Looking back: how far have we come?

Tanjona Ramiadantsoa

with materials form Calistus Ngonghala

Genesis

E^2M^2: Ecological and Epidemiological Modeling in Madagascar

January 13-20 & January 22, 2018

Centre ValBio, Ranomafana National Park & Institut Pasteur de Madagascar, Antananarivo

Applications available at:

http://metcalflab.princeton.edu/e2m2-application/

Deadline: Wednesday, November 1, 2017

We are pleased to announce the second annual *E^2M^2: Ecological and Epidemiological Modeling in Madagascar* clinic, to be held January 13-20, 2018 at Centre ValBio, Ranomafana National Park, Madagascar, with a mandatory closing session to follow at Institut Pasteur de Madagascar on January 22. The clinic will be a ten-day intensive workshop aimed to provide an introduction to the use of dynamical models in understanding ecological and epidemiological data.

Students will participate in a series of interactive lectures and computer-based tutorials and learn to fine-tune model-based research questions, develop clear model frameworks and corresponding equations, and fit models to real-world data. All students will work closely with peers and faculty to develop a research plan for an ongoing or existing project integrating dynamical modeling with data collection and/or analysis in a

Outline

- Research question and interest
- Mechanistic model
- Statistical model
- R

Research question

Abstract

Title: "Modeling polymerization of branched actin filaments"

In this project we will construct Master equations and Fokker Planck equations for the stochastic process of the (de)polymerization of actin filaments. The basic case of a single straight filament with simplified dynamics was considered in class during the biological physics course. This project will extend this case to consider multiple (de)polymerizing filaments, stochastic nucleation of new filaments and filament branches. Where possible, the stochastic differential equations will be solved analytically. When this is not possible we will solve the equations numerically. The results will be compared to data from biochemistry experiments and stochastic simulations done by member of the team at the University of Sheffield.

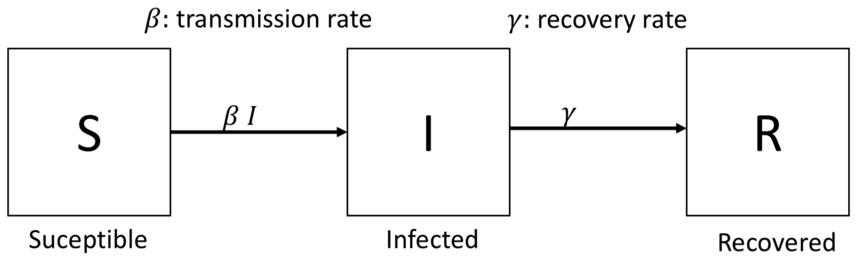
Research question

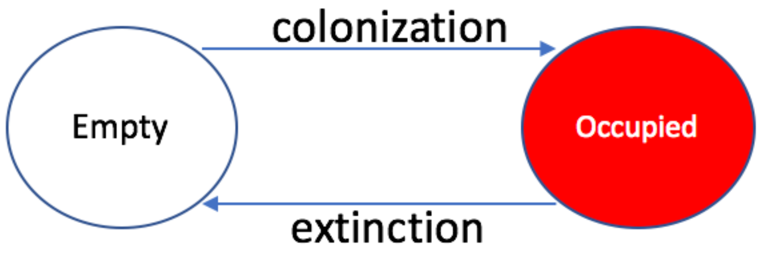
How…? = mechanistic model

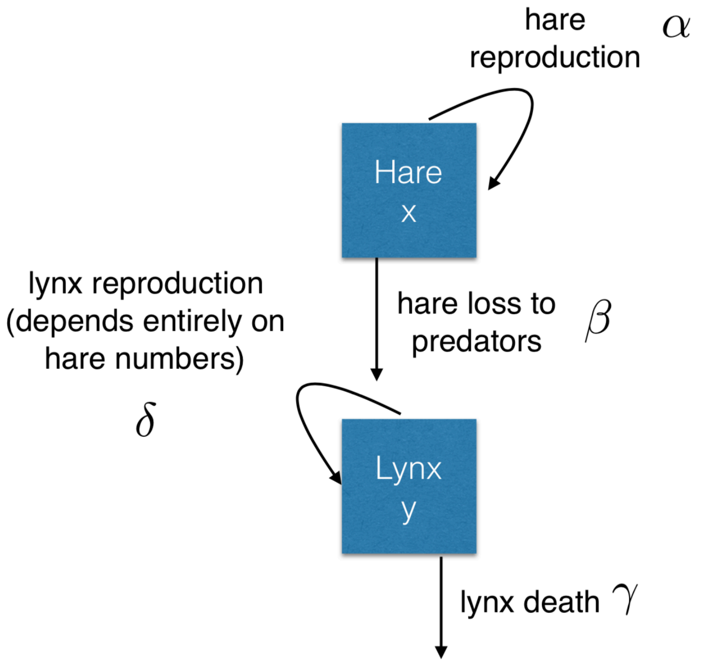
Does…? = statistical model

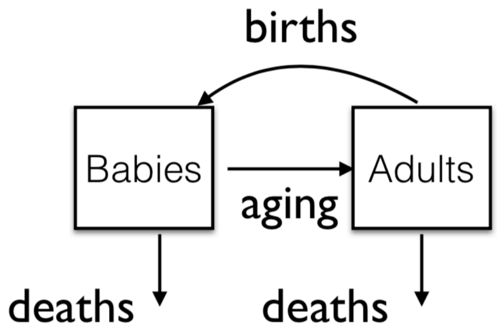
Mechanistic model: **process**-driven

- Compartmental model
- State: box
- Process: arrow









Equivalence?

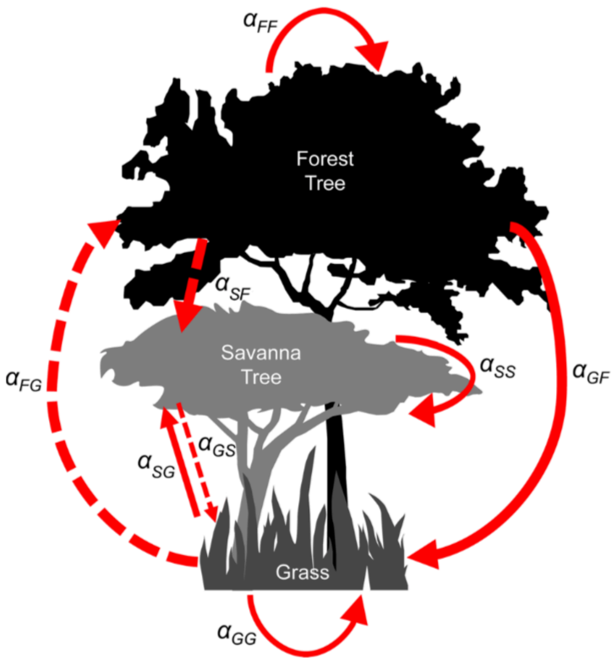


Figure 1. A pictorial representation of the model. *Thicker arrows* represent stronger limitation (larger competition coefficient values). The pathway of *different dashed arrows* represents a potential opportunity for savannas to facilitate their own growth via reduced limitation on grasses, which repels forest trees. For parameter definitions, see Table 1.

~

$$dS/dt = r_S S(1 - \alpha_{SG}G - \alpha_{SF}F - \alpha_{SS}S) + \xi_1$$

$$dF/dt = r_F F(1 - \alpha_{FG}G - \alpha_{FF}F) + \xi_2$$

$$dG/dt = r_S G(1 - \alpha_{GS}S - \alpha_{GF}F - \alpha_{GG}G) + \xi_3$$

~

The model is a formulation of the classic Lotka–Volterra competition model, with notation following Chesson (2000). The model aims to capture the dynamics of vegetation in areas with annual precipitation ranging from about 800–2000 mm, where savanna and forest potentially constitute alternative self-reinforcing states maintained primarily through differences in fire dynamics (for example, Sankaran and others 2005; Staver and others 2011). The model has three functional groups: grasses (G), savanna trees (S), and forest trees (F), and their abundance is expressed in terms of vegetation cover per unit ground at the spatial resolution of an area that would fit one fully grown forest tree and multiple individual grasses. We assume that cover is proportional to the size of a single tree in this patch. Together, the three groups do not necessarily sum to 100%, because vegetation can have multiple, overlapping layers and systems with more species often achieve greater biomass/cover than monocultures, through complementary use of resources or other mechanisms (Tilman and others 2014).

