Extending ANOVA and ANCOVA Analyses using SEM

Jim Grace

U.S. Department of the Interior
U.S. Geological Survey

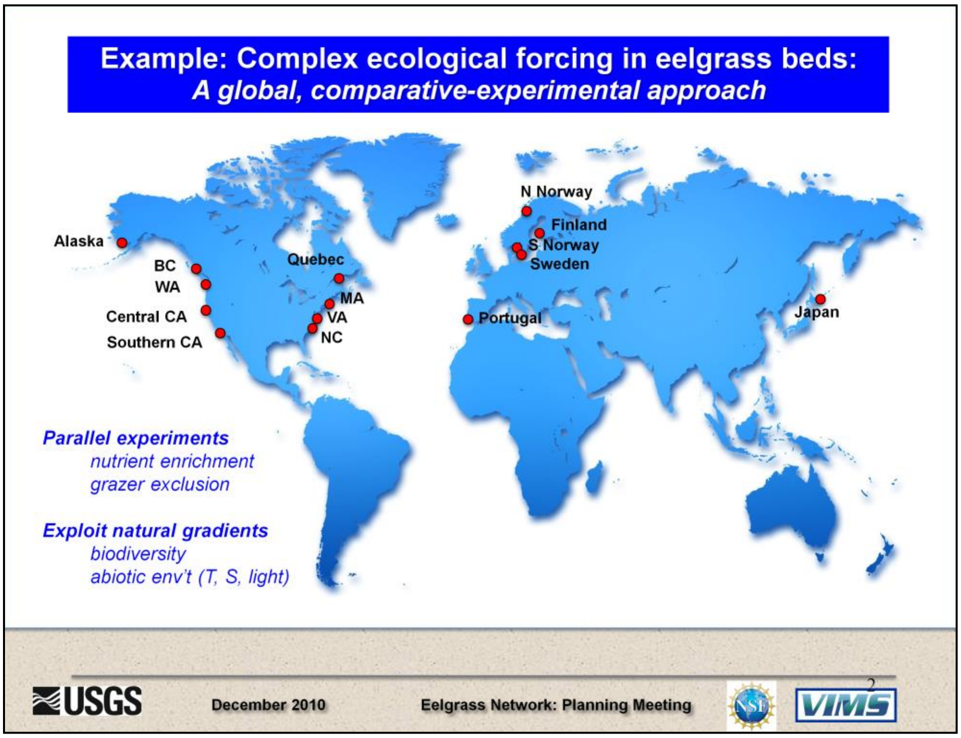
1

In this module I consider an example where randomized experiments were used to study effects. Here I try to show how SEM permits additional understanding to be developed.

An appropriate citation for this material is

Whalen, M.A., Duffy, J.E. and Grace, J.B. 2013. Temporal shifts in top-down versus bottom-up control of epiphytic algae in a seagrass ecosystem. *Ecology* 94:510-520.

Notes: IP-056512; Support provided by the USGS Climate & Land Use R&D and Ecosystems Programs. I would like to acknowledge formal review of this material by Jesse Miller and Phil Hahn, University of Wisconsin. Many helpful informal comments have contributed to the final version of this presentation. The use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Last revised 20141216. Questions about this material can be sent to sem@usgs.gov.



These data come from a global experiment being conducted on seagrasses.

Data from:
Field-based Experimental Study of the Importance of Small Herbivores in a Seagrass Ecosystem:

Matthew A Whalen and J Emmett Duffy

Whalen, Duffy, and Grace, 2013. *Ecology* 94:510-520.
(http://www.esajournals.org/doi/abs/10.1890/12-0156.1)

