Why include covariates in a MARSS model?

- You want to explain correlation in observation errors across sites or auto-correlation in time

Auto-correlated observation errors

Model your v(t) as a AR-1 process
Difficult numerically

Or if know what is causing the auto-correlation, include that as a covariate.

Correlated observation errors across sites (y rows)

Use a R matrix with off-diagonal terms
Difficult numerically

Or if know what is causing the correlation, include that as a covariate

Types of covariates

- Numerical
  - Continuous (eg, temperature, salinity)
  - Discrete (eg, counts)
- Categorical
  - Before/After
  - North/South
  - January, February, March, …

Covariates occur in state, obs or both

State equation

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u} + \boxed{\mathbf{C}\mathbf{c}_t} + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathrm{MVN}(0, \mathbf{Q})$$

(eg, nutrients affects growth, high temps kill)

Observation equation

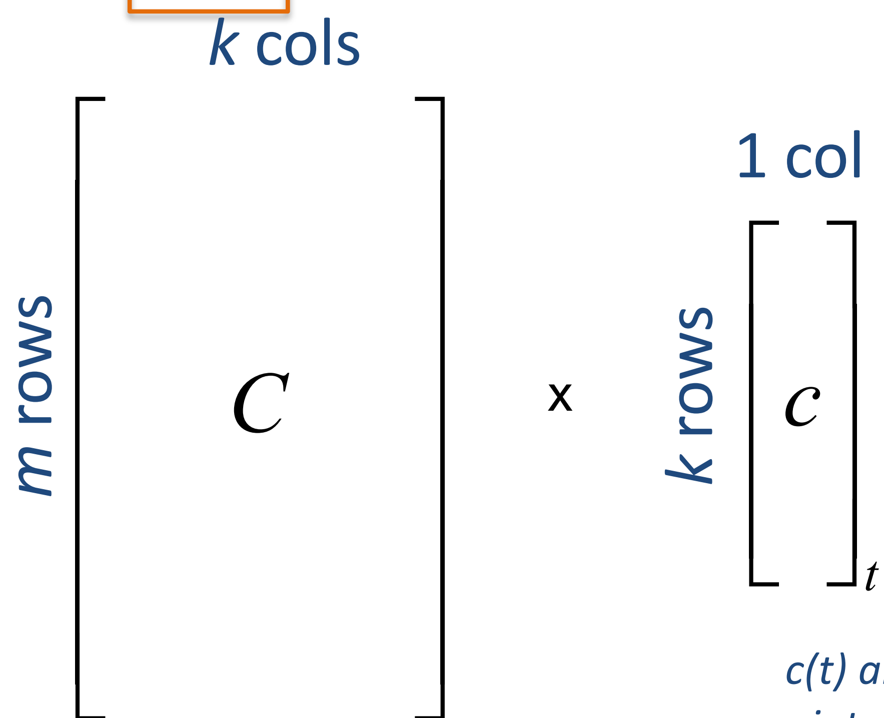
$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \boxed{\mathbf{D}\mathbf{d}_t} + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathrm{MVN}(0, \mathbf{R})$$

(eg, vegetation obscures individuals,
temperature affect behavior making animals visible)

Covariates occur in state, obs or both

State equation

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q})$$



C is the effect of cov on state

c(t) are the covariates at time t

m is number of states; *k* is number of covariates

Covariates occur in state, obs or both

Observation equation

$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{D}\mathbf{d}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{R})$$

$\mathbf{D} = \big[\ \big]$ — *k* cols, *n* rows

$\mathbf{d}_t = \big[\ \big]_t$ — 1 col, *k* rows

D is the effect of cov on state

d(t) are the covariates at time t

n is number of obs; *k* is number of covariates

Covariate effects can differ or not

Different effects

$$\mathbf{C} = \begin{bmatrix} C_{P,1} & C_{N,1} \\ C_{P,2} & C_{N,2} \\ C_{P,3} & C_{N,3} \end{bmatrix}$$

Same effect

$$\mathbf{C} = \begin{bmatrix} C_{P} & C_{N} \\ C_{P} & C_{N} \\ C_{P} & C_{N} \end{bmatrix}$$

$$\mathbf{c}_t = \begin{bmatrix} Precipitation \\ Nitrogen \end{bmatrix}_t$$

Covariates can be seasons or periods

State equation

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{Q})$$

Observation equation

$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{D}\mathbf{d}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{R})$$

Seasonal or periodical effects

For example, effects of “season” on 3 states (3 rows)

Winter Spring Summer Autumn | Winter Spring Summer Autumn ... Winter Spring Summer Autumn

$$
\begin{bmatrix}
C_1 & C_2 & C_3 & C_4 \\
C_1 & C_2 & C_3 & C_4 \\
C_1 & C_2 & C_3 & C_4
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & \ldots & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & \ldots & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 & 1
\end{bmatrix}
$$

t = 1 2 3 4 .....

