Introduction to structural equation modelling

Frank Pennekamp

11/11/2020

Research questions

- Ecology is about interactions between organisms and their environments.
- We often have ideas how things could be connected in ecological systems.
- To test these ideas, we need a way to dissect when they occurr for a reason versus randomness.
- We use statistics to understand, when connections are non-random.
- Research questions are about understanding cause and effect.

Why do we need to use SEM in Ecology

To better understand the relationship among variables in complex ecosystems: - We often have multiple observed variables - We want to test and evaluate multivariate causal relationships - Test direct and indirect effects on assumed causal relationships - Incorporate observed and latent variables - Include interaction terms can test main effects and interaction effects

Why do we need to use SEM in Ecology

- System thinking
  - "Understanding the whole rather than the parts in isolation"
- Examine relative influences and responses
- In SEM specification of a model is based on theory and research to determine and validify a proposed causal process and/or model.

Two goals of SEM:
1) Understand the patterns of correlation/covariance among a set of variables.
2) Explain as much of their variance as possible with the model specified (Kline, 1998).

Introduction to the dataset

We will use a dataset collected at the Cedar Creek Ecosystem Science Reserve. Includes ecosystems and species of North American forests and grassland.

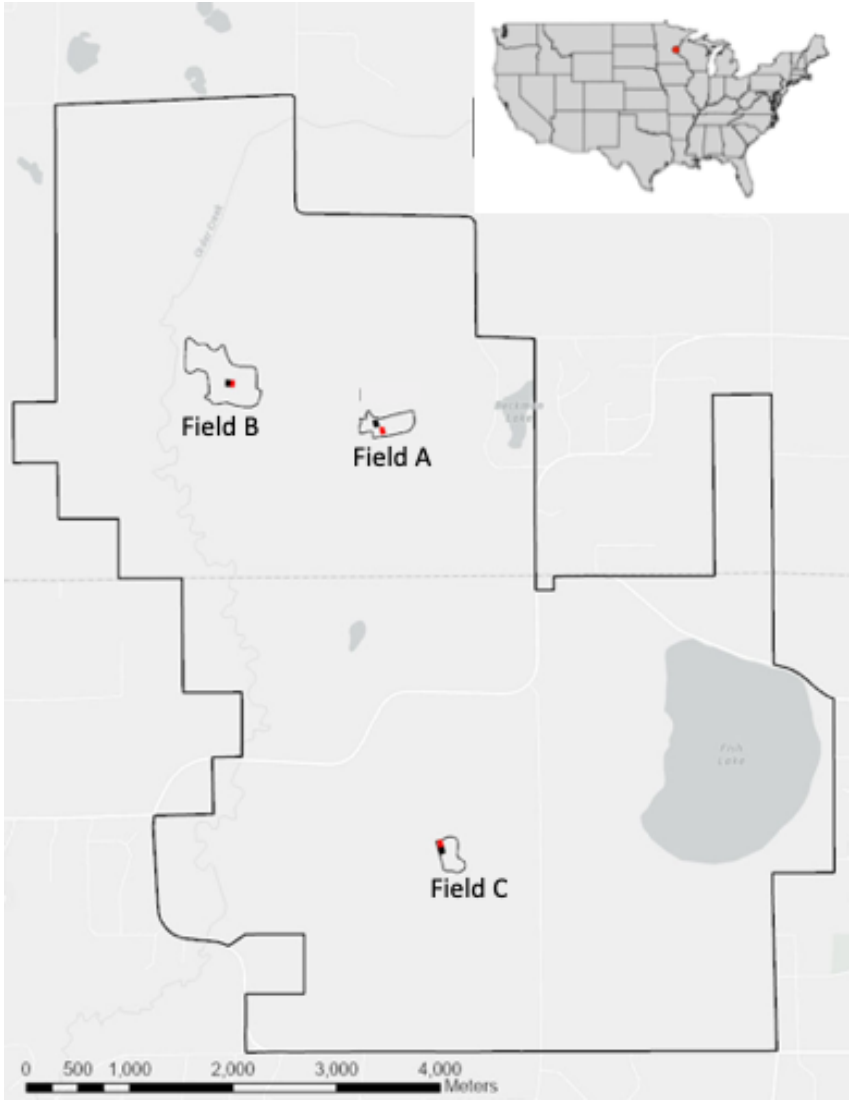
Experiments examine long-term consequences of human-driven environmental changes ecosystem responses to: - Biodiversity loss - Nitrogen deposition - Elevated carbon dioxide - Warming and changes in precipitation - Exotic species invasions

