

Covariates in Time Series Models

FISH 507 – Applied Time Series Analysis

Eli Holmes

26 Jan 2021

Topics for today

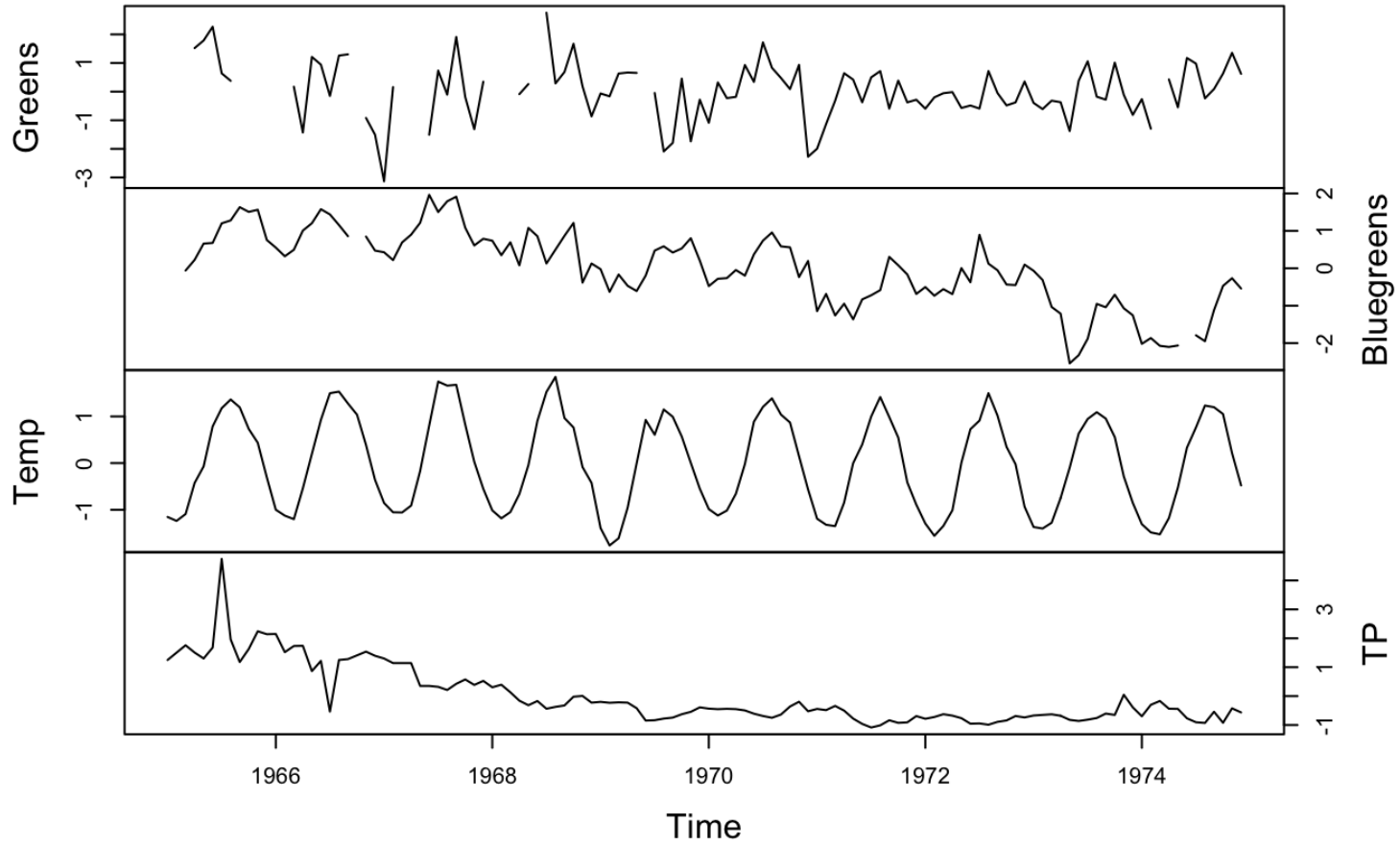
Covariates

- Why include covariates?
- Multivariate linear regression on time series data
- Covariates in MARSS models
- Seasonality in MARSS models
- Missing covariates

Why include covariates in a model?

- You want to forecast something using covariates
- We are often interested in knowing the cause of variation
- Covariates can explain the process that generated the patterns
- Covariates can help deal with problematic observation errors
- You are using covariates to model a changing system

Lake WA plankton and covariates



Covariates in time series models

- Multivariate linear regression for time series data
- Linear regression with ARMA errors
- ARMAX
 - MARSS models with covariates = ARX plus covariates
 - aka Vector Autoregressive Models with covariates and observation error

Multivariate linear regression for time series data

Can you do a linear regression with time series data (response and predictors)? Yes, but you need to be careful. Read Chapter 5 in [Hyndman and Athanasopoulos 2018](#)

- Diagnostics that need to be satisfied
 - Residuals are temporally uncorrelated
 - Residuals are not correlated with the predictor variables
- Be careful regarding spurious correlation if both response and predictor variables have trends

Linear regression with autocorrelated errors

The `xreg` argument in `Arima()` and `arima()` allows you to fit linear regressions with autocorrelated errors. Read Chapter 9 in [Hyndman and Athanasopoulos 2018](#) on Dynamic Regression.

A linear regression with autocorrelated errors is for example:

$$y_t = \alpha + \mathbf{Dd}_t + \nu_t$$
$$\nu_t = \theta_1\nu_{t-1} + \theta_2\nu_{t-2} + e_t$$

