Modeling Interactions

Jim Grace

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

This module considers how to model interactive effects and illustrates how to include interaction terms in structural equation models. The approach used here can be contrasted with the handling of interactions using the multigroup approach, which I currently discuss in the tutorial "Lavaan Options" at www.nwrc.usgs.gov/SEM.

A general citation for this material is

Grace, J.B. 2006. Structural Equation Modeling and Natural Systems. Cambridge University Press. Cambridge, UK.

Notes: IP-064930; Support provided by the USGS Climate & Land Use R&D and Ecosystems Programs. I would like to acknowledge formal review of this material by Chih-Han Chang, John Hopkins University and Mary Donovan, University of Hawaii. Many helpful informal comments have contributed to the final version of this presentation. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Questions about this material can be sent to sem@usgs.gov.

Last revised 15.06.17.

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.

One approach is to include an interaction term in the model. Note, in the equation below, y_3 is influenced by x_1, x_2, and the product of x_1 and x_2 (x_1*x_2), which represents the combined (interactive) effect

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

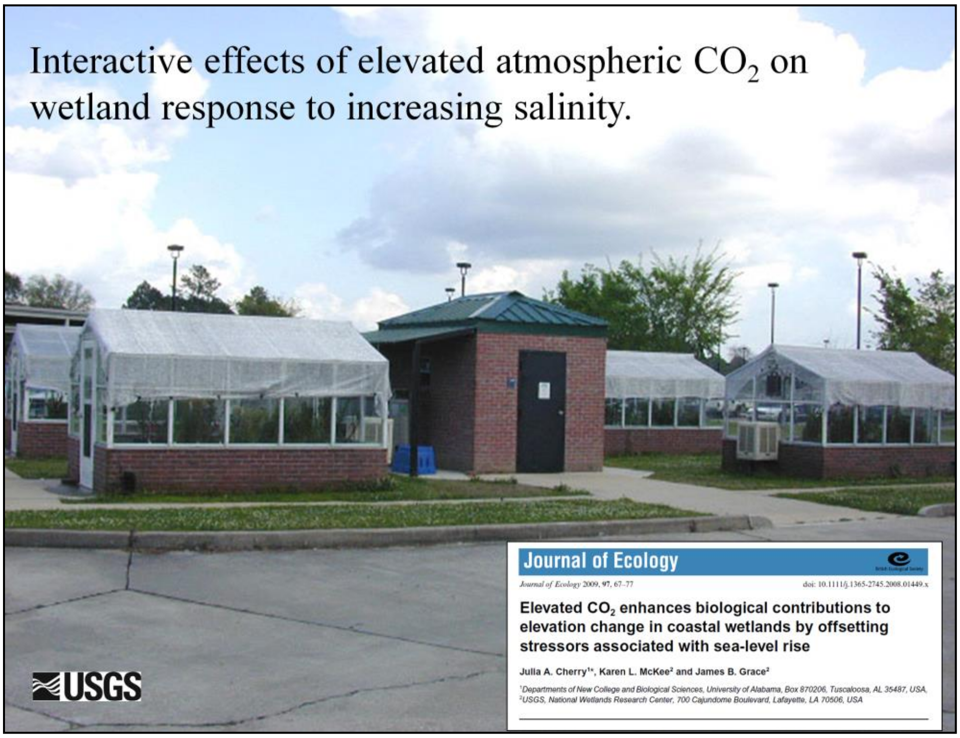
Alternative approach to interactions when predictors are categorical is to use multigroup modeling. The approach here can be used with continous variables.

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.



CO_2 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 CO_2 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

Treatments:

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

Responses:

- production by C_3 and C_4 species
- combined root production
- sediment elevation change

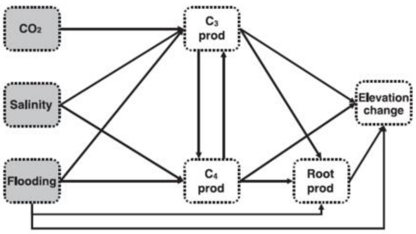
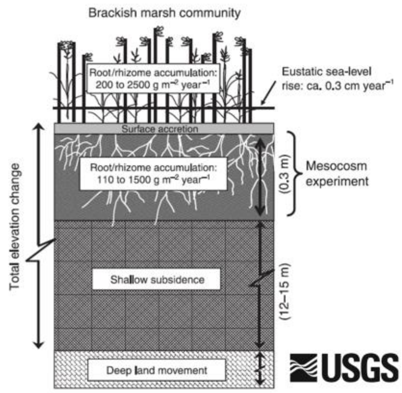


Fig. 2. Conceptual construct model presenting hypothesized direct and indirect effects of treatments (shaded boxes) on biotic variables and elevation change.



USGS

We had an a priori meta-model for this analysis. It was actually a little more involved than the one shown here, and was simplified because 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.

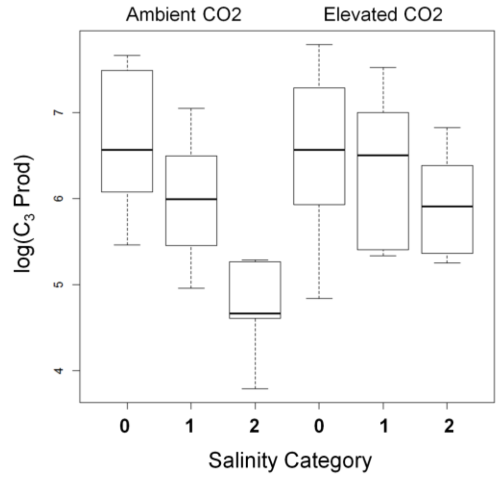
C_3 species was *Schoenoplectus americanus*.

