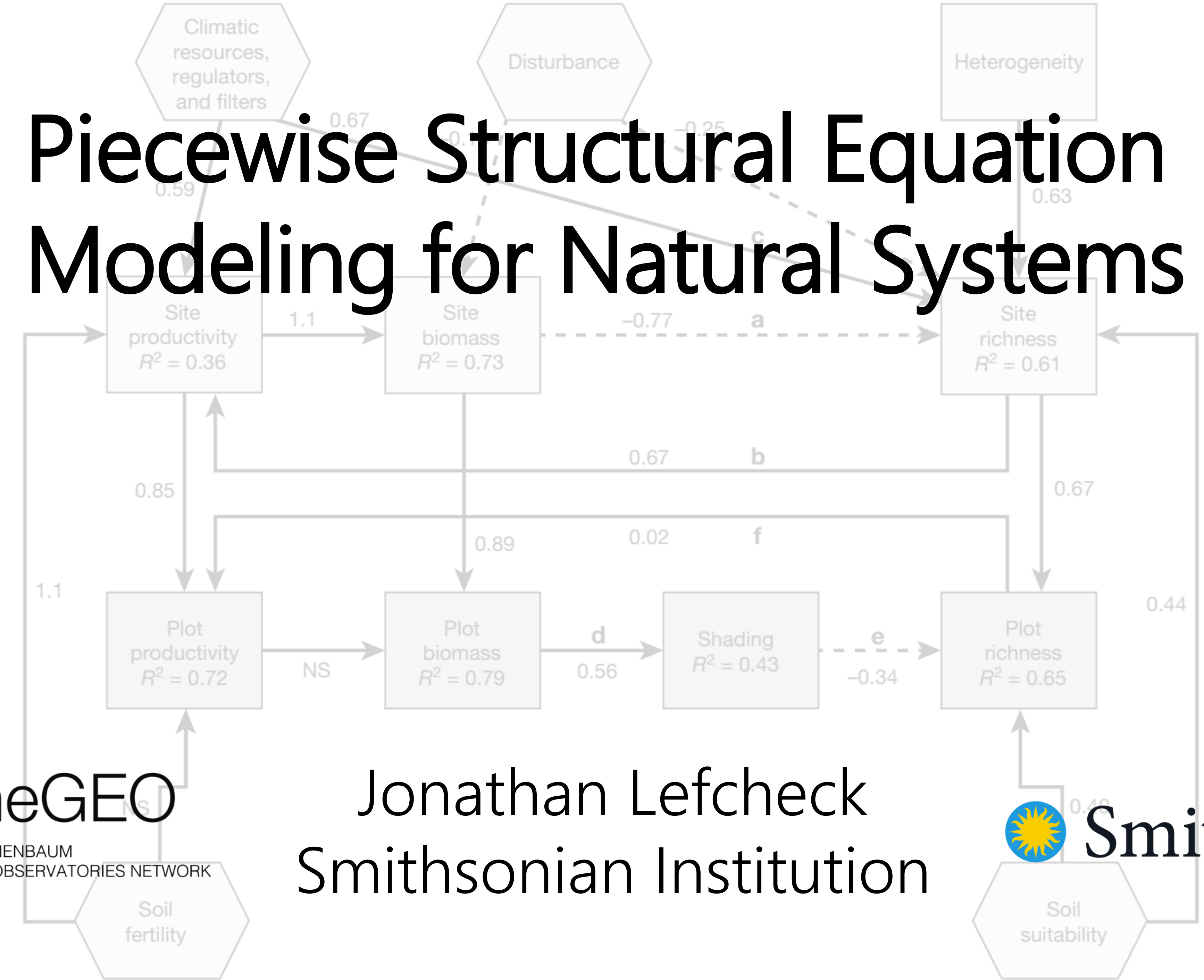
Piecewise Structural Equation Modeling for Natural Systems



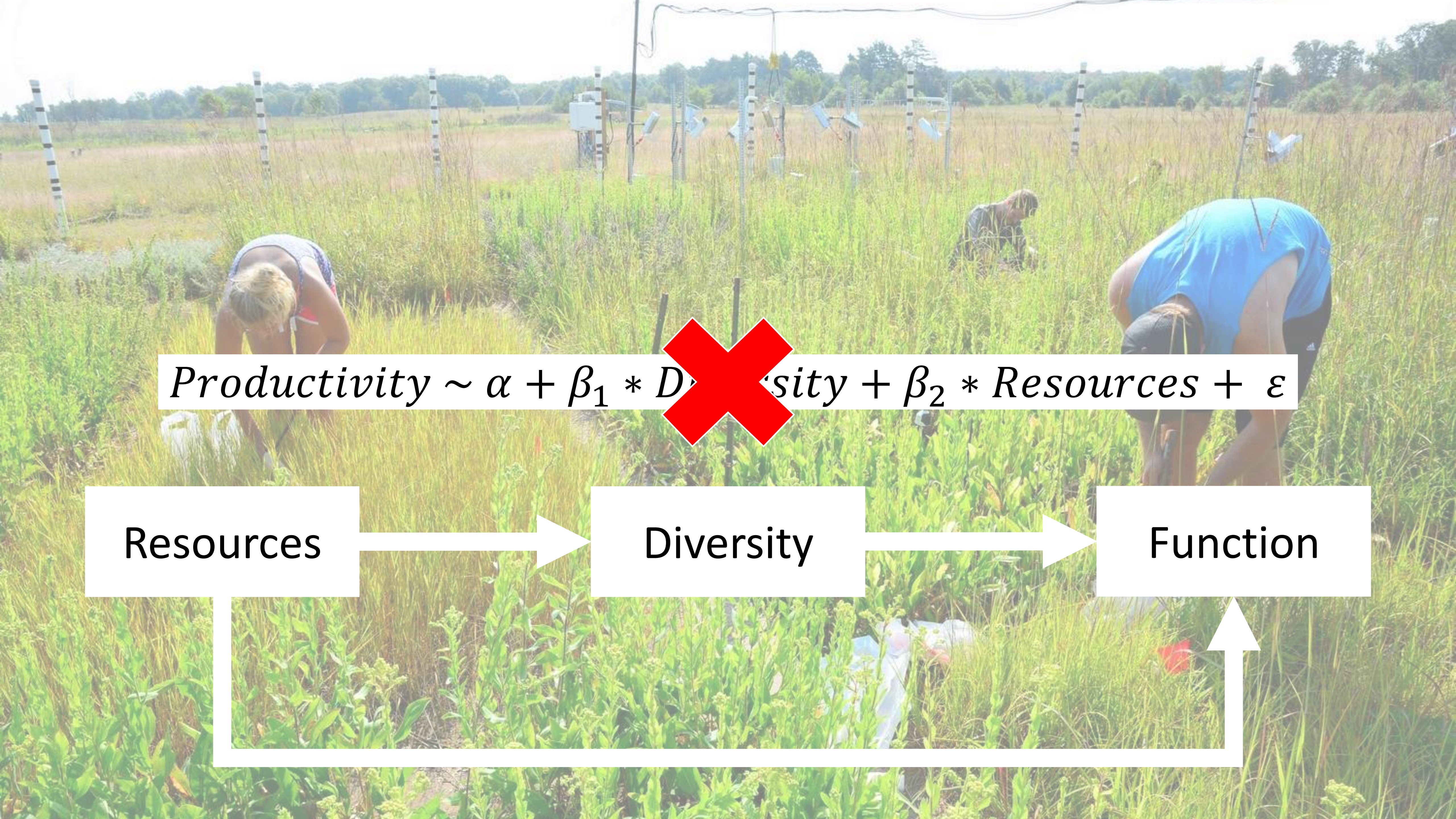
MarineGEO
THE TENNENBAUM
MARINE OBSERVATORIES NETWORK

