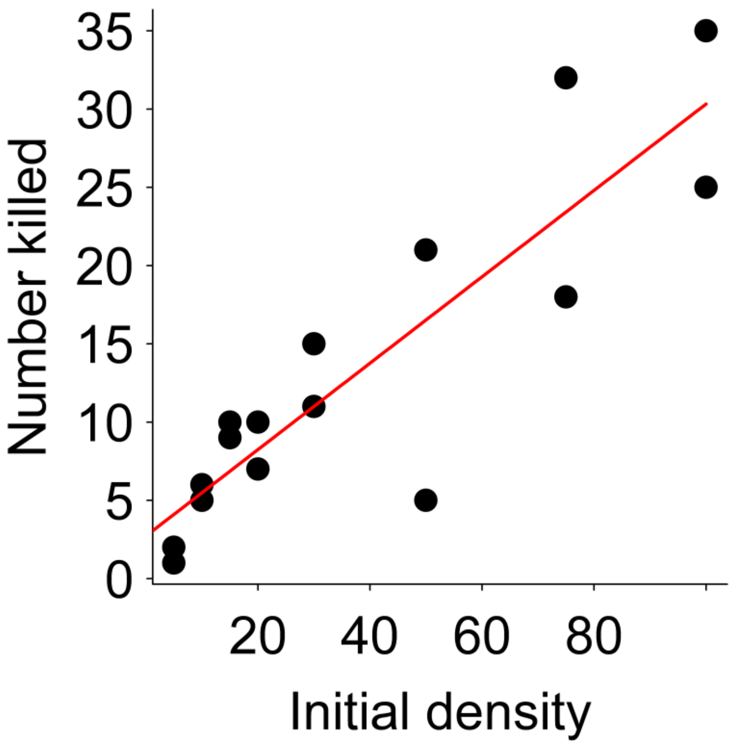
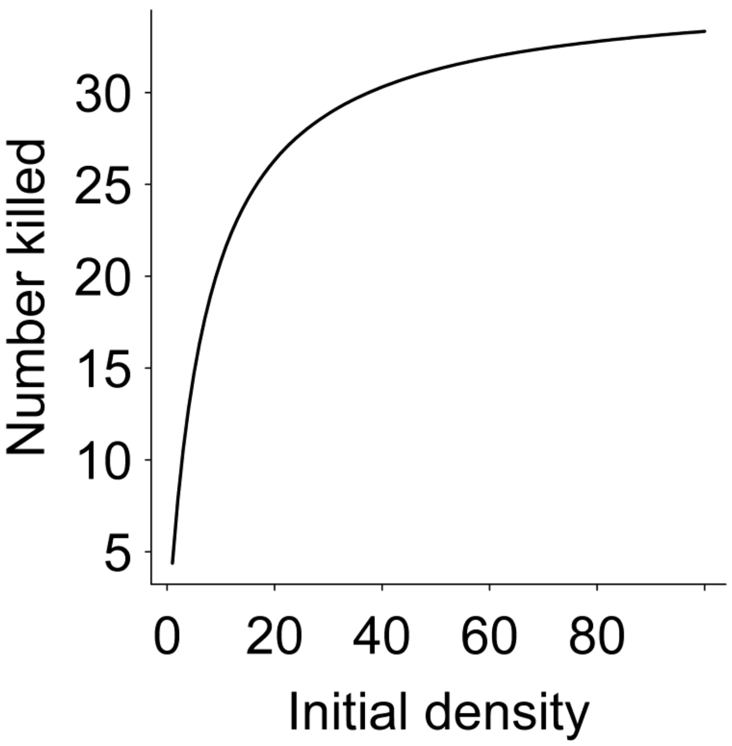
An introduction to ecological modelling

1. Approaches to Modelling and Statistical Models of Ecological Data

What is an ecological model?



Statistical models, such as linear regression.
E.g. predator-prey functional response
Data from Vonesh & Bolker (2005) *Ecology*



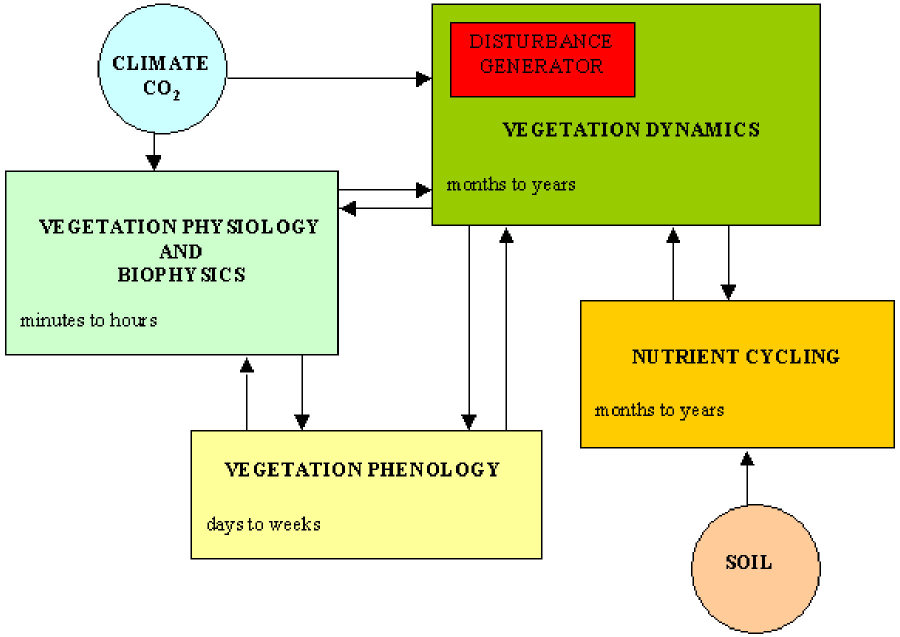
Simple theoretical models.
Predator-prey functional response again,
but this time without data

What is an ecological model?

(More) complex, mechanistic or process-based models

Represent the key processes underlying a system

Examples include global climate models and vegetation models



Course Objectives

An appreciation for the role that modelling plays in ecology (with a focus on studies that rely heavily on modelling)

An overview of the types of modelling approaches available

Information to help pick the right model for a particular question

Examples of the application of different types of ecological models

No need to understand all of the maths!

Lecture Objectives

Reasons for building ecological models

Trade-offs in different types of models

Statistical models beyond classical statistics

- General applications of maximum likelihood
- Bayesian statistics
- Mixed-effects models

Why model?

Ethical considerations

Large-scale studies – sample size = 1

Predicting the future

No model is perfect

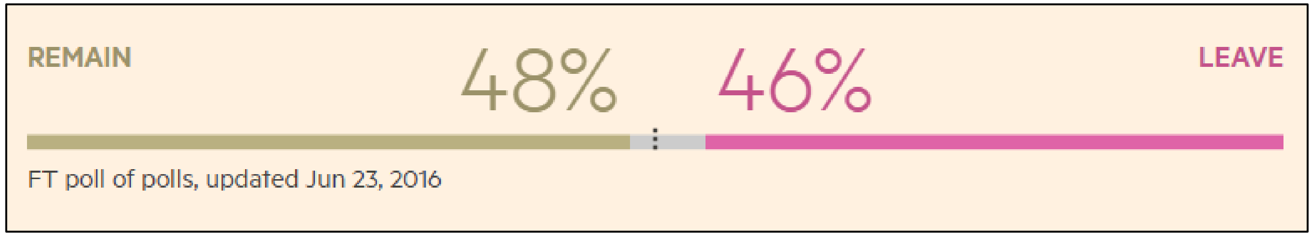
“Since all models are wrong the scientist cannot obtain a ‘correct’ one by excessive elaboration”

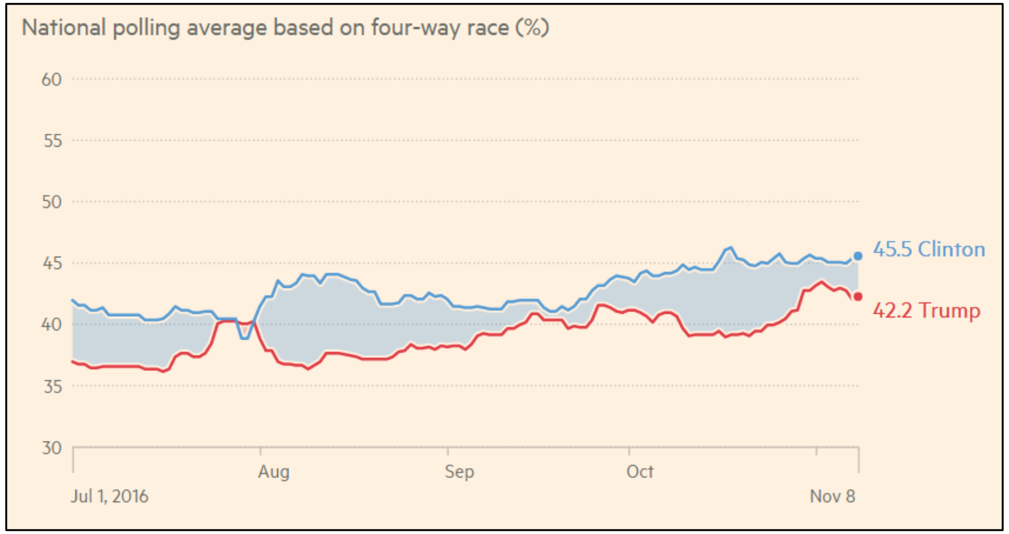
George Box (1976). Science and statistics. *Journal of the American Statistical Association* **71**: 791-799

“Essentially, all models are wrong, but some are useful”

George Box & Norman Draper (1987). *Empirical Model-Building and Response Surfaces*. Wiley.

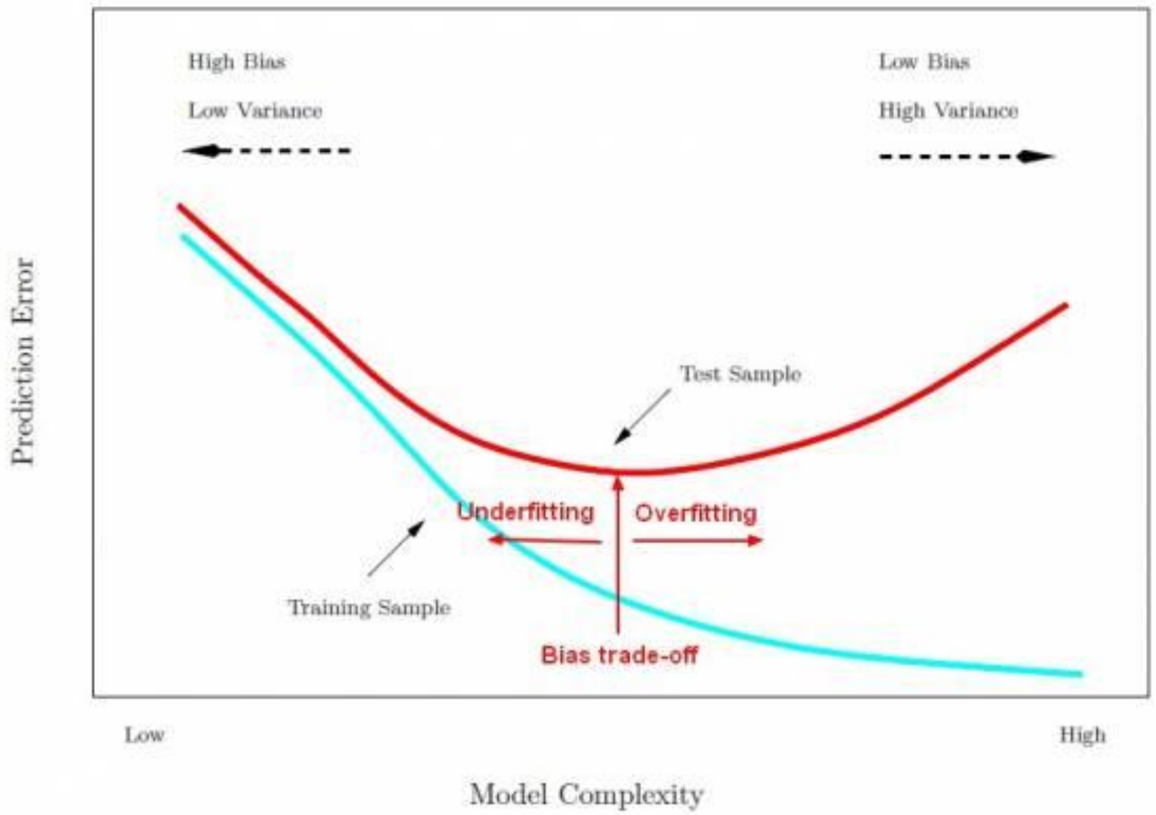
Even the most sophisticated modelling approaches won't work if model/assumptions/data are wrong