Seasonal or periodical effects

For example, effects of "season" on 3 states

$$\begin{bmatrix} C_1 & C_2 & C_3 & C_4 \\ C_1 & C_2 & C_3 & C_4 \\ C_1 & C_2 & C_3 & C_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 & 1 \end{bmatrix}$$

$\mathbf{c}_t$

Seasonal or periodical effects

For example, effects of "season" on 3 states

$$
\begin{bmatrix}
C_1 & C_2 & C_3 & C_4 \\
C_1 & C_2 & C_3 & C_4 \\
C_1 & C_2 & C_3 & C_4
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & \dots & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & \dots & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 & 1
\end{bmatrix}
$$

$\mathbf{c}_{t+1}$

Non-factor seasons or periods

Treating season as a factor means we have a parameter for each 'season'. 4 in the previous example. What if the factor were 'month'? Then we'd have 12 parameters!

- We can also estimate "season" via a nonlinear model
- Two common options:
  1) Cubic polynomial
  2) Fourier frequency

Season as a polynomial

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q})$$

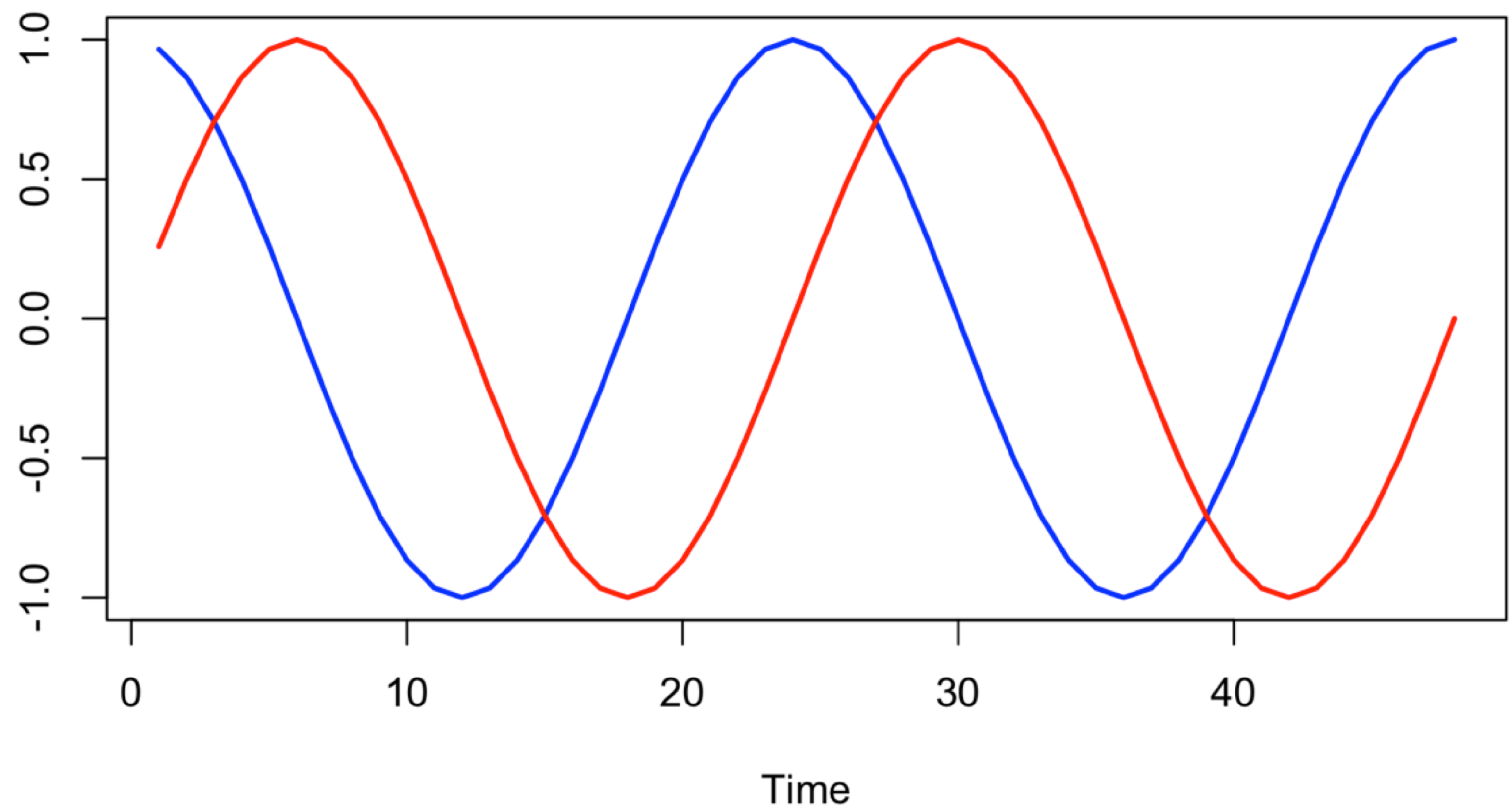
For months: $\mathbf{C}\mathbf{c}_t = b_1 m_t + b_2 m_t^2 + b_3 m_t^3$

