6. Autocorrelation

Overview

1. Independence
2. Modeling Autocorrelation
3. Temporal Autocorrelation Example
4. Spatial Autocorrelation Example

6.1 Independence

6.1 Independence. Model assumptions

- All linear models (including SEM) assume that errors are *independent*, i.e., uncorrelated

- Correlation can arise from:
  - Poor experimental design (Hurlbert 1984)
  - Natural phenomenon (e.g., temp & latitude, etc.)
  - Subsampling the same replicate
  - Spatial non-independence (sites closer together in space are more alike)
  - Temporal non-independence (periods closer in time are more alike, repeatedly sampling the same replicate)

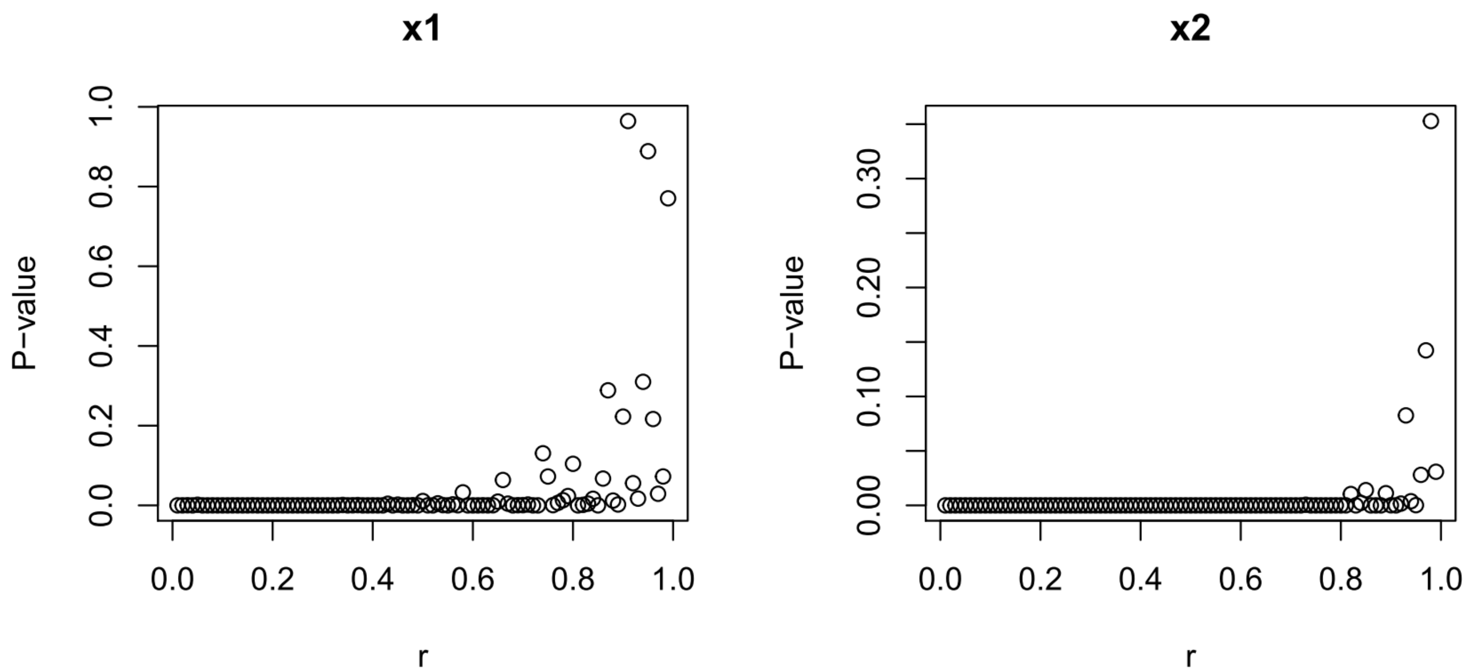
6.1 Independence. Two kinds of correlations

1. Correlations among predictors

2. Correlations among observations

6.1 Independence. Collinearity

- Correlated predictors inflate standard errors & influence significance tests

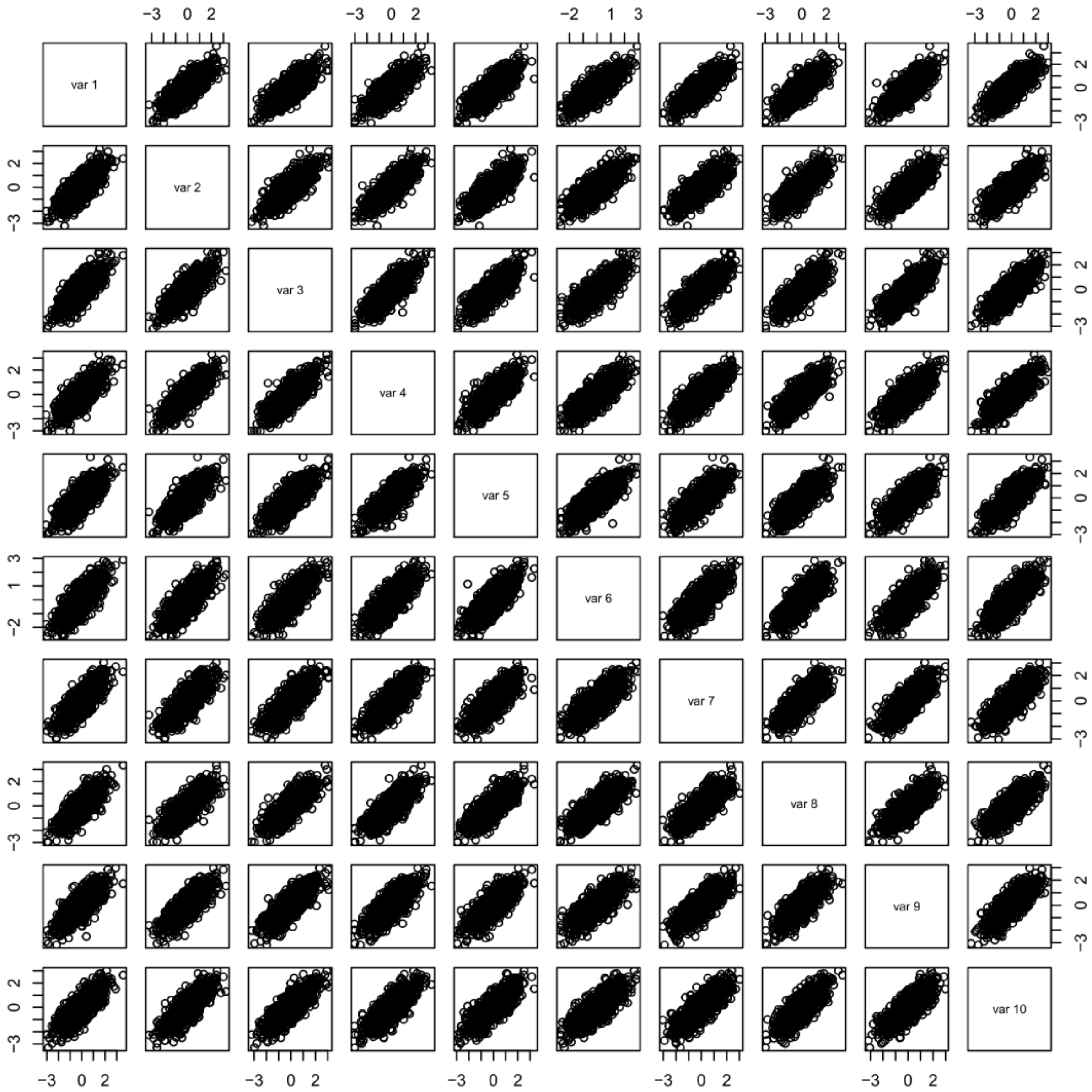


6.1 Independence. Diagnosing collinearity

- SPLOM
- Variance inflation factors
- Moran's *I*

6.1 Independence. SPLOM

```
pairs(data)

# Also see:
library(psych)
pairs.panel(data)

library(GGally)
ggpairs(data)
```