Introduction to the dataset

Understanding the recovery of a grassland for two decades following an intensive agricultural disturbance under ambient and elevated nutrient conditions.

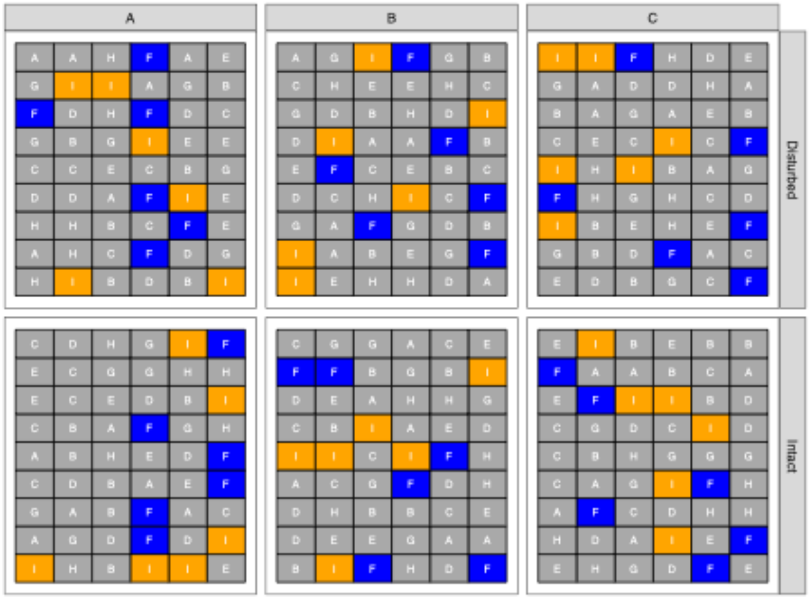
1) How has aboveground biomass changed as a function of disturbance (disking) and nutrient addition?
2) How are these effects mediated by diversity?

