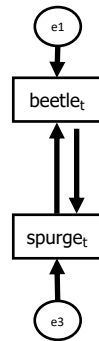




Wibbly Wobbly Timey-wimey SEM

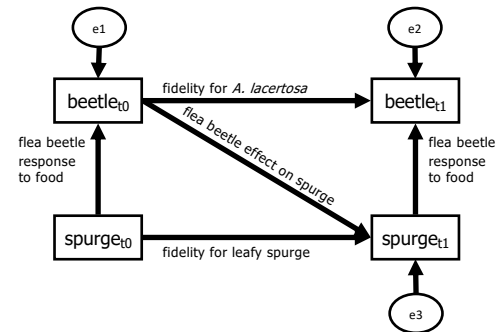
1. Killing non-recursivity: panel models
2. Autocorrelation
3. More exotic structures you already know

Problem of Non-Recursive Models

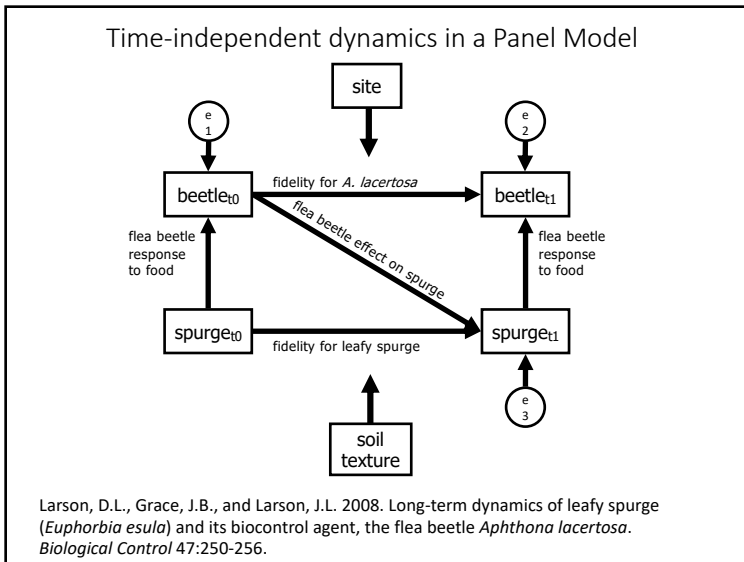
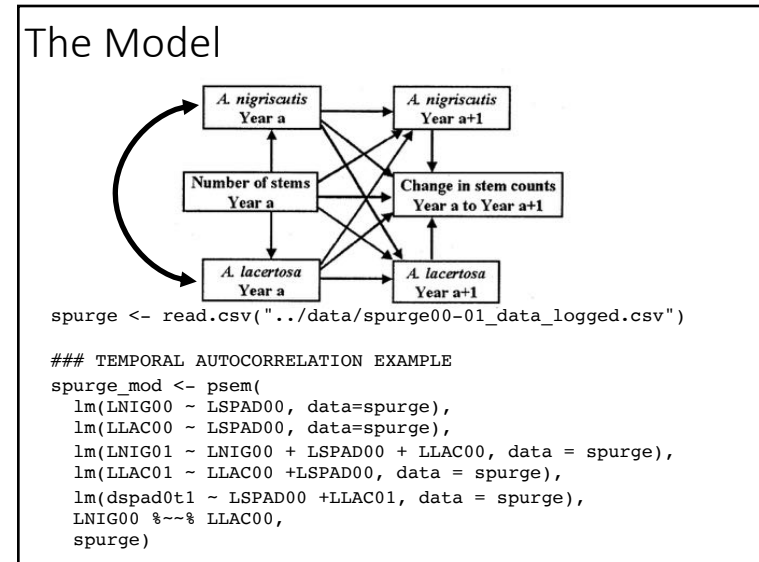
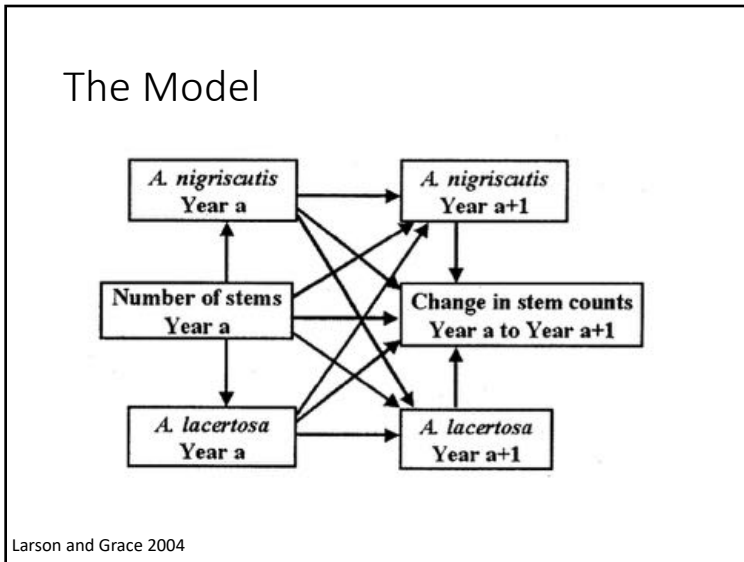