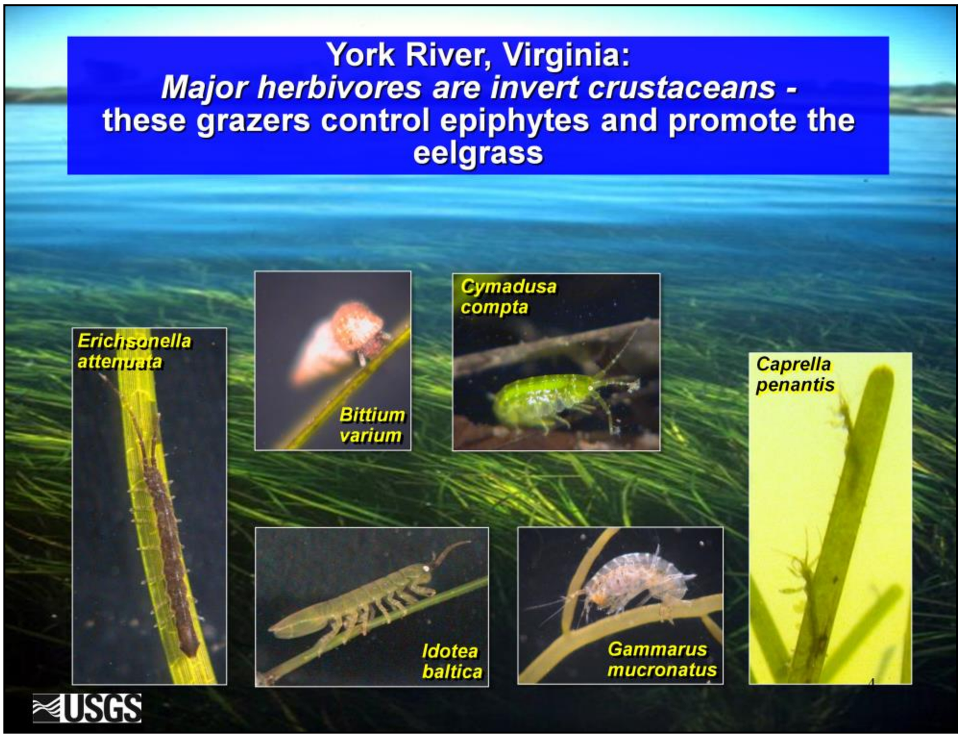



WILLIAM & MARY VIMS
VIRGINIA INSTITUTE OF MARINE SCIENCE
SCHOOL OF MARINE SCIENCE

USGS

3

More specifically, these are from a study in Virginia.



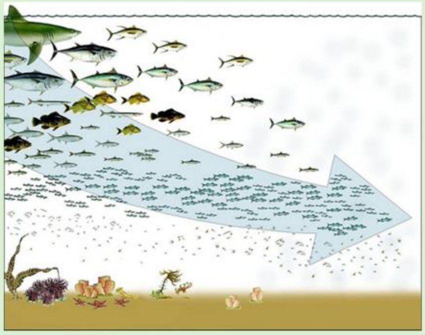
It is all about microcrustaceans grazing on the epiphytes that live on eelgrasses, a particularly important seagrass.

If grazers don’t keep epiphytes grazed down, they lead to the death of the seagrasses, causing the base of the ecosystem to collapse.

The Big Question

Are seagrasses controlled by bottom-up forces or trophic cascade?

Consumers
↓
Producers
↑
Resources



Subtext: Is nutrient runoff or overfishing causing seagrass declines?

5

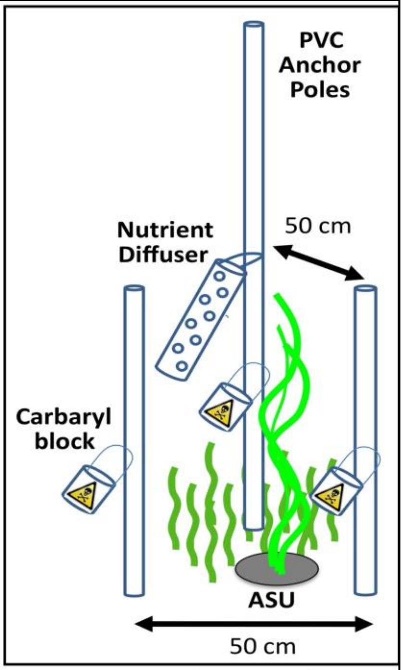
Part of the big deal is a question of what may be causing eelgrass declines worldwide and the broader implications of this issue.