$$\begin{bmatrix} C_1 & C_2 & C_3 \\ C_1 & C_2 & C_3 \\ C_1 & C_2 & C_3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & \dots & 12 \\ 1 & 4 & 9 & \dots & 144 \\ 1 & 8 & 27 & \dots & 1728 \end{bmatrix} \begin{matrix} m \\ m^2 \\ m^3 \end{matrix}$$

t = 1 2 3 .....

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 (which we will not cover here)



Season as a Fourier series

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_t + \mathbf{u} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{Q})$$

Our new covariates at time t

$$\mathbf{C}\mathbf{c}_t = C_1 \sin(2\pi t/p) + C_2 \cos(2\pi t/p)$$

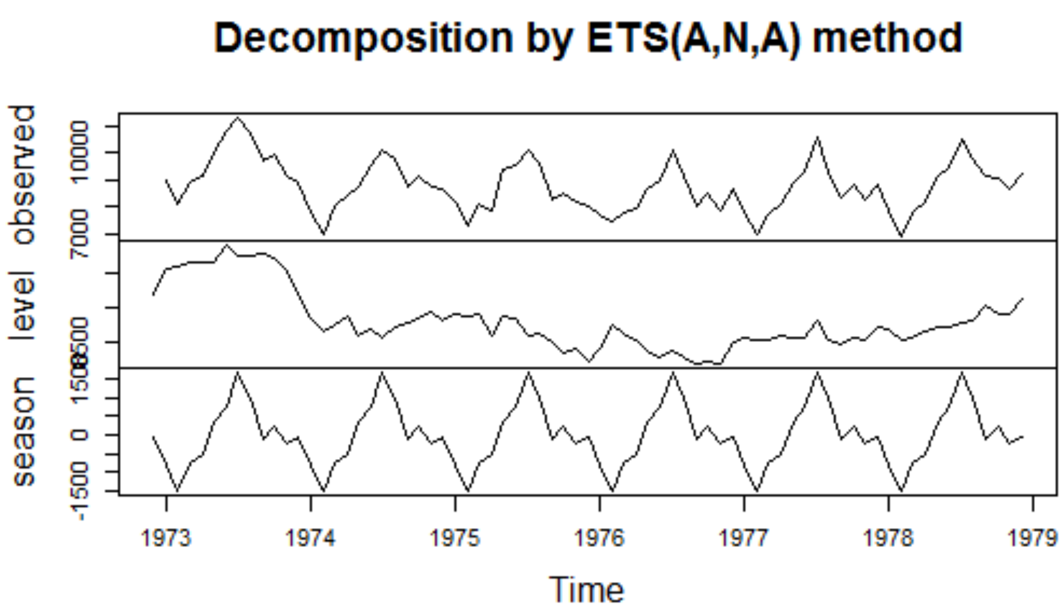
$$\begin{bmatrix} C_1 & C_2 \\ C_1 & C_2 \\ C_1 & C_2 \end{bmatrix} \begin{bmatrix} \sin\left(\frac{2\pi t}{p}\right) \\ \cos\left(\frac{2\pi t}{p}\right) \end{bmatrix}_t$$

t is time step (1, 2, 3, ... , number of data points)

p is period (e.g., 12 months per year so p=12)

