An introduction to modelling natural systems

Shane Blowes

German Centre for Integrative Biodiversity Research,

Martin Luther University

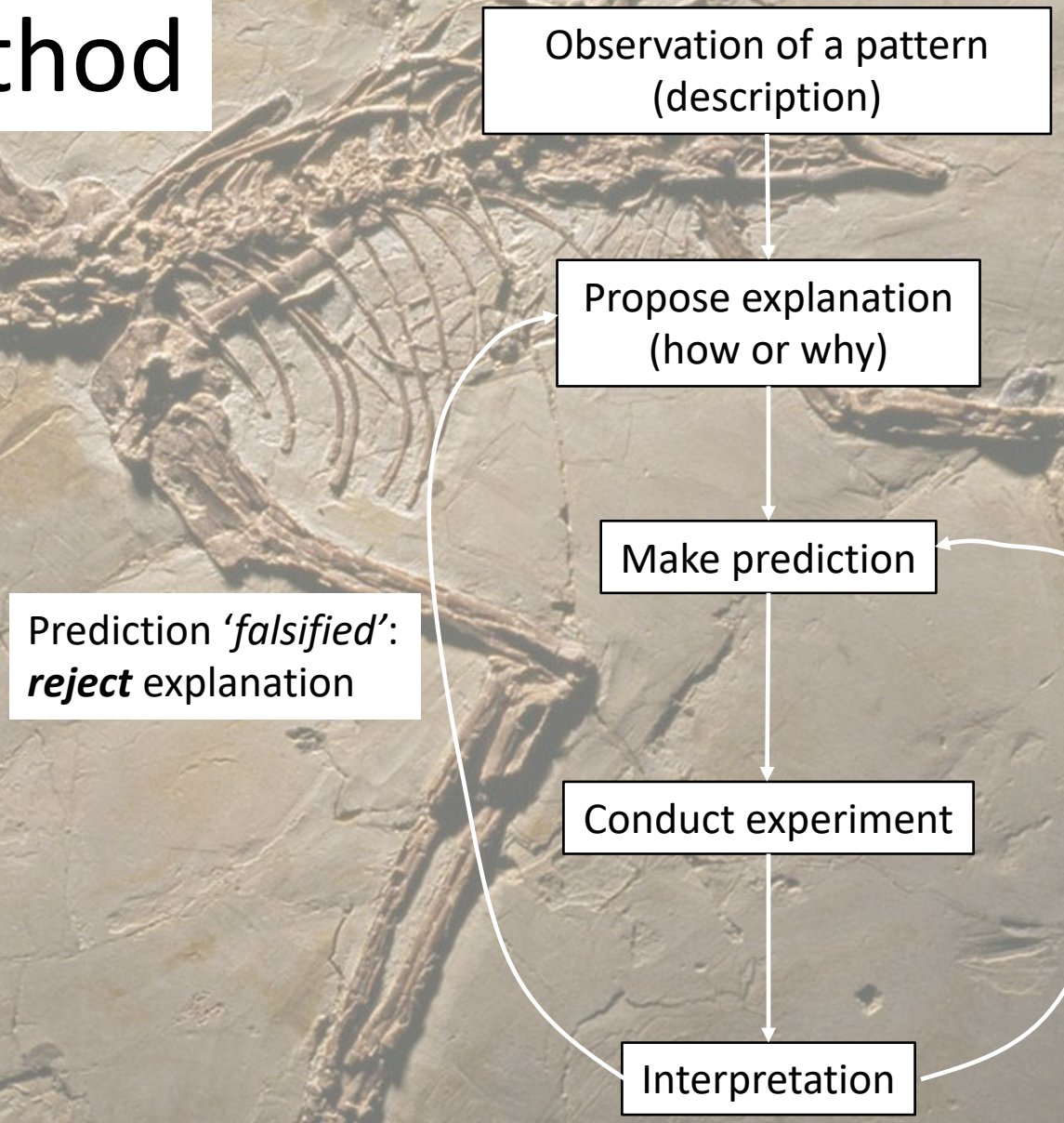
shane.blowes@idiv.de

Outline

- Models and the Scientific Method
