AI For The Environment, week 9

From AI to ecological models

Ella Browning and Rory Gibb

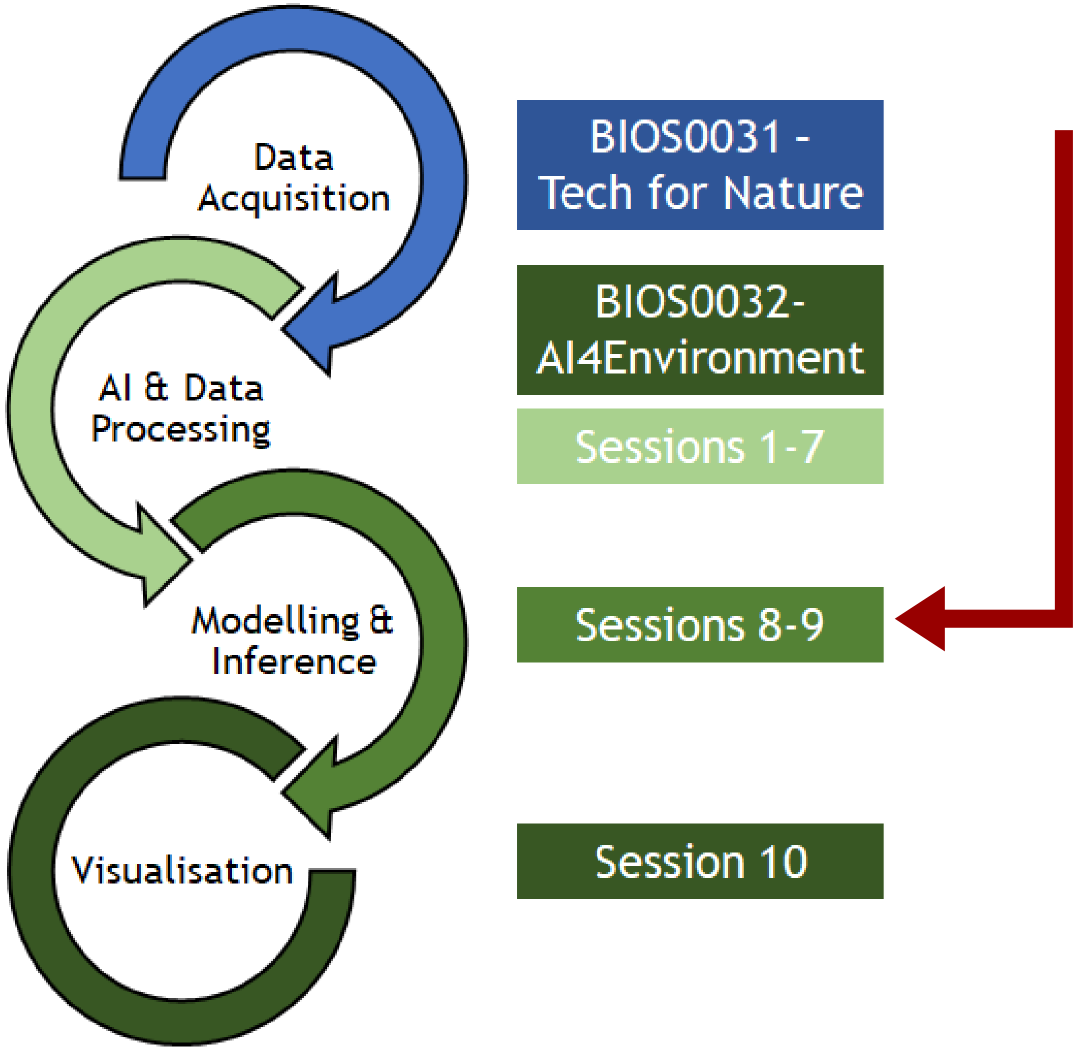
People & Nature Lab / Centre for Biodiversity and Environment Research, UCL

Introductions

Dr Rory Gibb - UCL

Dr Ella Browning - UCL

Today's session



Session 9:

- Linking research questions, to hypotheses, to data, to models.
- Environmental data and GIS in R.
- Principles of statistical modelling in ecology.
- Widely used methods and techniques.

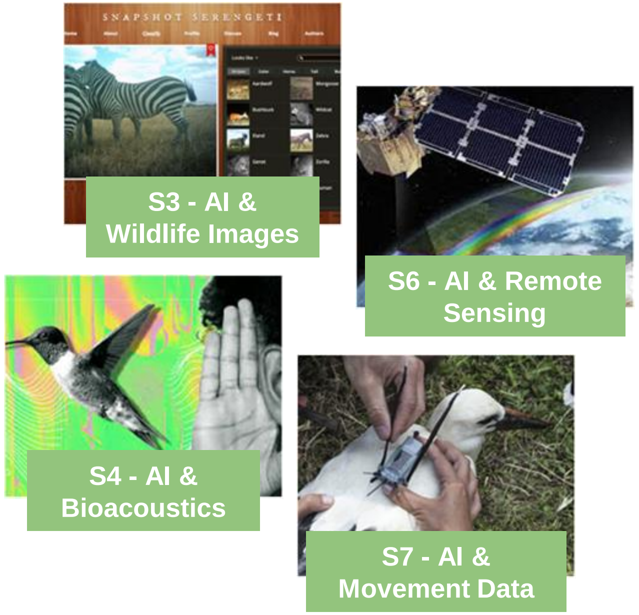
Session 10:

- Explanatory (causal) and predictive modelling.
- Confounding and sampling bias.
- Autocorrelation in space, time and phylogeny.
- Spatial and temporal models.

Today's session

Research question

Data collection & processing (sensors and AI)



Data in space and time

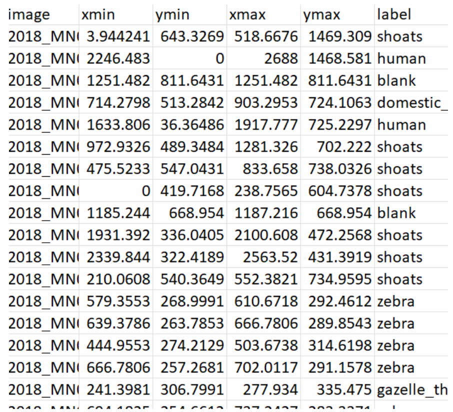
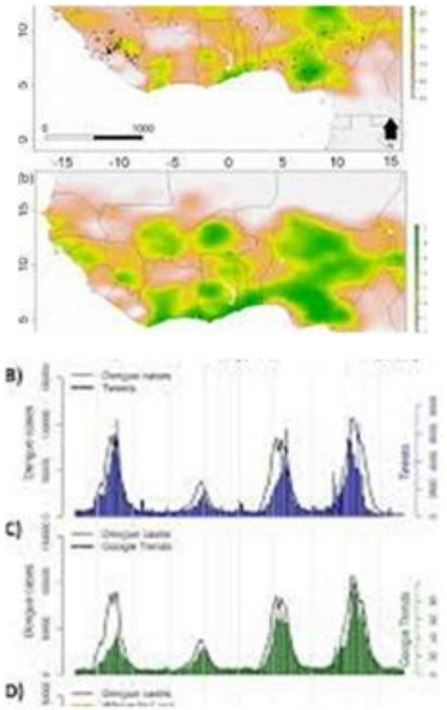


image	xmin	ymin	xmax	ymax	label
2018_MN	3.944241	643.3269	518.6676	1469.309	shoats
2018_MN	2246.483	0	2688	1468.581	human
2018_MN	1251.482	811.6431	1251.482	811.6431	blank
2018_MN	714.2798	513.2842	903.2953	724.1063	domestic_
2018_MN	1633.806	36.36486	1917.777	725.2297	human
2018_MN	972.9326	489.3484	1281.326	702.222	shoats
2018_MN	475.5233	547.0431	833.658	738.0326	shoats
2018_MN	0	419.7168	238.7565	604.7378	shoats
2018_MN	1185.244	668.954	1187.216	668.954	blank
2018_MN	1931.392	336.0405	2100.608	472.2568	shoats
2018_MN	2339.844	322.4189	2563.52	431.3919	shoats
2018_MN	210.0608	540.3649	552.3821	734.9595	shoats
2018_MN	579.3553	268.9991	610.6718	292.4612	zebra
2018_MN	639.3786	263.7853	666.7806	289.8543	zebra
2018_MN	444.9553	274.2129	503.6738	314.6198	zebra
2018_MN	666.7806	257.2681	702.0117	291.1578	zebra
2018_MN	241.3981	306.7991	277.934	335.475	gazelle_th

Ecological inference and prediction



Learning objectives for week 9

By the end of today you will…

- Understand the importance of clearly defined research questions and hypotheses for answering ecological questions.
- Understand the fundamentals of spatial data processing and GIS in R.
- Understand key principles of statistical modelling in ecology.
- Be familiar with common data and model types in ecological research, as well as important sources of bias impacting ecological data.

Structure of the day

Lecture 1 AM ~45min

- Defining a research question and hypotheses
- Combining ecological with spatial and temporal environmental data

Workshop 1 AM ~1hr

- Introduction to geographical and spatial data processing and analysis in R

Lecture 2 PM ~ 45min

- Linking ecological questions to models
- Widely-used modelling tools for ecological inference and prediction

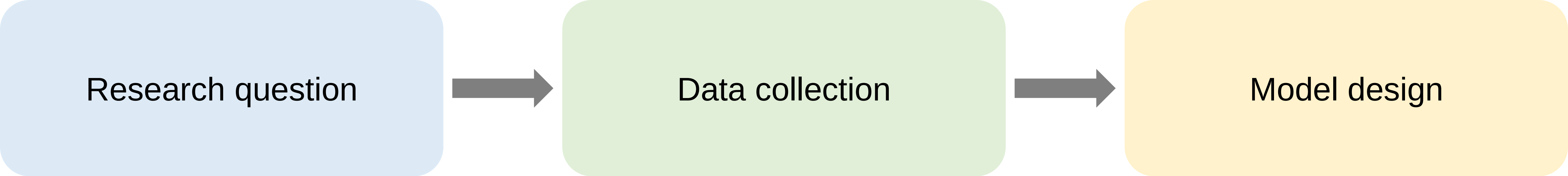
Workshop 2 PM ~2hr

- Combining camera trap and spatial environmental data to understand species occurrence
- Generalised linear and mixed effects models in R

Morning lecture

From research question to ecological and environmental data

Basic Principles for Ecological Modelling



Start with a question

Research question

- A good research question should be **F**easible, **I**nteresting, **N**ovel, **E**thical and **R**elevant (the FINER criteria) Hulley et al. (2007). Designing clinical research. Lippincott Williams & Wilkins
- If appropriate, define hypotheses ***a priori***
  - fundamental to robust experimental study design and causal inference.

ACADEMIC PRACTICE IN ECOLOGY AND EVOLUTION

Ecology and Evolution Open Access WILEY

When are hypotheses useful in ecology and evolution?

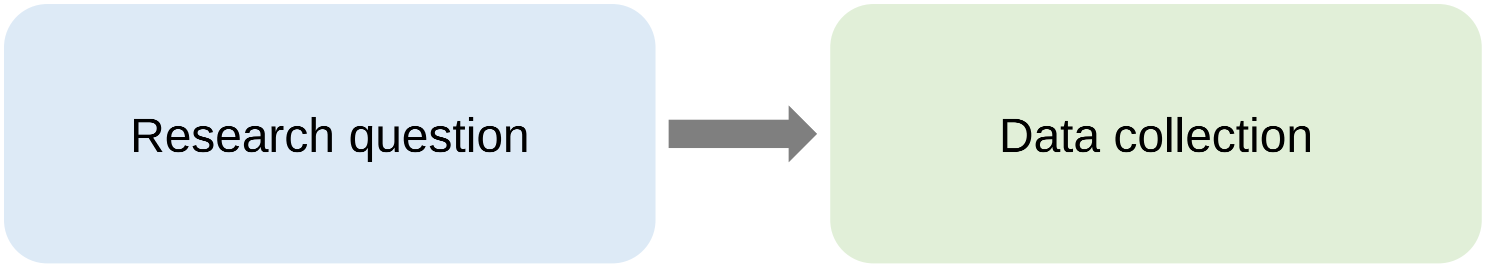
Matthew G. Betts[1] | Adam S. Hadley[1] | David W. Frey[1] | Sarah J. K. Frey[1] | Dusty Gannon[1] | Scott H. Harris[1] | Hankyu Kim[1] | Urs G. Kormann[1] | Kara Leimberger[1] | Katie Moriarty[2] | Joseph M. Northrup[1,3] | Ben Phalan[1] | Josée S. Rousseau[1] | Thomas D. Stokely[1] | Jonathon J. Valente[1] | Chris Wolf[1] | Diego Zárrate-Charry[1]

Pers Soc Psychol Rev. 1998;2(3):196-217. doi: 10.1207/s15327957pspr0203_4.

HARKing: hypothesizing after the results are known

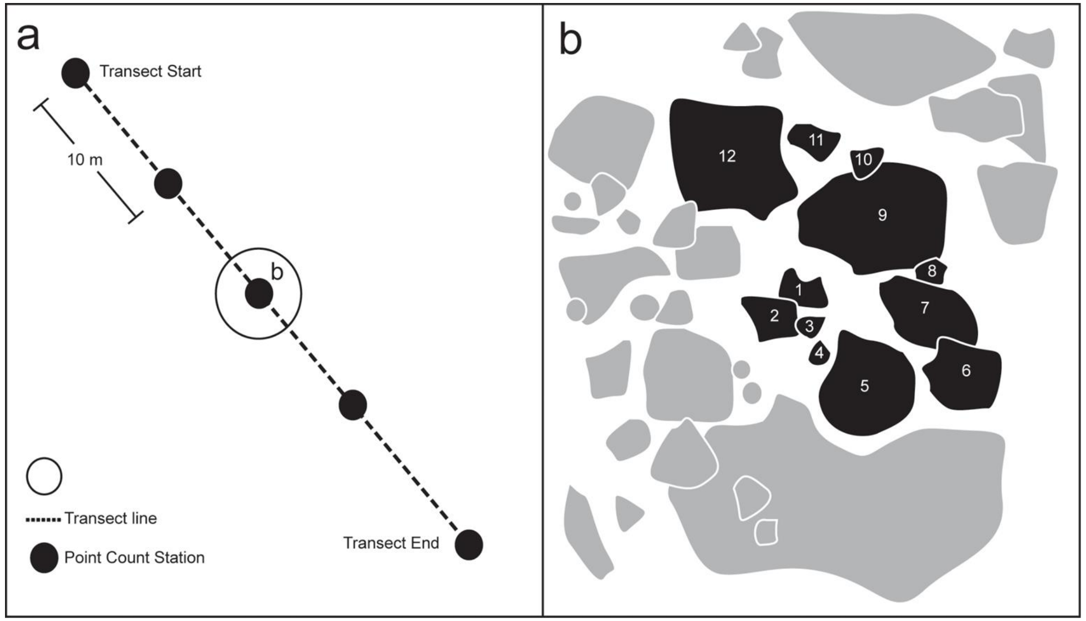
N L Kerr [1]

Data collection



Types of ecological data: Structured data

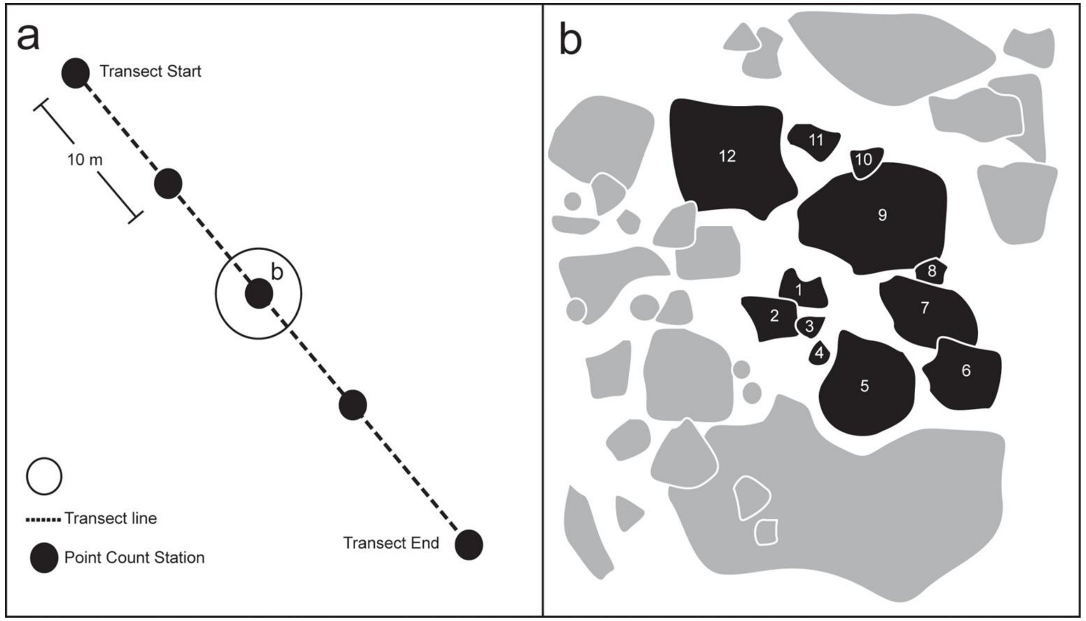
- Collected using a well defined, randomised, stratified sampling strategy
- Representative sample to reflect the patterns & characteristics of the target population/ community
- Repeatable data collection methods
- Unbiased (?)