Fitting in R

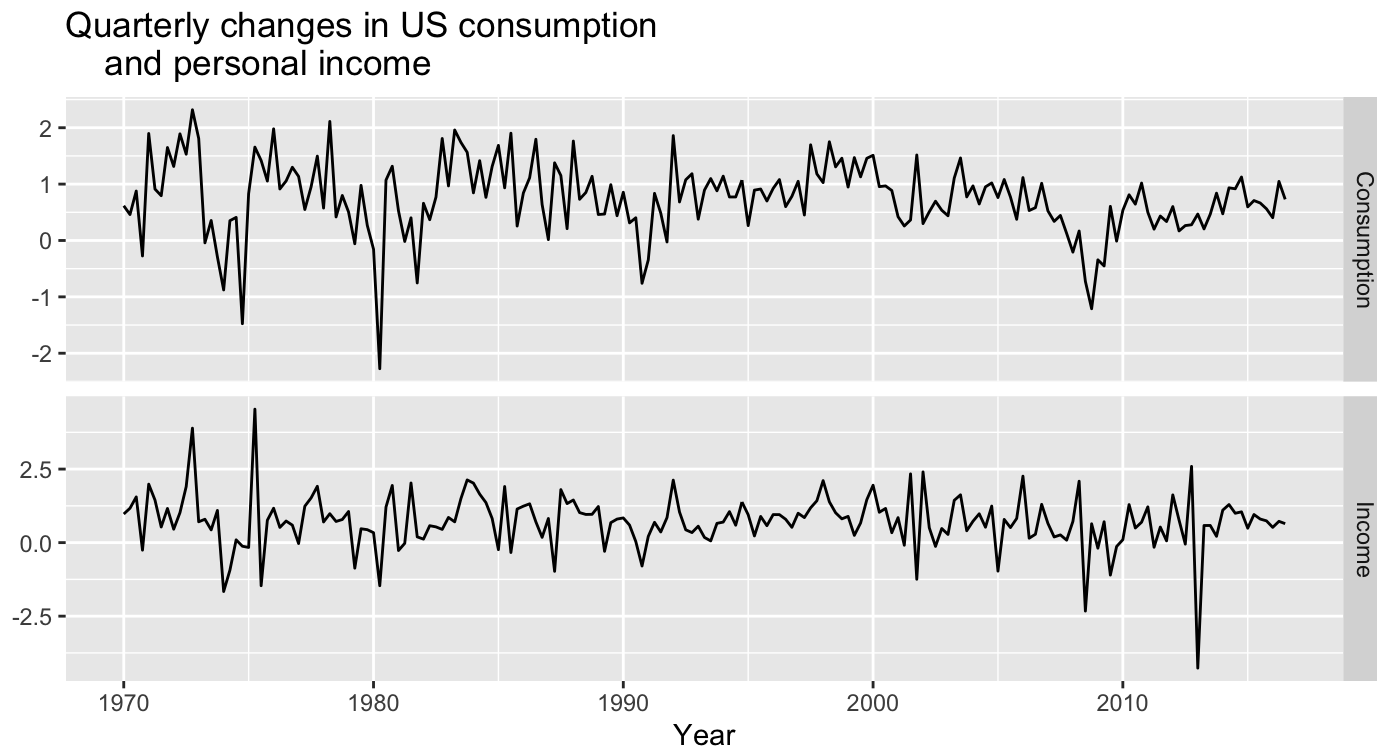
`Arima()`

```
fit <- Arima(y, xreg=d, order=c(1,1,0))
```

`auto.arima()`

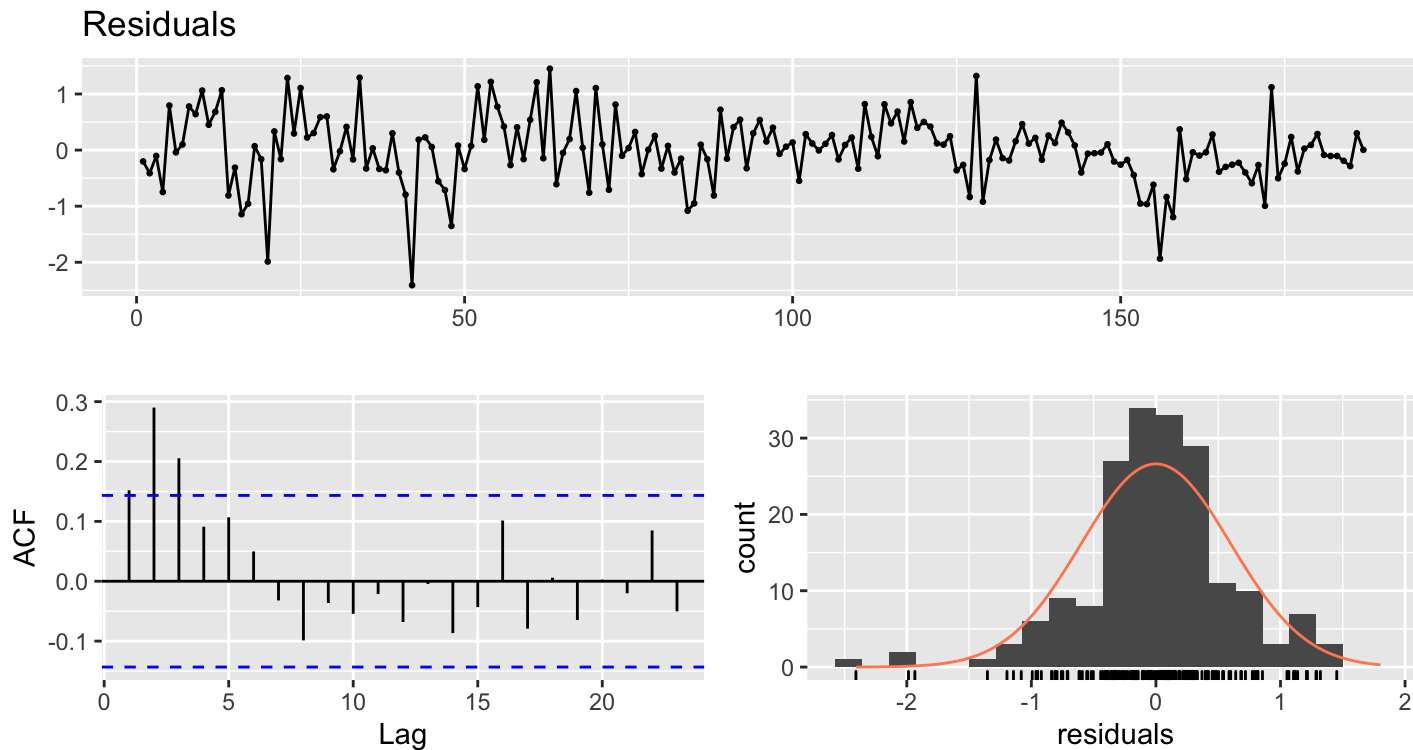
```
fit <- auto.arima(y, xreg=x)
```


Example from Hyndman and Athanasopoulos 2018



A simple regression has problems

```
y <- uschange[, "Consumption"]; d <- uschange[, "Income"]  
fit <- lm(y~d)  
checkresiduals(fit)
```



```
##
```