Introduction to the dataset



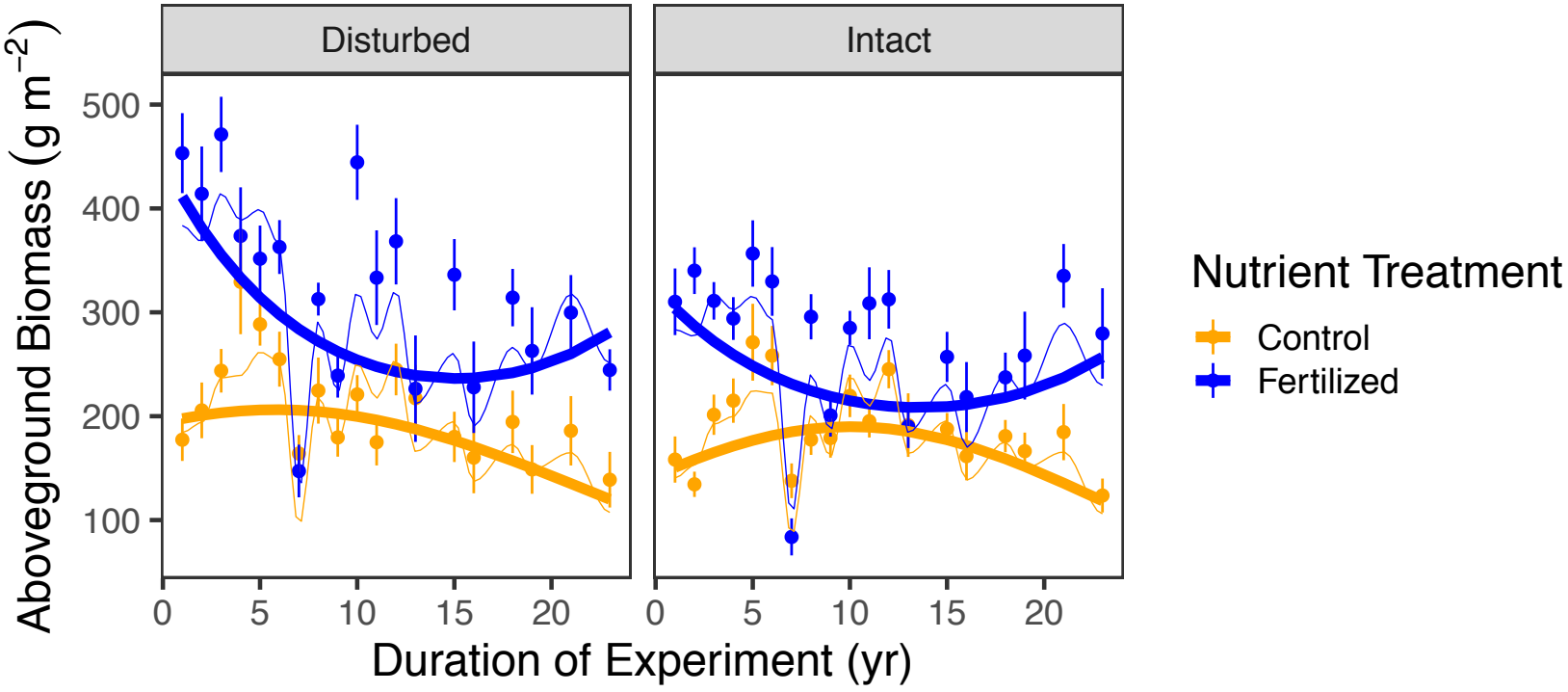
Map showing location of the study site (Cedar Creek Ecosystem Science Reserve), the location of the three study fields within the reserve, and location of the 35 x 55 m intact (black) and disturbed (red) plots within each field.

Introduction to the dataset



Location of the 4 x 4 m nutrient treatment plots within each 35 x 55 m Intact or Disturbed plot within each of three fields (A, B, and C). Letters indicate the nutrient treatments, and the colored plots are treatments that are the focus of the analyses presented here: Control (orange) and 9.5 g N m-2 yr-1 (blue).

Introduction to the dataset



Effect of soil disturbance (disking) and nutrient enrichment on live, aboveground plant biomass. Colors indicate nutrient addition treatment: Control and NPK+ (all nutrients plus 9.5 g N m-2 yr-1).

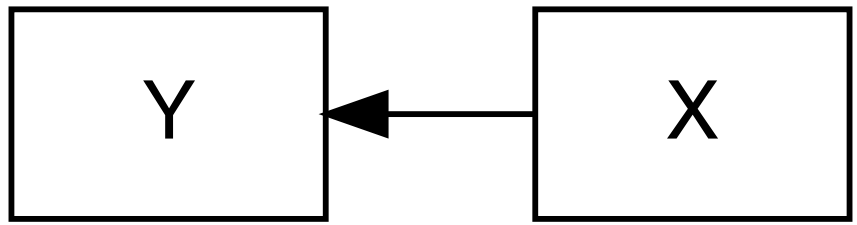
Question of interest

Present one research question based on the data. This question only NEEDS to include 2 variables. Biomass -> richness and evenness