Preliminary Study:
Virginia site

Experimental Design:

Treatments:
- pesticide
- nutrient addition
- combination
- controls

8 reps @ 5 trts = 40 plots



Pesticide effects:

Crustaceans: reduced 58-96%
Algal biomass: increased 130-748%

Nutrients: nonsignificant effects

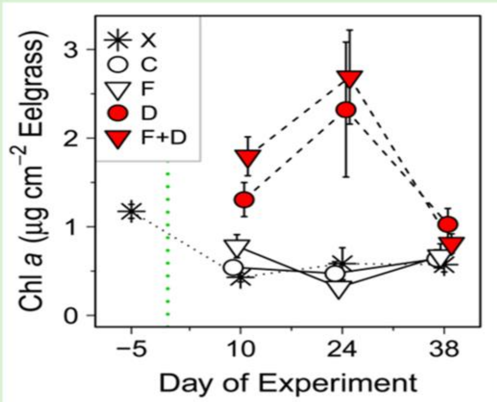
USGS

VIMS

6

Here is the part of the experimental study discussed in this example.

A Primary ANOVA result: Means for pesticide effect on epiphytes



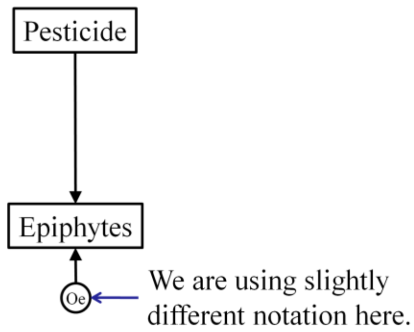
7

Anova results provide limited information.

Illustration of ANOVA-type model

```
# Read Whalen Seagrass Data
w.dat <- read.csv("WhalenData.csv")
```

```
# ANOVA Model
anovaModel <- 'epiphytes ~ pesticides'
```



8

An anova, in its most basic form, is a very simple model. The simplicity is created by the physical control in combination with randomization.

Illustration of ANOVA-type model (cont.)

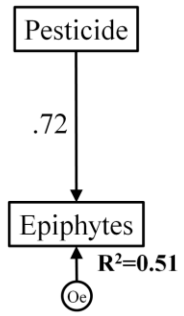
```
# Fit ANOVA Model
anovaFit <- sem(anovaModel, data=w.dat)

# Get Results
summary(anovaFit, standardized=T, rsq=T)
```

```
                    Est    SE      Z      P  Std.all
Regressions:
  epiphytes ~
    pesticides    0.998  0.154  6.48  0.000  0.716

Variances:
    epiphytes     0.227  0.051                0.488

R-Square:
    epiphytes     0.512
```

USGS

Here are the net-effect results. Note that the information extracted is similar to that obtained from an ANOVA. The main difference is that we are now treating treatment levels as points on a continuum (regression perspective) instead of simply testing for whether treatment means differ.

Results

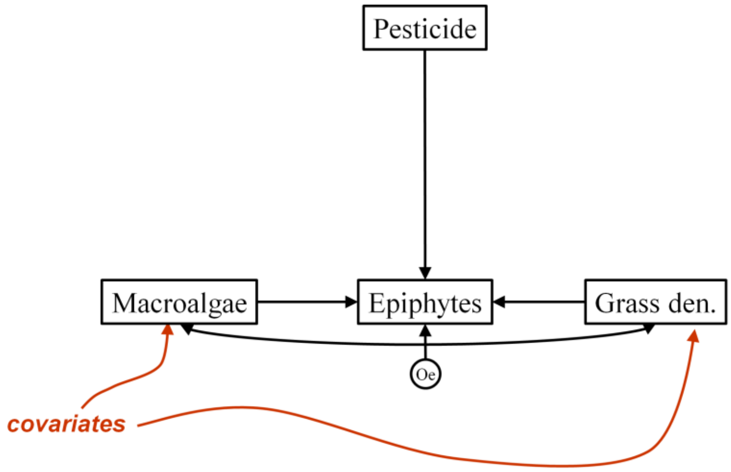


USGS

10

And shown graphically.

Illustration of ANCOVA-type model



Note: in ANCOVA, covariates are not allowed to correlate with treatment variables.

