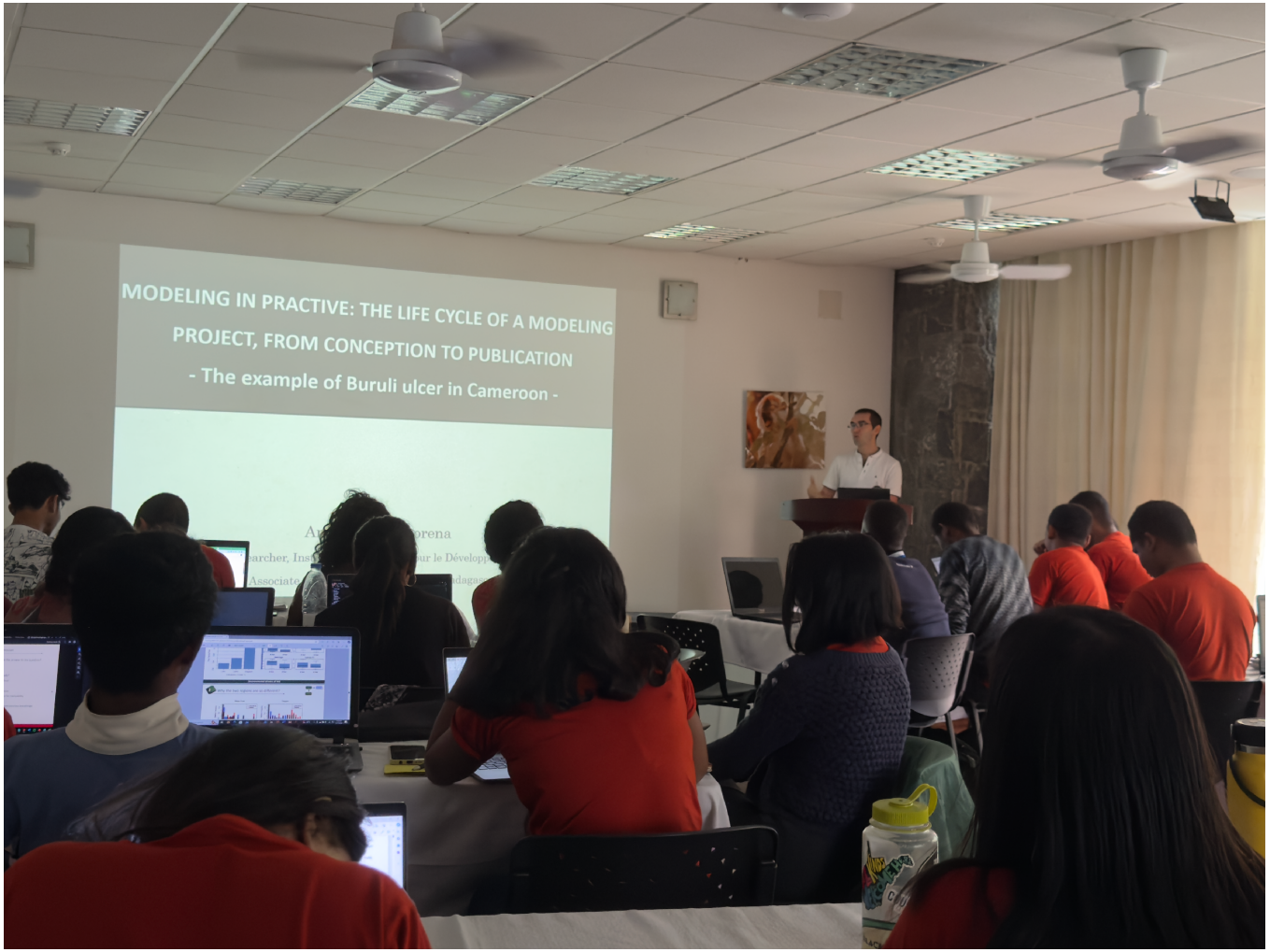
So many models, what have we learned?

Tanjona Ramiadantsoa
E2M2 2024


MODELING IN PRACTIVE: THE LIFE CYCLE OF A MODELING
PROJECT, FROM CONCEPTION TO PUBLICATION
- The example of Buruli ulcer in Cameroon -

Outline

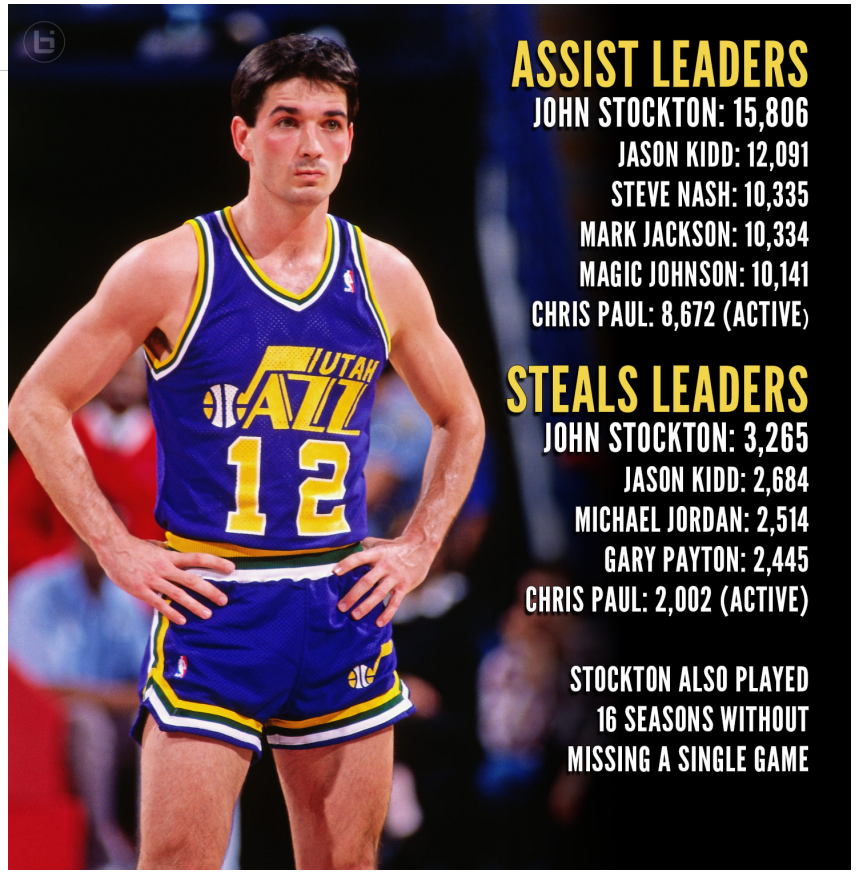
- The intangibles
- Advanced concepts
- Mechanistic modeling
- Statistical modeling
- Beyond this workshop

Outline

- **The intangibles**
- Advanced concepts
- Mechanistic modeling
- Statistical modeling
- Beyond this workshop

The intangibles

John Stockton



Reading a Scientific paper

Pando *et al. BMC Public Health* (2023) 23:1511
https://doi.org/10.1186/s12889-023-16425-w

BMC Public Health

RESEARCH ARTICLE Open Access

A social network analysis model approach to understand tuberculosis transmission in remote rural Madagascar

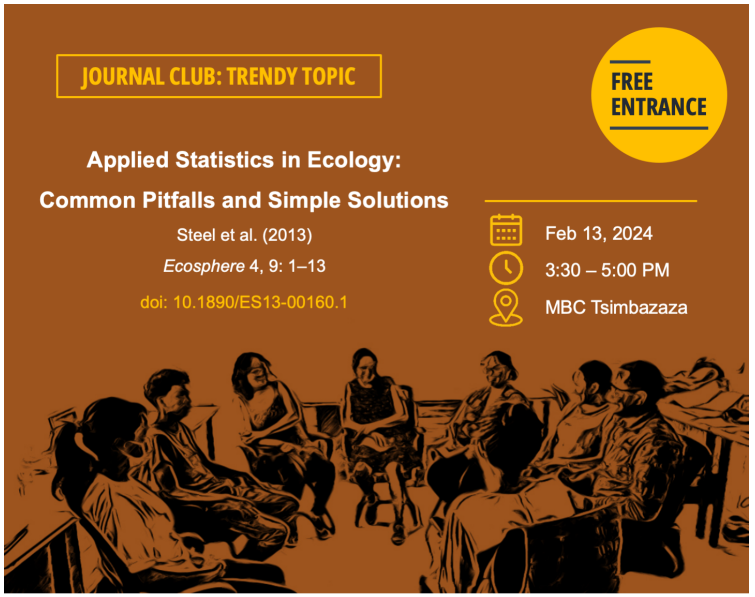
Christine Pando[1], Ashley Hazel[2], Lai Yu Tsang[1], Kimmerling Razafindrina[3], Andry Andriamiadanarivo[3], Roger Mario Rabetombosoa[3,4], Ideal Ambinintsoa[3], Gouri Sadananda[5], Peter M. Small[1], Astrid M. Knoblauch[4,6,7], Niaina Rakotosamimanana[4] and Simon Grandjean Lapierre[4,8,9*]

Abstract

Background Quality surveillance data used to build tuberculosis (TB) transmission models are frequently unavailable and may overlook community intrinsic dynamics that impact TB transmission. Social network analysis (SNA) generates data on hyperlocal social-demographic structures that contribute to disease transmission.

Methods We collected social contact data in five villages and built SNA-informed village-specific stochastic TB transmission models in remote Madagascar. A name-generator approach was used to elicit individual contact networks. Recruitment included confirmed TB patients, followed by snowball sampling of named contacts. Egocentric network data were aggregated into village-level networks. Network- and individual-level characteristics determining contact formation and structure were identified by fitting an exponential random graph model (ERGM), which formed the basis of the contact structure and model dynamics. Models were calibrated and used to evaluate WHO-recommended interventions and community resiliency to foreign TB introduction.

Journal club at MBC



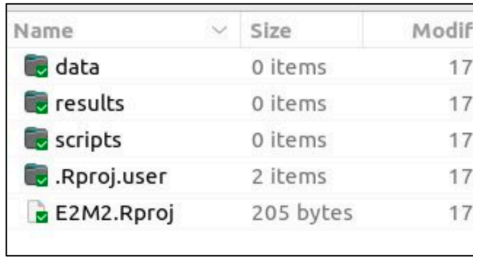
CALIFORNIA ACADEMY OF SCIENCES

MADAGASCAR BIODIVERSITY CENTER

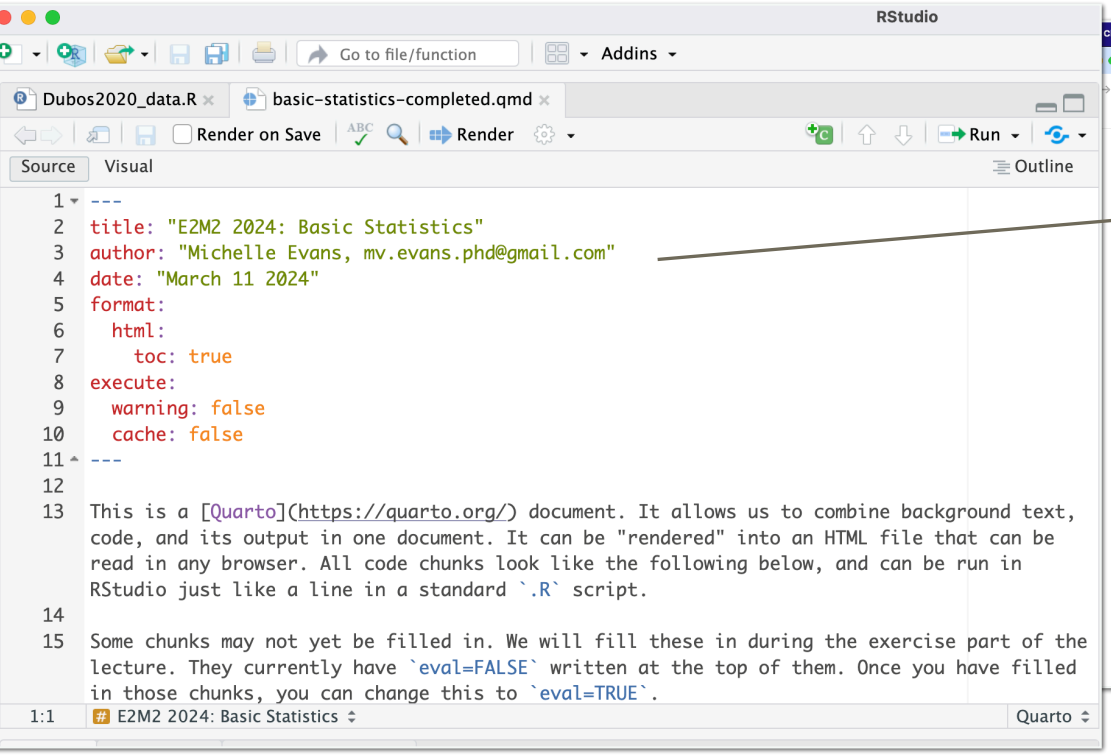
Examples on how to organize a project

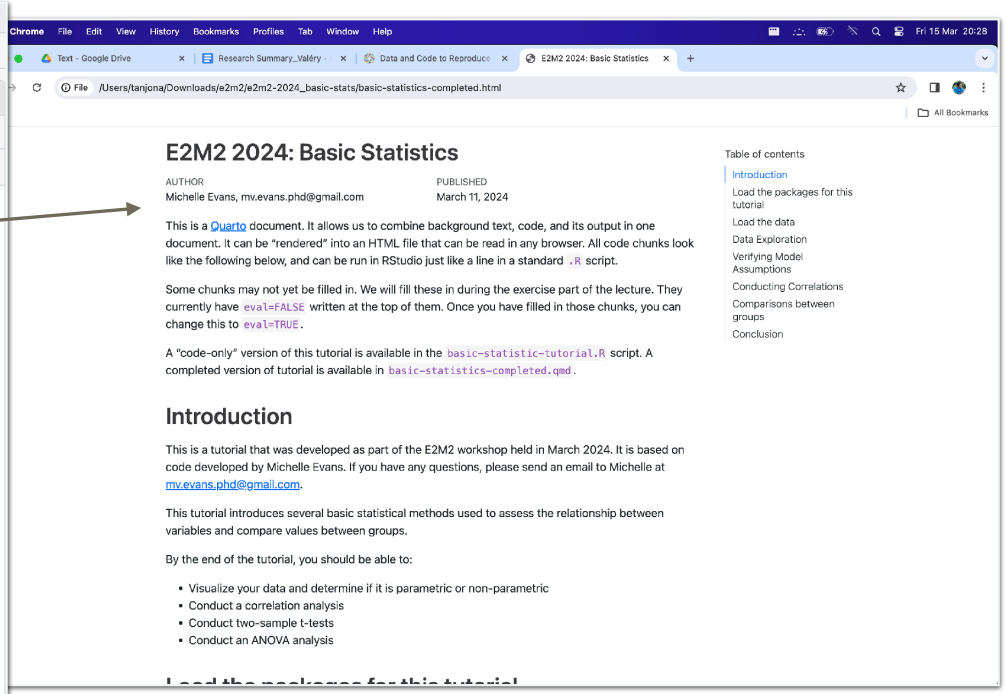
Create a folder structure

1. **Scripts:** all your .R files go here
   Tous les fichiers .R sont ici
2. **Data:** All of your data goes here. It is best to make two subdirectories: 'raw' and 'clean'
   Les données sont ici. Le meilleur practique est de créer deux sous-dossiers: `brut` et `nettoyé`
3. **Results:** Results of your analysis will go here. This includes tables of summary statistics, figures, and results of statistical tests
   Les résultats des analyses sont ici. Cela inclut les tableaux des statistiques sommaires, les figures, et les résultats des analyses

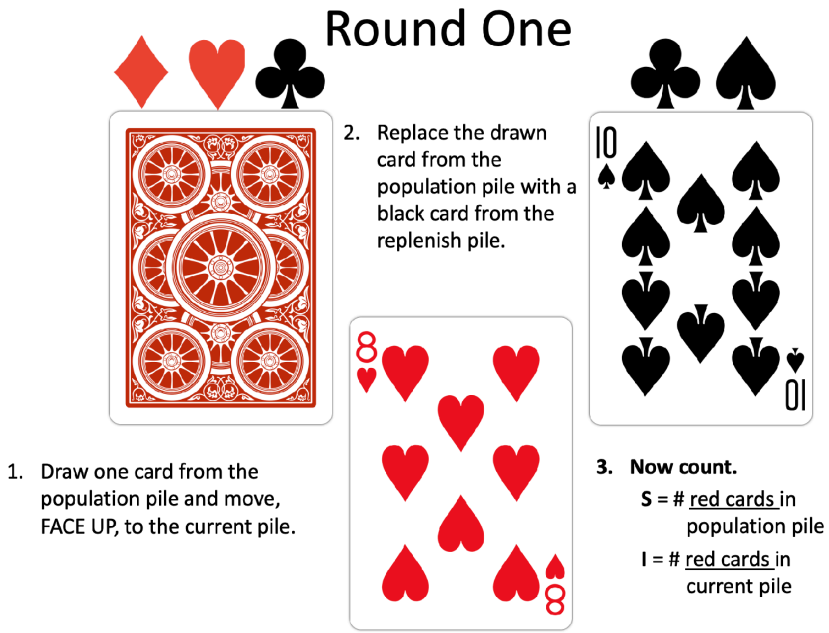


Combine texts and R codes and then generate a HTML





Using games to explain a concept



The Intangibles

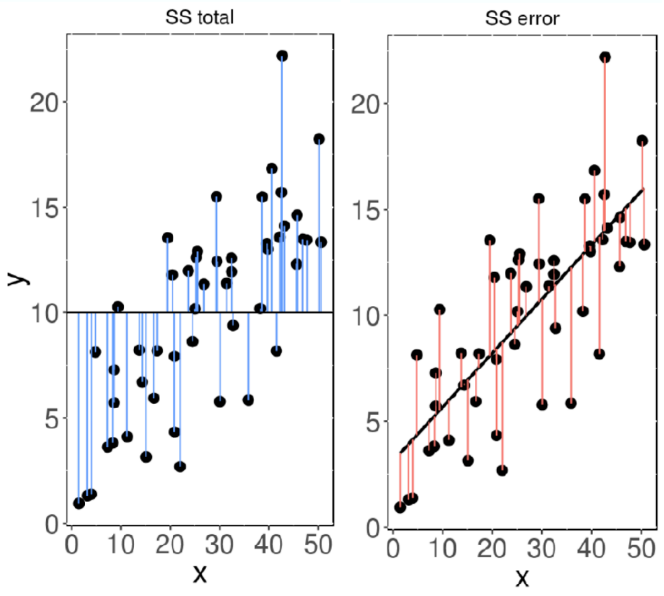
- Diverse research topics and meet new friends
- Improved your english: listening, speaking, and reading
- Extensive tutorial, R code, debugging, and good coding practice
- Asking a scientific question

Outline

- The intangibles
- **Advanced concepts**
- Mechanistic modeling
- Statistical modeling
- Beyond this course

Understanding model fit with R-squared

Definition r^2



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

❶ Our model, which is:
$y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon$

❹ Precision of the estimates

❺ p-values

❷ These are the differences between predictions from the model and the real y values

❸ Regression coefficients: β_0, β_1, and β_2

❻ Average distance of the model predictions from the real y values

```
Call:
lm(formula = y ~ x + z, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-2.79127 -0.71294  0.05412  0.73468  2.53241

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.2265     0.1577   1.437   0.1541
x             0.2797     0.1165   2.402   0.0182 *
z             0.1596     0.2097   0.761   0.4484
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.04 on 97 degrees of freedom
Multiple R-squared:  0.06047,   Adjusted R-squared:  0.0411
F-statistic: 3.122 on 2 and 97 DF,  p-value: 0.04855
```

❼ the proportion of variance in y explained by the predictors x and z (in this case, 6% of the variance in y can be explained by x and z)

❽ the proportion of variance in y explained by the predictors x and z beyond what we get if we added a random variable to the model

❾ tests if any of the predictors is related to y (here $p < 0.05$ means that at least 1 predictor is related to y)

What are some measures of model fit that you could use?