- What are models?
- Types and uses of models

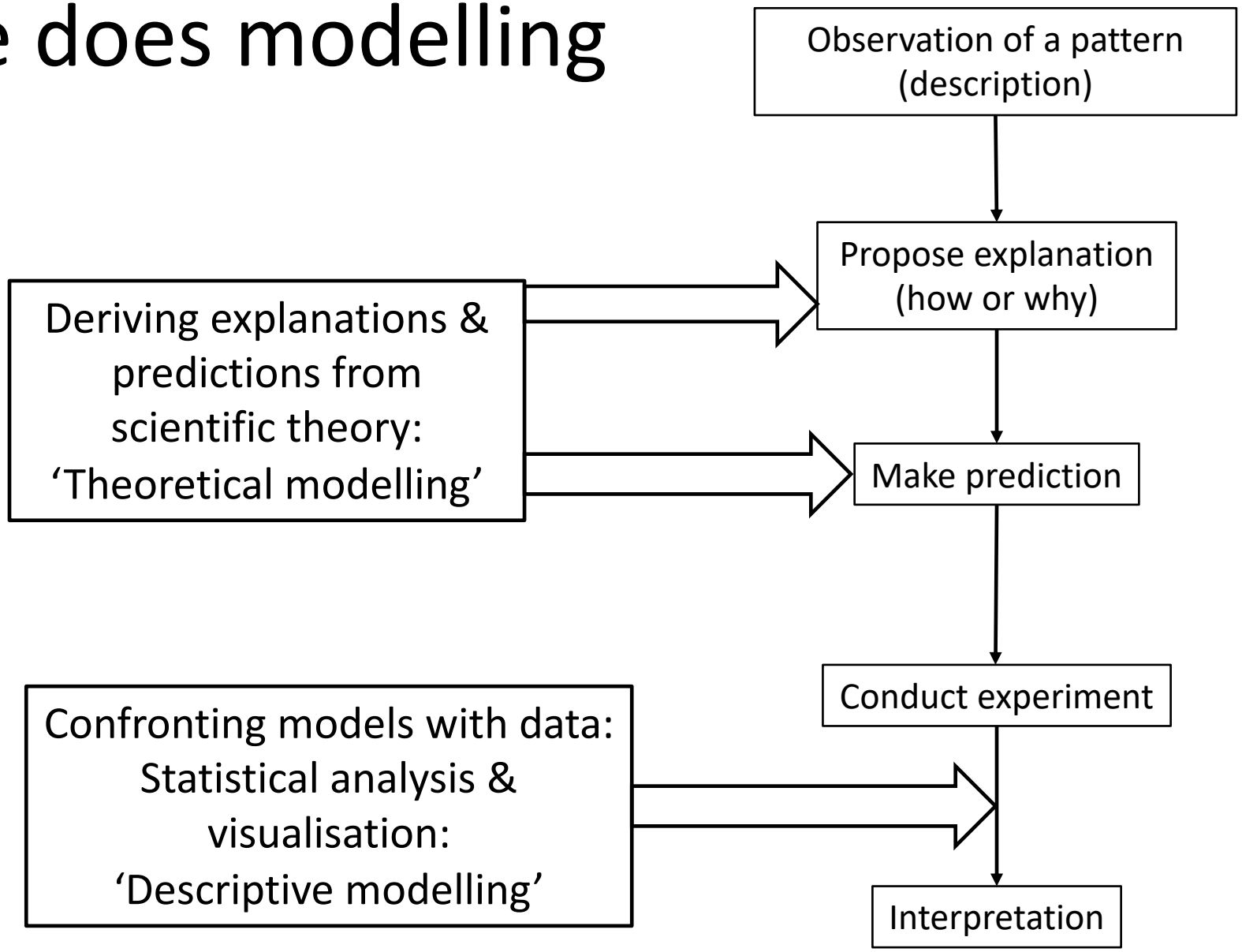
- Three types of descriptive models

Scientific Method



Carey 2004 *A beginner's guide to the Scientific Method*

Where does modelling fit in?



