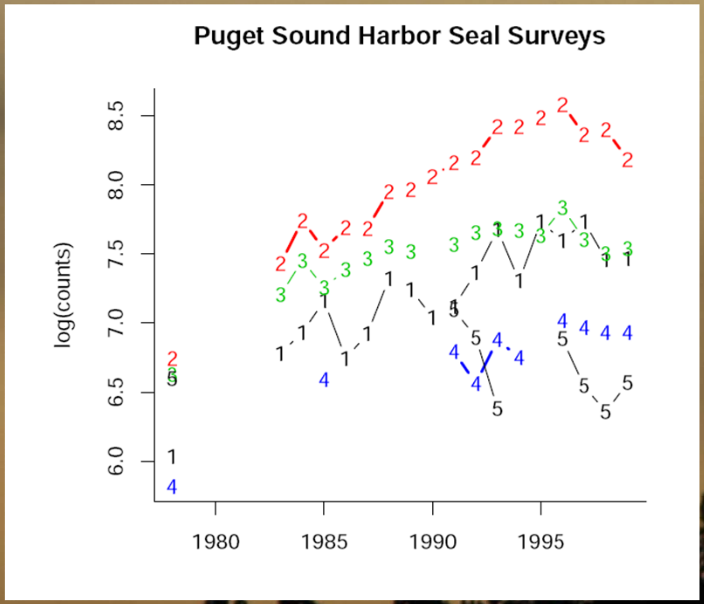
Place names you'll see in the code



Computer lab #1: Get used to multivariate modeling

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

- RShowDoc("Chapter_SealTrend.R",package="MARSS")
- Work through through the chapter and text on your own.



Computer lab #2: Use model select to test hypotheses about subpop structure

Chapter 8: Identifying spatial structure and covariance in harbor seals on the west coast of the USA

- RShowDoc("Chapter_SealPopStructure.R",package="MARSS")
- Work through to section 8.4

2000km

Structure of the observation models

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ x_s \end{bmatrix}_t + \begin{bmatrix} 0 \\ a_2 \\ 0 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}_t$$

observations | Z matrix | subpopulation size | bias | noise

Shortcut for the Z matrix

$$
\text{count} \quad
\begin{array}{r}
\texttt{Coastal Estuaries} \\
\texttt{Olympic Peninsula} \\
\texttt{Str. Juan de Fuca} \\
\texttt{San Juan Islands} \\
\texttt{Eastern Bays} \\
\texttt{Puget Sound} \\
\texttt{CA.Mainland} \\
\texttt{CA.ChannelIslands} \\
\texttt{OR North Coast} \\
\texttt{OR South Coast} \\
\texttt{Georgia Strait}
\end{array}
=
\begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
x_{wa.or,t} \\
x_{ps,t} \\
x_{ca,t}
\end{bmatrix}
+ a + v
$$

Z matrix

factor(c(“or.wa”,”or.wa,”ps”,”ps”,”ps”,”ps”,”ca”,”ca”,”or.wa”,”or.wa”,”ps”))

Model selection for ARMA models

H	Q	delta.AICc	AIC.weight
NC+Strait+PS+SC	diagonal and equal	0.00	0.886
NC+Strait+PS+SC	diagonal and unequal	4.15	0.112
N-S	diagonal and unequal	12.67	0.002
N-S	diagonal and equal	14.78	0.001
coast+PS	diagonal and equal	31.23	0.000
coast+PS	diagonal and unequal	33.36	0.000
stock	diagonal and equal	34.01	0.000
stock	diagonal and unequal	36.84	0.000
panmictic	diagonal and equal	48.28	0.000
panmictic	diagonal and unequal	48.28	0.000
site	diagonal and equal	56.36	0.000
site	diagonal and unequal	57.95	0.000

Measuring the support for different model structures

The likelihood: the probability of the data given your statistical model

Data: 1, 1, 1, 2, 1

Model 1: I produced these data by drawing from a cup with 1s and 2s in it. P(1)=p and p=0.5
Likelihood 1: 1/2 x 1/2 x 1/2 x 1/2 x 1/2 = 0.03125

Model 2: Same cup but p = 0.75
Likelihood 2: 3/4 x 3/4 x 3/4 x 3/4 x 3/4 = 0.2373

The likelihood of the data under "good" models is higher than under the "bad" models.

Why not just use likelihood as our metric?

Because if you add enough parameters and structure, you can overfit the data. The model fits the data perfectly and is completely useless.

Data: 1, 1, 1, 2, 1

"Best" *model* : I have 5 cups. I estimate a different p for each cup. Result. Cup 1 p=1, cup 2 p=1, cup 3 p=1, cup 4 p=0 and cup 5 p=1
Likelihood 3 : 1 x 1 x 1 x 1 x 1 = 1

Model fits the data perfectly but is useless.

The basic way we correct for this is to "penalize" for the number of parameters we estimate

There are many "penalty" functions. AIC is based on a specific penalty = 2 times the number of estimated parameters.

$$AIC = -2 \times \log(L(\theta \mid x)) + 2 \times k$$

More parameters means the likelihood should be higher – more flexibility to fit the data better

Penalty for the number of estimated parameters

AIC Basics

AIC basics for the case studies

- Smaller is “better”
- Only the relative difference matters
- A difference of less than 3 means the models have similar support
- A difference greater than 10 is big
- Always use the same data for each model

AICc is a small sample size correcton

- AIC

$$AIC = -2 \times \log(L(\theta \mid x)) + 2 \times k$$

- AICc (small sample AIC)

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

n = sample size

Burnham & Anderson (2002)

AIC weight

the Akaike weight gives the probability that a given model would have the lowest AIC for a replicate equivalent dataset.

$$w_i = \frac{\exp\left[-\frac{1}{2}\Delta_i\right]}{\sum_{i=1}^{m}\exp\left[-\frac{1}{2}\Delta_i\right]}$$

relative likelihood

sum of relative likelihoods

H	Q	delta.AICc	AIC.weight
NC+Strait+PS+SC	equalvarcov	0.00	0.976
site	equalvarcov	7.65	0.021
NC+Strait+PS+SC	unconstrained	11.47	0.003
NC+Strait+PS+SC	diagonal and equal	23.39	0.000
NC+Strait+PS+SC	diagonal and unequal	27.53	0.000
N-S	unconstrained	32.61	0.000
N-S	diagonal and unequal	36.06	0.000
N-S	equalvarcov	36.97	0.000
stock	equalvarcov	37.82	0.000
N-S	diagonal and equal	38.16	0.000