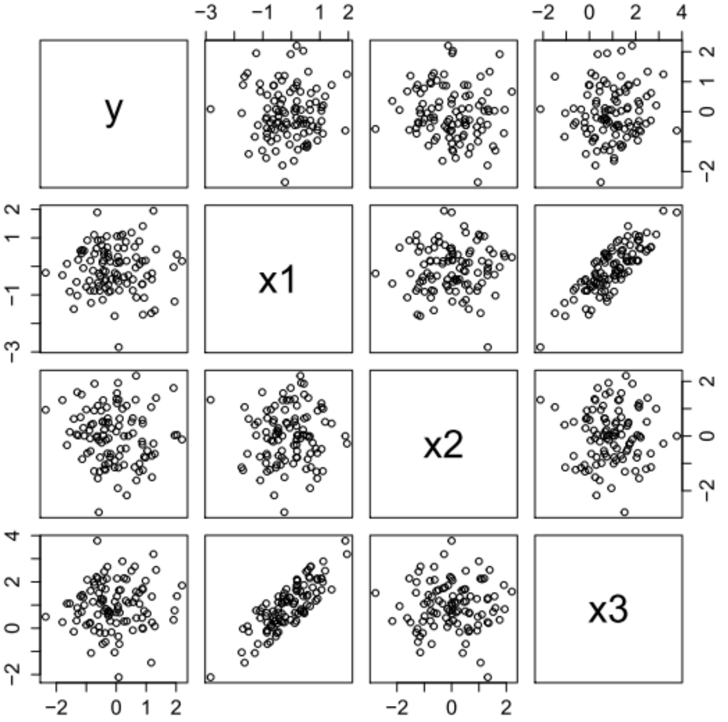


6.1 Independence. Variance inflation factors

- The proportion of variance in a predictor explained by all other predictors in the model: $1/(1 - R^2_j)$

- VIF < 2 is ideal (Zuur et al. 2009)

```
data = data.frame(
  y = rnorm(100, 0, 1),
  x1 = rnorm(100, 0, 1),
  x2 = rnorm(100, 0, 1)
)

data$x3 = data$x1 + runif(100, 0, 2)

pairs(data)
```



6.1 Independence. Variance inflation factors

```
model = lm(y ~ x1 + x2 + x3, data)

vif(model)

 x1         x2         x3
3.220394 1.004194 3.216532

# Drop correlated terms
model2 = update(model, . ~ . -x3)

vif(model2)

x1         x2
1.001204 1.001204
```

6.1 Independence. Moran's *I*

Estimates correlation based on distance matrix

```
random.sp.data = data.frame(
  y = rnorm(100, 0, 1),
  lat = sample(seq(36, 38, 0.1), 100, replace = T),
  long = sample(seq(-72, -70, 0.1), 100, replace = T)
)

order.sp.data = data.frame(
  y = c(runif(20, 0, 1), runif(20, 2, 10), runif(20, 10, 20),
    runif(20, 20,  40), runif(20, 40, 80)),
  lat = seq(36, 38, 0.02)[-1],
  long = seq(-72, -70, 0.02)[-1]
)

# Get spatial distance matrix
random.sp.dist = sp::spDists(as.matrix(random.sp.data[, c("long",
    "lat")]), longlat = T)

order.sp.dist = sp::spDists(as.matrix(order.sp.data[, c("long",
    "lat")]),  longlat = T)
```

6.1 Independence. Moran's *I*

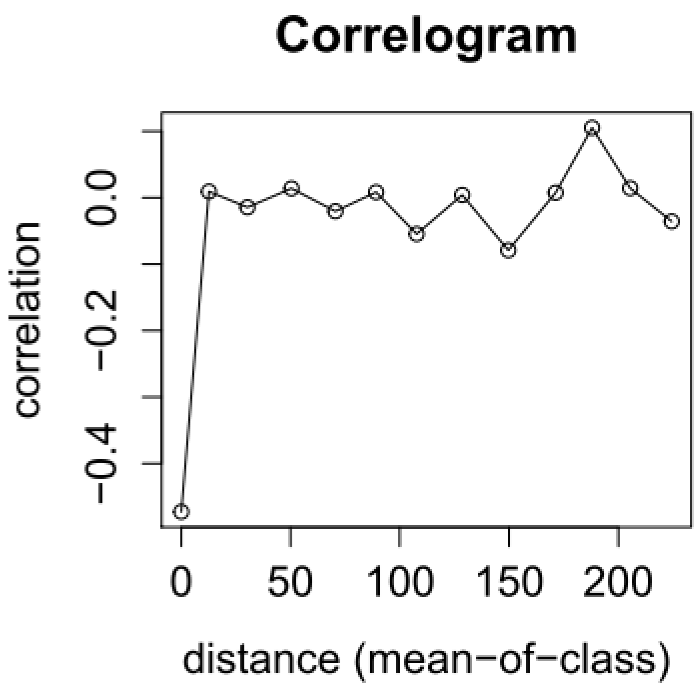
```
Moran.I(random.sp.data$y,
    random.sp.dist)

$observed
[1] -0.01325627

$expected
[1] -0.01010101

$sd
[1] 0.005655542

$p.value
[1] 0.5769092
```

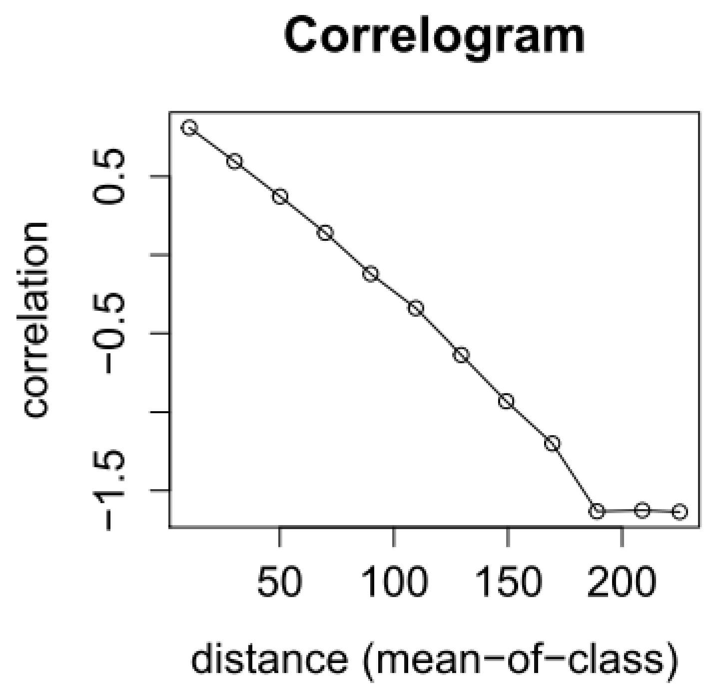
```
Moran.I(order.sp.data$y,
    order.sp.dist)

$observed
[1] -0.4392363

$expected
[1] -0.01010101

$sd
[1] 0.008284921

$p.value
[1] 0
```

6.1 Independence. Correlogram

Estimates correlation with increasing lag (distance) among data points

```
random.correlog =
   correlog(random.sp.data$lat, random.sp.data$long,
   z = random.sp.data$y, increment = 20, resamp = 10, latlon = T)

order.correlog =
   correlog(order.sp.data$lat, order.sp.data$long,
   z = order.sp.data$y, increment = 20, resamp = 10, latlon = T)
```





6.1 Independence. Treating collinearity

- Centering and scaling
- Variable selection
- PCA
- Model correlations / covariances