Larson, DL and Grace, JB (2004) Temporal Dynamics of Leafy Spurge (*Euphorbia esula*) and Two Species of Flea Beetles (*Aphthona* spp.) Used as Biological Control Agents. *Biological Control* 29:207–214.

Longitudinal Studies – Time-Step (Panel) Model for Lagged Effects



Larson, DL and Grace, JB (2004)



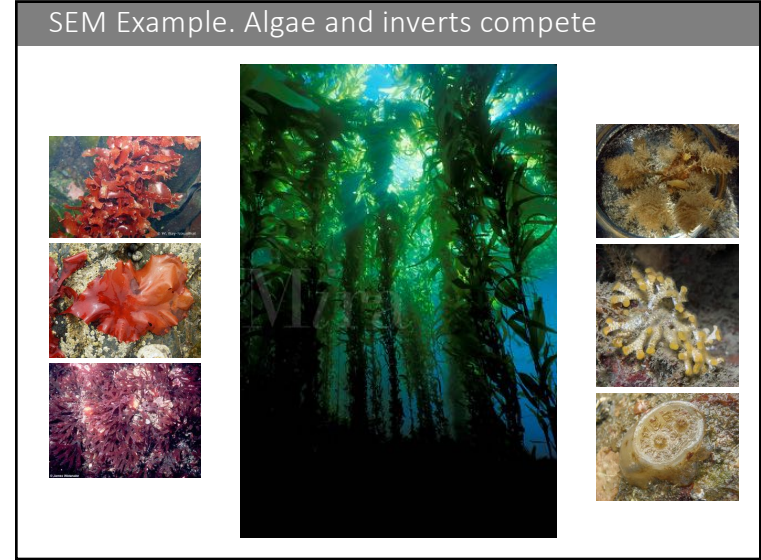
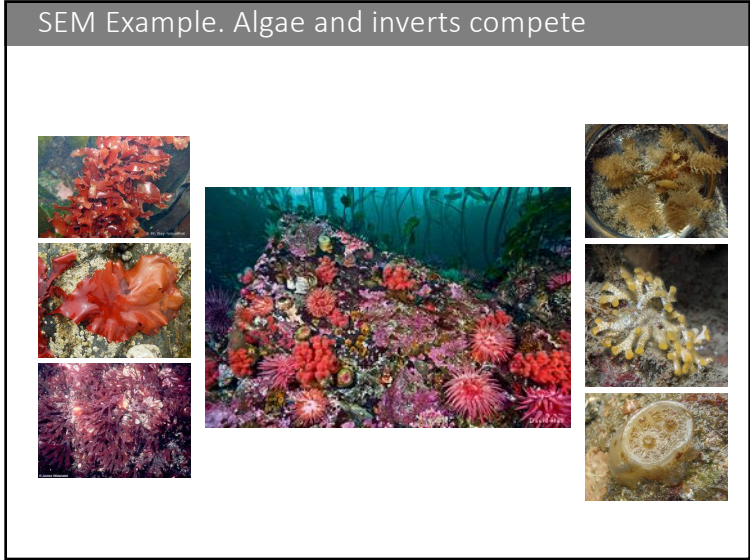
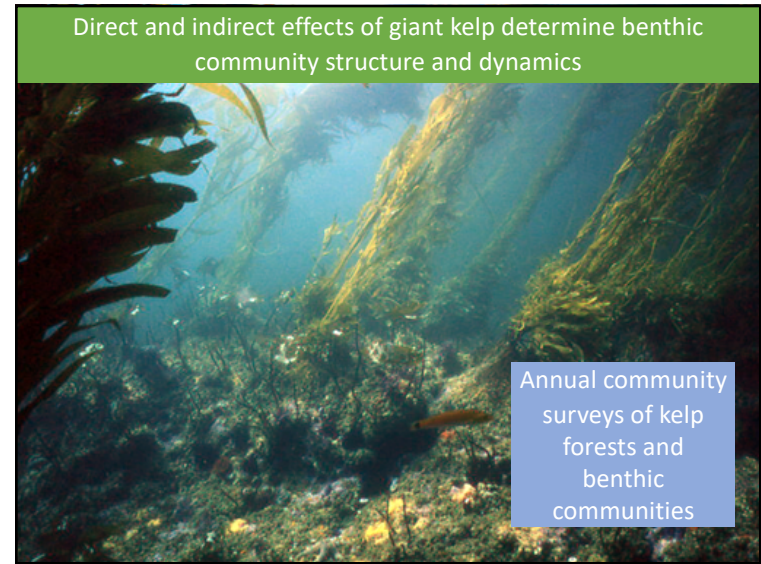
- ### Wibbly Wobbly Timey-wimey SEM
1. Killing non-recursivity: panel models
 2. Autocorrelation
 3. More exotic structures you already know