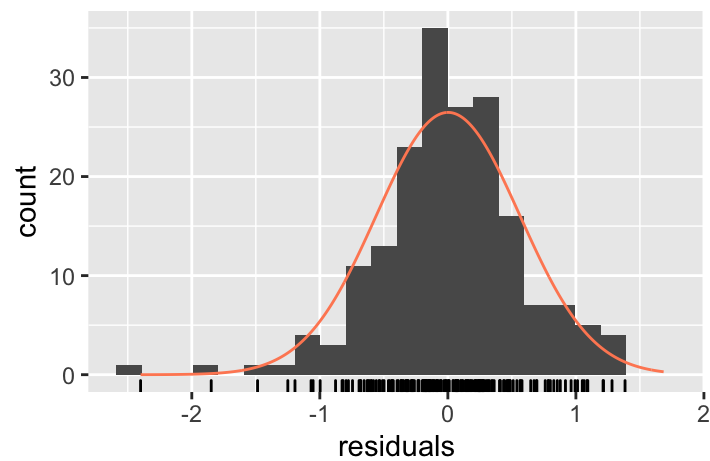
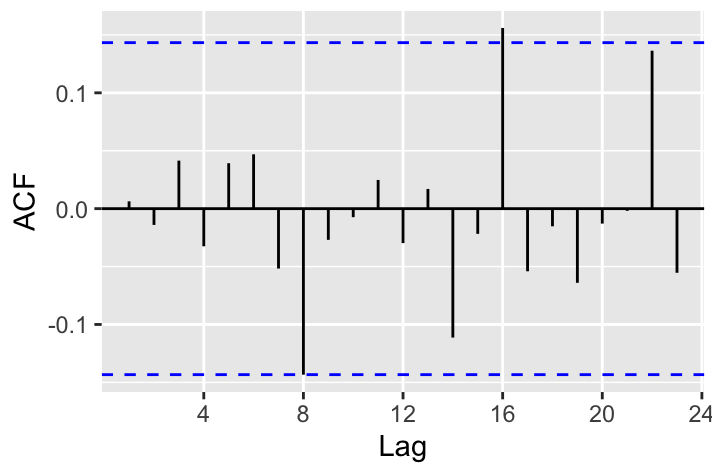
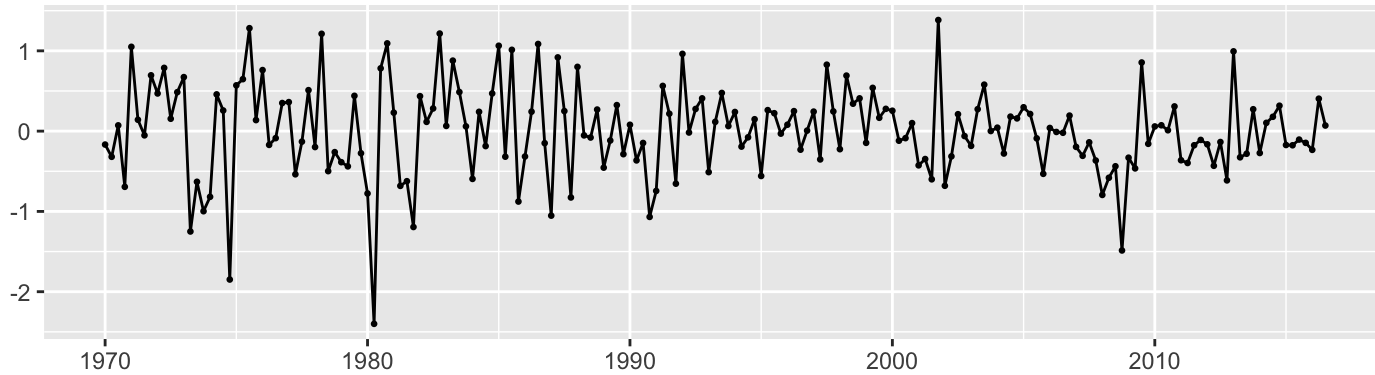
```
## Breusch-Godfrey test for serial correlation of order up to 10
```

```
##
```

Let `auto.arima()` find best ARMA model

```
fit <- auto.arima(y, xreg=d) # It finds a ARMA(1,0,2) is best.  
checkresiduals(fit)
```

Residuals from Regression with ARIMA(1,0,2) errors



ARMAX

$$x_t = bx_{t-1} + \boxed{C\epsilon_t + w_t}$$

where w_t can be moving average process. $w_t = e_t + \theta e_{t-1}$.

Covariates in univariate state-space models

$$x_t = bx_{t-1} + \mathbf{C}\mathbf{c}_t + w_t$$

$$y_t = x_t + \mathbf{D}\mathbf{d}_t + v_t$$

Now we can model how covariates affect the hidden process.

Covariates in univariate state-space models

$$x_t = x_{t-1} + u + \boxed{\mathbf{C}c_t} + w_t$$
$$y_t = x_t + v_t$$

Random walk with drift. How does covariate affect the drift term?

Example. You have tag data on movement of animals in the ocean. How does water temperature affect the speed (jump length) of the movement.

Covariates in univariate state-space models

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

$$y_t = x_t + \boxed{\mathbf{D}d_t + v_t}$$

How does covariate affect observation error relative to our stochastic trend.

Example. You are tracking population size using stream surveys. Turbidity affects your observation error.

Covariates in MARSS models

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

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

Covariates in state process

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

$\mathbf{C}\mathbf{c}_t$ The covariate is in \mathbf{c}_t and the effect is in matrix \mathbf{C} .

Example. lat/lon movement data so \mathbf{x} and \mathbf{y} are 2 dimensional (our lat and lon values).

Covariates in the state process

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_t = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{t-1} + \begin{bmatrix} C_a & C_b \\ C_a & C_b \end{bmatrix} \begin{bmatrix} Temp \\ TP \end{bmatrix}_t + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_t$$

The model for x_t in site 1 (or species 1) is:

$$x_{1,t} = x_{1,t-1} + C_a \times Temp_t + C_b \times TP_t + w_{1,t}$$

There is an effect of the prior x_t and an effect of temp and wind.