The dynamics of these three functional groups are given by the following three equations:

$$dS/dt = r_S S(1 - \alpha_{SG}G - \alpha_{SF}F - \alpha_{SS}S) + \xi_1$$

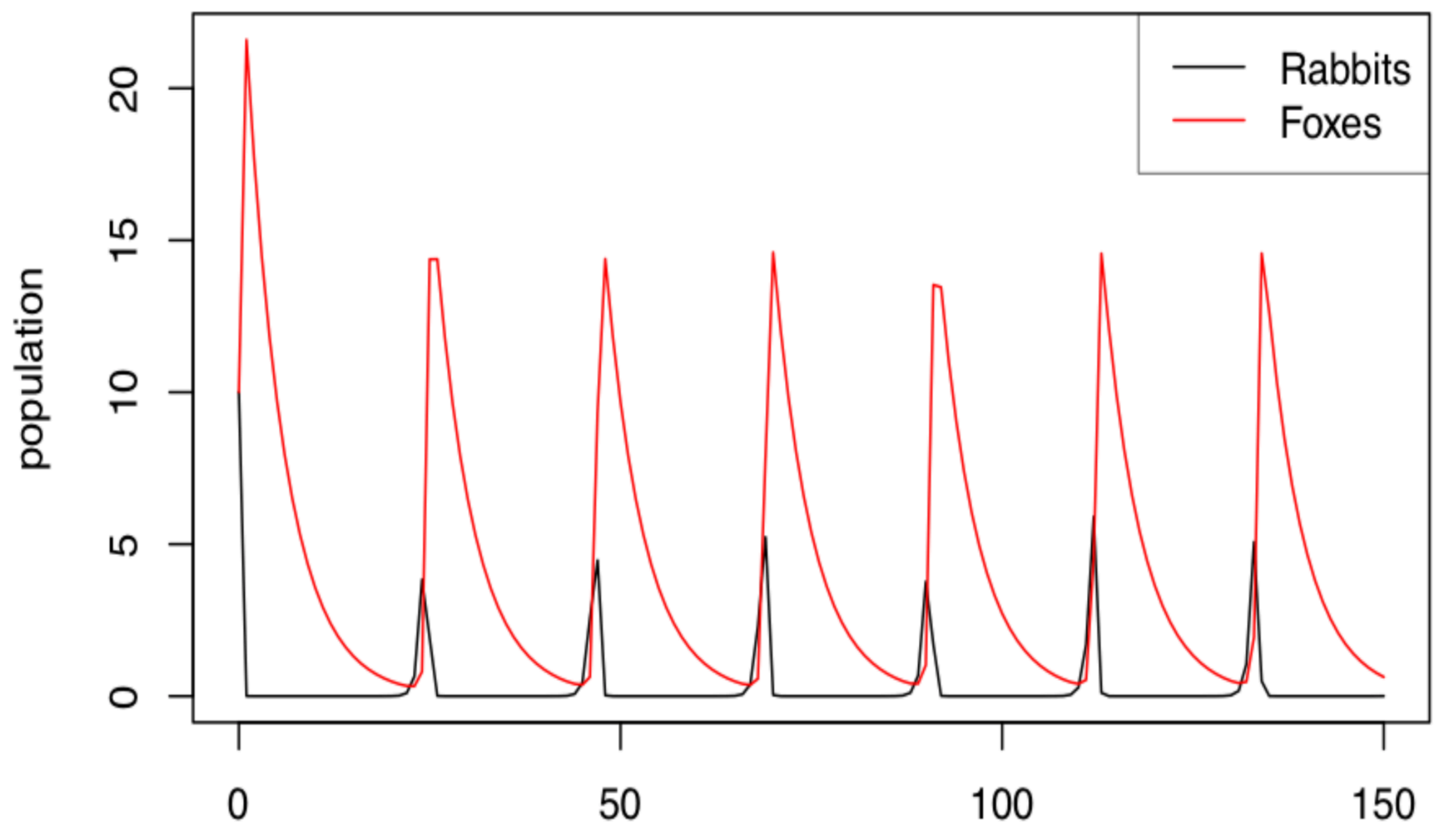
$$dF/dt = r_F F(1 - \alpha_{FG}G - \alpha_{FF}F) + \xi_2$$

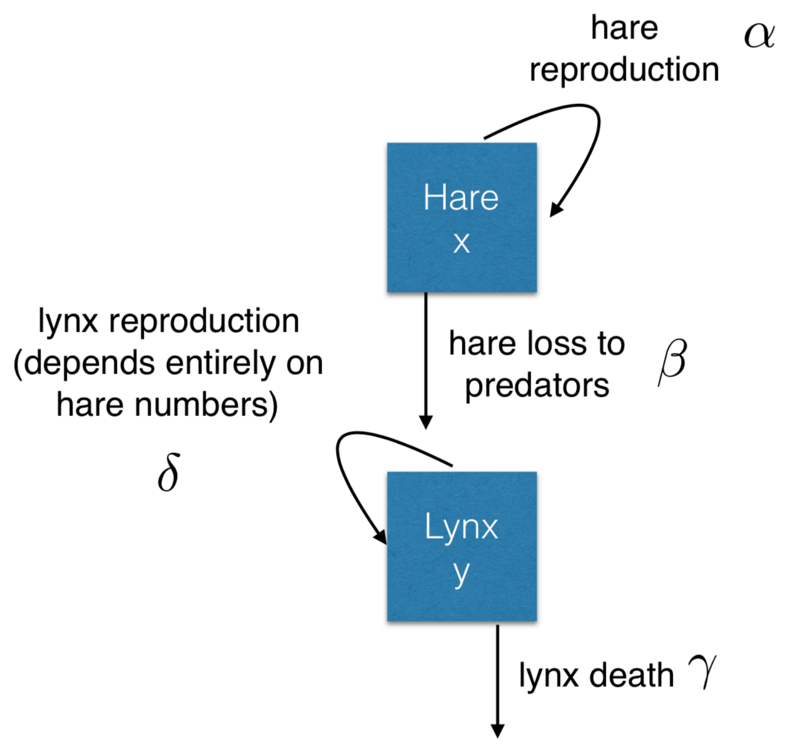
$$dG/dt = r_S G(1 - \alpha_{GS}S - \alpha_{GF}F - \alpha_{GG}G) + \xi_3$$

r_S, r_F, and r_G are maximum growth rates. α_{ij} are competition coefficients, representing the competitive effect of species j on species i, with larger values indicating that species j reduces species i's growth more. The α_{ii} coefficients, therefore, are self-limitation coefficients that can be interpreted as 1/carrying capacity. Each equation is also driven by independent additive Gaussian white noise ξ_{1-3}, accounting for sources of uncertainty associated with disease, herbivory, resource variability, and other stochastic factors (for example, Ridolfi and others 2011).