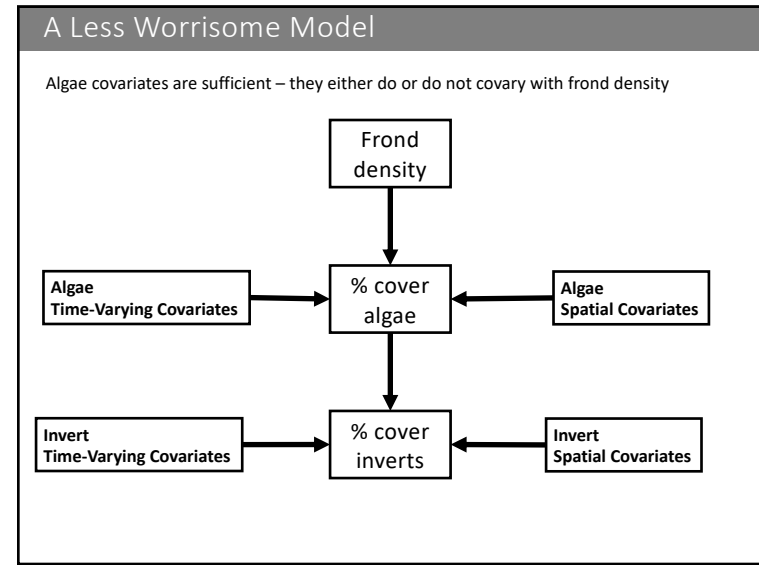
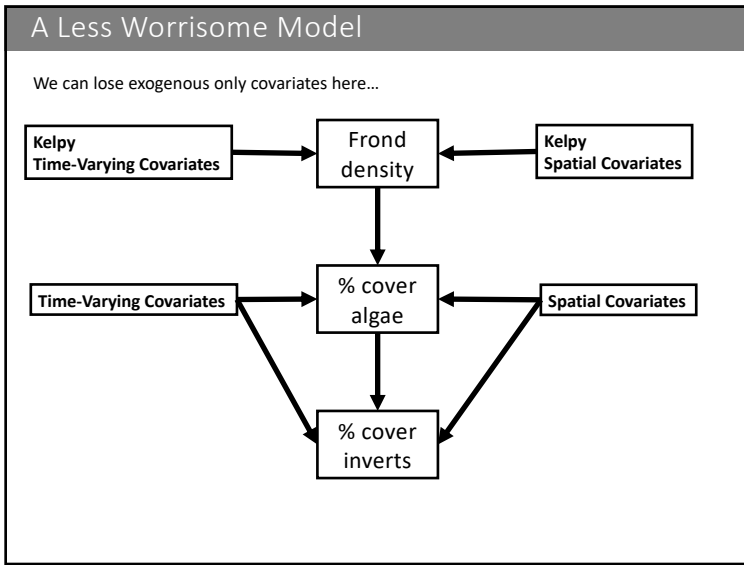
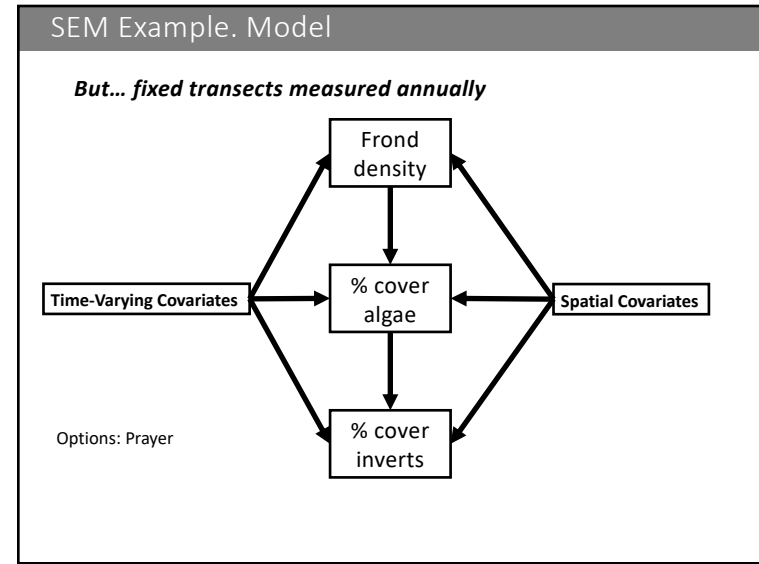
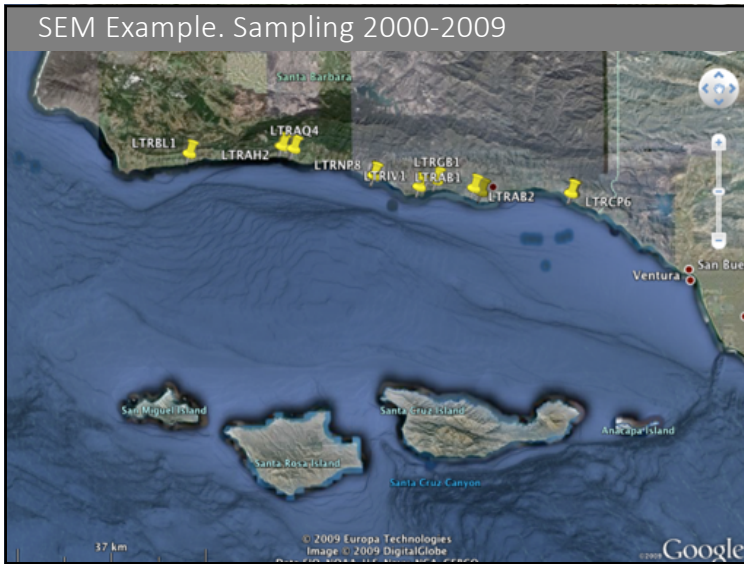
Why not to think about autocorrelation