The structure of \mathbf{C}

The structure of \mathbf{C} can model different effect structures

Effect of temp and TP is the same

$$\begin{bmatrix} C & C \\ C & C \end{bmatrix} \begin{bmatrix} Temp \\ TP \end{bmatrix}_t$$

Effect of temperature and TP is different but the same across sites, species, whatever the row in \mathbf{X} is

$$\begin{bmatrix} C_a & C_b \\ C_a & C_b \end{bmatrix} \begin{bmatrix} Temp \\ TP \end{bmatrix}_t$$

Effect of temperature and TP is different across sites or whatever the row in \mathbf{X} is

$$\begin{bmatrix} C_{a1} & C_{b1} \\ C_{a2} & C_{b2} \end{bmatrix} \begin{bmatrix} Temp \\ TP \end{bmatrix}_t$$

Effect of temperature is the same across sites but TP is not

$$\begin{bmatrix} C_a & C_{b1} \\ C_a & C_{b2} \end{bmatrix} \begin{bmatrix} Temp \\ TP \end{bmatrix}_t$$

Covariate in the observation process

eg, vegetation obscures individuals, temperature affects behavior making animals more or less visible

$$y_t = \mathbf{Zx}_t + \mathbf{a} + \mathbf{Dd}_t + \mathbf{w}_t$$

Covariates in the observation process

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_t + \begin{bmatrix} D_a & D_b \\ D_a & D_b \\ D_a & D_b \end{bmatrix} \begin{bmatrix} temp \\ wind \end{bmatrix}_t + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_t$$

In this case the covariate does not affect the *state* x . It only affects the observation of the state.

The model for y_t in site 1 is:

$$y_{1,t} = x_{1,t} + D_a \times temp_t + D_b \times wind_t + v_{1,t}$$

The structure of **D**

The structure of **D** can model many different structures of the effects.

Effect of temp and wind is the same across sites 1 & 2 but different for site 3. In site 3, temp has an effect but wind does not

$$\begin{bmatrix} D_a & D_b \\ D_a & D_b \\ D_c & 0 \end{bmatrix} \begin{bmatrix} temp \\ wind \end{bmatrix}_t$$

Why include covariates in a MARSS model?

- We want to understand how covariates drive the hidden process.
- We want to test hypotheses for what caused a perturbation or change in the dynamics.
- We want to forecast using covariates.
- We want to model the *autocorrelation in the process errors* using the known driver.

Why include covariates in a model?

Auto-correlated observation errors

- Model your v_t as a AR-1 process. hard numerically with a large multivariate state-space model
- If know what is causing the auto-correlation, include that as a covariate. Easier.

Correlated observation errors across sites or species (y rows)

- Use a **R** matrix with off-diagonal terms. really hard numerically
- If you know or suspect what is causing the correlation, include that as a covariate. Easier.

“hard numerically” = you need a lot of data

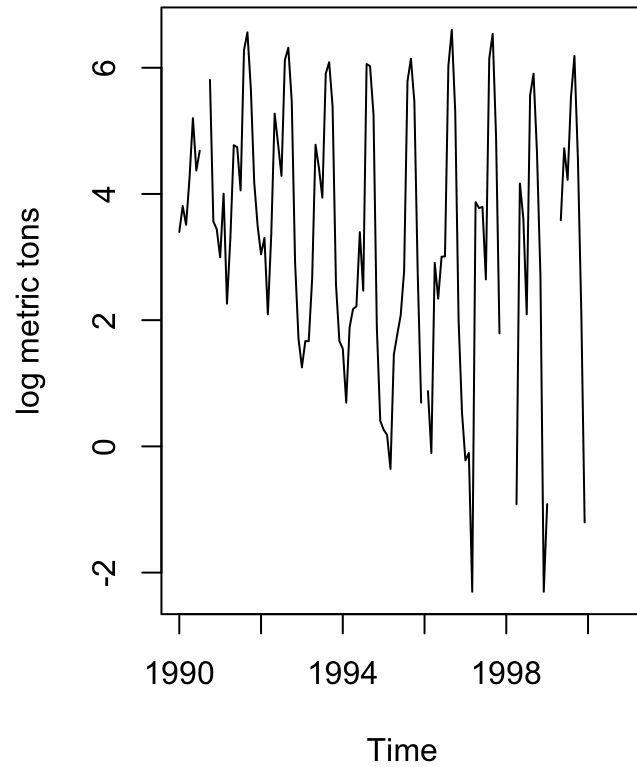
Let's work through an example

`lec_07_covariates.R`

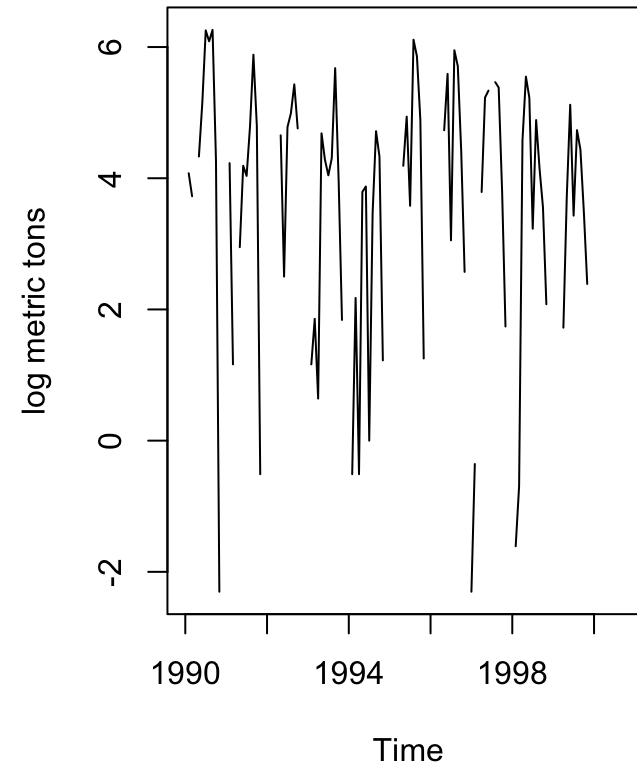
Follows Chapter 8: <https://nwfsc-timeseries.github.io/atsa-labs>

Seasonality

WA chinook landings



OR chinook landings



Seasonality

- Different approaches to modeling seasonality
 - Factors
 - Polynomials
 - Sines and cosines (Fourier series)

Monthly factors

Introduce 12 covariates: January, February, etc. If t is in January the January covariate is 1 otherwise it is 0.

$$x_t = x_{t-1} + u + \mathbf{C}\mathbf{c}_t + w_t$$

$$\mathbf{C}\mathbf{c}_t = \alpha_m$$

where m is month at time t .