Ratajczak et al. 2017

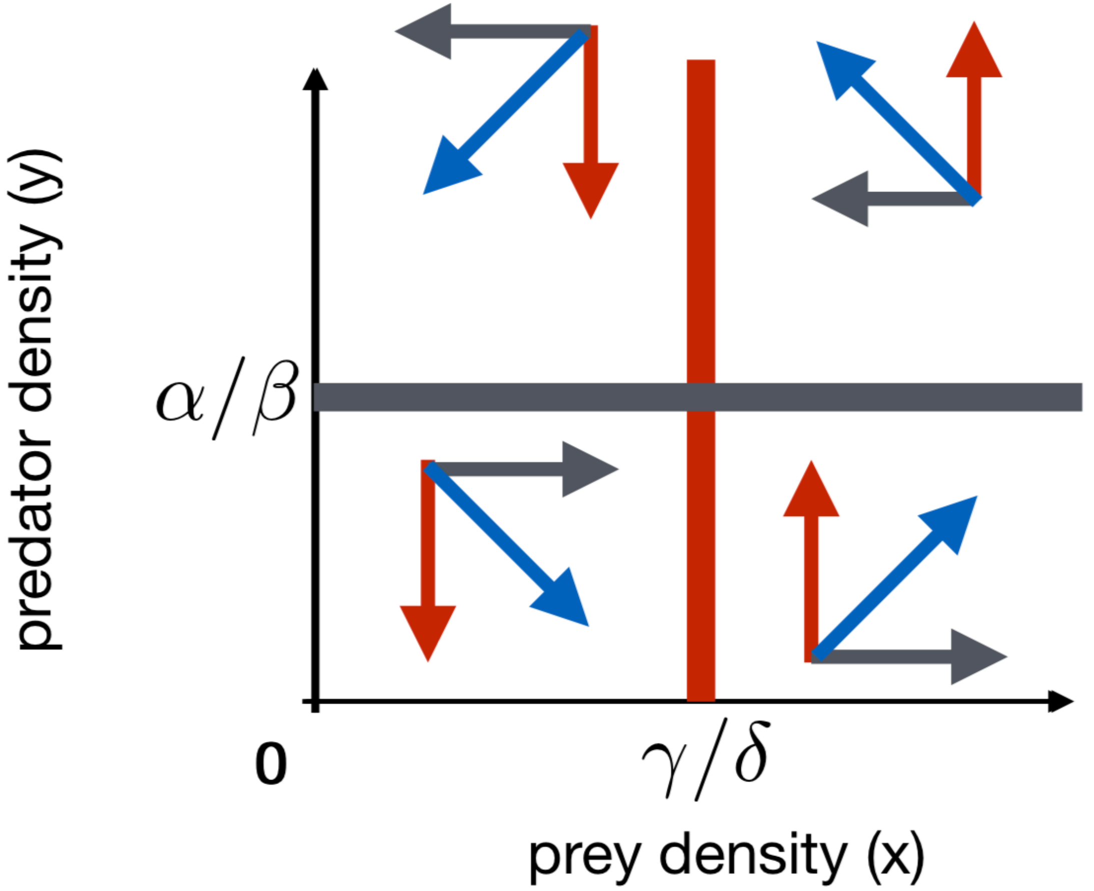
Analyses: forward simulation





See R later

Analyses: phase plane



Analyses: equilibrium

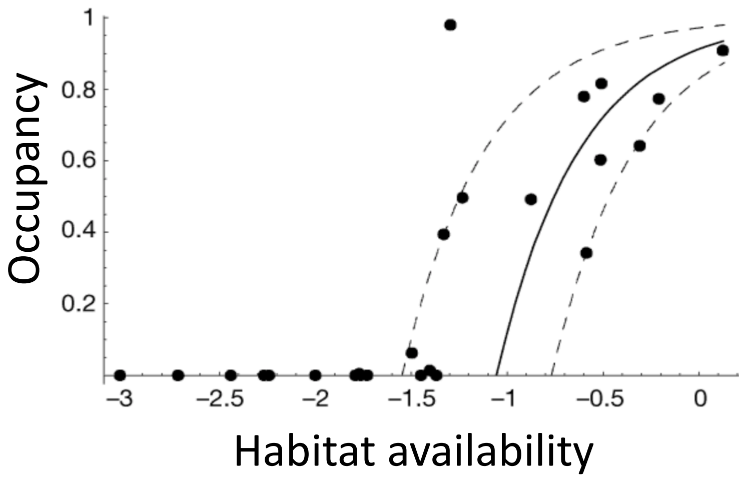
- Equilibria

$$\frac{dp}{dt} = 0 \iff cp\,(1 - p) - ep = 0$$

$$\iff p^* = 0 \text{ or } p^* = 1 - \frac{e}{c}$$

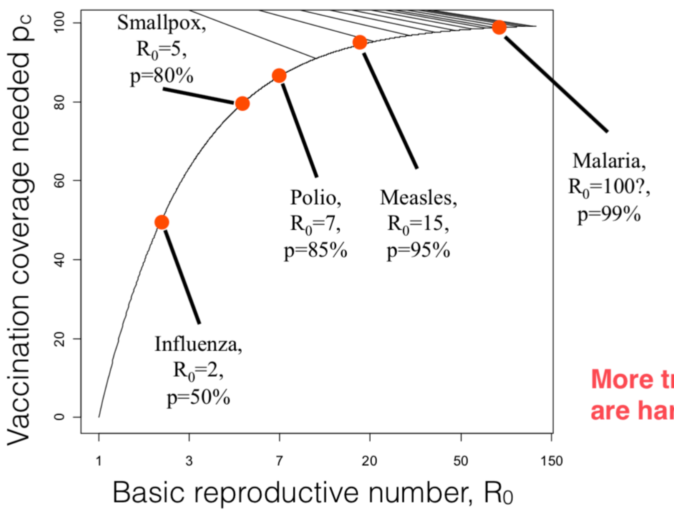
Insights

- Exponential increase
- Cyclic dynamic
- Extinction threshold
- Basic reproductive number R_0
- Effective reproductive number R_E
- Vaccination cover p_c
- …



Hanski & Ovaskainen, 2000, *Nature*

Same logic as without births: $p_c = 1 - \frac{1}{R_0}$



More transmissible diseases are harder to eradicate

Families of models

Discrete models

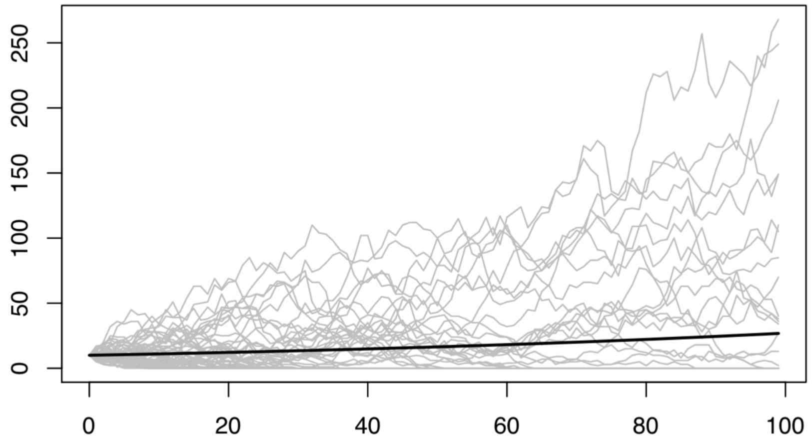
Structured population model

$$\mathbf{n_{t+1}} = \mathbf{A}\,\mathbf{n_t}$$

Continuous models

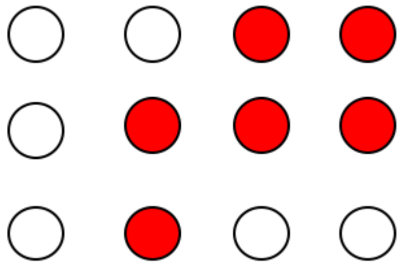
$$\frac{dS(t)}{dt} = -\beta S(t)I(t)$$
$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t)$$
$$\frac{dR(t)}{dt} = \gamma I(t)$$