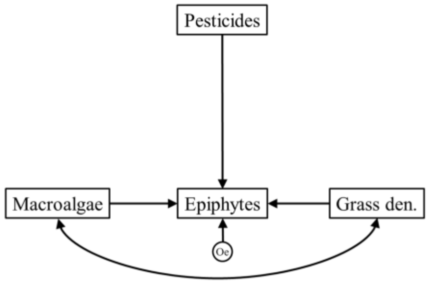
11

There are two covariates in this study, a macro alga and the density of eelgrass. We have not anticipation about what the macroalgae might do, but we expect greater eelgrass density to promote epiphytes by buffering water movement and physical damage to epiphytes.

In ANCOVA, the covariates are supposed to be uncorrelated with the treatment, which holds true in this case.

Illustration of ANCOVA-type model

```
# Simple ANCOVA Model
ancovaModel <- 'epiphytes ~ pesticides
                        + macroalgae + grass'
```

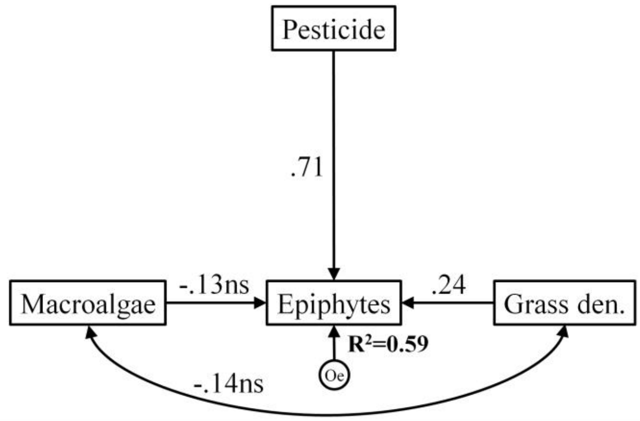


USGS

12

A simple ANCOVA here.

Results (visual)



1. Variance explanation for epiphytes improves.
2. Grass density promotes epiphyte development.
3. Macroalgae have nonsignificant negative effect on epiphytes.

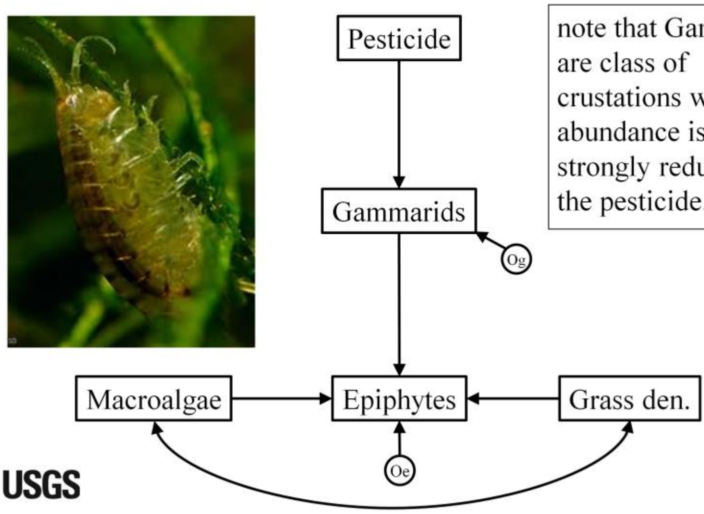
USGS

13

And the results

The test of mediation

Does reduction of Gammarids explain promotion of epiphytes by pesticide?



note that Gammarids are class of crustations whose abundance is most strongly reduced by the pesticide.

14

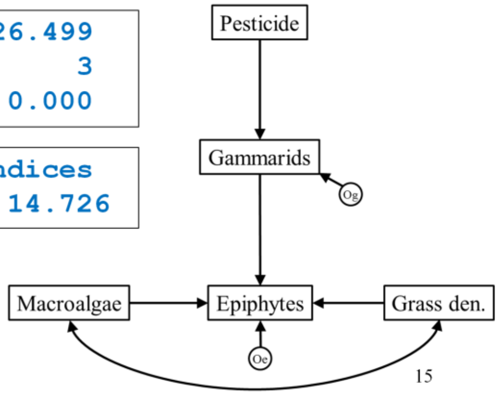
Here we perform the test of mediation with one of the microcrustaceans, the Gammarids.

