Integrating monitoring data across different data types, locations and habitat types

Christian Damgaard

Bioscience

Aarhus University

Conceptual talk

Discussion of models that may be used to integrate data

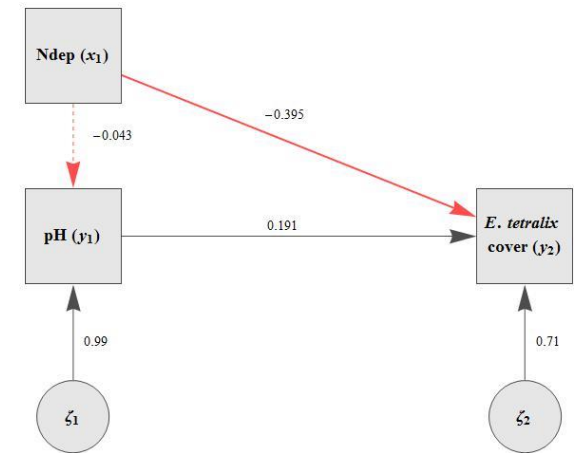Difficult to get funding for the work – why is that?

Structural Equation Modelling

Hypothesized causal mechanisms are specified in graphical models

Tests for conditional independence

Estimation of direct and indirect effects

Ecological predictions with a quantified degree of uncertainty

Correlation vs. causality
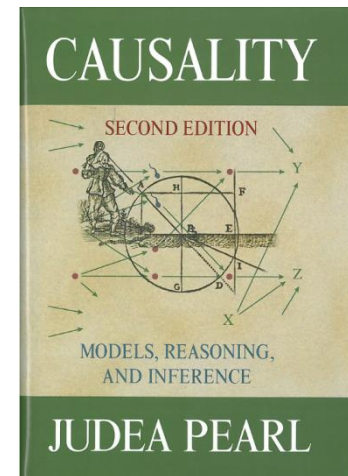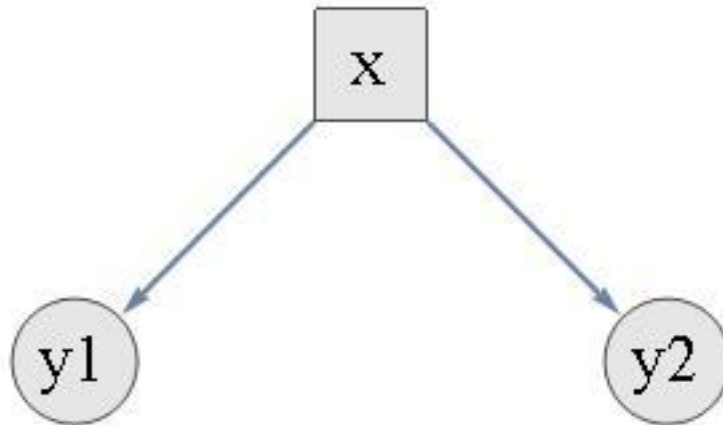
y1 and y2 are correlated

$$P(y1|y2) = P(y2|y1)\frac{P(y1)}{P(y2)}$$

Correlation vs. causality

If both y1 and y2 are regulated by x then the *residual* variation of y1 and y2 may be independent