Deterministic and stochastic models

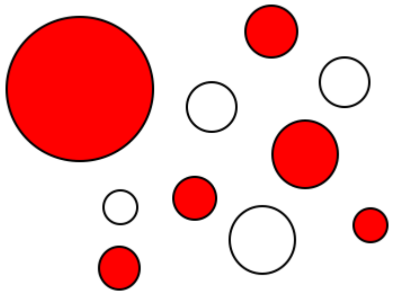


Spatial models

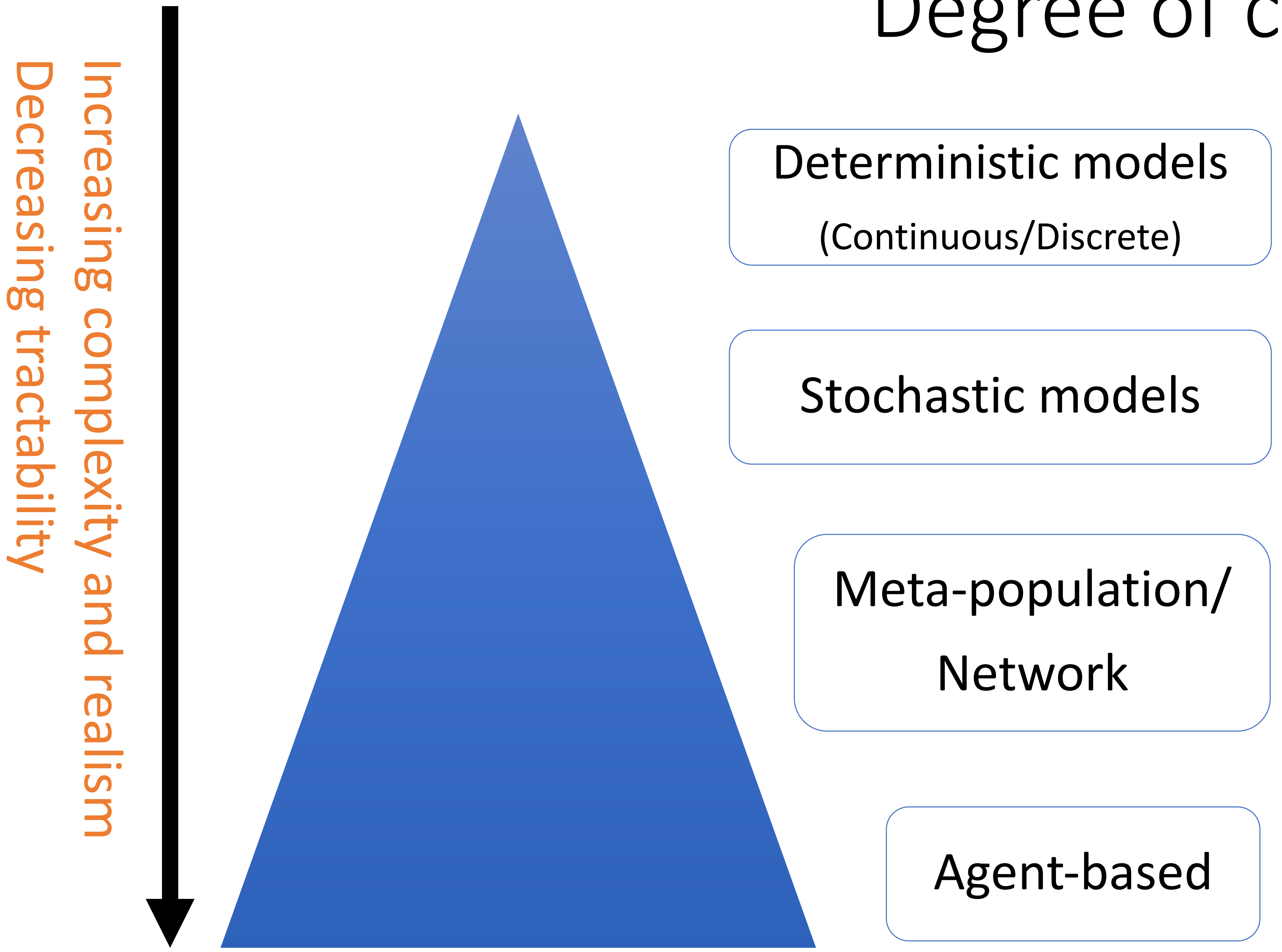
Spatially implicit: homogenous



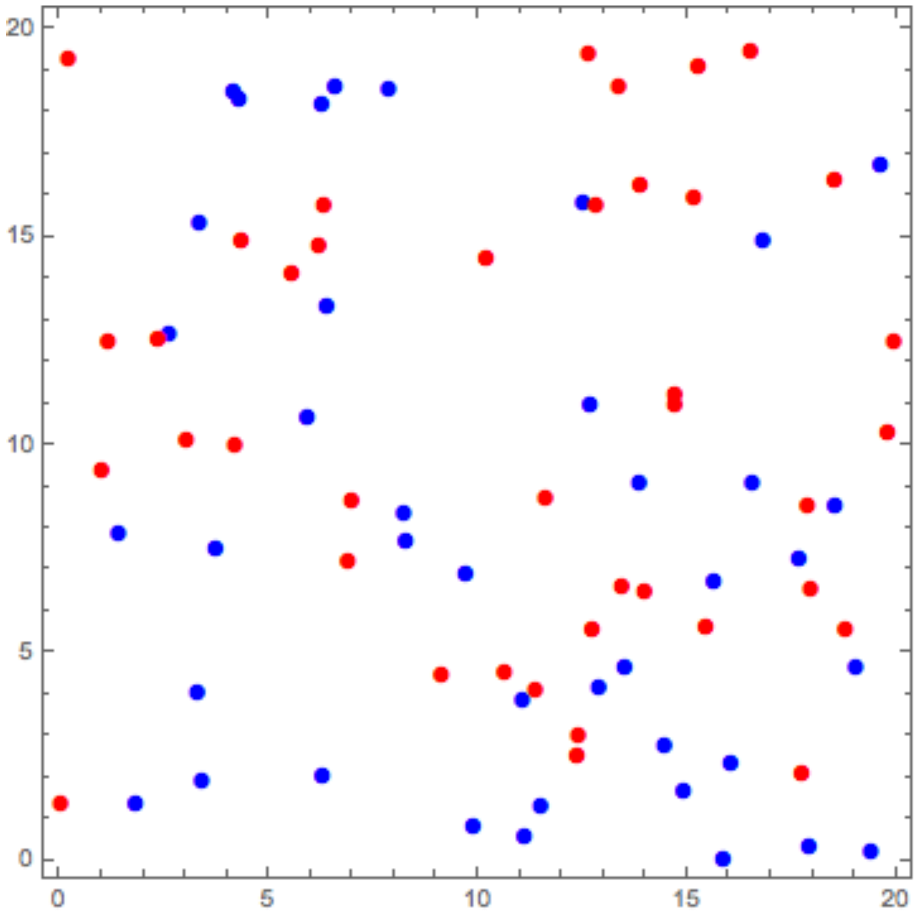
Spatially realistic: heterogeneous



Degree of complexity



Agent based: continuous landscape



Outline

- Science: research question
- Mechanistic model
- Statistical model
- R

Statistical modeling: **data**-driven

POSITIVE PROOF OF GLOBAL WARMING

Correlation does not imply causation

Regression families and assumptions

1. Univariate Linear Models

↓

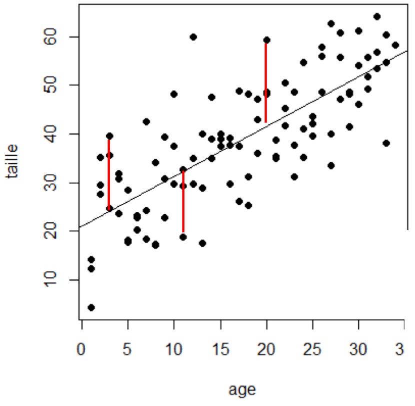
2. Multivariate Linear Models

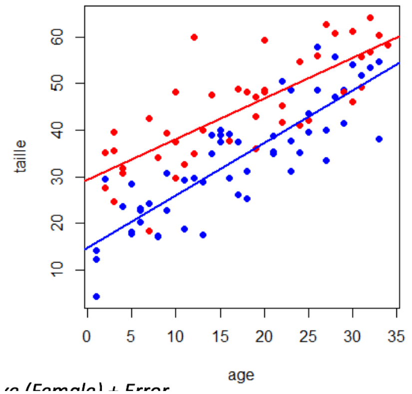
↓

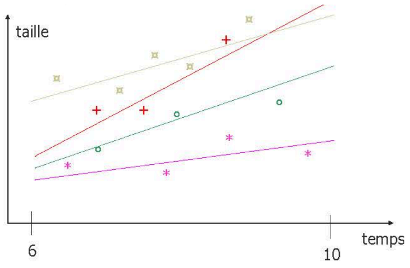
3. Generalized Linear Models

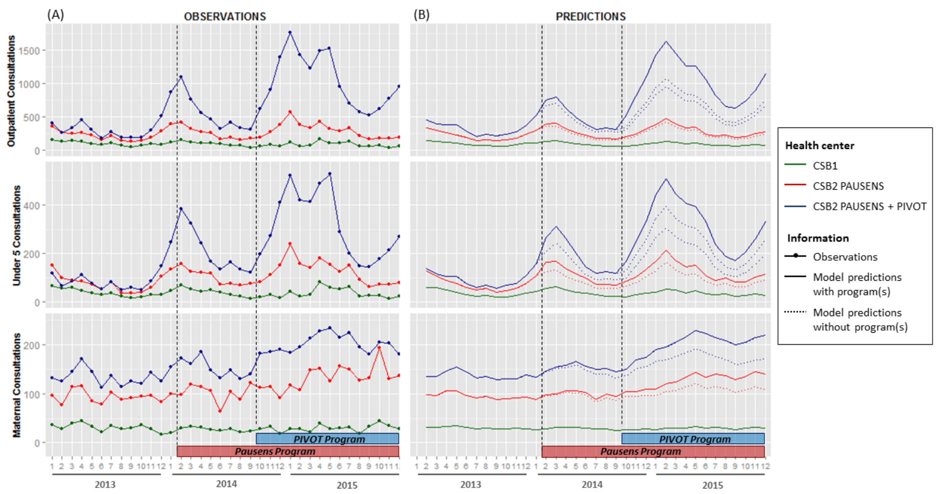
↓

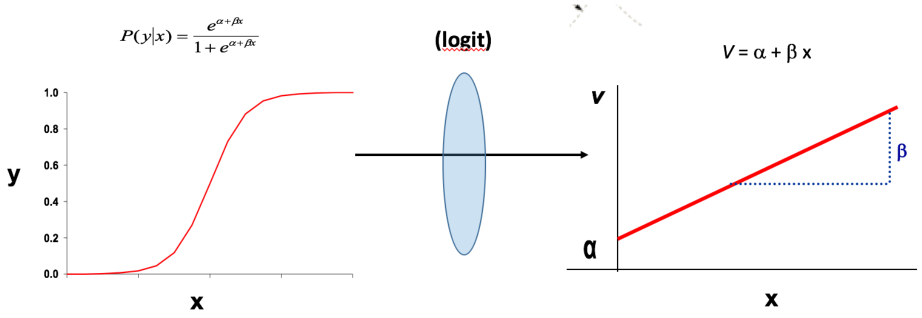
4. Generalized Linear Mixed Models → 5. Evaluating trends and effects over time