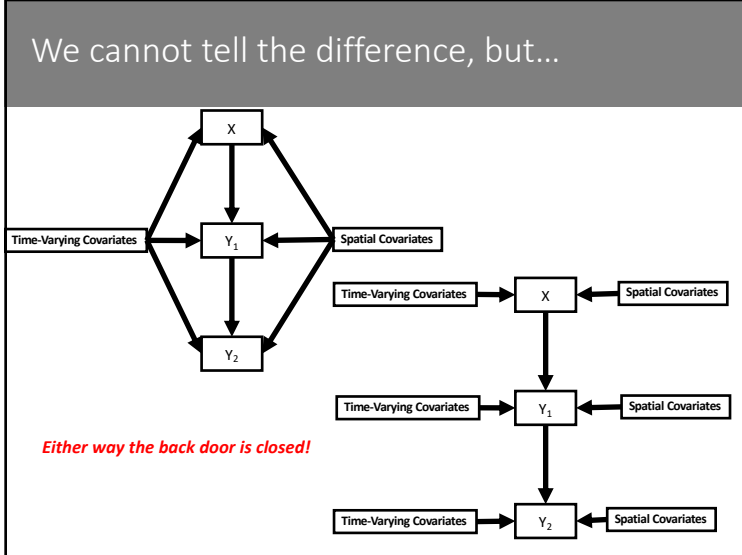
There are two key issues regarding space:

- (1) Are there things to learn about the other factors that could explain variations in the data that vary spatially?
- (2) Do we have nonindependence in our residuals?

Recent reference on the subject:
Hawkins, BA (2011) Eight (and a half) deadly sins of spatial analysis. *Journal of Biogeography*. doi:10.1111/j.1365-2699.2011.02637.x