R-squared (R-carré)

Least squares (Moindres carrés)

Maximum likelihood (Maximum de vraisemblance)

(manakaiky indrindra ny tena izy)

Likelihood Vraisemblance

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$l(\theta|x) = \log L(\theta|x)$$

$$\text{AIC} = N * ln(\frac{SS_e}{N}) + 2K$$

$\boldsymbol{N}$: *Number of observations*
$\boldsymbol{SS_e}$: *Sum square of errors*
$\boldsymbol{K}$: *Number of parameters*

AIC

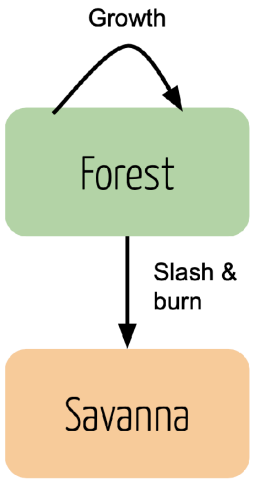
(uses least squares or log-likelihood but penalizes by number of fitted parameters)

$$AIC = -2\ln(L) + 2k$$

Constructing discrete and continuous time models

1. Construct a model

Développement d'un modèle



Régénération forestière

Forest regrowth

Perte du forêt due à tavy

Forest lost to S&B

$$\frac{dF}{dt} = rF\frac{K-N}{K} - \gamma FS$$

$$\frac{dS}{dt} = \gamma FS\frac{K-N}{K}$$

Savanna gained by S&B

Augmentation de la savane due à tavy

Solve a differential equation in R

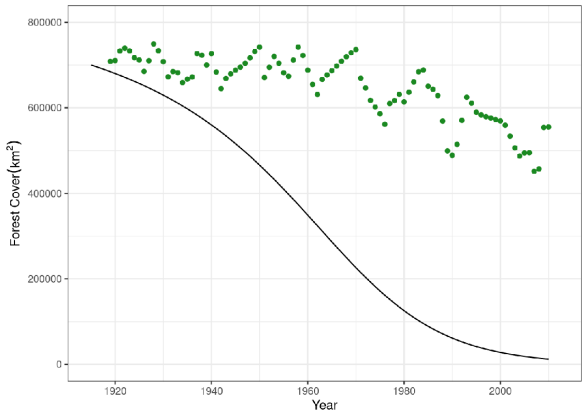
We define the model as a function that can be read by the ODE solver:

```
ForestSavannaCont <- function(t,y,parms){
  # The with() function gives access to the named values of parms within the
  # local environment created by the function
  with(c(as.list(y),parms),{
    N = For + Sav
    dFordt <- r*((K-N)/K)*For - gamma*For*Sav
    dSavdt <- gamma*((K-N)/K)*For*Sav

    #Return forest to compare with data
    list(c(dFordt, dSavdt))
  })
}
```

```
## Intialize population
Mada.start <- c(For = 450000, #sq. km of forest in Mada
                Sav = 100000)  #sq. km of savanna in Mada

## Set up time-steps and units - here go for slightly longer to allow model to equilibra
times <- seq(1900,2010,by=1)

## Define parameters
values <- c(r=1.01,
            gamma = .00000045,
            K = 900000)
```

We can then run the model and plot it with the data:

```
slash.mod1<- data.frame(lsoda(y = Mada.start, times = times, func = ForestSavannaCont,
                              parms = values))
names(slash.mod1) = c("yr", "forest", "savanna")

base.plot +
  geom_line(data = slash.mod1, aes(x = yr, y = forest)) +
  xlim(c(1915, 2012))
```



It looks like the model is overestimating the forest conversion rate. Let's adjust the rate by fitting the model to our dataset

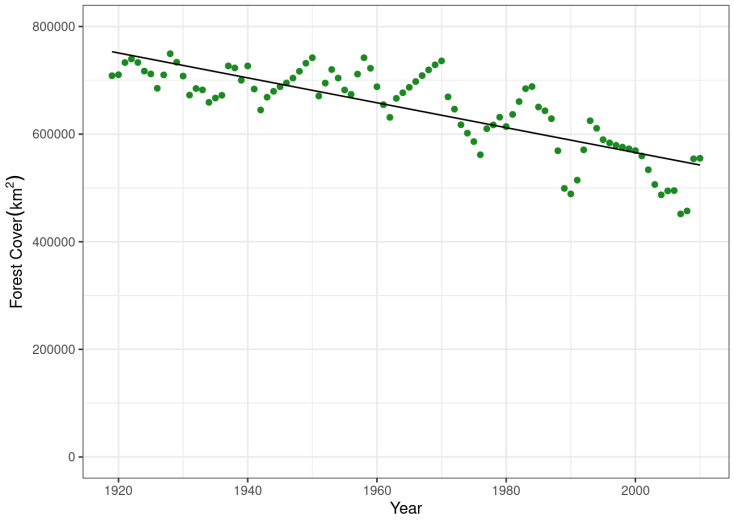
Optimization

Automatic optimization

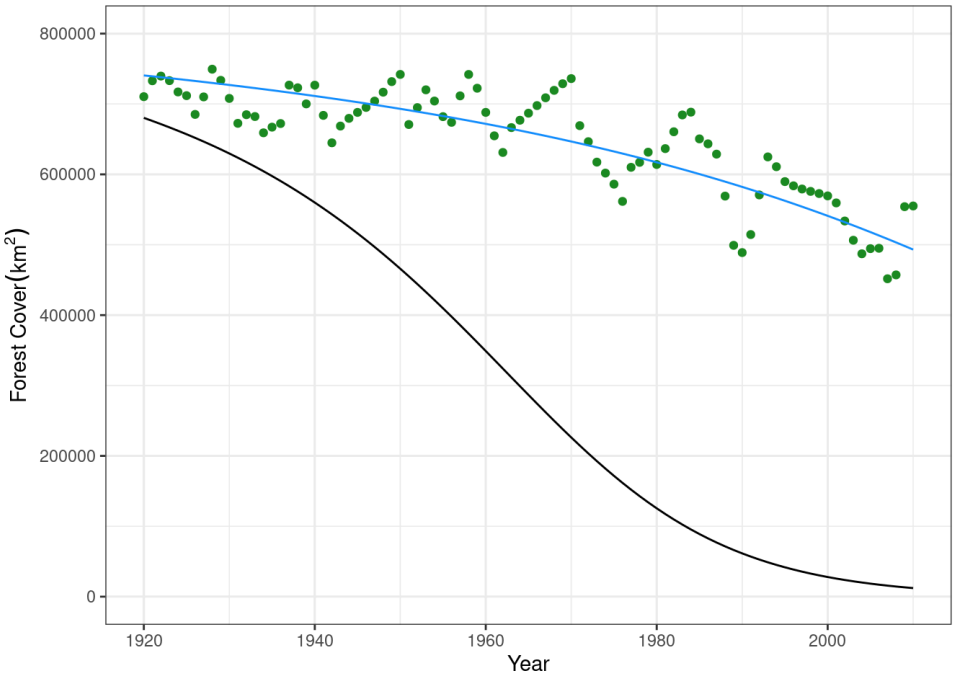
The second function is a wrapper that works to minimize the output (i.e. the sum of squares) from the previous function. By minimizing the output, we will find the optimum fit. This is done using the `optim` function in R, which searches the parameter space to find the values that result in the minimum sum of squares.

```
wrap_fit = function(guess.slope, guess.int, xguess, ydata){

  par <- c("m" = guess.slope, "b" = guess.int)


  out <- optim(par=par, fn = lst_sq, xval = xguess, ytrue = ydata)

  m.fit = out$par['m']

  b.fit = out$par['b']

  return(list(m.fit, b.fit))
}
```

```
#predict from this model
good.predictions <- mxb_line(m = good.fit[[1]], x = treedata$yr, b = good.fit[[2]])

#plot this data
base.plot +
  geom_line(aes(x = treedata$yr, y = good.predictions))
```



Optimization



Semi-manual optimization

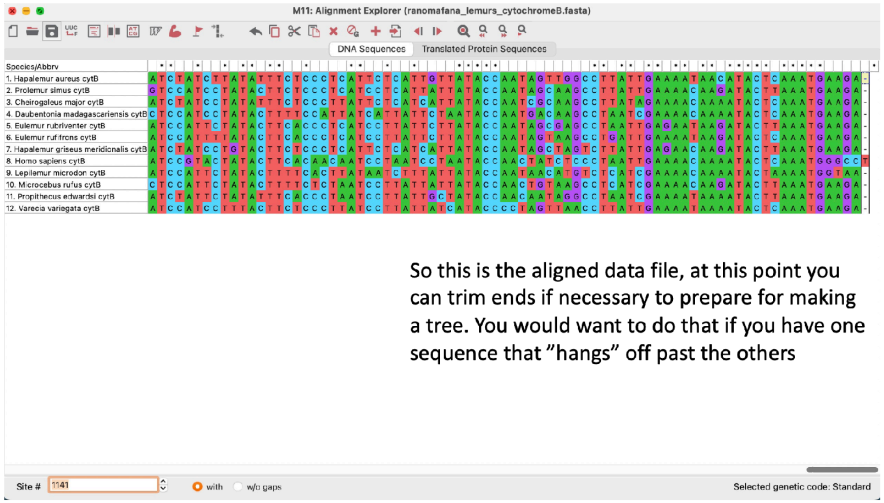
Now we write the wrapper function that minimizes the least squares to fit the optimum fit. It manually explores a range of values for the slashing rate and then chooses the one which results in the lowest sum of squares.

```
wrap_mech <- function(xstart, times, r, K, data){

  ## make a sequence of guesses for the rate
  slash.list = seq(from = 0, to = .000001, by = .0000001)

  ## then, search for the minimum
  sum.sq.lst = list()
  for (i in 1:length(slash.list)){
    sum.sq.lst[[i]] = sum_sq_mech(gamma=slash.list[i], xstart=xstart, r=r, K=K, times=ti
  }
  sum.sq.lst = c(unlist(sum.sq.lst))

  plot(slash.list, sum.sq.lst, type="b")

  ##find the minimum
  fit.slash = slash.list[sum.sq.lst==min(sum.sq.lst)]

  return(fit.slash)
}
```

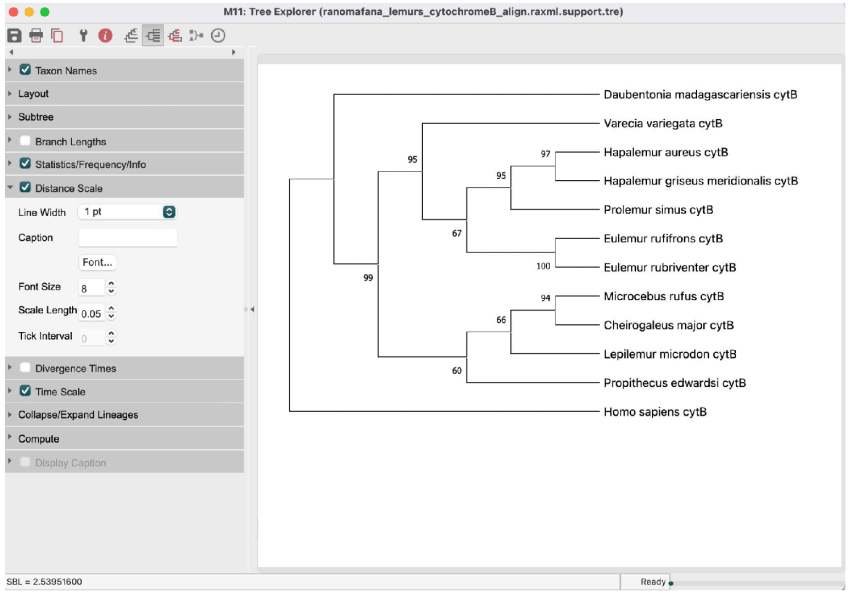
Then we can fit the model.

```
h <- wrap_mech(xstart = Mada.start, times = times, r = 1.01, K = 900000, data = treedata)
```

Construct a Phylogenetic Tree from DNA sequence



So this is the aligned data file, at this point you can trim ends if necessary to prepare for making a tree. You would want to do that if you have one sequence that "hangs" off past the others



We want a file of the tree saved that can be read in R, s newick. Click export trees…choose format newick, and customize in R

You can customize in MEGA too…it's just more limited

Outline

- The intangibles
- Advanced concepts
- **Mechanistic modeling**
- Statistical modeling
- Beyond this workshop

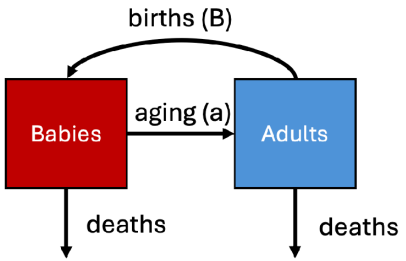
Mechanistic model: thinking about causation

Basic population models

The basic population model

Compartmental models (mechanistic models)

1. Populations are divided into compartments
2. Individuals within a compartment are homogenously mixed
3. Compartments and transition rates are determined by biological systems
4. *Rates of transferring between compartments are expressed mathematically*



$$n_{t+1} = An_t$$

A			n_t		n_{t+1}
$s_b(1-a)$	B	x	n_b	=	$s_b(1-a)n_b+bn_a$
S_ba	s_a		n_a		$S_ban_b+s_an_a$

Population growth will depend on population structure!

La croissance démographique dépendra de la structure de la population

Two species model

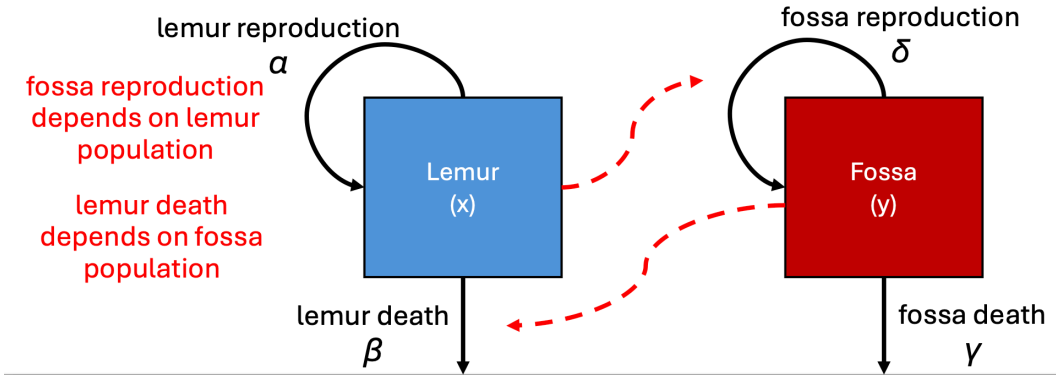
The predator-prey model

Compartmental models (mechanistic models)

1. Populations are divided into compartments
2. Individuals within a compartment are homogenously mixed
3. Compartments and transition rates are determined by biological systems
4. *Rates of transferring between compartments are expressed mathematically*

Some Assumptions:

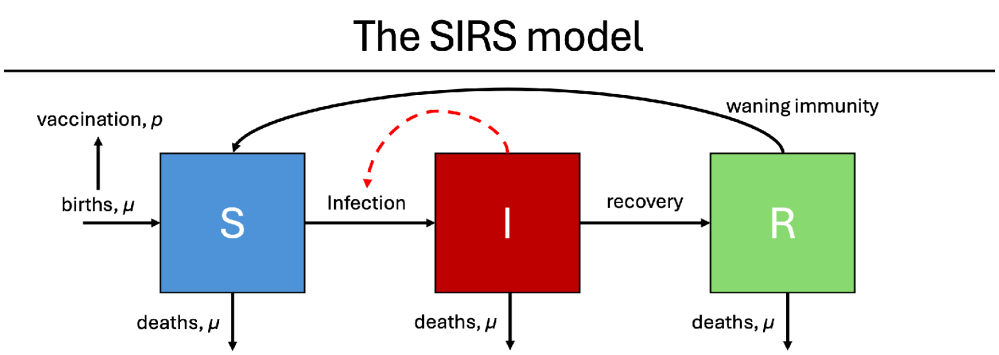
- The lemur has unlimited food supply
- The lemur only dies from being eaten by a fossa
- The fossa is totally dependent on a single prey species as its only food supply



$$\frac{dx}{dt} = x(\alpha - \beta y)$$

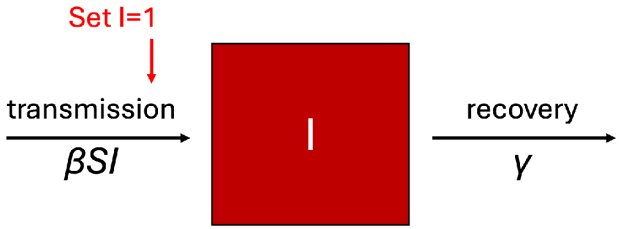
$$\frac{dy}{dt} = y(\delta x - \gamma)$$

Epidemiological model and R0



What do we change if immunity wanes? Recovered individuals become susceptible

Que change-t-on si l'immunité diminue? Les individus récupérés deviennent sensibles



$$R_0 = \beta N/\gamma$$

The average number of persons infected by an infectious individual when everyone is susceptible (S=100%, or S=1, start of an epidemic)

Le nombre moyen de personnes infectés par un individu infectueux quand tout le monde est sensible (S=100%, au début d'une épidémie)

Discrete vs. Continuous time

- Biological assumptions
  - Bacteria multiplication vs. seasonal breeding
  - Daily number of fish captured vs. the height of Ikopa river during a rainy season
- Tradition: what does exist that is similar to what I do
- Underlying behavior, mathematical toolkits, ease of simulation
- Familiarity

Other "family" of models

- Spatial or non-spatial model
- With or without age/stage-structure
- Stochastic or deterministic model

Inclusion or exclusion of processes

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Responses of generalist and specialist species to fragmented landscapes

Tanjona Ramiadantsoa [a,b,*], Ilkka Hanski [a,1], Otso Ovaskainen [a,c]

[a] Faculty of Biological and Environmental Sciences, University of Helsinki, Finland
[b] Department of Ecology, Evolution, and Behaviour, University of Minnesota, Twin Cities, USA
[c] Centre for Biodiversity Dynamics, Department of Mathematical Sc

ARTICLE INFO

Article history:
Received 15 August 2017
Available online xxxx

Keywords:
Habitat loss
Habitat fragmentation
Heterogeneous habitat

ABST

Empirical
species n
contrasti
of mecha
joint effe
between
assumpti

Table 2
A summary of the 14 model variants included in this paper. The column "Hump-shaped pattern" describes whether the model variant predicts that the generalist species achieves its maximal prevalence at intermediate level of habitat availability.