Roberts et al. 2016 PLOS ONE

Types of ecological data: Structured data

- Collected using a well defined, randomised, stratified sampling strategy
- Representative sample to reflect the patterns & characteristics of the target population/ community
- Repeatable data collection methods
- Unbiased (?)
- Smaller quantities of these data available
- Expensive - resources & infrastructure



Roberts et al. 2016 PLOS ONE

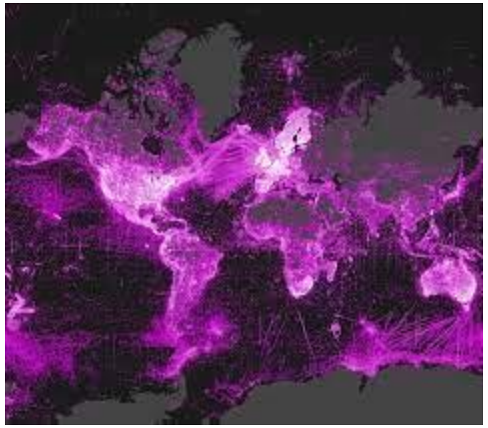
Types of ecological data: Unstructured data

- Non-representative sampling
- Often much larger quantities available
- Often combine observations or data from numerous sources
    - (many different data generating processes in the same dataset!)
- Typical of many citizen science datasets (some exceptions)
- Subject to many biases (taxonomic, geographical, detectability)
    - modelling can <u>attempt</u> to account for these

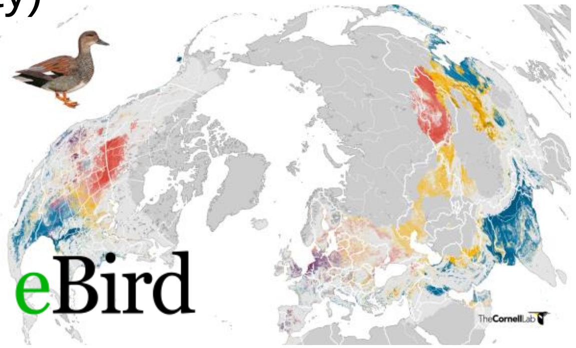
The Global Virome, in One Network

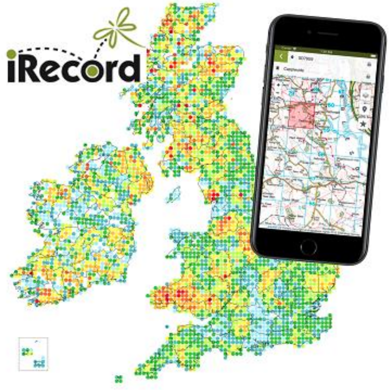
GBIF

NBN National Biodiversity Network



BRC Biological Records Centre





Understanding the data-generating process

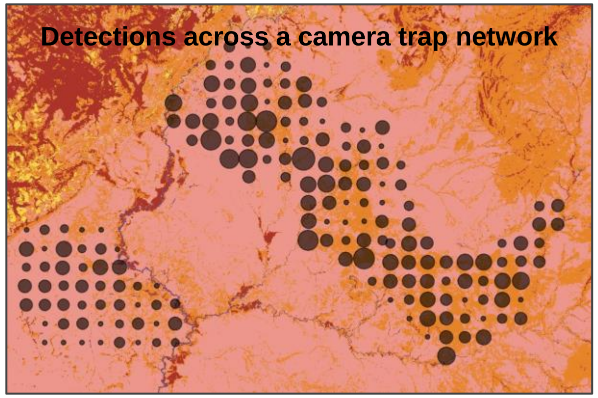
- How the data were generated affects the kind of questions that can be answered, and analysis approach
- Unstructured data are *typically* the lowest common denominator - e.g. presence only data
- Structured data are *typically* richer in their information - e.g. abundance data

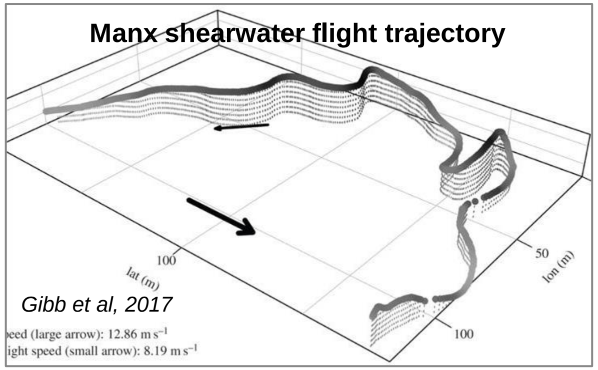
Unstructured data	**Structured**
Ecologically information poor - e.g. presence only	Ecologically information rich - e.g. abundance
Unrecorded sampling effort	Sampling effort well recorded
Non-systematic sampling	Systematic sampling
Not stratified	Stratified
No controls	Controls included (if appropriate)

From AI to ecological data, across space, time and phylogeny

Species detections, abundance or movement trajectories

(e.g. via camera trap; bioacoustic sensor; GPS logger)



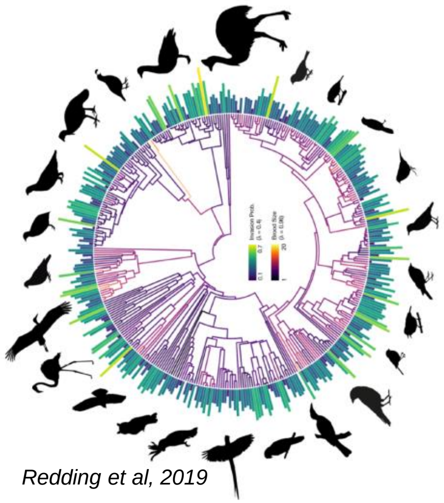


Precise and explicit in space and time

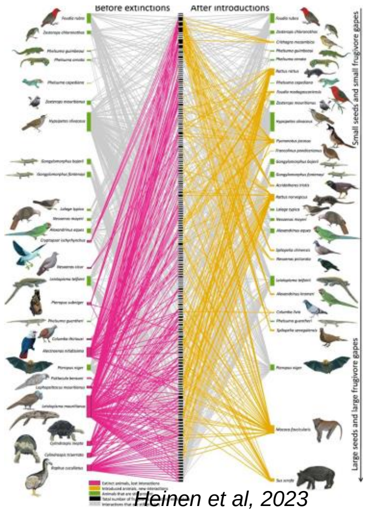
Species or population level traits or interaction networks

(e.g. via text mining of scientific literature or databases; derived from survey data)

Invasive bird traits



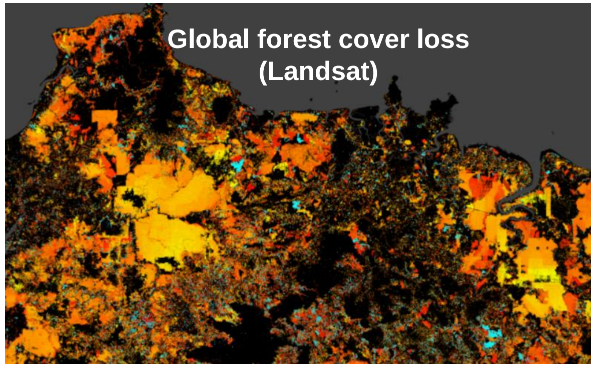
Food web

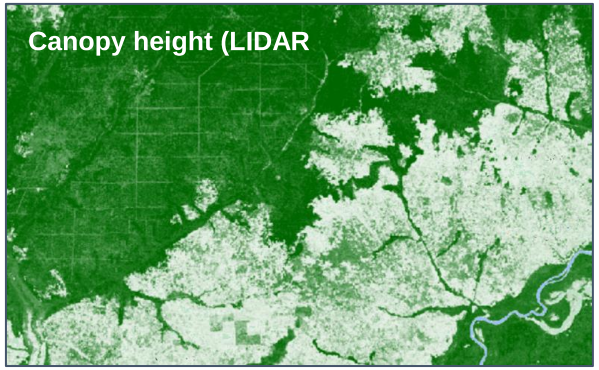


Often coarser or implicit in space and time

Gridded eco-environmental conditions

(e.g. via processing satellite imagery, drone footage or LIDAR)





Precise and explicit in space and/or time

Linking ecological with environmental data in space and time

Linking ecological to environmental data

- What are the environmental and biotic factors that cause or are associated with a particular ecological pattern or process?
- Location in space/time enables the linking of our observations to other sources of environmental or social data.

https://nedc.nz/

Spatial data types

Point

Discrete georeferenced locations of phenomena in space with XY coordinates

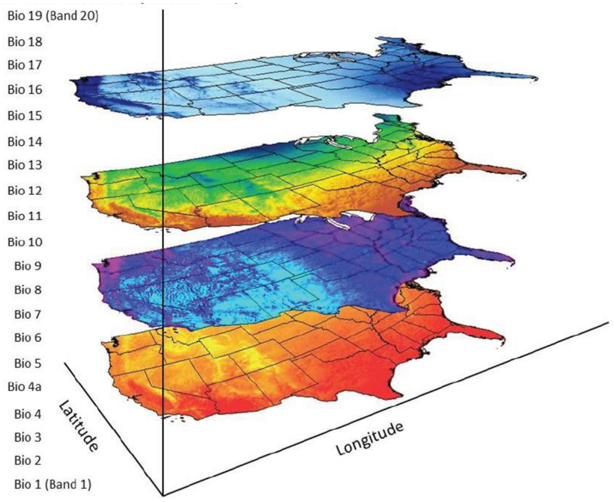
(e.g. species occurrence; sensor location; GPS locations)

Polygon

Two-dimensional bounded areas

(e.g. country or administrative boundaries; protected areas; habitat types/ecoregions)

Raster



https://pubs.usgs.gov/ds/691/ds691.pdf

Grid cells over space with values representing continuous or discrete phenomena

(e.g. land cover types; climatic variables; population density; pesticide input)

Spatial structures in ecological data

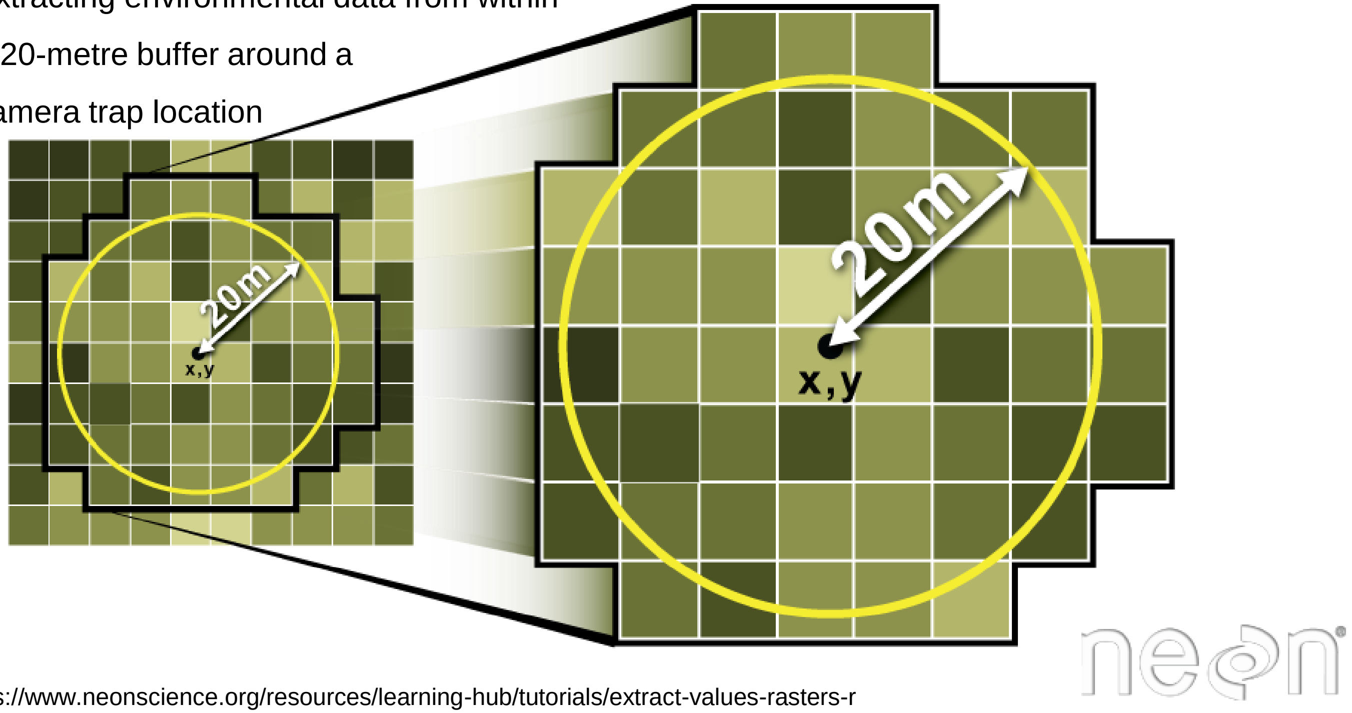
Ecological data are *often* associated in space with point locations or polygon boundaries (because we can't feasibly sample every grid cell of a raster!)

Linking survey observations to raster data

- The location and date/time of ecological observations, can be linked these to raster data about the local socio-environmental conditions.

- A wealth of raster data sources are openly-available on land-cover, vegetation, climate, socioeconomic factors, human pressure…

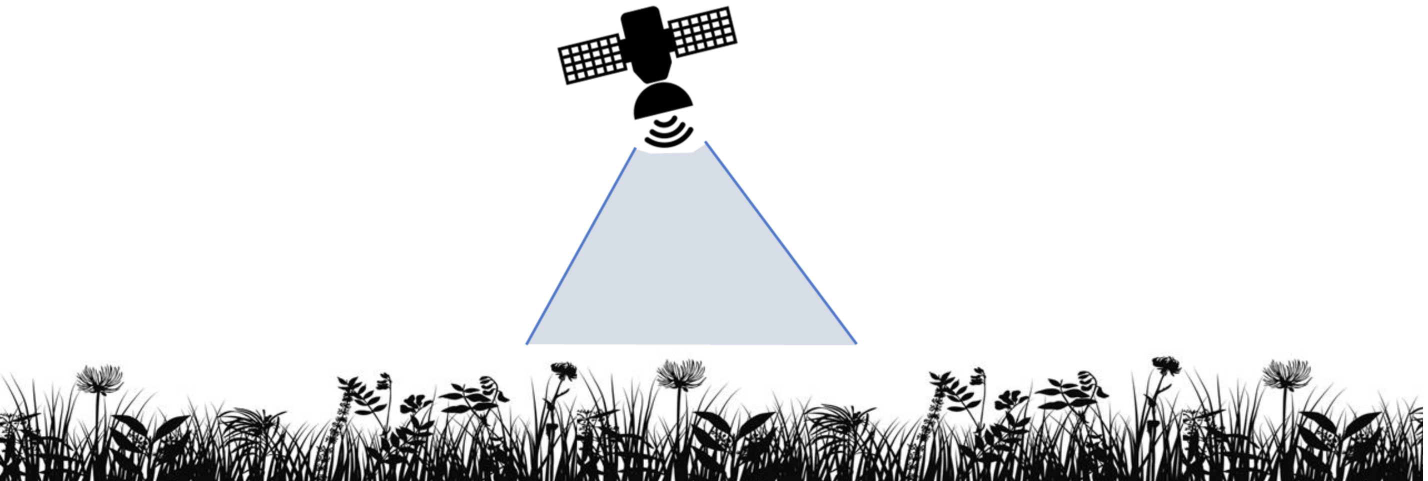
Extracting environmental data from within a 20-metre buffer around a camera trap location



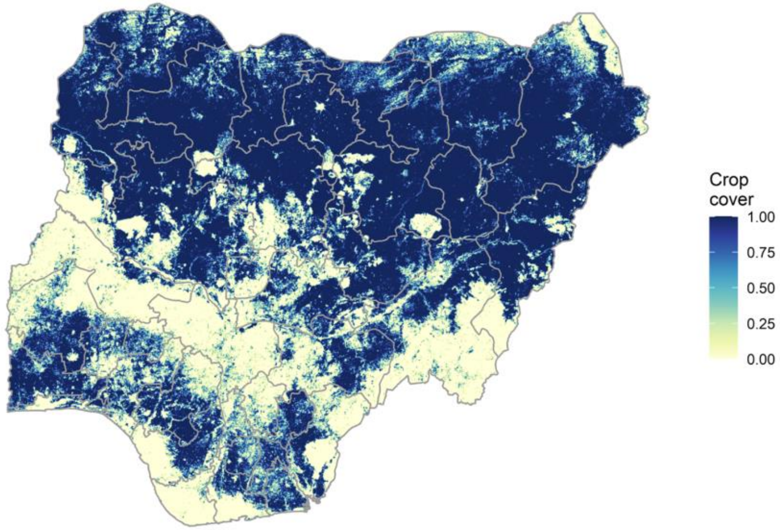
https://www.neonscience.org/resources/learning-hub/tutorials/extract-values-rasters-r

Land cover and land use

- Gridded data products describing aspects of land cover, derived from satellite imagery
- Quality affected by classification uncertainty or environmental factors such as cloud cover
- Discrete (e.g. habitat type) or continuous (e.g. NDVI, proportion vegetation cover)
- Vary in spatial (from 30m to >10km) and temporal resolutions (from weekly to decadal)

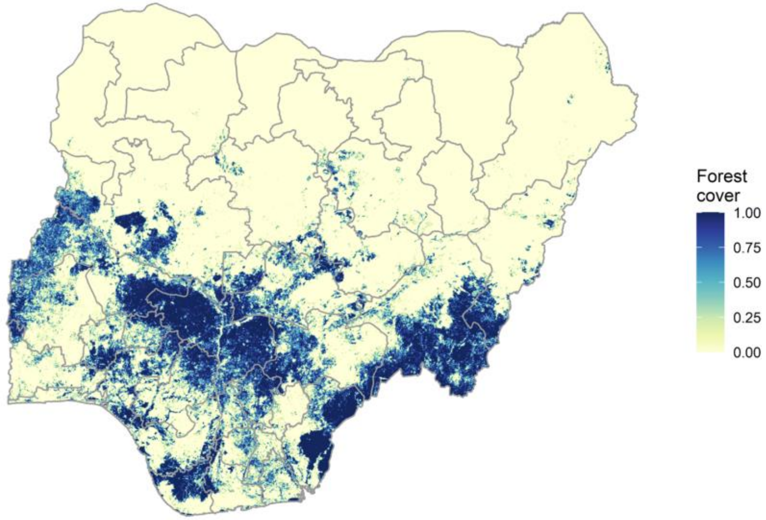


Land cover and land use



Nigeria cropland 2018

(ESA-CCI)

