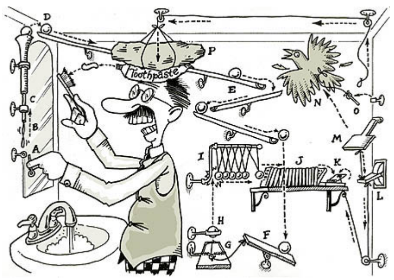
Taking Apart the Pieces of SEM:

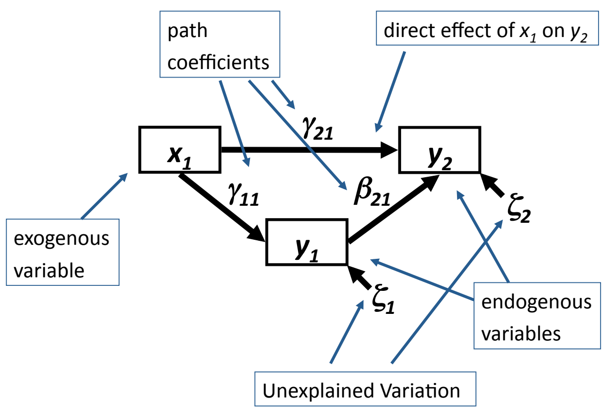
Path Diagrams, Path Coefficients, and Model Construction



Outline of the Basics

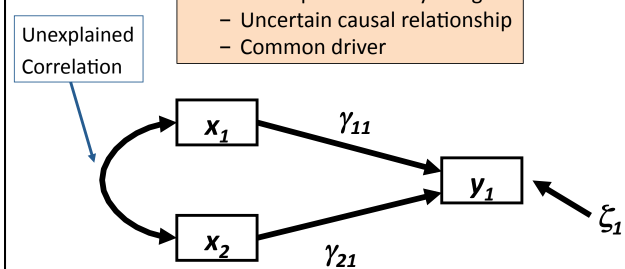
1. Terminology and housekeeping
   – Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data

Basic Terminology



Representing Unexplained Correlation

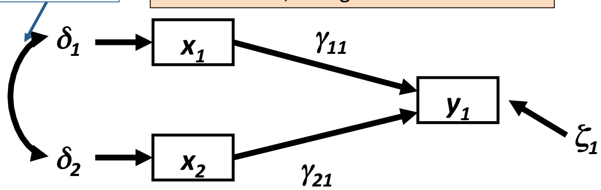
- This is multiple regression!
- <-> Can represent many things
  - Uncertain causal relationship
  - Common driver



Representing Unexplained Correlation

- Really, a correlation between residual variance
- Convention: show correlation between endogenous errors but not exogenous – still there, though!

Unexplained Correlation



Path Diagrams and Causality

1. Sewell Wright's intention was to describe (1) causal relationships and (2) strength of associations.

2. Explicit consideration of causation languished for 70 years Judea Pearl and others have revived it in the science of artificial intelligence.

3. Pearl argues that regular mathematics is unable to express the needed expressions to represent causation. "=" versus "→"

Pearl, J. 2009. Causality. Cambridge University Press (2nd ed)

Practical Criteria for Supporting Causal Assumptions

1. A manipulation of *x* would result in a subsequent change in the values of *y*

2. OR the values of *x* serve as <u>indicators</u> for the values of *x* that existed when effects on *y* were being generated.

3. Models are properly specified to extract causal information

Can my model be fit? *Identification*

$3 = a + b$
$4 = 2a + b$

a and b have unique solutions

Identified

$3 = a + b + c$
$4 = 2a + b + 3c$

a, b, and c have no unique solution

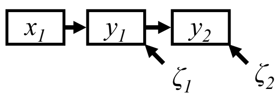
Underidentified

$3 = a + b$
$4 = 2a + b$
$7 = 3b + a$

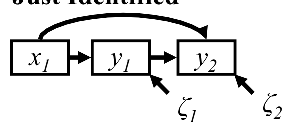
Overidentified

Structure of Models and Identification

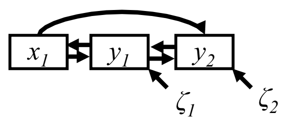
Unsaturated or Overidentified



Saturated or Just Identified

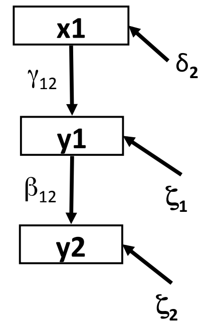


Oversaturated or Underidentified



Identification in SEM

of Parameters v. # of Variables



We need enough information to estimate unique parameters

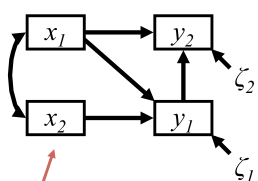
params ≤ # variables * (# variables+1)/2