Lavaan code and results

```
# SEM Model 1 "sem1"

sem1 <- 'epiphytes ~ macroalgae + grass + gammarids
         gammarids ~ pesticide'
```

```
Chi-square                 26.499
Degrees of freedom              3
P-value                     0.000
```

```
# Select Modification Indices
gammarids ~ macroalgae   14.726
```

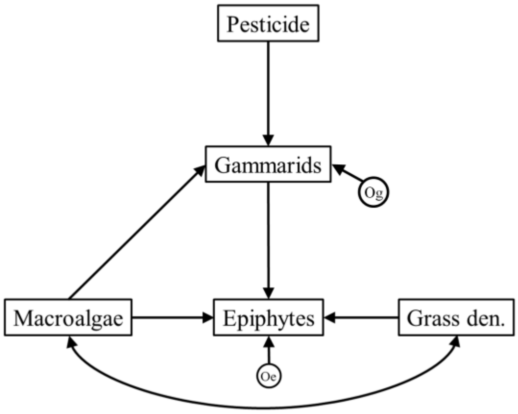


So, we should add path from macroalgae to gammarids.

USGS

Results suggest something missing from model.

Modifying our model: adding needed linkages




```
# New Model - SEM Model 2 "sem2"

sem2 <- 'epiphytes ~ macroalgae + grass + gammarids
         gammarids ~ pesticide + macroalgae'
```

An important discovery is an effect of macroalgae on Gammarids.

Results

```
Chi-square                  8.136
Degrees of freedom              2
P-value                     0.017
```

```
# Chi-square difference test
anova(sem1.fit, sem2.fit)

Chisq-diff = 18.363,
df-dif     =  1
p          = < 0.001
```

```
# Select Modification Indices
gammarids ~ grass      3.319
epiphytes ~ pesticide 4.205
```

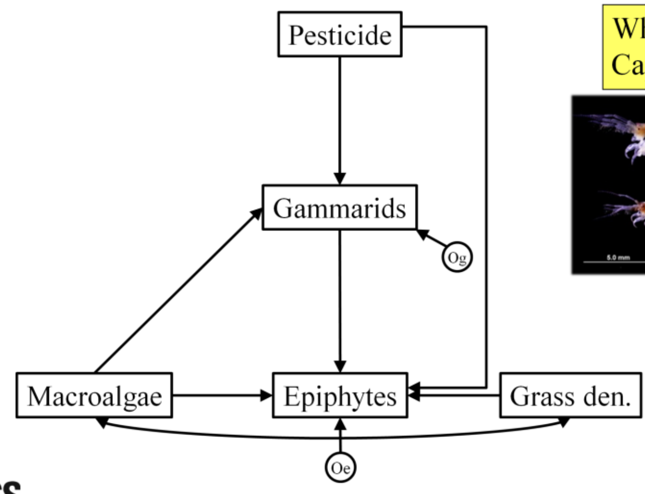
USGS

17

Model still missing another link, though the link added in model 2 definitely improved model fit dramatically. Modification indices suggest a remaining direct path from pesticide to epiphytes.

We can go further.

What is mediating the remaining effect of pesticide on epiphytes?



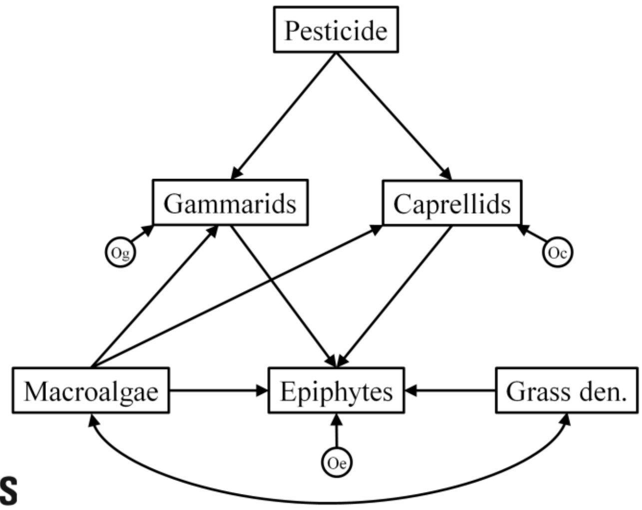
What about Caprellids?

