USGS
science for a changing world

Modeling Interactions

Jim Grace

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

1

This module illustrates the inclusion of interaction terms in models and the summarization of their effects using composites. The approach used here can be contrasted with the handling of interactions using the multigroup approach.

A general citation for this material is

Grace, J.B. and Bollen, KA. 2008. Representing general theoretical concepts in structural equation models: the role of composite variables. Environmental and Ecological Statistics 15:191-213.

(http://www.odum.unc.edu/content/pdf/Bollen%20Grace%20Bollen%20(preprint%202008)%20Environ%20and%20Ecol%20Stats.**pdf**)

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 17.02.08.

Source: https://www.usgs.gov/centers/wetland-and-aquatic-research-center/science/quantitative-analysis-using-structural-equation

How might we include interactive effects within a model?

By "interactive effects", we mean non-additive relations where one predictor affects the influence of another.

Note, in the equation below, y_3 is influenced by x_1, x_2, and the product of x_1 and x_2.

$$y_3 = \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 (x_1 * x_2)$$

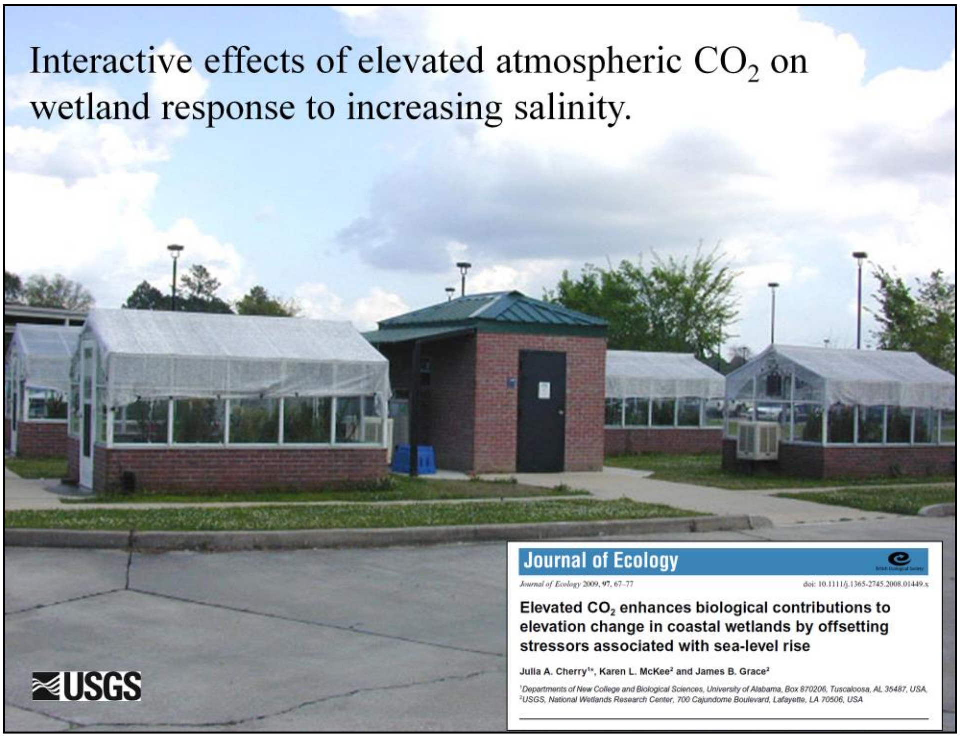
That product term ($x_1 * x_2$) is representing the combined (interactive) effect.

Note, an alternative approach to interactions when predictors are categorical is to use multigroup modeling. The approach here can be used with continous variables.

USGS 2

The mathematics of interactions is similar to that of polynomial regression.

Note that in contrast to formal multigroup analysis, here we can deal with interactions involving continuous or semi-continuous variables.



CO2 control greenhouses were used for this study in a split-plot design. Classical ANOVA analyses were performed first. The split-plot feature was handled in the classical analyses, but is ignored here in the illustration.

The example used here was extracted from:

Cherry, J.A., McKee, K.L., and Grace, J.B. 2009. Elevated CO2 enhances biological contributions to elevation change in coastal wetlands by offsetting stressors associated with sea-level rise. *Journal of Ecology* 97:67-77.