Feb 7th Forecasting with Exponential Smoothing Models

- We'll talk about modeling time-varying seasonal effects at that time.
- Exponential smoothing models are related to Dynamic Linear Models, which Mark will cover in Week 5



Dealing with missing covariates

- Drop years / shorten time series to remove missing values
- Interpolate missing values
- Develop process model for the covariates
  - Allows us to incorporate observation error into the covariates (known or unknown)
  - Allows us to interpolate but NOT treat that interpolated value as known. It is an estimated value that has uncertainty.

Dealing with missing covariates

$$\begin{bmatrix}\mathbf{x}^{(v)}\\ \mathbf{x}^{(c)}\end{bmatrix}_t = \begin{bmatrix}\mathbf{B}^{(v)} & \mathbf{C}\\ 0 & \mathbf{B}^{(c)}\end{bmatrix}\begin{bmatrix}\mathbf{x}^{(v)}\\ \mathbf{x}^{(c)}\end{bmatrix}_{t-1} + \begin{bmatrix}\mathbf{u}^{(v)}\\ \mathbf{u}^{(c)}\end{bmatrix} + \mathbf{w}_t,$$

$$\mathbf{w}_t \sim \mathrm{MVN}\left(0, \begin{bmatrix}\mathbf{Q}^{(v)} & 0\\ 0 & \mathbf{Q}^{(c)}\end{bmatrix}\right)$$

$$\begin{bmatrix}\mathbf{y}^{(v)}\\ \mathbf{y}^{(c)}\end{bmatrix}_t = \begin{bmatrix}\mathbf{Z}^{(v)} & \mathbf{D}\\ 0 & \mathbf{Z}^{(c)}\end{bmatrix}\begin{bmatrix}\mathbf{x}^{(v)}\\ \mathbf{x}^{(c)}\end{bmatrix}_{t} + \begin{bmatrix}\mathbf{a}^{(v)}\\ \mathbf{a}^{(c)}\end{bmatrix} + \mathbf{v}_t,$$

$$\mathbf{v}_t \sim \mathrm{MVN}\left(0, \begin{bmatrix}\mathbf{R}^{(v)} & 0\\ 0 & \mathbf{R}^{(c)}\end{bmatrix}\right)$$

(v) are the variates (data)

(c) are the covariates

Dealing with missing covariates

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{u} + \mathbf{w}_t, \text{ where } \mathbf{w}_t \sim \text{MVN}(0, \mathbf{Q})$$
$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{v}_t, \text{ where } \mathbf{v}_t \sim \text{MVN}(0, \mathbf{R})$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(v)} \\ \mathbf{x}^{(c)} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}^{(v)} & \mathbf{C} \\ 0 & \mathbf{B}^{(c)} \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}^{(v)} \\ \mathbf{u}^{(c)} \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} \mathbf{Q}^{(v)} & 0 \\ 0 & \mathbf{Q}^{(c)} \end{bmatrix}$$
$$\mathbf{y} = \begin{bmatrix} \mathbf{y}^{(v)} \\ \mathbf{y}^{(c)} \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}^{(v)} & \mathbf{D} \\ 0 & \mathbf{Z}^{(c)} \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} \mathbf{a}^{(v)} \\ \mathbf{a}^{(c)} \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} \mathbf{R}^{(v)} & 0 \\ 0 & \mathbf{R}^{(c)} \end{bmatrix}$$

See Holmes, Ward and Scheuerell (2018) "MARSS User Guide" for a discussion and example of how to do this.

Example: You measure temperature in 2 locations with one location having 2 different sensors. You want a composite of those. You have missing values.

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

$$\begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ a_2 \\ a_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$R = \begin{bmatrix} b & 0 & 0 \\ 0 & \sigma & c \\ 0 & c & \varphi \end{bmatrix}$$

Dealing with missing covariates

- You may want to fit the covariate model separately so that fit to the process is not driven by the fit to the covariates.
- Then fix the parameters of the covariate part of the model when fitting.
- Why not use the estimated state as your 'covariate'?
  - You can, but then uncertainty in the covariate doesn't propagate into your fit uncertainty.

Topics for the computer lab

- Fitting multivariate state-space models with covariates
  - Seasonal effects
- Fitting candidate model sets