6.2 Modeling Autocorrelation

6.2 Modeling Autocorrelation. Options

1. Specify fixed correlation structure

2. Fit random effects

6.2 Modeling Autocorrelation. Correlation 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	$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$
First-Order Autoregressive	AR(1)	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$

nlme: Linear and Nonlinear Mixed Effects Models

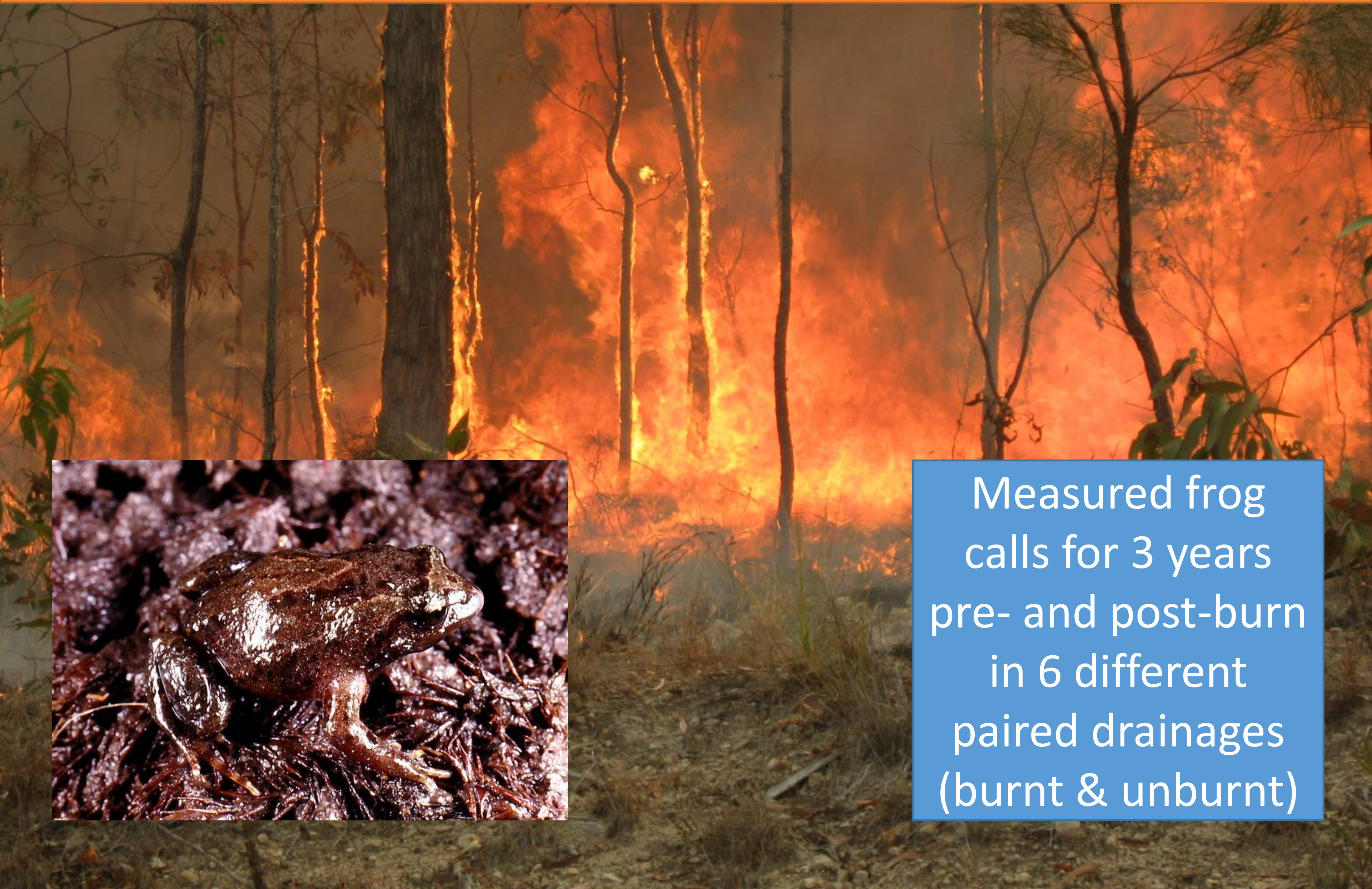
```
install.packages("nlme")
library(nlme)
```

6.2 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).
corCompSymm
compound symmetry structure corresponding to a constant correlation.
corExp
exponential spatial correlation.
corGaus
Gaussian spatial correlation.
corLin
linear spatial correlation.
corRatio
Rational quadratics spatial correlation.
corSpher
spherical spatial correlation.
corSymm
general correlation matrix, with no additional structure.

Driscoll & Roberts (1997): Impact of burning on recovery of Walpole frog (*Geocrinia lutea*)



Measured frog calls for 3 years pre- and post-burn in 6 different paired drainages (burnt & unburnt)

6.2 Modeling Autocorrelation. Driscoll Example

How does the difference in frog calls between burnt and unburnt plots change through time?

→ Fit within block (drainage) correlation structure to account for temporal autocorrelation within blocks

6.2 Modeling Autocorrelation. Driscoll Example

```
driscoll = read.csv("driscoll.csv")

# Fit and test different correlation structure
driscoll.none = gls(CALLS ~ YEAR, data = driscoll,
    na.action = na.omit)

driscoll.unstr = gls(CALLS ~ YEAR, data = driscoll,
    na.action = na.omit,
    correlation = corSymm(form = ~ 1 | DRAINAGE))

driscoll.cs = gls(CALLS ~ YEAR, data = driscoll,
    na.action = na.omit,
    correlation = corCompSymm(form = ~ 1 | DRAINAGE))

driscoll.ar1 = gls(CALLS ~ YEAR, data = driscoll,
    na.action = na.omit,
    correlation =   corAR1(form = ~ 1 | DRAINAGE))
```

6.2 Modeling Autocorrelation. Driscoll Example

```
# Compare models
anova(driscoll.none, driscoll.unstr, driscoll.cs, driscoll.ar1)

 Model                df      AIC      BIC    logLik   Test   L.Ratio p-value
driscoll.none      1  3 117.8460 119.9702 -55.92300
driscoll.unstr     2  6 110.5396 114.7879 -49.26980 1 vs 2 13.306407  0.0040
driscoll.cs        3  4 111.6621 114.4943 -51.83105 2 vs 3  5.122516  0.0772
driscoll.ar1       4  4 110.8742 113.7064 -51.43708


summary(driscoll.none)$tTable

                 Value Std.Error   t-value    p-value
(Intercept) -10.656863  5.280095 -2.018309 0.06181010
YEAR          5.416667  2.434004  2.225414 0.04181391

summary(driscoll.ar1)$tTable

                 Value Std.Error   t-value     p-value
(Intercept) -12.039670  4.265776 -2.822387 0.012864953
YEAR          5.696191  1.537009  3.706023 0.002112864
```

6.2 Modeling Autocorrelation. Options

1. Specify correlation structure
2. Fit random effects
3. Both??

Correlation structure accounts for long-term temporal trend

Random structure accounts for random variation among time points

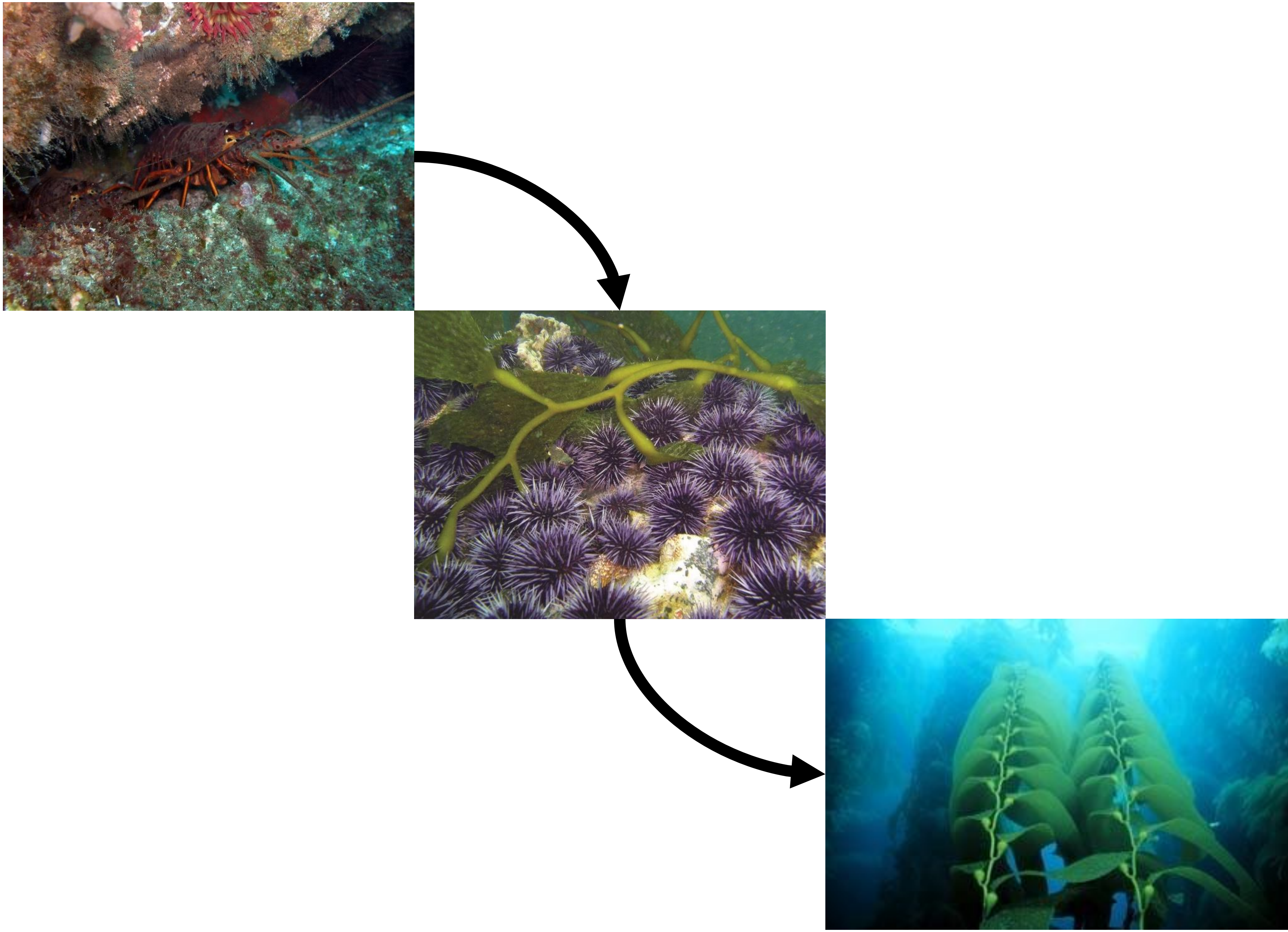
See comments: https://dynamicecology.wordpress.com/2015/11/04/is-it-a-fixed-or-random-effect/