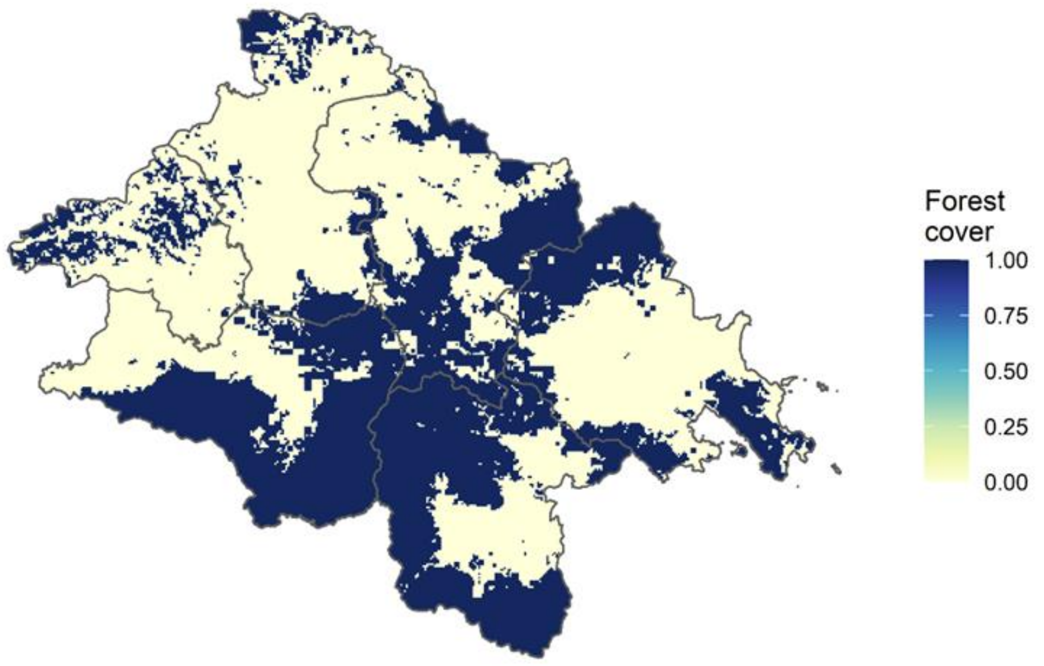


Nigeria forest 2018

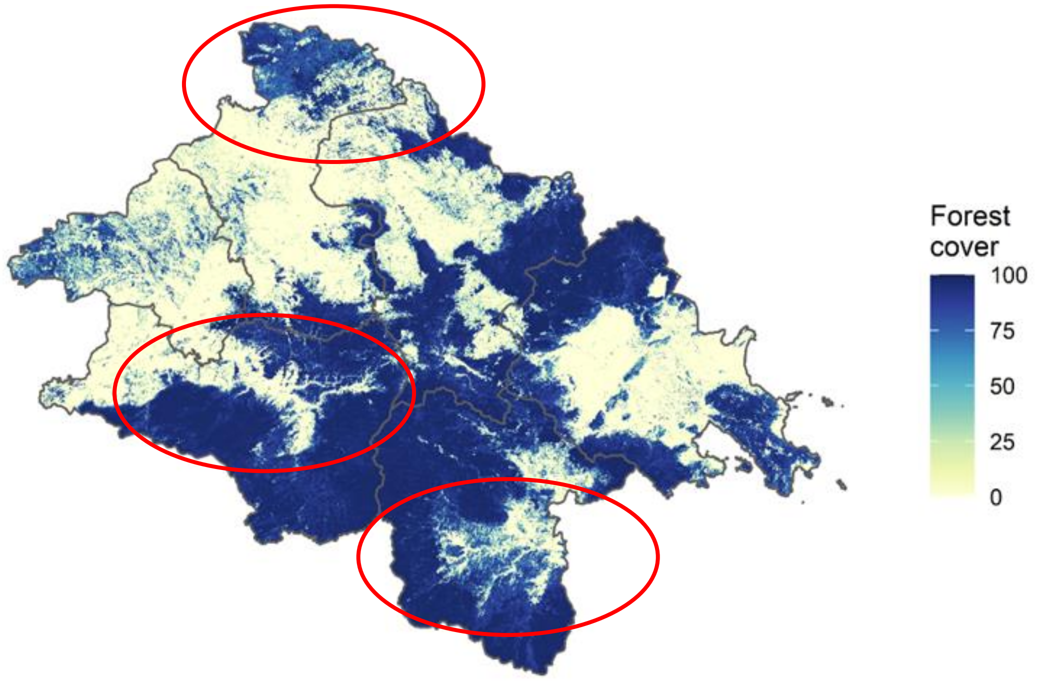
(ESA-CCI)

Why resolution matters: two views of the same area

ESA-CCI (300-metre grid cells)



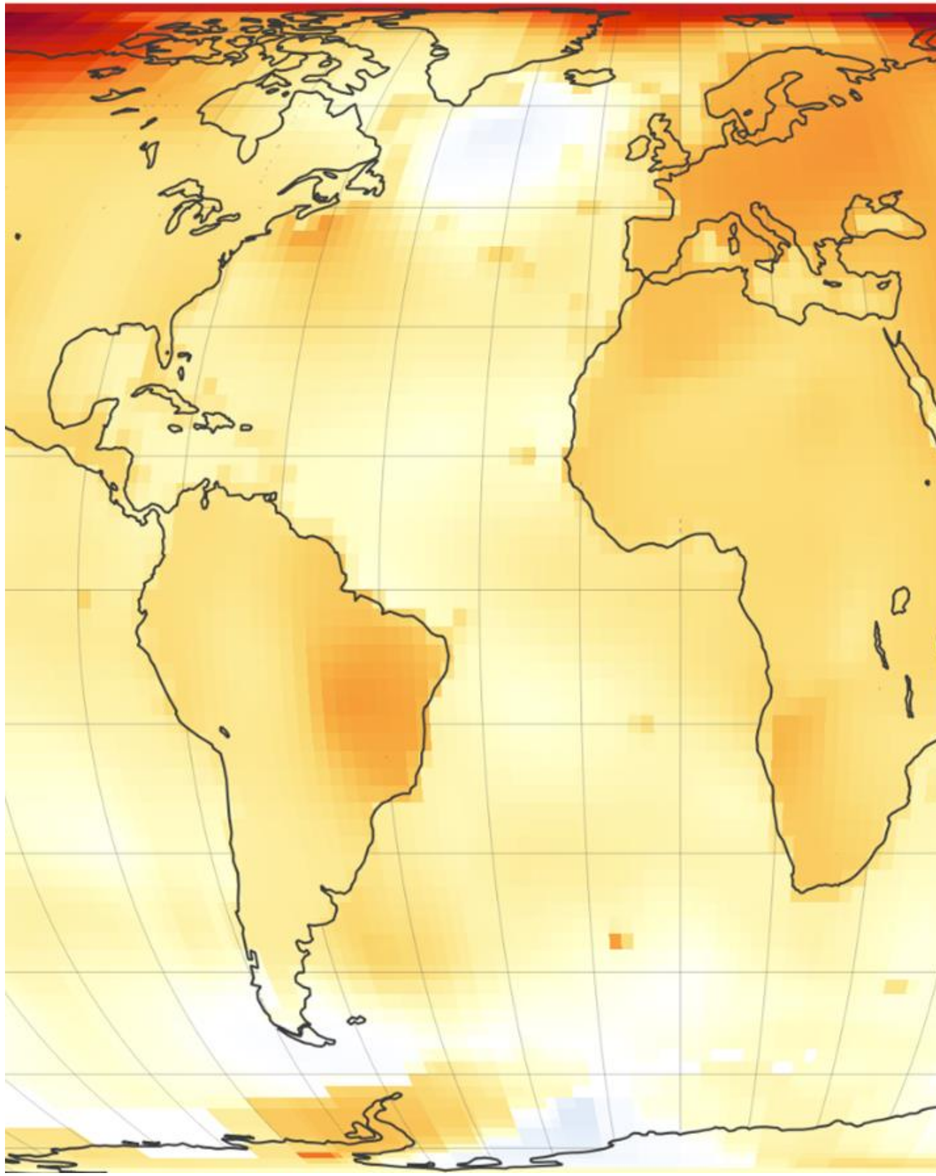
Global Forest Change (30-metre grid cells)



The lower spatial resolution and discrete land cover class scheme of ESA-CCI data classifies many areas with substantial tree cover as “non forest”

Climate data

- Temperature, precipitation, humidity, wind, potential evapotranspiration, drought/flood indicators, El Nino…
- Diverse sources: weather station observations, interpolation, downscaling, satellite sensors, climate models, reanalysis…
- Range of temporal resolutions, from **hourly** to **multi-decadal**
  - often feasible to match ecological observations to existing climate conditions



Google Earth Engine

A planetary-scale platform for Earth science data & analysis

Powered by Google's cloud infrastructure

CHELSA

Climatologies at high resolution for the earth's land surface areas

WorldClim

National Oceanic and Atmospheric Administration

U.S. Department of Commerce

EARTHDATA

OPEN ACCESS FOR OPEN SCIENCE

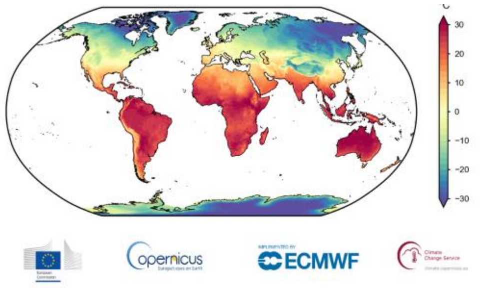
Copernicus

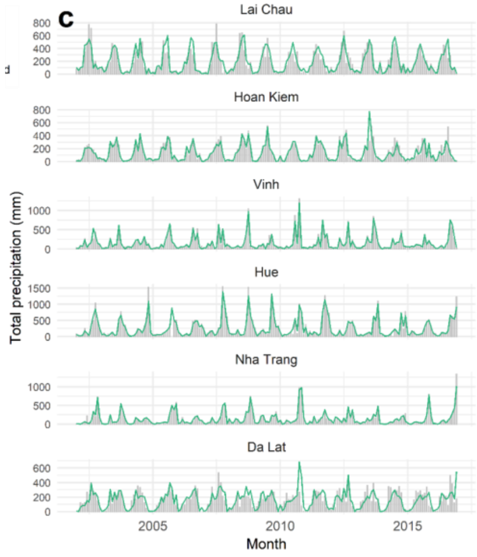
Europe's eyes on Earth

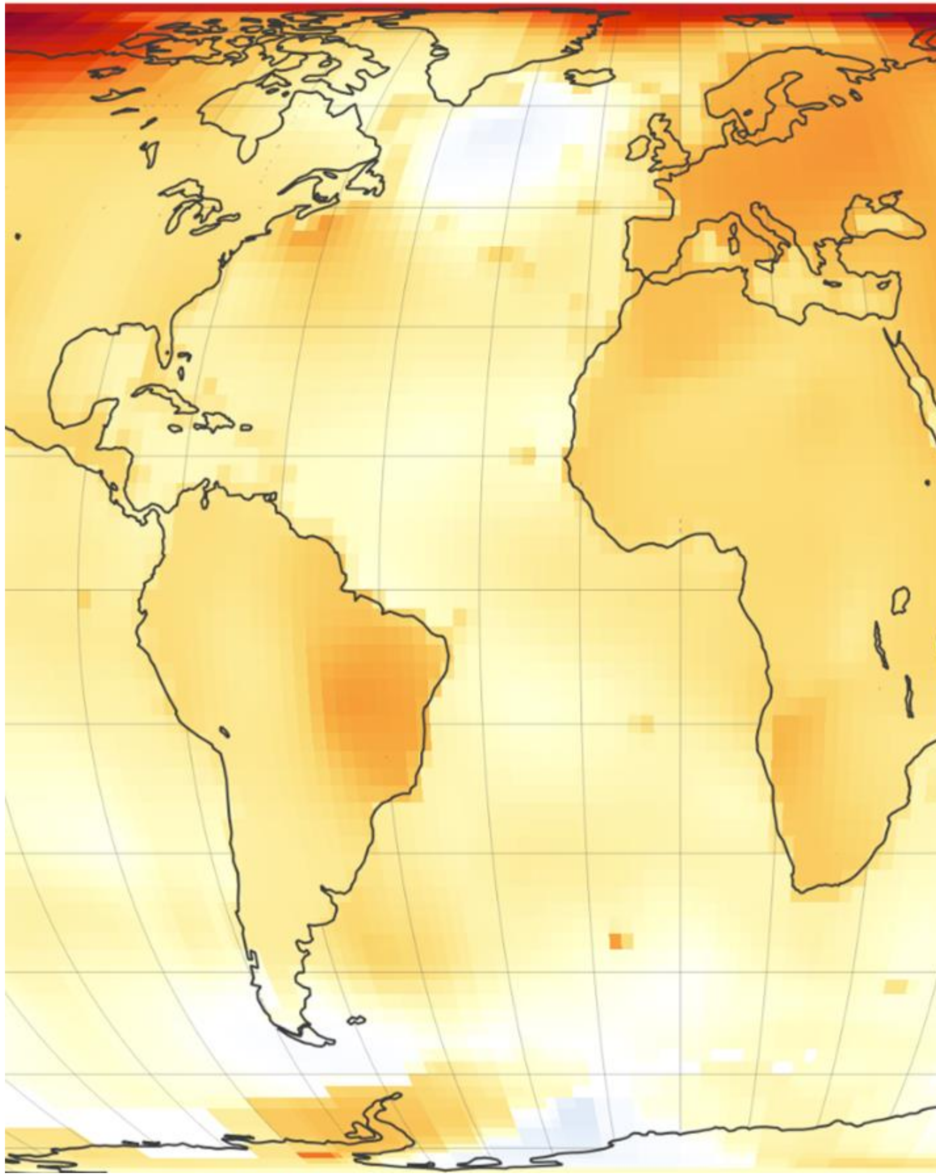
Climate Change Service

Daily or monthly climatic conditions

- **Interpolated weather station** observations (e.g. CRU, CHIRPS)
- **Satellite products** (e.g. MODIS daily rainfall or temperature)
- **Climate reanalysis** (e.g. ERA5) - powerful emerging approaches combining observations with physical models of the climate system



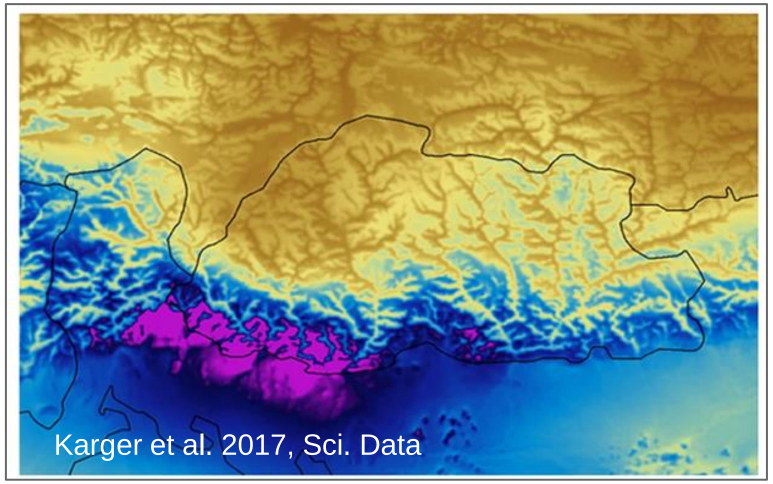




Survey locations + Monthly ERA5 rasters → Time series of precipitation at survey locations