Modelling and Science

- Science is a dialogue between
  - what's really out there (observations)
  - what we are looking for (e.g., framework to make sense of observations, accurate predictions)
- Requires good observations: experimental design
  - probability sampling (e.g., random sampling)
  - systematic measurements
- Being mindful of preconceptions: modelling
  - specify clearly what assumptions are being made
  - rigorously deriving predictions from assumptions
  - using data to evaluate predictions

What are models?

- A model is a set of assumptions about how something in the natural world works

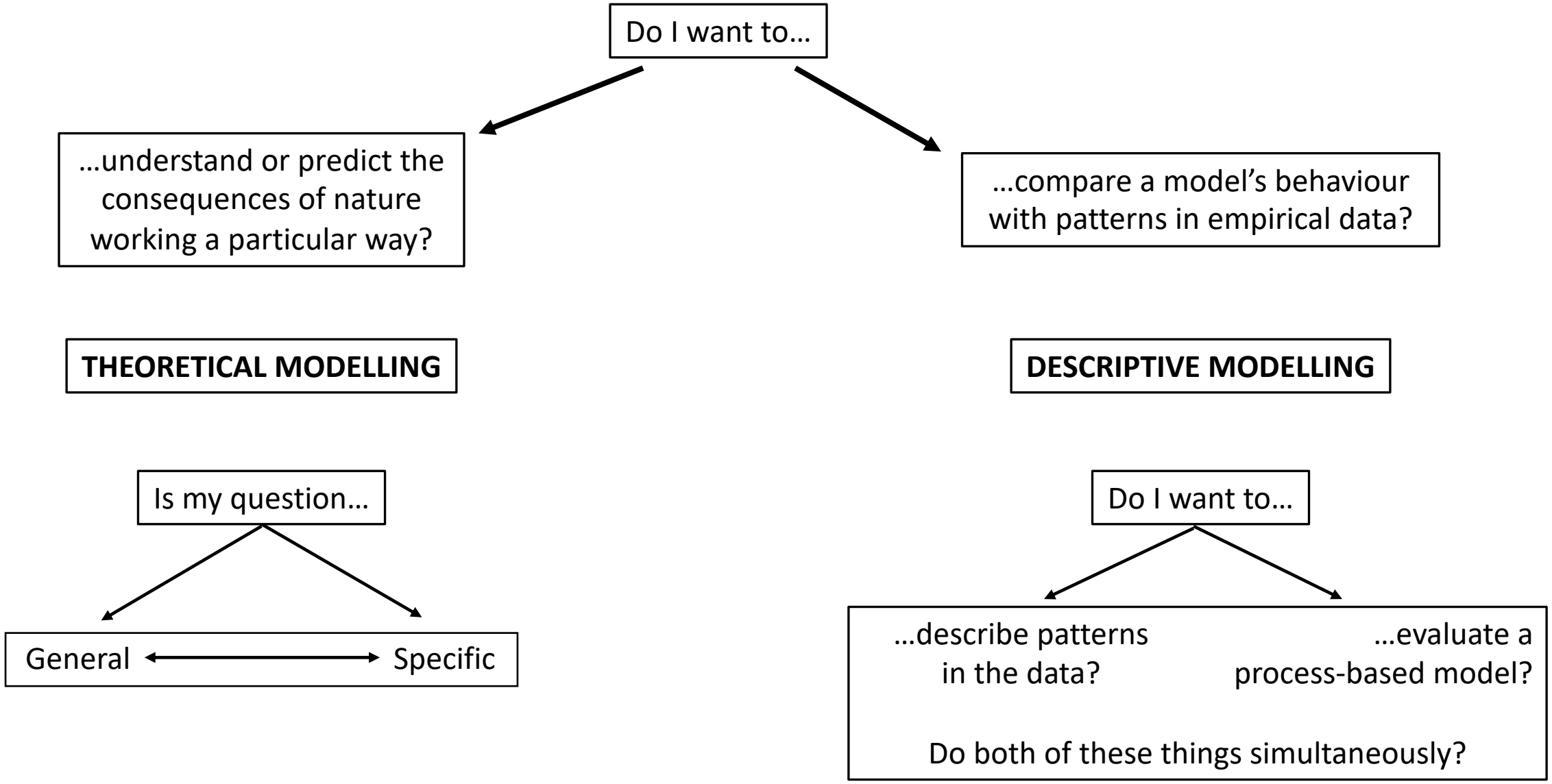
- Kinds of models
  - Conceptual (e.g., verbal or graphical) models
  - Scale models
  - Mathematical models