Model variant	Competition	Stochasticity	Space	Aggregation	Hump-shaped pattern	Results shown in
1	No	No	No	NA	No	Fig. 2D
2	No	No	Yes	No	No	Fig. S2B
3	No	No	Yes	Yes	No	Fig. S2A
4	No	Yes	No	No	No	Fig. S2D
5	No	Yes	No	Yes	No	Fig. S2C
6	No	Yes	Yes	No	No	Fig. 3A
7	No	Yes	Yes	Yes	No	Fig. S2E
8	Yes	No	No	NA	No	Fig. 2EF
9	Yes	No	Yes	No	No	Fig. 4B
10	Yes	No	Yes	Yes	No	Fig. 4A
11	Yes	Yes	No	No	Yes	Fig. 4D
12	Yes	Yes	No	Yes	Yes	Fig. 4C
13	Yes	Yes	Yes	No	Yes	Fig. 3D
14	Yes	Yes	Yes	Yes	Yes	Fig. 3C

It is all about your question

Context and question

Initiative Covax : Madagascar reçoit un premier lot de 250 000 doses de vaccins

How to allocate to doses for each region to minimize death due to COVID19?

RESEARCH Open Access

Prioritizing COVID-19 vaccination efforts and dose allocation within Madagascar

Fidisoa Rasambainarivo[1,2*], Tanjona Ramiadantsoa[3,4,5], Antso Raherinandrasana[6,7], Santatra Randrianarisoa[2], Benjamin L. Rice[1,8], Michelle V. Evans[5], Benjamin Roche[5], Fidiniaina Mamy Randriatsarafara[7,9], Amy Wesolowski[10] and Jessica C. Metcalf[1,11]

Abstract

Model description

SARS-CoV-2 transmission model of Madagascar

We constructed an age-structured, stochastic SEAIR (susceptible, exposed, asymptomatic infection, symptomatic infection, and removed) transmission model by expanding previous work [16, 17] (see Supplementary figure S1). With this model, we simulated the trajectory of SARS-CoV-2 cases in each of the 22 regions of Madagascar under different assumptions about vaccination deployment among the regions (detailed below). For each

16. Evans MV, Garchitorena A, Rakotonanahary RJL, Drake JM, Andriamihaja B, Rajaonarifara E, et al. Reconciling model predictions with low reported cases of COVID-19 in Sub-Saharan Africa: insights from Madagascar. Glob Health Action. 2020;13:1816044.
17. Roche B, Garchitorena A, Roiz D. The impact of lockdown strategies targeting age groups on the burden of COVID-19 in France. Epidemics-neth. 2020;33:100424.

Scenarios on doses allocations

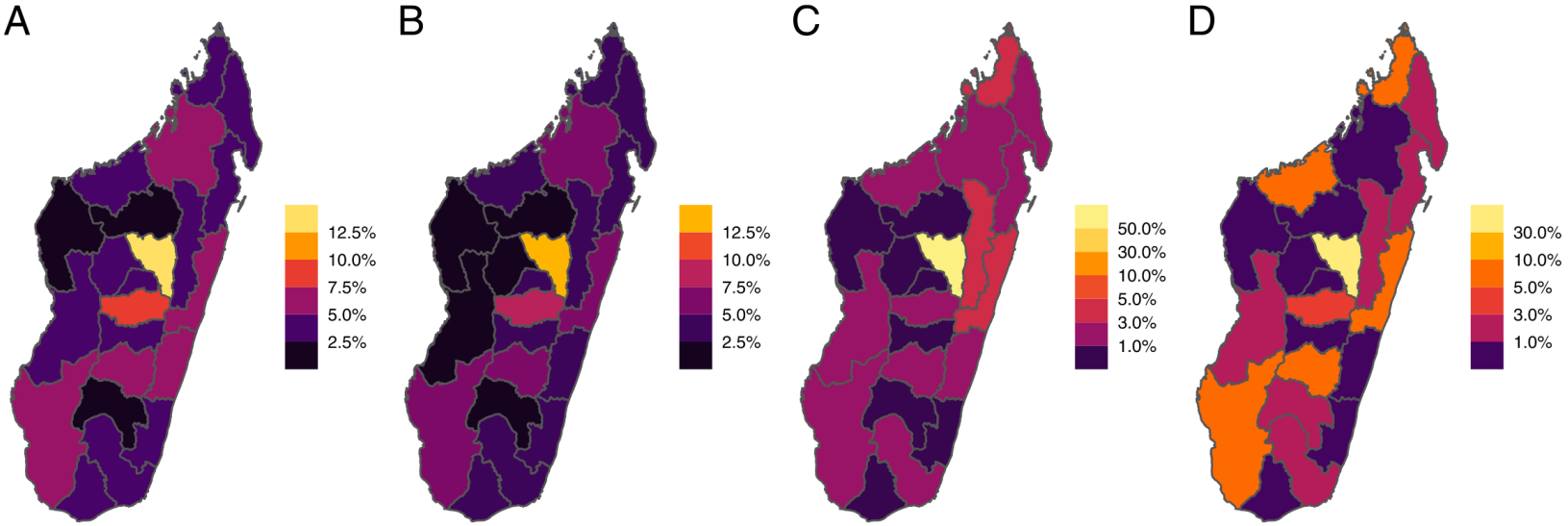
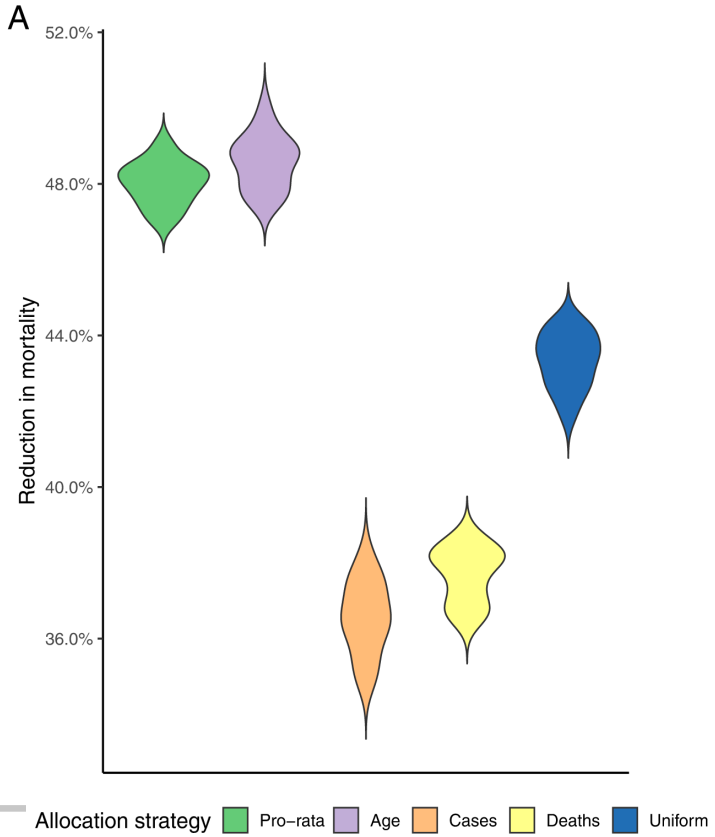


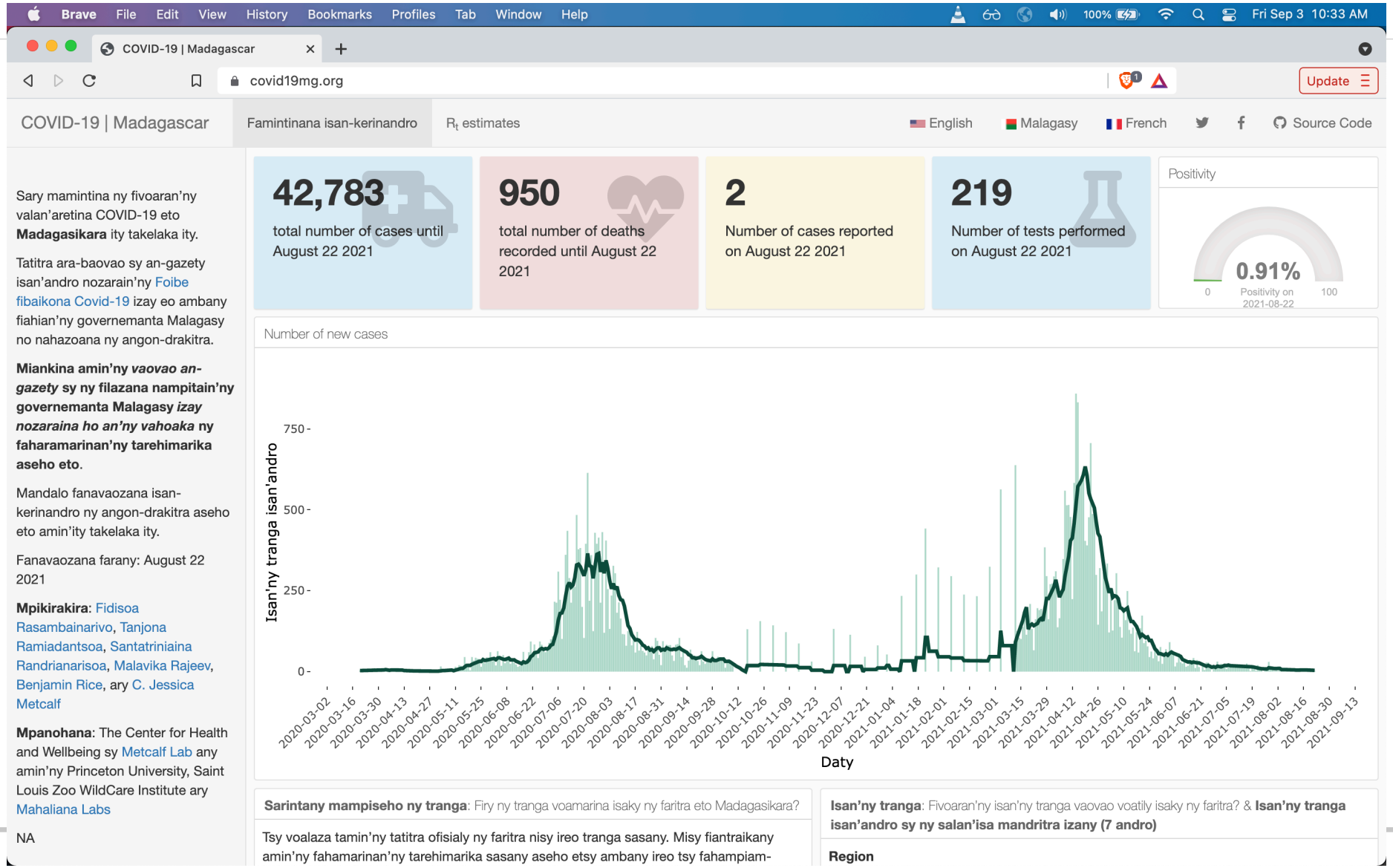
Fig. 2 The proportion of total doses distributed by region. Assuming that the total vaccine supply is 20% of the entire population, we explored various distribution strategies. The proportion of doses per region is shown based on each prioritization scheme: (A) doses are distributed to regions based on population size (pro-rata), (B) doses are allocated based on the distribution of people aged over 60 years between the regions (age), (C) doses are distributed to regions based on the number of cases reported (cases), (D) doses are distributed to regions based on the number of deaths reported (deaths)

Key results

How to allocate to doses for each region to minimize death due to COVID19?

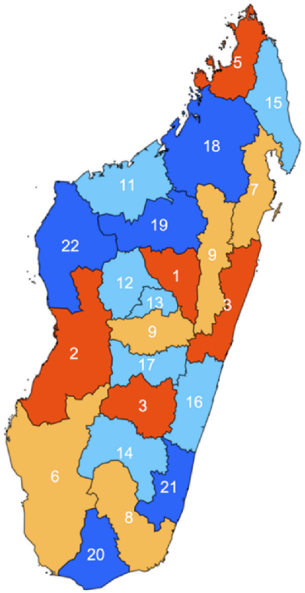


Most effective allocation is based on the number of people over 60 years old or population size

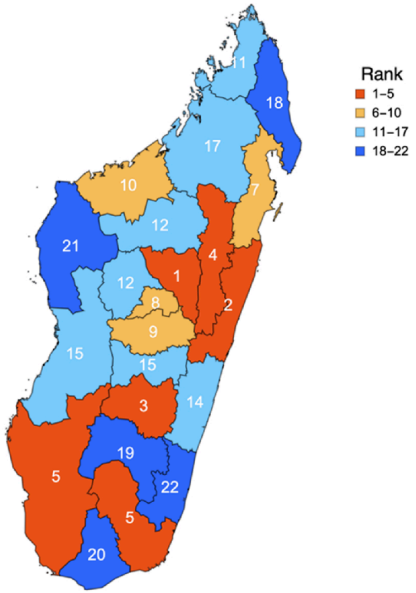

Brave File Edit View History Bookmarks Profiles Tab Window Help
100% Fri Sep 3 10:33 AM
COVID-19 | Madagascar
covid19mg.org
Update
COVID-19 | Madagascar
Famintinana isan-kerinandro
Rt estimates
English
Malagasy
French
Source Code
Sary mamintina ny fivoaran'ny valan'aretina COVID-19 eto Madagasikara ity takelaka ity.
Tatitra ara-baovao sy an-gazety isan'andro nozarain'ny Foibe fibaikona Covid-19 izay eo ambany fiahian'ny governemanta Malagasy no nahazoana ny angon-drakitra.
Miankina amin'ny vaovao an-gazety sy ny filazana nampitain'ny governemanta Malagasy izay nozaraina ho an'ny vahoaka ny faharamarinan'ny tarehimarika aseho eto.
Mandalo fanavaozana isan-kerinandro ny angon-drakitra aseho eto amin'ity takelaka ity.
Fanavaozana farany: August 22 2021
Mpikirakira: Fidisoa Rasambainarivo, Tanjona Ramiadantsoa, Santatriniaina Randrianarisoa, Malavika Rajeev, Benjamin Rice, ary C. Jessica Metcalf
Mpanohana: The Center for Health and Wellbeing sy Metcalf Lab any amin'ny Princeton University, Saint Louis Zoo WildCare Institute ary Mahaliana Labs
NA
42,783
total number of cases until August 22 2021
950
total number of deaths recorded until August 22 2021
2
Number of cases reported on August 22 2021
219
Number of tests performed on August 22 2021
Positivity
0.91%
0
Positivity on 2021-08-22
100
Number of new cases
Isan'ny tranga isan'andro
750
500
250
0
2020-03-02
2020-03-16
2020-03-30
2020-04-13
2020-04-27
2020-05-11
2020-05-25
2020-06-08
2020-06-22
2020-07-06
2020-07-20
2020-08-03
2020-08-17
2020-08-31
2020-09-14
2020-09-28
2020-10-12
2020-10-26
2020-11-09
2020-11-23
2020-12-07
2020-12-21
2021-01-04
2021-01-18
2021-02-01
2021-02-15
2021-03-01
2021-03-15
2021-03-29
2021-04-12
2021-04-26
2021-05-10
2021-05-24
2021-06-07
2021-06-21
2021-07-05
2021-07-19
2021-08-02
2021-08-16
2021-08-30
2021-09-13
Daty
Sarintany mampiseho ny tranga: Firy ny tranga voamarina isaky ny faritra eto Madagasikara?
Tsy voalaza tamin'ny tatitra ofisialy ny faritra nisy ireo tranga sasany. Misy fiantraikany amin'ny fahamarinan'ny tarehimarika sasany aseho etsy ambany ireo tsy fahampiam-
Isan'ny tranga: Fivoaran'ny isan'ny tranga vaovao voatily isaky ny faritra? & Isan'ny tranga isan'andro sy ny salan'isa mandritra izany (7 andro)
Region

Context and question

Rank for 1st case



Rank for 5th case



Can we predict the order of reported arrival of COVID19 using mobility data?

Existing human mobility data sources poorly predicted the spatial spread of SARS-CoV-2 in Madagascar

Tanjona Ramiadantsoa [a,b,c,*], C. Jessica E. Metcalf [d,e], Antso Hasina Raherinandrasana [f,g], Santatra Randrianarisoa [h], Benjamin L. Rice [d,i], Amy Wesolowski [j], Fidiniaina Mamy Randriatsarafara [g,k], Fidisoa Rasambainarivo [d,h]

[a] *Department of Life Science, University of Fianarantsoa, Madagascar*
[b] *Department of Mathematics, University of Fianarantsoa, Madagascar*
[c] *Department of Integrative Biology, University of Wisconsin-Madison, WI, USA*
[d] *Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA*
[e] *Princeton School of Public and International Affairs, Princeton University, NJ, USA*
[f] *Surveillance Unit, Ministry of Health of Madagascar, Madagascar*
[g] *Faculty of Medicine, University of Antananarivo, Madagascar*

Model description

$$S(i,t+1) = S(i,t) - \beta\, S(i,t)I(i,t)/N(i,t)$$

$$E(i,t+1) = E(i,t) + \beta\, S(i,t)I(i,t)/N(i,t) - \alpha E(i,t)$$

$$I(i,t+1) = I(i,t) + \alpha E(i,t) - \gamma I(i,t)$$

$$R(i,t+1) = R(i,t) + \gamma I(i,t)$$

In practice, we used a hierarchical approach to calculate the number of individuals moving across the regions. First, we fixed the total number of individuals moving per unit of time (X). Then, we used a vector $P = (P_1, \ldots, P_{22})$, where $P_k \geq 0$ and $\sum_{k=1}^{22} P_k = 1$, to calculate the number of individuals leaving each region $(X_1, \ldots, X_{22}) \sim$ Multinomial(X, P). Finally, for each region i, we used a vector $M_i = (M_1, \ldots, M_{22})$, where $M_{i.} \geq 0$ and $\sum_{j=1}^{22} M_{ij} = 1$ and $M_{ii} = 0$ as we ignored mobility within regions, to calculate the number of individuals leaving region i and entering region j where x_{ij} is the j^{th} entry from $x_{i.} \sim$ Multinomial(X_i, M_i). The technical details of obtaining the vectors P and M_i from the four mobility matrices are reported in the supplement.