Jonathan Lefcheck
Smithsonian Institution

Smithsonian

$$Diversity \sim \alpha + \beta_1 * Resources + \varepsilon$$

Resources → Diversity



$$Productivity \sim \alpha + \beta_1 * \cancel{Diversity} + \beta_2 * Resources + \ \varepsilon$$

$$\begin{tikzcd} \text{Resources} \arrow[r] \arrow[rr, bend right] & \text{Diversity} \arrow[r] & \text{Function} \end{tikzcd}$$


Abiotic
Resources
Diversity
Function


Abiotic
Resources
Diversity
Function

Structural Equation Modeling

Structural Equation Modeling

There is hypothesized underlying *structure* to nature (a cause and an effect)…

Structural

There is hypothesized underlying *structure* to nature (a cause and an effect)…

Equation

…that can be translated to a series of mathematical *equations*…

Modeling

Structural

There is hypothesized underlying *structure* to nature (a cause and an effect)…

Equation

…that can be translated to a series of mathematical *equations*…

Modeling

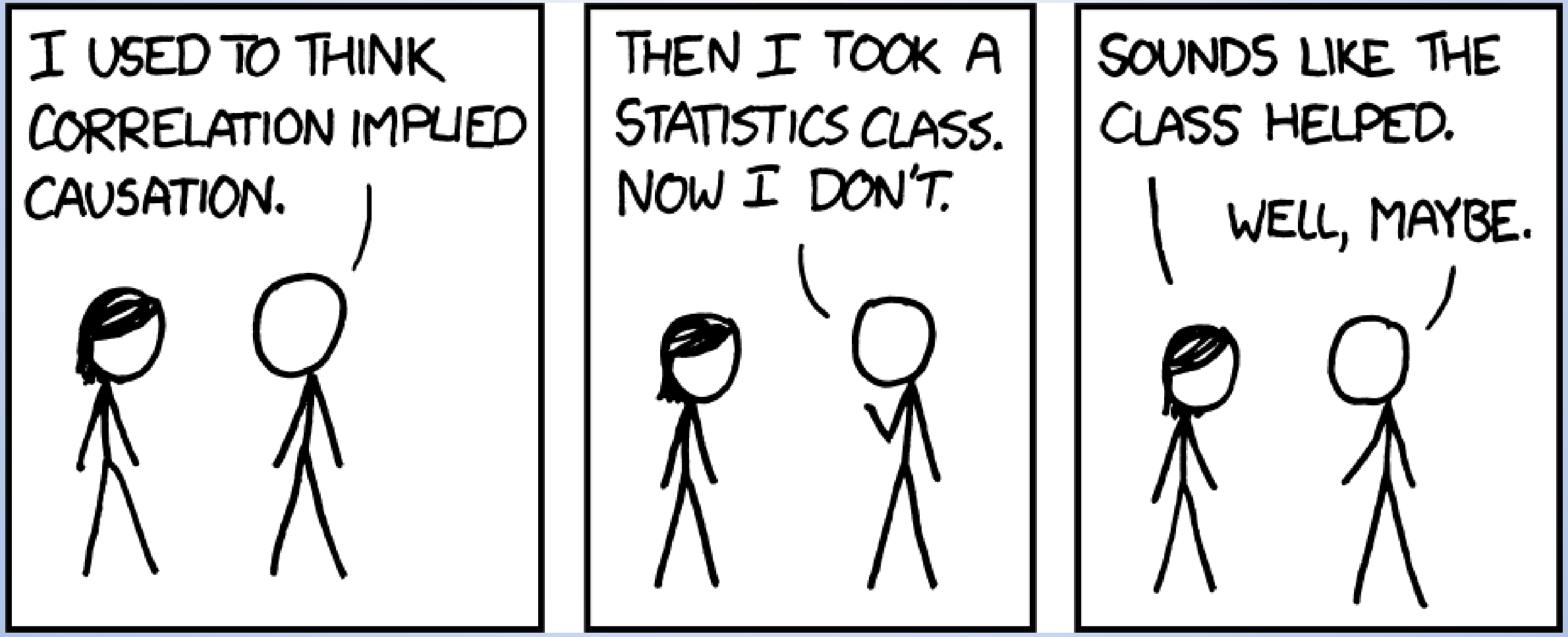
…which can be *modeled* against data to support or refute the proposed structure

Structural Equation Modeling: The Basics

- A series of equations united in a single *causal* network
  - Can investigate the flow of effects through a system (i.e., indirect effects)
- What the f$*# is causality?

Structural Equation Modeling: The Basics

- *"Correlation does not imply causation"*



Structural Equation Modeling: The Basics

"In fact, with a few exceptions, correlation <u>does</u> imply causation. [A] simple correlation implies an *unresolved* causal structure, since we cannot know which is the cause, which is the effect, or even if both are common effects of some third, unmeasured variable."

-Bill Shipley,
Cause and Correlation in Biology, 2004

Elspeth Kirkman
@Karminker

Major causal inference issues in our house atm. Two year old thinks coats make the air colder so keeps putting hers on when she’s too hot 🤦‍♀️

09:50 · 12/11/2022 · Twitter for iPhone

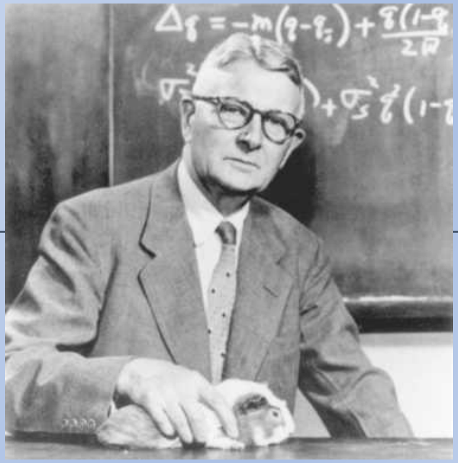
442 Retweets **39** Quote Tweets **7,642** Likes

Structural Equation Modeling: The Basics

- How can we determine *causality*?
  - Design and conduct experiments
  - Develop theory
  - Build knowledge from past observation
  - Test hypotheses

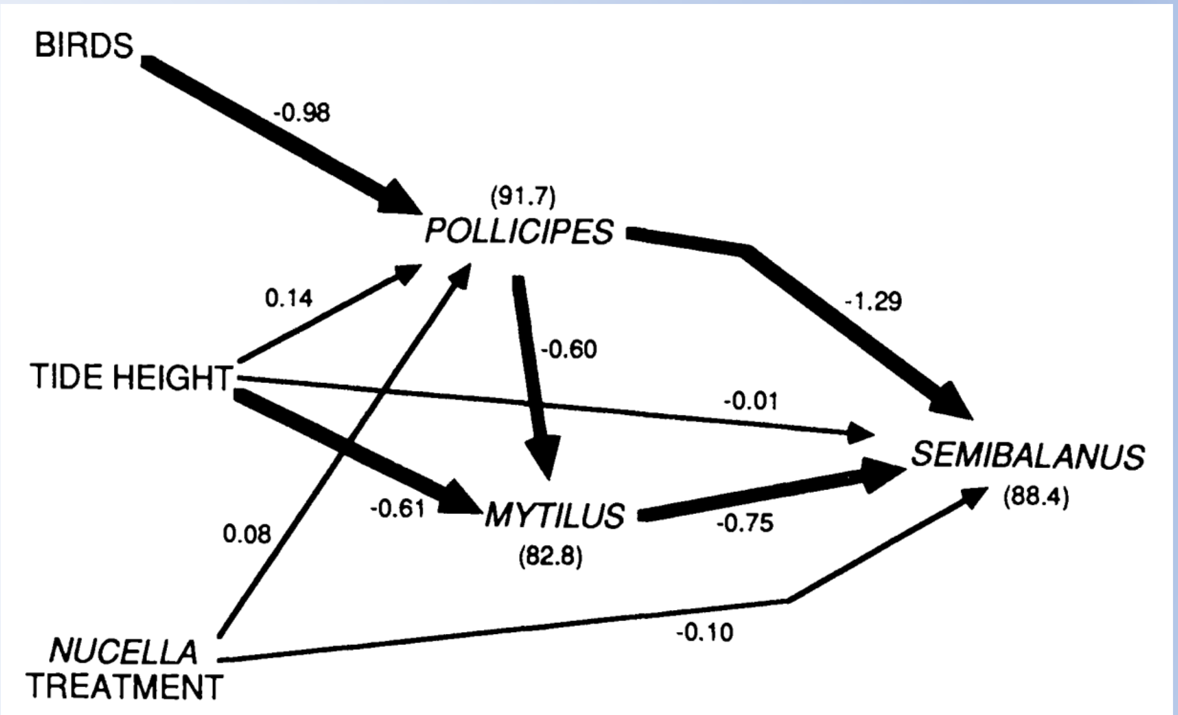
"The *combination* **of knowledge of correlations with knowledge of causal relations,** to obtain certain results, **is a different thing from the *deduction* of causal relations from correlations.**"