Source: Financial Times

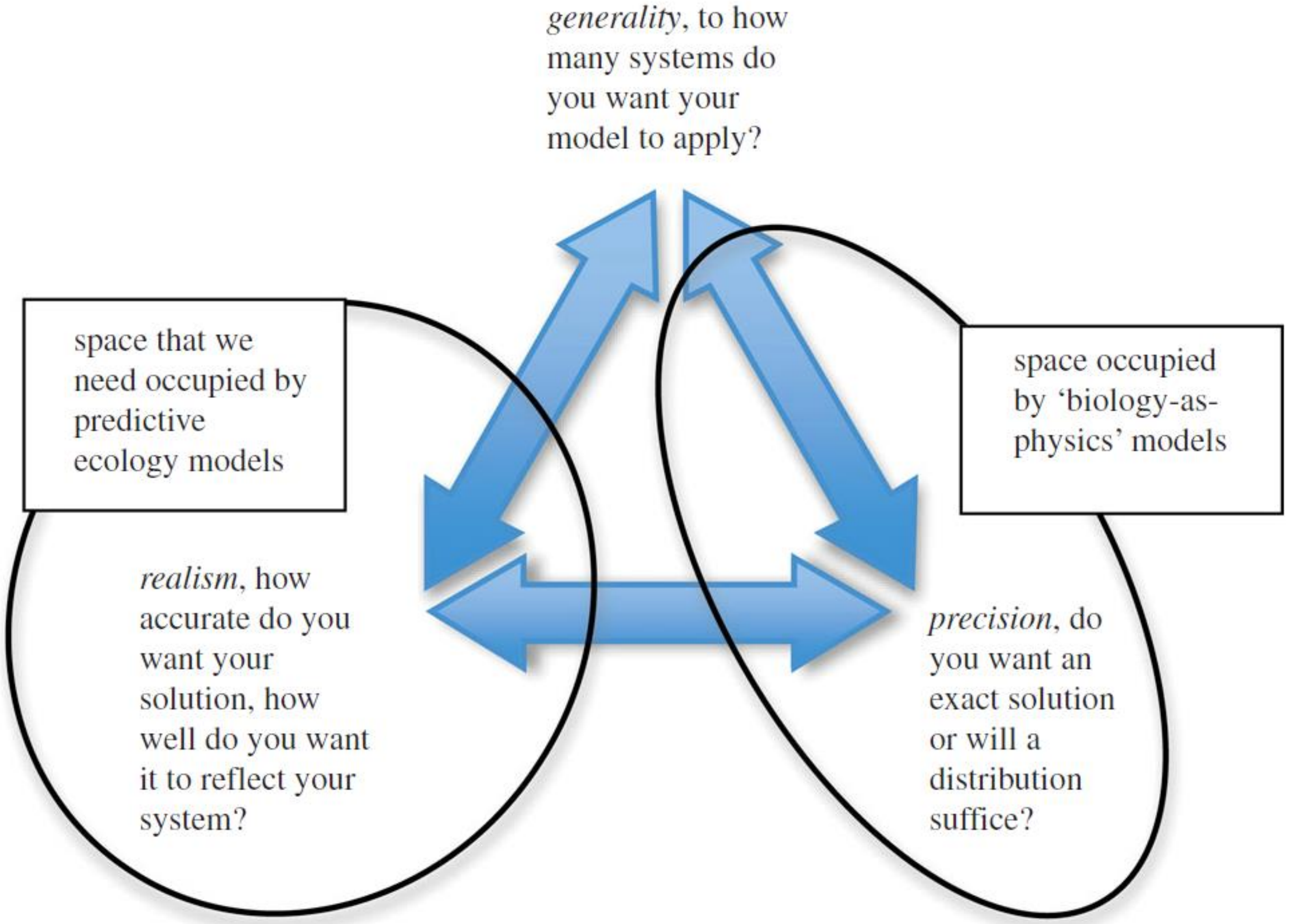
Model complexity, fit and predictive ability (statistical models)



More complex statistical models tend to fit the data used to build them better

But they risk overfitting the data, and have lower predictive power

Model realism, generality and precision (theoretical/mechanistic models)



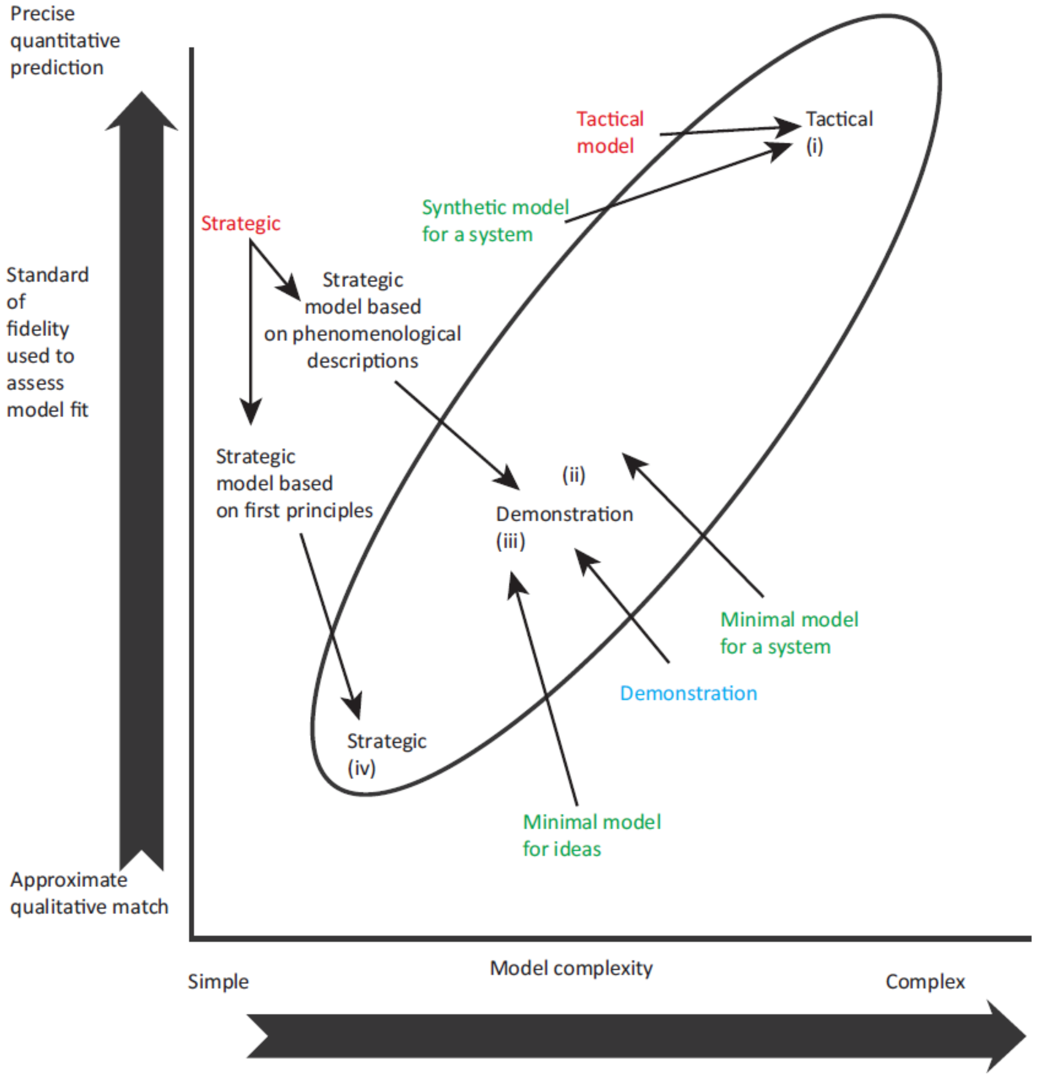
Precise models are useful for gaining understanding

Predictive models need to be more realistic (and possibly general)

'Physics envy' and the search for precise, general models

Evans (2012). *Philosophical Transactions of the Royal Society B* **367**: 181-190

Model complexity and model precision (theoretical/mechanistic models)

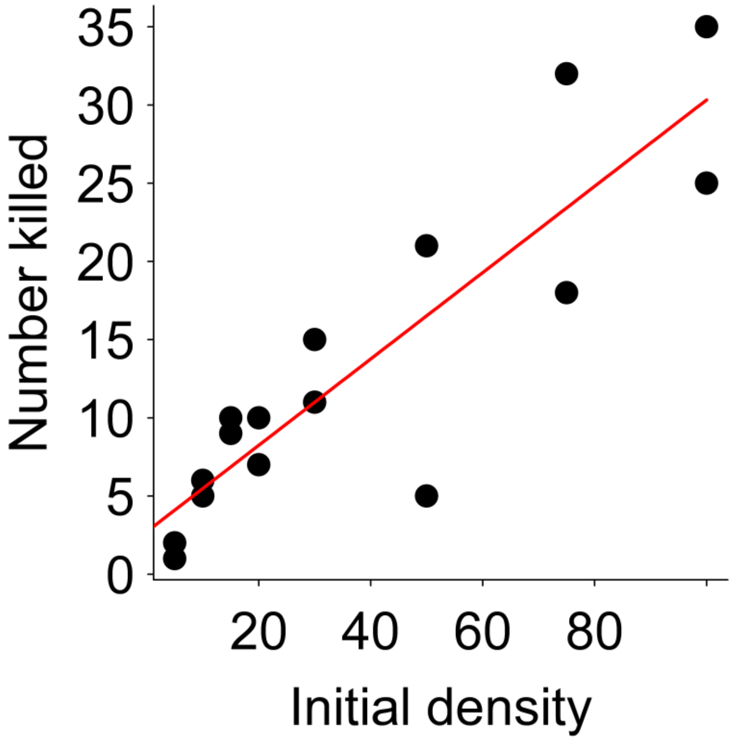


More complex models tend to be more precise

Simple models might be more general (but not if they miss important processes)

Evans et al. (2013). *TREE* **28**: 578-583

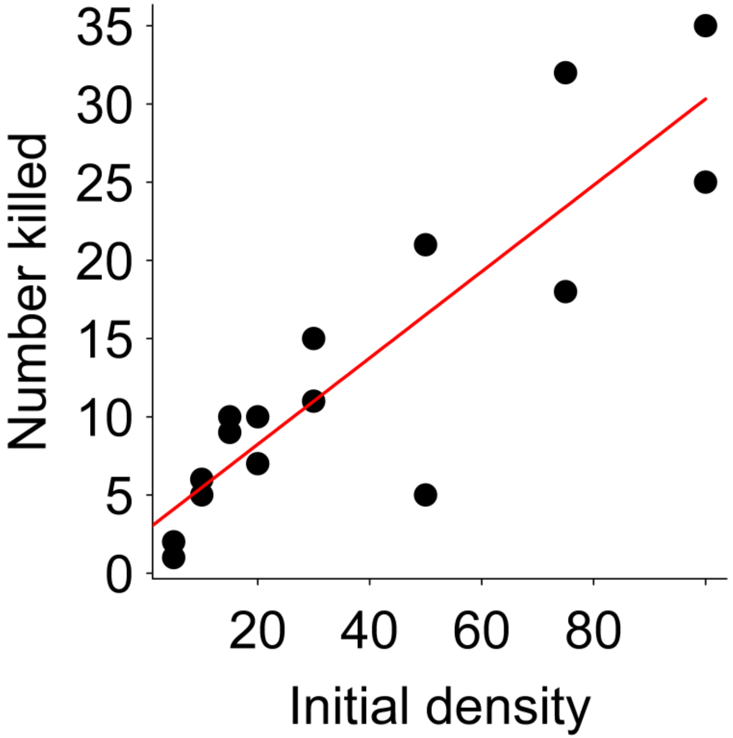
Statistical models



Statistical models:

- Maximum likelihood
- Frequentist vs. Bayesian statistics
- Hierarchical data and mixed-effects models

Maximum likelihood



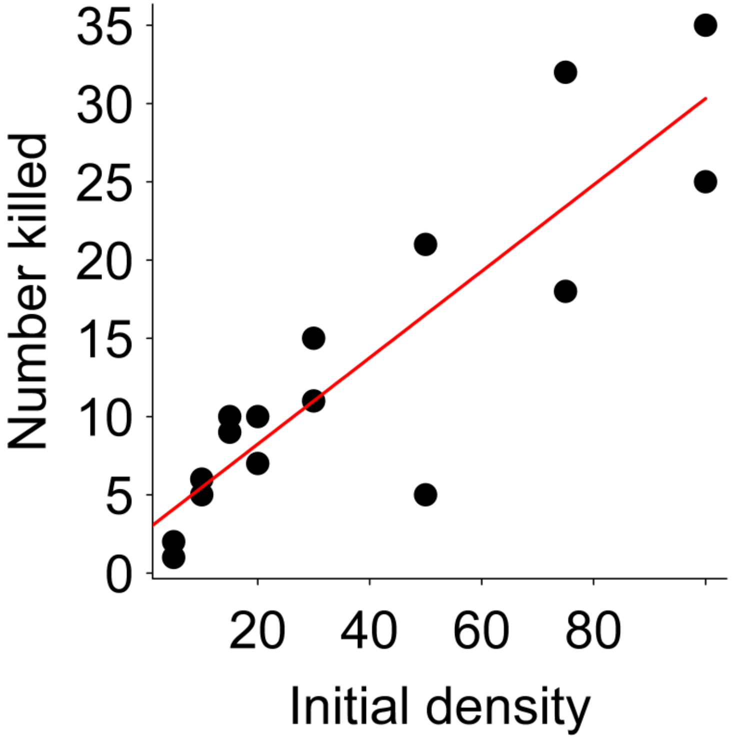
Statistical models, such as linear regression.
E.g. predator-prey functional response
Data from Vonesh & Bolker (2005) *Ecology*

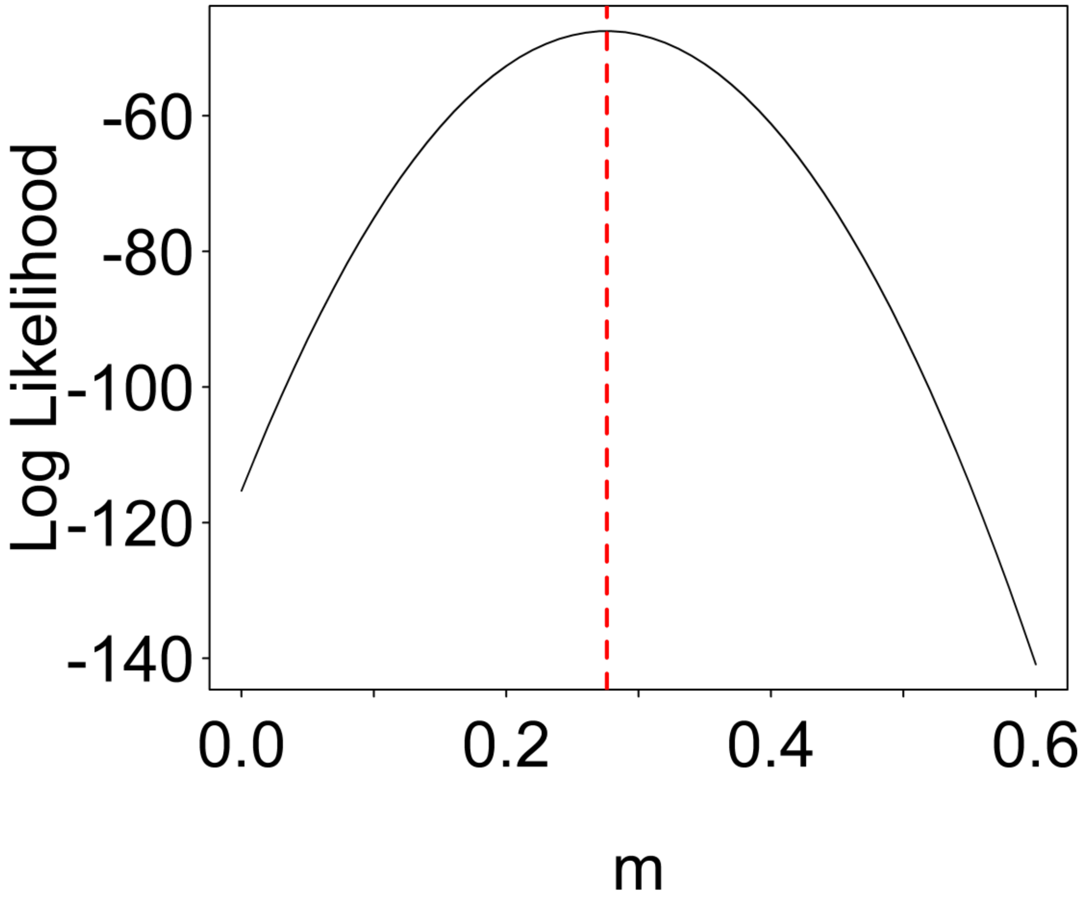
Much of classical statistics is concerned with finding the best estimate for parameters

E.g. in the case of a linear regression, the slope and intercept

Often this is done by maximizing (log) likelihood

Maximum likelihood: a simple linear regression example



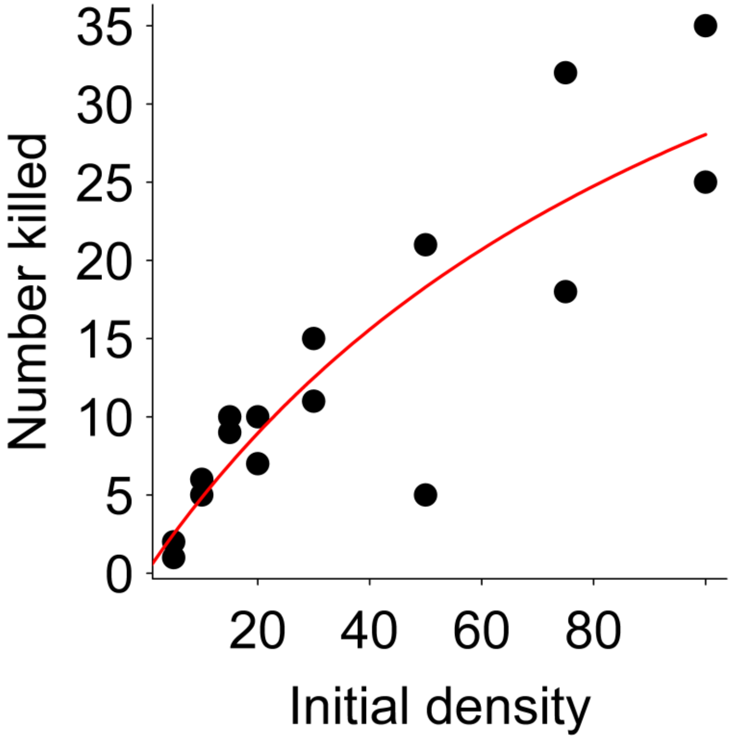


Killed = m × Initial + c + ε

ε ~ N(0, σ)

Killed – (m × Initial + c) ~ N(0, σ)

Maximum likelihood: more complex models



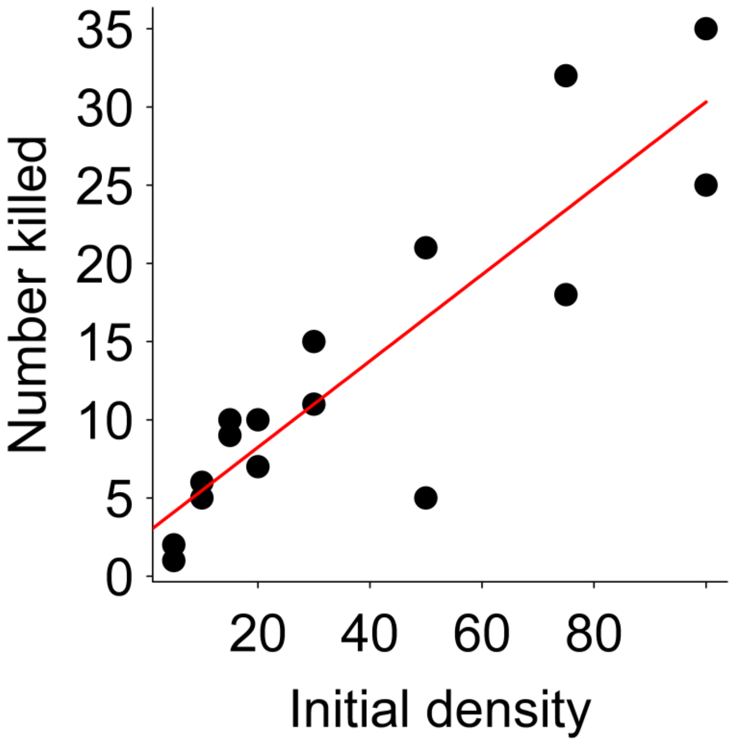
Predator-prey functional response
Data from Vonesh & Bolker (2005) *Ecology*

Type II functional response:

$$P_{death} = \frac{a}{1 + aHN}$$

$$N_{killed} = \frac{aN}{1 + aHN}$$

Finding the maximum likelihood



Predator-prey functional response
Data from Vonesh & Bolker (2005) *Ecology*

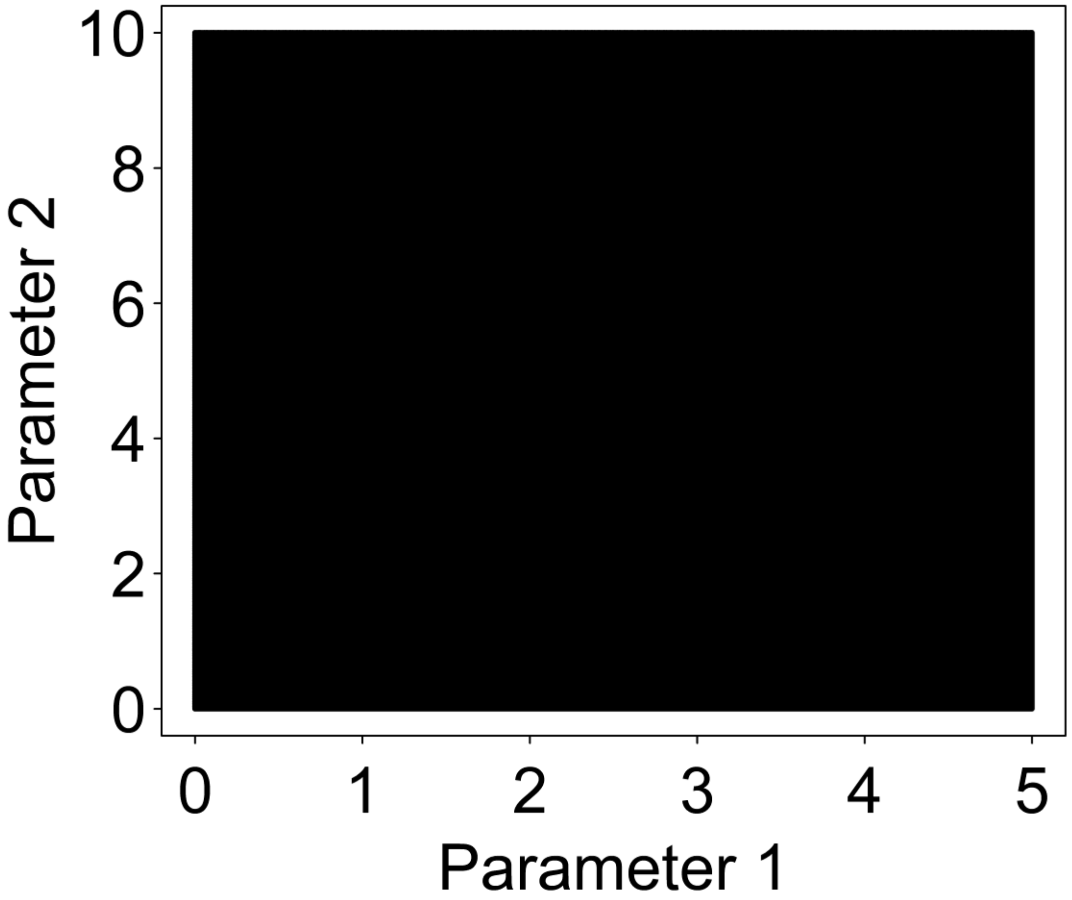
Analytical solutions, for example ordinary least squares regression