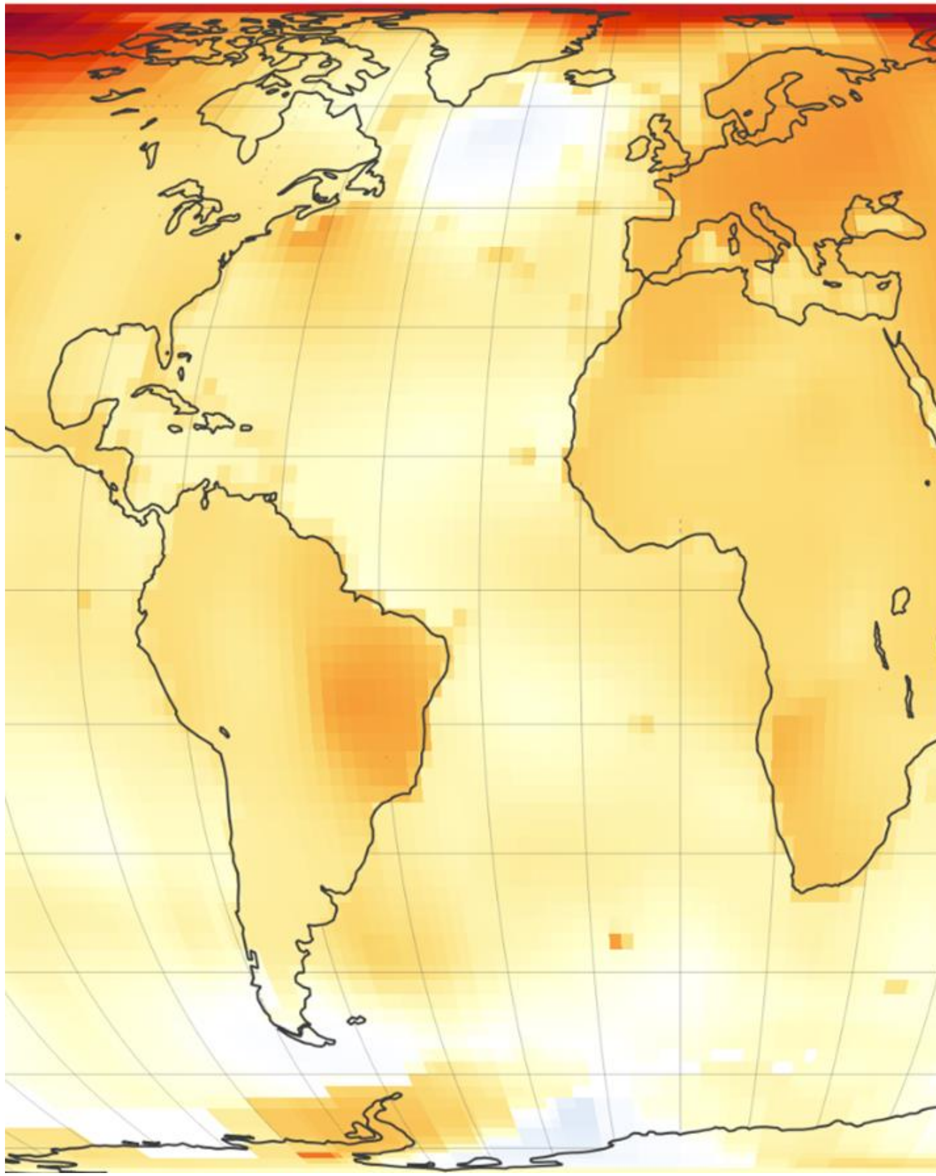
Climatologies

- Averaged climate conditions in each location, typically calculated over multi-decadal periods (e.g. 1970-2010).
- “Bioclimatic variables” describe how average climatic characteristics vary over space
- Often used in species distribution modelling to capture the climatic limits of a species range (a proxy for its climatic tolerances/“niche”)



Mean annual precipitation, Bhutan

(high-resolution downscaled climatology from CHELSA)



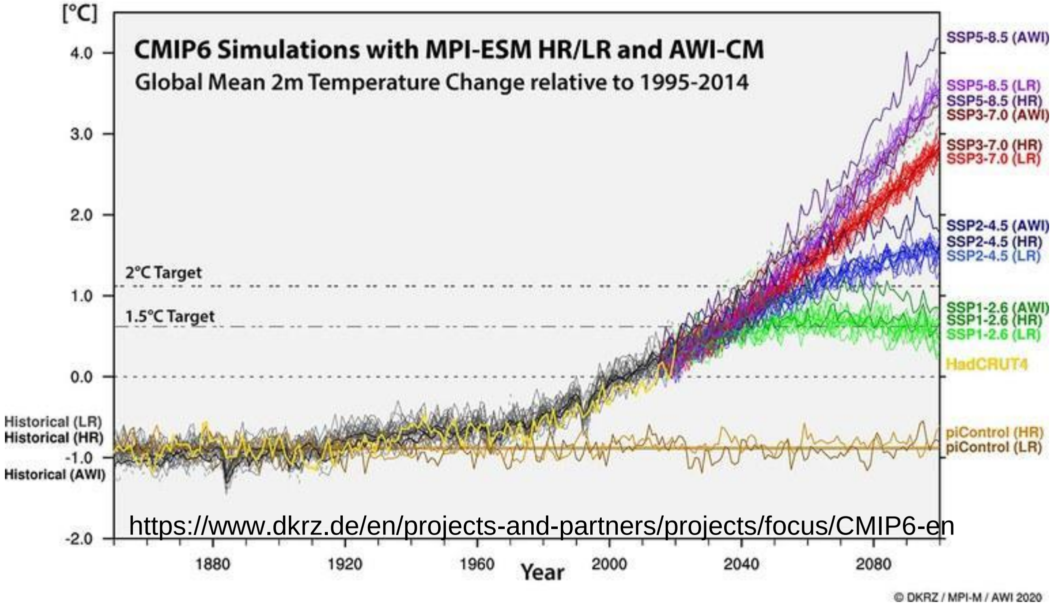
CHELSA
Climatologies at high resolution for the earth's land surface areas

WorldClim

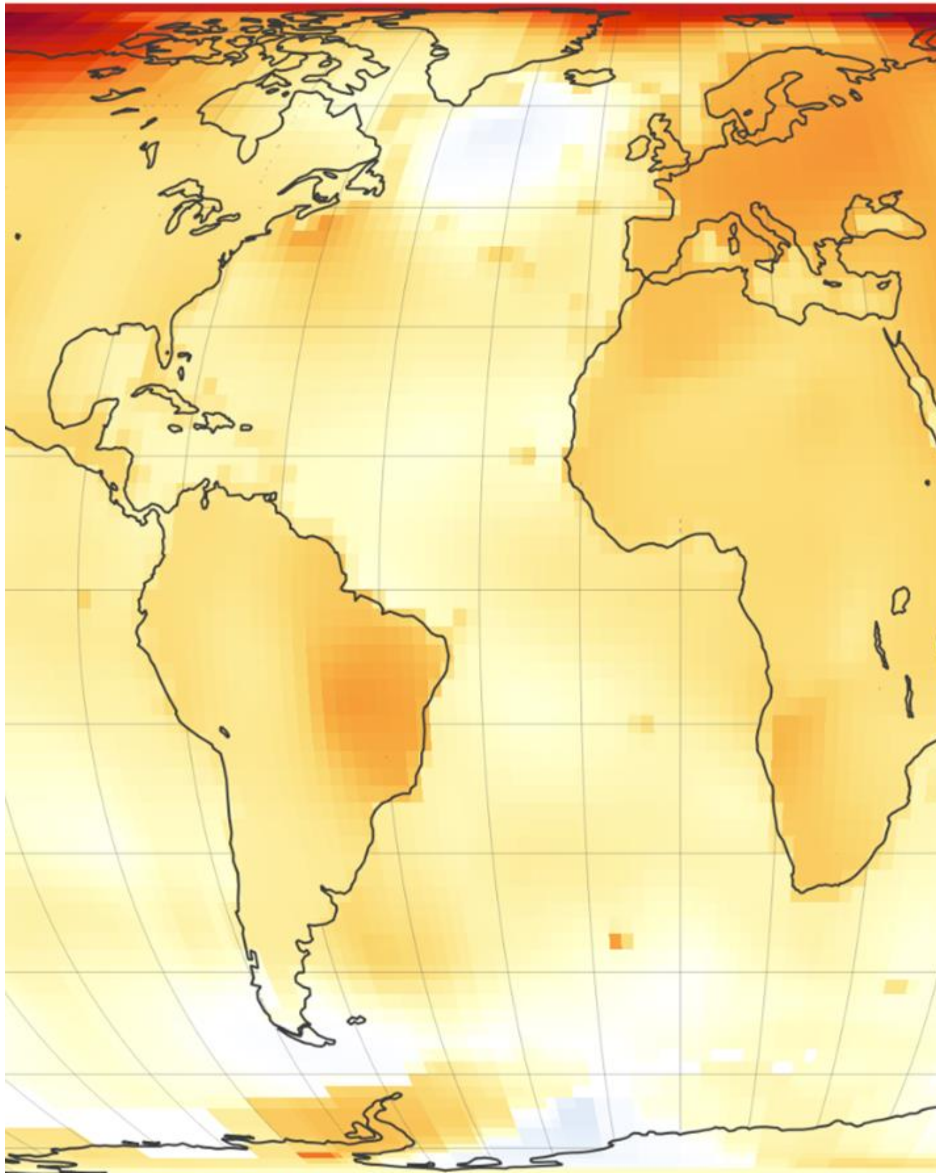
Climate Change Service

Future climates under emissions scenarios

- **General circulation models (GCMs)** - models of the climate system under different future emissions (RCP) and socioeconomic (SSP) scenario pathways.
- Projections form the basis for the IPCC Assessment Reports on climate change.
- Downscaled future climatologies provide data for ecological/biodiversity impact studies.



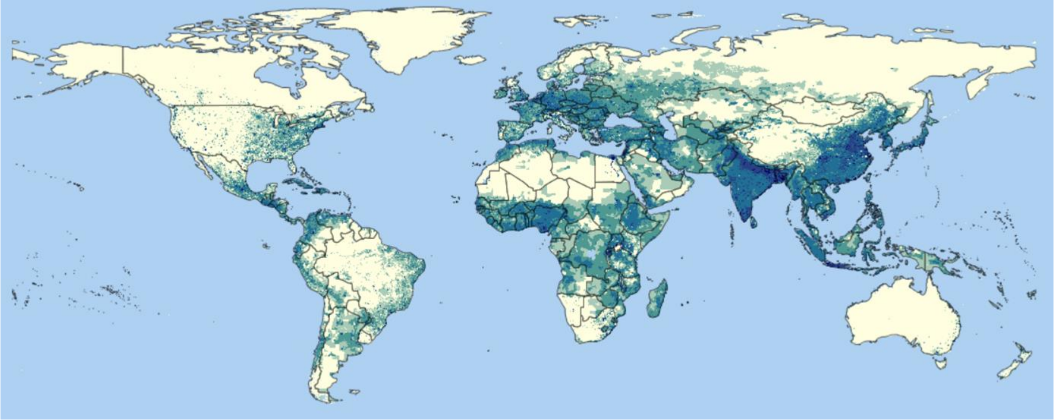
https://www.dkrz.de/en/projects-and-partners/projects/focus/CMIP6-en



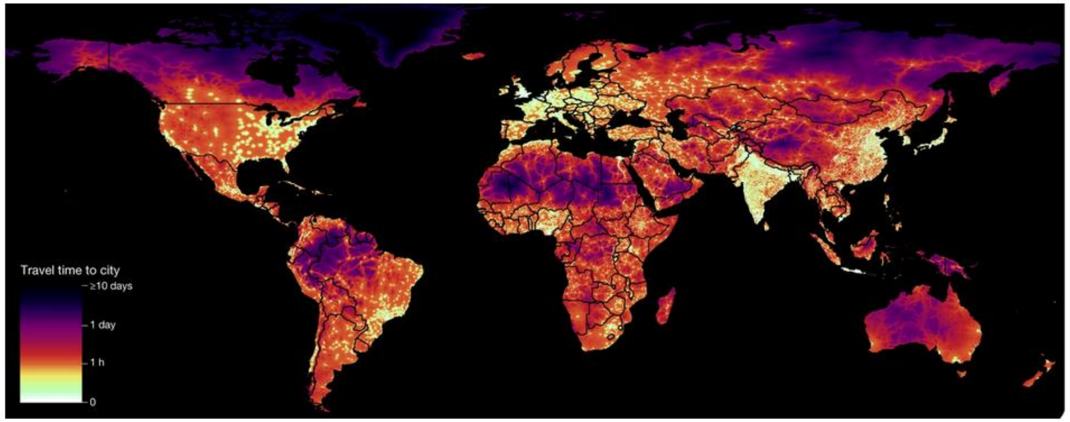
Demographic and socioeconomic data

- A wealth of data on human demographic and socioeconomic indicators over space and time.
- Important to account for the human dimension in (socio-)ecological analyses, as well as other types of analysis (e.g. health and disease)
- Often originally based on census or survey data, but gridded products also developed for many indicators.

Global human population (SEDAC)



Connectivity (travel time to nearest city)



map
malaria atlas project

WorldPop
Open Spatial Demographic Data and Research

Socioeconomic Data and Applications Center (SEDAC)
A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIESIN at Columbia University

Useful resources for spatial & environmental data

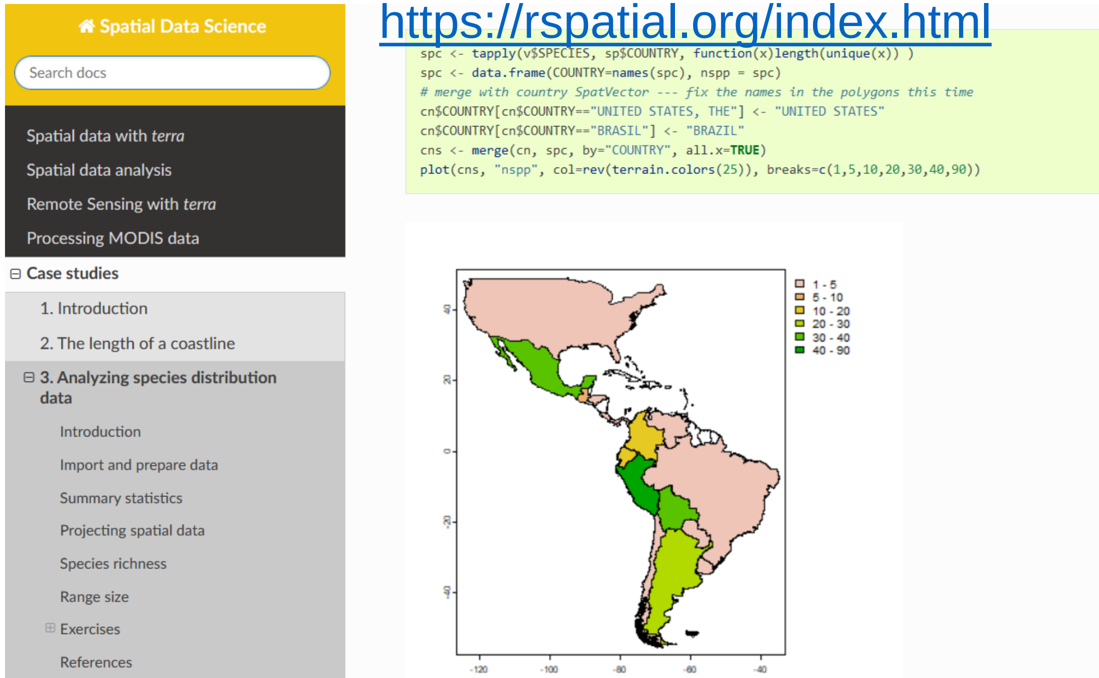
- Coordinate reference systems
  - https://www.spatialreference.org/
- Climatologies/ weather
  - https://chelsa-climate.org/
  - https://www.worldclim.org/
  - https://www.earthdata.nasa.gov/
  - https://www.metoffice.gov.uk/research/climate/maps-and-data/data/index
  - https://openweathermap.org/
- Satellite derived land-cover/ vegetation/ hydrological
  - https://www.earthdata.nasa.gov/
  - https://earth.esa.int/eogateway/catalog
  - https://www.ceh.ac.uk/data
- Demographic/ socioeconomic
  - https://www.worldpop.org/
  - https://sedac.ciesin.columbia.edu/

Spatial data processing and analysis in R

- Ever-growing ecosystem of R packages for reproducible GIS workflows, analysis and visualisation (e.g. *sf, terra, raster, exactextractr),* as well as ecology-specific applications.
- Rspatial.org is a good start for core principles and approaches to geographic data (including fiddly concepts like map projections).

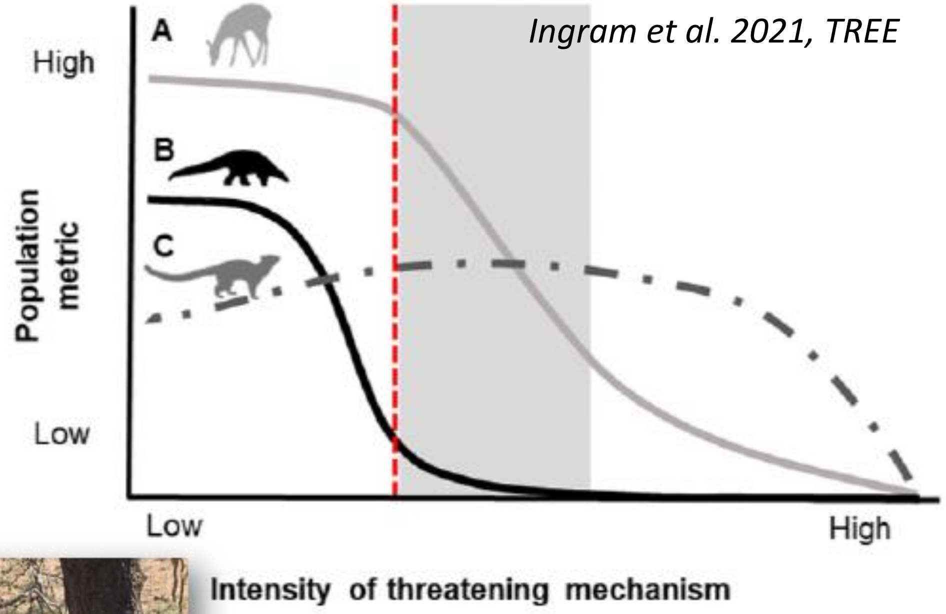
We will be exploring some of these topics in today's workshops.

https://rspatial.org/index.html



Workshops: camera trap data from the Masai Mara

- Camera trap data collected in the Masai Mara, Kenya, in 2018.
- Part of the WWF-funded **Biome Health Project**, focused on understanding biodiversity responses to landscape pressures.
- **Key pressure = livestock grazing.**
- We will analyse a subset of these data to answer questions about spatial drivers of species occurrence.