Discrete time	Stochastic	Spatial explicit	No age-structure

Mobility scenarios

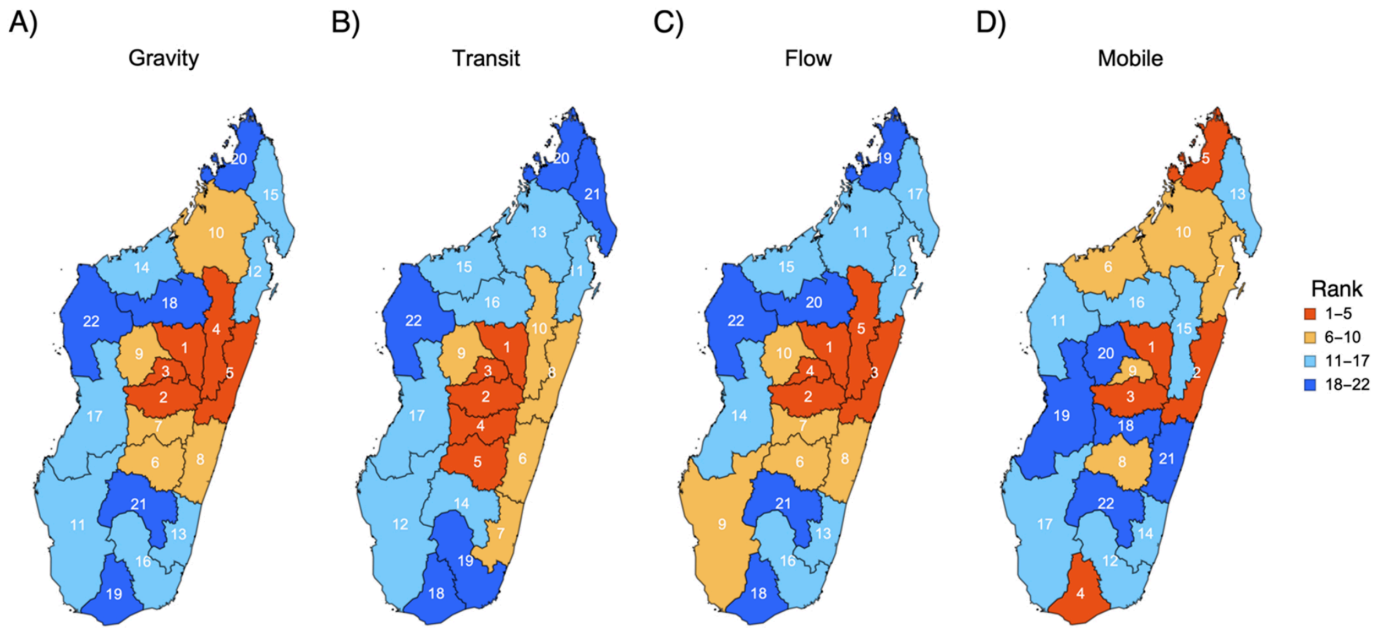


Fig. 3. A comparison of the four main mobility matrices used. The regions are ranked according to the number of individuals entering the region per day using the various mobility matrices: A) Euclidean, B) transit, C) Internal Migration Flow (Flow), and D) mobile phone data.

Key result

Can we predict the order of reported arrival of COVID19 using mobility data? NOOO

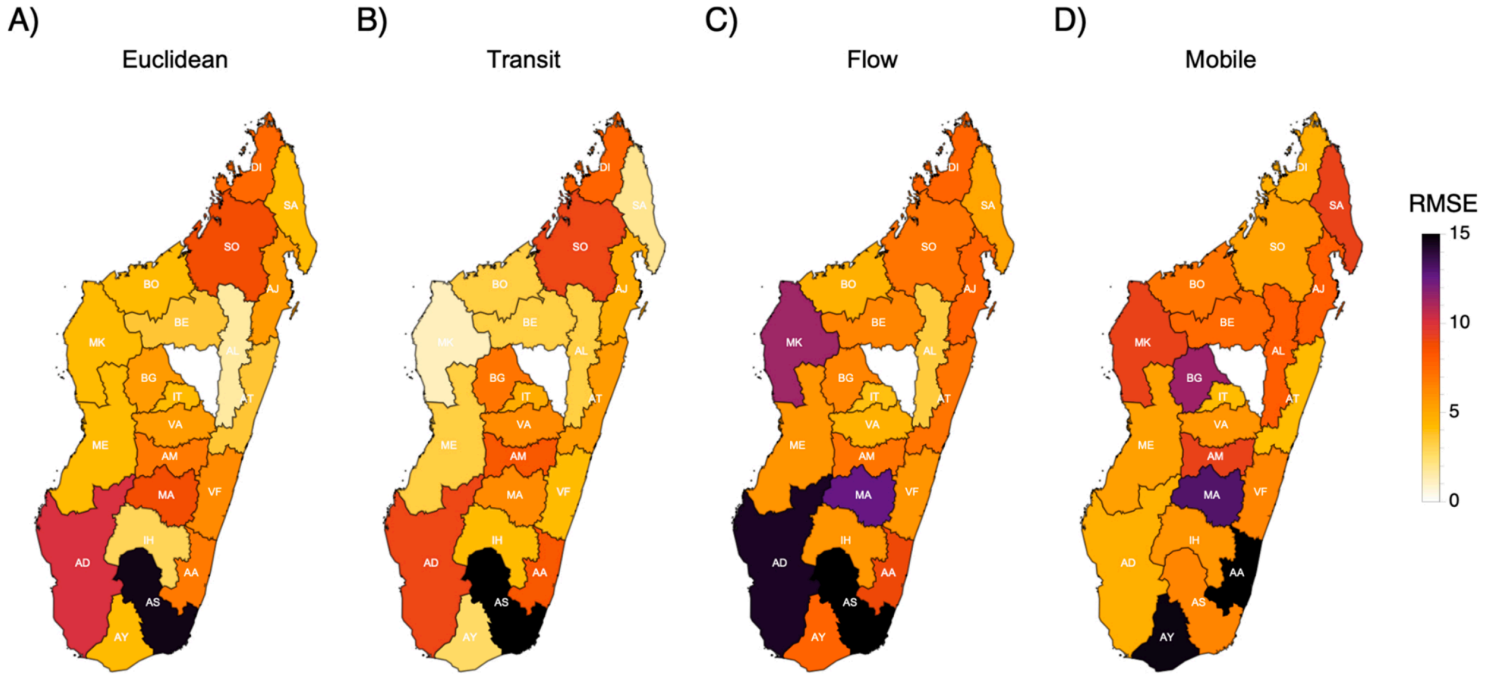


Fig. 5. The root mean squared error (RMSE) in predicting the rank of the reported fifth case for each region. The regions are colored based on the accuracy of the ranks for the reported fifth case for each region using various mobility matrices: A) Euclidean, B) transit, C) Internal Migration Flow (Flow), and D) mobile phone data.

Context and question

COVID-19 testing | March 31, 2020

Who should we test to minimize the spread of the COVID-19 in two villages in SAVA?
(Tests are limited)

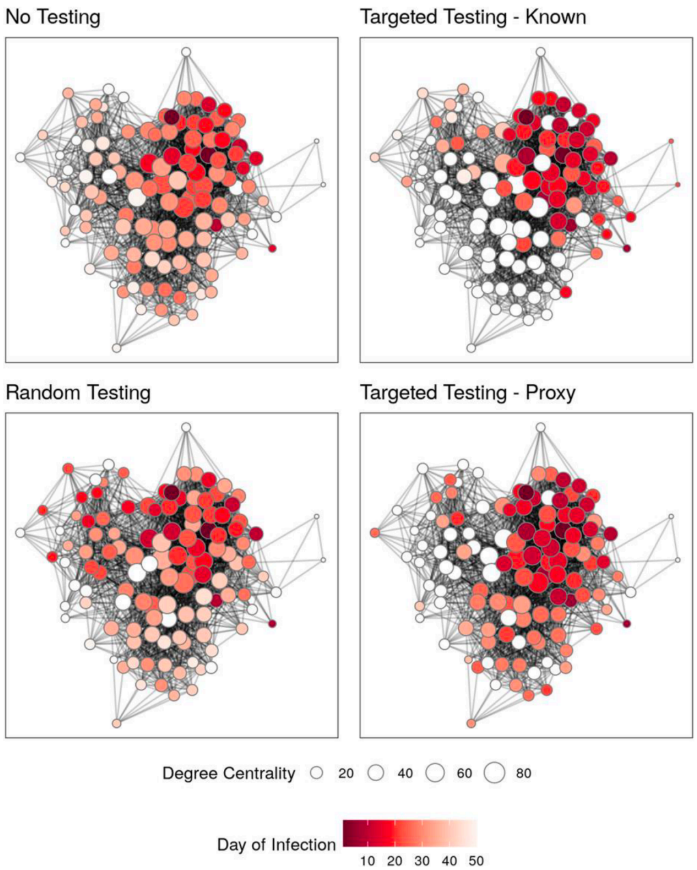
Sociodemographic Variables Can Guide Prioritized Testing Strategies for Epidemic Control in Resource-Limited Contexts

Michelle V. Evans,[1] Tanjona Ramiadantsoa,[1] Kayla Kauffman,[2,3,4] James Moody,[5] Charles L. Nunn,[2,3] Jean Yves Rabezara,[6] Prisca Raharimalala,[7] Toky M. Randriamoria,[8,9] Voahangy Soarimalala,[8,10] Georgia Titcomb,[4,11,12] Andres Garchitorena,[1,13] and Benjamin Roche[1]

[1]Maladies Infectieuses et Vecteurs : Écologie, Génétique, Évolution et Contrôle, Université Montpellier, CNRS, IRD, Montpellier, France; [2]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA; [3]Duke Global Health Institute, Durham, North Carolina, USA; [4]Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California, USA; [5]Department of Sociology, Duke University, Durham, North Carolina, USA; [6]Department of Science and Technology, University of Antsiranana, Antsiranana, Madagascar; [7]Andapa, Madagascar; [8]Association Vahatra, Antananarivo, Madagascar; [9]Zoologie et Biodiversité Animale, Domaine Sciences et Technologies, Université d'Antananarivo, Antananarivo, Madagascar; [10]Institut des Sciences et Techniques de l'Environnement, Université de Fianarantsoa, Fianarantsoa, Madagascar; [11]Marine Science Institute, University of California, Santa Barbara, California, USA; [12]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA; and [13]Pivot, Ifanadiana, Madagascar

Background. Targeted surveillance allows public health authorities to implement testing and isolation strategies when

An example of result

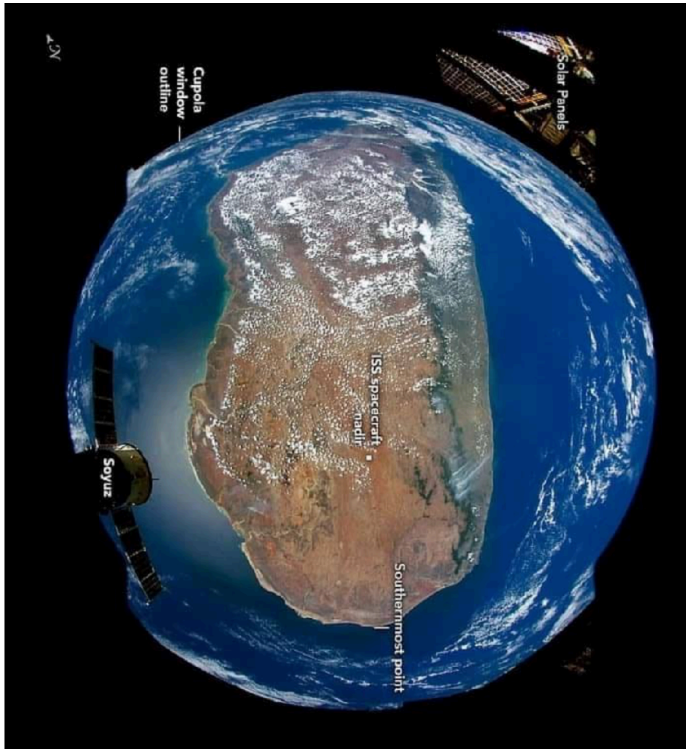


Who should we test to minimize the spread of the COVID-19 in two villages in SAVA?
(Tests are limited)

Testing regularly people with lots of contact is most effective in reducing and slowing the number of infected individuals

Examples in ecology

"We have destroyed the island,
almost not forest left"



Taken by the International Space Station (2020)

"Proofs and consequences,
Look at these erosions all over the island"

http://www.traveladventures.org/continents/africa/tsiroanomandidy-ankavandra09.html

Lavaka

Nontrivial responses of vegetation to compound disturbances: A case study of Malagasy grasslands

Tanjona Ramiadantsoa[1,2,3] & Cédrique L. Solofondranohatra[4,5]

[1] Mention Sciences de la Vie, Université de Fianarantsoa, Fianarantsoa 301, Madagascar
E-mail: ramiadantsoa@wisc.edu

[2] Mention Mathématiques, Université de Fianarantsoa, Fianarantsoa 301, Madagascar

[3] Department of Integrative Biology, University of Wisconsin-Madison, Wisconsin 53706, USA

drying could have indirectly precipitated the demise of megaherbivores by shrinking the extent of grazing-maintained grassland. These mechanisms add to the list of factors that potentially drove the extinction of Madagascar's megaherbivores. Mechanistic approaches like the one we present here are underrepresented compared to traditional empirical and statistical (correlative) approaches in

Forest

Fire-maintained grasslands

Aristida tenuissima

Grazing-maintained grasslands

Paspalum conjugatum

Photos: Maria Vorontsova

Forest

Fire-maintained grasslands

Grazing-maintained grasslands

- Without disturbance forest expands by outcompeting grasses
- With disturbance:
  - Fire kills forest and fire-maintained grassland expands
  - Grazing kills seedlings and grazing-maintained grasslands
- With high precipitation, fire does not occur, forest expands
- With intermediate precipitation, fire occurs and kills forests, FMG expands
- With low precipitation, forest and no biomass accumulation, GMG expands
- With grazing by herbivores, GMG persists (no fire and no colonization)

Original model

THEORETICAL POPULATION BIOLOGY 18, 363–373 (1980)

Disturbance, Coexistence, History, and Competition for Space

ALAN HASTINGS

Department of Mathematics, University of California, Davis, California 95616

Received February 1, 1980

A simple model to elucidate the effect of disturbance on a large number of

$$\frac{dx_i}{dt} = D_i(x_i)\left(1 - \sum_{j=1}^{i} x_j\right) - \sum_{j=1}^{i-1} D_j(x_j)\, x_i - e(t)\, x_i\,, \qquad i = 1, n. \qquad (4)$$

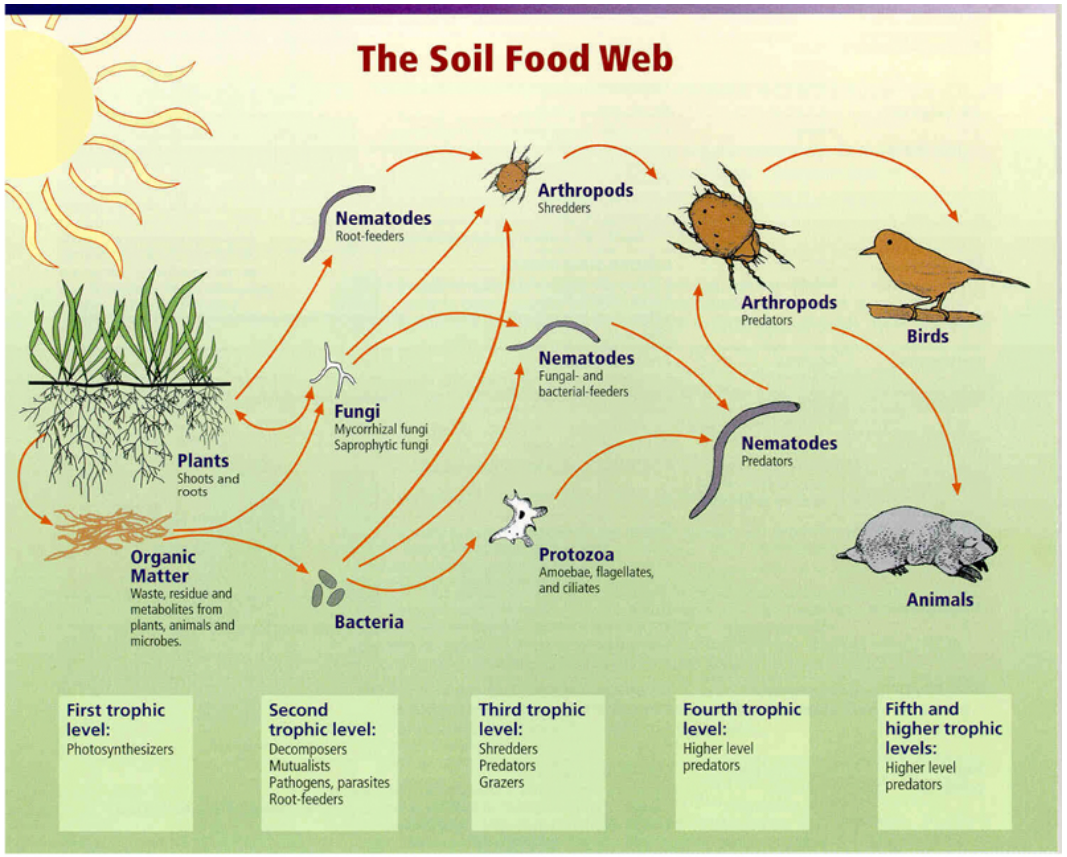
Continuous time	Deterministic	Spatial (implicit)	No age-structure

These assumptions lead to the following set of equations

$$\begin{cases} \frac{dX}{dt} = c_X X(1 - X - f\,Y - g\,Z) - m_X X \\ \frac{dY}{dt} = c_Y Y(1 - X - Y - g\,Z) - m_Y Y - (1 - f) c_X X\,Y \\ \frac{dZ}{dt} = c_Z Z(1 - X - Y - Z) - m_Z Z - (1 - g)(c_X X + c_Y Y)\,Z, \end{cases}$$

Lake Mýtvan, Iceland

Midge swarms (still no clear explanation of the phenomenon)



Relationships between soil food web, plants, organic matter, and birds and mammals
Image courtesy of USDA Natural Resources Conservation Service
http://soils.usda.gov/sqi/soil_quality/soil_biology/soil_food_web.html.

What are the effect of midge swarm on nitrogen cycle within the terrestrial foodweb?