$$y = mx + c + \varepsilon$$

$$m = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

$$c = \bar{Y} - m\bar{X}$$

Finding the maximum likelihood



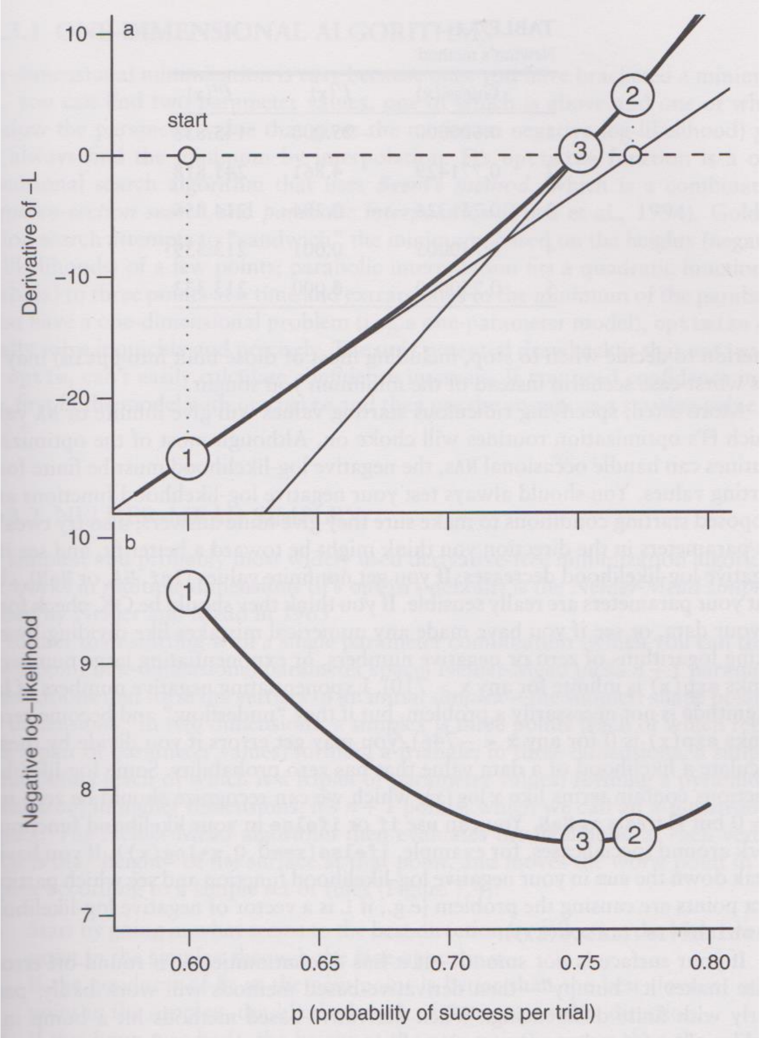
1,000,000 parameter combinations

Brute force search

Can be useful for simple models

But quickly becomes unmanageable

Finding the maximum likelihood



Derivative-based methods

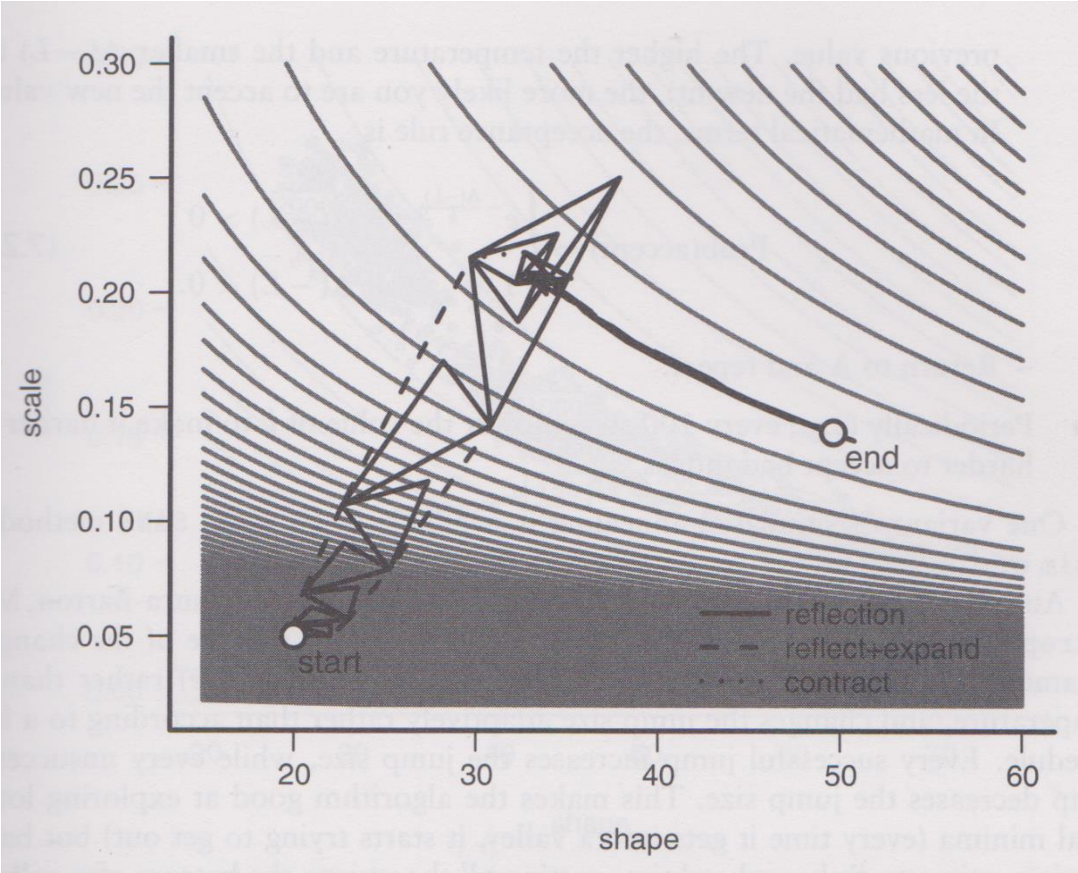
Find point at which derivative of likelihood function = 0

E.g. Newton method

Only work well with smooth likelihood surfaces

From Bolker (2008). *Ecological Models and Data in R*. Princeton University Press.

Finding the maximum likelihood



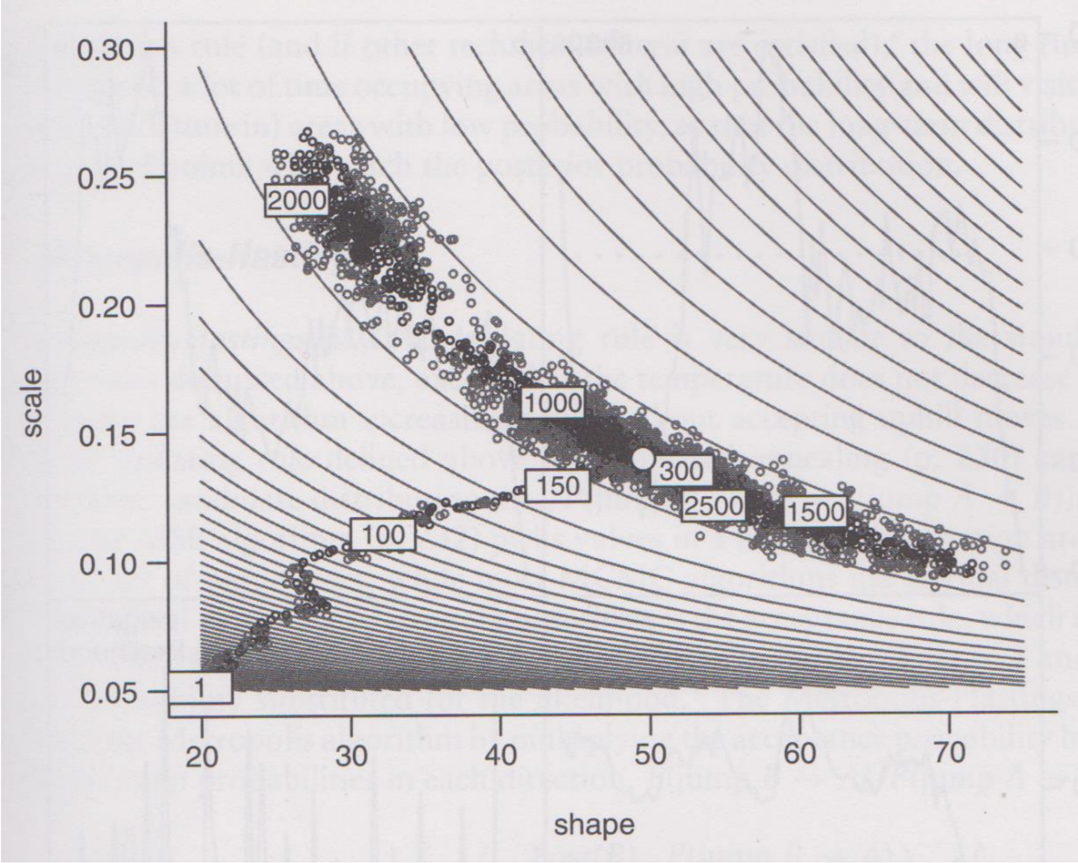
From Bolker (2008). *Ecological Models and Data in R*. Princeton University Press.

Derivative-free methods

Nelder-Mead simplex:

- For *n* parameters, based on n+1 combinations of parameters that form the vertices of a ‘simplex’
- This simplex is modified according to set rules

Finding the maximum likelihood



From Bolker (2008). *Ecological Models and Data in R*. Princeton University Press.