Ingram et al. 2021, TREE

CBER
Centre for Biodiversity & Environment Research

WWF

Workshops: camera trap data from the Masai Mara

Conservancies

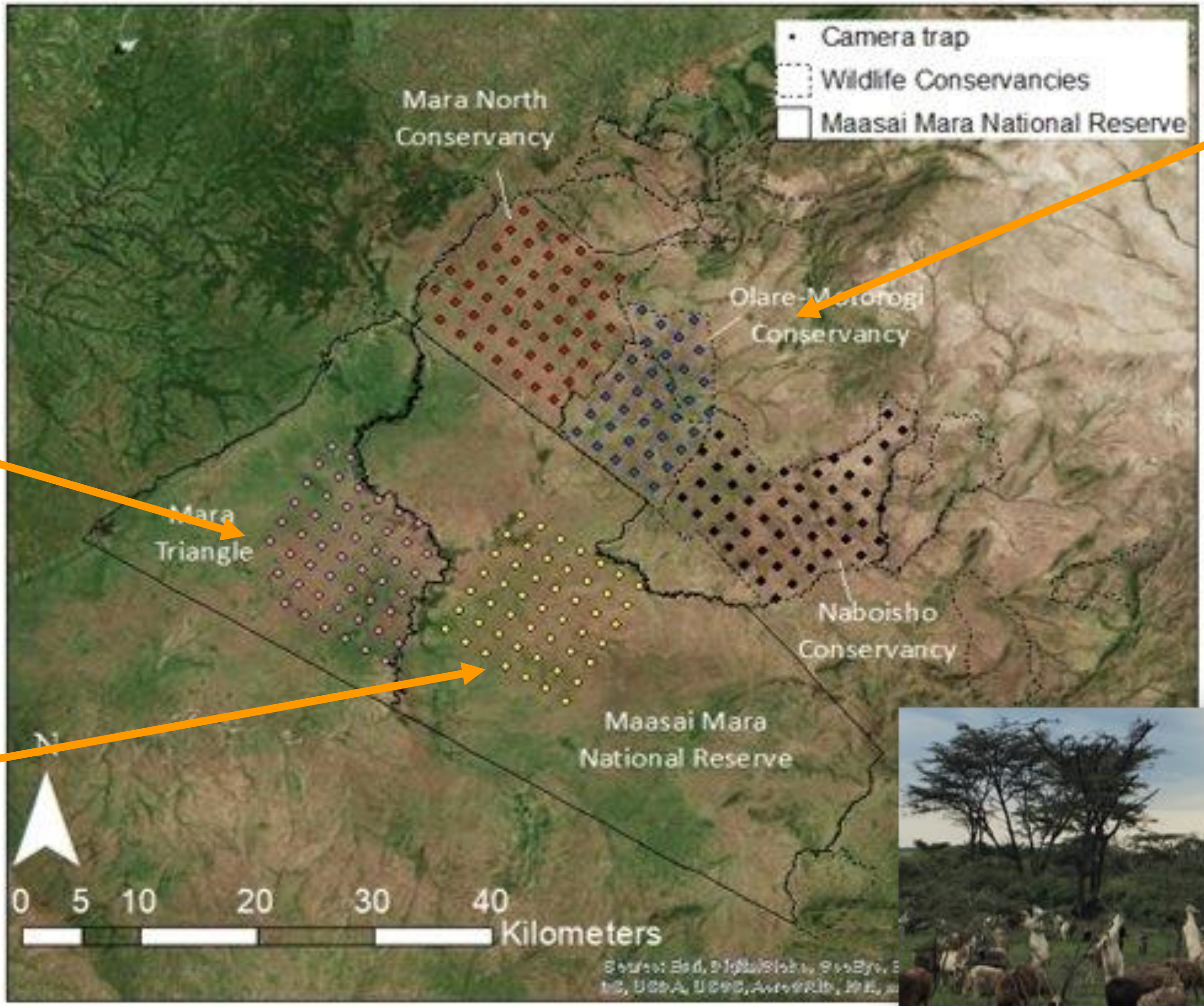
- Different grazing management regimes (e,g. rotational grazing)
- No burning

Mara Triangle

- Most strict
- Periodic grazing
- No burning

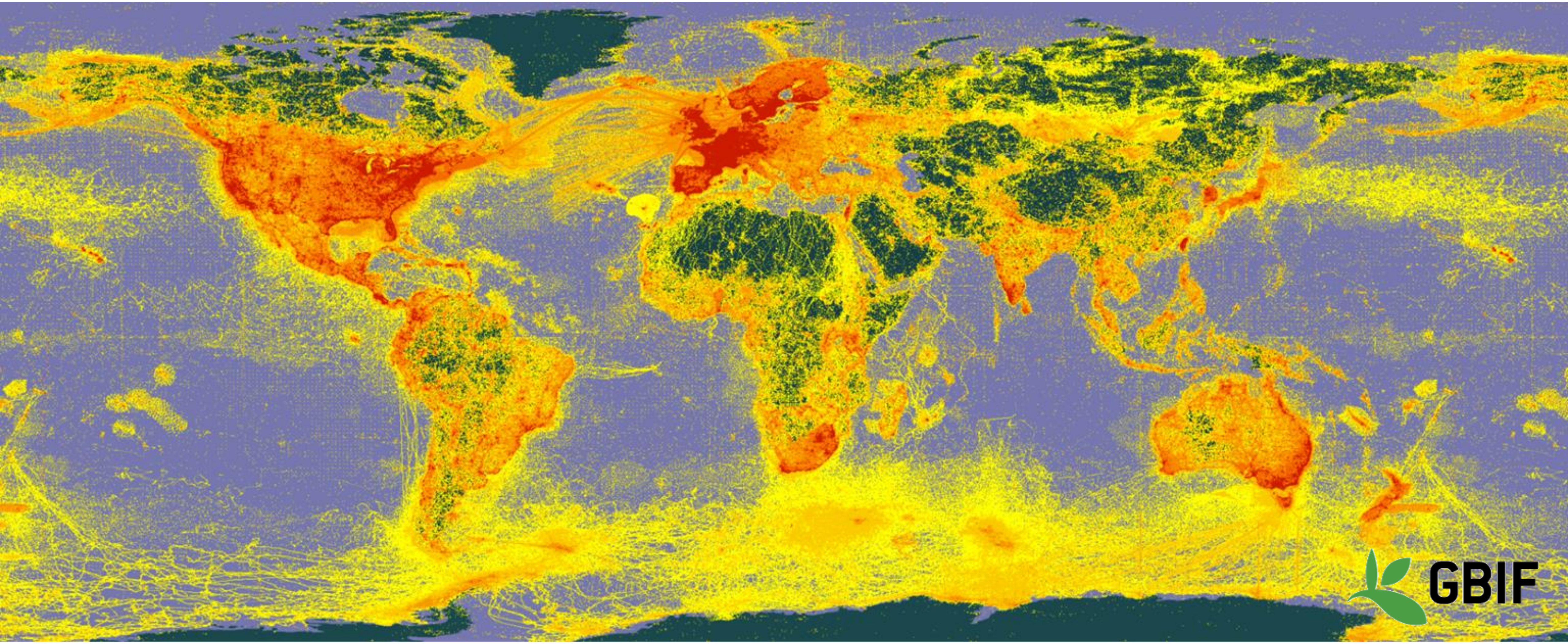
Mara Reserve

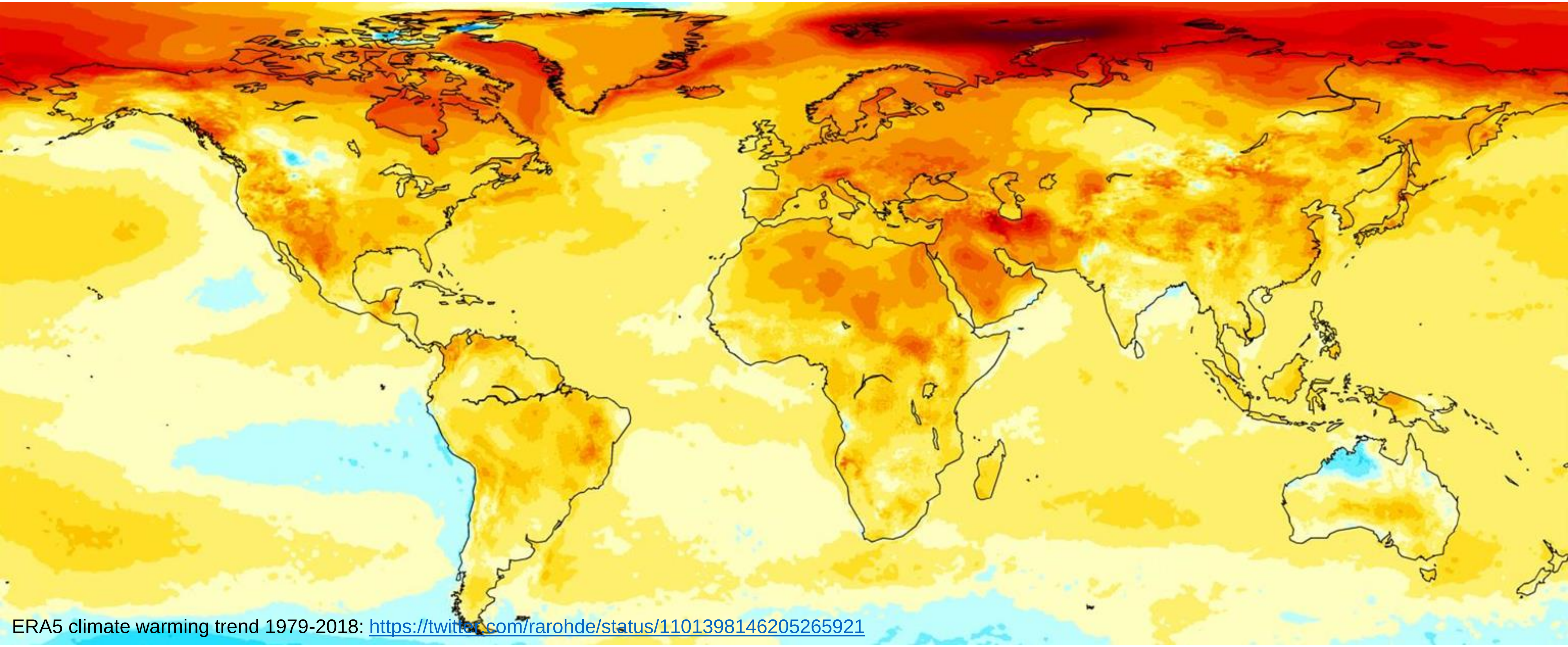
- Allowed grazing until recently; may be illegal grazing
- Recently started burning



Afternoon lecture

Principles of statistical modelling in ecology





ERA5 climate warming trend 1979-2018: https://twitter.com/rarohde/status/1101398146205265921

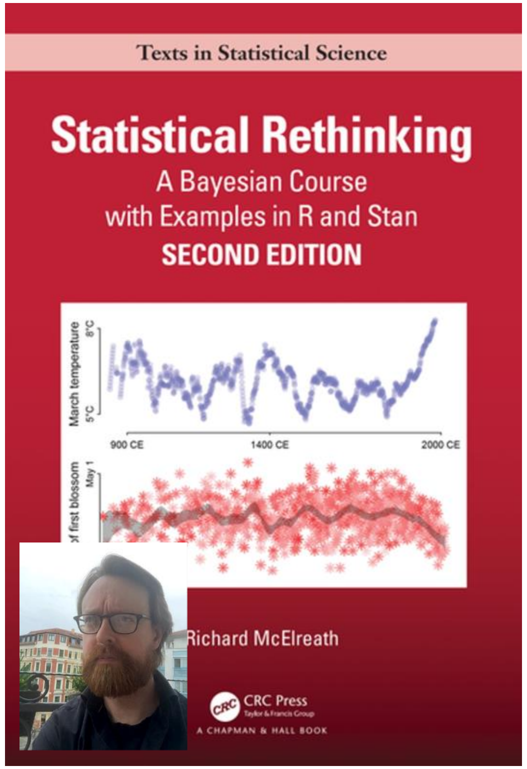
How to *think* about modelling analyses

Research question → Data collection → Model design

- There is no “correct way” to model any data - a huge range of general and ecology-specific modelling tools and softwares are out there (many in R and Python)

- **Focus on the scientific question and on how your data were generated**

Richard McElreath -
Statistical Rethinking lectures
and “Science Before Statistics”



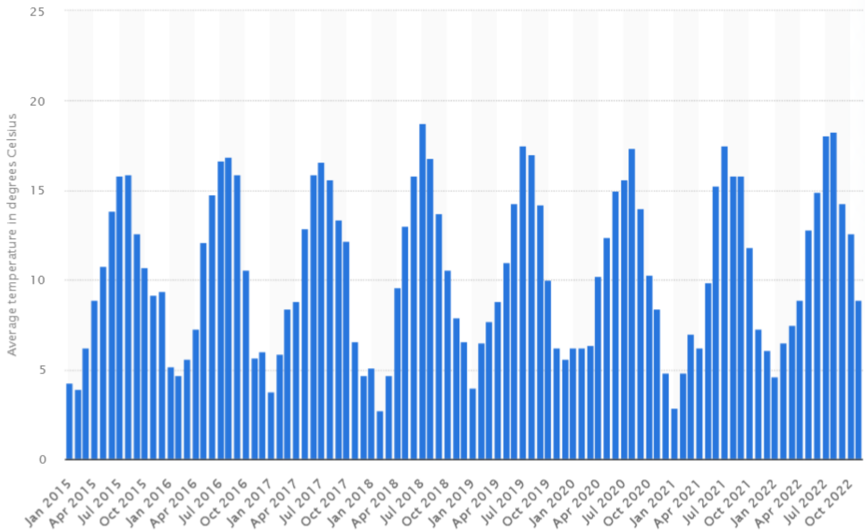
https://xcelab.net/rm/statistical-rethinking/

Three types of question

- **Descriptive (exploratory):** *“What winter weather variables are associated with springtime abundance of the Red Admiral butterfly?”*

- **Causal (hypothesis-led):** *“Do warmer temperatures between November-January lead to higher Red Admiral abundance in the following spring?”*

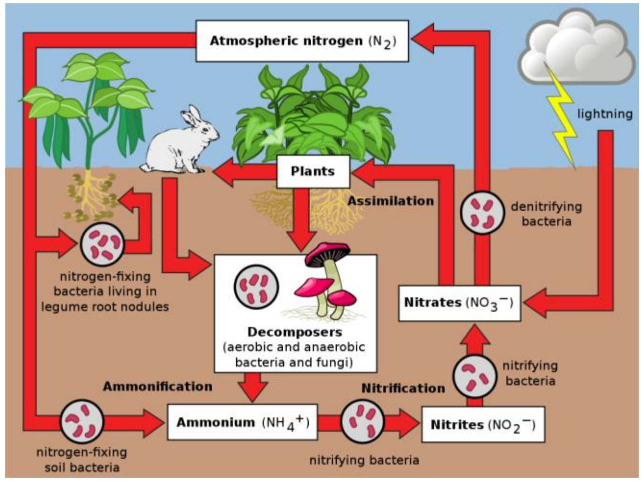
- **Predictive:** *“Can springtime Red Admiral abundance be accurately predicted based on winter weather variables?”*



What is a model?

What is a model?

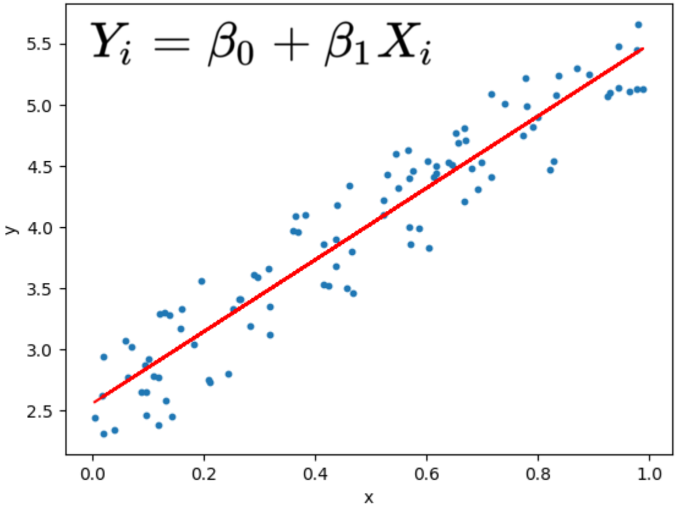
Conceptual



Conceptual representation of known/ hypothesised relationships in a system

Can make qualitative/ quantitative predictions about a system's expected behaviour

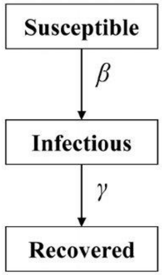
Statistical



Estimation of parameters of interest through fitting to observed data

Examples: linear regression; GL(M)Ms/GAMs

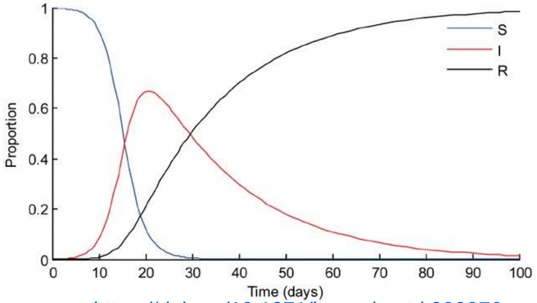
Mathematical (mechanistic)



$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$



https://doi.org/10.1371/journal.pntd.0000761

Mathematical description of a system used to explore and predict system behaviour

Examples: SIR model (disease dynamics); Lotka-Volterra (population dynamics)

What is a model?