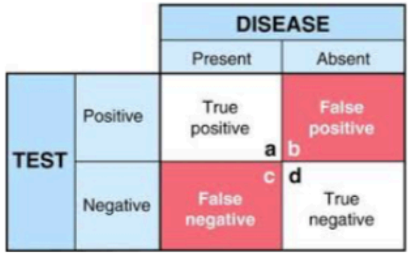


Occupancy model

ESTIMATE DISEASE PREVALENCE

Reality:

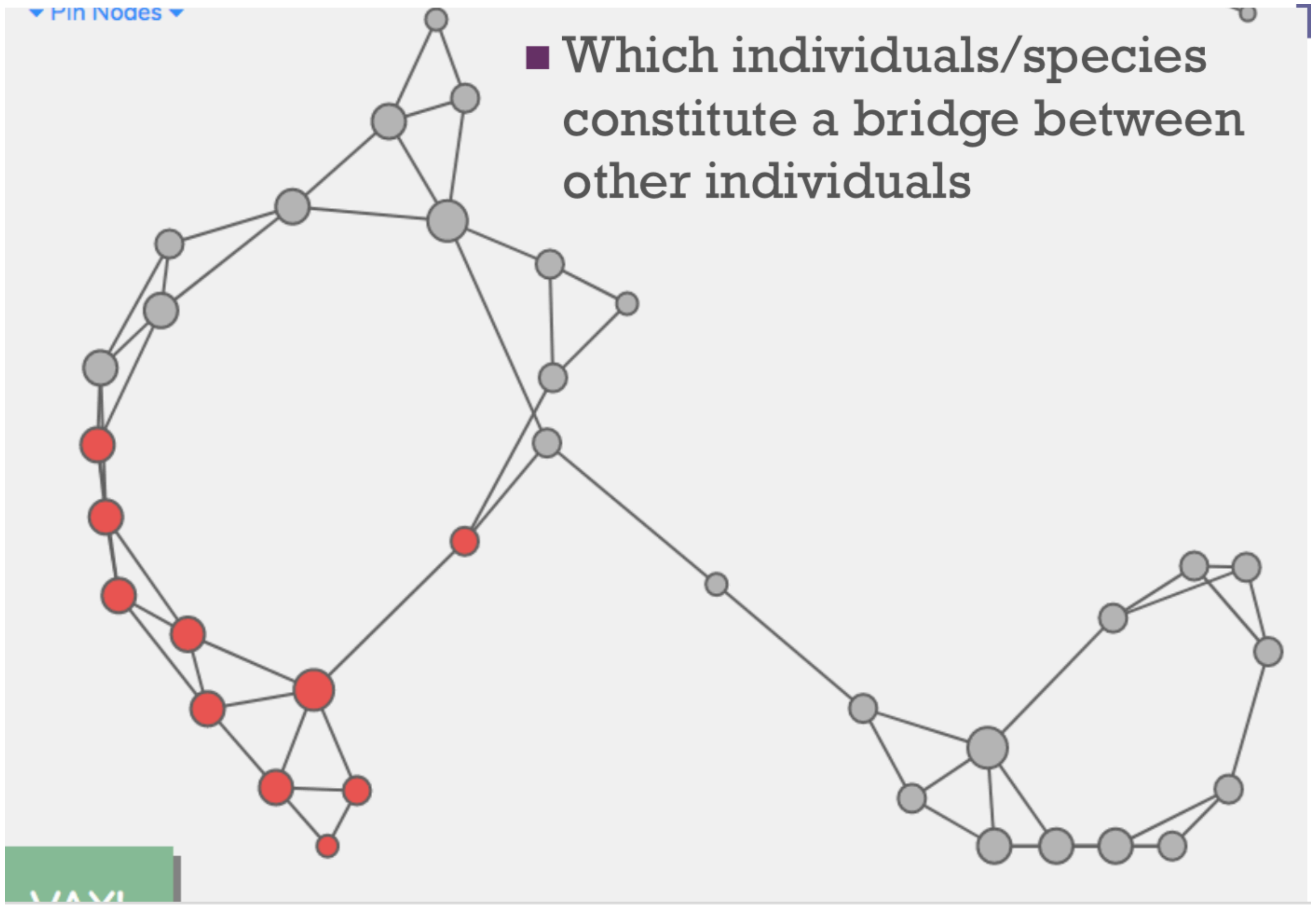
- Most tests are imperfect
- Result in occupancy/prevalence estimates that are biased low
- Underestimation of pathogen transmission rates
- Flawed predictions regarding infection dynamics and



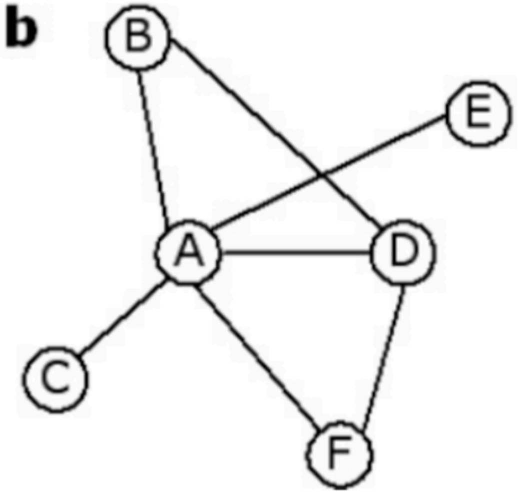
SINGLE SEASON OCCUPANCY

		Tested once	Occupancy
	n	Prevalence (95% CI)	
Galidia	29	0.48 (0.30-0.67)	0.83 (0.65-1.00)
Galidictis	12	0.92 (0.60-1.00)	1.00 (0.95-1.00)
Males	14	0.36 (0.14-0.64)	0.79 (0.54-1.00)
Females	27	0.48 (0.29-0.68)	0.94 (0.78-1.00)
Total	41	0.44 (0.29-0.60)	0.89 (0.72-1.00)