Derivative-free methods

Simulated annealing, e.g. Metropolis algorithm:

$$P(accept) = \begin{cases} e^{-\frac{\Delta(-L)}{k}} & if\ \Delta(-L) > 0 \\ 1 & if\ \Delta(-L) < 0 \end{cases}$$

k ('temperature') is reduced periodically to make moves to poor parameter values less likely

A note on selecting models: Information criteria

P-values are only applied in classical, frequentist statistics

An alternative is to use information criteria

These are all a measure of variation in the response variable explained, penalized by number of free parameters

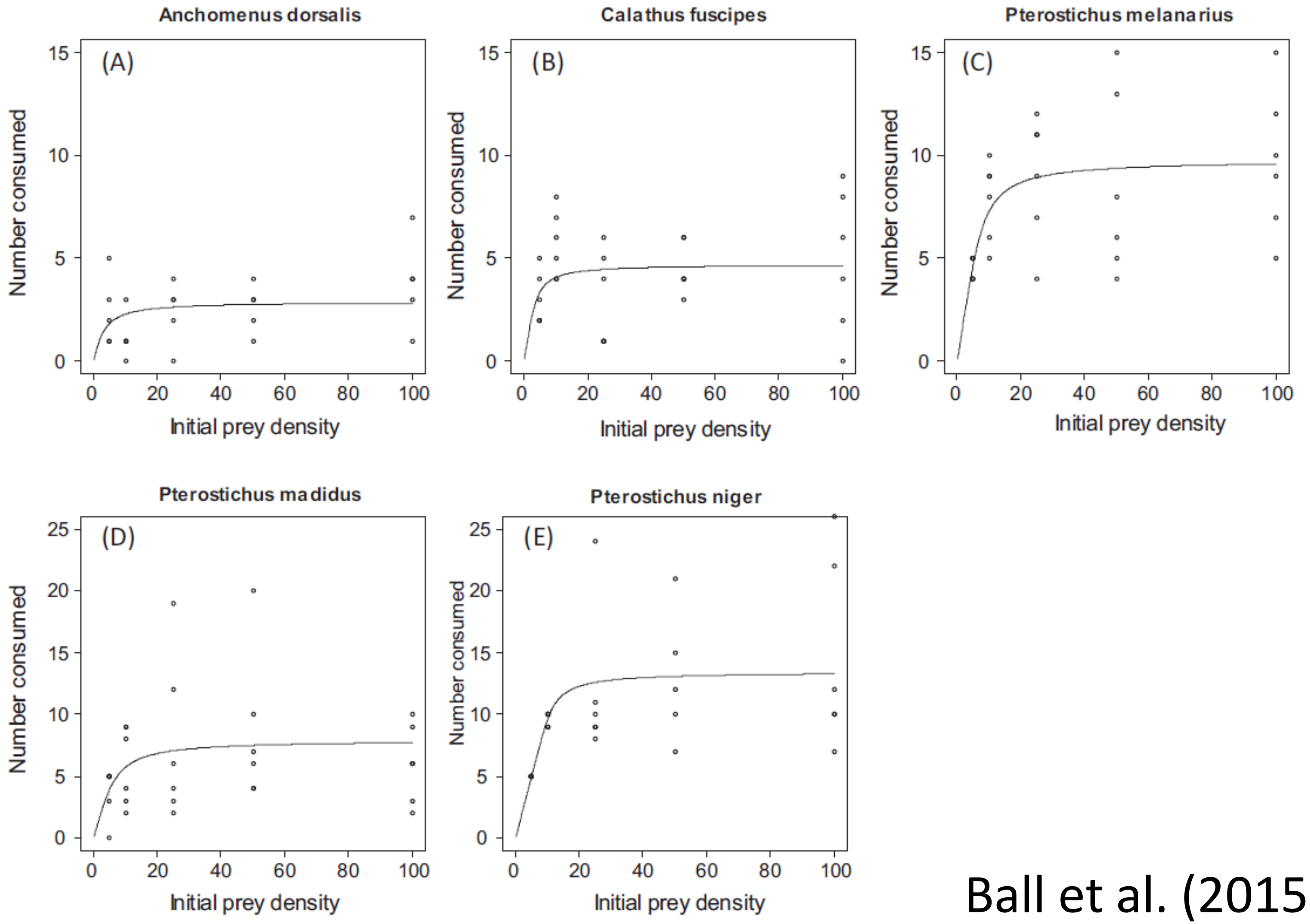
Akaike's Information Criterion (AIC):

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

L = likelihood, k = number of free parameters

For more on model selection, see Johnson & Omland (2004). *TREE* **19**: 101-108

Applications of maximum likelihood estimation: inferring functional responses



For 5 beetle species, modelled Type II functional responses:

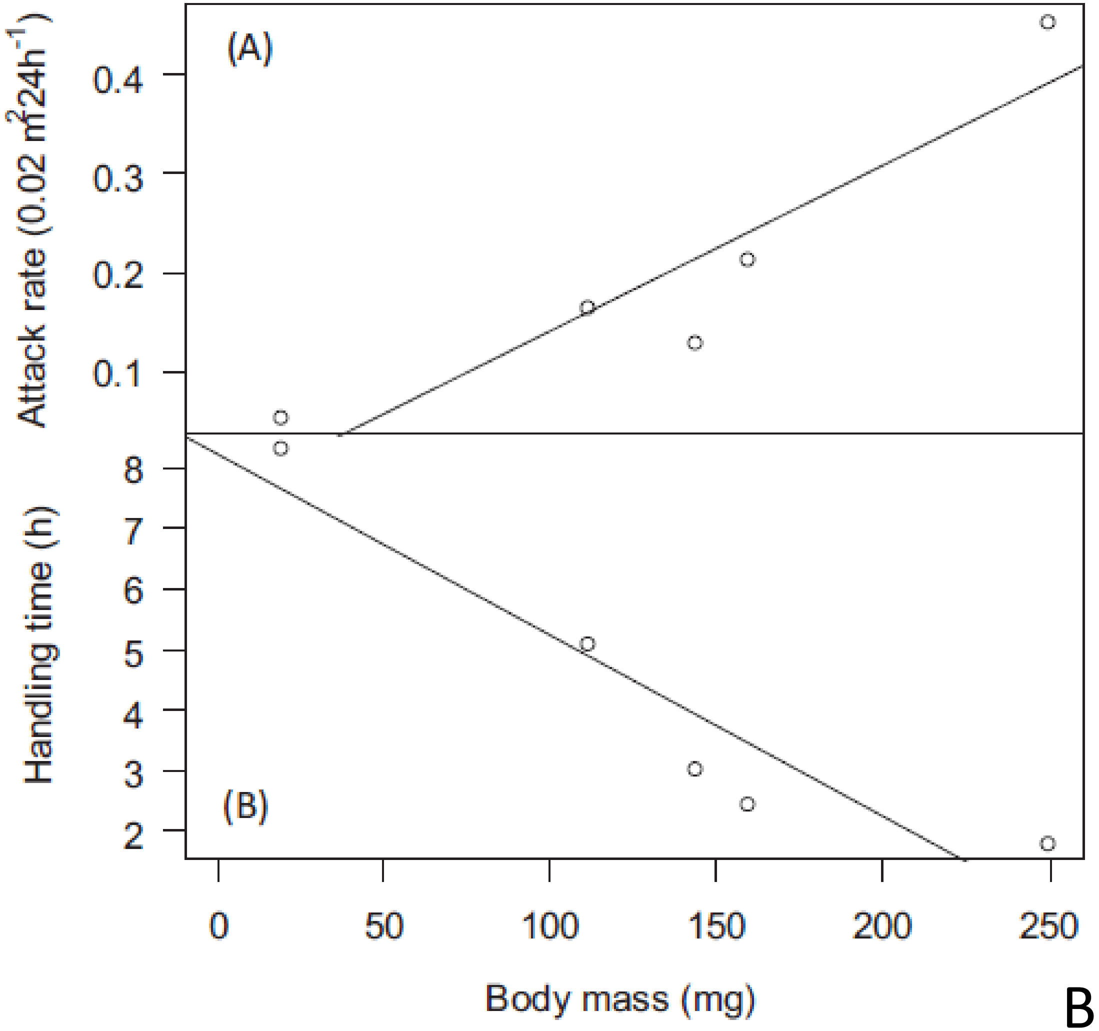
$$N_{killed} = \frac{aN}{1 + aHN}$$

a = attack rate, H = handling time, N = prey density

Maximum likelihood estimation

Ball et al. (2015). *Basic & Applied Ecology* **16**: 621-628

Applications of maximum likelihood estimation: inferring functional responses



Attack rate and handling time vary with predator body mass

Ball et al. (2015). *Basic & Applied Ecology* **16**: 621-628

Applications of maximum likelihood estimation: estimating occupancy and detection probability

Some species are very hard to detect

Might not show up during short surveys

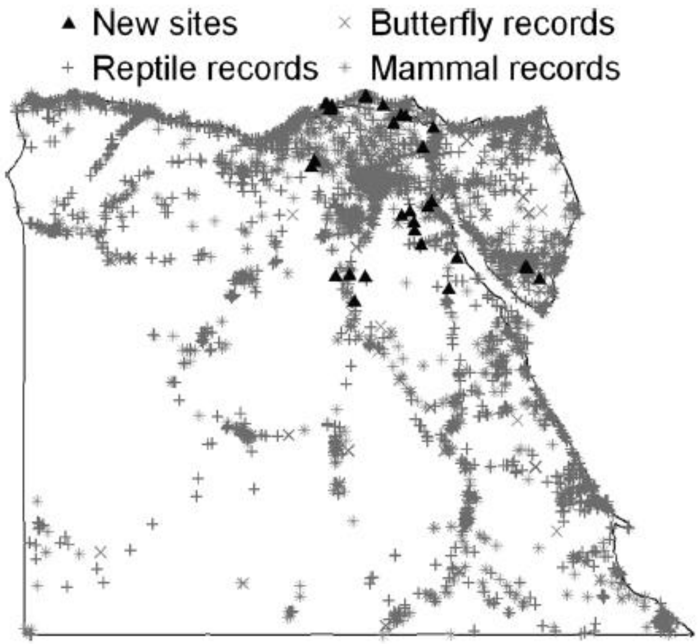
Can model separately probability of detection and probability of occupancy, given detection:

$$L = \left[\Psi^{n.} \prod_{t=1}^{T} p^{n_t} (1-p)^{n.-n_t}\right] \times \left[\Psi \prod_{t=1}^{T} (1-p) + (1-\Psi)\right]^{N-n.}$$

L = likelihood, Ψ = probability of occupancy, p = probability of detection in one visit, given occupancy, n. = number of sites with at least one detection, n_t = number of sites with detection on visit t, T = number of visits at each site

MacKenzie et al. (2002). *Ecology* **83**: 2248-2255

Applications of maximum likelihood estimation: estimating occupancy and detection probability



Assessing accuracy of species distribution models

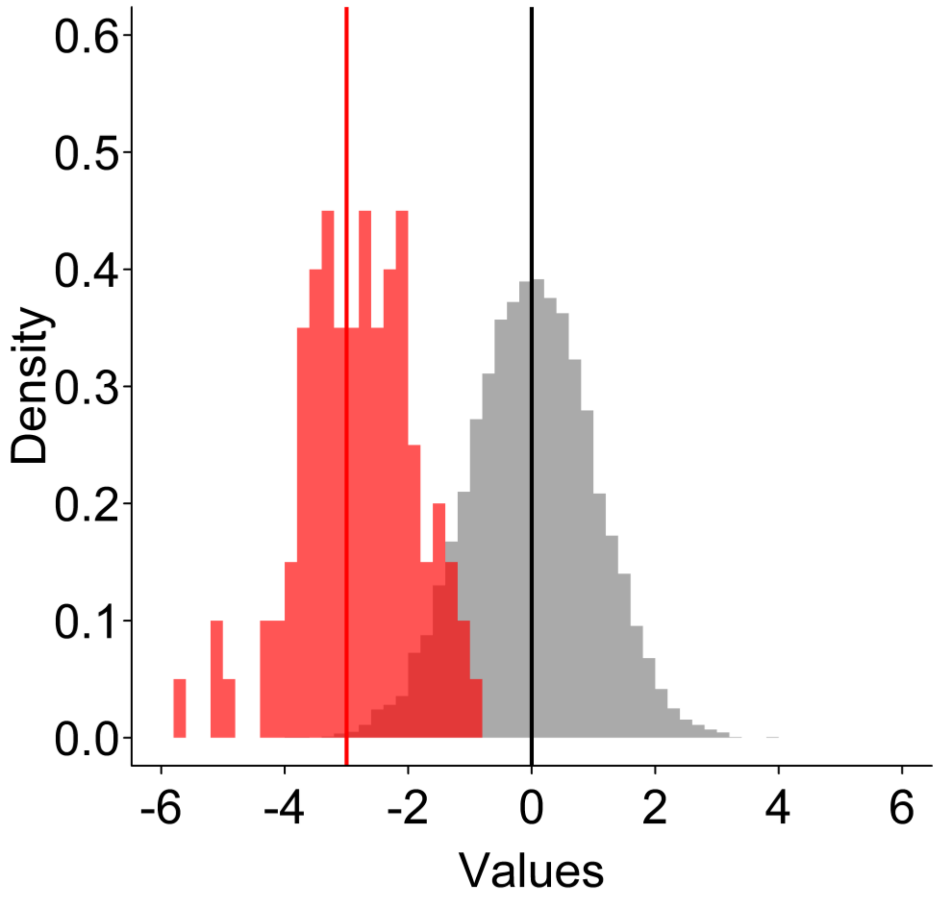
Surveyed 21 new sites, with 4 short transects per site