-Sewell Wright, 1923

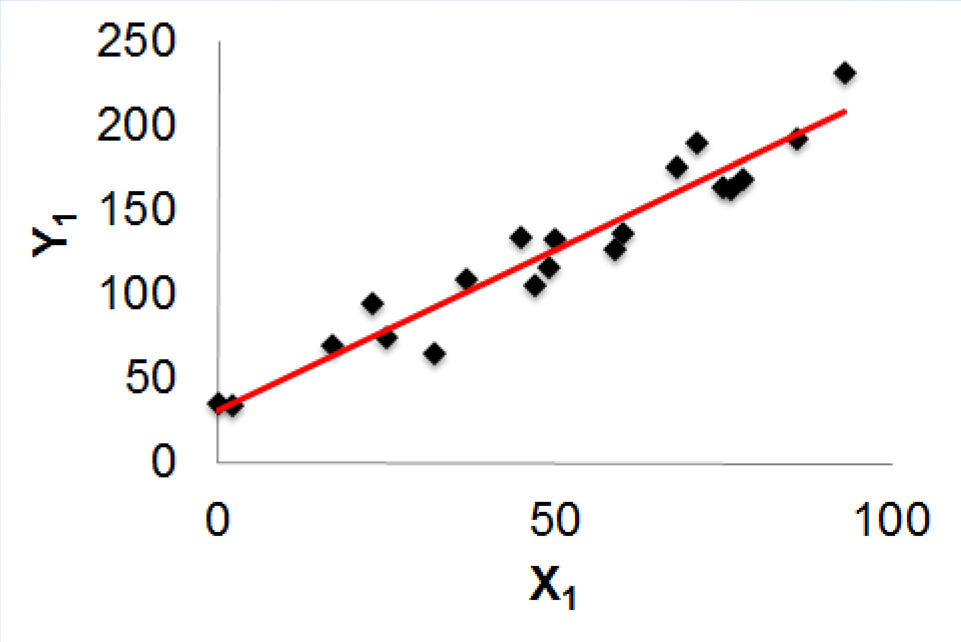


Structural Equation Modeling: The Basics

- How can we evaluate *causality*?
  - Build a hypothesized causal structure
  - Collect data
  - Test data against structure, extract inferences
  - Wash-rinse-repeat