= conditional independence

$$P(y2|x, y1) = P(y2|x) \ \ if \ y2|x \perp y1|x$$

Graphical models - SEM
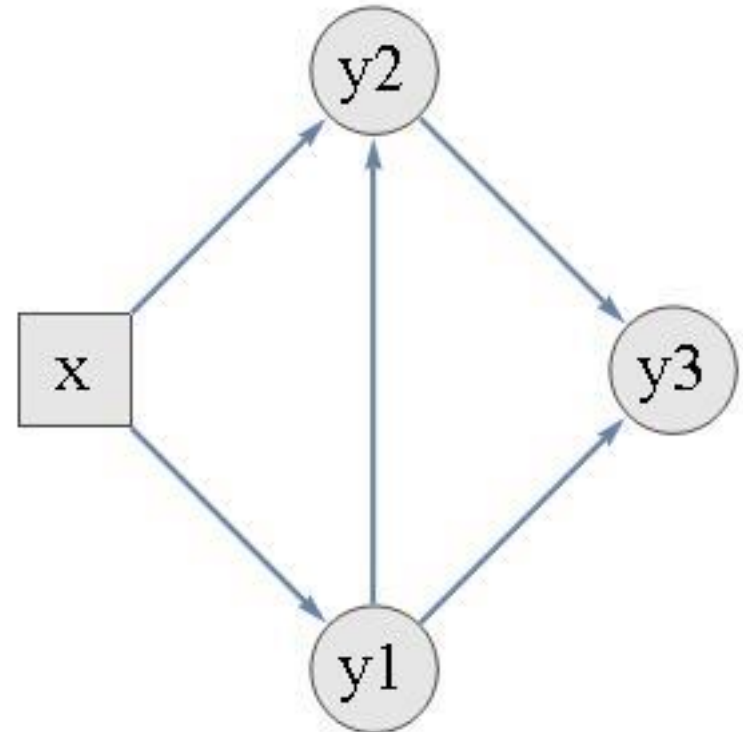
Graphical model of hypothesized causal relationships

X is an exogenous variable (outside the model)
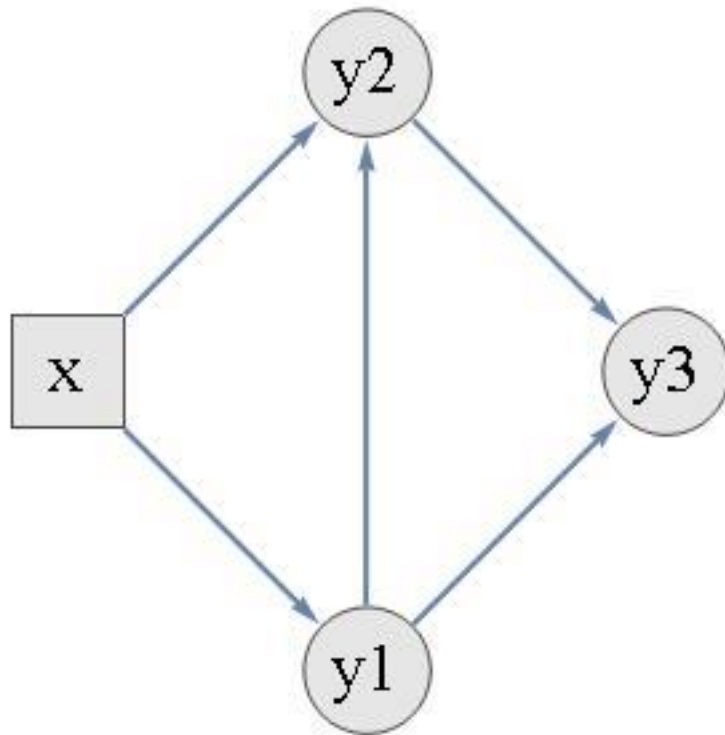
Y are endogenous variables

X has a direct effect on y1

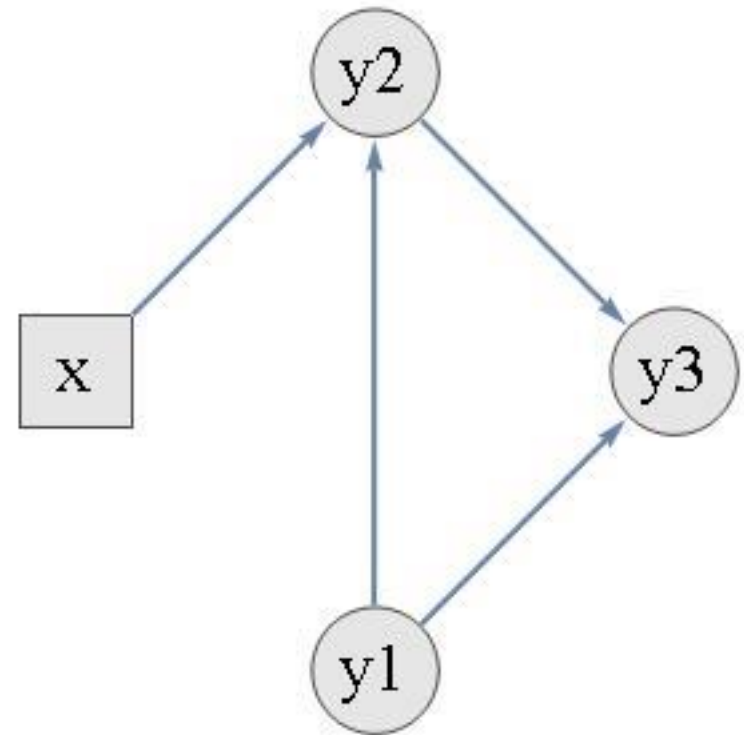**X has a direct effect on y2, but also an indirect effect mediated by y1**