T-Rule: $t \leq p(p+1)/2$

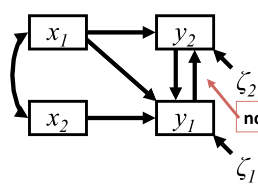
t=# params, p=# variables

Feedbacks

Recursive



Nonrecursive

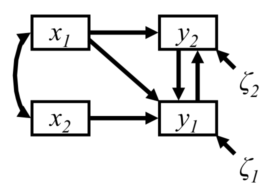


nonrecursive

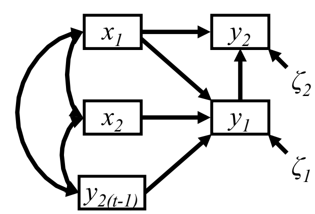
Recursive (each item in a series is directly determined by the preceding item).

Breaking Feedbacks using Time

Nonrecursive

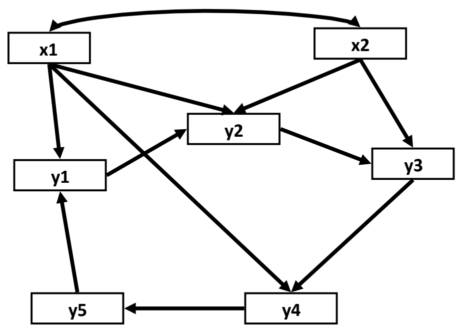


Recursive



- Feedback assumes equilibrium
- Often concerned with dynamic processes
- Solution: Build in the lag

What is this?

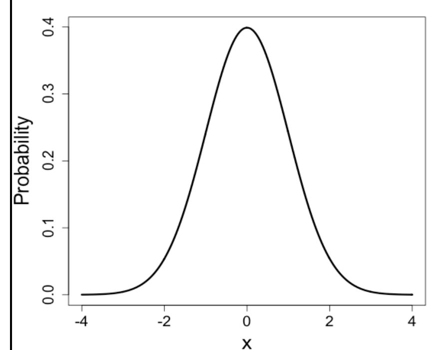


Questions?

Outline of the Basics

1. Terminology and housekeeping
   – Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data

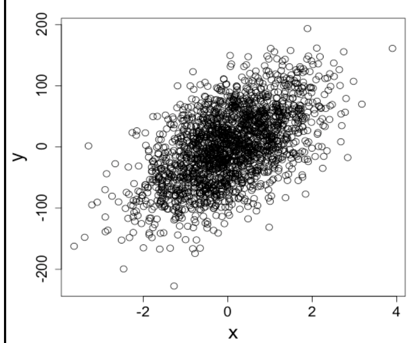
Variance of a Normal Distribution



$$VAR_x = \frac{\sum (x - \overline{x})^2}{n-1}$$

$$SD_x = \sqrt{VAR_x}$$

Covariance and Correlation



$$COV_{xy} = \frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{n-1}$$

$$r_{xy} = \frac{COV_{xy}}{SD_x SD_y}$$

Covariances and Correlations

We often use covariances to fit models, but standardized covariances – i.e. correlations – for interpretation.

$$r_{xy} = \frac{COV_{xy}}{SD_x \times SD_y}$$

r_{xy} = correlation/ std covariance
COV = covariances (ave cross product)
SD = std. dev. = *VAR*^1/2^