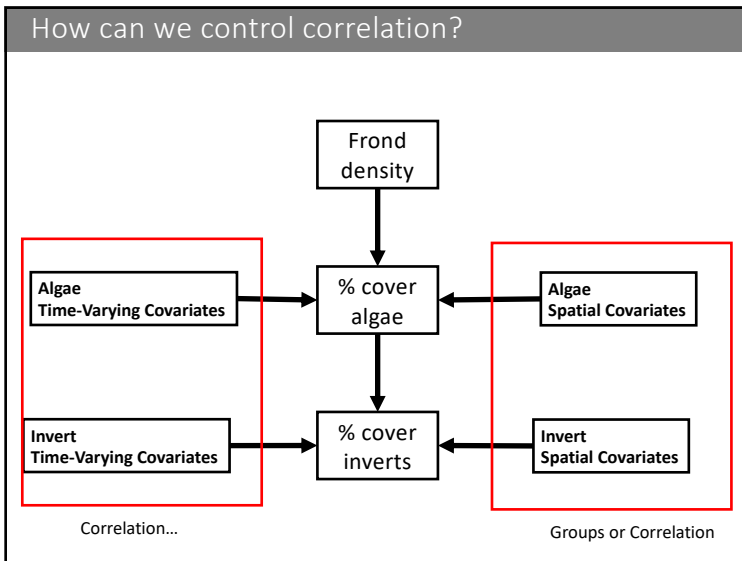






When should I worry?

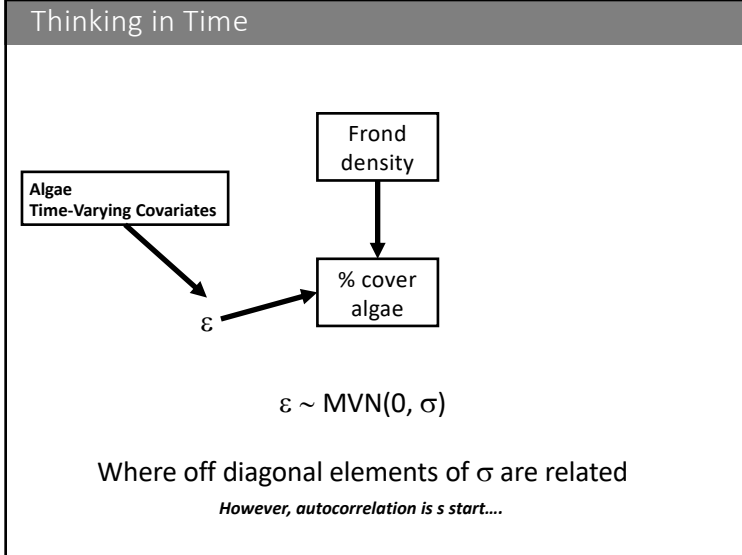
- If covariates enter by the back door, you should measure them?
 - Autocorrelation as a "cure"?
- If covariates are unique to variable, you can 'leave-out' as 'noise'
 - BUT – will affect accuracy of estimates
 - Random effects or autocorrelation needed



Thinking in Time

The diagram shows Frond density leading to % cover algae. Algae Time-Varying Covariates influence an error term ε, which also influences % cover algae.

If time varying covariates have lagged effects,
 $y = BX + \epsilon$ with ϵ being iid is no longer correct
Autocorrelation alone does not shut the back door
Include actual covariates as predictors where possible

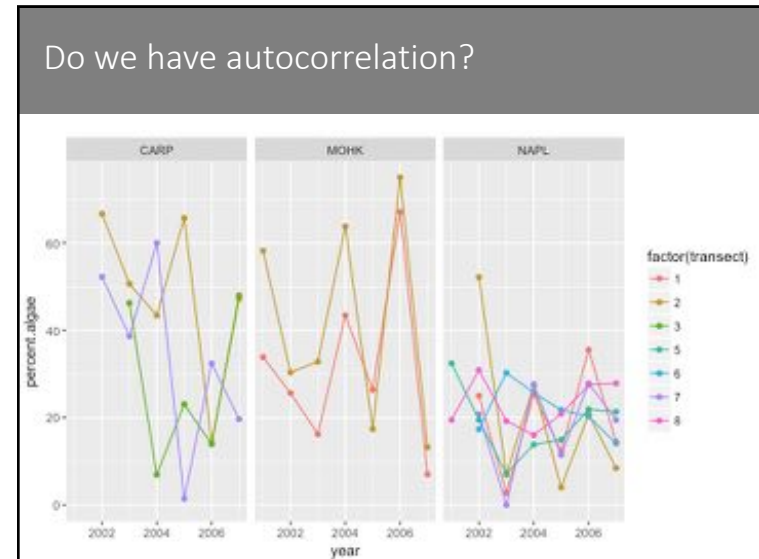


Modeling Autocorrelation. Structures

Correlations among sampling points follow a predetermined pattern

Compound Symmetry	CS	$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$
Unstructured	UN	$\begin{bmatrix} \sigma_{11}^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \end{bmatrix}$
First-Order Autoregressive	AR(1)	$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$