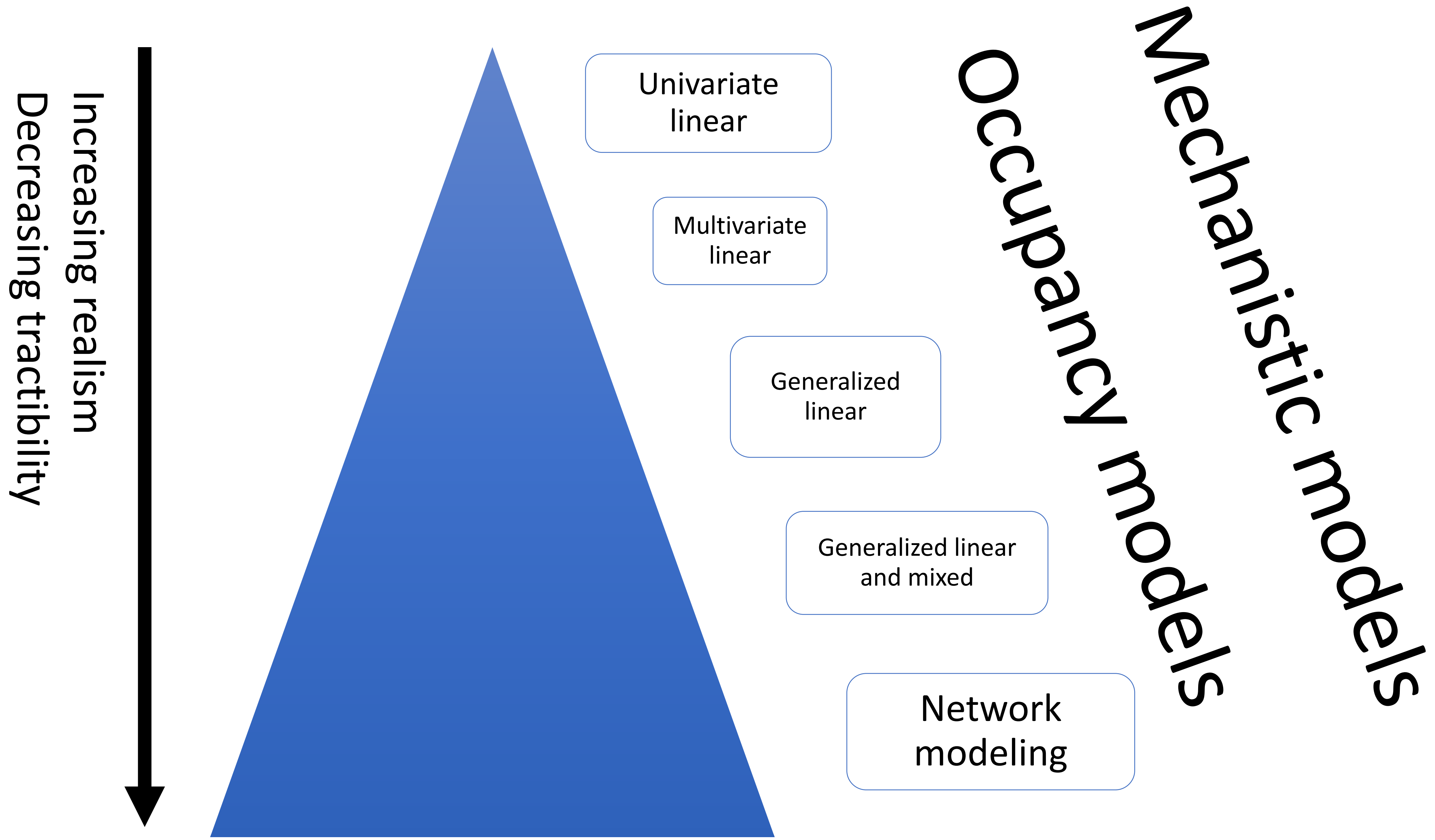
Network



- Which individuals/species constitute a bridge between other individuals

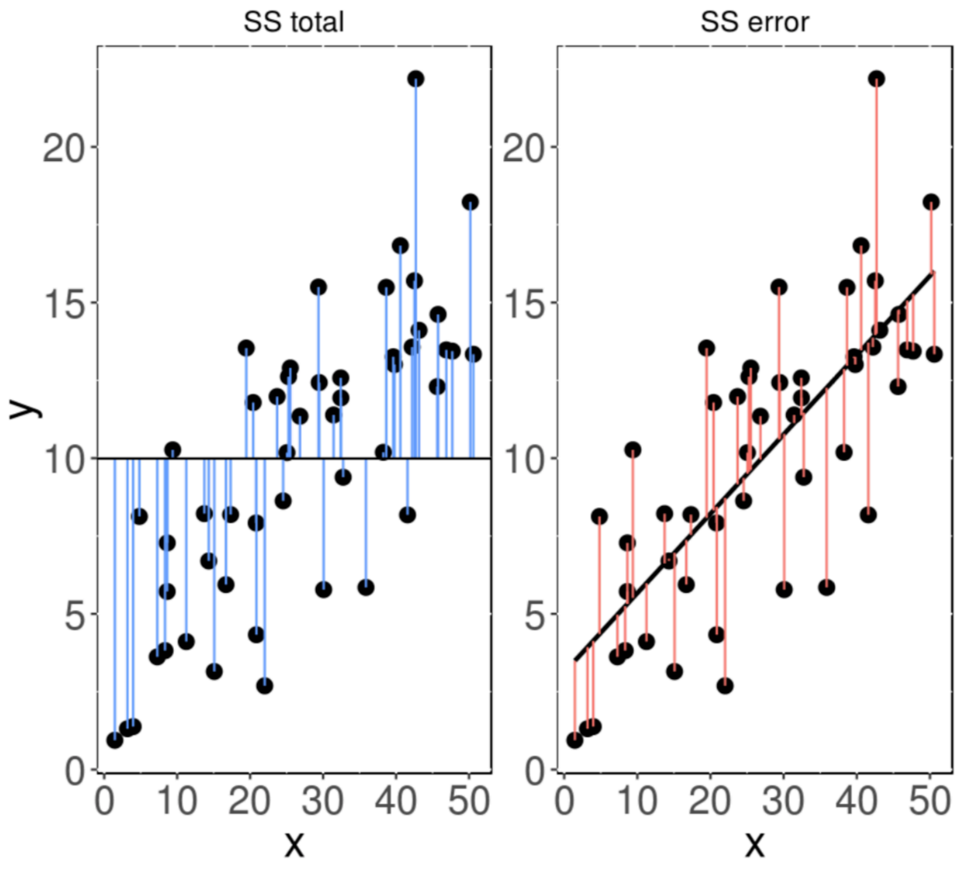


Individual	Degree
A	4
B	2
C	1
D	3
E	1
F	2


Increasing realism
Decreasing tractibility
Univariate linear
Multivariate linear
Generalized linear
Generalized linear and mixed
Network modeling
Occupancy models
Mechanistic models

Model fitting

Least square



Maximum likelihood

Mathematical statements of all detection histories are combined into model likelihood, such as:

L(ψ, p H 1 ,…, H 30) =Π 30 i=1 Pr(Hi)