Raw Covariance Matrix

	x_1	x_2	y_1
x_1	**3.2**		
x_2	**0.65**	**0.8**	
y_1	**1.98**	**1.19**	**4.8**

variance

covariance

Standardized Covariance Matrix

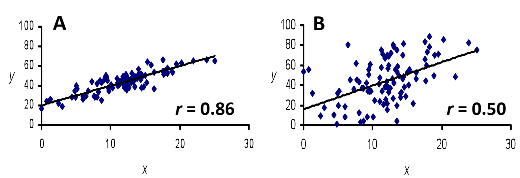
	x_1	x_2	y_1
x_1	**1.0**		
x_2	**0.40**	**1.0**	
y_1	**0.50**	**0.60**	**1.0**

correlation

Correlations and Regression

Correlations contain less information, but are directly comparable

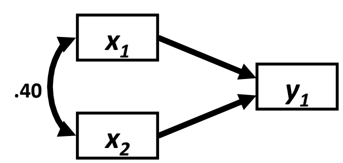
Example: these regressions having same unstandardized slope and intercept



Regression: y = a + bx

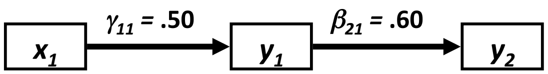
Standardized Coefficients: r = b * sd(x)/sd(y)

First Rule of Path Coefficients: the path coefficients for unanalyzed relationships (curved arrows) between exogenous variables are simply the correlations (standardized form) or covariances (unstandardized form).



	x_1	x_2	y_1
x_1	**1.0**		
x_2	**0.40**	**1.0**	
y_1	**0.50**	**0.60**	**1.0**

Second Rule of Path Coefficients: when variables are connected by a single causal path, the path coefficient is simply the standardized or unstandardized regression coefficient (note that a standardized regression coefficient = a simple correlation.)

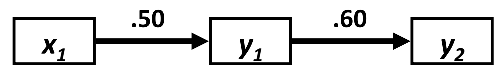


	x_1	y_1	y_2
x_1	1.0		
y_1	0.50	1.0	
y_2	0.30	0.60	1.0

γ (gamma) used to represent effect of exogenous on endogenous.

β (beta) used to represent effect of endogenous on endogenous.

Third Rule of Path Coefficients: strength of a compound path is the product of the coefficients along the path.

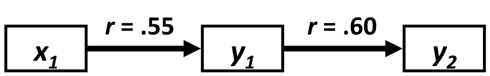


Thus, in this example the effect of x_1 on y_2 = 0.5 x 0.6 = 0.30

Since the strength of the indirect path from x_1 to y_2 equals the correlation between x_1 and y_2, we say x_1 and y_2 are **conditionally independent**.

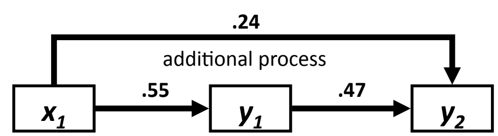
What does it mean when two separated variables are not conditionally independent?

	x_1	y_1	y_2
x_1	1.0		
y_1	0.55	1.0	
y_2	0.50	0.60	1.0



0.55 x 0.60 = 0.33, which is not equal to 0.50

The inequality implies that the true model is

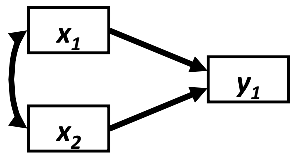


Fourth Rule of Path Coefficients: when variables are connected by more than one causal pathway, the path coefficients are "partial" regression coefficients.

Which pairs of variables are connected by two causal paths?

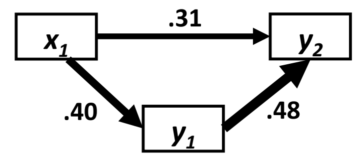
answer: x_1 and y_2 (obvious one), but also y_1 and y_2, which are connected by the joint influence of x_1 on both of them.

And for another case:



A case of shared causal influence: the unanalyzed relation between x_1 and x_2 represents the effects of an unspecified joint causal process. Therefore, x_1 and y_1 are connected by two causal paths. x_2 and y_1 likewise.

How to Interpret Partial Path Coefficients: The Concept of Statistical Control

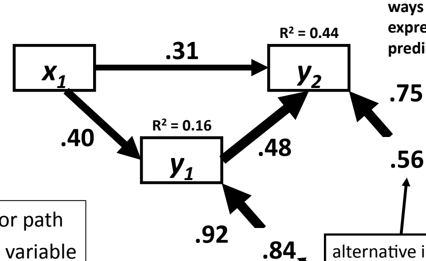


The effect of y_1 on y_2 is controlled for the joint effects of x_1.

With all other variables in model held to their means, how much does a response variable change when a predictor is varied?