Conceptual

Statistical

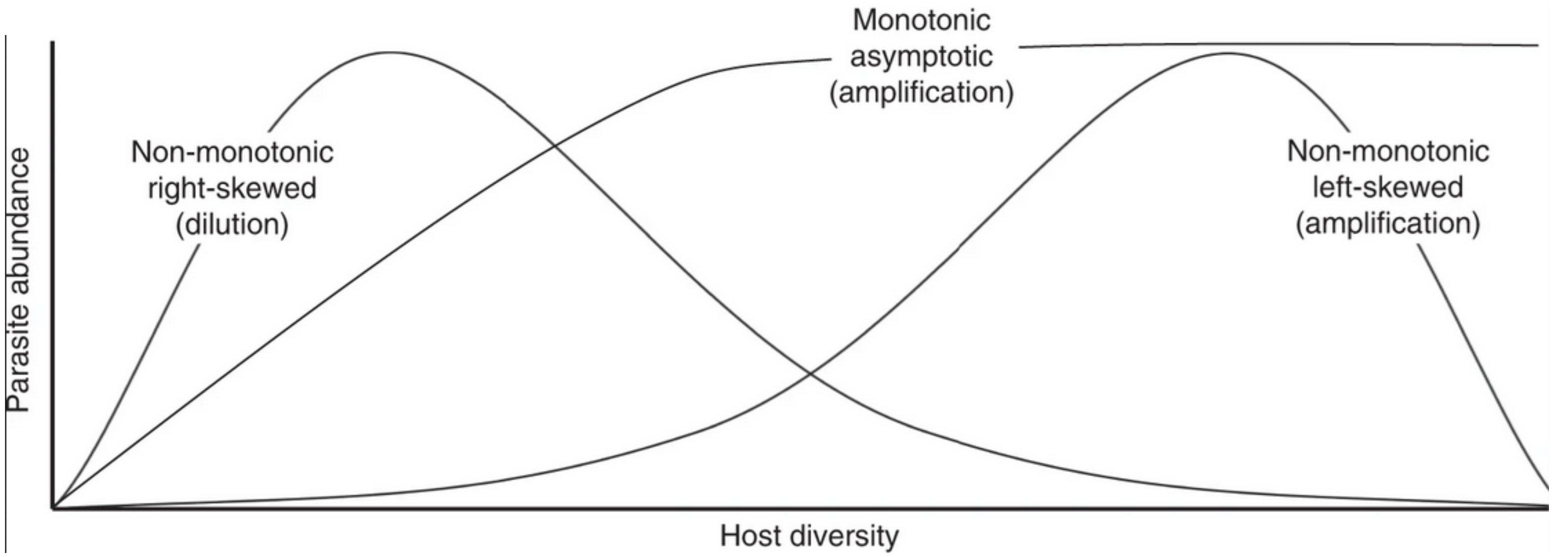
Mathematical (mechanistic)

From conceptual to statistical: the dilution effect model

- “Dilution effect” theory proposes that parasite prevalence increases as local species diversity is lost, vs. “amplification effects” which propose the opposite.
  (Rohr et al. 2020, Nat Eco Evo)

- Either might be possible depending on the system, and results have often been conflicting.

- Conceptual model outlines 3 hypotheses for the nonlinear shape of this relationship - these can be challenged with data!



https://doi.org/10.1038/s41467-019-13049-w OPEN

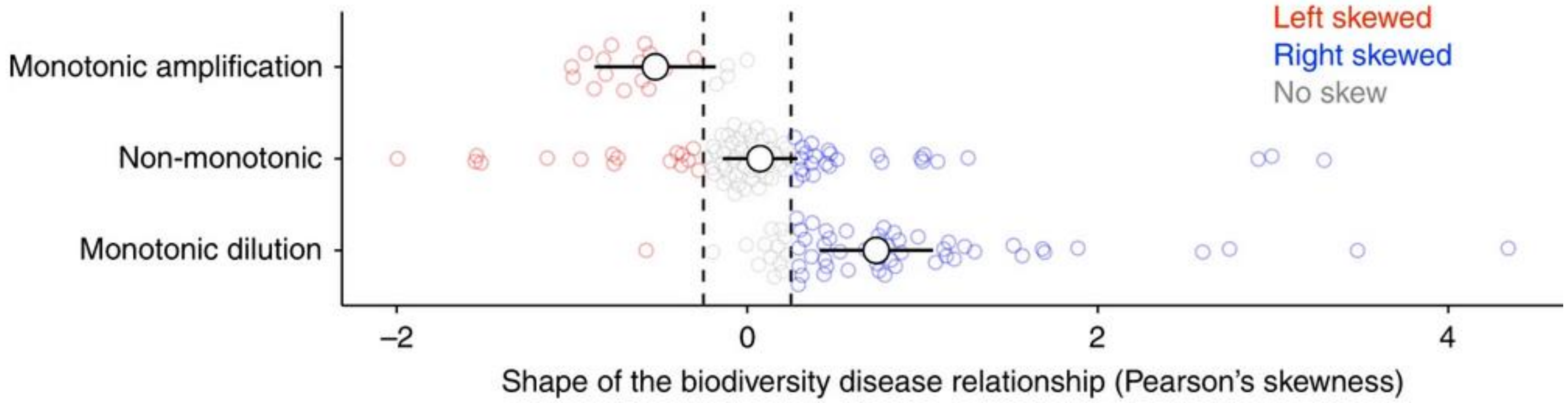
Measuring the shape of the biodiversity-disease relationship across systems reveals new findings and key gaps

Fletcher W. Halliday[1*] & Jason R. Rohr[2]

nature COMMUNICATIONS

From conceptual to statistical: the dilution effect model

- Meta-analysis of the shape of the diversity-disease relationship across 205 studies - do real-world observations follow the predictions of the conceptual model?
- Systems with dilution effects are significantly right skewed, and vice versa, as predicted by the conceptual model - the analysis estimates *parameters* to describe this shape.



https://doi.org/10.1038/s41467-019-13049-w OPEN

Measuring the shape of the biodiversity-disease relationship across systems reveals new findings and key gaps

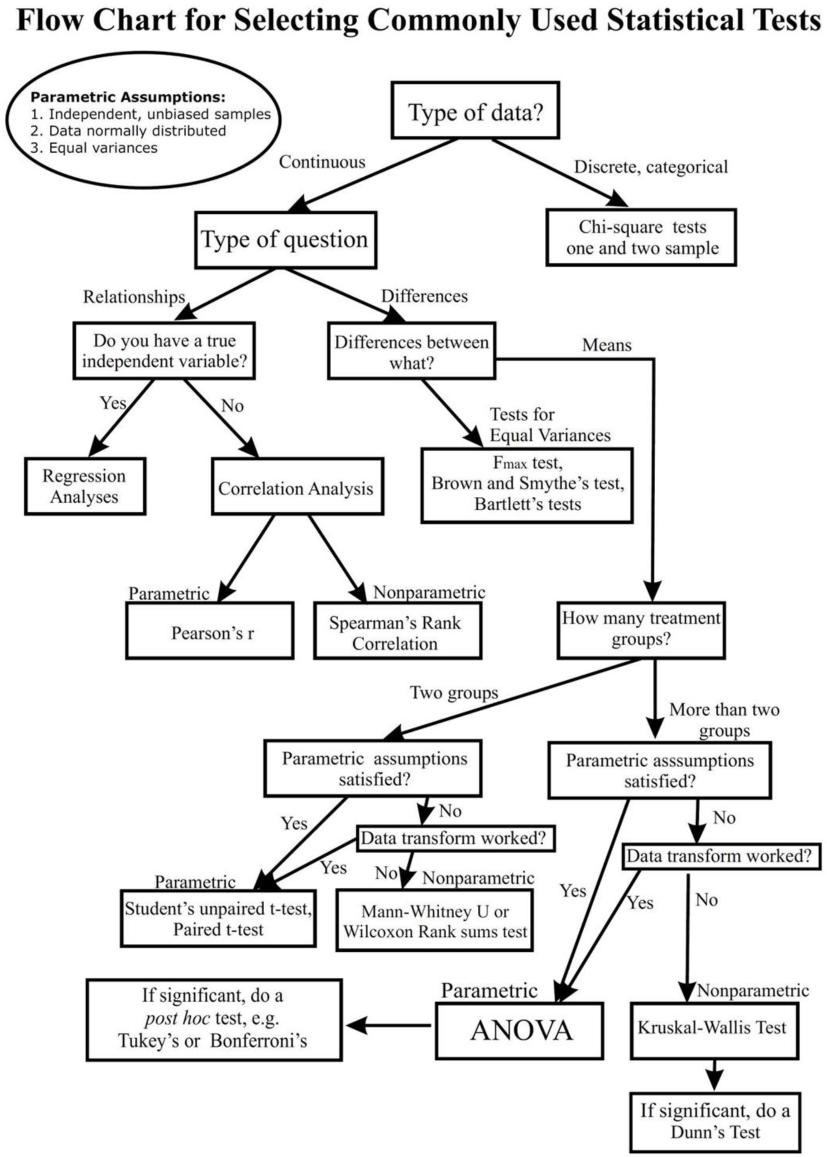
Fletcher W. Halliday[1*] & Jason R. Rohr[2]

nature COMMUNICATIONS

What is our conceptual model of the world?

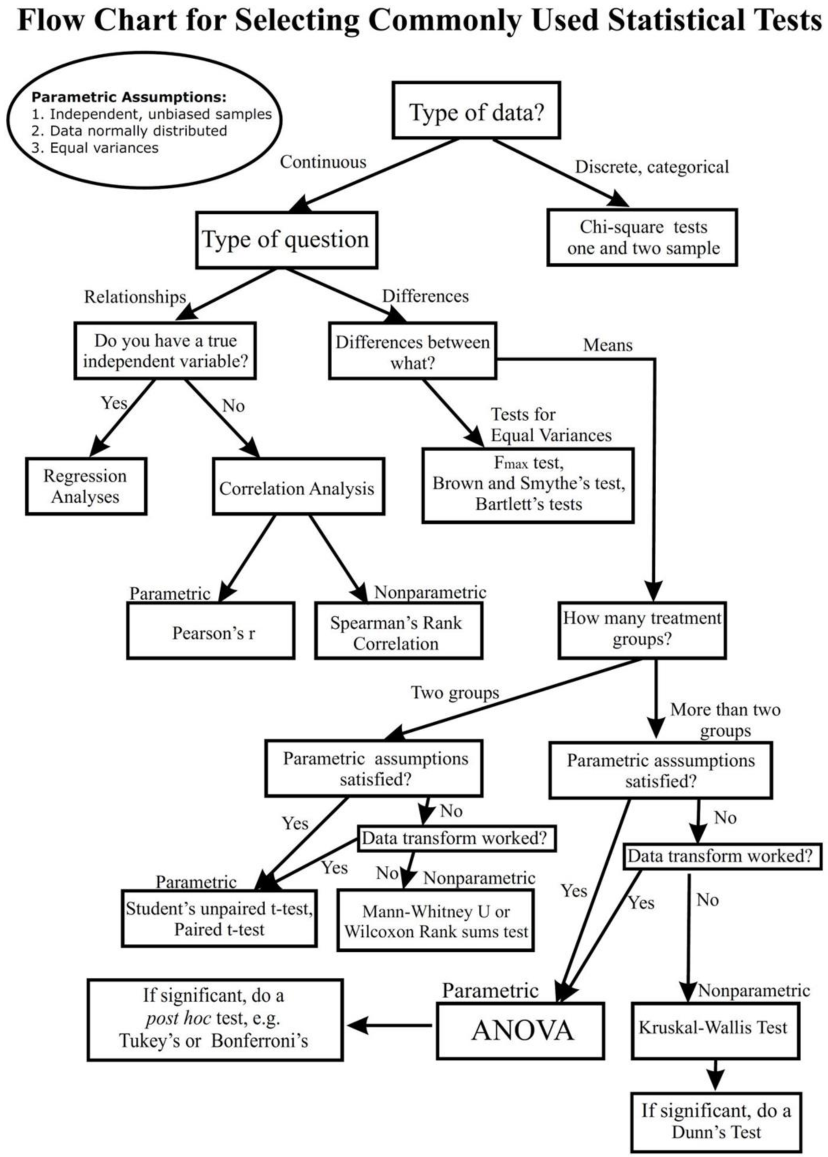
Conceptual **Statistical**

The ‘cookbook’ approach to teaching statistics



https://twitter.com/TrevorABranch/status/1082135735757504512

The ‘cookbook’ approach to teaching statistics



This approach ignores the most important things that help us ask the right question!

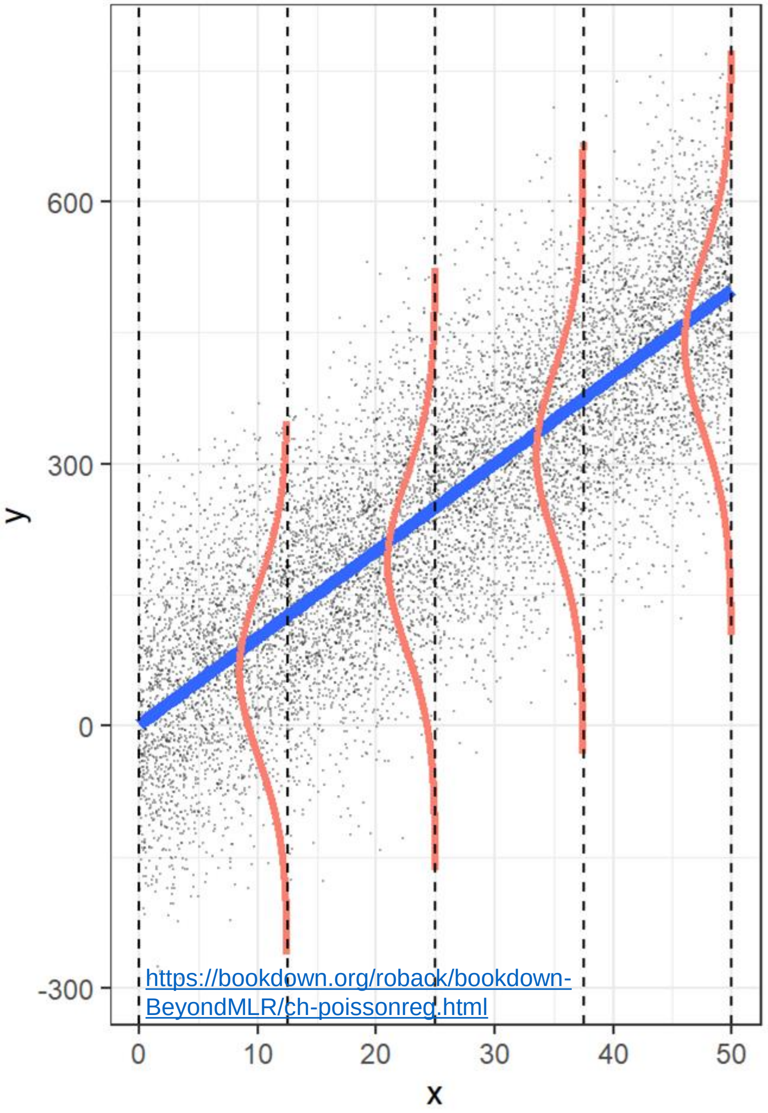
(1) Our scientific understanding of the system (the specific conceptual model)

(2) How the data were generated

https://twitter.com/TrevorABranch/status/1082135735757504512

Building blocks of a statistical model

Linear regression



https://bookdown.org/roback/bookdown-BeyondMLR/ch-poissonreg.html

General formula for linear regression

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

Intercept

Slope(s) for

covariates X

Residual error -
assumed independent, normally distributed and homoscedastic conditional on the model

Generalised linear models

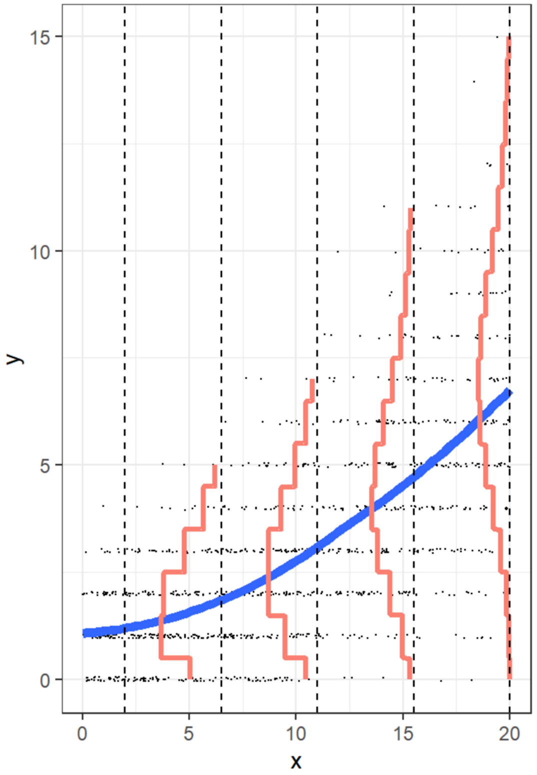
- Likelihood and link function relax assumptions around linearity and error distribution

- Still assume errors are independent conditional on the model!