Product of math equations forms the model likelihood for the observed data

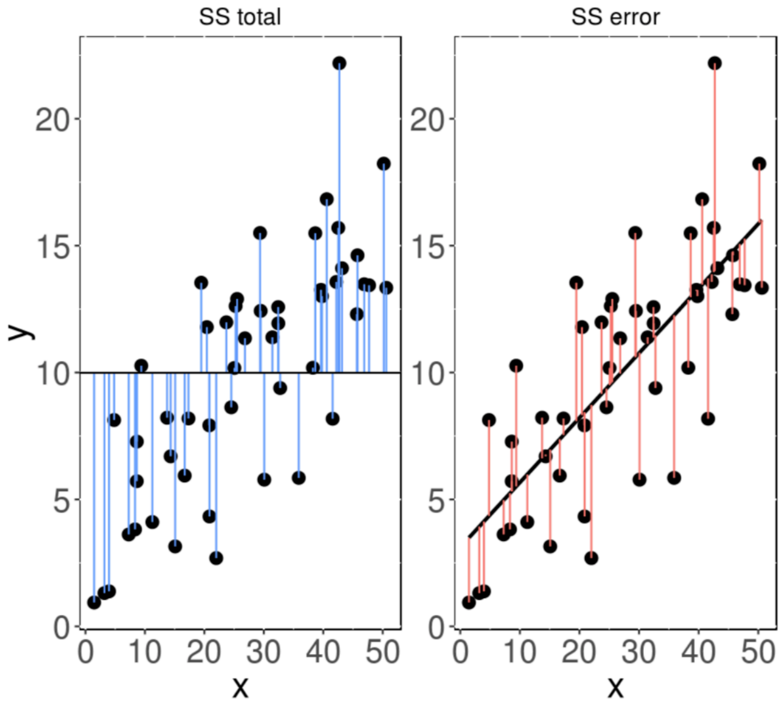
Using MLE determine Psi and p

Model fitting

Model evaluation and comparison

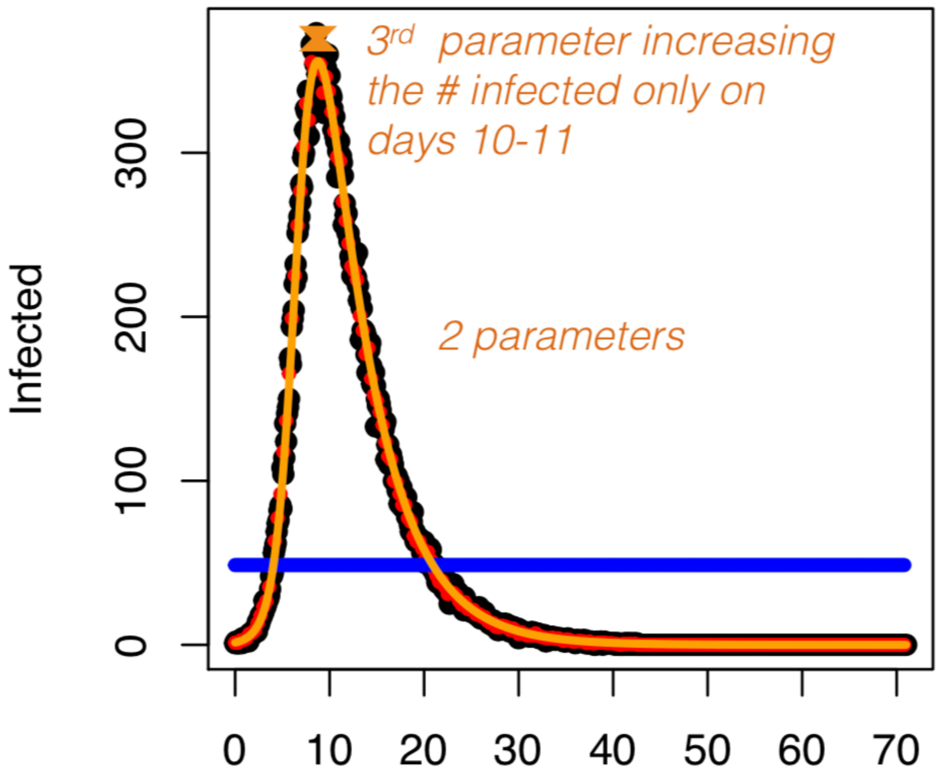
R^2

Definition r^2

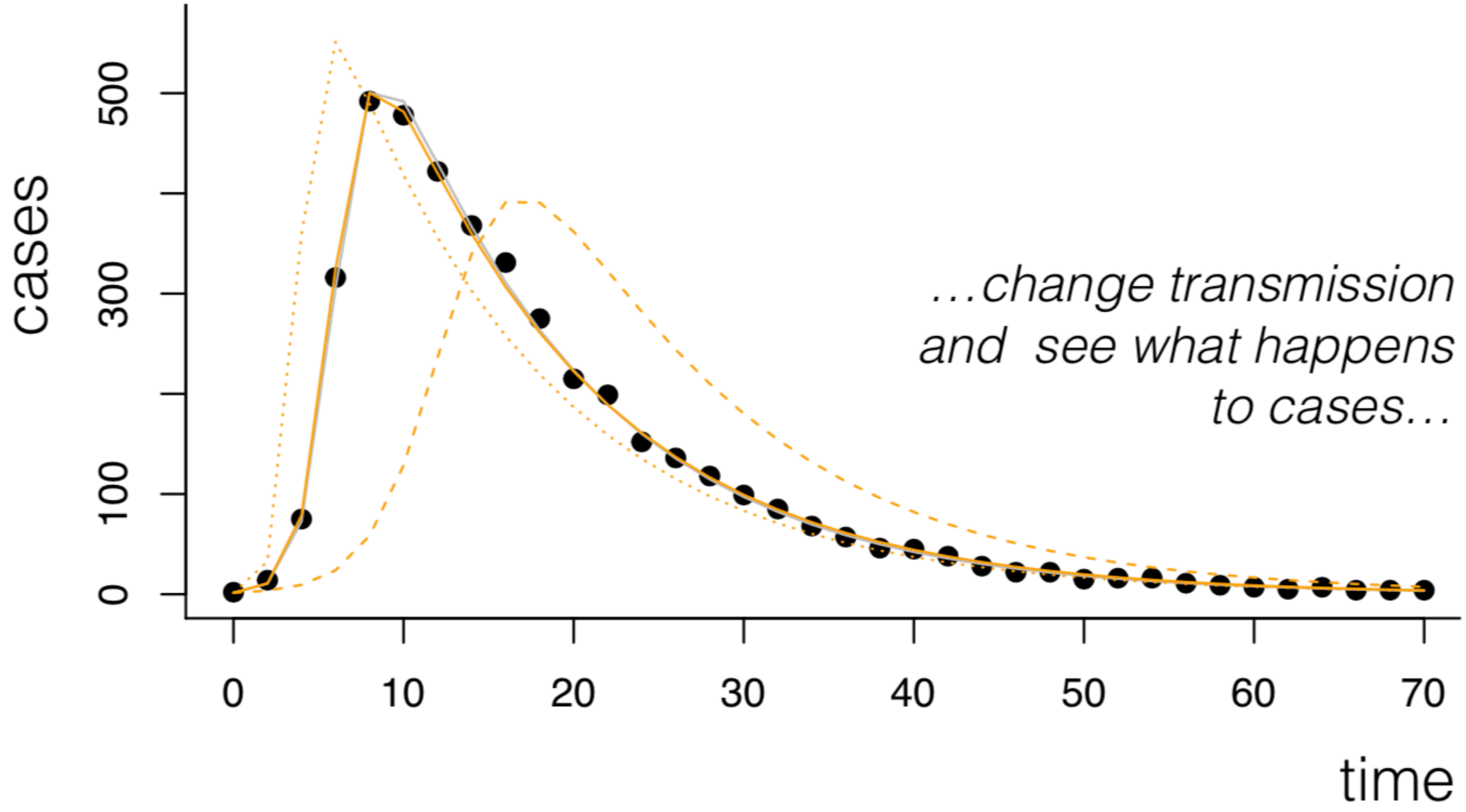


$$R^2 = 1 - \frac{SSE_p}{SST}$$

parsimony



Sensitivity analyses



Outline

- Science: research question
- Mechanistic model
- Statistical model
- R

R

Basics

```
# In R you can assign a value to a variable by using the "<-" operator.
# Run the following line of code by pressing "Control+Enter"

x <- 10

(1242-241.1)*32.21

# When you press control+Enter, the cursor automat
# in the console. This way, you can run scripts or
# pressing control+enter.

# The following code does some basic arithmetic.

x <- 2              # ask someone to guess a number
y <- x*9            # tell them to multiply it by 9
d1 <- floor(y/10)   # get the first digit  (floor(x
d2 <- y%%10         # get the second digit (a%%b gi
d1+d2-4             # tell them to add the digits a
```

```
require("Matrix")

# once you've included a package, you
# package. You can click the name of a
# help for that specific package. Here
# package

M <- Matrix(0, 2, 2, sparse=FALSE)
M[1,1] = 2; M[1,2] = 3
M[2,1] = 3; M[2,2] = 4;
M
det(M)
```

```
# Indeed we can make a new list by using the "c" (concatenate) command:

mylist <- c(1.1, 2.2, 3.3, 4.4, 5.5, 4.4, 3.3, 2.2, 1.1)
```

Intermediate

```
## ----Popdata, include=TRUE----------------------------------------
setwd("/Users/jessicametcalf/Downloads/E2M2-Dropbox-2018/E2M2.Metcalf.Lectures/Rcode/")
pop.data <- read.csv("WorldBankPop.csv")
head(pop.data[,1:10])


#####################     Arrange data   #####################
## For this we will need the package "dplyr"
## If you do have this packages yet, install it with the function
"install.packages(...)"

install.packages("dplyr")
install.packages("ggplot2")

## Then load it with require() or library()

require(dplyr)
require(ggplot2)
```

```
> a + 1
Error: object 'a' not found
>
```

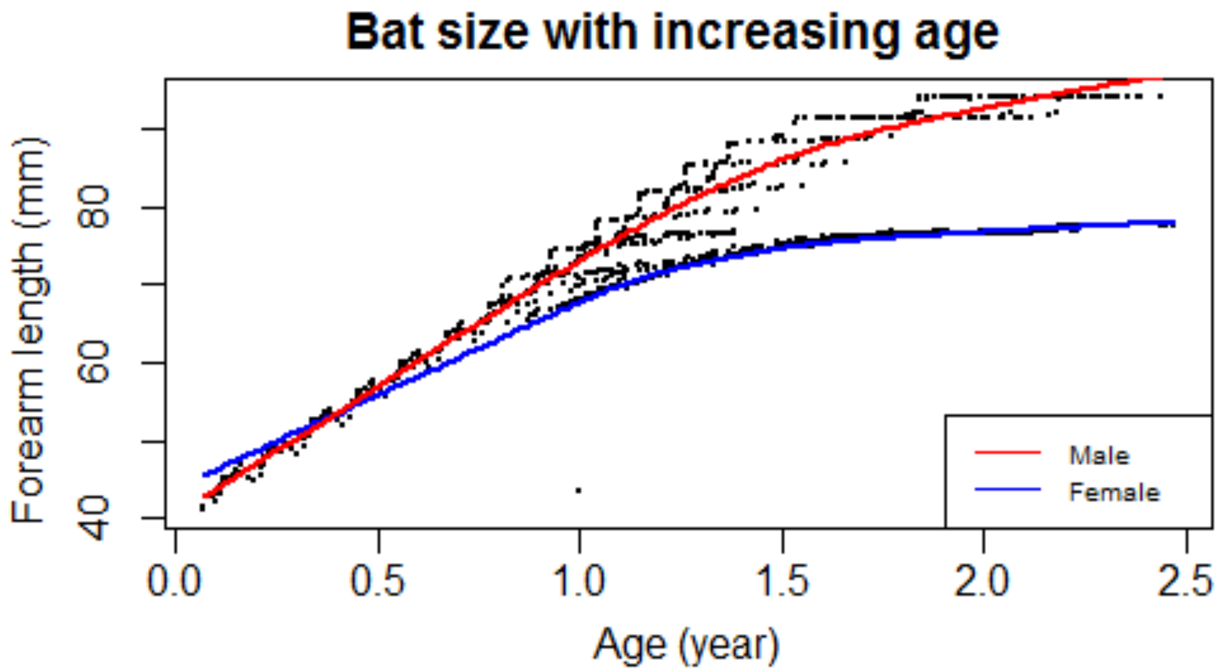
Intermediate

```
#-----------------------------------------------------
# 1. Data exploration: Plots and summary stadistics
#-----------------------------------------------------

# Distribution of outcome variable
hist(csb.data$outpatient, col='grey', main='',xlab='Number of outpatient visits per month')

# Outpatient visits for each health center (exploration of association for categorical variables)
boxplot(csb.data$outpatient~csb.data$csb, ylab='Number of outpatient visits per month')

# Outpatient visits after interventions were in place
boxplot(csb.data$outpatient~csb.data$int1, ylab='Number of outpatient visits per month')
boxplot(csb.data$outpatient~csb.data$int2, ylab='Number of outpatient visits per month')

# Correlation plots (exploration of association for quantitative variables)
plot(csb.data$staff,csb.data$outpatient, ylab='Number of outpatient visits per month')
abline(lm(outpatient~staff, data=csb.data))
plot(csb.data$ref,csb.data$outpatient, ylab='Number of outpatient visits per month')
abline(lm(outpatient~ref, data=csb.data))
```

```
# We'll check a multivariate model that includes number of medicall staff and
# (we don't consider a full model for simplicity and lack of sufficient observ
m1=glmer.nb(outpatient~ staff+ int1+ int2+ season+ (1 | csb), data=csb.data )
summary(m1)
```