Fifth Rule of Path Coefficients: paths from error variables represent prediction error (influences from other forces).

three different ways of expressing prediction error.

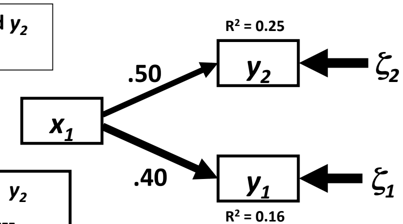


equation for path from error variable

$$= \sqrt{1 - R^2_{y_i}}$$

alternative is to show values for zetas, which = $1\text{-}R^2$

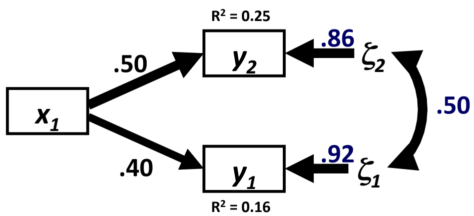
Now, imagine y_1 and y_2 are joint responses



	x_1	y_1	y_2
x_1	1.0		
y_1	0.40	1.0	
y_2	0.50	0.60	1.0

Implied correlation between y_1 and y_2 = 0.50 x 0.40 = 0.20.

Sixth Rule of Path Coefficients: unanalyzed residual correlations between endogenous variables are partial correlations or covariances.

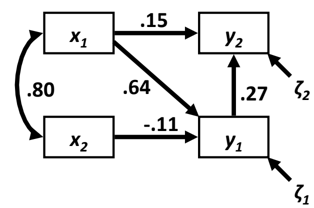


the partial correlation between y_1 and y_2 is typically represented as a **correlated error term**

This implies that some other factor is influencing y_1 and y_2

Note that total correlation between y_1 and y_2 =
0.50 x 0.40 + 0.86 x 0.50 x 0.92 = 0.60 (the observed corr)

Seventh Rule of Path Coefficients: **total effect** one variable has on another equals the sum of its direct and indirect effects.



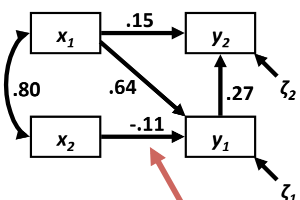
Total Effects:

	x_1	x_2	y_1
y_1	0.64	-0.11	---
y_2	0.32	-0.03	0.27

Eighth Rule of Path Coefficients:
sum of all pathways between two variables (directed and undirected) equals the correlation/covariance.

note: correlation between x_1 and y_1 = 0.55, which equals 0.64 - 0.80*0.11

Suppression Effect - when presence of another variable causes path coefficient to strongly differ from bivariate correlation.



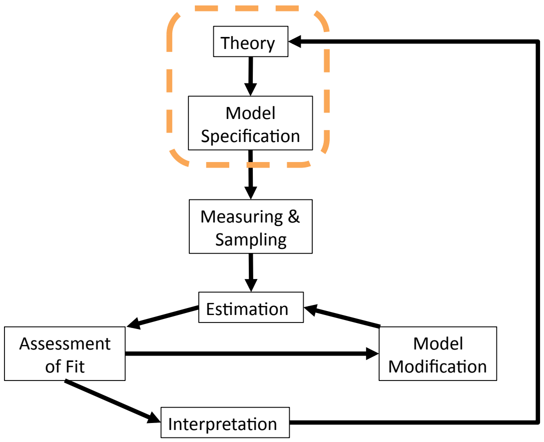
	x_1	x_2	y_1	y_2
x_1	1.0			
x_2	0.80	1.0		
y_1	0.55	0.40	1.0	
y_2	0.30	0.23	0.35	1.0

path coefficient for x_2 to y_1 very different from correlation,
(results from overwhelming influence from x_1.)

Outline of the Basics

1. Terminology and housekeeping
   – Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data

SEM as a Unifying Process



After Grace & Bollen 2005

The Continuum of SEM

structural equation modeling

exploratory/ model-building applications → **confirmatory/ hypothesis-testing/ model-comparison applications**

It all starts with an underlying model!

What are the goals of the analysis?

Purpose of modeling effort:
- discovery?
- testing hypotheses?
- making predictions?

Focus of modeling effort:
- driver focused?
- response focused?
- mediation focused?
- theory testing focused?

Span of inference:
- doing inferential estimation?
- learning about processes?