Graphical representation

Draw a model by hand

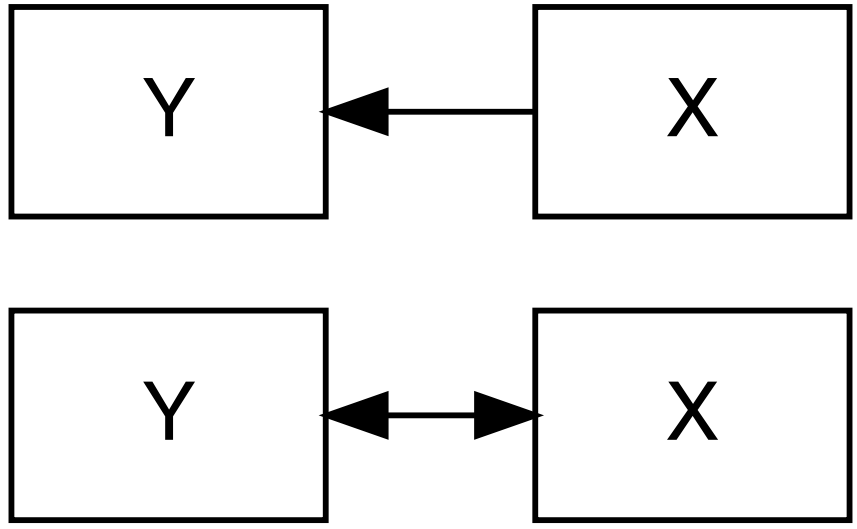
Introduce the graphical model representation (X -> Y):



General notation: - Variables are nodes.
- Arrows are causal relationships.

Causality

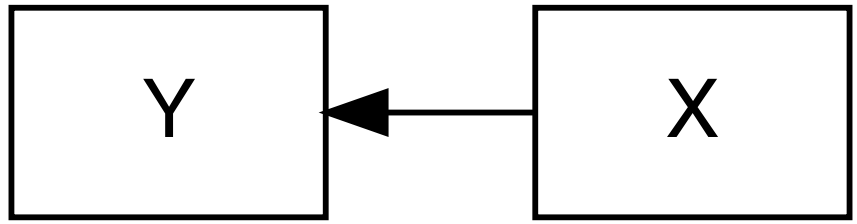
- What is a causal relationship?
- "Correlation does not imply causation"
- Causation indicates a relation between two variables in which one variable is affected by another.
- The arrow above indicates a causal relationship.
- Explain the concept and discuss with students the topic.
- Show the difference between direct effects and correlation together with the graphical representation.

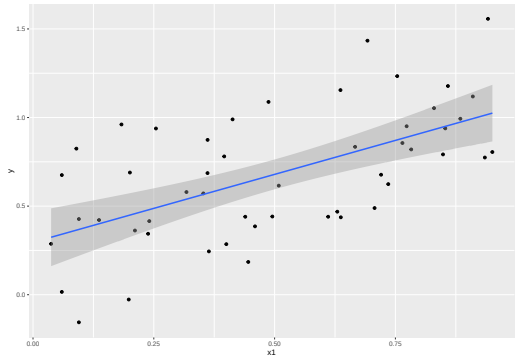


Simple statistical model

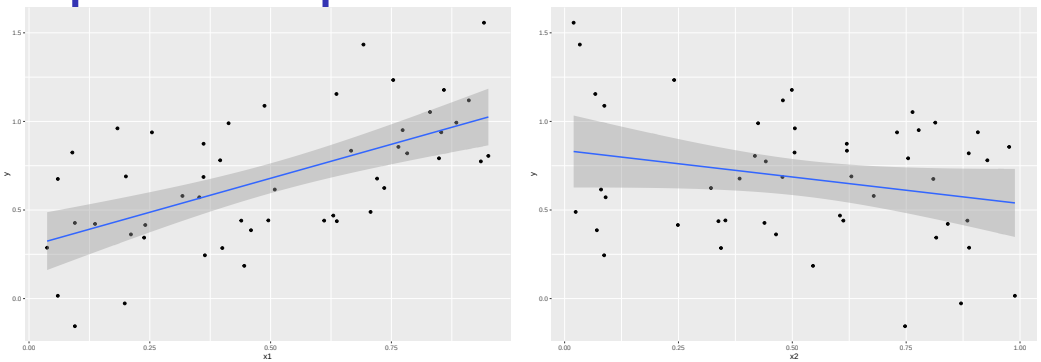
Present the bivariate model. Include the equation and a plot. It is important that students understand that all they know can be defined with a graphical representation.