6.3 Temporal Autocorrelation Example

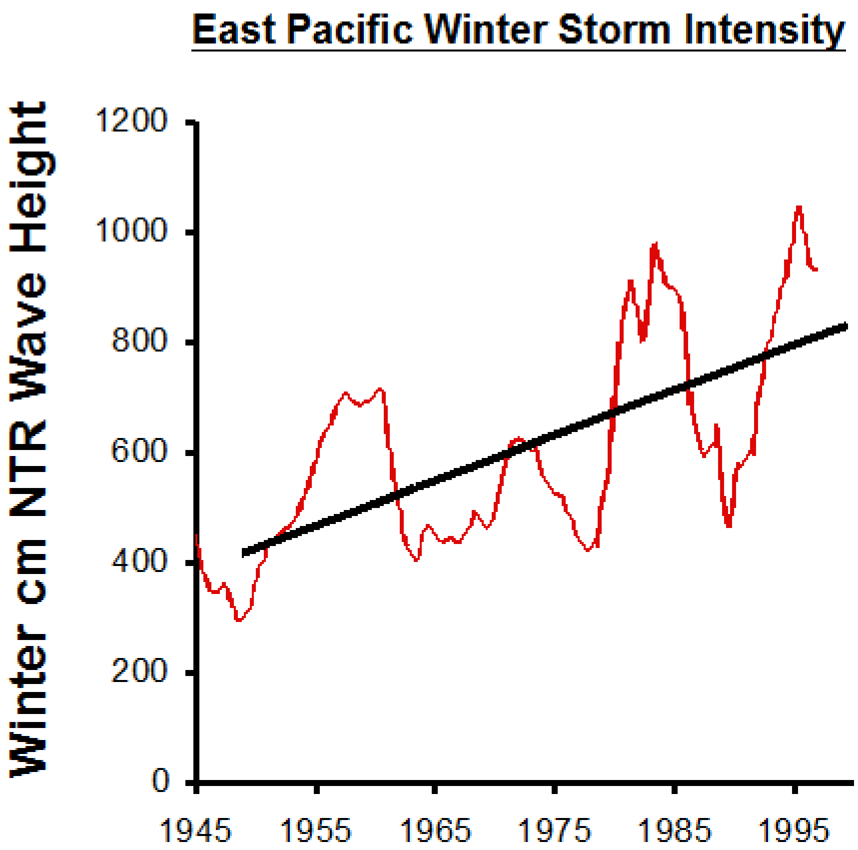
6_Autocorrelation.R

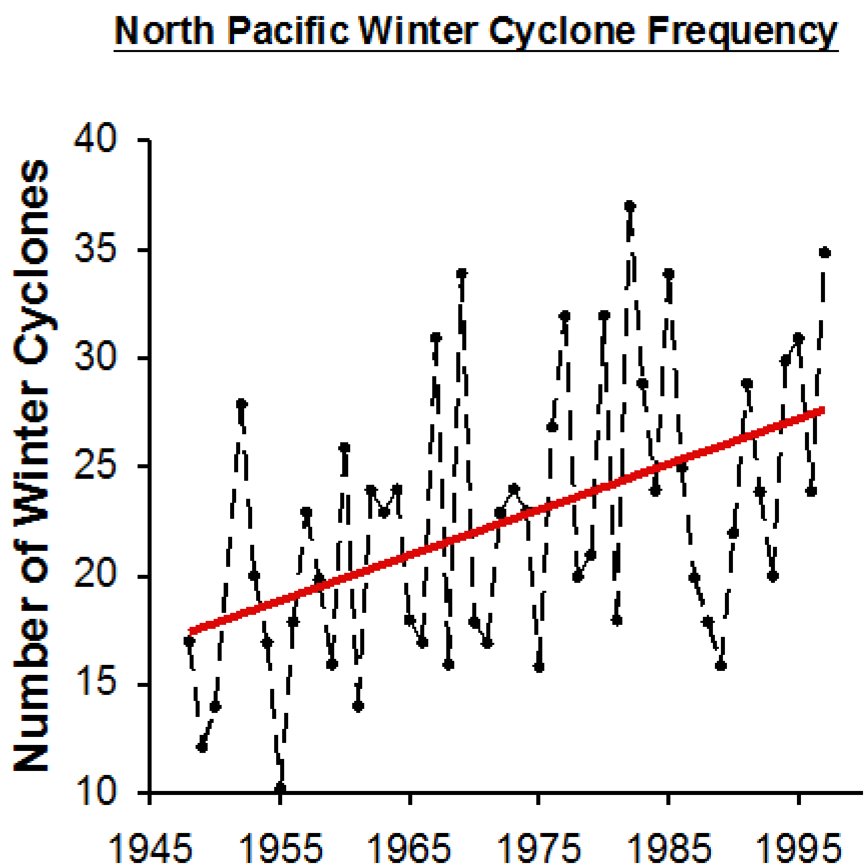
Byrnes et al (2011) – Disturbance impacts on food web structure in kelp forests