Graphical form

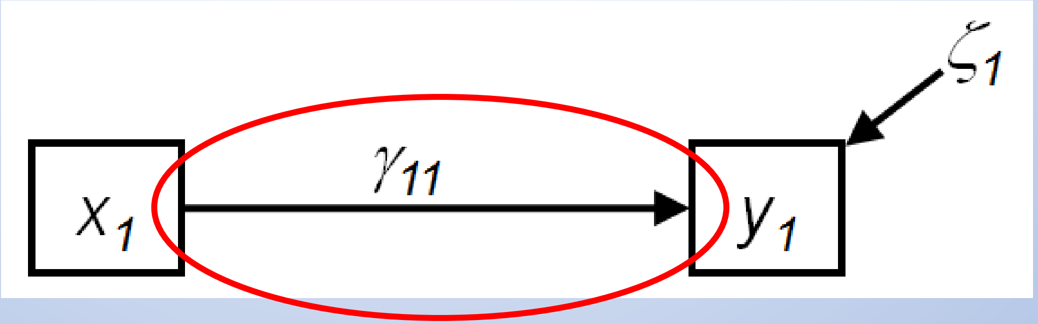


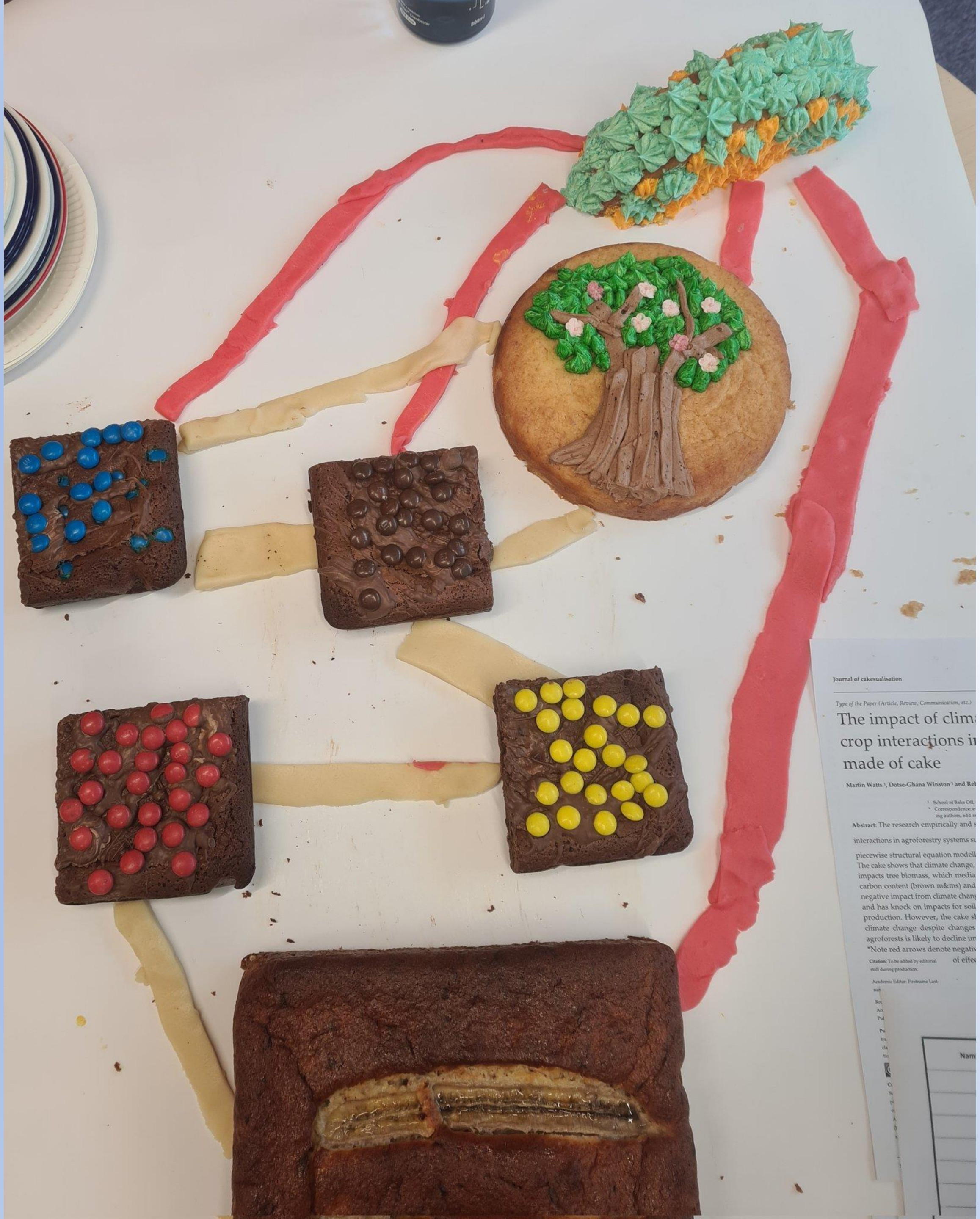
Equation form

$$y_1 \sim \alpha_1 + \beta_1 X_1 + \varepsilon$$

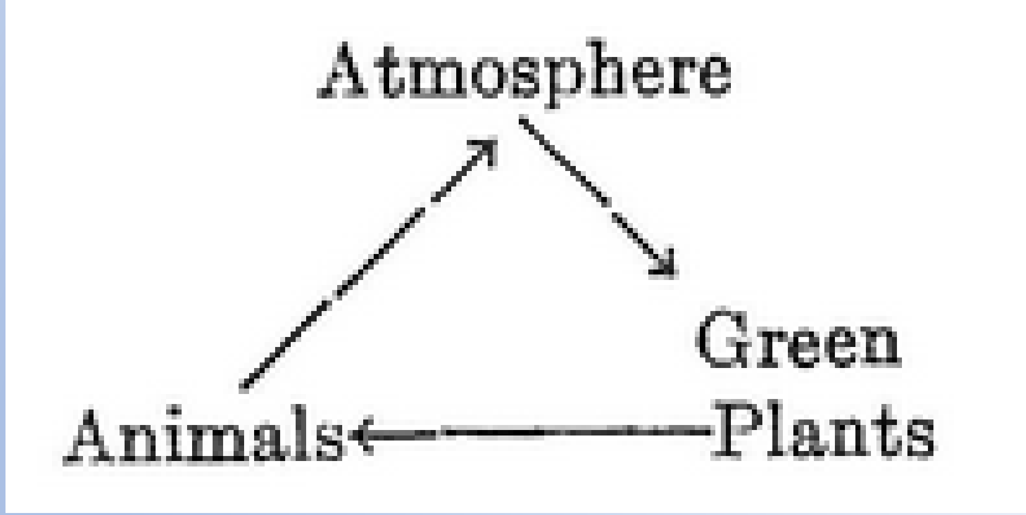
$$y_{11} = \alpha_1 + \gamma_{11} x_1 + \zeta_1$$

Graphical form





@HomeGardenMartn



Alfred Lotka,
"Elements of
Physical
Biology"
(1925, p. 221)

Bob Paine,
1966

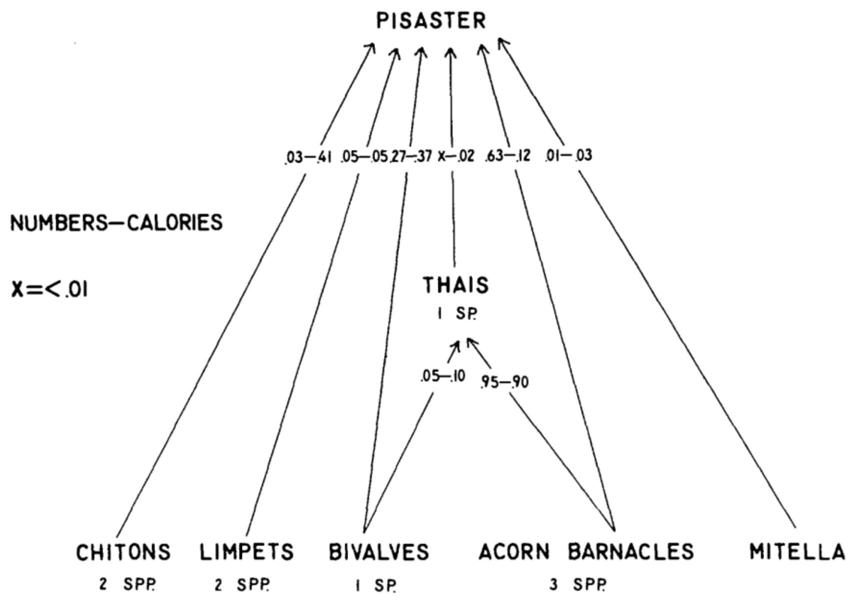


FIG. 1. The feeding relationships by numbers and calories of the *Pisaster* dominated subweb at Mukkaw Bay. *Pisaster*, N = 1049; *Thais*, N = 287. N is the number of food items observed eaten by the predators. The specific composition of each predator's diet is given as a pair of fractions; numbers on the left, calories on the right.

Structural Equation Modeling: The History

- First generation SEM = path analysis

- Second generation SEM = path analysis + factor analysis
  - Global estimation of variance-covariance matrix
  - *lavaan* package

- Third generation SEM = graph theory
  - Local estimation of each equation
  - Increased flexibility to accommodate messy ecological data
  - *piecewiseSEM* package

R

Grace & Keeley (2006): How does fire severity influence California shrublands?

N = 99 sites, measured post-fire

Focus just on plant community metrics

Ecological Applications, 16(2), 2006, pp. 503–514

A STRUCTURAL EQUATION MODEL ANALYSIS OF POSTFIRE PLANT DIVERSITY IN CALIFORNIA SHRUBLANDS

JAMES B. GRACE[1,4] AND JON E. KEELEY[2,3]

[1]*U.S. Geological Survey, National Wetlands Research Center, 700 Cajundome Boulevard, Lafayette, Louisiana 70506 USA*
[2]*U.S. Geological Survey, Western Ecological Research Center, Sequoia and Three Rivers Canyon National Parks, Three Rivers, California 93271 USA,*
[3]*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095 USA*

Abstract. This study investigates patterns of plant diversity following wildfires in fire-prone shrublands of California, seeks to understand those patterns in terms of both local and landscape factors, and considers the implications for fire management. Ninety study sites were established following extensive wildfires in 1993, and 1000-m^2 plots were used to sample a variety of parameters. Data on community responses were collected for five years following fire. Structural equation modeling (SEM) was used to relate plant species richness to plant abundance, fire severity, abiotic conditions, within-plot heterogeneity, stand age, and position in the landscape. Temporal dynamics of average richness response was also modeled. Richness was highest in the first year following fire, indicating postfire enhancement of diversity. A general decline in richness over time was detected, with year-to-year variation attributable to annual variations in precipitation. Peak richness in the landscape was found where (1) plant abundance was moderately high, (2) within-plot heterogeneity was high, (3) soils were moderately low in nitrogen, high in sand content, and with high rock cover, (4) fire severity was low, and (5) stands were young prior to fire. Many of these characteristics were correlated with position in the landscape and associated conditions. We infer from the SEM results that postfire richness in this system is strongly influenced by local conditions and that these conditions are, in turn, predictably related to landscape-level conditions. For example, we observed that older stands of shrubs were characterized by more severe fires, which were associated with a low recovery of plant cover and low richness. These results may have implications for the use of prescribed fire in this system if these findings extrapolate to prescribed burns as we would expect.

Key words: colonization; diversity; fire; heterogeneity; landscape; niche partitioning; prescribed burning; productivity; resource availability; species richness; structural equation modeling (SEM).