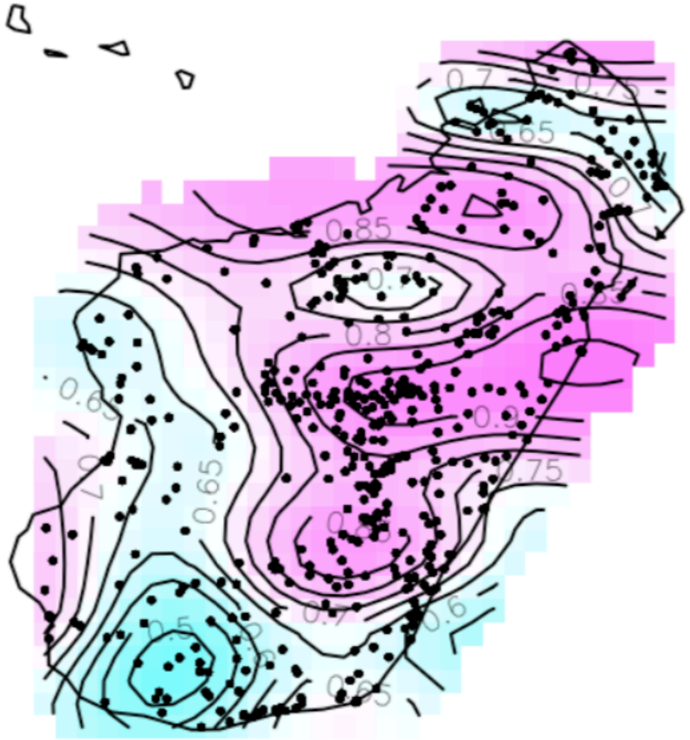
Intermediate

```
# The variable "Number of parasites" is count data and it's poisson dist
# This type of variable is typically modelled with poisson models
m6=glm(GIparasites~age+sexe+malaria, family='poisson', data=lemur.data)
summary(m6)

m7=step(m6)
summary(m7)
```

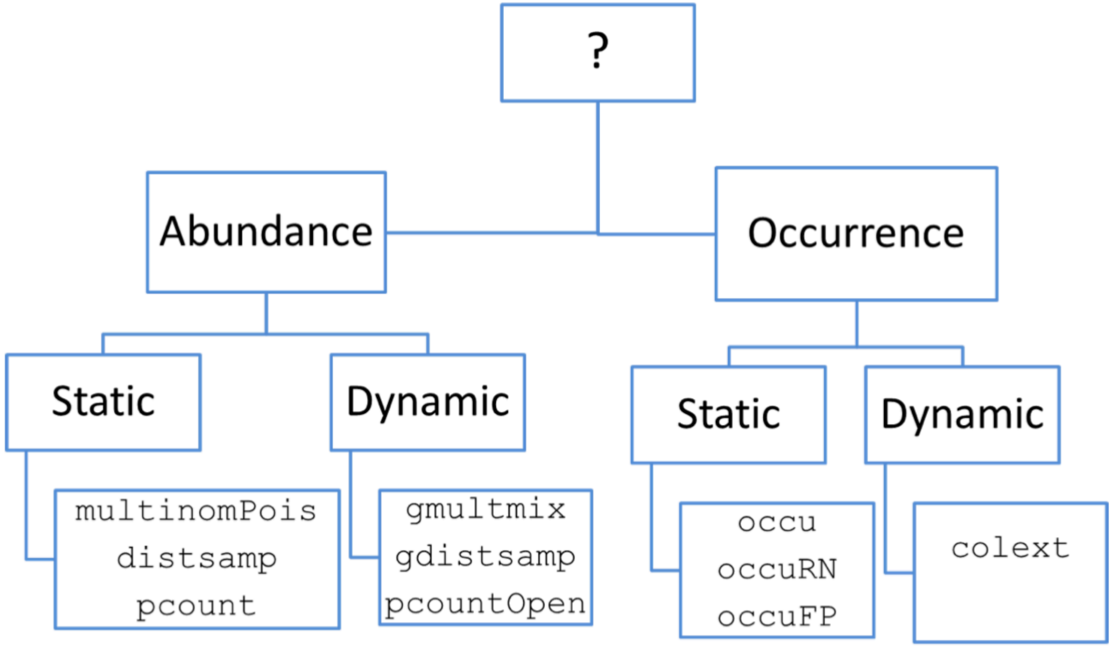


Advanced

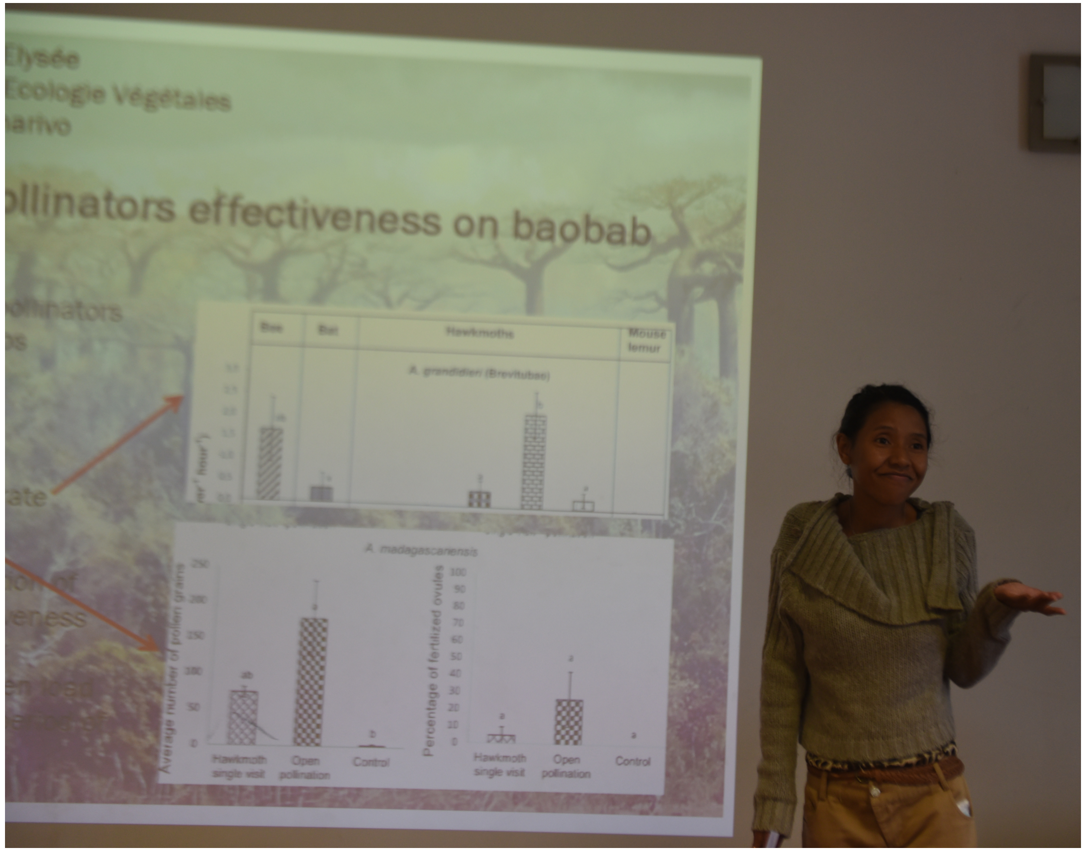
Add dsolve for LV model



WHICH MODEL SHOULD I USE (UNMARKED)?



Project presentation and completion



vi Manuscript writing and submission

- What are the main results that provide the answer to my question?
  - 1 to 3 graphs
  - 1 to 3 tables
- What is the journal that best fits my study?
  - Scope, audience, impact factor, math focus
- How do I present my manuscript?
  - Introduction: set the stage to your question
  - Methodology: describe explicitly all steps for replicability
  - Results: clear and concise
  - Discussion: explain how your study improves previous knowledge

You are now well equipped!

Any aminareo ny baolina!