C_4 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 simplified for this illustration, and as a result, the results presented here are slightly different from those in the original paper.

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 CO_2. So, we answer the original question,

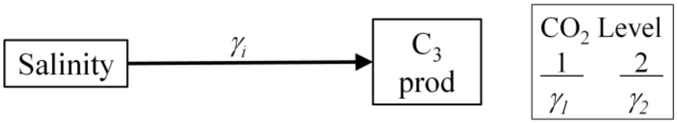
"Does elevated CO_2 enhance production of the C_3 species?"

with

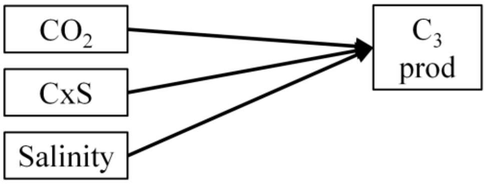
"Only at high salinities, where it appears to increase salinity tolerance."

Approaches to modeling an interaction.

Approach #1: Multi-group approach (refer to illustration elsewhere).



Approach #2: Interaction variable approach (approach illustrated here).



USGS

6

This slide contrast two different approaches to dealing with interactive effects.

The first of these is a multi-group approach. Here we develop a separate model for each CO_2 level so the gamma effect of salinity on C_3 production can be different between CO_2 levels. Multigroup modeling is currently illustrated as one of the topics in the tutorial “Additional Lavaan Options”.

The second approach to modeling interactions involves bringing in an “interaction” variable, which is usually just the product of the two other terms (multiply CO_2 times Salinity). This is the approach illustrated in this tutorial.

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.

Cherry_etal_Modeling_Int...

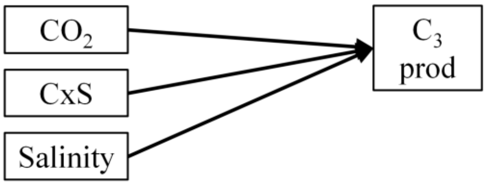
	pot	CO2	Salinity	C3prod	CxS
2	3	2	0	541.0658	0
3	5	2	0	940.9091	0
4	12	2	0	793.5737	0
5	23	1	0	597.6489	0
6	24	1	0	1933.542	0
7	29	1	0	343.5737	0
8	36	2	0	1308.621	0
9	37	2	0	1453.448	0
10	45	2	0	394.9843	0
11	51	1	0	710.0313	0
12	56	1	0	543.7304	0
13	57	1	0	1341.536	0
14	2	2	0	316.4577	0
15	10	2	0	882.2884	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

7

Raw data are provided in a supporting document "SEM-Modeling Interactions_Cherry_etal_data.csv"

Modeling the interaction: Step 1 – Evaluation of elements.




```
# 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

8

I recommend a two-step approach when modeling interactions. The first step, which simply includes the interaction variable (CxS in this case), allows us to test whether there is evidence for a significant interaction term. The second step, described in later slides, allows us to create a measure of the strength of the overall interactive effect. My approach to that second step is to include a composite. Composite variable modeling is described in tutorials with that key word in the title and should be consulted before proceeding in this tutorial.

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

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
```

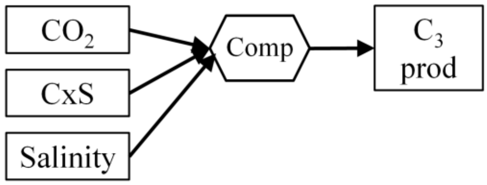
USGS

9

Results for the non-composited model show significant contributions by the salinity and salinity-by-CO2 interaction terms. We retain all three terms in the model for generality, because there is a significant interation.

Note also that in this small model, there are zero degrees of freedom and "perfect fit" (i.e., p-value = 1.0).

Modeling the interaction: Step 2 – The collective effect.




```
# 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

10

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). It is sometimes the case that when we make an arbitrary guess at the initial value for a parameter, like 1.0, convergence fails because the true value is very far from that (like -1.0). There are numerous remedies for this. One is to simply set the value for a different parameter in the model (in this case I used Salinity instead of CO2). Another approach would be to get the exact value from the non-composited model. If we had used

-0.084*CO2

which is the estimated parameter found in the previous slide, the model would have converged.

The tutorials on Modeling with Composites give more illustrations of this process.

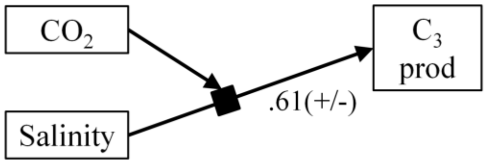
Results - Step 2.

```
                 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

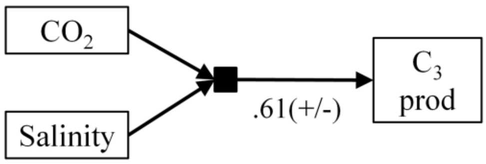
As with the model in step 1, individual raw coefficients ("Estimates") are not interpretable in the usual fashion because the parameters work together to represent the interaction. The combined effect of the predictors is the std.all value for the regression (0.606), which represents the collective effect of the interaction in standardized metric. I would not put too much stock in the sign of that value because this is not a linear relationship.

Here is how we chose to represent the interaction graphically.



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

Here is an alternative representation.



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

12

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.

More information can be found at
http://www.nwrc.usgs.gov/SEM

USGS

I hope this overview has been useful. For more information, go to our webpage or search for examples involving your subject of interest. Questions and comments can be sent to sem@usgs.gov. Please note I cannot guarantee responses to individual inquiries, but will definitely incorporate suggestions in future tutorials. – Thanks!