Monthly factors

$$\begin{array}{r}
 \text{WA} \\
 \text{OR}
 \end{array}
 \begin{array}{c}
 \text{month effects} \\
 \text{a 2 by 12 matrix}
 \end{array}
 \begin{bmatrix}
 C_1 & C_2 & C_3 & \dots & C_{12} \\
 C_1 & C_2 & C_3 & \dots & C_{12}
 \end{bmatrix}
 \begin{array}{c}
 \text{covariates} \\
 \text{a 12 by } T \text{ matrix}
 \end{array}
 \begin{bmatrix}
 1 & 0 & 0 & \dots & 0 \\
 0 & 1 & 0 & \dots & 0 \\
 0 & 0 & 1 & \dots & 0 \\
 \vdots & \vdots & \vdots & \vdots & \\
 0 & 0 & 0 & \dots & 1
 \end{bmatrix}
 \begin{array}{l}
 \text{Jan} \\
 \text{Feb} \\
 \text{Mar} \\
 \vdots \\
 \text{Dec} \\
 \text{month}
 \end{array}
 \begin{array}{l}
 1 \\
 2 \\
 3 \\
 \dots \\
 T
 \end{array}$$


```

TT <- nrow(chinook.month)/2
covariate <- matrix(0, 12, TT)
monrow <- match(chinook.month$Month, month.abb)[1:TT]
covariate[cbind(monrow,1:TT)] <- 1
covariate[,1:12]

```

```

##           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
## [1,]         1    0    0    0    0    0    0    0    0    0    0    0
## [2,]         0    1    0    0    0    0    0    0    0    0    0    0
## [3,]         0    0    1    0    0    0    0    0    0    0    0    0
## [4,]         0    0    0    1    0    0    0    0    0    0    0    0
## [5,]         0    0    0    0    1    0    0    0    0    0    0    0
## [6,]         0    0    0    0    0    1    0    0    0    0    0    0
## [7,]         0    0    0    0    0    0    1    0    0    0    0    0
## [8,]         0    0    0    0    0    0    0    1    0    0    0    0
## [9,]         0    0    0    0    0    0    0    0    1    0    0    0
## [10,]        0    0    0    0    0    0    0    0    0    1    0    0
## [11,]        0    0    0    0    0    0    0    0    0    0    1    0
## [12,]        0    0    0    0    0    0    0    0    0    0    0    1

```

WA and OR have different month effects.

```
C <- matrix(paste0(month.abb,rep(1:2,each=12)), 2, 12, byrow = TRUE)
```

```
C
##      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]  [,10]
## [1,] "Jan1" "Feb1" "Mar1" "Apr1" "May1" "Jun1" "Jul1" "Aug1" "Sep1" "Oct1"
## [2,] "Jan2" "Feb2" "Mar2" "Apr2" "May2" "Jun2" "Jul2" "Aug2" "Sep2" "Oct2"
##      [,11] [,12]
## [1,] "Nov1" "Dec1"
## [2,] "Nov2" "Dec2"
```

WA and OR have same month effects.

```
C <- matrix(month.abb, 2, 12, byrow = TRUE)
```

```
C
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
## [1,] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct" "Nov" "Dec"
## [2,] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct" "Nov" "Dec"
```

Season as a 3rd order polynomial

Introduce 3 covariates: m , m^2 and m^3 where m is month (1 to 12).

$$x_t = x_{t-1} + u + \mathbf{C}\mathbf{c}_t + w_t$$

$$\mathbf{C}\mathbf{c}_t = \beta_1 m + \beta_2 m^2 + \beta_3 m^3$$

where m is month at time t .

Season as polynomial

		covariates	
	month effects	a 3 by T matrix	
	a 2 by 3 matrix	$\begin{bmatrix} 1 & 2 & 3 & \dots & 12 \\ 1^2 & 2^2 & 3^2 & \dots & 12^2 \\ 1^3 & 3^3 & 3^3 & \dots & 12^3 \end{bmatrix}$	m m^2 m^3
WA	$\begin{bmatrix} C_1 & C_2 & C_3 \end{bmatrix}$		
OR	$\begin{bmatrix} C_1 & C_2 & C_3 \end{bmatrix}$	$1 \quad 2 \quad 3 \quad \dots \quad T$	month

```
TT <- nrow(chinook.month)/2
monrow <- match(chinook.month$Month, month.abb)[1:TT]
covariate <- rbind(monrow, monrow^2, monrow^3)
rownames(covariate) <- c("m", "m2", "m3")
covariate[,1:13]
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## m      1    2    3    4    5    6    7    8    9   10   11   12    1
## m2     1    4    9   16   25   36   49   64   81  100  121  144    1
## m3     1    8   27   64  125  216  343  512  729 1000 1331 1728    1
```

WA and OR have different seasonal pattern.

```
C <- matrix(paste0(c("m", "m2", "m3"), ".", rep(1:2, each=3)), 2, 3,  
            byrow = TRUE)
```

C

```
##      [,1] [,2] [,3]  
## [1,] "m.1" "m2.1" "m3.1"  
## [2,] "m.2" "m2.2" "m3.2"
```

WA and OR have same seasonal pattern.

```
C <- matrix(c("m", "m2", "m3"), 2, 3, byrow = TRUE)
```

C

```
##      [,1] [,2] [,3]  
## [1,] "m"  "m2" "m3"  
## [2,] "m"  "m2" "m3"
```

Season as a Fourier series

- Fourier series are paired sets of sine and cosine waves
- They are commonly used in time series analysis in the frequency domain