  (including computer simulations)

What are models?

- All models abstract and idealise
- All models are wrong
- We don’t ‘test’ models
  - test predictions
  - examine model assumptions
  - confrontation between empirical data and model predictions informs model development
  - Or, tells us what phenomena the model can and cannot help us understand

A modelling decision tree



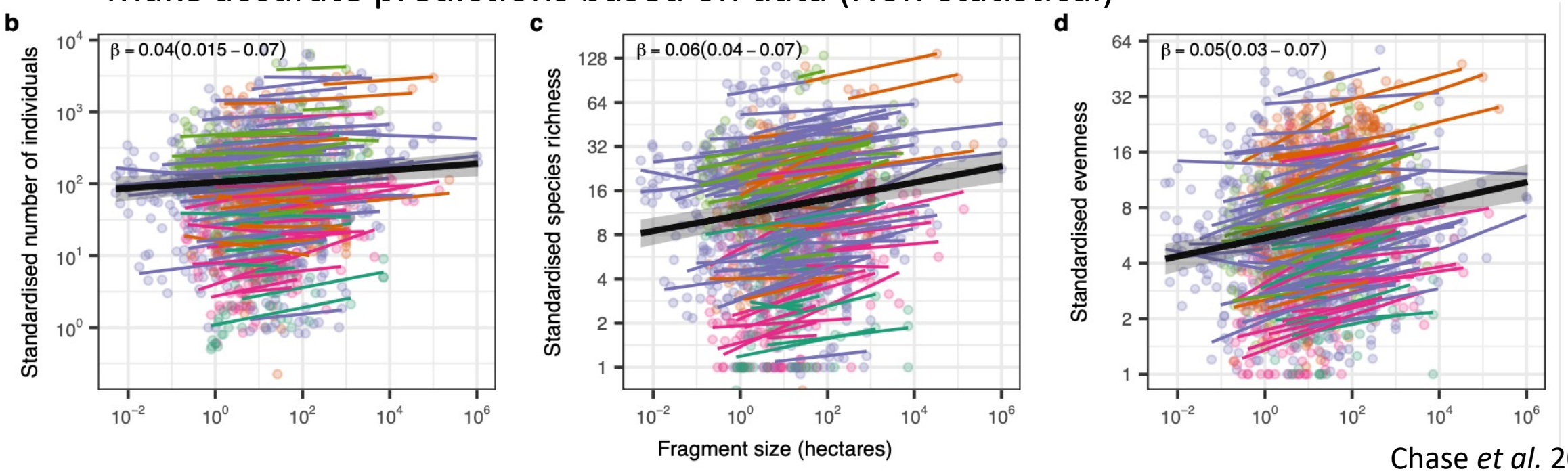
Different kinds of models

- Process-based models
  - represents a set of ideas about how nature works
  - used in theoretical and statistical modelling