USGS

18

Now we bring in the second most abundant type of micrograzer.

Our final model – complete mediation of pesticide.

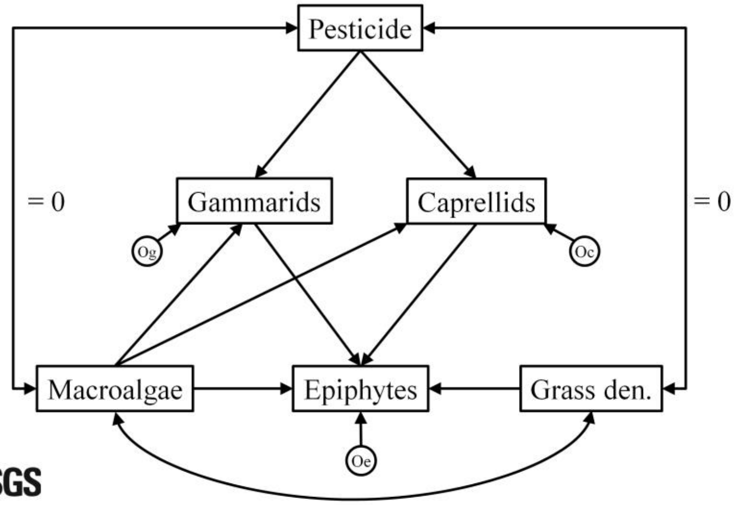


19

Finally, a fully-mediate model.

What if we wanted to include some constraits?

Here we force the correlations between treatment and covariates to equal 0.



USGS

20

Here we simple demonstrate setting exogenous correlations to zero. this permits more pure causal attribution (if it holds).

“sem5” model and results

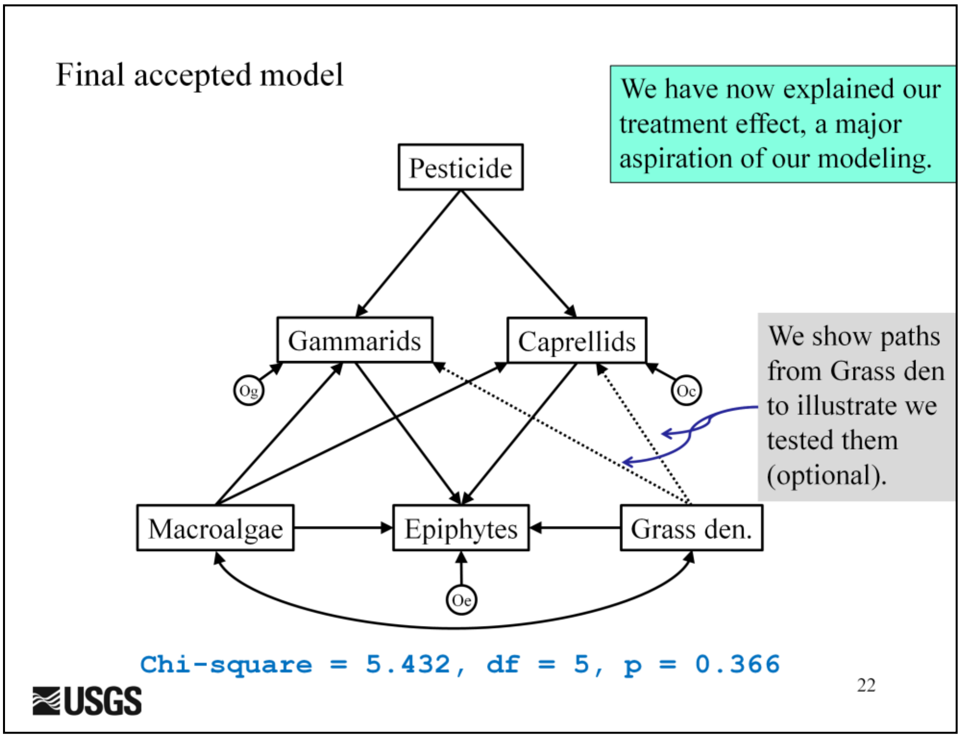
```
# SEM Model 5 “sem5”

sem5 <- 'epiphytes ~ macroalgae + grass + gammarids
                   + caprellids
          gammarids ~ pesticide + macroalgae
          caprellids ~ pesticide + macroalgae
          pesticide ~~ 0*macroalgae
          pesticide ~~ 0*grass'

sem5.fit <- sem(sem5, data=w.dat, fixed.x=F)
```

```
# Chi-square difference test
anova(sem4.fit, sem5.fit)

Chisq-diff =  1.363
df-dif     =  3
p          =  highly ns
```

```
# note we must
declare
"fixed.x=FALSE"
to work with
exogenous
correlations.
```