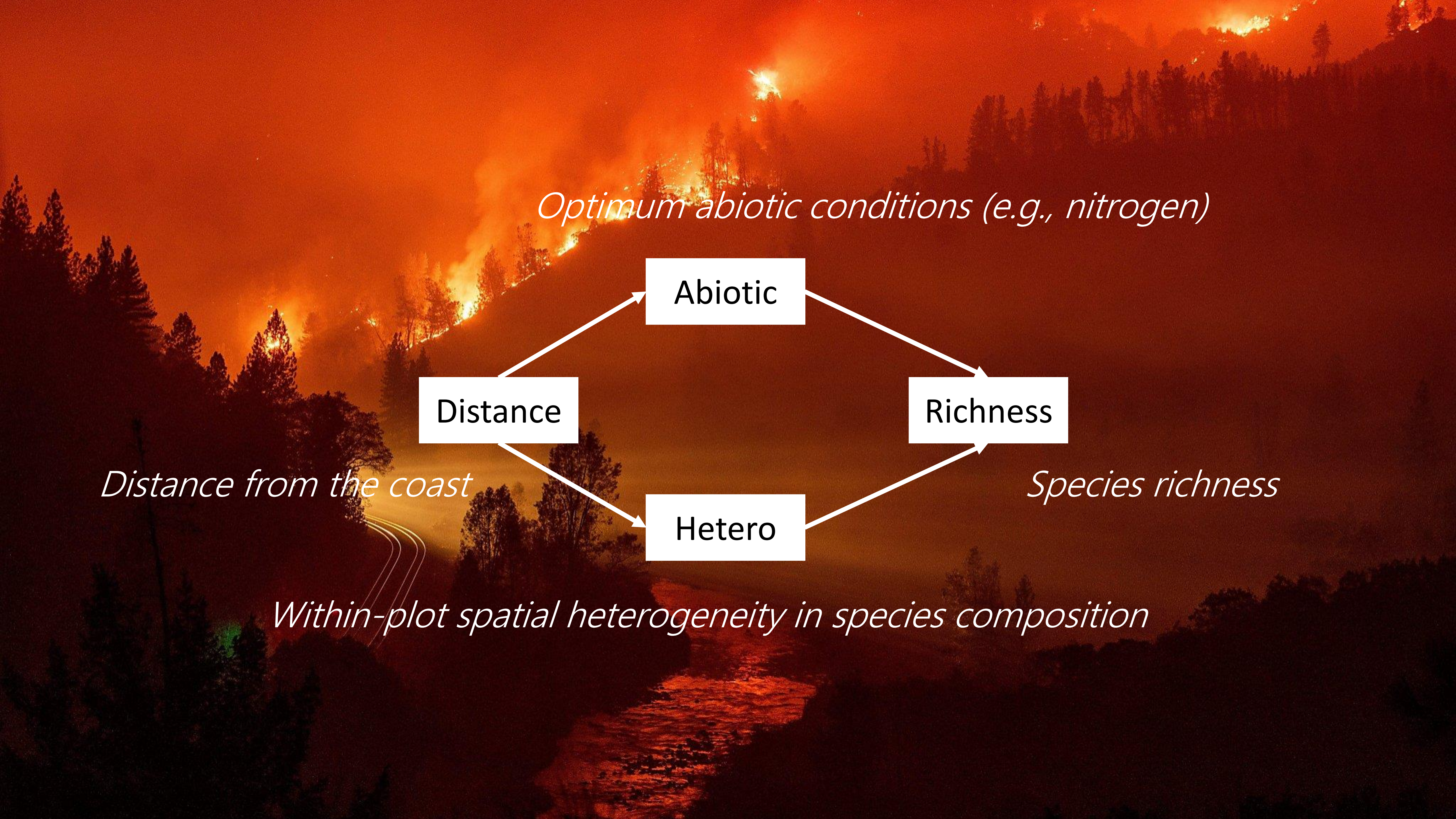
INTRODUCTION

There exists a large number of theories about individual mechanisms that can potentially influence species diversity (see reviews in Huston 1994, Palmer 1994, Rosenzweig 1995). Involved in these mechanisms are factors such as (1) resource availability and community productivity, (2) competition and facilitation, (3) spatial heterogeneity and dispersal, (4) disturbance and succession, (5) the influences of regional and local species pools, and (6) the role of stochastic factors. The challenge that now faces ecologists is to ascertain how these various forces work together in natural systems to regulate patterns and dynamics in diversity.

It is often unappreciated how limited conventional analytical approaches are when it comes to addressing questions about multiple controlling factors. Because of the intercorrelations among factors, piecewise analysis of systems using univariate procedures (those that consider one response variable at a time) are helpful for describing overall patterns, but leave unresolved many questions that relate to networks of interactions. Alternative approaches, such as structural equation modeling (Shipley 2000, Pugesek et al. 2003, Grace 2006), are designed for the study of multivariate hypotheses and apply more directly to understanding interacting systems. Structural equation modeling (SEM) involves the use of a multiequational framework to develop and test theoretically based models in order to understand responses controlled by multiple factors (Bollen 1989). Through the use of a simultaneous analysis procedure, SEM derives results that seek to account for the roles of multiple factors in a single analysis.

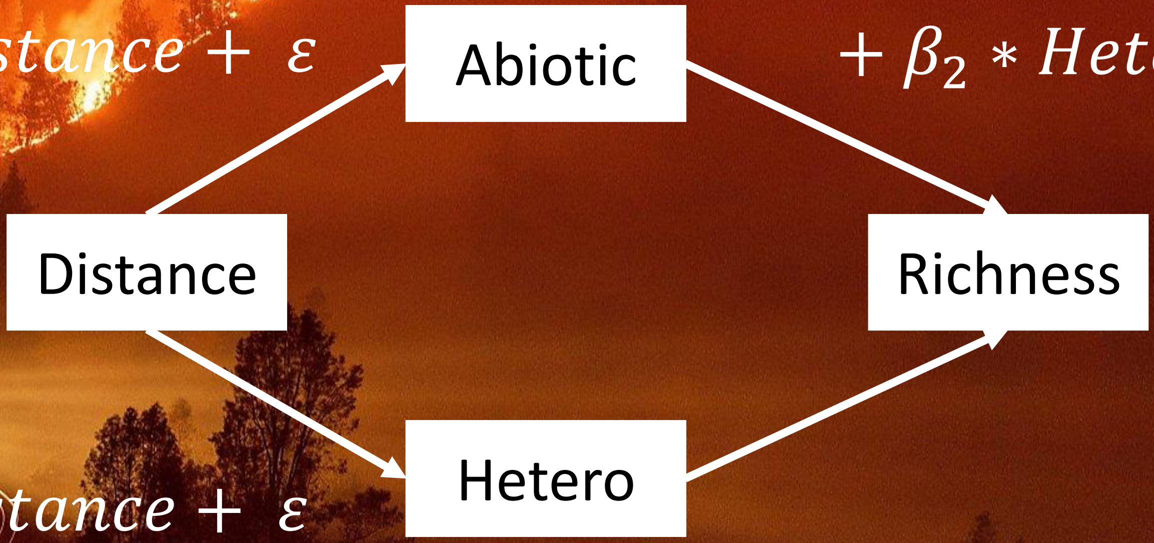
Commonly, SEM provides quite a different perspective by partitioning direct from indirect effects and thereby revealing a variety of mechanisms behind the overall patterns. When SEM is used in the study of natural systems, it is frequently shown that reliance on conventional univariate relationships can lead to misleading impressions. For example, Johnson et al. (1991) found that the addition of the herbicide atrazine to

Manuscript received 4 May 2005; revised 23 August 2005; accepted 29 August 2005. Corresponding Editor: C. L. Boggs.
[4] E-mail: Jim_Grace@usgs.gov


Optimum abiotic conditions (e.g., nitrogen)
Abiotic
Distance
Richness
Hetero
Distance from the coast
Species richness
Within-plot spatial heterogeneity in species composition