6.3 SEM Example. Kelp food webs



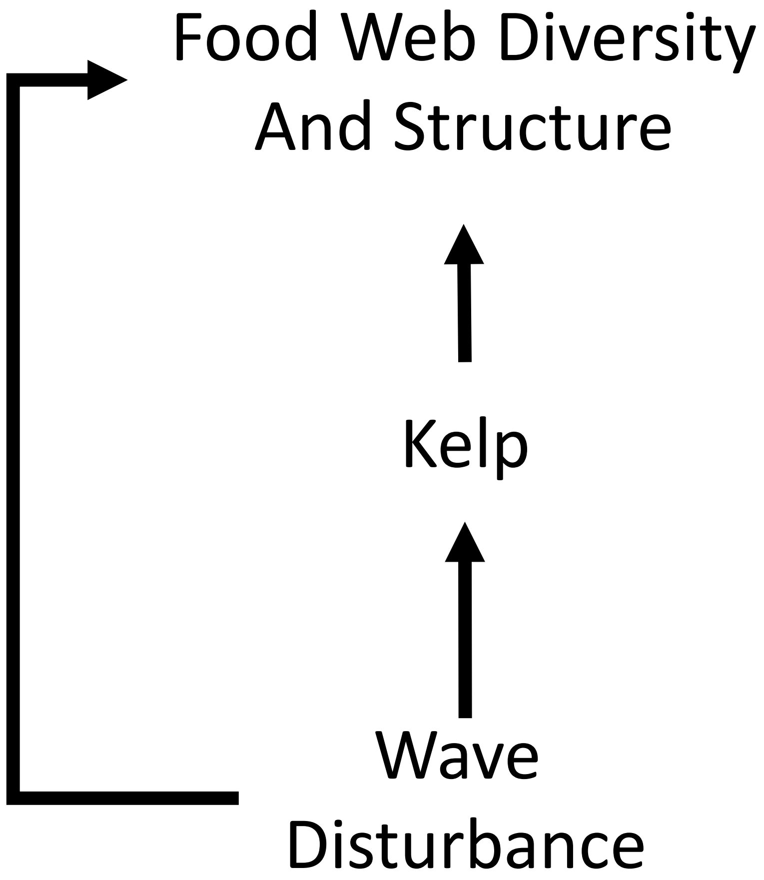
6.3 SEM Example. Increasing storms



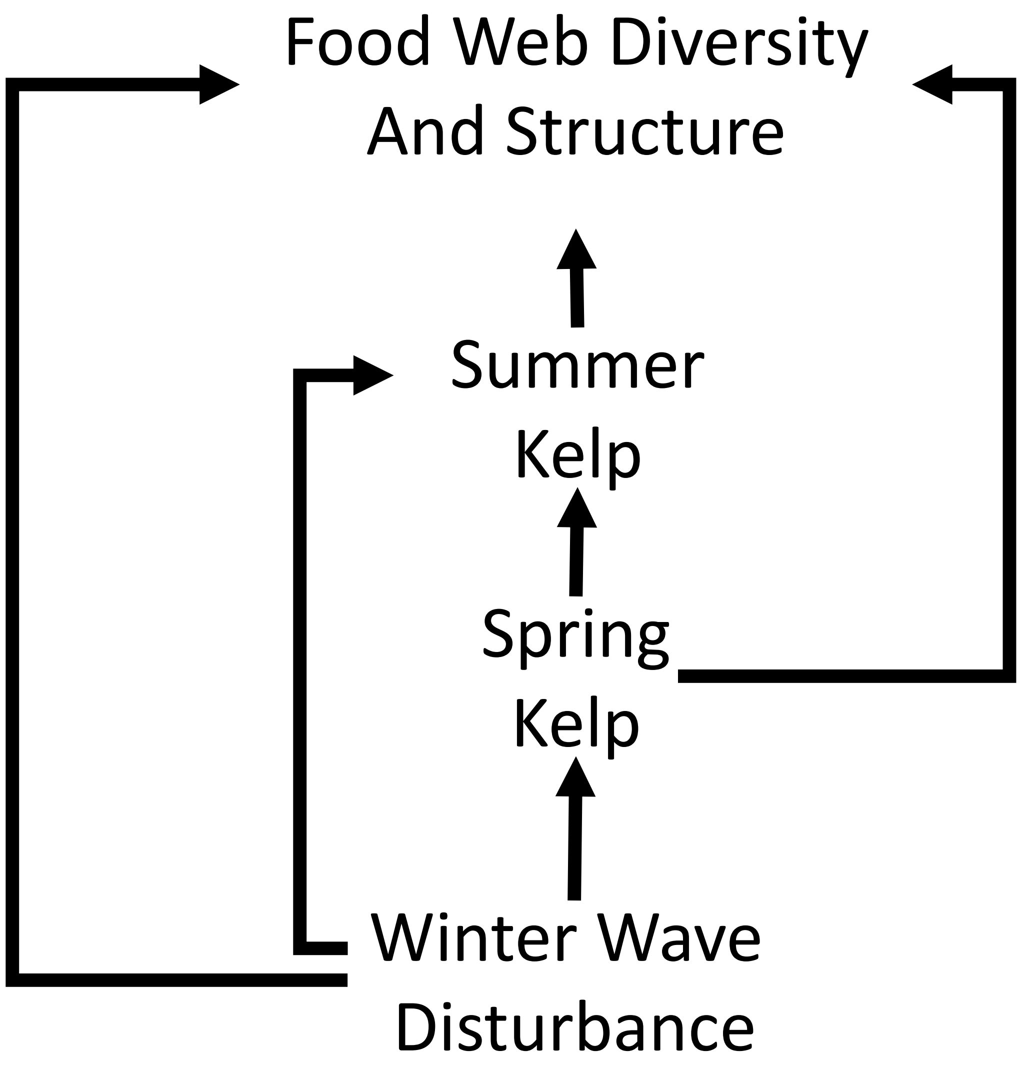


6.3 SEM Example. Sampling 2000-2009

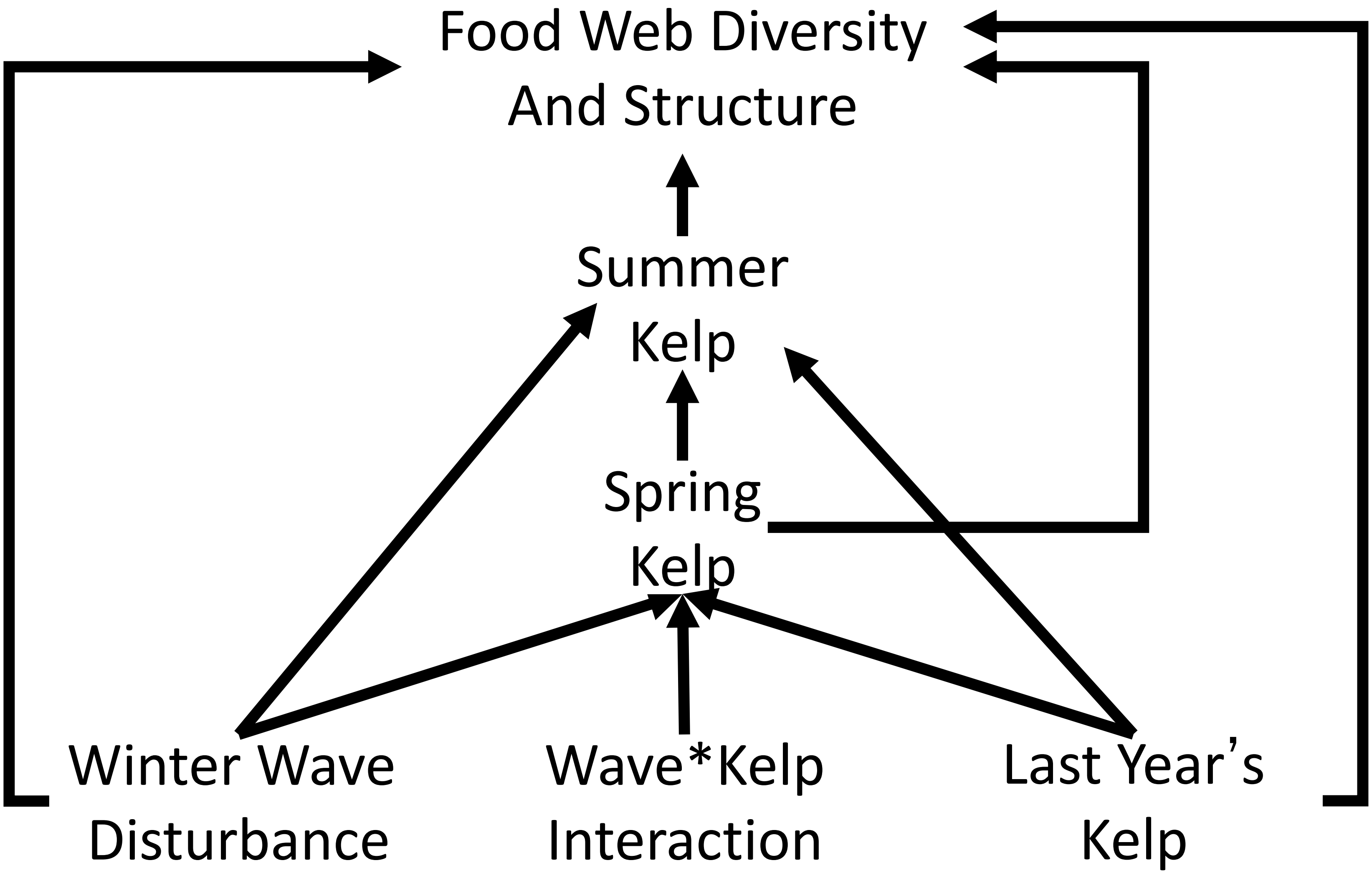
6.3 SEM Example. Meta-model



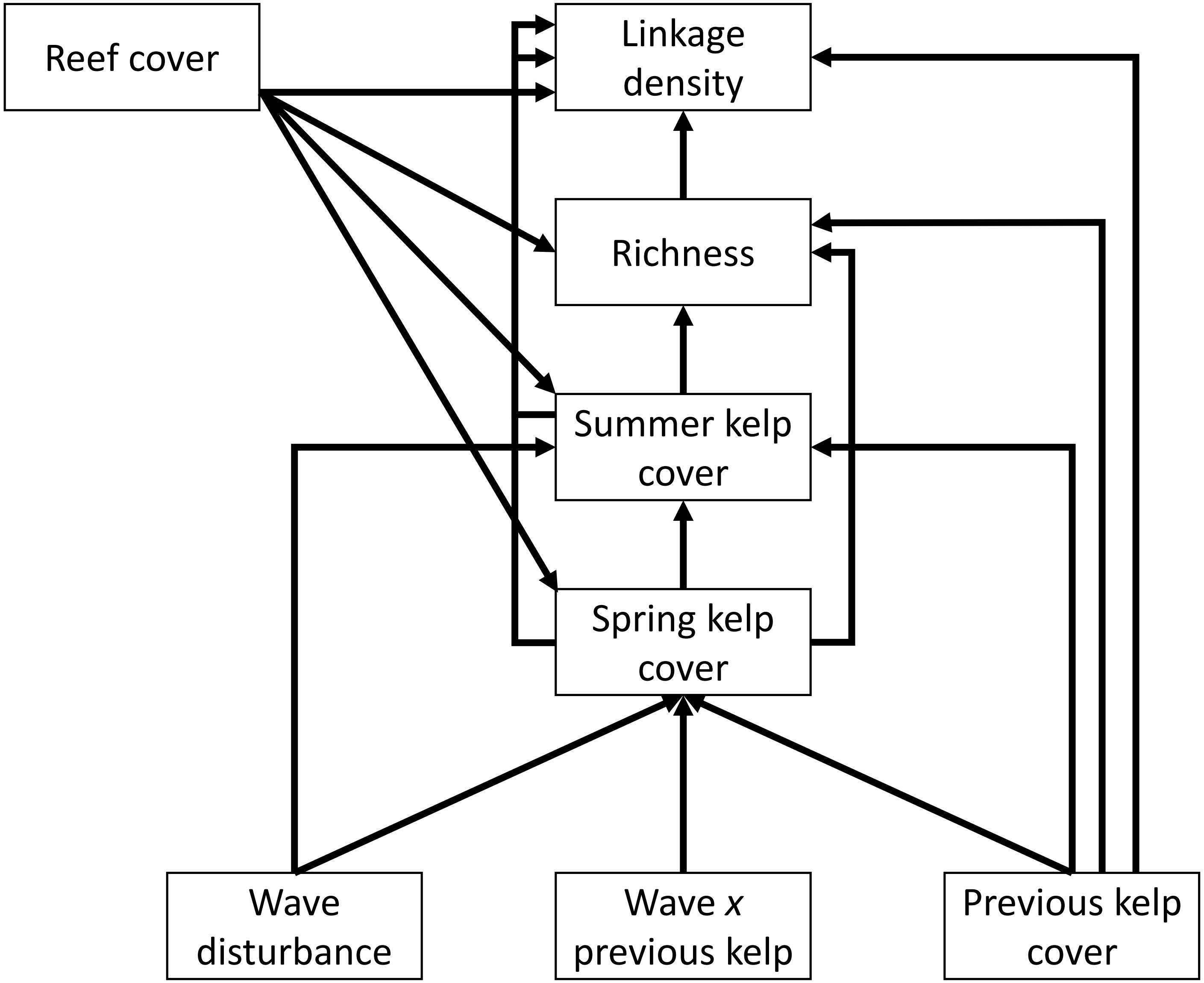
6.3 SEM Example. Meta-model



6.3 SEM Example. Meta-model



6.3 SEM Example. Full model



6.3 SEM Example. Formatting the data

```
# Read Byrnes data
byrnes = read.csv("byrnes.csv")

# Only examine complete cases (remove NAs)
byrnes = byrnes[complete.cases(byrnes), ]

# Log transform some variables (as in original analysis)
byrnes$prev.kelp = log(byrnes$prev.kelp + 1)

byrnes$spring.kelp = log(byrnes$spring.kelp + 1)

byrnes$summer.kelp = log(byrnes$summer.kelp + 1)
```