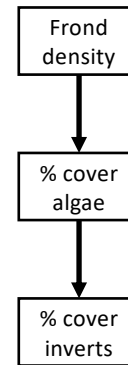
- ### Why do this? What if we did not incorporate autocorrelation?
- Estimates of coefficients will be unbiased
 - Although using temporal autocorrelation can yield different answers
 - BUT – standard errors will be biased
 - Usually too narrow



Modeling Autocorrelation. *nlme****nlme***: Linear and Nonlinear Mixed Effects Models

```
install.packages("nlme")
library(nlme)
```

SEM Example. Model

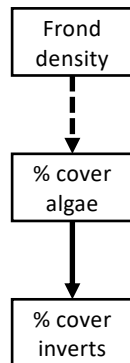


```
#Model
algae_mod <- lm(percent.algae ~ frond.density,
  data = arkema)

invert_mod <- lm(percent.inverts ~ percent.algae,
  data = arkema)

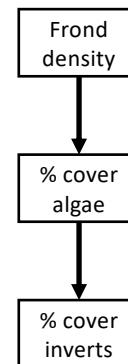
arkema.sem <- psem(
  algae_mod,
  invert_mod,
  arkema
)
```

SEM Example. Model



Response	Predictor	Estimate	Std.Error	DF	Crit.Value	P.Value	Std.Estimate
percent.algae	frond.density	-0.6900	0.4370	67	-1.5789	0.1191	-0.1894
percent.inverts	percent.algae	-0.8429	0.0803	67	-10.4903	0.0000	-0.7884 ***

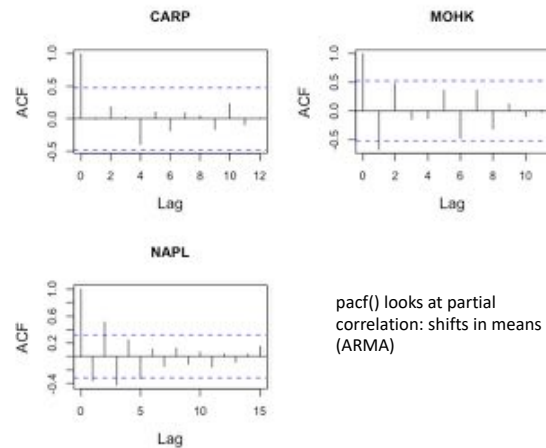
Look at autocorrelation



```
#Look at residuals by site
res_df <- data.frame(site_trans=arkema$site_trans,
  res = residuals(algae_mod))

par(mfrow=c(2,2))
for(asite in levels(res_df$site_trans)){
  subdata <- subset(res_df, res_df$site_trans==asite)
  acf(subdata$res, cex.lab=1.3, main=asite)
}
par(mfrow=c(1,1))
```

ACF Function Looks at Lag Correlation

Modeling Autocorrelation. *nlme***?corClasses**

...

corAR1

autoregressive process of order 1.

corARMA

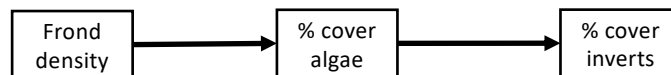
autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.

corCAR1

continuous autoregressive process (AR(1) process for a continuous time covariate).

...

Look at autocorrelation

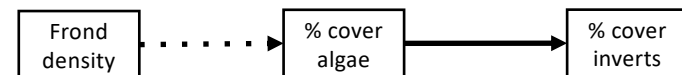


```

algae_mod_ac <- gls(percent.algae ~ frond.density, data = arkema,
  correlation = corAR1(form = ~ year | site_trans))

invert_mod_ac <- gls(percent.inverts ~ percent.algae, data = arkema,
  correlation = corAR1(form = ~ year | site_trans))
  
```

Look at autocorrelation



	Response	Predictor	Estimate	Std. Error	DF	Crit. Value	P. Value	Std. Estimate
1	percent.algae	frond.density	-0.7505	0.4412	69	-1.7011	0.0936	-0.2060
2	percent.inverts	percent.algae	-0.8159	0.0596	69	-13.6922	0.0000	-0.7631 ***

SEM Example. Your turn...

```

    graph LR
      A[Fron density] --> B[% cover algae]
      B --> C[% cover inverts]
  
```

Re-fit the model addressing the potential temporal autocorrelation with an AR1 Process and the spatial autocorrelation using a random effect of site/transect

Use lme

Use `arkema <- read.csv("../Data/arkema.csv")`

SEM Example. Your turn...

```

    graph LR
      A[Fron density] --> B[% cover algae]
      B --> C[% cover inverts]
  
```

```

arkema.sem3 <- psem(
  lme(percent.algae ~ frond.density, random = ~ 1 | site/transect,
    correlation = corCAR1(form = ~ year | site/transect),
    data = arkema),

  lme(percent.inverts ~ percent.algae, random = ~ 1 | site/transect,
    correlation = corCAR1(form = ~ year | site/transect),
    data = arkema),

  data = arkema
)
  
```

SEM Example. Your turn...

```

    graph LR
      A[Fron density] -- -0.36 --> B[% cover algae]
      B -- -0.74 --> C[% cover inverts]
  
```

	Response	Predictor	Estimate	Std.Error	DF	Crit.Value	P.Value	Std.Estimate
1	percent.algae	frond.density	-1.3262	0.4024	57	-3.2960	0.0017	-0.3640 **
2	percent.inverts	percent.algae	-0.7958	0.0621	57	-12.8101	0.0000	-0.7444 ***

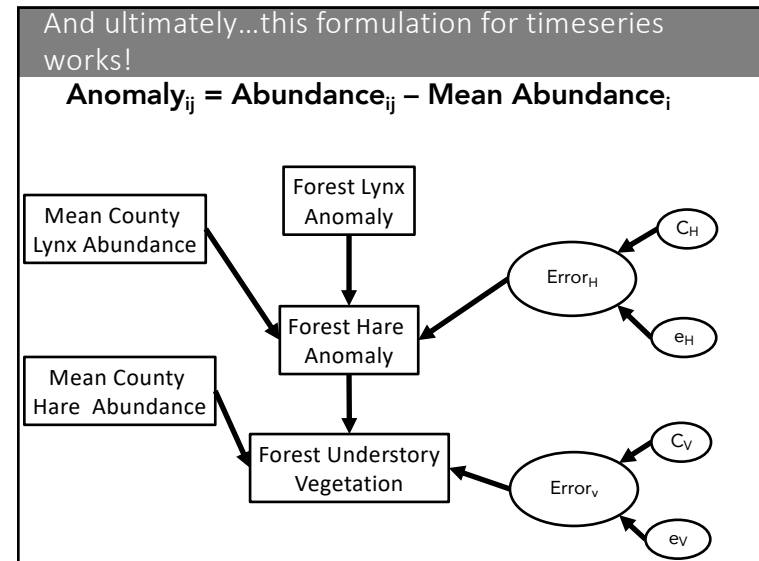
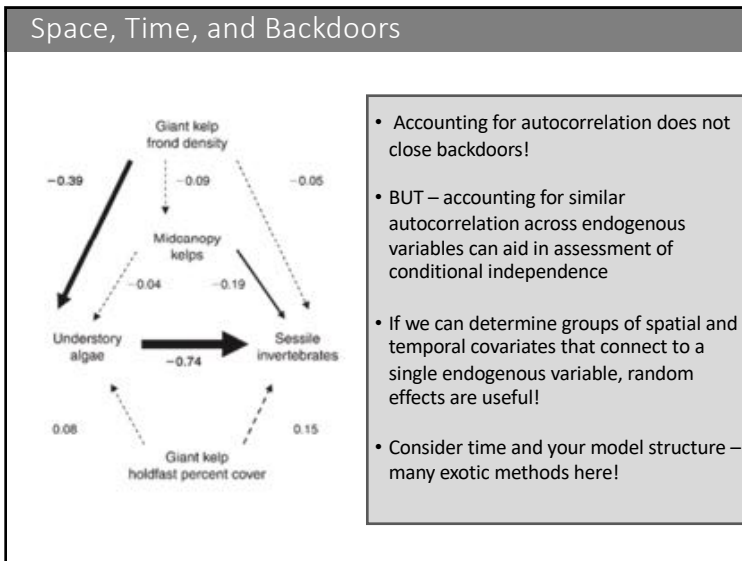
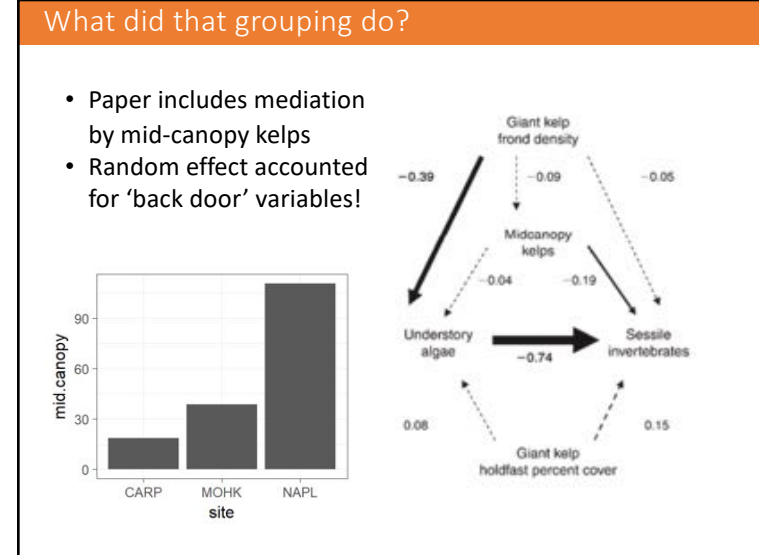
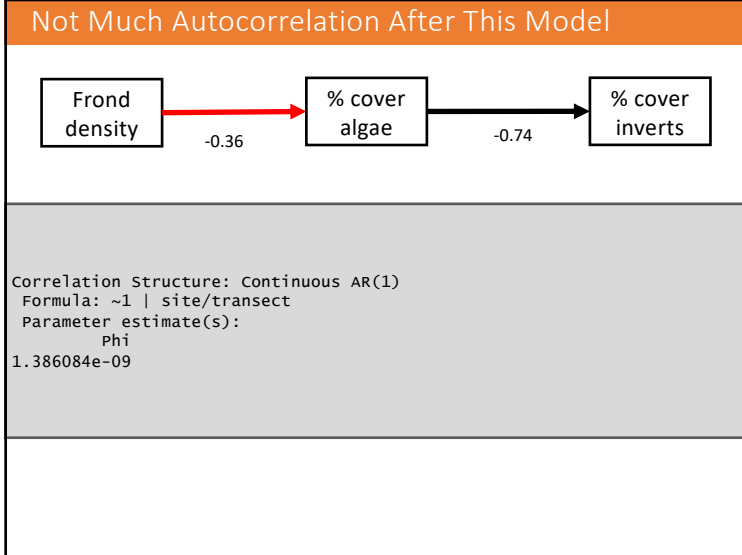
Algae Model

```

    graph LR
      A[Fron density] -- -0.36 --> B[% cover algae]
      B -- -0.74 --> C[% cover inverts]
  