Transient top-down and bottom-up effects of resources pulsed to multiple trophic levels

Matthew A. McCary [1,3,5] Joseph S. Phillips [2,4] Tanjona Ramiadantsoa [2] Lucas A. Nell [2] Amanda R. McCormick [2] and Jamieson C. Botsch [2]

[1]*Department of Entomology, University of Wisconsin, Madison, Wisconsin 53706 USA*
[2]*Department of Integrative Biology, University of Wisconsin, Madison, Wisconsin 53706 USA*

Citation: McCary, M. A., J. S. Phillips, T. Ramiadantsoa, L. A. Nell, A. R. McCormick, and J. C. Botsch. 2020. Transient top-down and bottom-up effects of resources pulsed to multiple trophic levels. Ecology 00 (00):e03197. 10.1002/ecy.3197

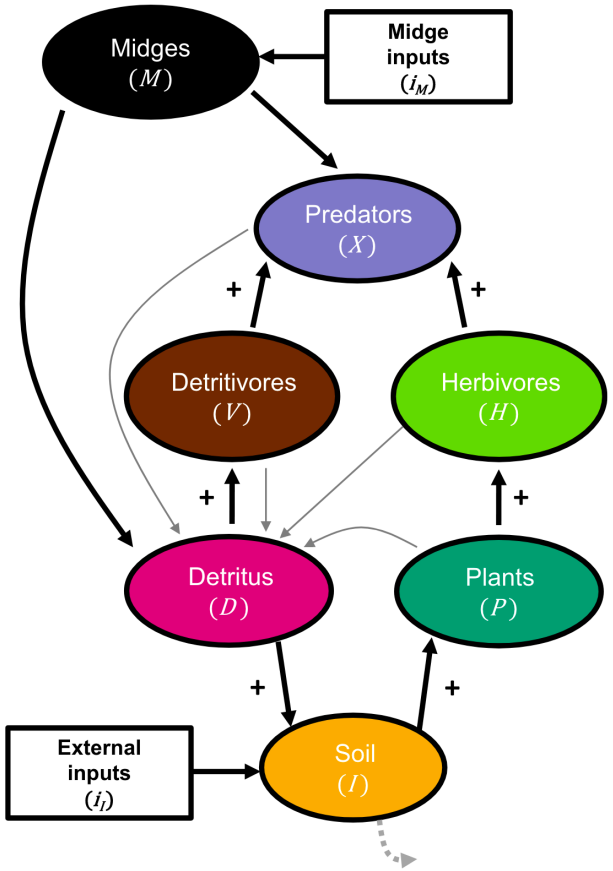
Abstract. Pulsed fluxes of organisms across ecosystem boundaries can exert top-down and bottom-up effects in recipient food webs, through both direct effects on the subsidized trophic levels and indirect effects on other components of the system. While previous theoretical and empirical studies demonstrate the influence of allochthonous subsidies on bottom-up and top-

Sources of the original models

We analyzed a hybrid community–ecosystem model as a system of ordinary differential equations (ODEs) combining classical consumer–resource dynamics (e.g., Rosenzweig and MacArthur 1963) with nutrient cycling via inorganic soil and detrital N pools (DeAngelis 1992, Loreau 2010). The model contains seven N pools: midges (*M*), inorganic soil N (*I*), detritus (*D*), plants (*P*), detritivorous arthropods (*V*), herbivorous arthropods (*H*), and predatory arthropods (*X*; Fig. 1). While the model does

Plant uptake of N from the soil-nutrient pool, detritivore consumption of detritus, and herbivore consumption of plants are all modeled as type-II functional responses (Holling 1959), whereby uptake saturates with increasing resource availability. Predator consumption of herbivores, detritivores, and midges is modeled as a multispecies type-II functional response (Murdoch 1969), whereby consumption of each of the pools satu-

Continuous time	Deterministic	Non-spatial	No age-structure



$$\frac{dI}{dt} = i_I + (1-l)\mu_D D - \frac{a_I I P}{1 + a_I h_I I} - \mu_I I$$

$$\frac{dD}{dt} = (1-l)\mu_M M + (1-l) \sum_{j \in \{P,V,H,X\}} \left(\mu_j + m_j j\right) j - \frac{a_D D V}{1 + a_D h_D D} - \mu_D D$$

$$\frac{dP}{dt} = \frac{a_I I P}{1 + a_I h_I I} - \frac{a_P P H}{1 + a_P h_P P} - (\mu_P + m_P P) P$$

$$\frac{dV}{dt} = \frac{a_D D V}{1 + a_D h_D D} - \frac{a_X V X}{1 + a_X h_X H + a_X h_X V + q a_X h_M M} - (\mu_V + m_V V) V$$

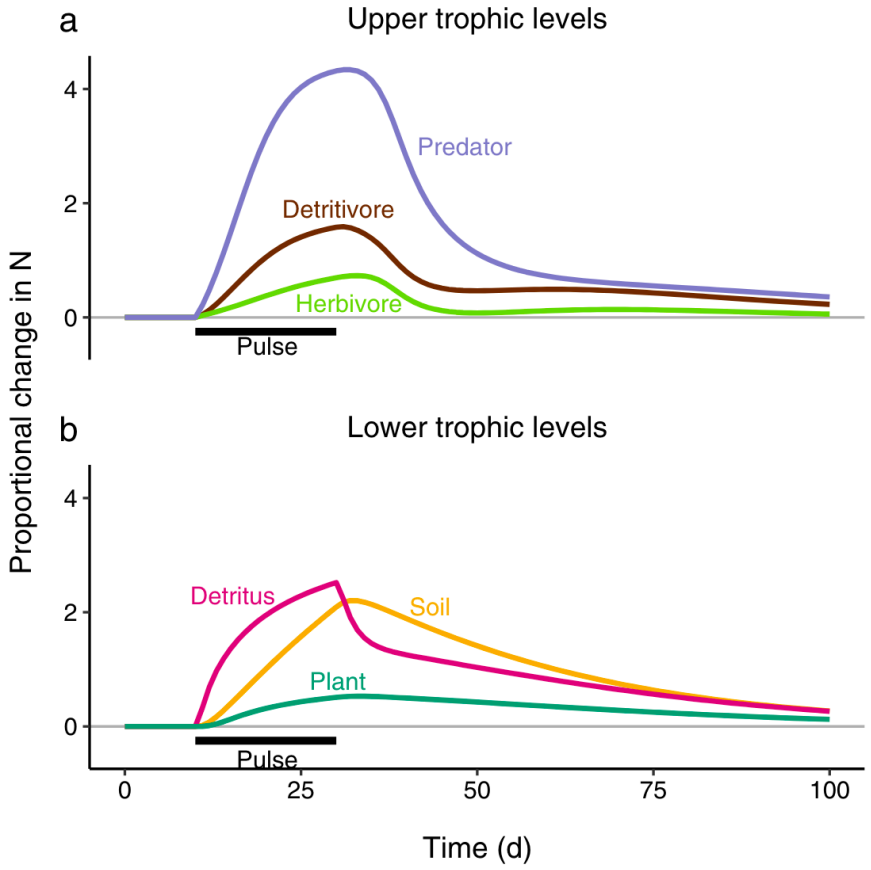
$$\frac{dH}{dt} = \frac{a_P P H}{1 + a_P h_P P} - \frac{a_X H X}{1 + a_X h_X H + a_X h_X V + q a_X h_M M} - (\mu_H + m_H H) H$$

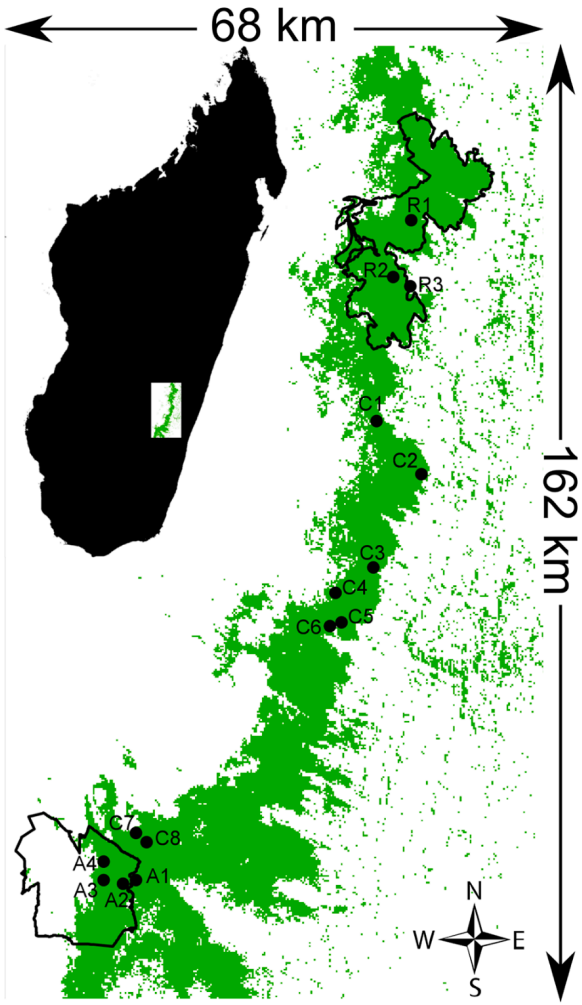
$$\frac{dX}{dt} = \frac{(a_X V + a_X H + q a_X M) X}{1 + a_X h_X H + a_X h_X V + q a_X h_M M} - (\mu_X + m_X X) X$$

$$\frac{dM}{dt} = i_{M(t)} - \frac{q a_X M X}{1 + a_X h_X H + a_X h_X V + q a_X h_M M} - \mu_M M$$

(1)

An example of result





How does the occupancy of a species decline if the corridor is deforested ?

RESEARCH ARTICLE

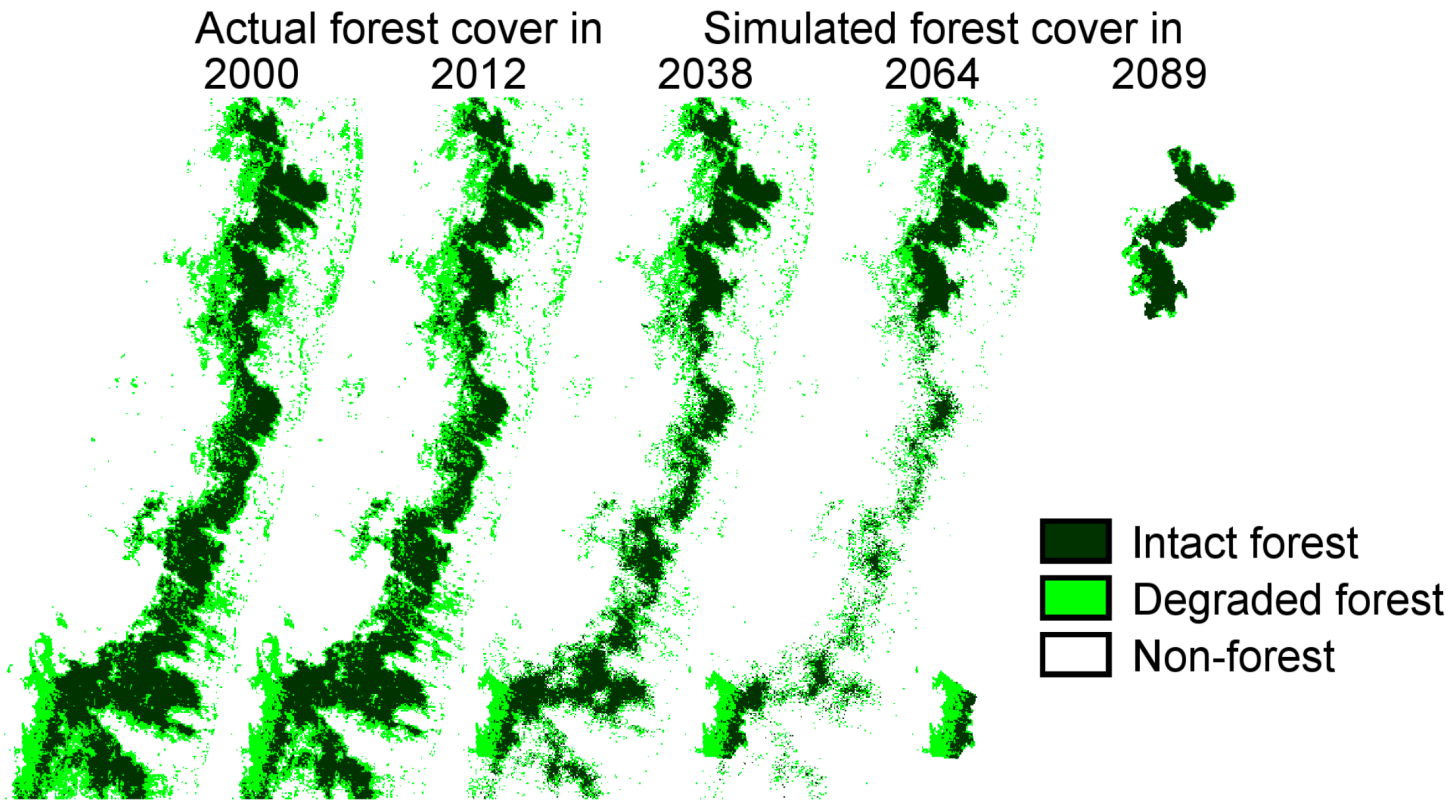
Large-Scale Habitat Corridors for Biodiversity Conservation: A Forest Corridor in Madagascar

Tanjona Ramiadantsoa[1¤]*, Otso Ovaskainen[1], Joel Rybicki[2,3], Ilkka Hanski[1]

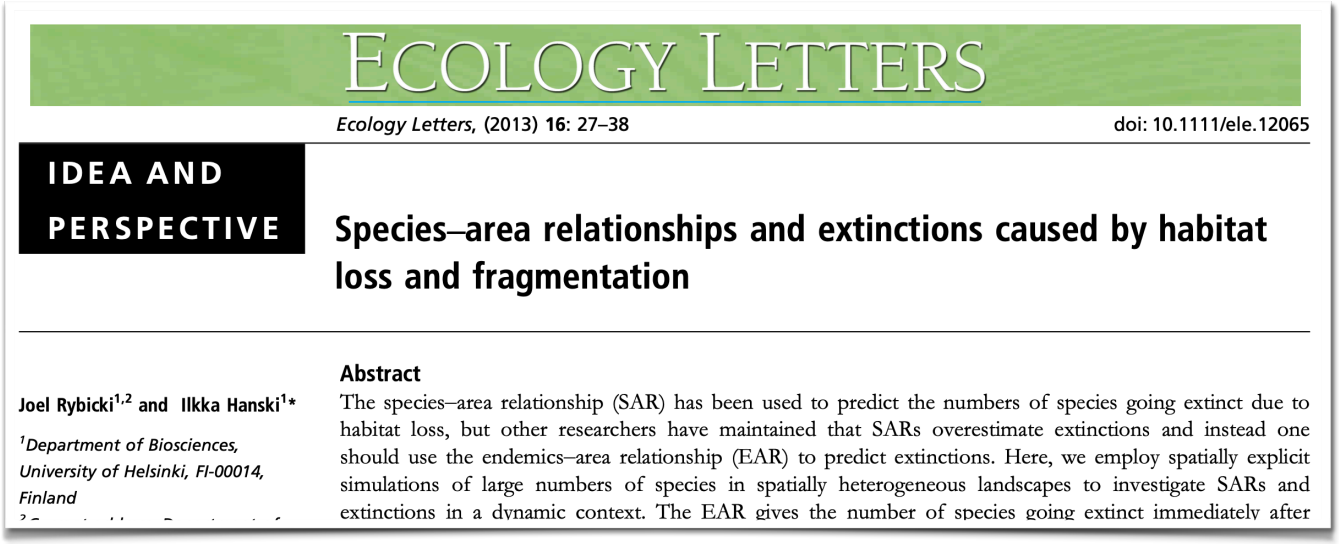
1 Metapopulation Research Centre, Department of Biosciences, University of Helsinki, Helsinki, Finland, **2** Department of Computer Science, Aalto University, Espoo, Finland, **3** Department of Algorithms and Complexity, Max Planck Institute for Informatics, Saarbrücken, Germany

¤ Current Address: Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, United States of America
* tramiada@umn.edu

Simulated deforestation scenarios



Sources of the original models



ECOLOGY LETTERS

Ecology Letters, (2013) **16**: 27–38 doi: 10.1111/ele.12065

IDEA AND PERSPECTIVE

Species–area relationships and extinctions caused by habitat loss and fragmentation

Joel Rybicki[1,2] and Ilkka Hanski[1]*

[1]Department of Biosciences, University of Helsinki, FI-00014, Finland

Abstract

The species–area relationship (SAR) has been used to predict the numbers of species going extinct due to habitat loss, but other researchers have maintained that SARs overestimate extinctions and instead one should use the endemics–area relationship (EAR) to predict extinctions. Here, we employ spatially explicit simulations of large numbers of species in spatially heterogeneous landscapes to investigate SARs and extinctions in a dynamic context. The EAR gives the number of species going extinct immediately after

Model description

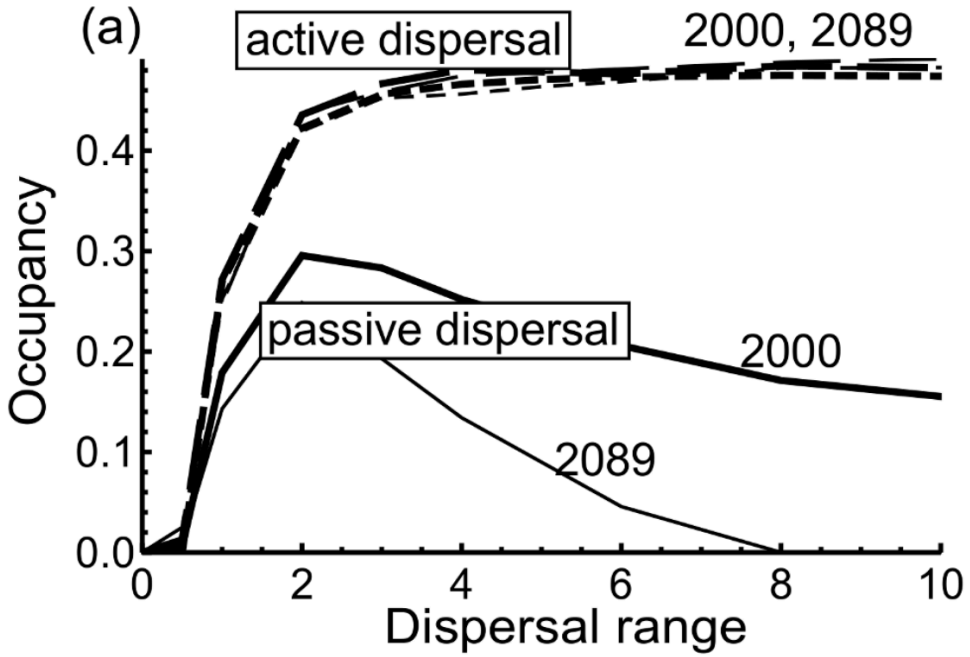
In this section, we describe the computational model, while the mathematical definitions and further technical details are presented in the online supporting information.