X has only indirect effects on y3

Manipulating a factor (y1) in conservation / restauration

$$P(y1, y2, y3|x)$$

$$P(y2, y3|x, do(y1))$$

The causal relationships are best resolved from longitudinal data

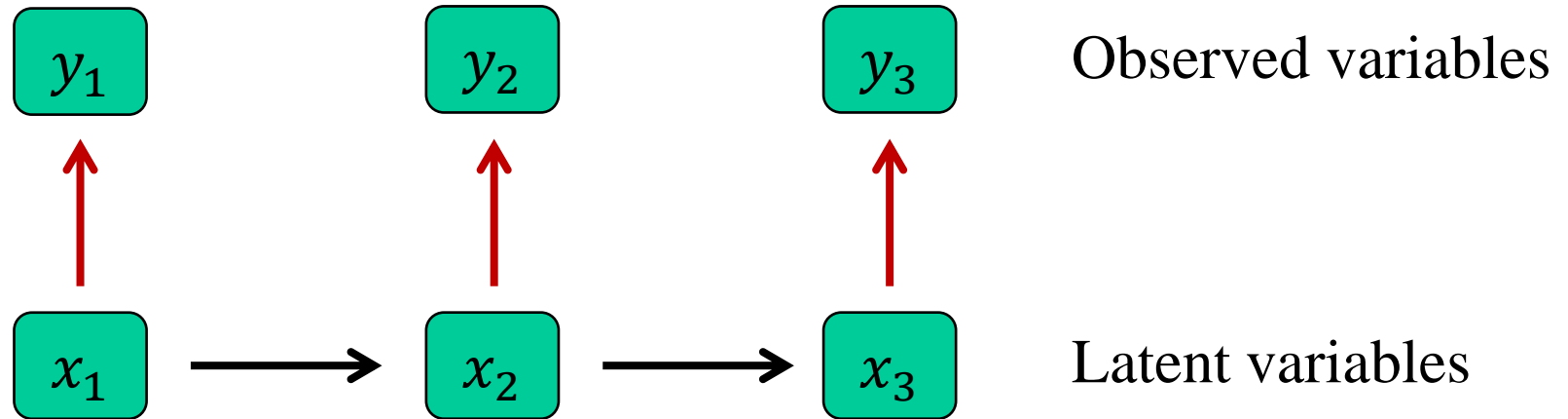**The events of today control what happens tomorrow**

Conditional independence

$\qquad$ **Markov chain (no history or time lag)**

$$P\left(x_{t=1}, x_{t=2}, x_{t=3}, x_{t=4}\right) =$$

$$P(x_{t=1})P(x_{t=2}|x_{t=1})P(x_{t=3}|x_{t=2})P(x_{t=4}|x_{t=3})$$

SEM – state-space model

$$y_1 \qquad y_2 \qquad y_3 \qquad \text{Observed variables}$$

$$x_1 \longrightarrow x_2 \longrightarrow x_3 \qquad \text{Latent variables}$$

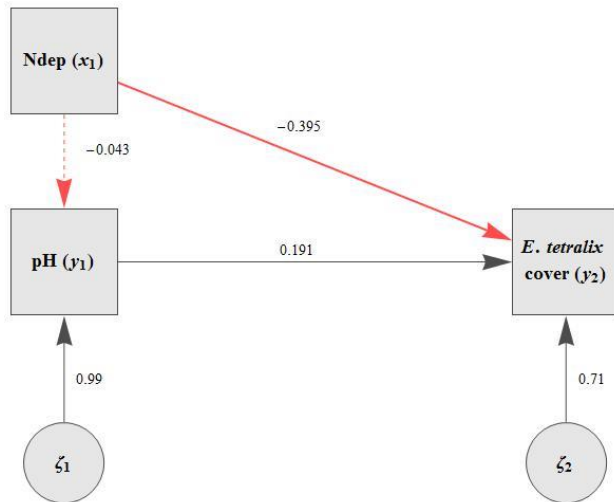Process: $x_i \sim N(f(x_{i-1}), \theta)$