This article was featured in Nature News April 9, 2009, featured in Nature Climate Change Research Highlights May 5, 2009, and was a USGS Science Newsroom Pick. http://www.nature.com/climate/2009/0905/full/climate.2009.32.html

Study Design

<u>Treatments</u>:
- CO_2 (ambient = 380 ppm and elevated = 720 ppm)
- salinity (0, 5, 10, 15, and 20‰ sea salts)
- flooding (drained, intermittently flooded, and flooded)

<u>Responses</u>:
- production by C_3 and C_4 species
- combined root production
- sediment elevation change

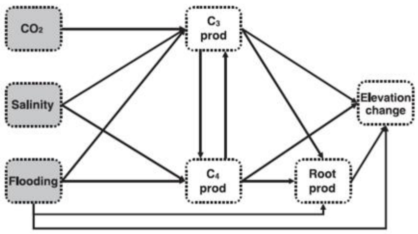
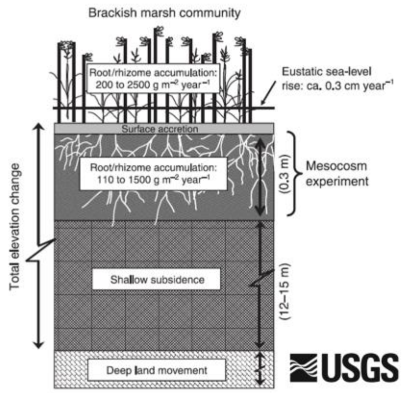


Fig. 2. Conceptual construct model presenting hypothesized direct and indirect effects of treatments (shaded boxes) on biotic variables and elevation change.



We had an a priori meta-model for this analysis. It was actually a little more involved than this, and was simplified as the soil chemistry data was uninformative.

The biology in this case is that the plant builds soil with their organic material, allowing natural marshes to keep pace with rising sea-levels.

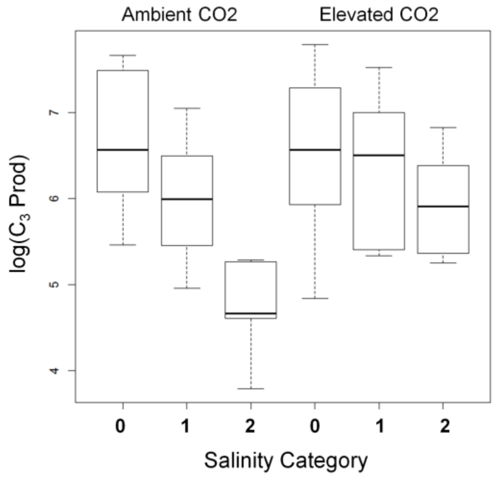
C3 species was *Schoenoplectus americanus*.

C4 species was *Spartina patens*.

In this example, we omit the flooding effect and simplify the salinity variable to 3 levels (0, 1, and 2 for low, medium, and high).

Note also that the data were adjusted slightly so the simplified analysis results are consistent with those from the full dataset.

Classical analyses and inspections revealed an interactive effect of CO_2 on plant response to salinity.



USGS

Ability of C3 plant to tolerate high salinities enhanced by CO_2. No effect of CO_2 at low salinity.

5

It is critical that you identify the nature of the interactive effect (usually through visualizations) in order to support the interpretation. This figure shows how production drops off faster at higher salinities in ambient CO2. So, we answer the original question, “Does elevated CO2 enhance production of the C3 species?” with “Only at high salinities, where it appears to increase salinity tolerance.

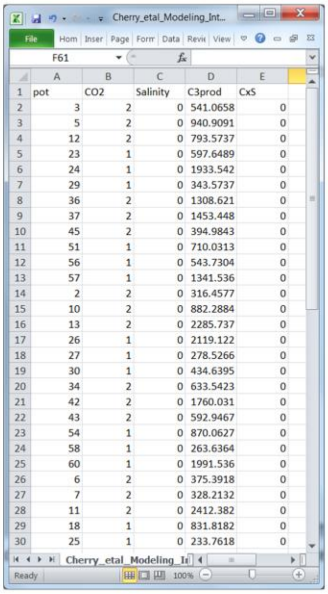
How do we model this interactive effect?

The data (note data provided in notes of this slide).

Note we use “dummy variable” coding for CO_2, $CO_2 = (0,1)$,

while salinity is ordered categorical (0, 1, 2)

Interaction variable, CxS, is the simple product of CO_2 and salinity level.