- **Poisson regression**: count data, Poisson likelihood *(error variance equal to expected value)*, log link function *(assume exponential relationship between X and Y)*

Poisson

(e.g. butterfly counts)



Likelihood function
(describes error distribution)

$$Y_i \sim Pois(\lambda_i)$$

Link function
(describes shape of relationship between X and Y)

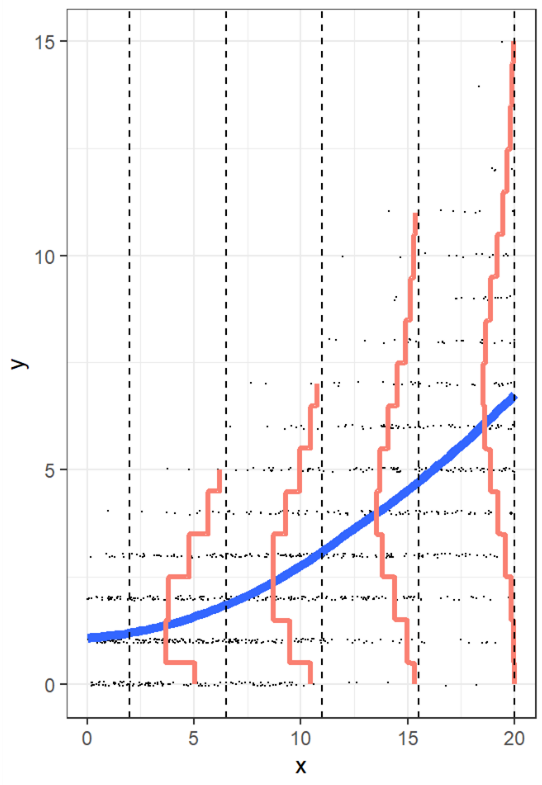
$$\log(\lambda_i) = \beta_0 + \beta_1 X_{i1}$$

Generalised linear models

- Likelihood and link function relax assumptions around linearity and error distribution
- Still assume errors are independent conditional on the model!
- **Poisson regression**: count data, Poisson likelihood *(error variance equal to expected value)*, log link function *(assume exponential relationship between X and Y)*
- **Logistic regression**: binary outcome (1/0), binomial likelihood *(probability of success)*, logit link *(assume linear relationship between X and log odds of Y)*

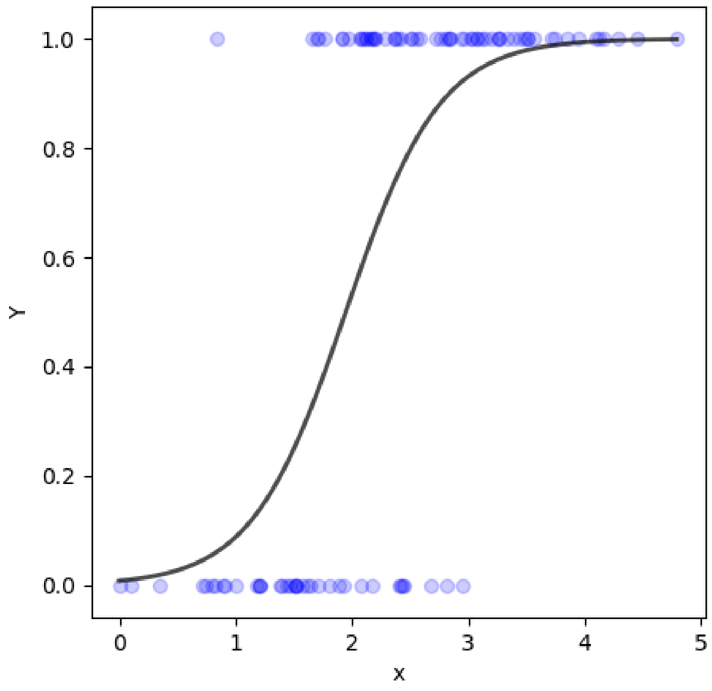
Poisson

(e.g. butterfly counts)



Logistic (binomial)

(e.g. hatching success)



Likelihood function
(describes error distribution)

Link function
(describes shape of relationship between X and Y)

$$Y_i \sim Pois(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 X_{i1}$$

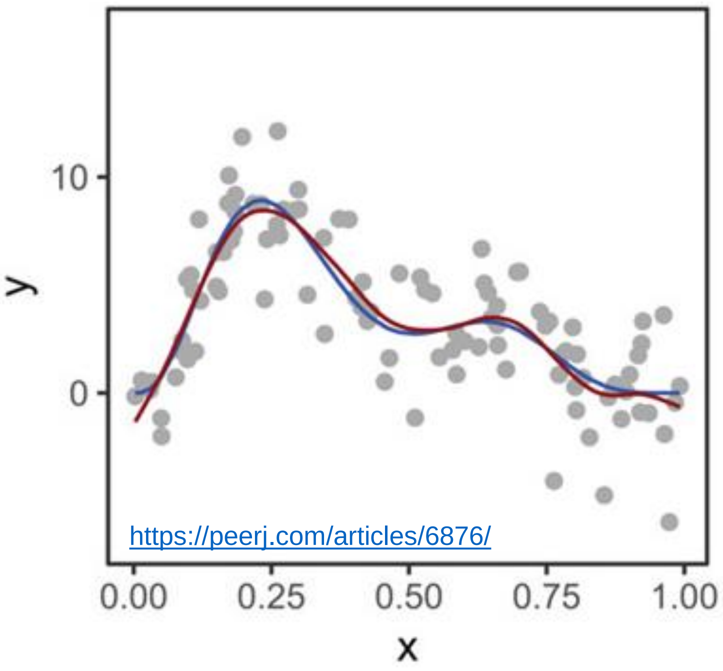
$$Y_i \sim Binom(p_i)$$

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_{i1}$$

Extensions to nonlinearity, multilevel data, space and time

Nonlinear

(e.g. generalised additive models)



$$Y_i = f(\eta_i) = \beta_0 + f(X_{i1})$$

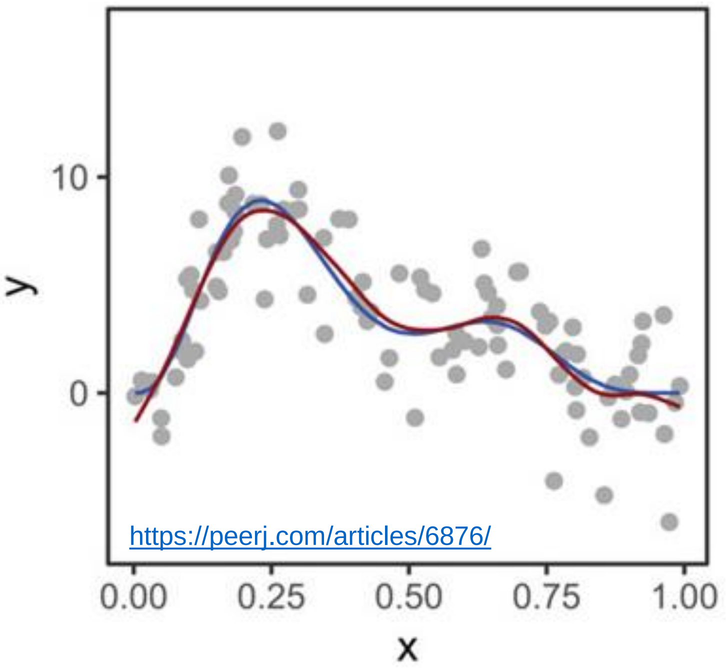
Nonlinear function of covariates X

(e.g. penalised splines)

Extensions to nonlinearity, multilevel data, space and time

Nonlinear

(e.g. generalised additive models)



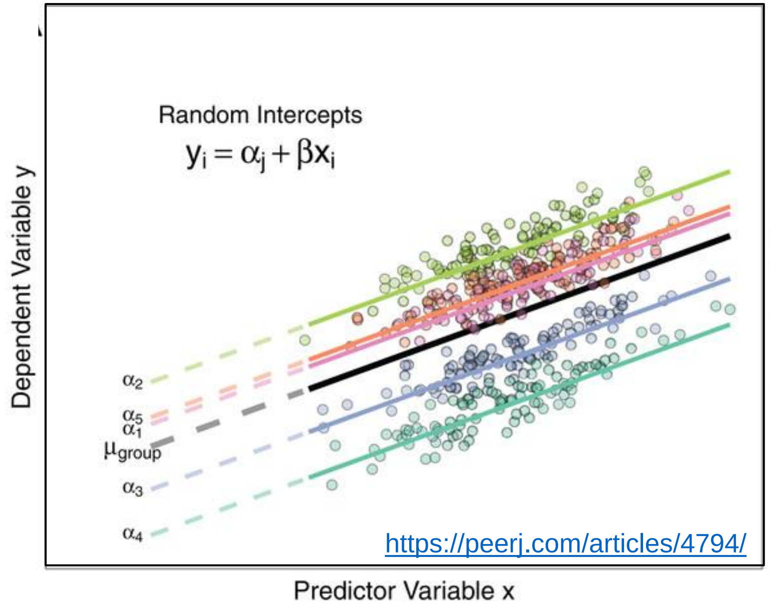
$$Y_i = f(\eta_i) = \beta_0 + f(X_{i1})$$

Nonlinear function of covariates X

(e.g. penalised splines)

Multilevel (mixed effects)

(e.g. GLMMs; random slopes or intercepts)



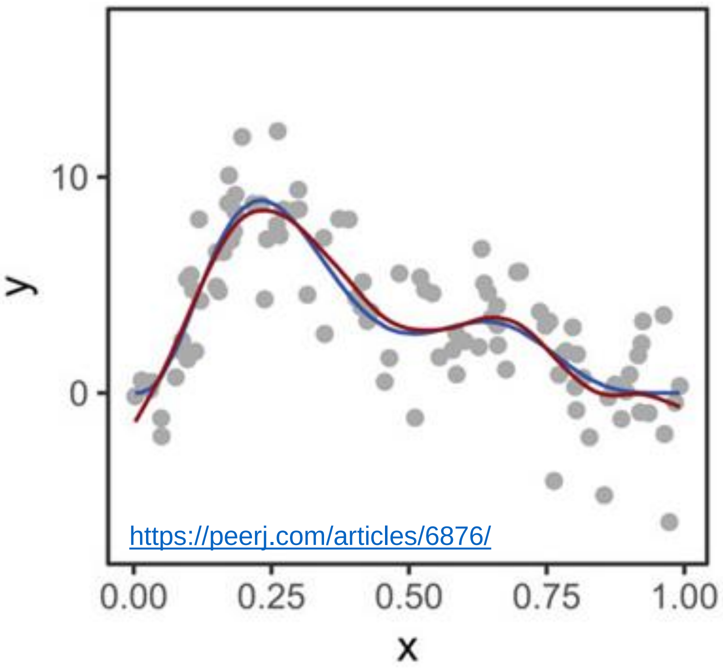
$$Y_i = f(\eta_i) = \beta_0 + \beta_1 X_{i1} + \alpha_{s(i)}$$

Random intercept for study site *S*

(intercepts vary to account for hierarchical structure; pools information across sites)

Extensions to nonlinearity, multilevel data, space and time

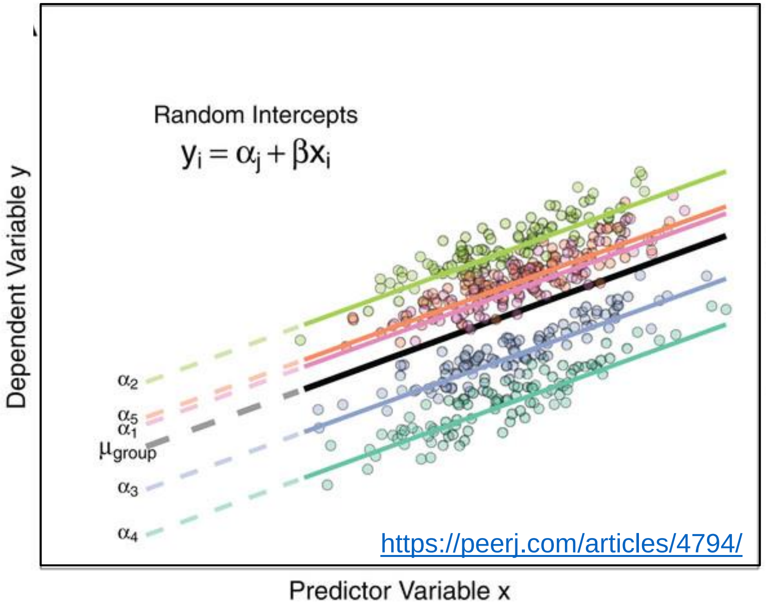
Nonlinear
(e.g. generalised additive models)



$$Y_i = f(\eta_i) = \beta_0 + f(X_{i1})$$

Nonlinear function of covariates X
(e.g. penalised splines)

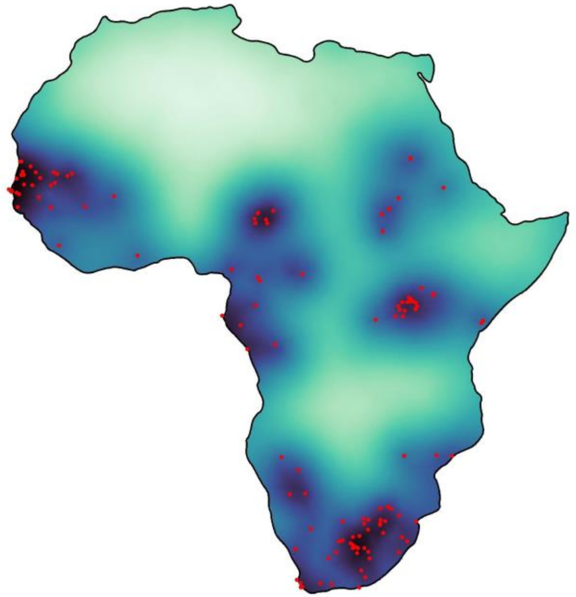
Multilevel (mixed effects)
(e.g. GLMMs; random slopes or intercepts)



$$Y_i = f(\eta_i) = \beta_0 + \beta_1 X_{i1} + \alpha_{s(i)}$$

Random intercept for study site *S*
(intercepts vary to account for hierarchical structure; pools information across sites)

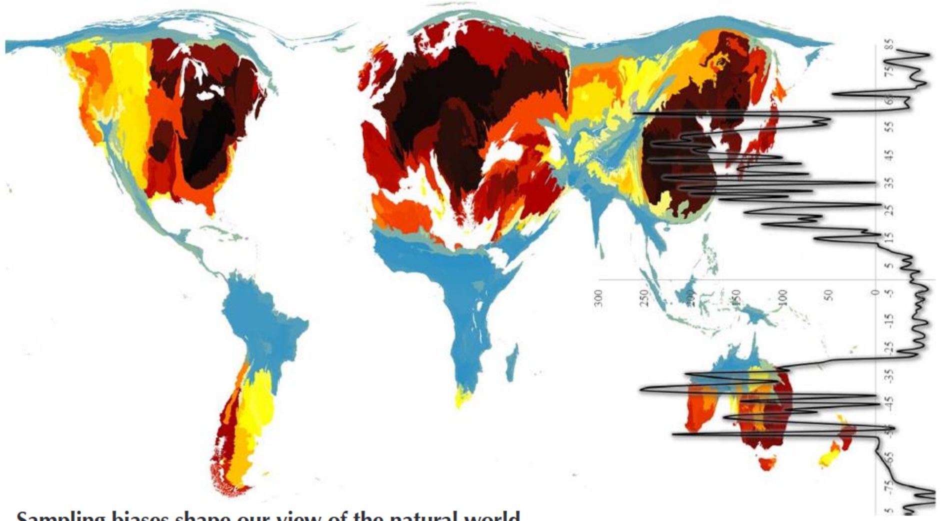
Spatial/temporal
(e.g. conditional autoregressive; ARIMA; Gaussian processes)



$$Y_i = f(\eta_i) = \beta_0 + \beta_1 X_{i1} + s_i$$

Spatially or temporally-structured effect
(observations that are closer together are more closely related)

Real worlds are noisy (and so are data)

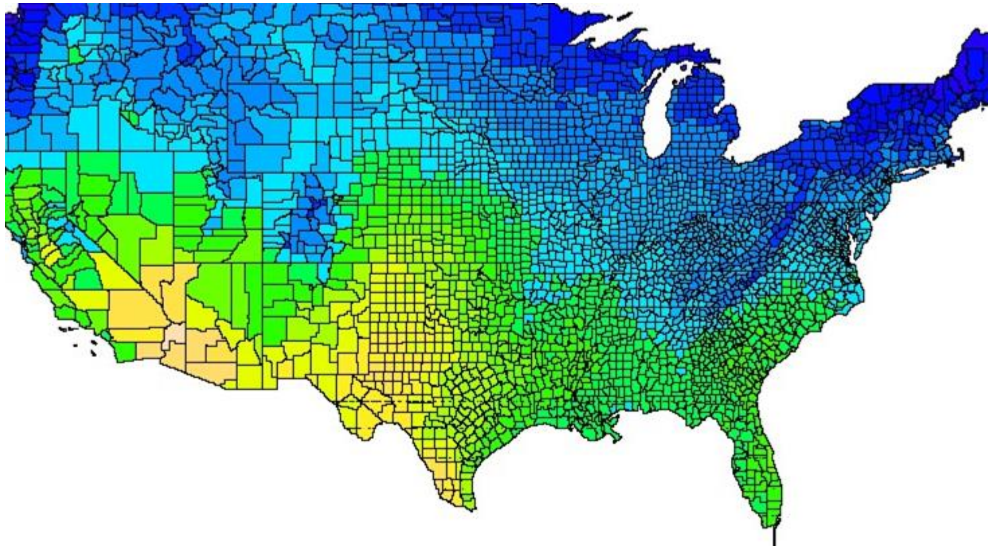


Sampling biases shape our view of the natural world

Alice C. Hughes, Michael C. Orr, Keping Ma, Mark J. Costello, John Waller, Pieter Provoost, Qinmin Yang, Chaodong Zhu and Huijie Qiao

Dependency in space and time

(observations closer together are more closely related)



Sampling biases shape our understanding

(historical and evolving biases in survey effort confound our knowledge of pattern and process in ecology)

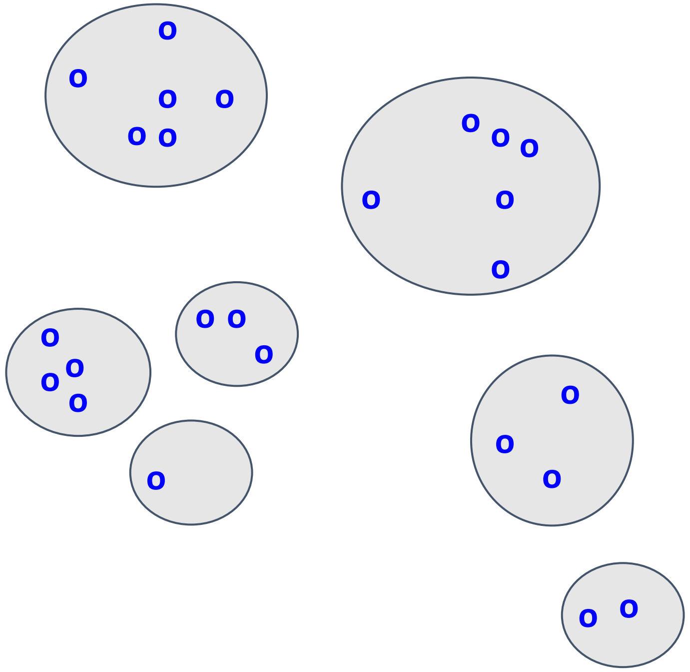
Observation is imperfect

(ecological data often contain a mixture of true and false zeroes, and we often do not measure key variables acting on the system)