The Ebb & Flow of Model Building

- Start with big ideas and basic theory
- Focus ideas on a targeted area
- Expand conceptual model to encompass the details of the problem
  - Be thorough, otherwise you may miss important elements of suppression or confounding variables
- Prune unnecessary details
- Confront your model with data and expand and contract it as needed…

The Structural Equation Meta-Model (SEMM)

Concept A

Causal Process 1

Concept C

Causal Process 2

Causal Process 3

Concept B

My Research Program as a Meta-Model

Food Web Structure

Ecosystem Function

Environmental Change

Targeting Your Question

Food Web Structure

Ecosystem Function

Environmental Change

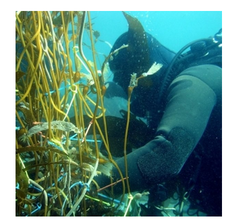
Targeting Your Question

Focused on driver & Mediator

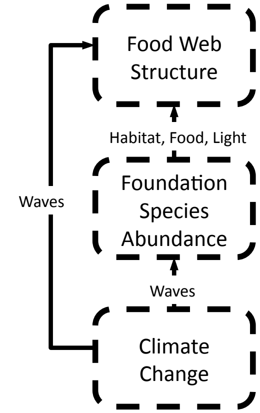
Food Web Structure

Foundation Species Abundance

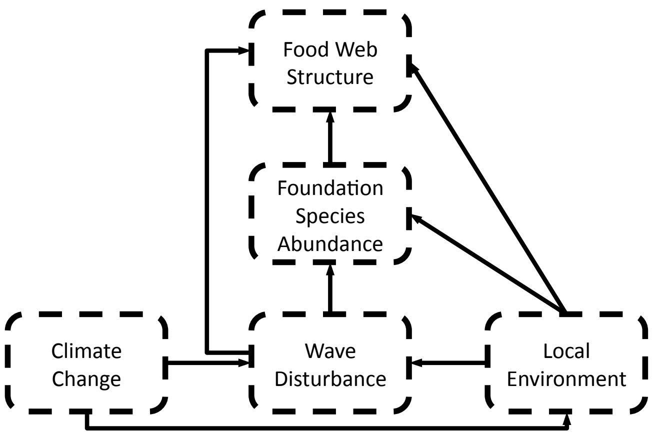
Climate Change



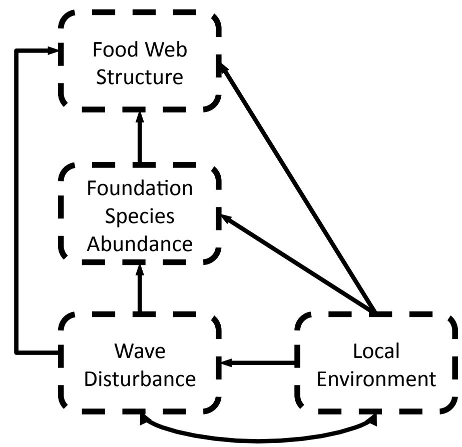
Identify Relevant Processes



Take a Step Back…



Remove Drivers Not of Interest



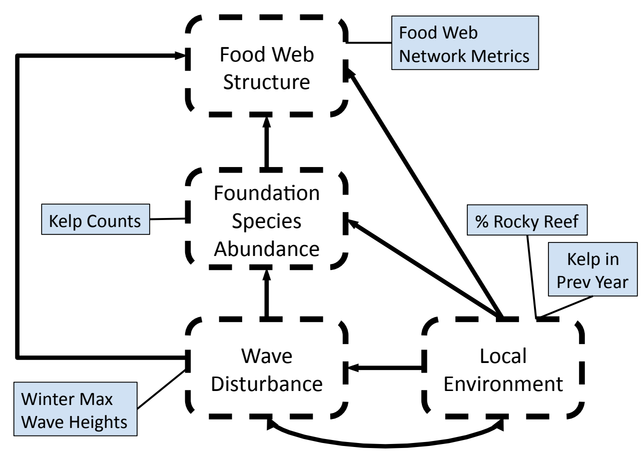