Season as a Fourier series

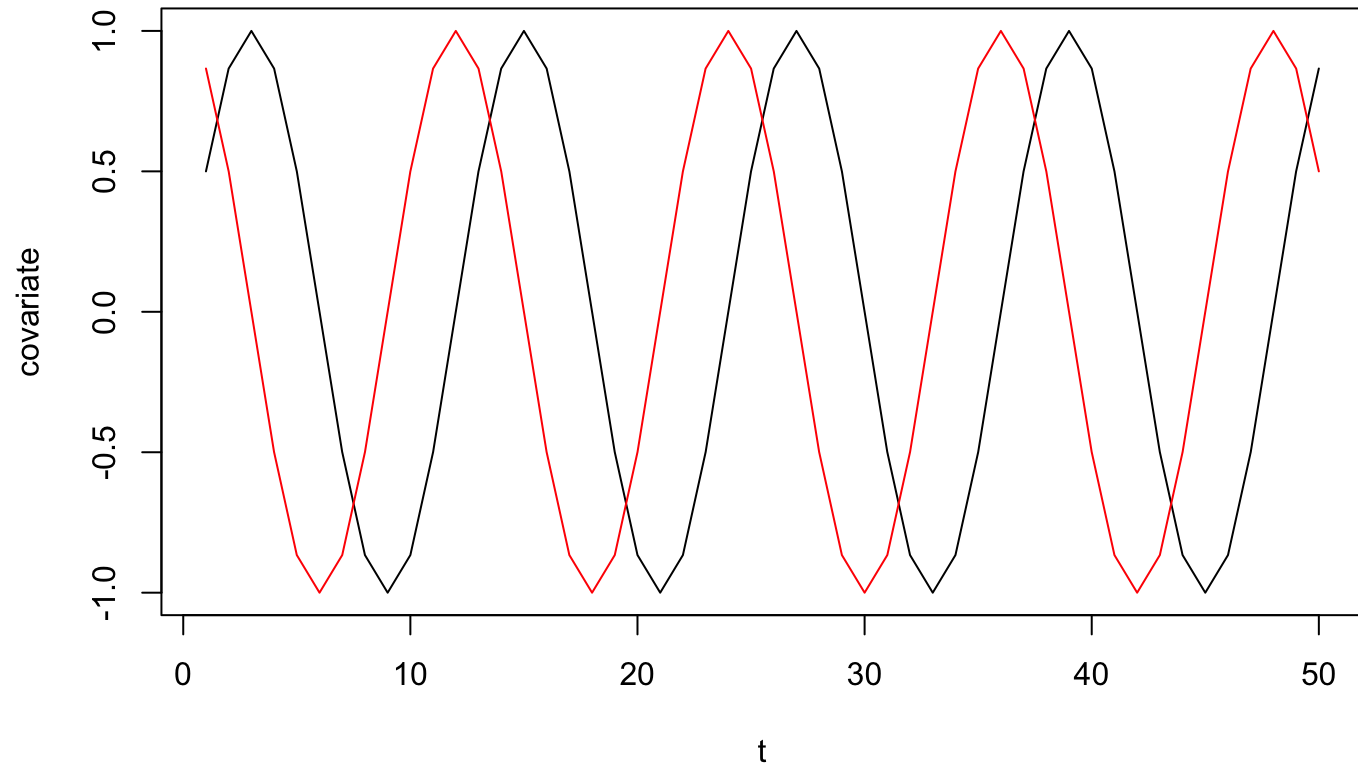
Introduce 2 covariates: $\sin(2\pi t/p)$, $\cos(2\pi t/p)$ where p is period (12 for monthly) and t is the time step (1 to T).

$$x_t = x_{t-1} + u + \mathbf{C}\mathbf{c}_t + w_t$$

$$\mathbf{C}\mathbf{c}_t = \beta_1 \sin(2\pi t/p) + \beta_2 \cos(2\pi t/p)$$

where p is 12 (for monthly).

covariates sines and cosines



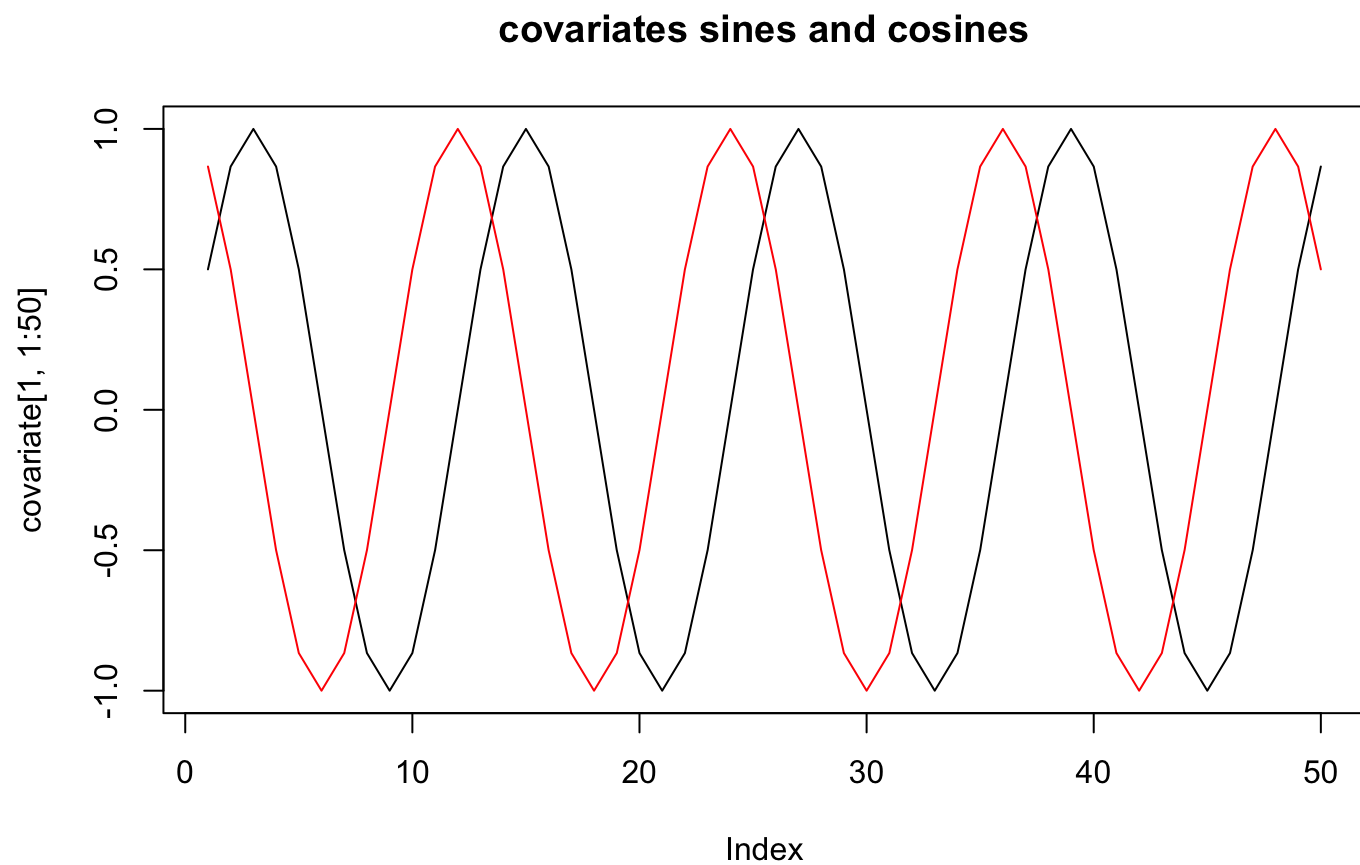
Season as a Fourier series

covariates

a 2 by T matrix

$$\begin{array}{l} \text{WA} \\ \text{OR} \end{array} \begin{bmatrix} C_1 & C_2 \\ C_1 & C_2 \end{bmatrix} \begin{bmatrix} \sin\left(\frac{2\pi 1}{12}\right) & \sin\left(\frac{2\pi 2}{12}\right) & \dots & \sin\left(\frac{2\pi T}{12}\right) \\ \cos\left(\frac{2\pi 1}{12}\right) & \cos\left(\frac{2\pi 2}{12}\right) & \dots & \cos\left(\frac{2\pi T}{12}\right) \end{bmatrix}$$

```
TT <- nrow(chinook.month)/2
covariate <- rbind(sin(2*pi*(1:TT)/12), cos(2*pi*(1:TT)/12))
plot(covariate[1,1:50], type="l")
lines(covariate[2,1:50], col="red")
title("covariates sines and cosines")
```