Modelled detectability and average occupancy probability

Parameters estimated with maximum likelihood estimation

Newbold et al. (2010). *Oikos* **119**: 1326-1334

Classical frequentist statistics



Null hypothesis: e.g. normal distribution with mean 0

Collect some observations to test our hypothesis – calculate likelihood ($\mathcal{L}$) of data given null hypothesis:

$$P(D|H_0)$$

Find parameters that maximise the (log) likelihood:

$$P(D|H)$$

Bayesian Statistics: Bayes' Rule

The Reverend Thomas Bayes

Try to find the probability of the hypothesis given the data, instead of the probability of the data given the hypothesis, i.e.:

$$P(H|D)$$

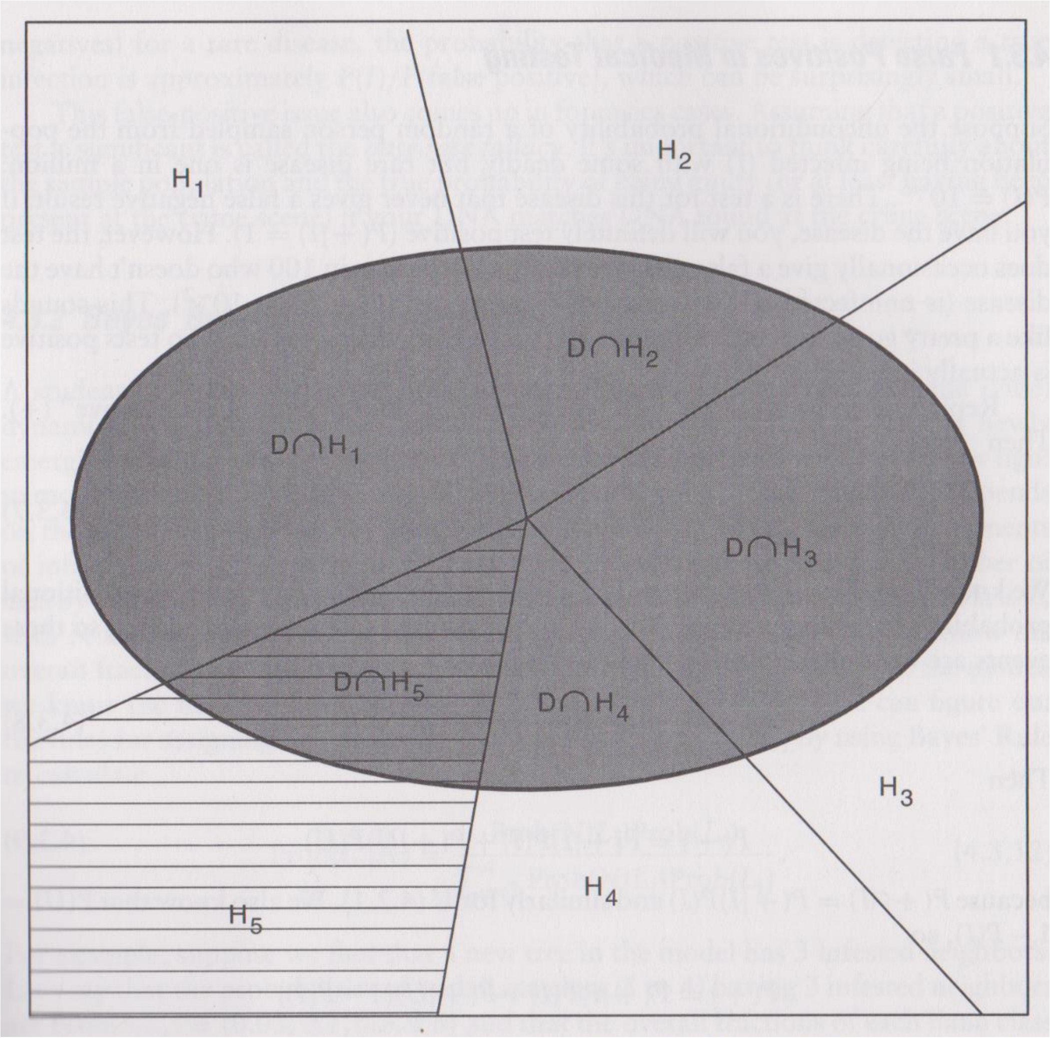
Likelihood ($\mathcal{L}$)

Bayes' Rule:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

?

Bayesian Statistics



Assume all possible hypotheses are known

P(D) is the sum of the dark grey areas, i.e.:

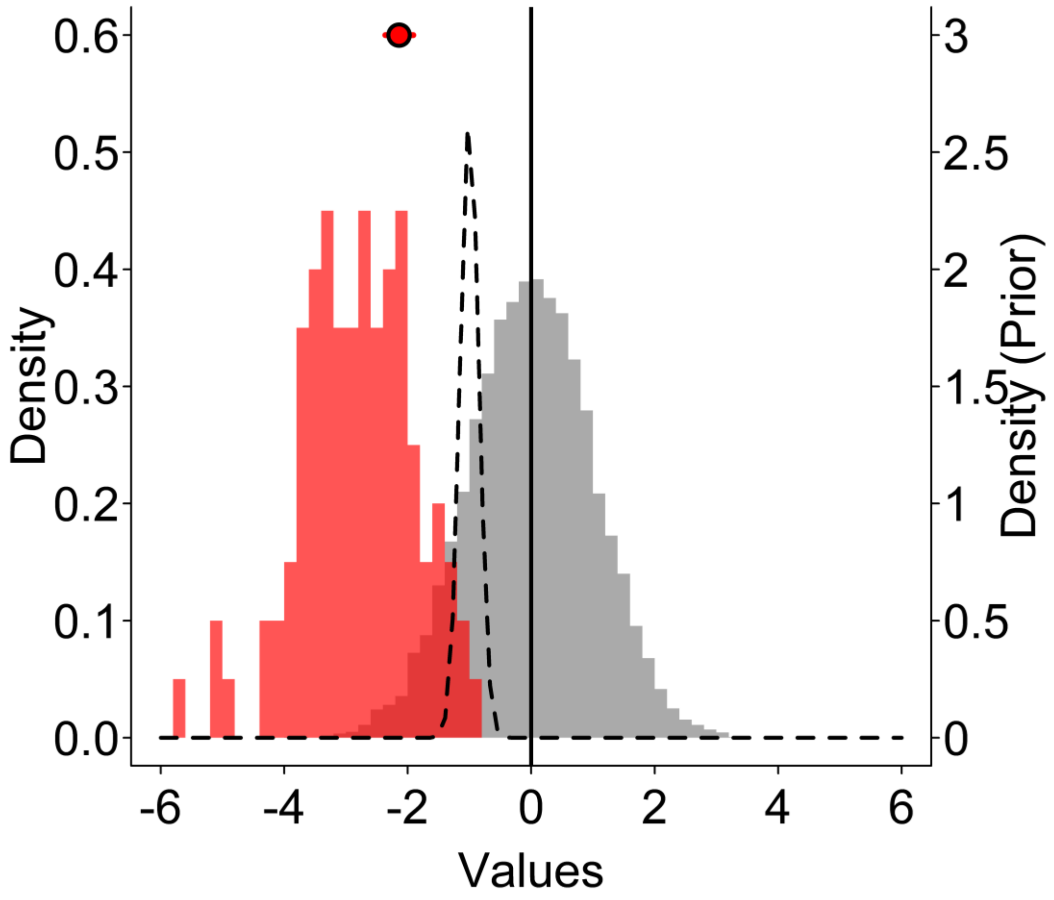
$$P(D) = \sum_{j=1}^{N} P\left(D \cap H_j\right)$$

$$P(D) = \sum_{j=1}^{N} P\left(D \middle| H_j\right) P(H_j)$$

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_{j=1}^{N} P\left(D \middle| H_j\right) P(H_j)}$$

Prior probabilities

Bayesian Statistics: Prior Probabilities



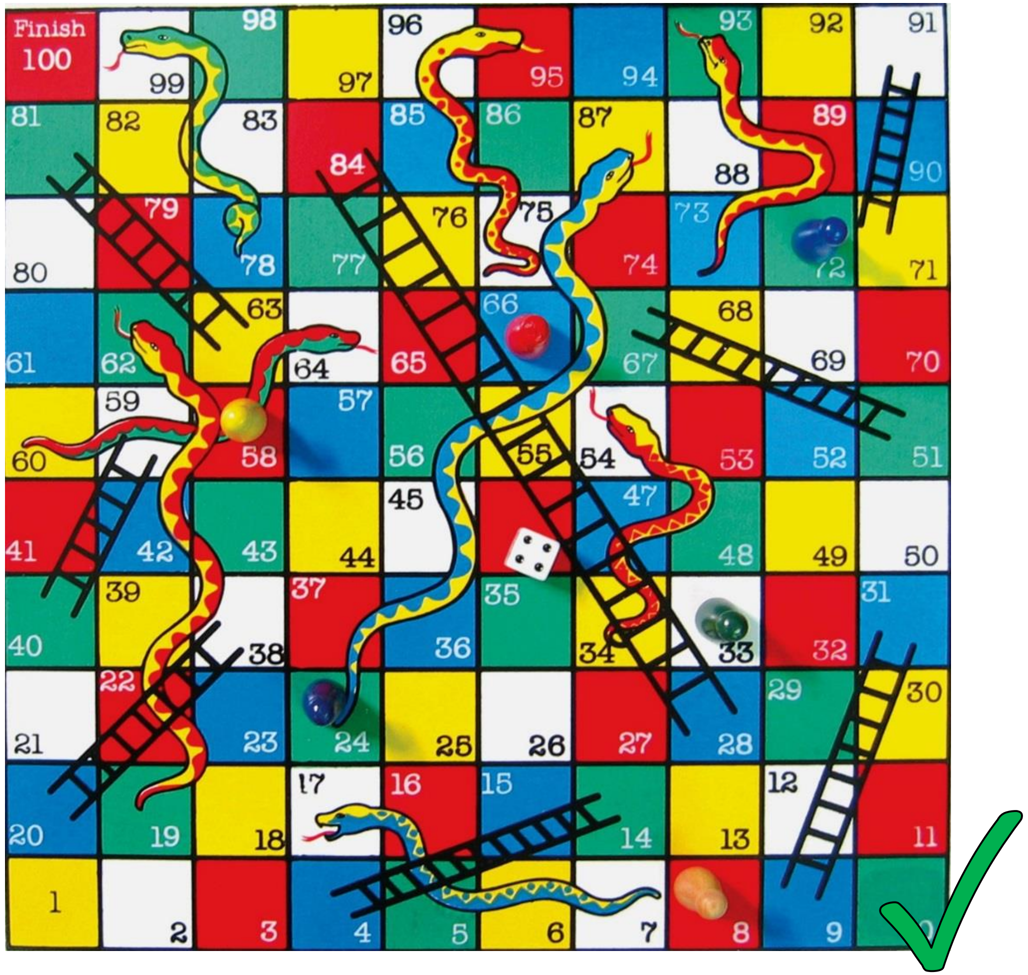
Often an 'uninformative' or 'flat' prior is used