Mixed-effects (multilevel/hierarchical models)

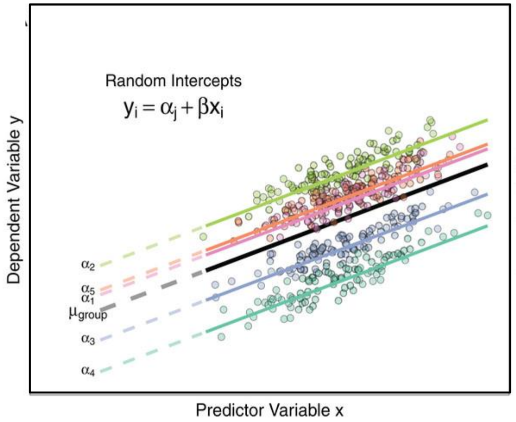
- Data often contain some clustered/nested structure (e.g. space, time, phylogeny, individual, population)
- **Observations within each cluster may be non-independent -** account for this by allowing intercepts/slopes to vary across levels (*"random effects"*)

Ecological study of bird abundance across an island archipelago



O = Abundance record of bird species X
Sampled across 7 islands *s* (s = 1…7)

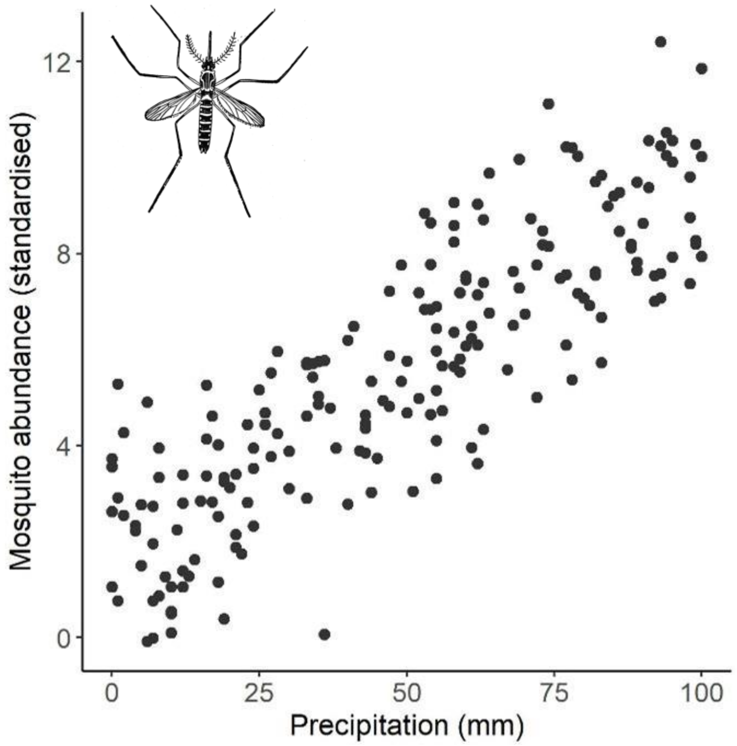
$$Y_i \sim Pois(\lambda_i)$$
$$\log(\lambda_i) = \beta_0 + \beta_1 X_1 + \alpha_s$$
$$\alpha_s \sim N(0, \sigma)$$



Island-level intercepts are modelled as a population described by a normal distribution with variance σ (how variable between islands?)

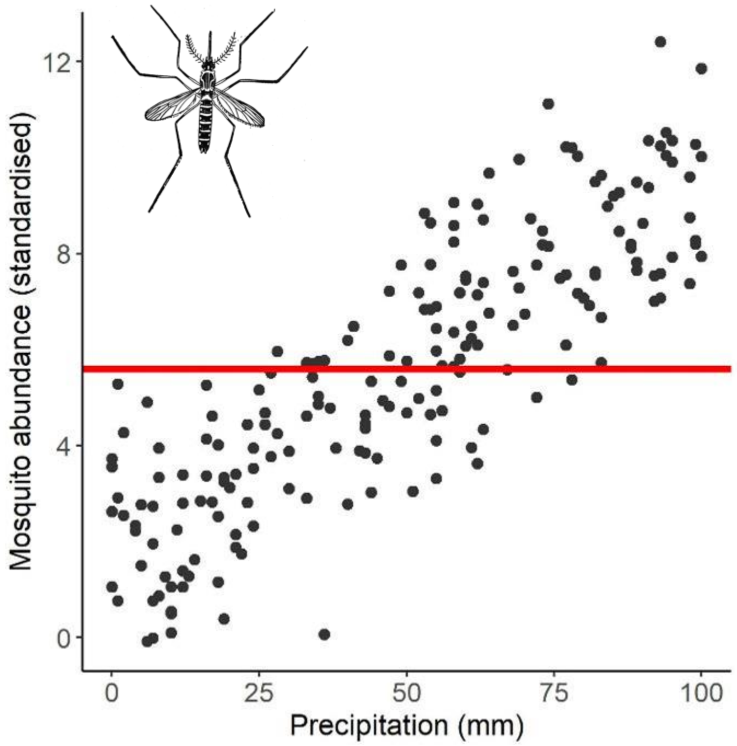
A model within a model - hence "hierarchical"!
(σ is a '*hyperparameter*')

What do we mean by “good” model fit?



- Toy example: we have surveyed mosquito abundance in relation to the previous month’s precipitation.
- **How do we model these data to gain generalisable knowledge about this relationship?**
- *(e.g. if we want to predict abundance next month based on this historical pattern?)*

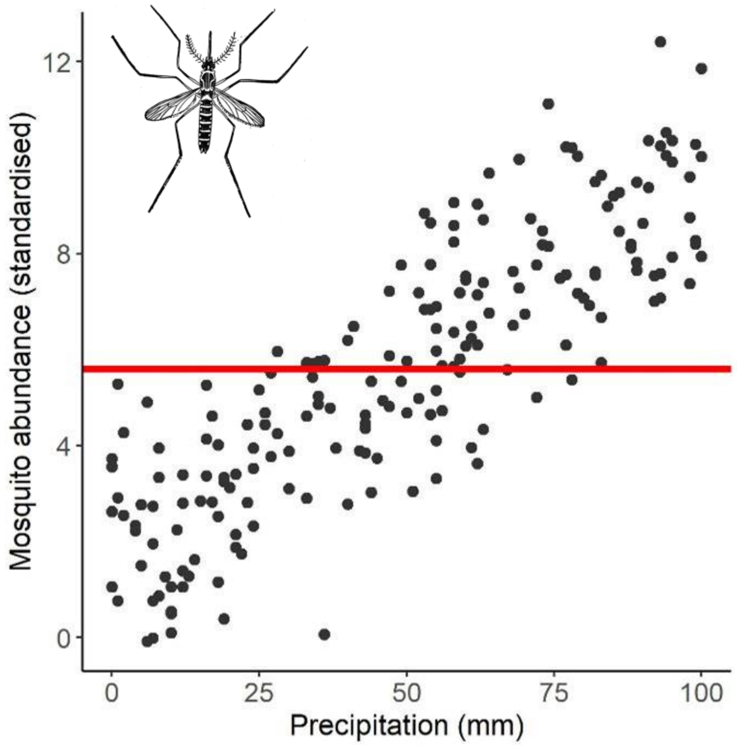
What do we mean by “good” model fit?



Simplest possible model
(‘null’ model)

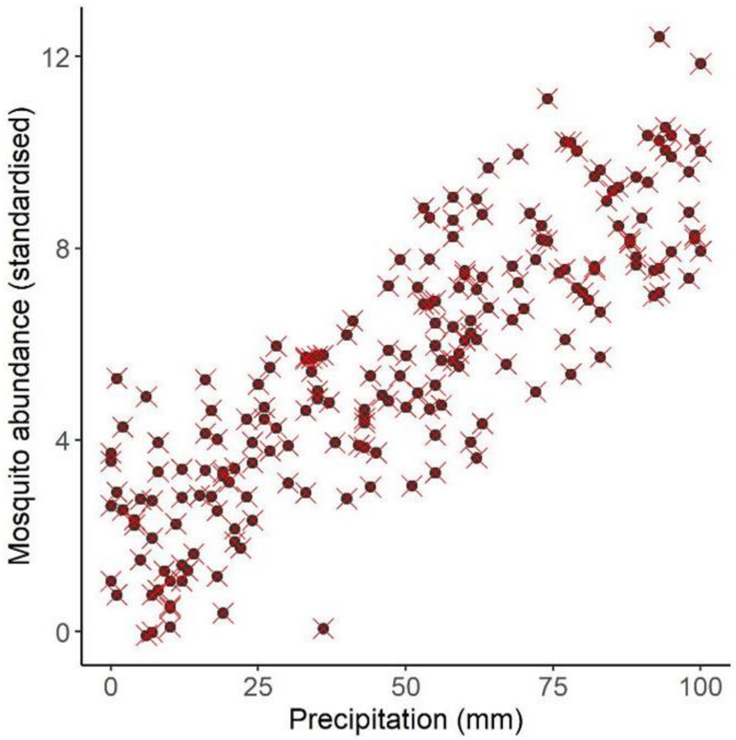
- 1 parameter (intercept)
- Maximum residual error
- Learning very little from the data so not generalisable

What do we mean by “good” model fit?



Simplest possible model
(‘null’ model)

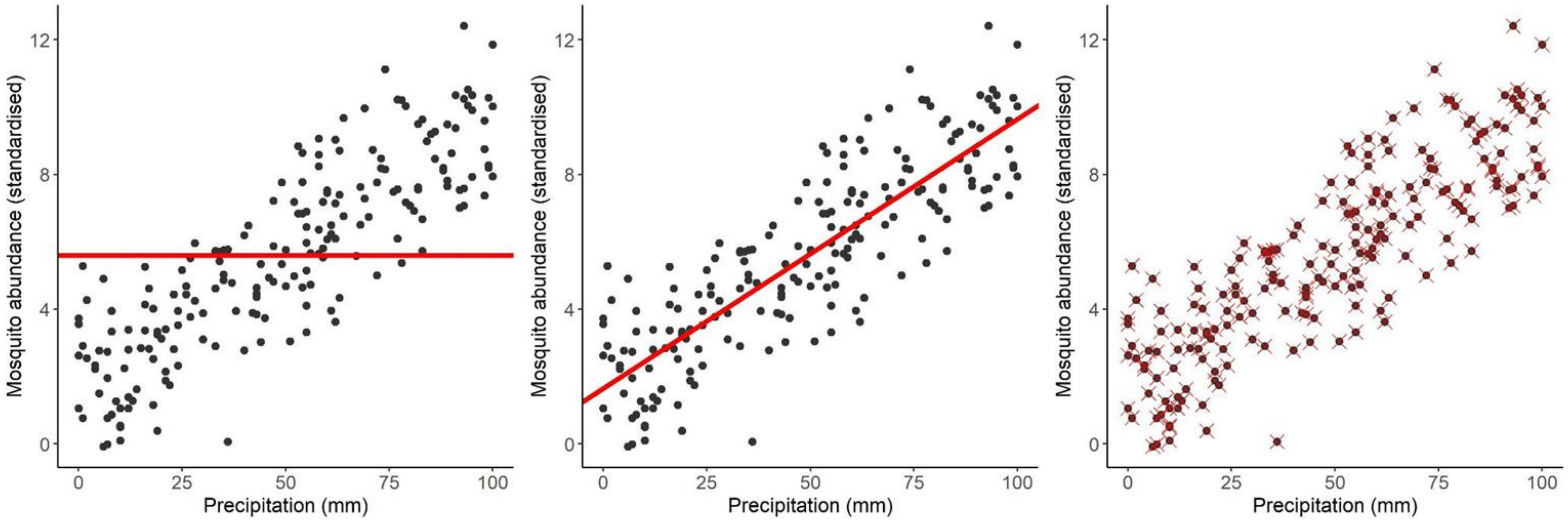
- 1 parameter (intercept)
- Maximum residual error
- Learning very little from the data so not generalisable



Most complex model
(‘saturated model’)

- 1 parameter <u>per observation</u>
- No residual error
- Learning too much from this specific dataset - not generalisable

What do we mean by “good” model fit?



‘Goldilocks’ model

(not too simple, not too complex)

Underfitting ← → **Overfitting**

- Learned enough from the data for (potentially) generalisable inference

From question, to data, to model design

Research question

Analysis objective (description/ inference/ prediction)? Which covariates? Expected shape of relationships (e.g. linear/ nonlinear)? Confounders?

Model design

Data collection (*how were the data generated?*)

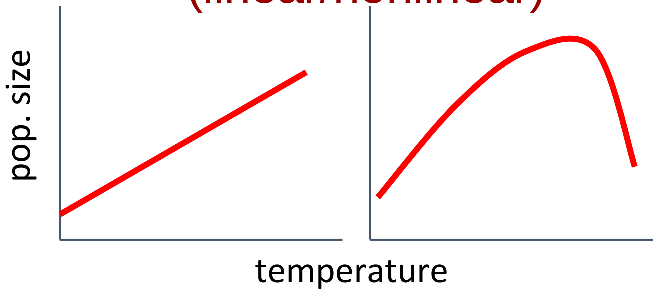
Structured survey vs unstructured? Continuous/discrete data? Sampling design (random, non-random, nested)? Spatial/temporal structure? Measurement error? Selection bias?

Using scientific understanding to inform model design

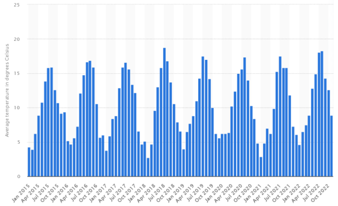
“What winter weather variables are associated with springtime abundance of the Red Admiral butterfly?”

What variables? (temperature, precip, humidity…?)

What shape of relationships do we expect? (linear/nonlinear)



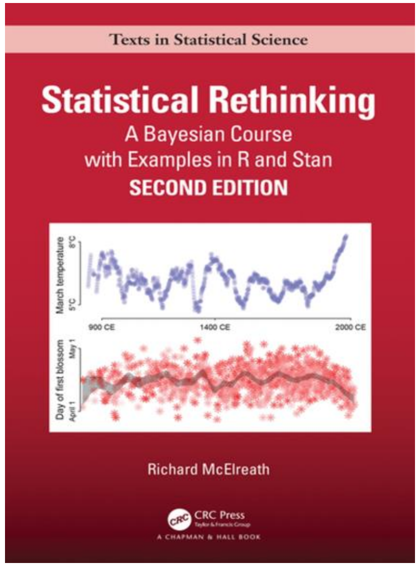
What timescale of climate - multi-month averages? Transient extremes?



What other factors might be impacting abundance, or interacting with weather?

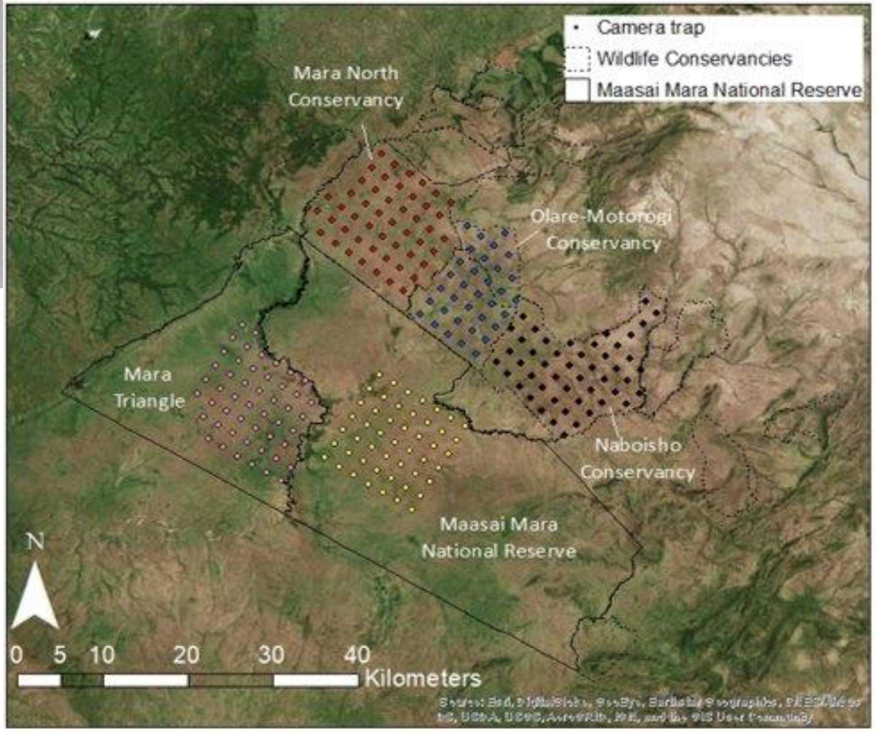
Models are powerful but unaware

- Statistical models are powerful machines for helping us to understand and predict the world, but lack insight.
- The model only knows what you tell it!



https://xcelab.net/rm/statistical-rethinking/

Afternoon workshop: effects of grazing on hare occurrence



Cape hare (*Lepus capensis*)

https://www.inaturalist.org/guide_taxa/271513