WA and OR have different seasonal pattern.

```
C <- matrix(paste0(c("s", "c"), "."), rep(1:2, each=2)), 2, 2, byrow = TRUE)
C
```

```
##      [,1] [,2]
## [1,] "s.1" "c.1"
## [2,] "s.2" "c.2"
```

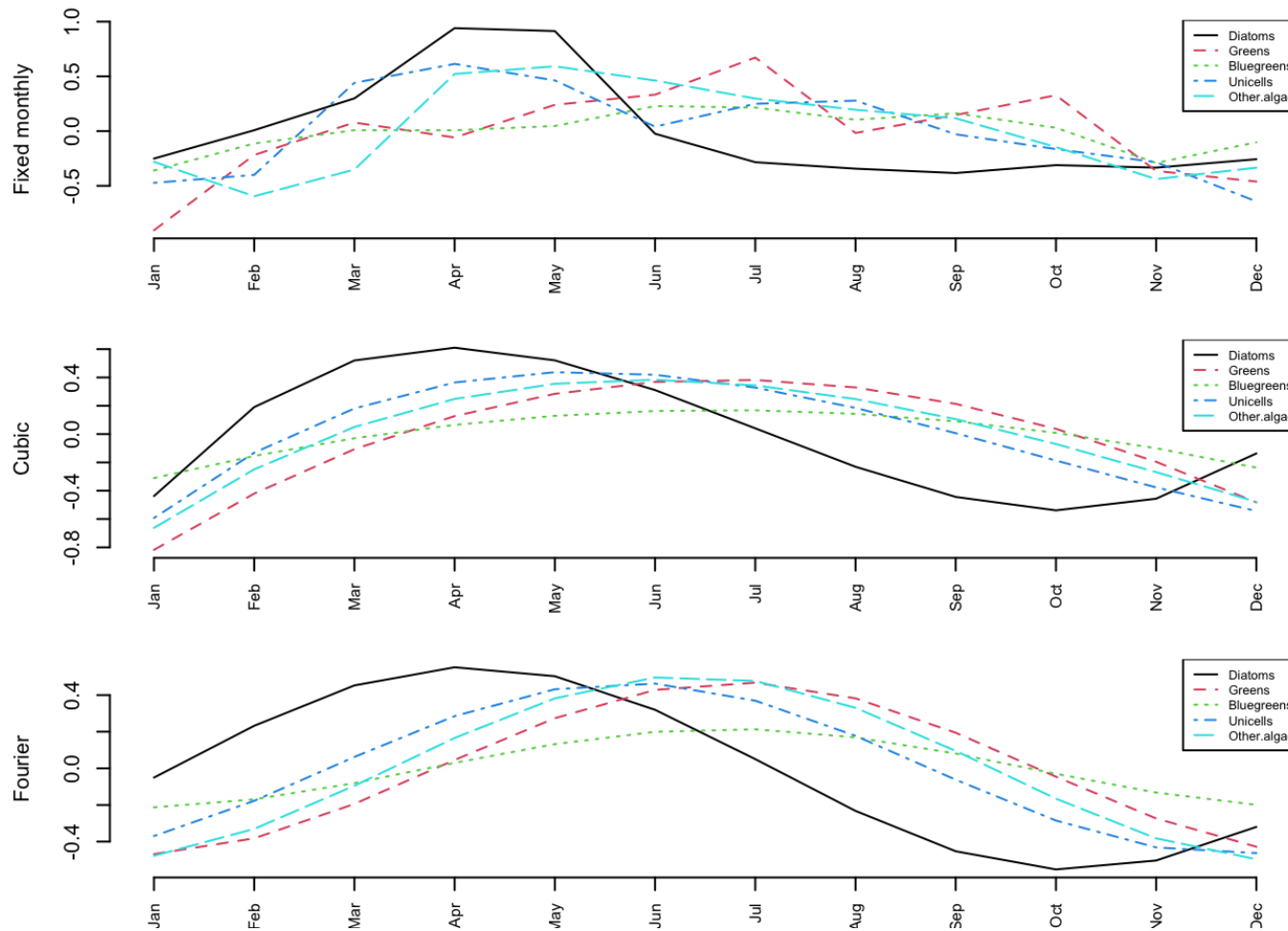
WA and OR have same seasonal pattern.

```
C <- matrix(c("s", "c"), 2, 2, byrow = TRUE)
C
```

```
##      [,1] [,2]
## [1,] "s"  "c"
## [2,] "s"  "c"
```