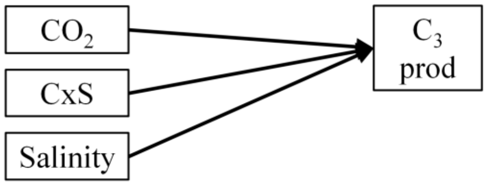


	A	B	C	D	E
1	pot	CO2	Salinity	C3prod	CxS
2	3	2	0	541.0658	0
3	5	2	0	940.9091	0
4	12	2	0	793.5737	0
5	23	1	0	597.6489	0
6	24	1	0	1933.542	0
7	29	1	0	343.5737	0
8	36	2	0	1308.621	0
9	37	2	0	1453.448	0
10	45	2	0	394.9843	0
11	51	1	0	710.0313	0
12	56	1	0	543.7304	0
13	57	1	0	1341.536	0
14	2	2	0	316.4577	0
15	10	2	0	882.2884	0
16	13	2	0	2285.737	0
17	26	1	0	2119.122	0
18	27	1	0	278.5266	0
19	30	1	0	434.6395	0
20	34	2	0	633.5423	0
21	42	2	0	1760.031	0
22	43	2	0	592.9467	0
23	54	1	0	870.0627	0
24	58	1	0	263.6364	0
25	60	1	0	1991.536	0
26	6	2	0	375.3918	0
27	7	2	0	328.2132	0
28	11	2	0	2412.382	0
29	18	1	0	831.8182	0
30	25	1	0	233.7618	0

USGS

Raw data. Semi-colons are end of line markers.
pot,CO2,Salinity,C3prod,CxS;
3,2,0,541.0658307,0; 5,2,0,940.9090909,0; 12,2,0,793.5736677,0;
23,1,0,597.6489028,0; 24,1,0,1933.54232,0; 29,1,0,343.5736677,0;
36,2,0,1308.62069,0; 37,2,0,1453.448276,0; 45,2,0,394.984326,0;
51,1,0,710.031348,0; 56,1,0,543.7304075,0; 57,1,0,1341.53605,0;
2,2,0,316.4576803,0; 10,2,0,882.2884013,0; 13,2,0,2285.736677,0;
26,1,0,2119.122257,0; 27,1,0,278.5266458,0; 30,1,0,434.6394984,0;
34,2,0,633.5423197,0; 42,2,0,1760.031348,0; 43,2,0,592.9467085,0;
54,1,0,870.0626959,0; 58,1,0,263.6363636,0; 60,1,0,1991.53605,0;
6,2,0,375.3918495,0; 7,2,0,328.2131661,0; 11,2,0,2412.382445,0;
18,1,0,831.8181818,0; 25,1,0,233.7617555,0; 28,1,0,1876.018809,0;
33,2,0,2201.724138,0; 35,2,0,125.0783699,0; 44,2,0,249.6865204,0;
48,1,0,1785.109718,0; 9,1,0,565.5172414,0; 52,1,1,398.1191223,1;
1,2,1,644.0438871,2; 8,2,1,1844.043887,2; 15,2,1,221.3166144,2;
16,1,1,1147.805643,1; 20,1,1,187.6175549,1; 22,1,1,290.1253918,1;
32,2,1,690.9090909,2; 39,2,1,1090.438871,2; 40,2,1,206.2695925,2;
46,1,1,432.6018809,1; 50,1,1,141.3793103,1; 59,1,1,1008.777429,1;
4,2,2,589.8119122,4; 9,2,2,271.4733542,4; 14,2,2,212.539185,4;
17,1,2,110.5015674,2; 19,1,2,43.26018809,2; 21,1,2,192.3197492,2;
31,2,2,499.2163009,4; 38,2,2,190.2821317,4; 41,2,2,916.4576803,4;
47,1,2,99.05956113,2; 53,1,2,196.5517241,2; 55,1,2,100,2

Modeling the interaction: Step 1.




```
# specify model
mod.1 <- 'ln.C3prod ~ CO2 + Salinity + CxS'

# fit model
mod.1.fit <- sem(mod.1, data=dat2)

# request output
summary(mod.1.fit, rsq=T)
```

USGS 7

As is typical in nonlinear modeling where a composite will be used, we first run the model without the composite.

Note that we log transformed the responses in this example, which normalized errors.

Results - Step 1.