Measurement: $y_i \sim \rho(x_i, \tau)$

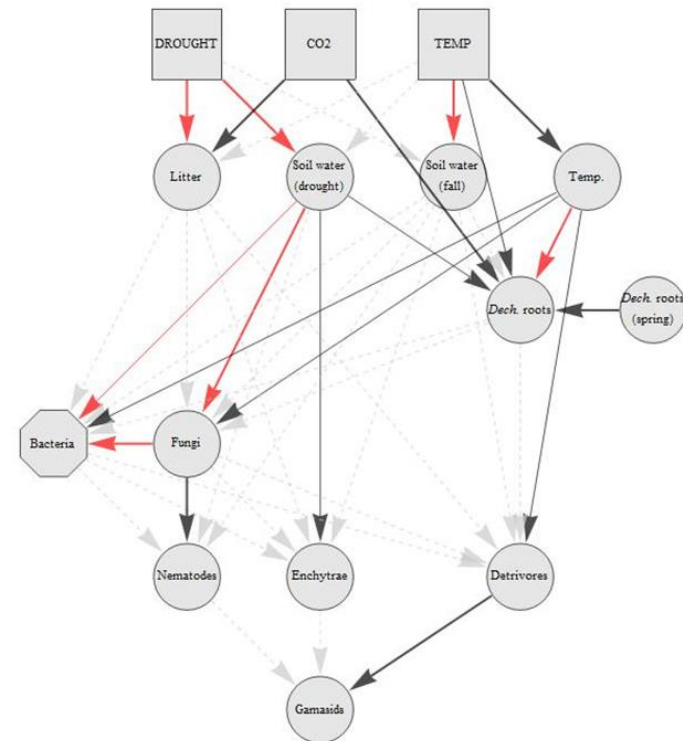Separation of process error and sampling error

SEM - examples

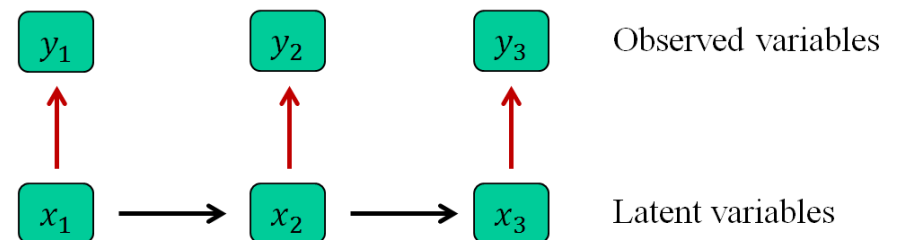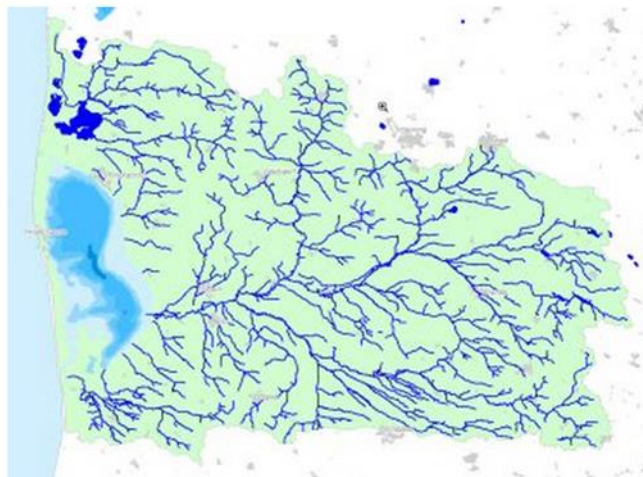## Nitrogen deposition and acidification

Climate change

Integration of different ecological processes

In order to study causality and the effect of restoration at the ecosystem level we need to integrate different ecological processes in one modelling framework

Models should be explicit in space and time

Process: $x_i \sim N(f(x_{i-1}), \theta)$
Measurement: $y_i \sim \rho(x_i, \tau)$

A network of data and knowledge

N-deposition: