

Supporting Information. Grace, J. B., and K. M. Irvine. 2020. Scientist's guide to developing explanatory statistical models using causal analysis principles. Ecology.

Appendix S4

Marine Food Web SEM Analyses

Data and R code for these analyses can be found in "Data S4". Analyses were run using R version 3.5.3 (R Core Team 2019).

Box S1. R code for ANOVA-type Models represented in Figure 6.

```
dat <- read.csv("AppendixS4_data.csv ")
names(dat)
attach(dat)

#####
# CONVENTIONAL ANALYSIS
#####

##### MODELS IN MANUSCRIPT
### Model A Simple ANOVA
modA <- aov(Epiphytes ~ Treatment)
drop1(modA, ~., test="F")

### Model B Simple MANOVA
Ygc <- cbind(Gammarids,Caprellids)
modB <- manova(Ygc ~ Treatment)
summary.manova(modB)

### Model C Full MANOVA
Yegc <- cbind(Epiphytes,Gammarids,Caprellids)
modC <- manova(Yegc ~ Treatment)
summary.manova(modC)

### Model D MANCOVA
modD <-manova(Ygc ~ Treatment + Epiphytes)
summary.manova(modD)

### Model E ANCOVA #1
modE <- aov(Epiphytes ~ Treatment + Macroalgae + Seagrassdens)
drop1(modE, ~., test="F")

### Model F ANCOVA #2
modF <- aov(Epiphytes ~ Treatment + Gammarids + Caprellids)
drop1(modF, ~., test="F")

### Model G ANCOVA #3
modG <- aov(Epiphytes ~ Treatment + Gammarids + Caprellids +
Macroalgae + Seagrassdens)
drop1(modG, ~., test="F")
```

Box S2. R code for ANOVA-type Models represented in Figure 6 run as linear models.

```
# Load library "lm.beta" - has function "lm.beta" to standardize coeffs
library(lm.beta)

### Model A Simple ANOVA
modA.lm <- lm(Epiphytes ~ Treatment)
summary(lm.beta(modA.lm))

### Model B Simple MANOVA
Ygc <- cbind(Gammarids,Caprellids)
modB.lm <- lm(Ygc ~ Treatment)
summary(lm.beta(modB.lm))

### Model C Full MANOVA
Yegc <- cbind(Epiphytes,Gammarids,Caprellids)
modC.lm <- lm(Yegc ~ Treatment)
summary(lm.beta(modC.lm))

### Model D MANCOVA
modD.lm <-lm(Ygc ~ Treatment + Epiphytes)
summary(lm.beta(modD.lm))

### Model E ANCOVA #1
modE.lm <- lm(Epiphytes ~ Treatment + Macroalgae + Seagrassdens)
summary(lm.beta(modE.lm)) # request standardized coefficients

### Model F ANCOVA #2
modF.lm <- lm(Epiphytes ~ Treatment + Gammarids + Caprellids)
summary(lm.beta(modF.lm))

### Model G ANCOVA #3
modG.lm <- lm(Epiphytes ~ Treatment + Gammarids + Caprellids +
Macroalgae + Seagrassdens)
summary(lm.beta(modG.lm))
```

Box S3. Results for ANOVA-type models in Box S2 (Part 1 of 3).

```
library(piecewiseSEM)

### Model A Simple ANOVA
> modA.lm <- lm(Epiphytes ~ Treatment)

Coefficients:
              Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept) -0.65328      0.00000      0.09987  -6.541 1.03e-07 ***
Treatment    0.99767      0.71575      0.15791   6.318 2.09e-07 ***
---
Residual standard error: 0.4893 on 38 degrees of freedom
Multiple R-squared:  0.5123, Adjusted R-squared:  0.4995
F-statistic: 39.92 on 1 and 38 DF, p-value: 2.088e-07

### Model B Simple MANOVA
Ygc <- cbind(Gammarids,Caprellids)
modB.lm <- lm(Ygc ~ Treatment)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
[1,]      3.2519      0.2545  12.776 2.50e-15 ***
[2,]     -2.1959      0.4024  -5.456 3.17e-06 ***
---
Residual standard error: 1.247 on 38 degrees of freedom
Multiple R-squared:  0.6004, Adjusted R-squared:  0.5898
F-statistic: 57.08 on 1 and 38 DF, p-value: 4.414e-09

### Model C Full MANOVA
Yegc <- cbind(Epiphytes,Gammarids,Caprellids)
modC.lm <- lm(Yegc ~ Treatment)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
[1,]    -0.6533      0.2734  -2.389  0.0219 *
[2,]     0.9977      0.4323   2.308  0.0266 *
---
Residual standard error: 1.339 on 38 degrees of freedom
Multiple R-squared:  0.7468, Adjusted R-squared:  0.7401
F-statistic: 112.1 on 1 and 38 DF, p-value: 6.838e-13
```

Box S4. Results for ANOVA-type models in Box S2 (Part 2 of 3).

```

### Model D MANCOVA
modD.lm <-lm(Ygc ~ Treatment + Epiphytes)

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
[1,]    2.8927     0.3551   8.145 8.97e-10 ***
[2,]   -1.6473     0.5515  -2.987 0.00498 **
[3,]   -0.5499     0.3956  -1.390 0.17284
---
Residual standard error: 1.193 on 37 degrees of freedom
Multiple R-squared:  0.6437, Adjusted R-squared:  0.6244
F-statistic: 33.42 on 2 and 37 DF, p-value: 5.122e-09

### Model E ANCOVA #1
modE.lm <- lm(Epiphytes ~ Treatment + Macroalgae + Seagrassdens)

Coefficients:
      Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept)  -1.1154         0.0000     0.2595  -4.298 0.000125 ***
Treatment      0.9761         0.7003     0.1503   6.496 1.52e-07 ***
Macroalgae    -0.1230        -0.1017     0.1318  -0.934 0.356638
Seagrassdens   0.6097         0.2355     0.2812   2.168 0.036818 *
---
Residual standard error: 0.4637 on 36 degrees of freedom
Multiple R-squared:  0.585, Adjusted R-squared:  0.5504
F-statistic: 16.91 on 3 and 36 DF, p-value: 5.082e-07

### Model F ANCOVA #2
modF.lm <- lm(Epiphytes ~ Treatment + Gammarids + Caprellids)

Coefficients:
      Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept) -0.11060         0.00000     0.28808  -0.384   0.703
Treatment    0.64326         0.46148     0.23670   2.718   0.010 *
Gammarids   -0.09928        -0.20144     0.09234  -1.075   0.289
Caprellids  -0.15824        -0.21899     0.09537  -1.659   0.106
---
Residual standard error: 0.4621 on 36 degrees of freedom
Multiple R-squared:  0.5878, Adjusted R-squared:  0.5535
F-statistic: 17.11 on 3 and 36 DF, p-value: 4.499e-07

```

Box S5. Results for ANOVA-type models in Box S2 (Part 3 of 3).

```
### Model G ANCOVA #3
modG.lm <- lm(Epiphytes ~ Treatment + Gammarids + Caprellids +
Macroalgae + Seagrassdens)

Coefficients:
              Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept)  -0.37790      0.00000      0.31629  -1.195 0.240436
Treatment      0.21825      0.15658      0.25558   0.854 0.399106
Gammarids     -0.26845     -0.54466      0.10181  -2.637 0.012524 *
Caprellids    -0.24635     -0.34092      0.08864  -2.779 0.008809 **
Macroalgae     0.36969      0.30572      0.17086   2.164 0.037607 *
Seagrassdens   0.99497      0.38422      0.26024   3.823 0.000536 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3953 on 34 degrees of freedom
Multiple R-squared:  0.7151, Adjusted R-squared:  0.6732
F-statistic: 17.07 on 5 and 34 DF, p-value: 1.964e-08
```

Box S6. R code for SE Models shown in Figure 7 using 'piecewiseSEM' package.

```
#####
##### piecewiseSEM modeling#####
#####
library(piecewiseSEM)

### Model A - a priori model
pw.modA <- psem(lm(Gammarids ~ Treatment, data=dat),
               lm(Caprellids ~ Treatment, data=dat),
               lm(Epiphytes ~ Gammarids + Caprellids + Macroalgae
                 + Seagrassdens, data=dat))

dSep(pw.modA)

### Model B.1 - add links 'Gammarids ~ Macroalgae'
### and 'Caprellids ~ Macroalgae based on modification indices
pw.modB.1 <- psem(lm(Gammarids ~ Treatment + Macroalgae, data=dat),
                 lm(Caprellids ~ Treatment + Macroalgae, data=dat),
                 lm(Epiphytes ~ Gammarids + Caprellids + Macroalgae
                   + Seagrassdens, data=dat))

dSep(pw.modB.1)

### Model B.final - final model, with 'Gammarids ~ Seagrassdens' added
pw.modB.final <- psem(lm(Gammarids ~ Treatment + Macroalgae
                       + Seagrassdens, data=dat),
                     lm(Caprellids ~ Treatment + Macroalgae, data=dat),
                     lm(Epiphytes ~ Gammarids + Caprellids + Macroalgae
                       + Seagrassdens, data=dat))

dSep(pw.modB.final)
summary(pw.modB.final)
```