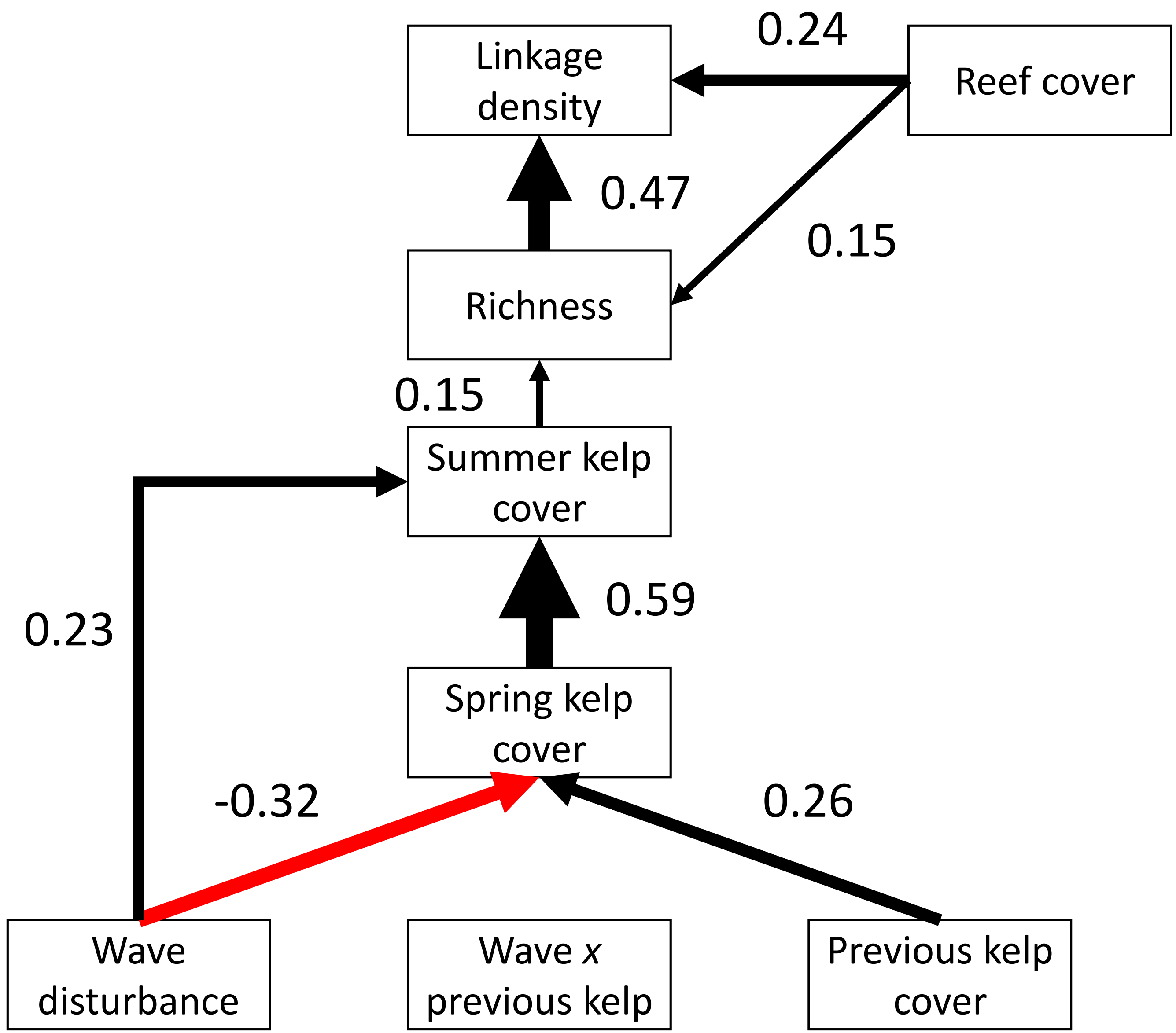
6.3 SEM Example. Fit the model

```
# Create model with correlation structure
byrnes.sem <- psem(

  spring.kelp <- gls(spring.kelp ~ wave * prev.kelp + reef.cover,
    correlation = corCAR1(form = ~ YEAR | SITE /   TRANSECT),
    data = byrnes),

  kelp <- gls(summer.kelp ~ wave * prev.kelp + reef.cover + spring.kelp,
    correlation = corCAR1(form = ~ YEAR | SITE / TRANSECT),
    data =  byrnes),

  richness <- gls(richness ~ summer.kelp + prev.kelp + reef.cover +
    spring.kelp,
    correlation = corCAR1(form = ~ YEAR | SITE / TRANSECT),
    data = byrnes),

  linkdensity <- gls(linkdensity ~ richness + summer.kelp + prev.kelp +
    reef.cover + spring.kelp,
    correlation = corCAR1(form = ~ YEAR | SITE / TRANSECT),
    data = byrnes)

)
```