Example Section 8.6

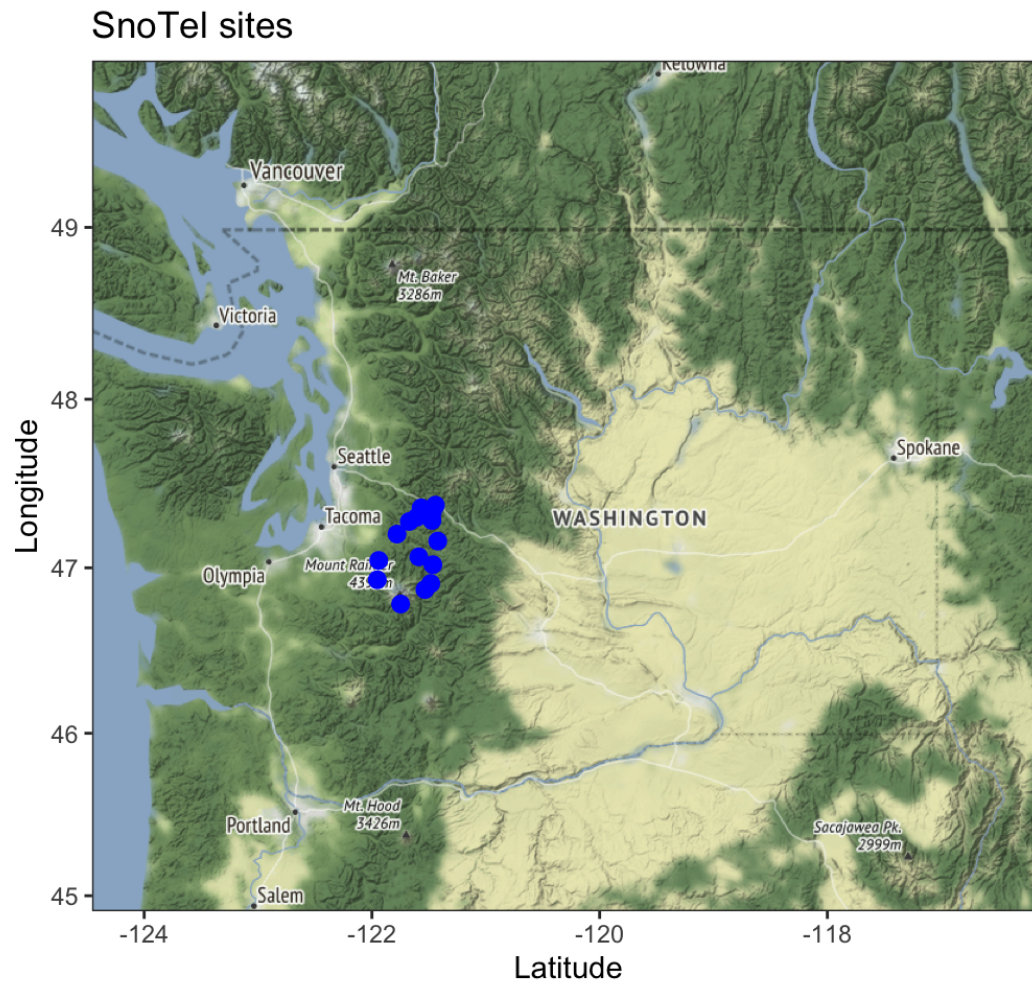
Seasonality of Lake WA plankton



Missing covariates

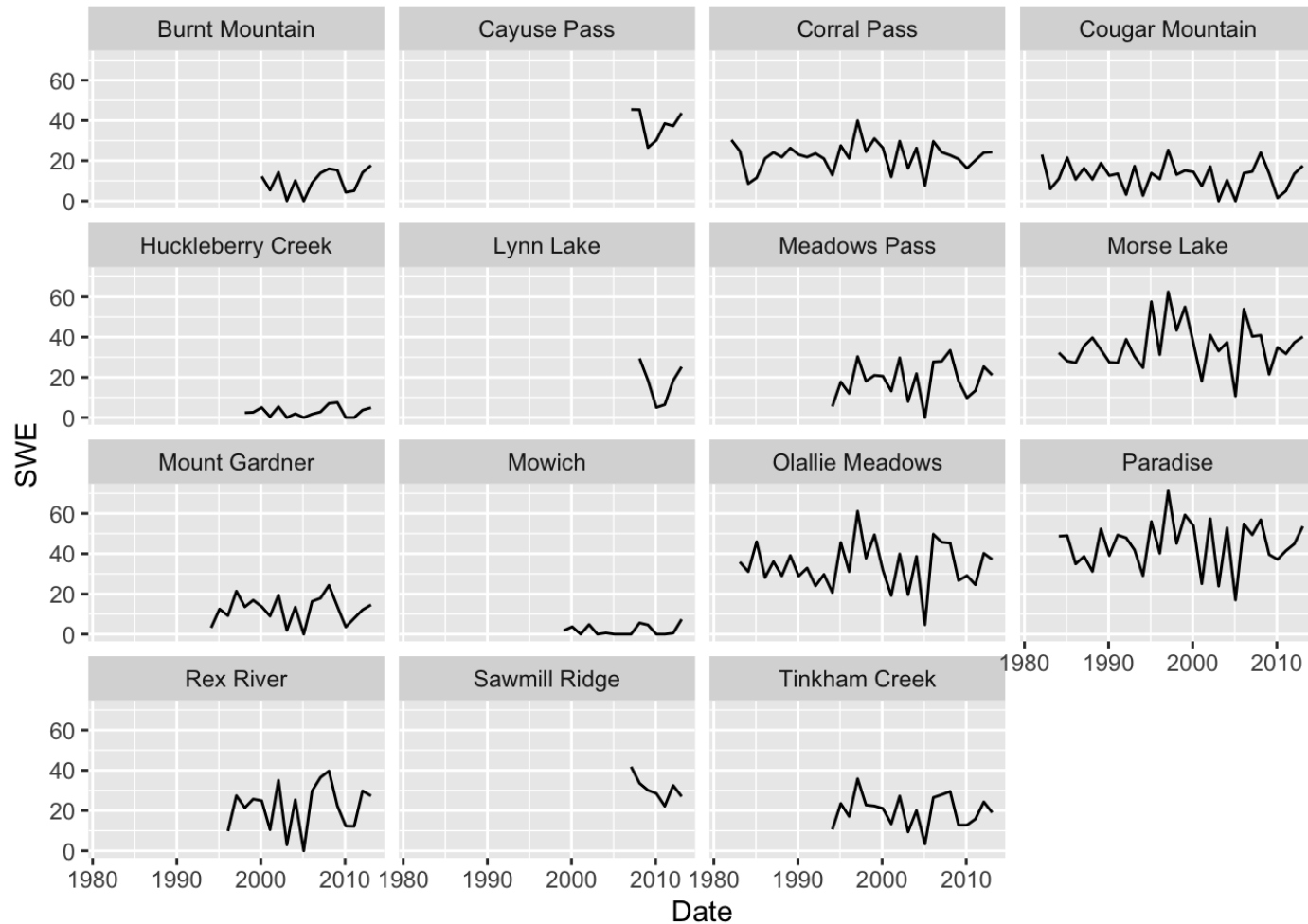
SNOTEL Example Chapter 11

<https://nwfsc-timeseries.github.io/atsa-labs/example-snotel-data.html>



Snow Water Equivalent (snowpack)

February snowpack estimates



Use MARSS models Chapter 11

- Accounts for correlation across sites and local variability