Literature Cited

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Box S7. Select results from SEM code in Box S6: d-separation test results.

```
> dSep(pw.modA)
      Independ.Claim Test.Type DF Crit.Value      P.Value
1  Epiphytes ~ Treatment + ...      coef 34  0.8539570 3.991056e-01
2  Gammarids ~ Macroalgae + ...      coef 37  5.3596554 4.622501e-06 ***
3  Caprellids ~ Macroalgae + ...      coef 37  3.6578135 7.870962e-04 ***
4  Gammarids ~ Seagrassdens + ...      coef 37  0.9795621 3.336642e-01
5  Caprellids ~ Seagrassdens + ...      coef 37  0.6943843 4.917786e-01
6  Caprellids ~ Gammarids + ...      coef 37  2.8369984 7.346110e-03 **

> dSep(pw.modB.1)
      Independ.Claim Test.Type DF Crit.Value      P.Value
1  Epiphytes ~ Treatment + ...      coef 34  0.8539570 0.39910564
2  Gammarids ~ Seagrassdens + ...      coef 36  2.2045815 0.03395976 *
3  Caprellids ~ Seagrassdens + ...      coef 36  1.3723662 0.17844018
4  Caprellids ~ Gammarids + ...      coef 36  0.7714145 0.44549545

dSep(pw.modB.final)
      Independ.Claim Test.Type DF Crit.Value      P.Value
1  Epiphytes ~ Treatment + ...      coef 34  0.8539570 0.3991056
2  Caprellids ~ Seagrassdens + ...      coef 36  1.3723662 0.1784402
3  Caprellids ~ Gammarids + ...      coef 35  0.3277904 0.7450229

Global goodness-of-fit:
  Fisher's C = 5.873 with P-value = 0.438
```

Box S8. Results from SEM code in Box S6: final model coefficients and R-squares.

```
> # request range-standardized coefficients
> coefs(pw.modB.final, standardize = "range")
      Response      Predictor Estimate Std.Error DF Crit.Value P.Value Std.Estimate
1  Gammarids      Treatment    -2.0993   0.2100 36    -9.9952  0.0000    -0.3933 ***
2  Gammarids      Macroalgae     1.0857   0.1842 36     5.8948  0.0000     0.4562 ***
3  Gammarids Seagrassdens     0.8665   0.3930 36     2.2046  0.0340     0.2283 *
4  Caprellids      Treatment    -0.7837   0.2441 37    -3.2112  0.0027    -0.2500 **
5  Caprellids      Macroalgae     0.7745   0.2117 37     3.6578  0.0008     0.5540 ***
6  Epiphytes      Gammarids     -0.3398   0.0579 35    -5.8707  0.0000    -0.7029 ***
7  Epiphytes      Caprellids    -0.2636   0.0860 35    -3.0652  0.0042    -0.3202 **
8  Epiphytes      Macroalgae     0.4572   0.1362 35     3.3567  0.0019     0.3974 **
9  Epiphytes Seagrassdens     1.0690   0.2444 35     4.3738  0.0001     0.5825 ***

> # request r-squares
> rsquared(pw.modB.final)
      Response      family      link method R.squared
1  Gammarids gaussian identity    none 0.8030290
2  Caprellids gaussian identity    none 0.4122171
3  Epiphytes gaussian identity    none 0.7090204
```

Box S9. Bonus code for those that wish to be convinced that conditioning on a collider induces correlation between the parent variables.

```
#####  
##### CONDITIONING ON A COLLIDER SIMULATION STUDIES  
##### Three-variable model A-> M <- B  
#####  
### Simulate the Data  
n = 10000  
set.seed(1)  
  
### create independent exogenous predictors  
A <- rnorm(n, 0, 0.25)  
B <- rnorm(n, 0, 0.30)  
  
# coefficients for model  
bexd = 0.60  
bexp = -0.50  
bcmp = -0.55  
bend = -0.35  
benc = -0.45  
  
# collider is M variable  
M <- 0.60*A + 0.30*B + rnorm(n, 0, 0.20)  
  
dat <- data.frame(A, B, M)  
print(cor(dat), digits=4) # note A and B effectively uncorrelated  
  
##### Regression Models  
### Model 1: M ~ A + B  
summary(lm(M ~ A + B))  
  
### Model 2: A ~ M + B - Illustrates that conditioning A on M induces (non-  
causal) association between A and B  
summary(lm(A ~ M + B))  
  
##### Two-Stage Conditioning  
## Stage 1 conditioning of A and B on M  
A.res1 <- resid(lm(A ~ M))  
B.res1 <- resid(lm(B ~ M))  
  
## Residual Correlation Illustrates conditioning A and/or B on M induces  
correlation between A.res and B.res  
## this extends to induced correlations between A.res and B, and A and B.res  
cor.test(A.res1, B.res1)  
cor.test(A.res1, B)  
cor.test(A, B.res1)
```