lm(y~x1)



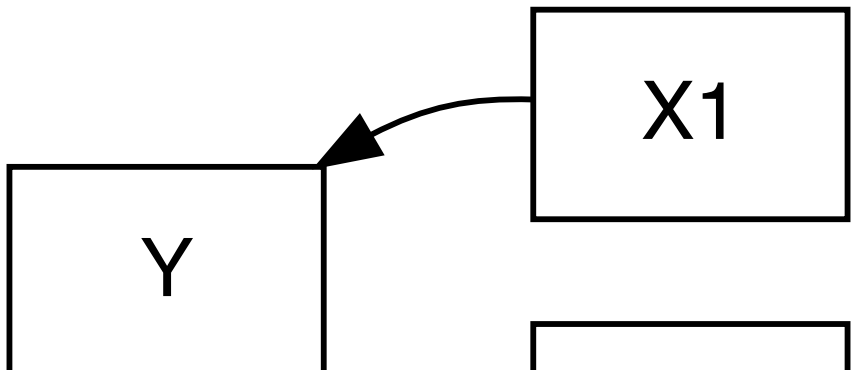


Multiple independent variables



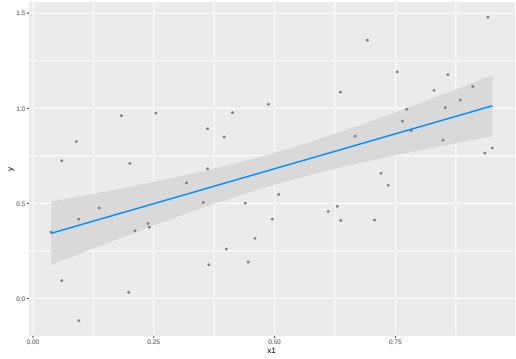
- Discuss with students other variables that may affect B.
- Introduce a new variable and the question that it answers
- Add nutrient variable
- Show the graphical model (X1 -> Y, X2 -> Y).
- Include the equation and the two partial plots.
- Be clear this is a multiple regression model.

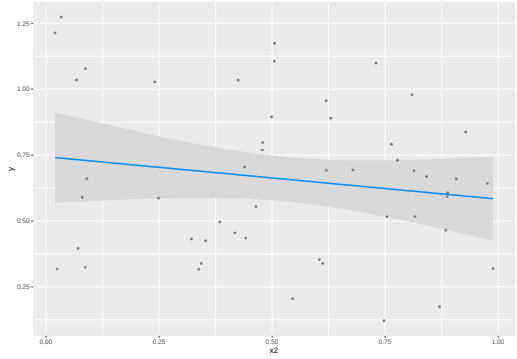
lm(y~x1+x2, data=dat)