- Descriptive model
  - a mathematical formula proposed to ‘fit’ an observed pattern in nature
  - only used in statistical modelling

Examples

- Theoretical modelling
  - What are the consequences of altered biological processes (e.g., reduced connectivity, changed resource availability) for biodiversity? (June 19 with Zach)
- Descriptive modelling
  - Test hypothesis of diversity and habitat fragment size relationship? (Statistical)
  - Make accurate predictions based on data (Non-statistical)



Chase *et al.* 2020 *Nature*

The perfect model (what we want)

- General
  - applies in a wide range of contexts
- Realistic
  - incorporates all the important processes
- Precise
  - predicts what happens in nature closely
- Simple
  - interactions can be comprehensively understood

Imperfect models (what we work with)

- Generality, realism, precision and simplicity all involve trade-offs
  - because nature is complex and variable
- For important questions, use different models that make different trade-offs

"Our truth…is the intersection of independent lies." --Richard Levins (theoretical ecologist)

Break

Part II

Descriptive models

{Trade-off}

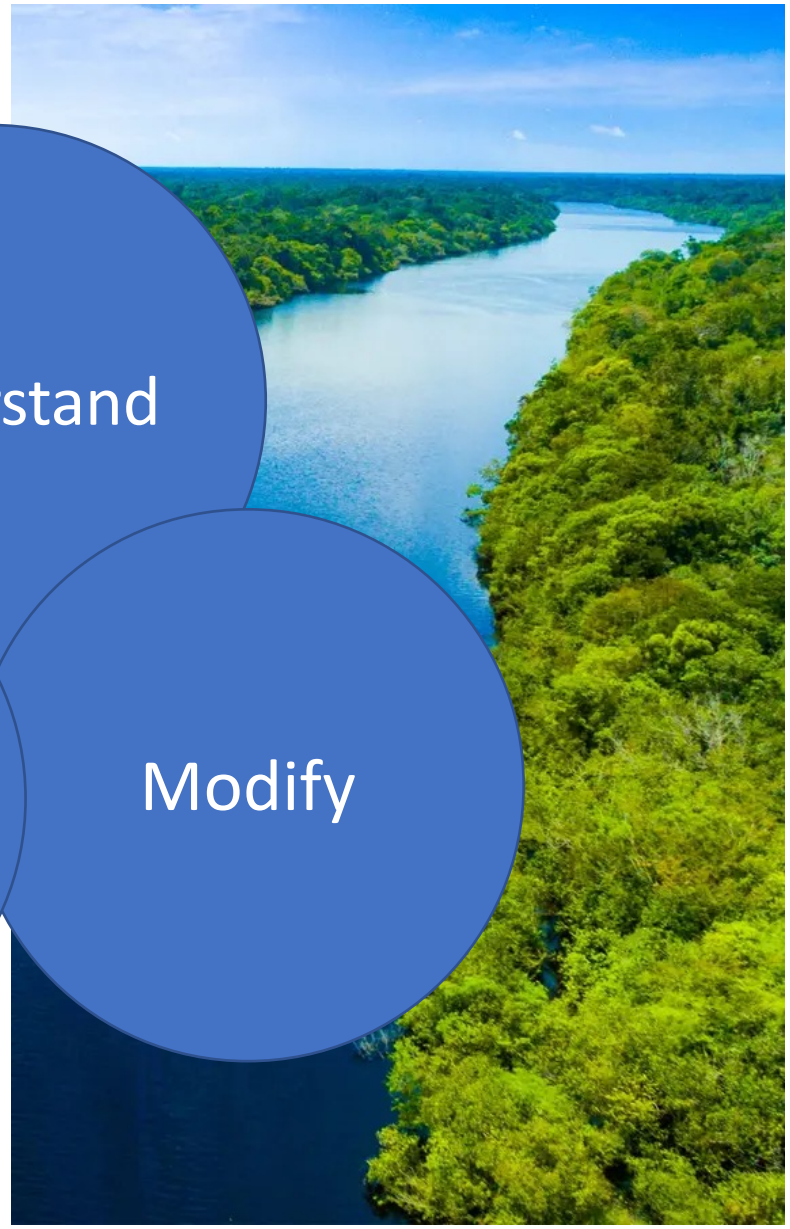
general, realistic, precise

models to

Understand

Predict

Modify



Levins 1966 *American Scientist*

Three part exploration of descriptive models

1. Parametric statistical models
2. Nonparametric statistical models
3. Nonparametric, non-statistical models

Underfitting

Overfitting

Regularisation

What are they?

What are they used for?

How to fit them?

Lucas 2020 *Ecological Monographs*

1. Parametric statistical models

Response ~ *f*(covariates),

- *f* defined in advance (functional form)
- Parametric: *f* expressed in terms of parameter(s)
- Statistical: include some kind of likelihood function, making model probabilistic (meaning e.g., estimates of uncertainty for predictions and parameter values can be made)
- Examples: Theoretical (process-based) models, Generalised Linear Models (GLMs), mixed effects models, time series models

Lucas 2020 *Ecological Monographs*

1. Parametric statistical models

$$y \sim Normal(\mu, \sigma),$$

$$\mu = \alpha + \beta x$$

2. Nonparametric statistical models

Response ~ *f*(covariates),

- Nonparametric: *f* not defined in advance (functional form unspecified)
- Statistical: include some kind of likelihood function, making model probabilistic (meaning e.g., estimates of uncertainty for predictions and parameter values can be made)
- Examples: Splines, Generalised Additive Models (GAMs), Gaussian Process regression