```

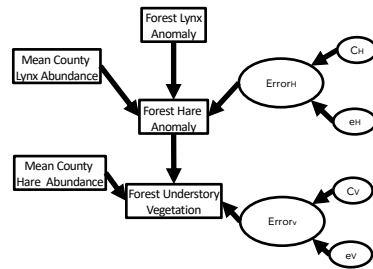
Random effects:
 Formula: ~1 | site
 (Intercept)
 StdDev: 11.16566

Formula: ~1 | transect %in% site
 (Intercept) Residual
 StdDev: 4.527033 14.46497



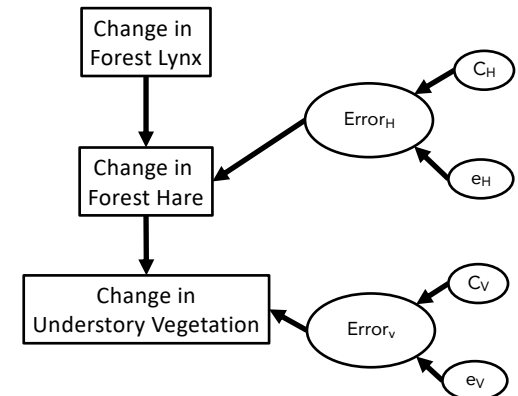
Causal Inference from Longitudinal Panel Data

$$\text{Anomaly}_{ij} = \text{Abundance}_{ij} - \text{Mean Abundance}_i$$



- This formulation DOES shut the back door
- See also fixed effect model
- See also Mundlak devices
- And change models...

Change Models Shut Back Doors



Why not to think about autocorrelation

There are two key issues regarding space:

- (1) Are there things to learn about the other factors that could explain variations in the data that vary spatially?
- (2) Do we have nonindependence in our residuals?

Recent reference on the subject:

Hawkins, BA (2011) Eight (and a half) deadly sins of spatial analysis. Journal of Biogeography. doi:10.1111/j.1365-2699.2011.02637.x

Wibbly Wobbly Timey-wimey SEM

1. Killing non-recursivity: panel models
2. Autocorrelation
3. More exotic structures you already know