Model-estimated mean is similar to before

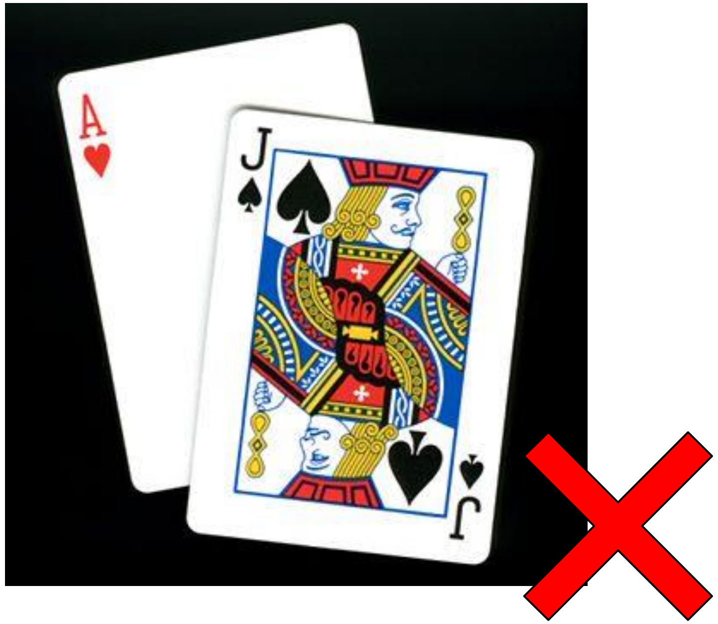
Or we can incorporate some prior knowledge/expectation

This shifts the model-estimated mean toward the prior distribution

Parameter sampling in Bayesian statistics: Markov Chain Monte Carlo



Markov process: transition probability depends only on system's current state, not on its history



Parameter sampling in Bayesian statistics: Markov Chain Monte Carlo

MCMC rule:

$$\frac{Post(A)}{Post(B)} = \frac{P(B \to A)P(accept\ A|B)}{P(A \to B)P(accept\ B|A)}$$

Ensures that parameter estimates reflect the posterior probability distribution, rather than honing in on the maximum-likelihood estimate

Example methods:

- Metropolis-Hastings
- Gibbs sampler

Bayesian and frequentist models compared

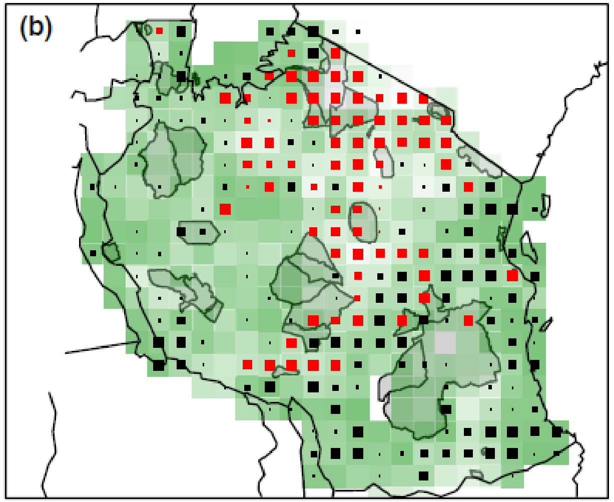
Models of ant species richness as a function of habitat, elevation and latitude

Table 4 Parameter estimates for the additive model (eqn 3) predicting ant species richness from habitat, elevation, and latitude

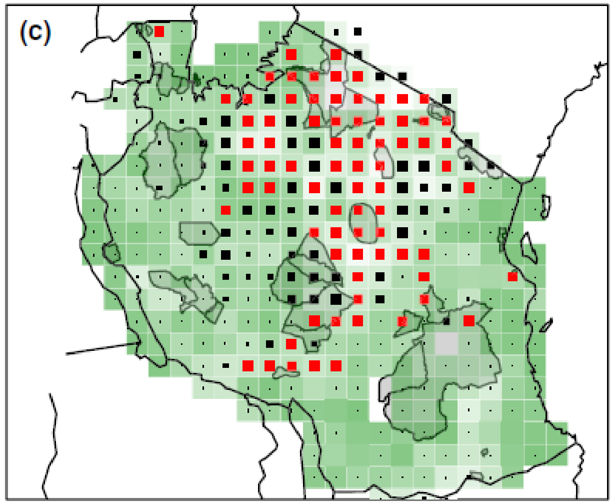
		Bayesian models		
	Classical model (maximum likelihood estimate)	Posterior mode, non-informative prior	Posterior mode, informative prior	Averaged model, non-informative prior
$\hat{\beta}_0$	11.95 (2.65) [6.81,17.73]	11.49 (1.87) [7.89, 15.32]	12.18 (2.22) [6.89, 16.33]	12.03 (2.65)
$\hat{\beta}_1$	−0.24 (0.06) [−0.36, −0.11]	−0.23 (0.04) [−0.31, −0.14]	−0.24 (0.05) [−0.33, −0.12]	−0.24 (0.06)
$\hat{\beta}_2$	−0.001 (0.0003) [−0.002, −0.0004]	−0.001 (0.0004) [−0.002, −0.0004]	−0.001 (0.0004) [−0.002, −0.0004]	−0.001 (0.0004)
$\hat{\beta}_3$	0.64 (0.06) [0.44, 0.75]	0.64 (0.12) [0.40, 0.88]	0.63 (0.12) [0.40, 0.84]	0.64 (0.12)

Ellison (2004). *Ecology Letters* **7**: 509-520

Applications of Bayesian modelling approaches: occupancy and detection again



Survey effort

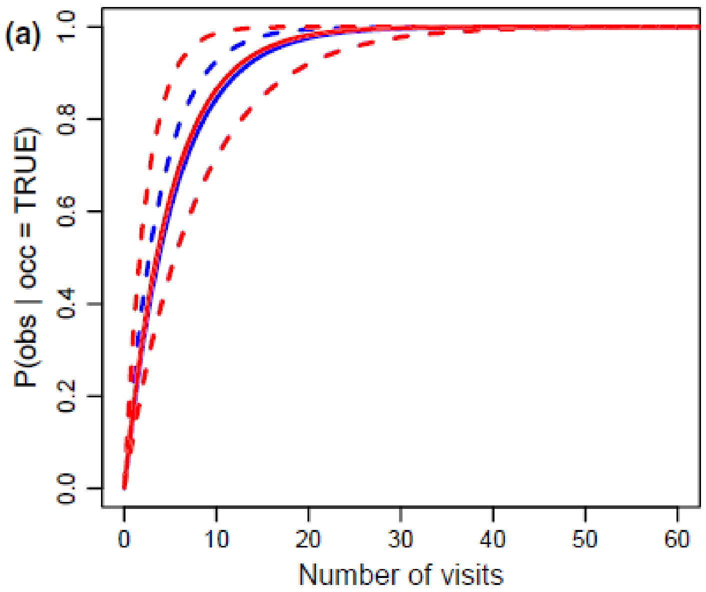


Probability of occupancy

Beale et al. (2013). *Ecology Letters* **16**: 1061-1068

Hierarchical occupancy model: probability of occupancy and probability of detection

Accounted for survey effort



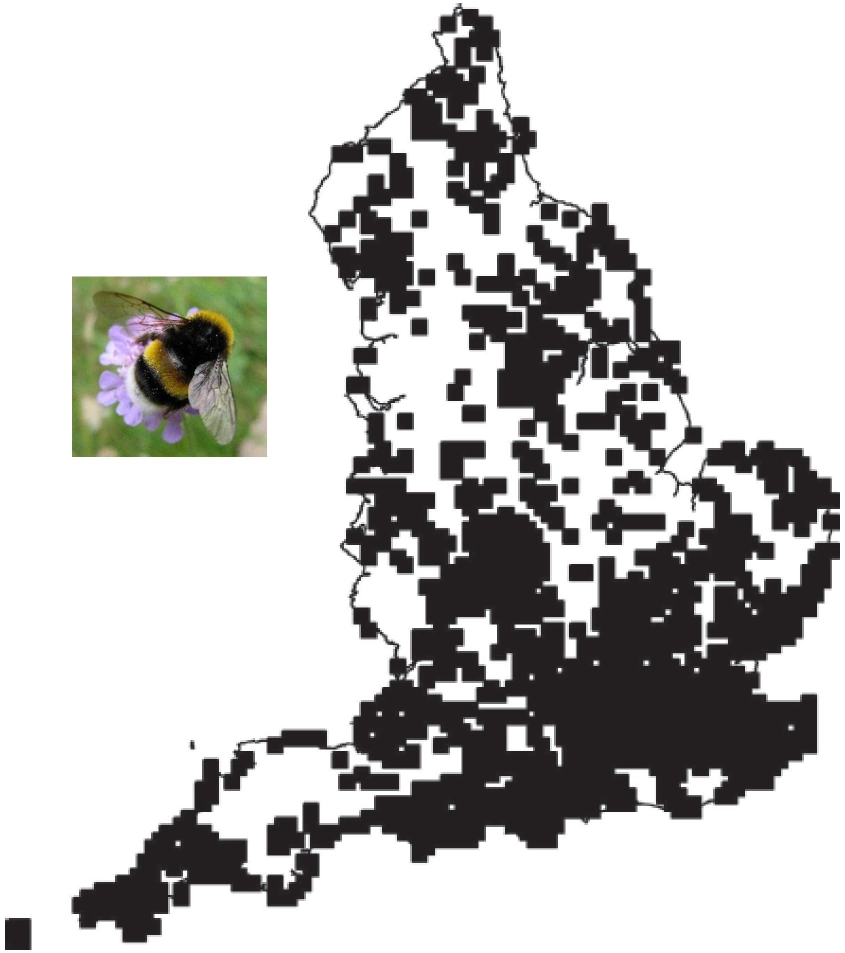
Applications of Bayesian modelling approaches: occupancy and detection again

Wild bee occurrence data from 1994 to 2010

Bayesian occupancy-detection model

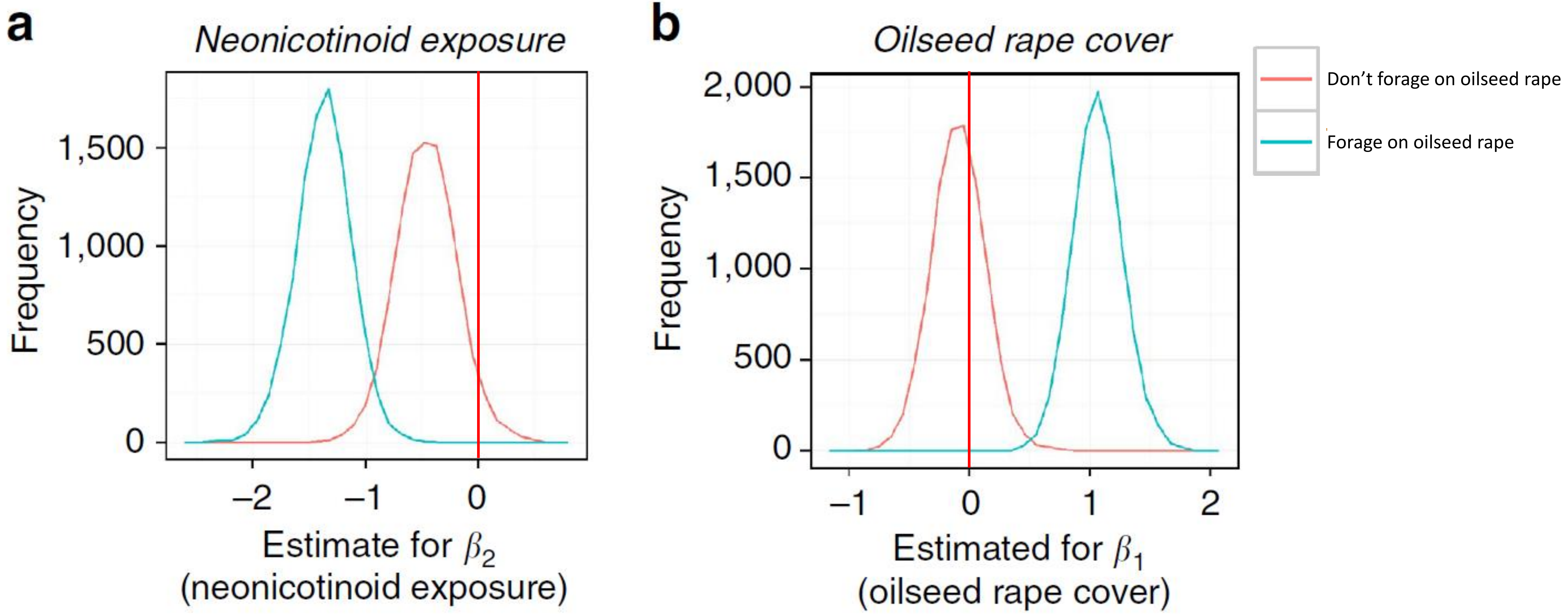
Probability of detection a function of survey effort (number of species recorded)