Lucas 2020 *Ecological Monographs*

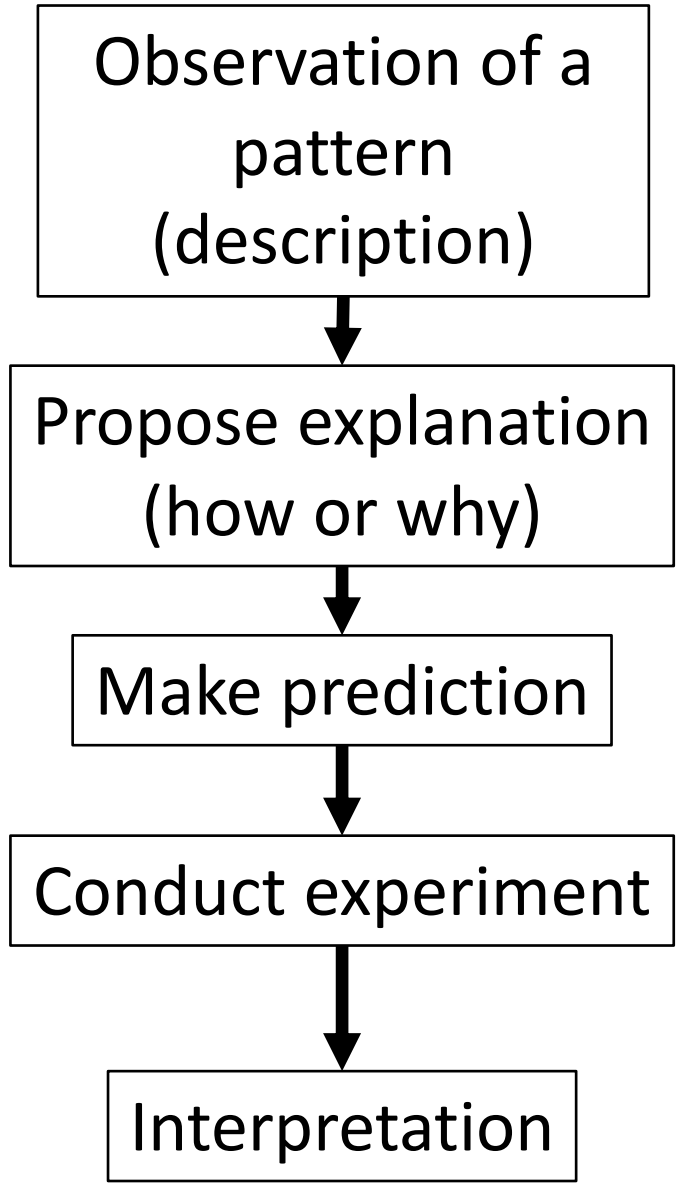
2. Nonparametric statistical models

$$y \sim Normal(\mu, \sigma),$$

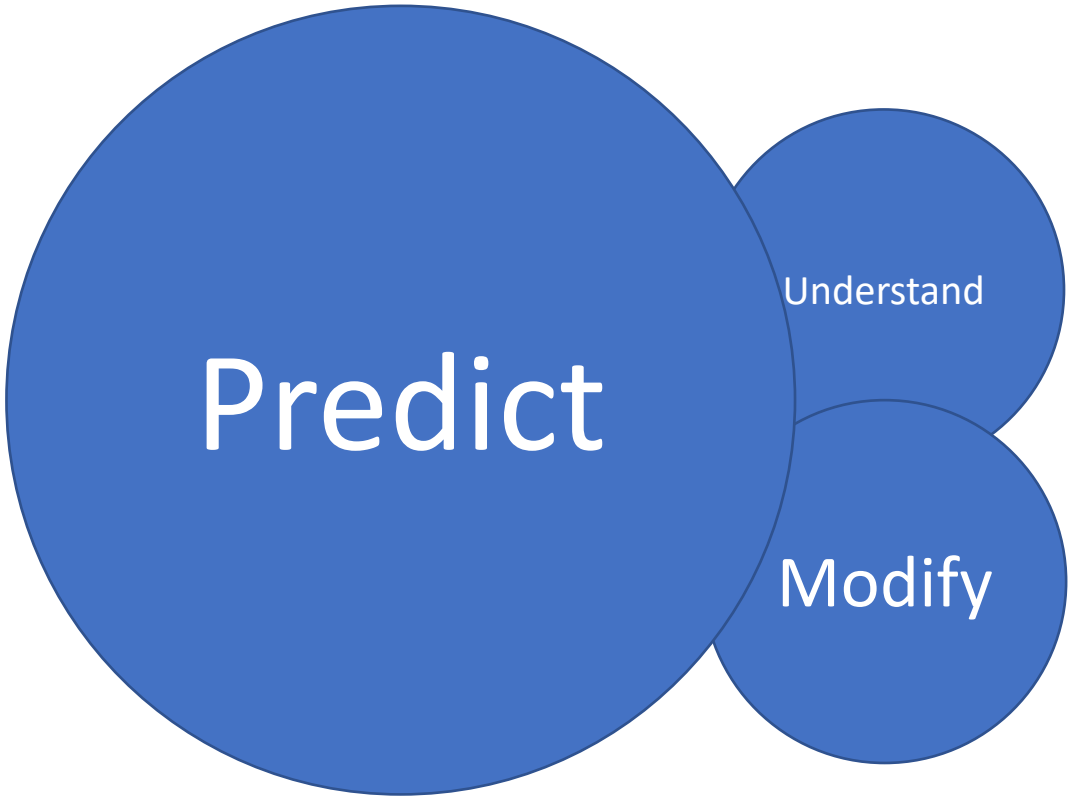
$$\mu = f(x)$$

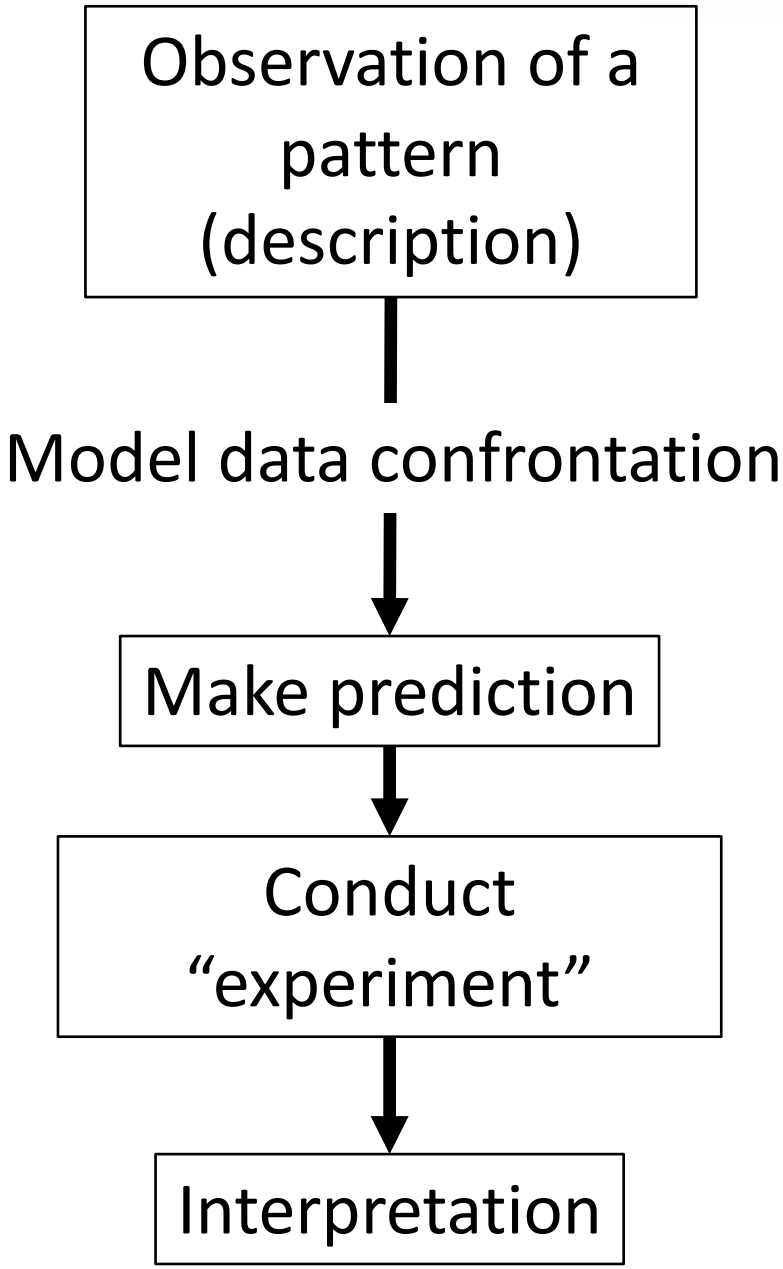
3. Nonparametric, non-statistical models

Recall our first model of the scientific method



3. Nonparametric, non-statistical models





3. Nonparametric, non-statistical models

Response ~ *f*(covariates),

- Nonparametric: *f* not defined in advance (functional form unspecified)
- Non-statistical: not fully probabilistic (meaning e.g., that they can have many parameters with no estimates of uncertainty)
  - Uncertainty typically estimated using bootstrap
- Examples: decision trees, Random Forests, boosted regression trees, neural networks, support vector machines,

Lucas 2020 *Ecological Monographs*

3. Nonparametric, non-statistical models

- Machine Learning
  - Response (Machine Learning [ML] equivalent: labels)
  - Predictors or covariates (ML: features)
  - Data (ML: instances)
- Differs from broader field of statistics
  - Less emphasis of estimating interpretable parameters
  - Predictions driven by data (not necessarily related to theory)
  - Focus on prediction, not understanding

Lucas 2020 *Ecological Monographs*

Break

Afternoon: Computer exercises