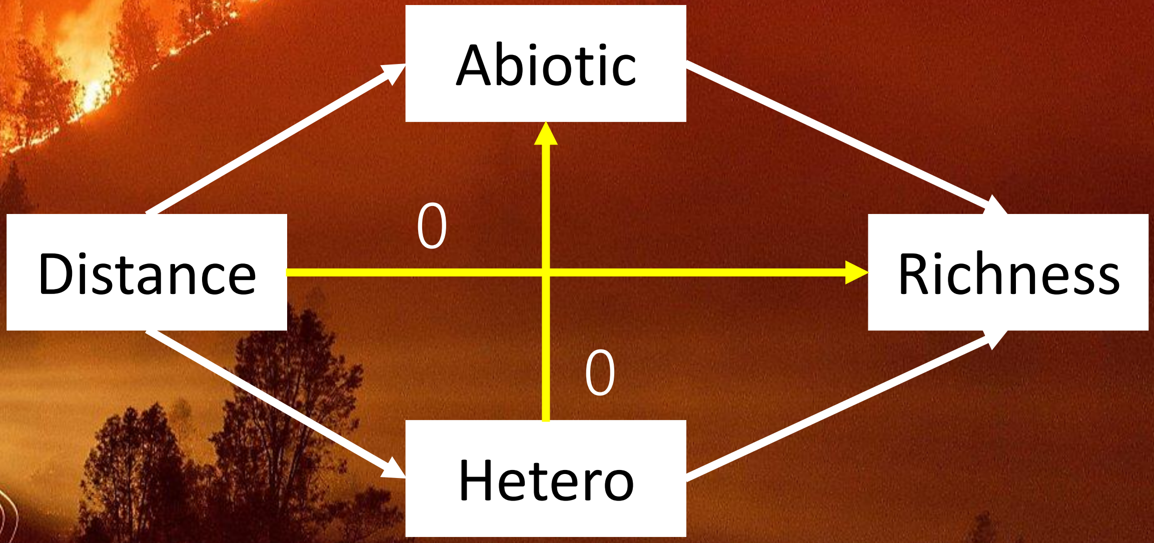
$$Abiotic \sim \alpha + \gamma_1 * Distance + \varepsilon$$