Multiple independent variables

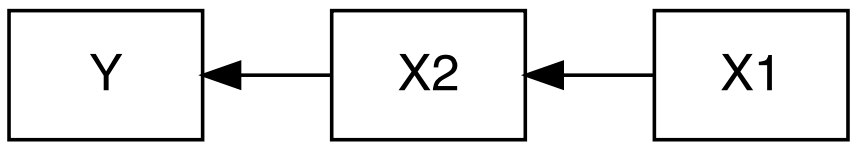
```
## [[1]]

##
## [[2]]
```





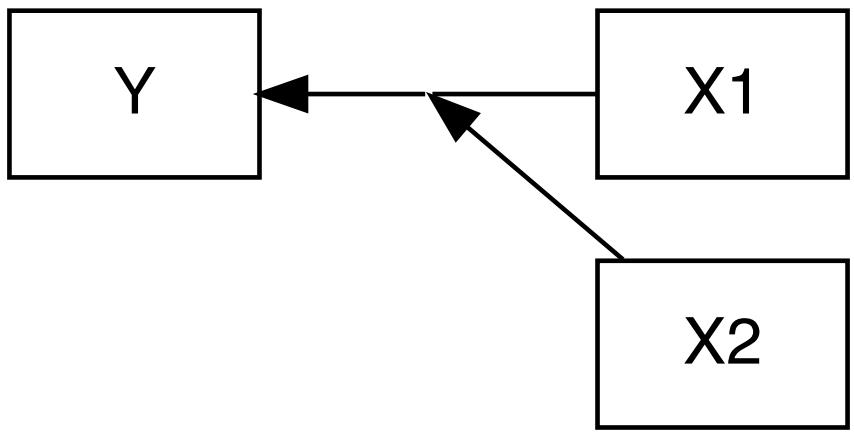
Indirect effect

- All the models we have seen so far, show direct effects.
- A direct effect is a directional relation between two variables, e.g., independent and dependent variables.
- Indirect effect is the effect of an independent variable on a dependent variable through one or more intervening or mediating variables.
- Include partial plots for both to illustrate the indirect effect.



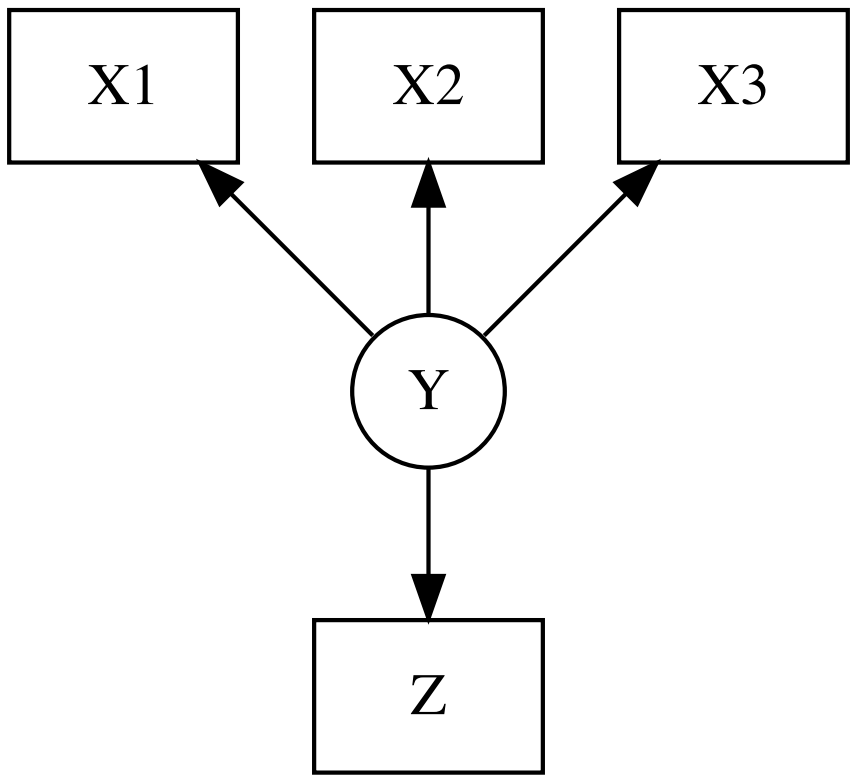
Interaction questions (Moderation)

- Present another question involving a new variable with an interactive effect (X3 -> (Y -> Z)).
- Show the graphical representation.
- Explanation of interactive effects.
- Include partial plots to illustrate the interaction effect.
- Make clear that this is the same as an interaction in traditional regression models.