6.3 SEM Example. Goodness-of-fit

```
Goodness-of-fit:

  Global model: Fisher's C = 1.912 with P-value = 0.752 and on 4 degrees of
freedom

  Individual R-squared:
       Response R.squared
  spring.kelp      0.17
  summer.kelp      0.34
     richness      0.04
  linkdensity      0.37
```

6.3 SEM Example. Full model



6.3 SEM Example. Inferences

- Wave action reduces spring kelp cover
- Spring and summer kelp cover enhances richness, and therefore food web structure
- Wave action reduces food web structure:
    - -0.32 * 0.59 * 0.15 * 0.47 = -0.01

6.3 SEM Example. Your turn…

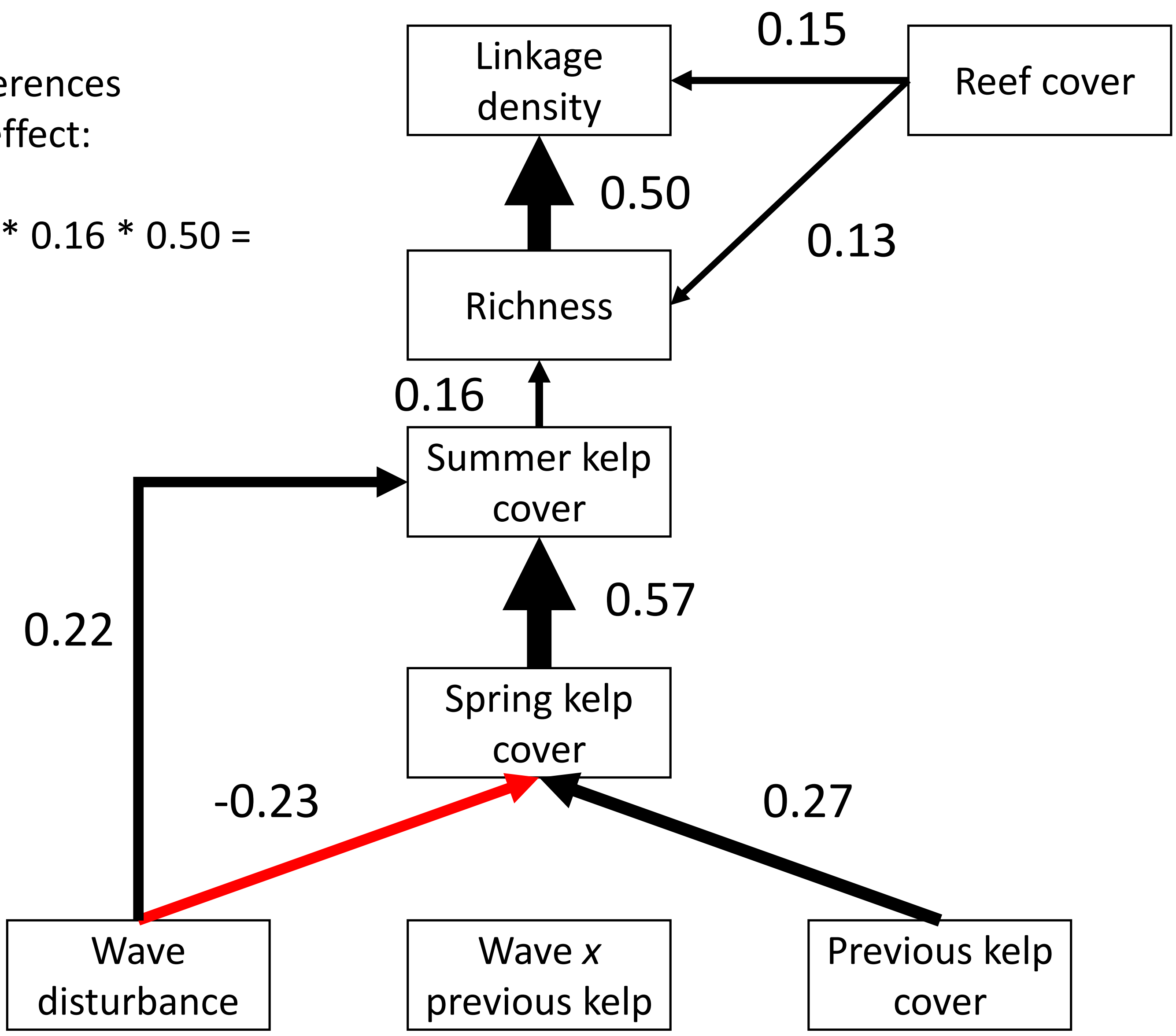
Re-fit the model addressing the hierarchical structure of the sampling (hint: random effect of transect within site) *<u>and</u>* the temporal autocorrelation!

Compare using AIC…

6.3 SEM Example. Your turn...

- Same inferences
- Indirect effect:

-0.23 * 0.57 * 0.16 * 0.50 =
-0.01



6.3 SEM Example. Your turn…

- AIC (temporal structure only) = 61.912
- AIC (temporal & hierarchical structure) = 79.028
- GLS model incorporating temporal lag for transects within site sufficient (ΔAIC > 2)
- Random variation after accounting for temporal lag not significant
- Same answer (estimates) anyways

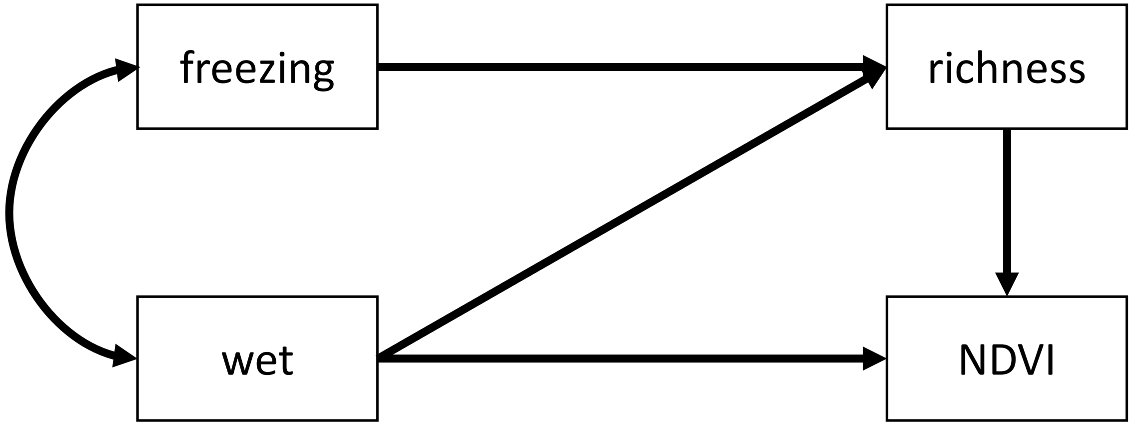
6.4 Spatial Autocorrelation Example

6_Autocorrelation.R

Harrison & Grace (2007) - Drivers of forest productivity in California

Spatially explicit forest and environmental data

6.4 SEM Example 2. “Independent”




```
boreal.sem <- psem(

  lm(richness ~ freezing, data = boreal),

  lm(NDVI ~ richness + freezing + wet, data = boreal),

  freezing %~~% wet,

  data = boreal

)
```