Meta-Model Your Research

Outline of the Basics

1. Terminology and housekeeping
   – Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data

Sampling of Rocky Reefs 2000-2009

Matching Data to Concepts

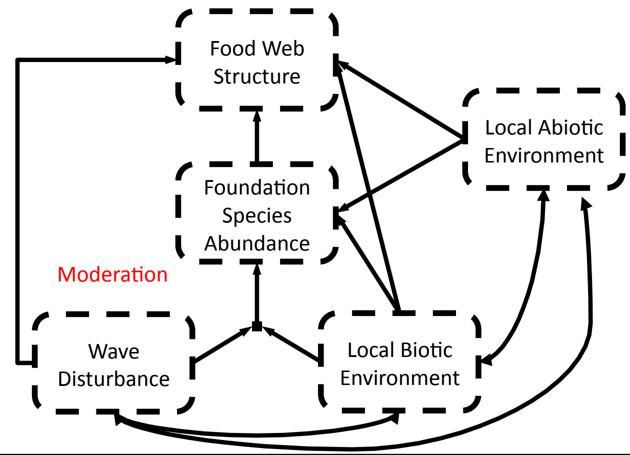


Confront Model with Natural History

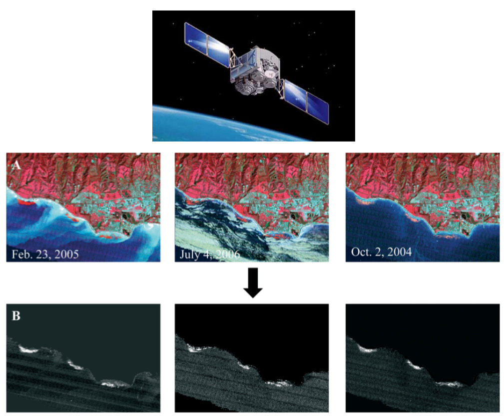
1. Kelp moderates disturbance
   – More Kelp = Smaller Disturbance?
   – BUT no effect on kelp that isn't present…

2. Kelp regrows quickly
   – Dense beds after storms if nutrients present

Incorporate Interacting Variables

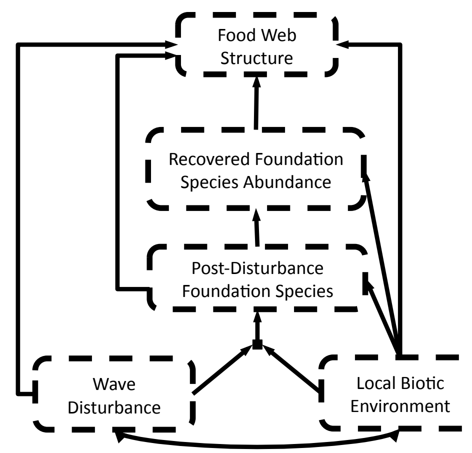


Measuring Realized Disturbance via Satellite Measurements

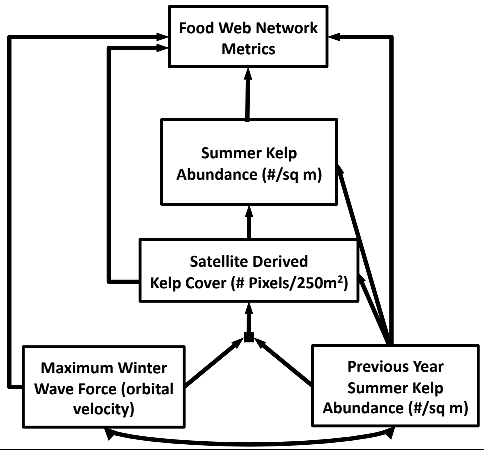


Cavanaugh et al 2011 MEPS

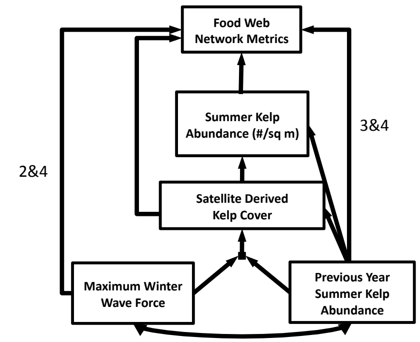
Incorporate Natural History of Disturbance



Model with Observed Variables



Consider Alternate Models



1. Full Model
2. No waves -> Food Web effect
3. No effect of previous year's kelp -> Food Web
4. No effect of either waves **or** previous kelp -> Food Web

Confront your models with data!

(then have lunch)