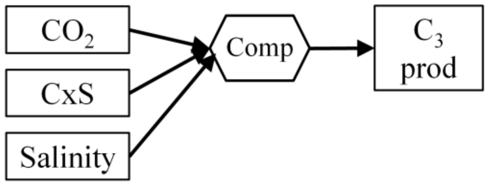
```
lavaan (0.5-15) converged normally after  1 iterations

  Number of observations                            60
  Estimator                                         ML
  Minimum Function Test Statistic                0.000
  Degrees of freedom                                 0
  P-value (Chi-square)                           1.000

                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  ln.C3prod ~
    CO2              -0.084    0.242   -0.345    0.730
    Salinity         -1.529    0.381   -4.016    0.000
    CxS               0.612    0.240    2.547    0.011

R-Square:
    ln.C3prod         0.367
```

USGS

8

Results for the non-composited model show significant effect of salinity and the interaction. We retain all three factors in the model for generality.

Modeling the interaction: Step 2.




```
# specify model
mod.2 <- 'Comp <~ CO2 + 1*Salinity + CxS
          ln.C3prod ~ Comp'

# fit model
mod.2.fit <- sem(mod.2, data=dat2)

# request output
summary(mod.2.fit, standardized=T, rsq=T)
```

USGS

9

If you are not familiar with composites, you should check out the module “Composites and Formative Indicators” first.

Recall, lavaan has a special operator for composites “<~”.

We could also create the composite scores by hand and then model.

In this case, the model had trouble converging when “1*” was applied to CO2, but was fine when specified as above (with “1*” times salinity).

Results - Step 2.

	Estimate	Std.err	Z-value	P(>|z|)	Std.all
Composites:					
Comp <~					
CO2	0.055	0.151	0.363	0.717	0.074
Salinity	1.000				2.156
CxS	-0.400	0.070	-5.713	0.000	-1.401
Regressions:					
ln.C3prod ~					
Comp	-1.529	0.381	-4.016	0.000	-0.606
R-Square:					
ln.C3prod	0.367				

USGS

10

Results for the composite model do not produce interpretable raw coefficients (Estimates). The combined effect of the predictors is the std.all value for the regression (0.606). I would not put too much stock in the sign of that value, as shown in the next slide.

Modeling the interaction: Step 2 – alternative approach.

```
## Model 3 - create composite by hand
# get parameters from uncomposited model
summary(mod.1.fit)

# compute composite scores
comp.hand <- 6.762 -0.084*CO2 -1.529*Salinity
+0.612*CxS

# add variable to data set
dat2$Comp.hand <- comp.hand

# specify model
mod.3 <- 'ln.C3prod ~ Comp.hand'

# fit model
mod.3.fit <- sem(mod.3, data=dat2)
```

USGS

11

Note, the intercept that shows up when we compute the composite comes from a more complete print out of results than is shown on slide 8. You can request meanstructure=TRUE in lavaan to get the intercepts to print.

Results - Step 2.

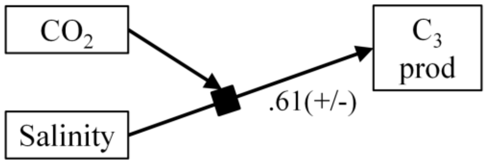
```
Regressions:
                   Estimate  Std.Err  Z-value  P(>|z|)   Std.all
  ln.C3prod ~
    Comp.hand         1.000    0.169    5.902    0.000    0.606

Variances:
                   Estimate  Std.Err  Z-value  P(>|z|)   Std.all
    ln.C3prod         0.552    0.101    5.477    0.000     0.633

R-Square:
                   Estimate
    ln.C3prod         0.367
```

USGS

12

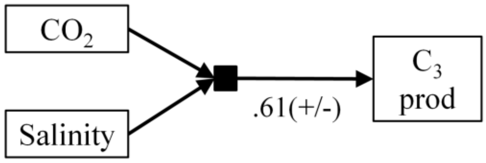
Results for the composite model do not produce interpretable raw coefficients (Estimates). The combined effect of the predictors is the std.all value for the regression (0.606). I would not put too much stock in the sign of that value, as shown in the next slide.

Here is how we chose to represent the interaction graphically.



Here we point the arrow from CO_2 to the effect of salinity to support the interpretation that CO_2 is modifying salinity effect.

Here is an alternative representation.



If we just wanted to say there was an interaction, we might present this way.

13

Rather than show the composite variable explicitly in this example, we chose to show in a simpler form.

Note we generally do not show the parameters for paths that make up the composite, only its net effect, and always in standardized form.