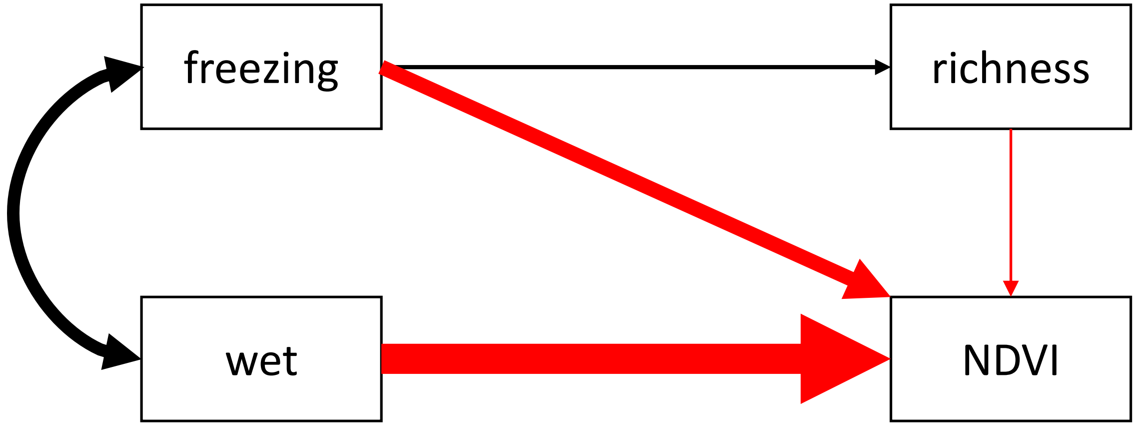
6.4 SEM Example 2. “Independent”

```
Goodness-of-fit:

  Global model: Fisher's C = 2.422 with P-value = 0.298 and on 2 degrees of
freedom

  Individual R-squared:
    Response R.squared
  richness      0.01
      NDVI      0.77
```

6.4 SEM Example 2. "Independent"

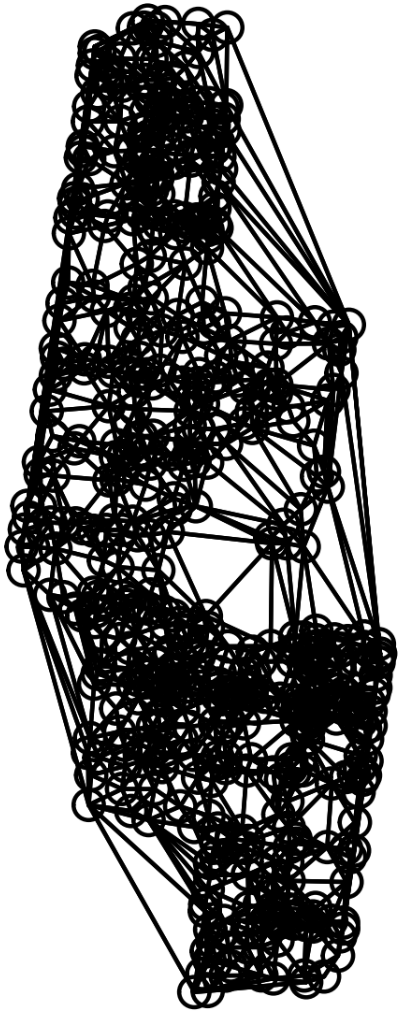
```
Structural Equation Model of boreal.sem

Call:
  richness ~ freezing
  NDVI ~ richness + freezing + wet
  freezing ~~ wet

    AIC      BIC
 18.422   52.65

Tests of directed separation:

          Independ.Claim Estimate Std.Error  DF Crit.Value P.Value
  richness  ~  wet + ...   -35.13   33.7123 530    -1.0421  0.2979

Coefficients:

    Response Predictor Estimate Std.Error  DF Crit.Value P.Value Std.Estimate
    richness  freezing   1.1707    0.5470 531     2.1402  0.0328       0.0925   *
        NDVI  richness  -0.0004    0.0002 529    -2.0862  0.0374      -0.0440   *
        NDVI  freezing  -0.0355    0.0023 529   -15.6564  0.0000      -0.3456 ***
        NDVI       wet  -4.2701    0.1329 529   -32.1406  0.0000      -0.7066 ***
  ~~freezing     ~~wet   0.2961        NA 531     7.1436  0.0000       0.2961 ***
  ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '' 1
```

6.4 SEM Example 2. “Independent”

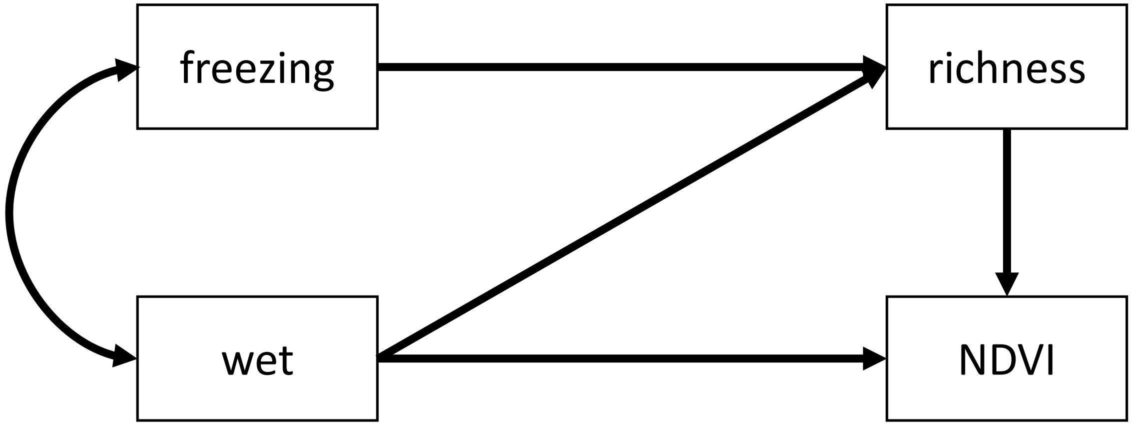


6.4 SEM Example 2. Spatial weights

```
coordinates(boreal) <- ~x+y

nb <- tri2nb(boreal)

plot(nb, coordinates(boreal))
```



6.4 SEM Example 2. Spatial weights




```
boreal.sem2 <- psem(

  lagsarlm(richness ~ freezing, data = boreal, listw =
   nb2listw(nb), tol.solve = 1e-11),

  lagsarlm(NDVI ~ richness + freezing + wet, data = boreal,
   listw = nb2listw(nb), tol.solve = 1e-11),

  freezing %~~% wet,

  data = boreal)
```

6.4 SEM Example 2. Spatial weights

```
Goodness-of-fit:

  Global model: Fisher's C = 1.774 with P-value = 0.412 and on 2 degrees of
freedom

  Individual R-squared:
    Response R.squared
  richness        NA
      NDVI        NA
```

6.4 SEM Example 2. Spatial weights

```
Structural Equation Model of boreal.sem2

Call:
  richness ~ freezing
  NDVI ~ richness + freezing + wet
  freezing ~~ wet

    AIC      BIC
 21.774   64.559

Tests of directed separation:

          Independ.Claim Estimate Std.Error DF Crit.Value P.Value
  richness  ~  wet + ... -25.6601   31.2746 NA    -0.8205  0.4119

Coefficients:

    Response Predictor Estimate Std.Error  DF Crit.Value P.Value Std.Estimate
    richness  freezing   0.6808    0.5094  NA     1.3364  0.1814       0.0538
        NDVI  richness  -0.0003    0.0001  NA    -1.8619  0.0626      -0.0338
        NDVI  freezing  -0.0207    0.0023  NA    -9.1514  0.0000      -0.2016 ***
        NDVI       wet  -3.1033    0.1519  NA   -20.4340  0.0000      -0.5135 ***
  ~~freezing     ~~wet   0.2961        NA 531     7.1436  0.0000       0.2961 ***
  ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '' 1
```

6.4 SEM Example 2. Comparison