Stochastic patch occupancy model

We use a spatially explicit stochastic patch occupancy model (SPOM). These models are well-established in spatial ecology and metapopulation modelling (Moilanen 1999; Ovaskainen & Hanski 2004). In our case, the patch network is represented by a finite regular grid, in which each cell represents a discrete habitat patch. Each cell is associated with a value h, $0 \leq h \leq 1$, which deter-

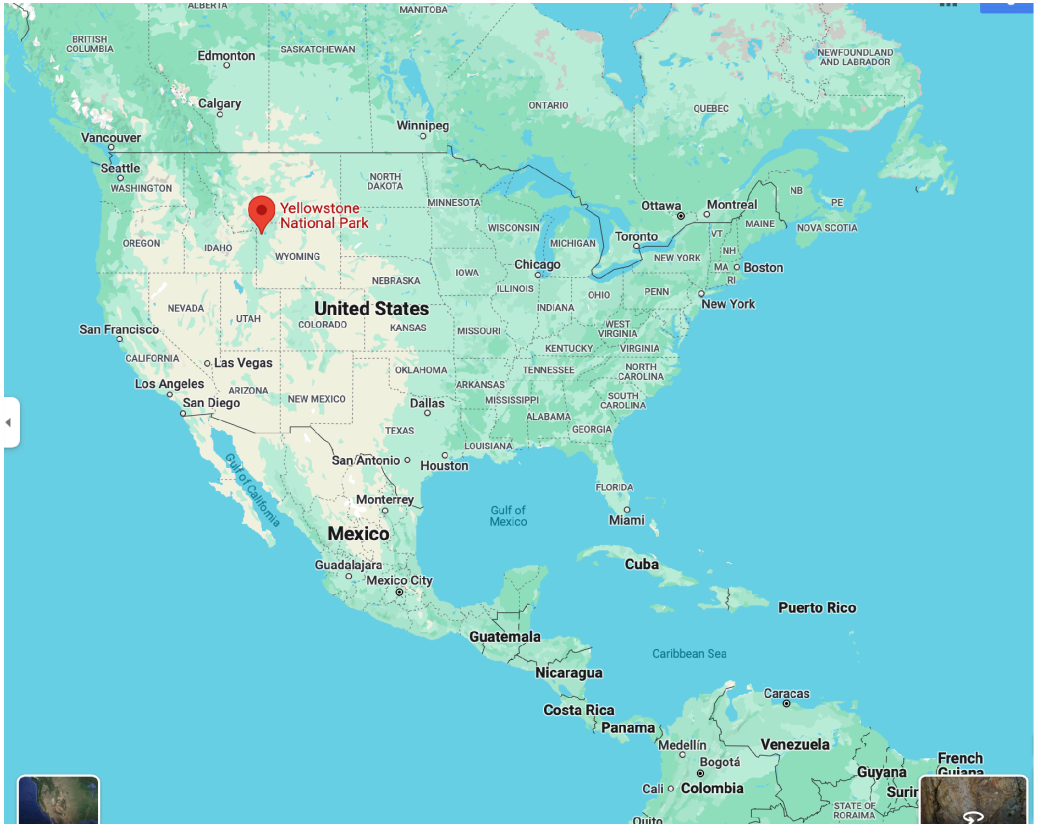
Discrete time	Stochastic	Spatial (discrete)	No age-structure

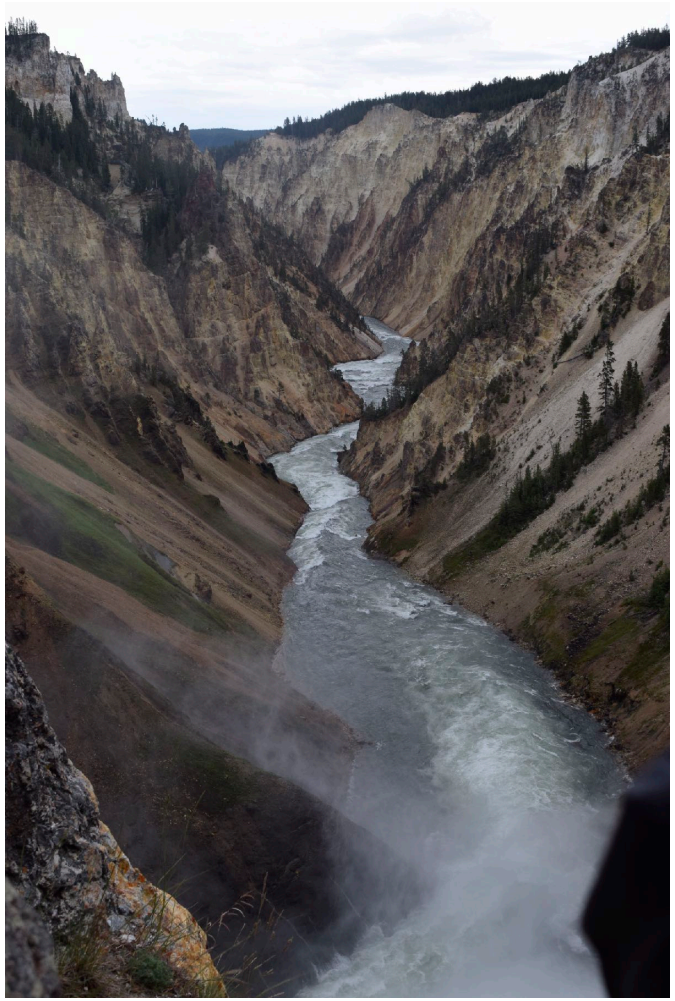
Example of result

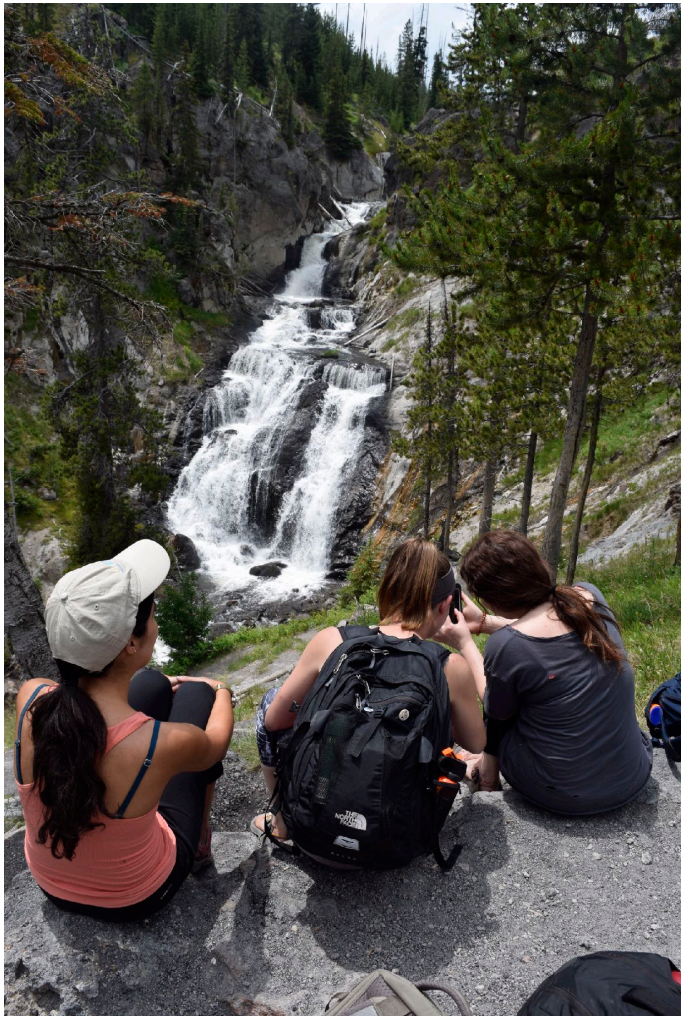


- Active dispersers are not affected by destruction of the corridor
- Passive disperser can go extinct in 2089, especially if they have a long dispersal range

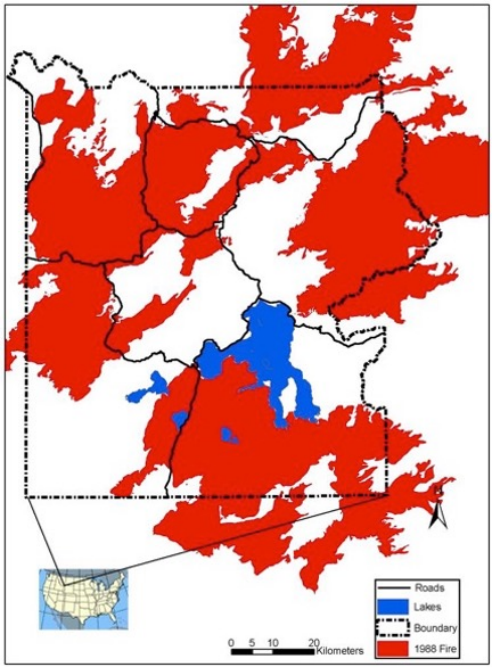
Yellowstone National Park (9000 km2 = 900000 ha)







1988 fire burned ~30%



Serotinous lodgepole pine needs fire to release seeds

ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Regeneration strategies and forest resilience to changing fire regimes: Insights from a Goldilocks model

Tanjona Ramiadantsoa[1,2] | **Zak Ratajczak**[1,3] | **Monica G. Turner**[1]

[1]Department of Integrative Biology, University of Wisconsin-Madison, Madison, Wisconsin, USA

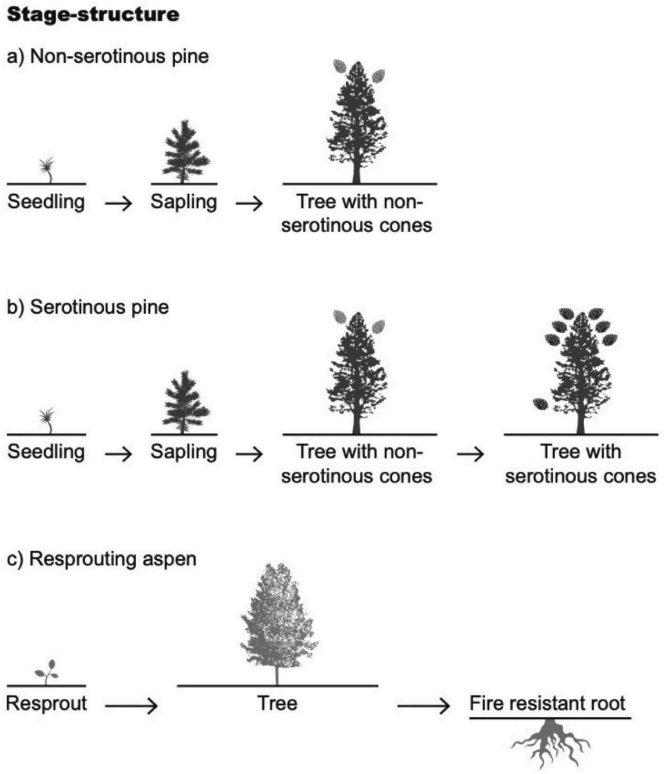
[2]Madagascar Biodiversity Center, Antananarivo, Madagascar

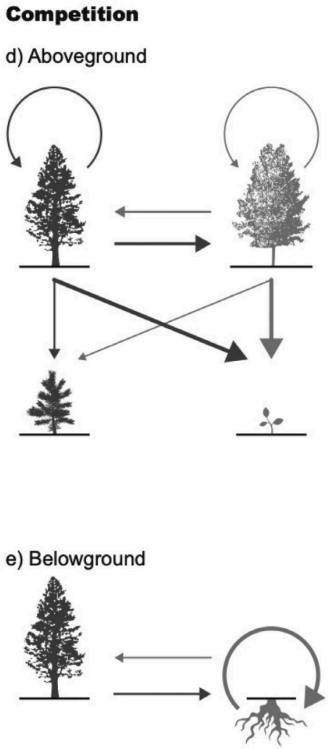
[3]Department of Biology, Kansas State University, Manhattan, Kansas, USA

Abstract

Disturbances are ubiquitous in ecological systems, and species have evolved a range of strategies to resist or rebound following disturbance. Understanding how the presence and complementarity of regeneration traits will affect community responses to disturbance is increasingly urgent as disturbance regimes shift beyond their historical ranges of variability. We define "disturbance niche" as a

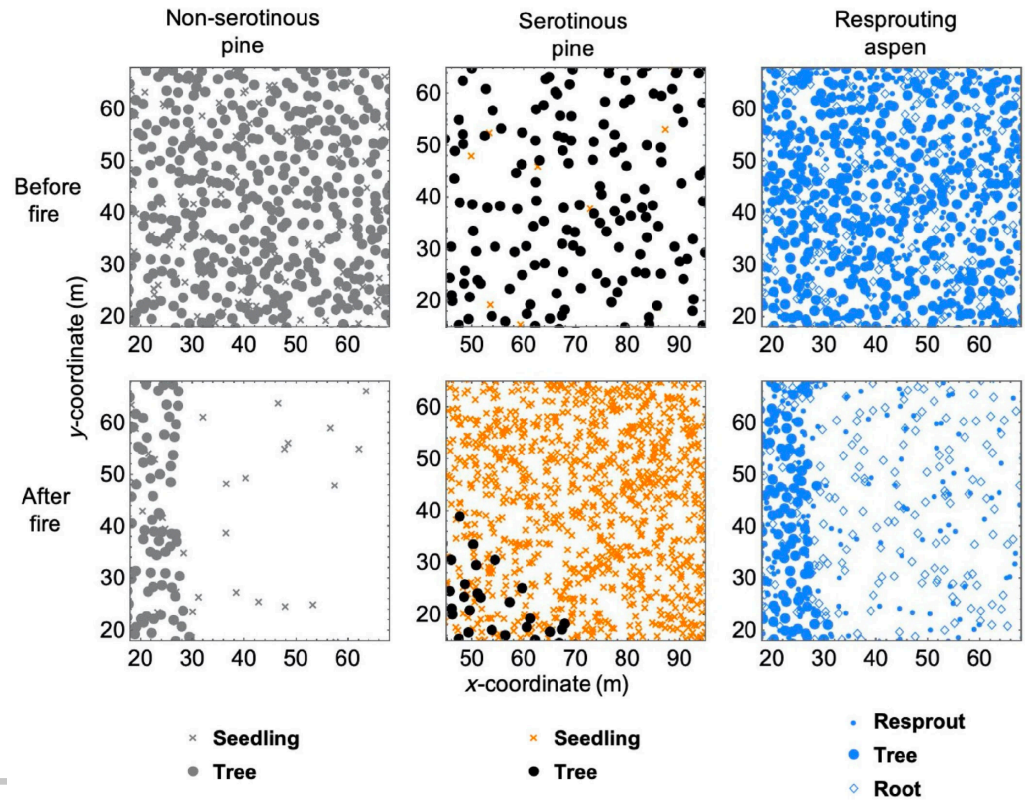
Continuous time	Stochastic	Spatial	With stage-structure

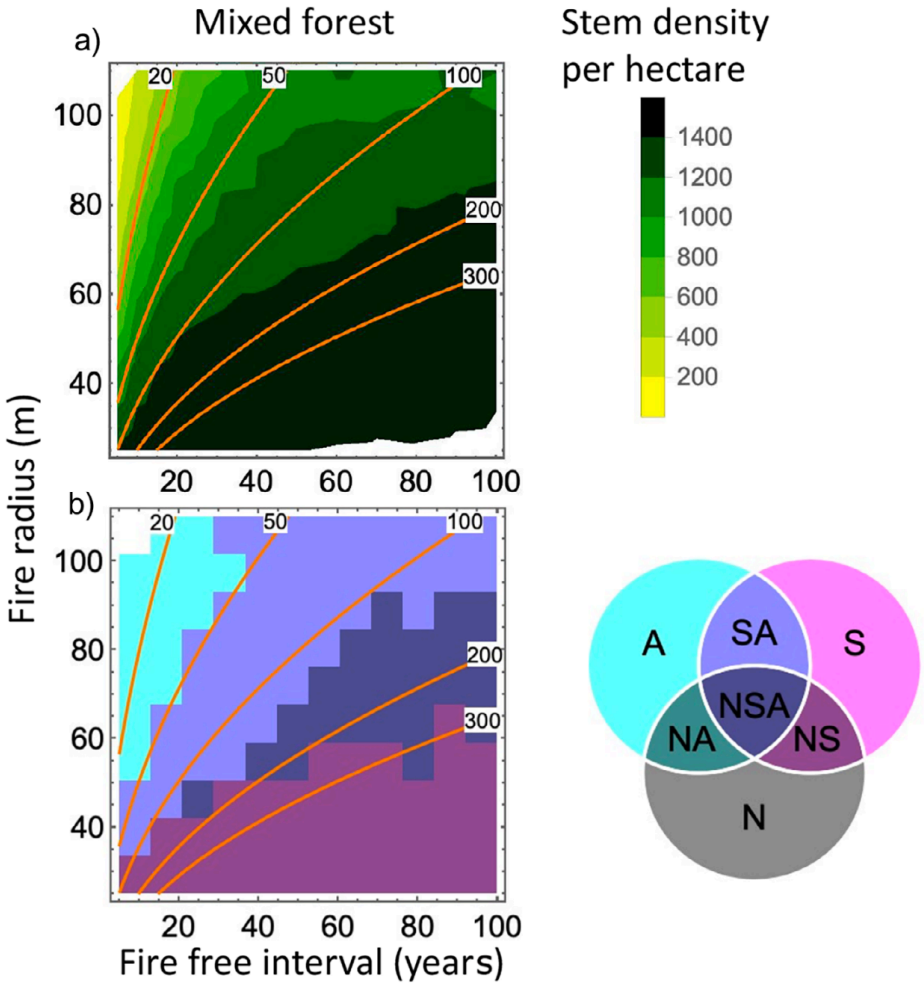




Technically, the model is a spatiotemporal point process, commonly used in theoretical ecology (Bolker & Pacala, 1997) with spatially explicit Lotka-Volterra competition (stronger competition when individuals are in closer proximity) and stage-structure (e.g., passing from seedling, to sapling, to multiple adult tree stages).