$$Richness \sim \alpha + \beta_1 * Abiotic + \beta_2 * Hetero + \varepsilon$$

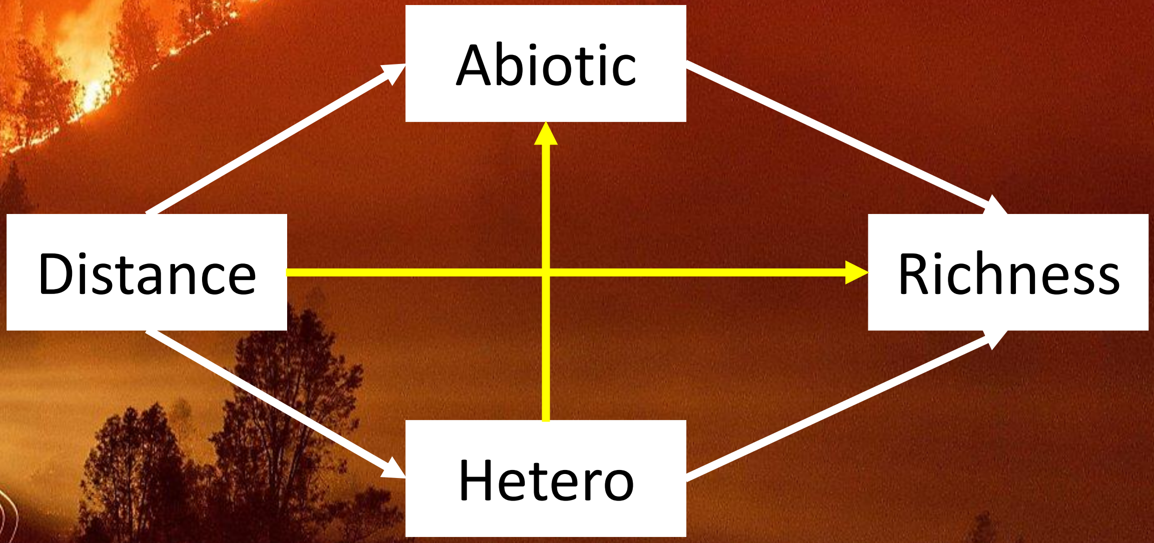


$$Hetero \sim \alpha + \gamma_2 * Distance + \varepsilon$$

1. Create list of structured equations
2. Conduct *d*-sep tests (evaluate fit)
3. Extract coefficients

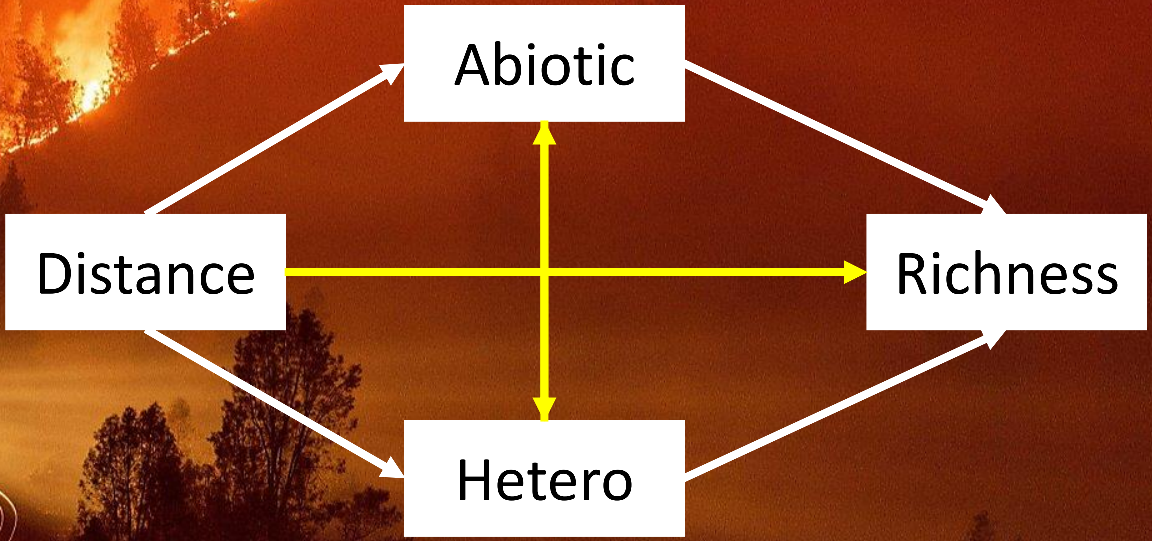


1. Create list of structured equations
2. **Conduct *d*-sep tests (evaluate fit)**
3. Extract coefficients



$C = -2 * \sum \ln(p_i)$

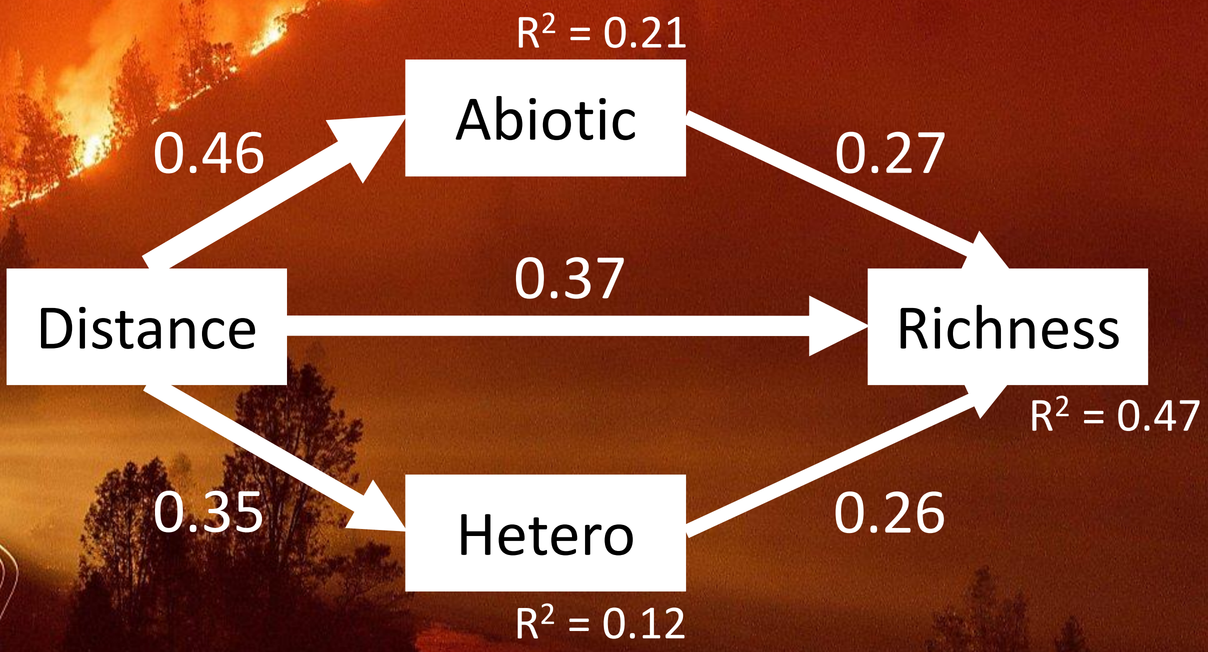
1. Create list of structured equations
2. **Conduct *d*-sep tests (evaluate fit)**
3. Extract coefficients



$$C = -2*\sum \ln(p_i)$$

1. Create list of structured equations
2. **Conduct *d*-sep tests (evaluate fit)**
3. Extract coefficients

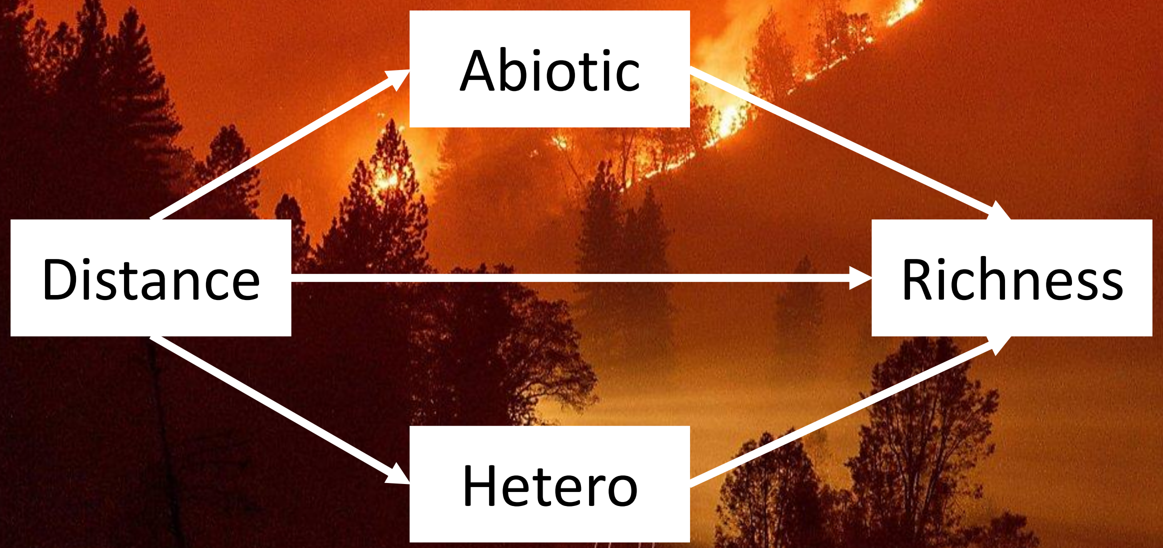
R



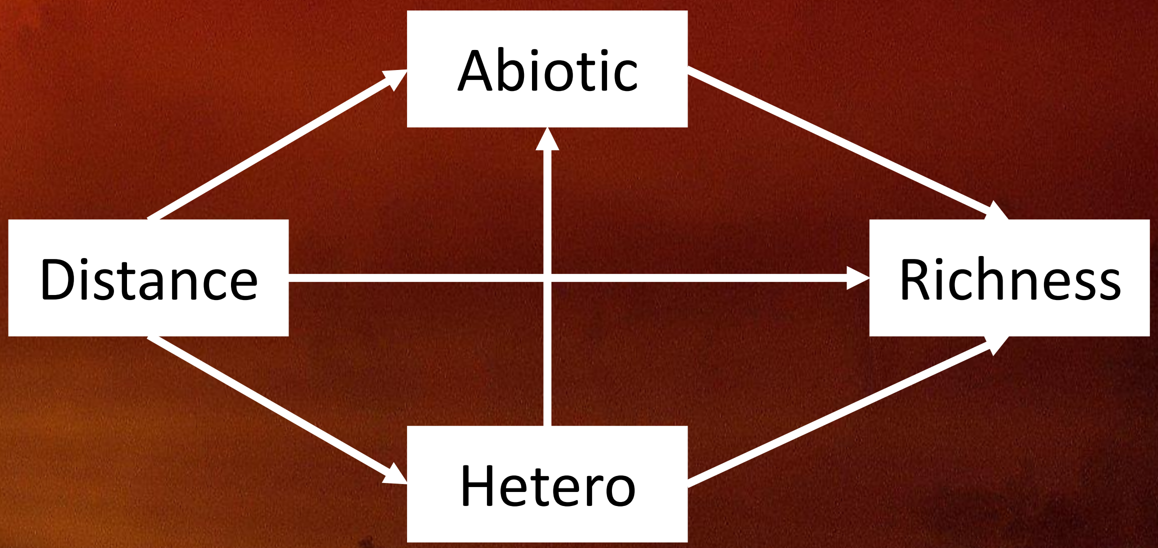
1. Create list of structured equations
2. Conduct *d*-sep tests (evaluate fit)
3. **Extract coefficients**