21

Code in red shows how we set correlations to zero.


Final accepted model
We have now explained our treatment effect, a major aspiration of our modeling.
Pesticide
Gammarids
Caprellids
Og
Oc
We show paths from Grass den to illustrate we tested them (optional).
Macroalgae
Epiphytes
Grass den.
Oe
Chi-square = 5.432, df = 5, p = 0.366
22
USGS

Final model.

USGS

Results

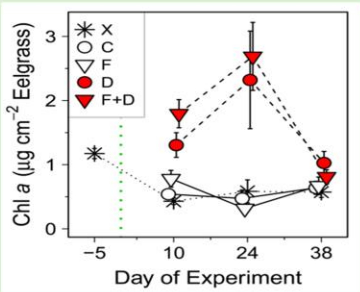
```
                Est.  Std.err  Z-val  P(>|z|)  Std.all
Regressions:
  epiphytes ~
    macroalgae   0.105    0.040   2.612    0.009    0.290
    grass        0.405    0.100   4.034    0.000    0.389
    gammarids   -0.329    0.057  -5.828    0.000   -0.663
    caprellids  -0.240    0.085  -2.834    0.005   -0.335
  gammarids ~
    pesticide   -2.053    0.215  -9.570    0.000   -0.748
    macroalgae   0.304    0.057   5.347    0.000    0.418
    grass        0.315    0.164   1.922    0.055    0.150
  caprellids ~
    pesticide   -0.748    0.231  -3.239    0.001   -0.393
    macroalgae   0.243    0.061   3.965    0.000    0.481
    grass        0.231    0.176   1.311    0.190    0.159
R-Square:
    epiphytes        0.645
    gammarids        0.756
    caprellids       0.411
```

23

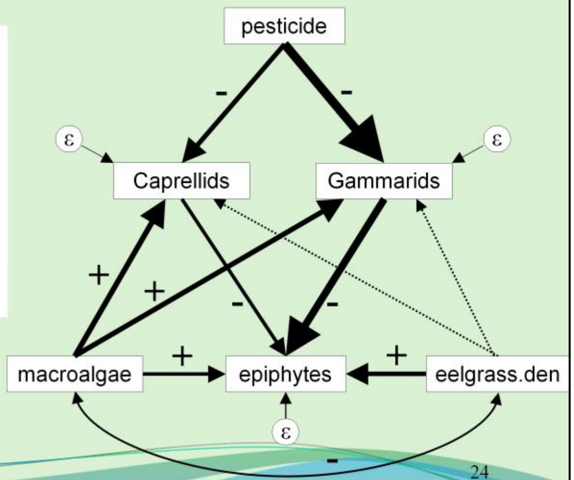
Here are the details of the estimates. Shown are raw parameter estimates (Est.), their standard errors (Std.err), associated Z-values (which are like likelihood-based t-values, the probabilities associated with the Z-values (P(>\z\), and the standardized parameter estimates (Std.all).

Our Inference

Our model results imply that behind this summary of mean responses



is a network of effects like this.



24

So, behind the standard anova result (on the left), lies a network of relationships going on.

Lessons about using SEM with experimental data

1. Test of mediation is neglected concept in biometrics.

2. Careful with classic ANCOVA; if we used mediating variables as covariates, results would indicate no significant treatment effect!

3. SEM easy to implement with simple experimental designs. With blocking, nested designs, etc., more work required for SEM analyses.

4. Recommend performing classic analyses along with SEM analyses and reporting both. Classical analyses can more easily detect interactions and in SEM you have to work to examine them (more on that later).

USGS

25

Just a few summary points.