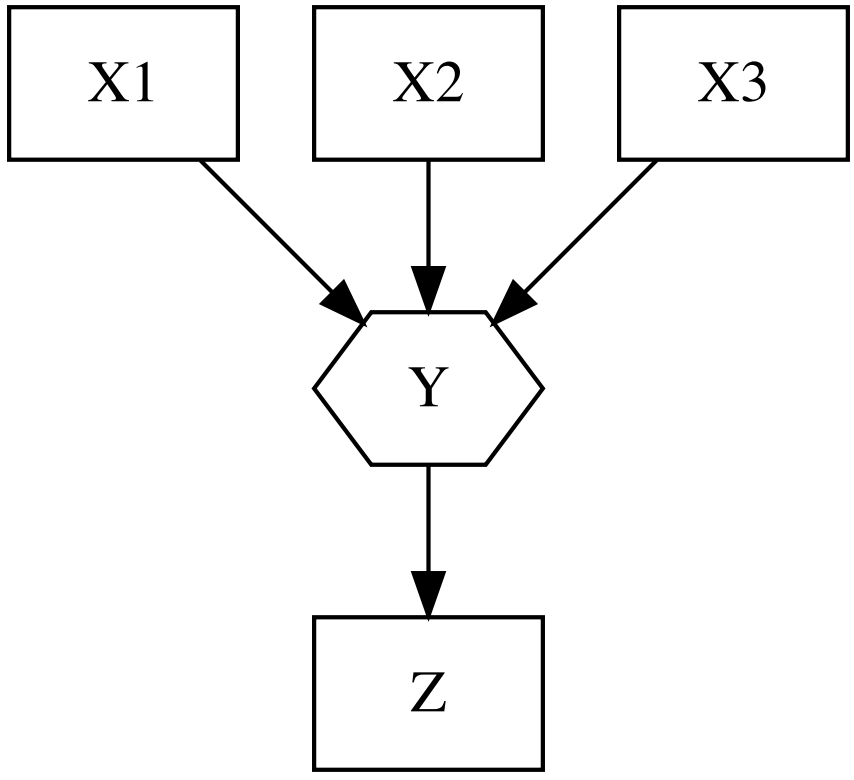
Latent variables

Until now everything was about paths but modelling also involves variables. A variable that is not directly measured is a latent variable (examples).



Composite variables

Composite variables specify the influences of collections of other variables (examples)



System level approach

Show a very complex model including multiple types of variables, arrows (direct and correlations) for the study system. The complex model is built based on submodels, each one addressing specific questions. We need a framework to analyze this together -> SEM
Discuss about complexity in ecology and if SEM is useful to answer your questions ("We need to do things as simple as possible, but no simpler").

Assumptions of SEM

Assumptions common to any statistical analysis (normality, amount of data, experimental design...). Assumptions specific to SEM (data in the same scale...).

Thinking about the model

Which variables include in the model - Supported by theory. - Ecologically meaningful. - Garbage in - garbage out concept (both data quality and ecologically meaningful).

Setting arrow direction. - Causal vs correlation. - Support for both directions, how to proceed (loops).

Good practice: Table of causal relationships before analysis.

Meta-model

Thinking about what is missing in the model, measurement errors etc. Present the concept of meta-model and impications for research. Include a graphical representation of the meta-model as what is missing in the model.

Modelling philosophy

Two approaches: from theory (hypothesis) to data (exploratories). Show as a continuum between two extremes. Importance of being aware about what the model is telling to us. Finall discussion about what SEM provides and the importance of system thinking in ecology.

SEM Steps

review the relevant theory and research literature to support model specification specify a model (e.g., diagram, equations) determine model identification (e.g., if unique values can be found for parameter estimation; the number of degrees of freedom, df, for model testing is positive) select measures for the variables represented in the model collect data conduct preliminary descriptive statistical analysis (e.g., scaling, missing data, collinearity issues, outlier detection) estimate parameters in the model assess model fit respecify the model if meaningful interpret and present results