$$\chi^2 = -2(\log(\mathcal{L}(M_1)) - \log(\mathcal{L}(M_2)))$$

Proposed model



Saturated model

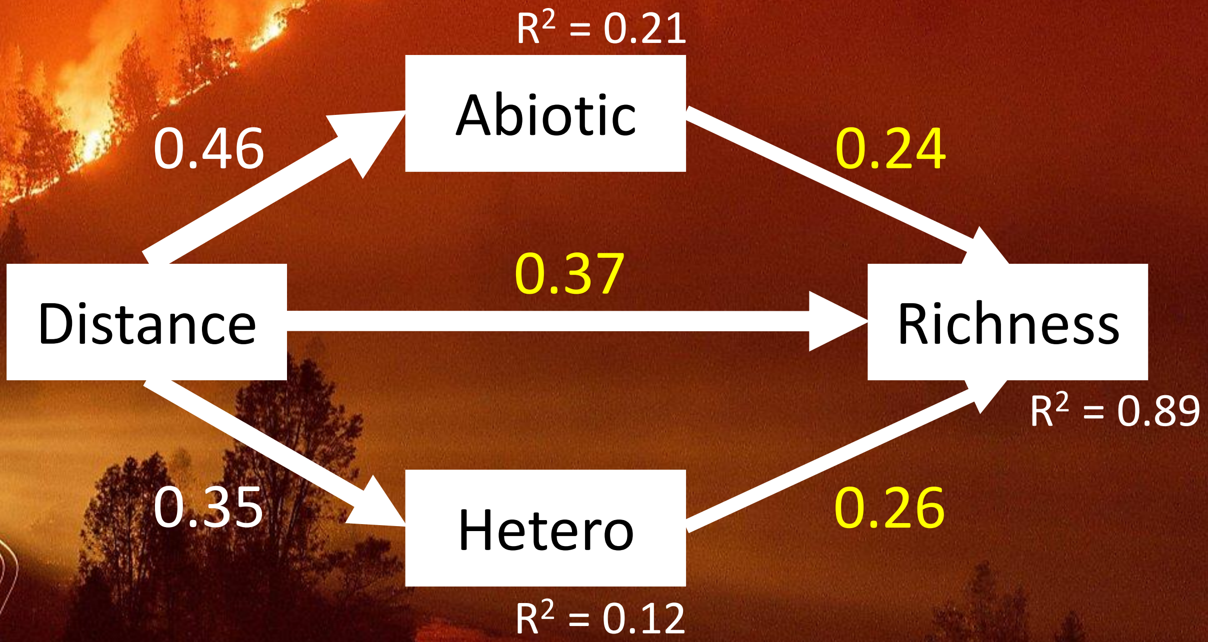


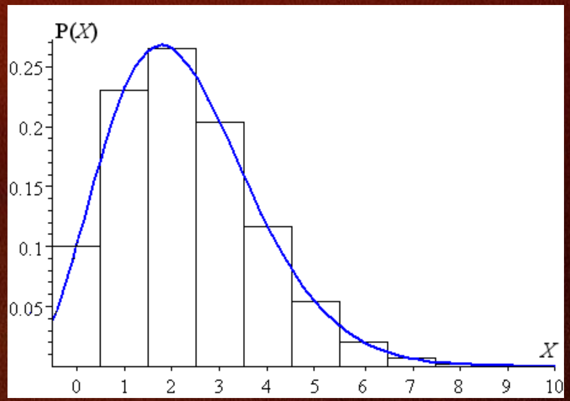
1. Create list of structured equations
2. **Conduct Chi-squared test (evaluate fit)**
3. Extract coefficients

R

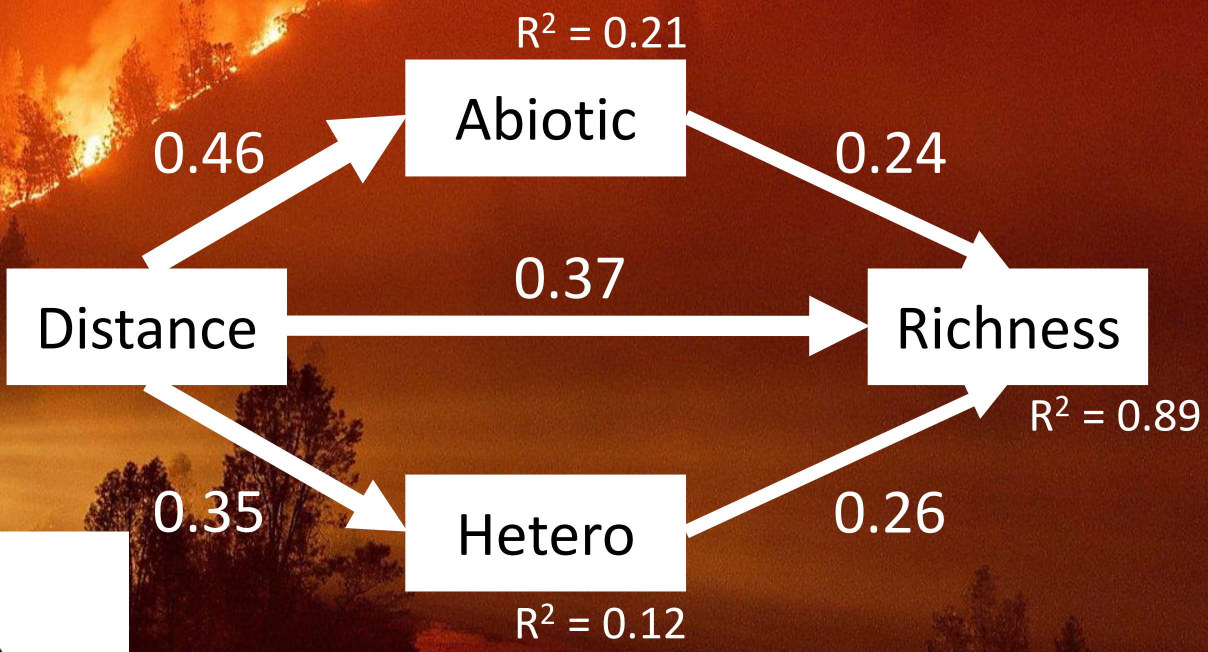
Piecewise SEM: Benefits over classical SEM

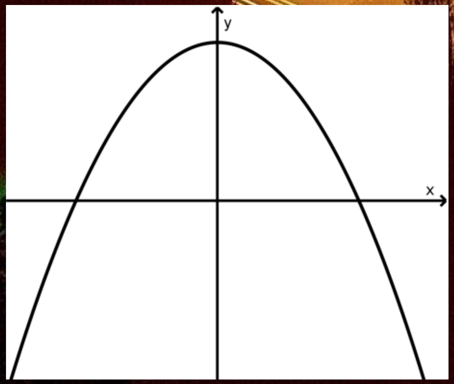
- Can incorporate different error distributions (e.g., Poisson, binomial, Gamma, etc.)
- Can address non-independence (e.g., random effects, spatial or temporal autocorrelation)
- Can model truly non-linear relationships (e.g., generalized linear and generalized additive models)
- Uses familiar syntax in R: if you can code it*, you can make an SEM out of it!

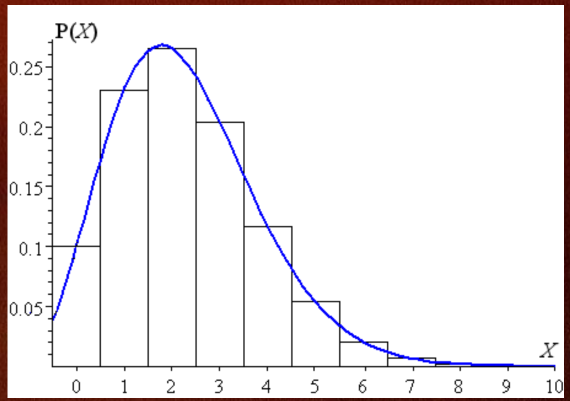




1. Create list of structured equations
2. Conduct *d*-sep tests (evaluate fit)
3. **Extract coefficients (GLM)**







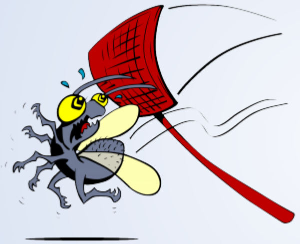
1. Create list of structured equations
2. Conduct *d*-sep tests (evaluate fit)
3. **Extract "coefficients" (GAM)**

Piecewise SEM: Benefits over classical SEM

- Greater parity with classical SEM now possible (e.g., Chi-squared tests) but not *totally* there yet
- Correlated errors (Shipley & Douma 2021, 2022)
- Latent variables (Shipley & Douma 2021)
- Full extension to non-linear applications

Piecewise SEM: Into the future

- Increase in supported model classes (added *glmmTMB* and *Sarlm* in this update)– what do you want to see?

- Patience & bug fixes! 

- Contact me:
  - Email: LefcheckJ@si.edu
  - Twitter: [@jslefche](https://twitter.com/jslefche)
  - E-book: https://jslefche.github.io/sem_book/