A snapshot of an output





- Total tree density gradually declines as fires become larger and more frequent
- Forest composition will change

Outline

- The intangibles
- Advanced concepts
- Mechanistic modeling
- **Statistical modeling**
- Beyond this workshop

Statistical model:
Thinking about correlation

Semantic

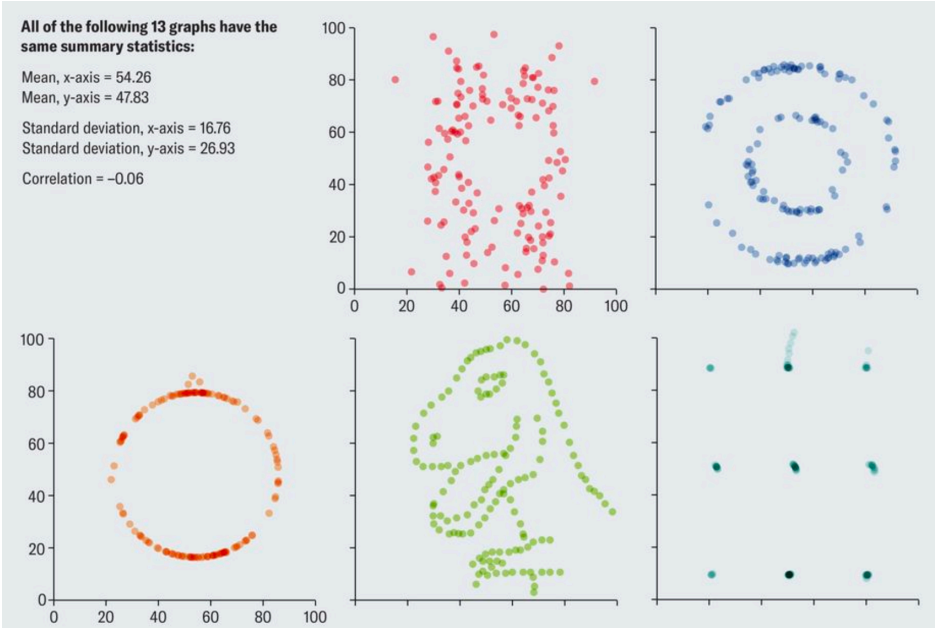
- Explanatory, predictor, independent, X
- Response, dependent, predicted, Y
- Type of variables:
  - Qualitative/Factor/Category/Group/Class
  - Quantitative

Basic statistics

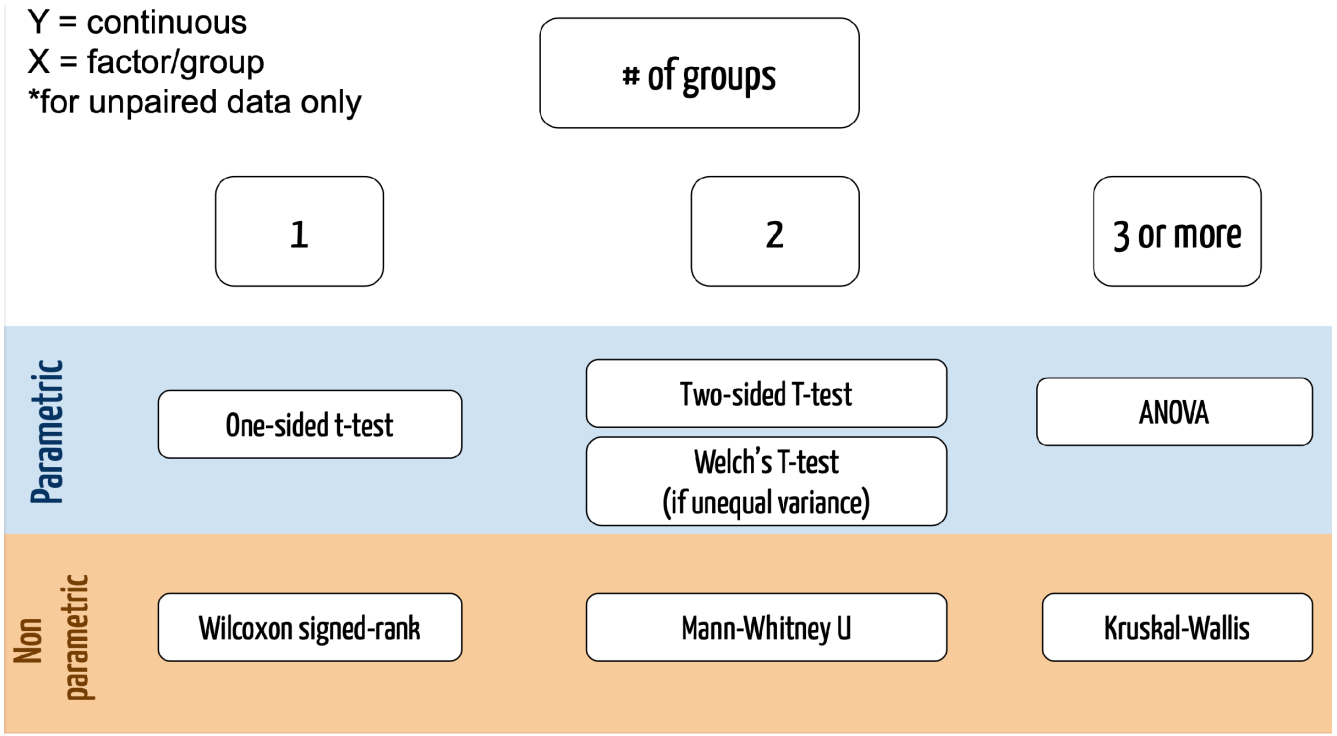
Revisiting correlation

Two most common types of correlations

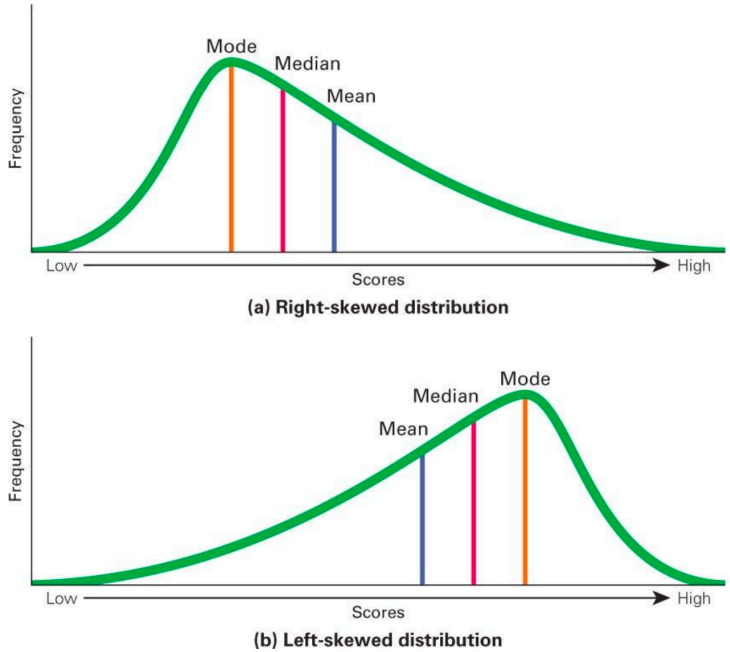
Pearson's Correlation	Spearman's Correlation
Normally-distributed data *Distribution normal*	Does not require normally distributed data *N'exige pas les données avec une distribution normal*
Linear relationship *Association linèare*	Correlation is based on rank, not linear relationship *La corrélation est basé sur leur ordre, pas une association linéaire*
Both variables numeric *Les deux variables doivent être numériques*	Both variables numeric *Les deux variables doivent être numériques*
"Mean"- based *Basé sur la moyenne*	"Median"-based *Basé sur la médiane*



A recipe for choosing a model



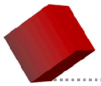
Difference between right-skewed and left-skewed distributions



Normal distribution is symmetric (no skew), i.e., mean = mode = median

Using shapiro.test(variable) in R

A tale of linear model: Part I



Simple linear regression

$$y = \alpha + \beta * x + \varepsilon$$

Response variable = Systematic component + Residual component

Systematic component → Intercept and explanatory variables

Residual component →
- Null mean
- Independence
- Fixed variance
- Normality

The R function to fit a linear model is lm() which uses the form
fitted.model <- lm(formula, data=data.frame)

A tale of linear model: Part I

- Linear model: $y = ax + b + e$ where $e \sim N(0, \sigma^2)$ or in R formula notation $y \sim x$
- Variants:
  - (Multivariate) $y = a_1x_1 + a_2x_2 + \cdots + a_nx_n + e$, or $y \sim x_1 + x_2 + \cdots + x_n$
  - (Still linear) $y = a_1x_1 + a_2x_2^2 + a_3\sqrt{x_3} + e$ or $y \sim x_1 + x_2^2 + \sqrt{x_3}$
  - (Interaction) $y = a_1x_1 + a_2x_2 + a_3x_1x_2 + e$ or $y \sim x_1 * x_2$

What assumptions do we need to consider for parametric tests?

Linearity : The relation between two variables is linear

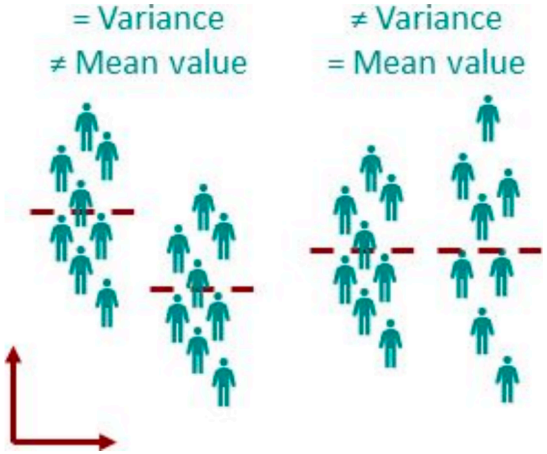
Independence : Each observation is independent of all other observations

Normality: The distribution of the data must be normal

Homogeneity of variance: The variance of subsets

or groups of data should be equal

Les variations des groupes de données doivent être égal



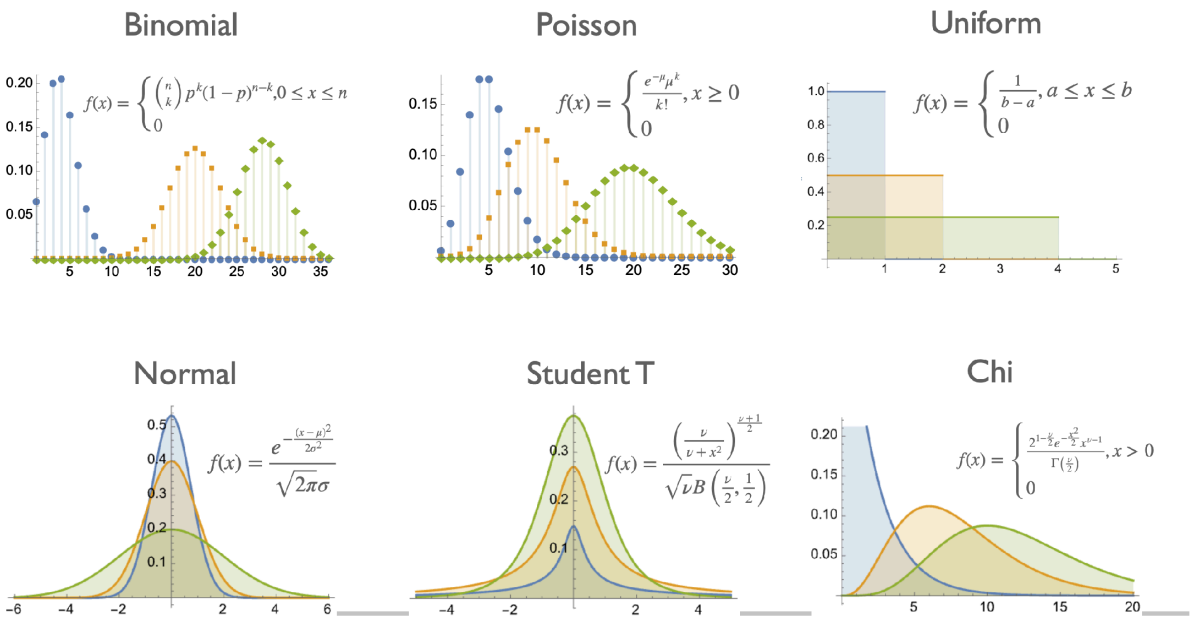
A tale of linear model: Part 2

Switching notation

- Linear model: $y = ax + b + e$ where $e \sim N(0, \sigma^2)$
- A math transformation
  - $m + N(0, \sigma^2) = N(m, \sigma^2)$
  - $y \sim ax + b + N(0, \sigma^2) \Leftrightarrow y \sim N(ax + b, \sigma^2)$

A tale of linear model: Part 2

When normality is not satisfied i.e., $y \sim N(ax + b, \sigma^2)$, pick a different distribution



A tale of linear model: Part 2

When normality is not satisfied i.e., $y \nsim N(ax + b, \sigma^2)$

- In other words, the response variable is not a real number
- If the response variable is a count (positive integer)
  - We can assume that $y \sim \text{Poisson}(\mu)$, μ is called the intensity and should be a positive number
  - The model becomes $y \sim \text{Poisson}(e^{ax+b})$, because the exponential is always positive
  - By the taking the reciprocal of exponential, log, we have a linear model $\log(y) = ax + b$

For Poisson process, the variance is equal to the mean. You need to check that the dispersion parameter is one, otherwise a different family (quasipoisson) might be more appropriate

A tale of linear model: Part 2

When normality is not satisfied i.e., $y \nsim N(ax + b, \sigma^2)$

- In other words the response variable is not a real number (could be negative and continuous)
- If the response variable is binary (presence or absence, 0 or 1)
  - We can assume that $y \sim$ Binomial$(1, p)$, where p is the probability of success after one trial
  - But p is a probability and needs to be between 0 and 1, we take $p = f(ax + b) = \frac{e^{ax+b}}{1 + e^{ax+b}}$, for any a, x, and b, p is always between 0 and 1
  - The reciprocal of f, is the logit function $\log(\frac{p}{1-p})$, we have a linear model $\text{logit}(p) = ax + b$

A tale of linear model: Part 3

When might we use GLMM?

Generalized linear mixed models include both fixed effects and random effects to allow for:

- *Repeated measures*
- *Temporal correlation*
- *Spatial correlation*
- *Heterogeneity*
- *Nested data*

We use the model glmer() in R

Model = glmer(formula, family=family, data=data)

A tale of linear model: Part 3

When independence is not satisfied i.e., $y \sim N(ax + b, \sigma^2(x))$

When might we use GLMM?

Generalized linear mixed models include both fixed effects and random effects to allow for:

- *Repeated measures*
- *Temporal correlation*
- *Spatial correlation*
- *Heterogeneity*
- *Nested data*

We use the model glmer() in R

Model = glmer(formula, family=family, data=data)

A tale of linear model: Part 3

Fixed or Random?

- Half of E2M2 students worked with Tanjona for their research questions and the other half worked with Sophie
- For each student, the response variable is happiness after the meeting, the explanatory variables are: age, sex, familiarity with mathematics, and the identity of the instructor
- We include instructor as fixed effect (Tanjona OR Sophie) if we want to compare if we want to know the difference in happiness after meeting with Tanjona and Sophie
- We choose instructor as random effect if we are not interested in comparing the two instructors but want to acknowledge that their happiness could depend on the instructor

Explaining a GLMM results table

```
> summary(m6)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: Negative Binomial(1.3844)  ( log )
Formula: GIparasites ~ treatment + (1 | year)
   Data: lemur.data

     AIC      BIC   logLik deviance df.resid
  3363.3   3379.3  -1677.7   3355.3      396

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.1448 -0.6781 -0.3015  0.5365  3.3872

Random effects:
 Groups Name        Variance Std.Dev.
 year   (Intercept) 0.01263  0.1124
Number of obs: 400, groups:  year, 4

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.2623     0.0743  43.906  < 2e-16 ***
treatment1   -0.5841     0.1640  -3.562 0.000368 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
           (Intr)
treatment1 -0.283
```

Form of your model

AIC for model selection

Random effect variance

Fixed effects

A tale of (non-)linear model: Part 4

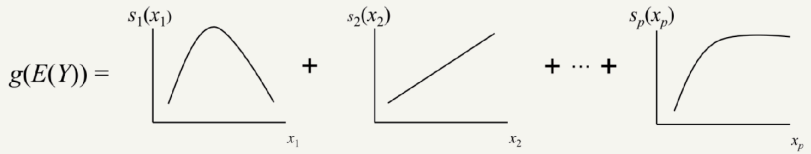
Most common options for smoothers
bs= "?"

Thin-plate = "tp"
Cubic regression = "cr" or "cs"
P-spline = "ps" or "cp" (cyclic version)
Random effects = "re"
And more….

When linearity is out of reach

Introducing GAMs

An additive model where the impact of the predictive or independent variables is captured through smoothing functions:



We can write the GAM structure as:

$$g(E(Y)) = \alpha + s_1(x_1) + \cdots + s_p(x_p),$$

***Importantly – you can control how smooth the predictor functions are!*

Introducing GAMs

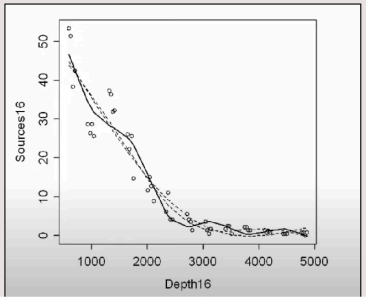
A spline curve is a is piecewise polynomial curve, i.e., it joins two or more polynomial curves. The locations of the joins are known as "knots".

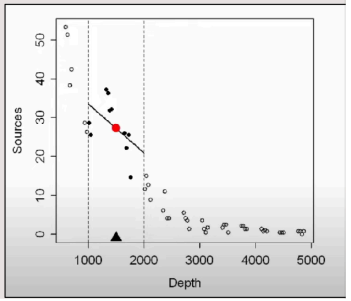
Three classes of smoothers:

- *Local regression (loess)*
- *Smoothing splines*
- *Regression splines (B-splines, P-splines, thin plate splines)*

```
gam <- gam(response ~ predictor+
s(smoother), k=#, bs="X"),  data=data,
family="family")
```

Helpful link!





A summary

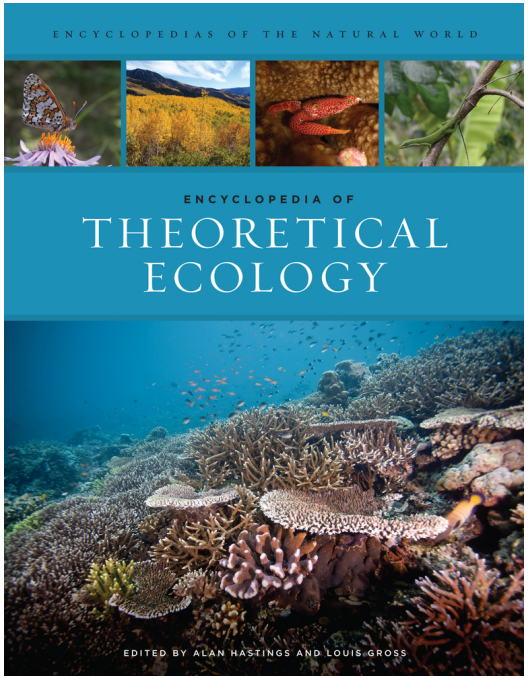
- ANOVA is a generalization of t-test (two or more categories of qualitative explanatory variable)
- lm is a generalization of an ANOVA (explanatory variable can be qualitative or quantitative)
- glm is a generalization of lm (response variable does not need to follow normal distribution)
- glmm is a generalization of glm (data points are correlated/non-independent/ pseudoreplication)
- gam relaxed the linearity of the explanatory variables

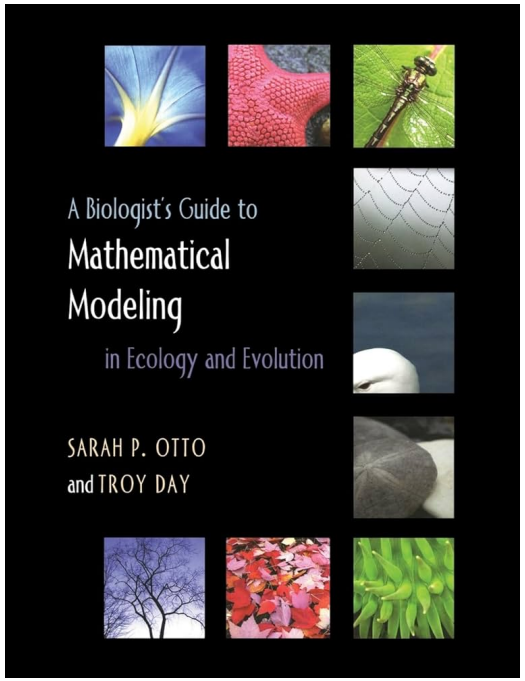
Outline

- The intangibles
- Advanced concepts
- Mechanistic modeling
- Statistical modeling
- **Beyond this workshop**

How to choose a mechanistic model

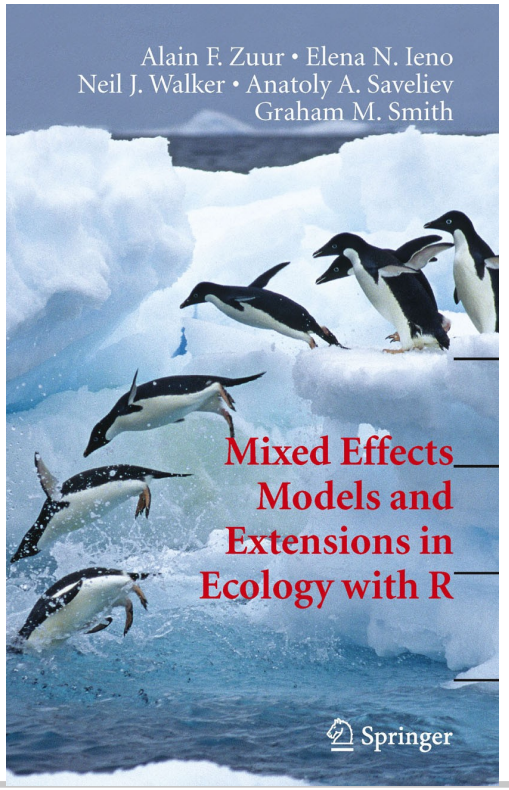
- Context and question
- Modify existing model
- Capacity to solve or/and simulate the model

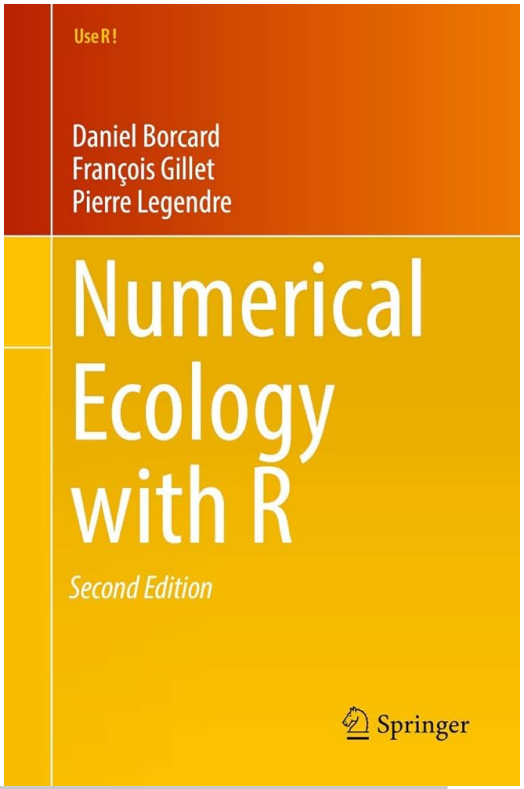


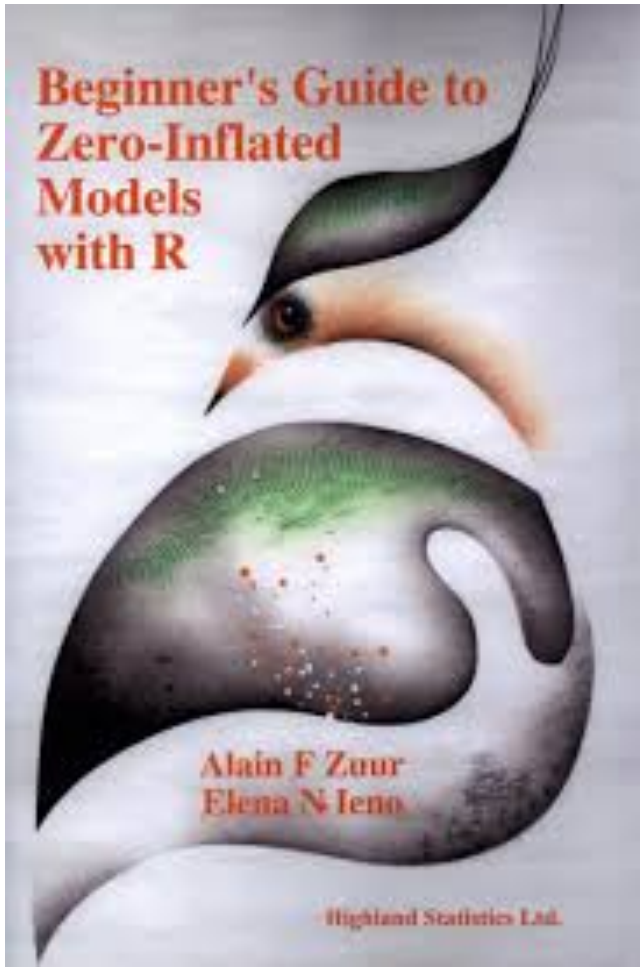


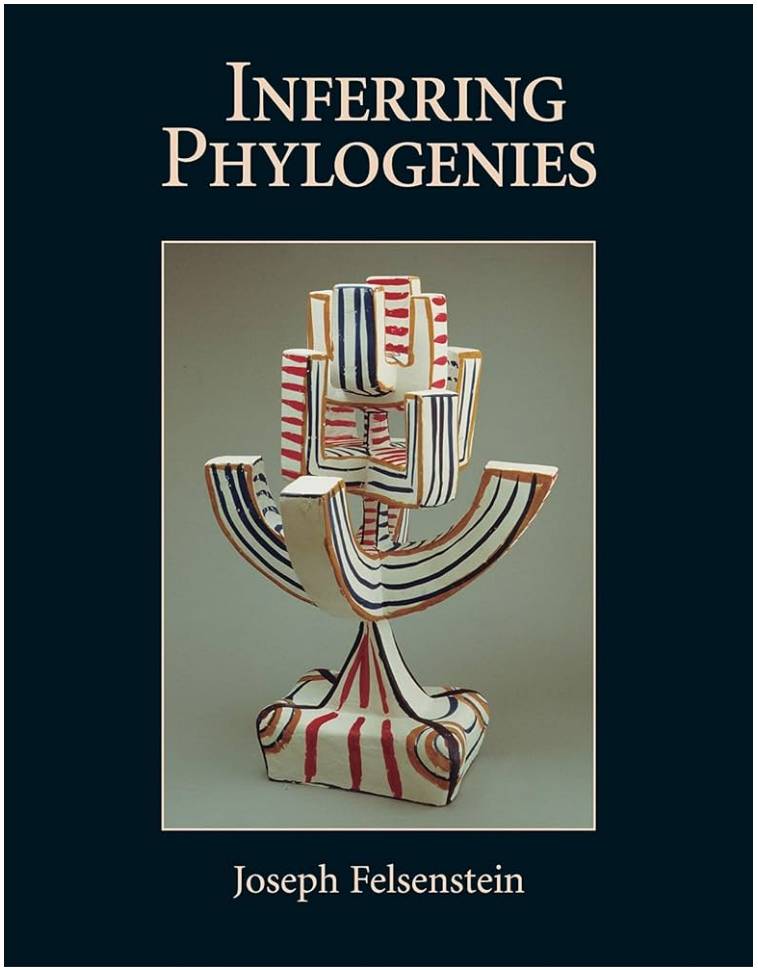
How to choose a statistical model

- Depends on the structure of your data
- Ideally, choose before collecting your data

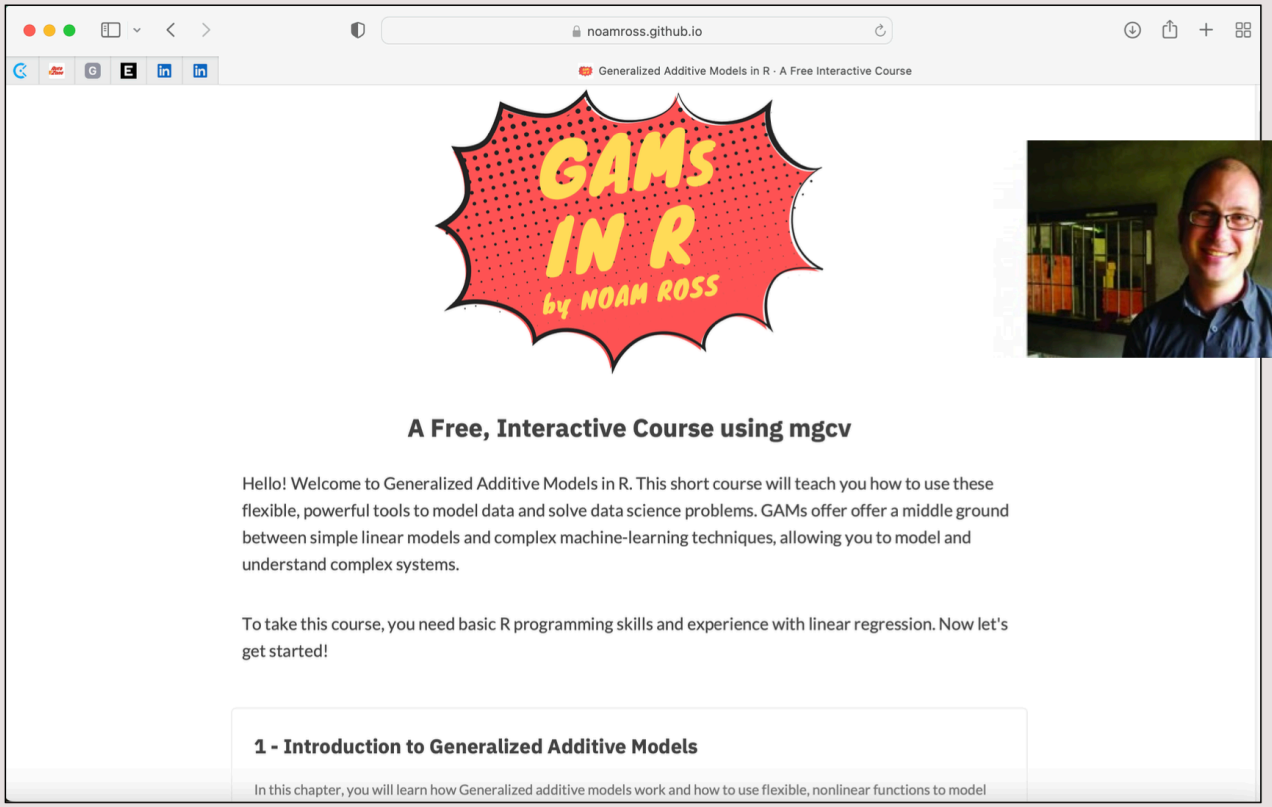





Beginner's Guide to
Zero-Inflated
Models
with R
Alain F Zuur
Elena N Ieno
Highland Statistics Ltd.

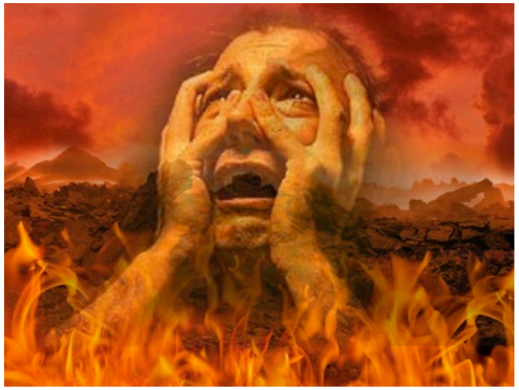

INFERRING
PHYLOGENIES
Joseph Felsenstein

Do you love GAMs now and you want to learn more?



https://noamross.github.io/gams-in-r-course/

p-value



The Earth Is Round ($p < .05$)

Jacob Cohen

After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists. This article reviews the problems with this practice, including its near-universal misinterpretation of p *as the probability that* H_0 *is false, the misinterpretation that its complement is the probability of successful replication, and the mistaken assumption that if one rejects* H_0 *one thereby affirms the theory that led to the test. Exploratory data analysis and the use of graphic methods, a steady improvement in and a movement toward standardization in measurement, an emphasis on estimating effect sizes using confidence intervals, and the informed use of available statistical methods is suggested. For generalization, psychologists must finally rely, as has been done in all the older sciences, on replication.*

sure how to test H_0, chi-square with Yates's (1951) correction or the Fisher exact test, and wonders whether he has enough power. Would you believe it? And would you believe that if he tried to publish this result without a significance test, one or more reviewers might complain? It could happen.

Almost a quarter of a century ago, a couple of sociologists, D. E. Morrison and R. E. Henkel (1970), edited a book entitled *The Significance Test Controversy*. Among the contributors were Bill Rozeboom (1960), Paul Meehl (1967), David Bakan (1966), and David Lykken (1968). Without exception, they damned NHST. For example, Meehl described NHST as "a potent but sterile intellectual rake who leaves in his merry path a long train of ravished maidens but no viable scientific offspring" (p. 265). They were, however, by no means the first to do so.

Cohen 1994

Statisticians also recommend to move away from AIC

Received: 6 November 2018 | Accepted: 18 January 2019
DOI: 10.1111/2041-210X.13159

COMMENTARY

Methods in Ecology and Evolution

I can see clearly now: Reinterpreting statistical significance

Jonathan Dushoff[1] | Morgan P. Kain[1] | Benjamin M. Bolker[1,2]

[1]Department of Biology, McMaster University, Hamilton, Ontario, Canada
[2]Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

BIOLOGY LETTERS

royalsocietypublishing.org/journal/rsbl

Review

Cite this article: Halsey LG. 2019 The reign of the *p*-value is over: what alternative analyses could we employ to fill the power vacuum? *Biol. Lett.* **15**: 20190174.

Animal behaviour

The reign of the *p*-value is over: what alternative analyses could we employ to fill the power vacuum?

Lewis G. Halsey

University of Roehampton, London SW15 4JD, UK

LGH, 0000-0002-0786-7585

The *p*-value has long been

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 1–19: Editorial
https://doi.org/10.1080/00031305.2019.1583913

Taylor & Francis

EDITORIAL

OPEN ACCESS

Moving to a World Beyond "$p < 0.05$"

Some of you exploring this special issue of *The American Statistician* might be wondering if it's a scolding from pedantic statisticians lecturing you about what *not* to do with *p*-values, without offering any real ideas of what *to do* about the very hard problem of separating signal from noise in data and making decisions under uncertainty. Fear not. In this issue, thanks to 43 innovative and thought-provoking papers from forward-looking statisticians, help is on the way.

special issue of *The American Statistician*. Authors were explicitly instructed to develop papers for the variety of audiences interested in these topics. If you use statistics in research, business, or policymaking but are not a statistician, these articles were indeed written with YOU in mind. And if you are a statistician, there is still much here for you as well.

The papers in this issue propose many new ideas, ideas that in our determination as editors merited publication to enable

Nothing wrong with these concepts, (uniformed) people just misuse them

Take-home messages

Scientific questions

- Be inspired
- Be reasonable

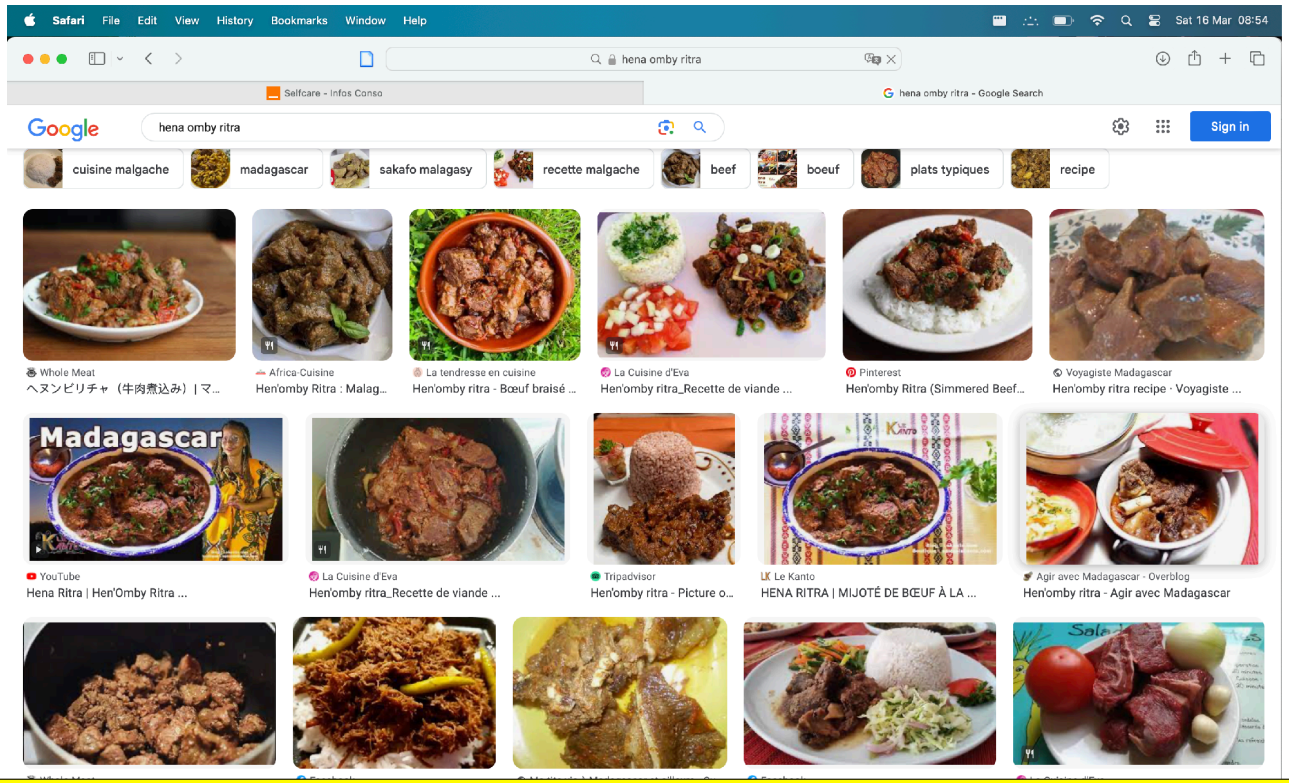
Hypothesis and data

- Be precise especially about the control
- Be thoughtful

Models

- Be knowledgable

How to choose a model
There is no perfect rule: be knowledgable



Ask a person how they make hena ritra and each one will give you different recipe


CENTRE