Persistence probability a function of oilseed rape cover and neonicotinoid exposure



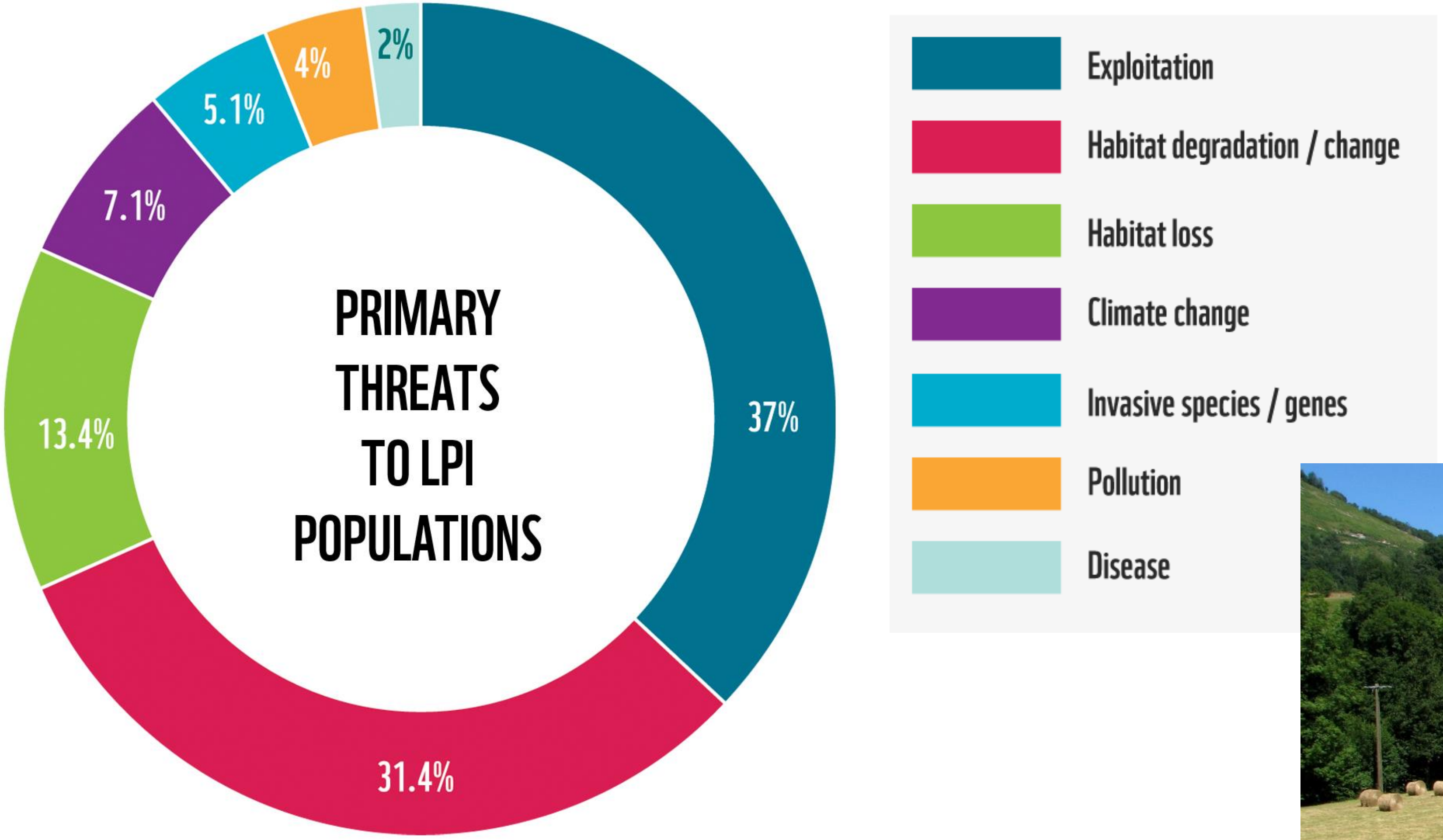
Woodcock et al. (2016). *Nature Communications* **7**: 12459

Applications of Bayesian modelling approaches: occupancy and detection again



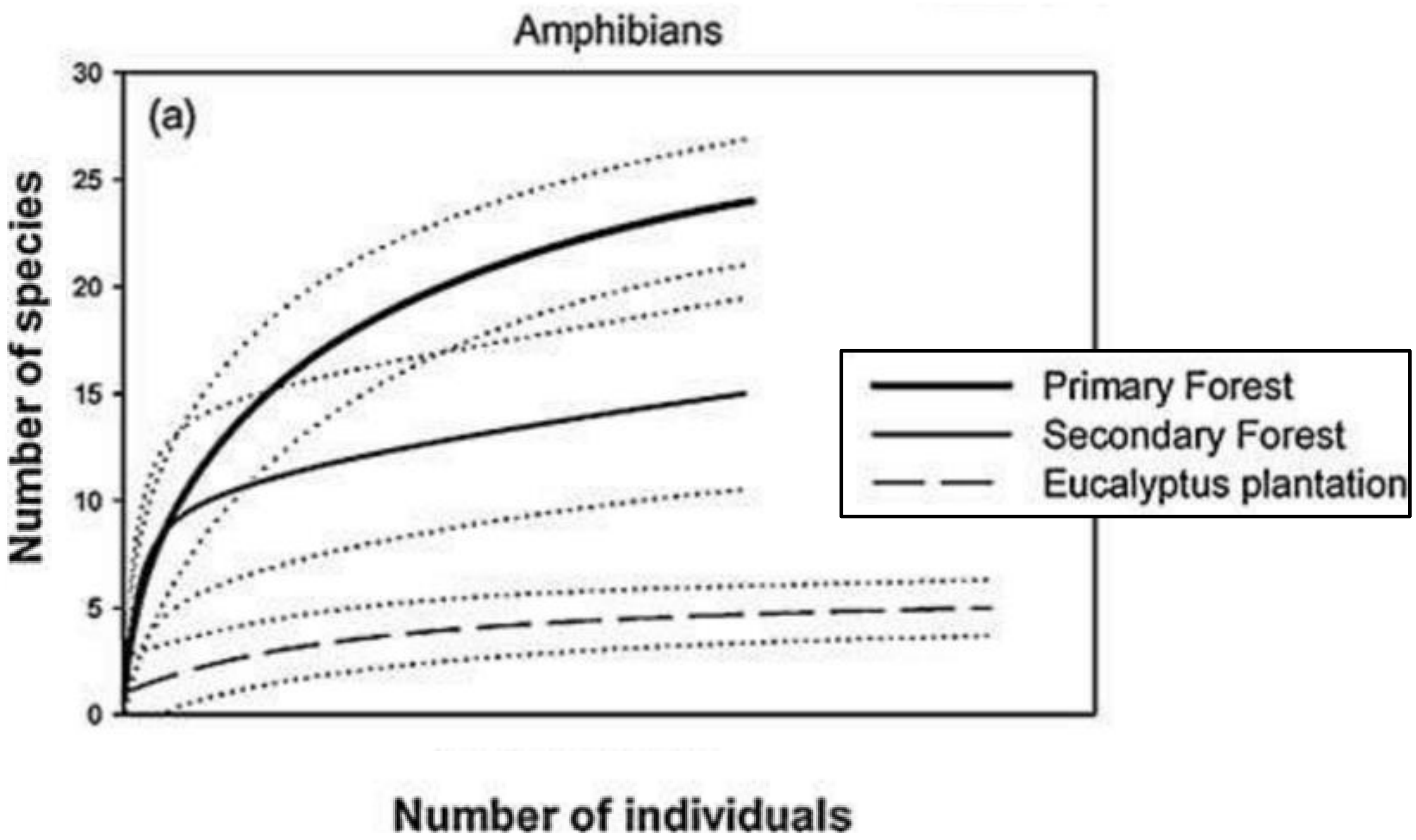
Woodcock et al. (2016). *Nature Communications* **7**: 12459

Land-use change presents the greatest threat to biodiversity currently



WWF (2014). *Living Planet Report*

There are lots of small-scale studies of land-use impacts on biodiversity



Gardner et al. (2007). *Conservation Biology* **21**: 775-787

Applications of Bayesian modelling approaches: impacts of land use on bird biodiversity

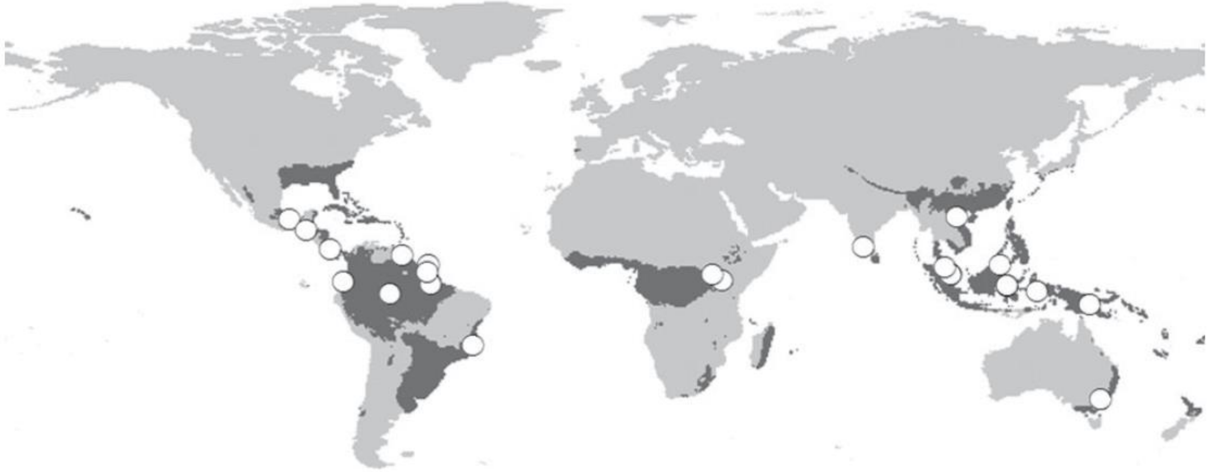
24 published studies of bird communities in tropical forest

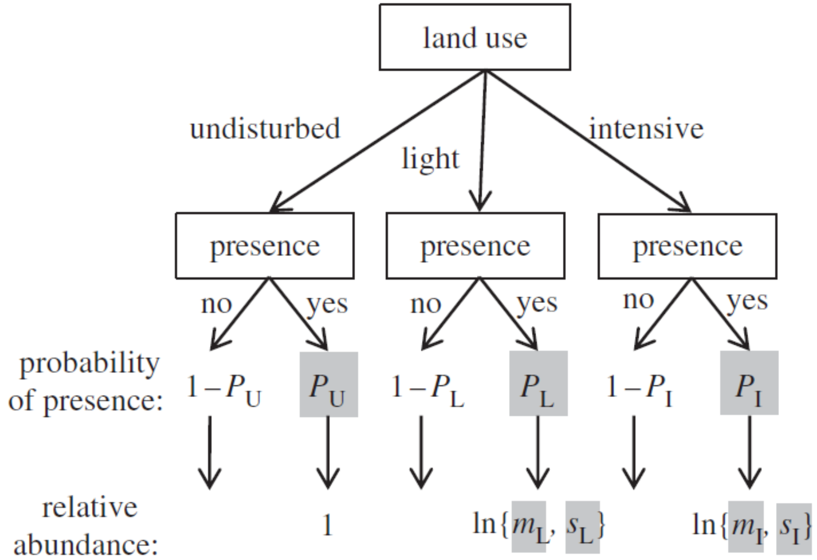
Three levels of land-use intensity (undisturbed, light, intensive)

Considered effects of species traits on response to land-use intensity

Lots of gaps in data, so user-defined likelihood function

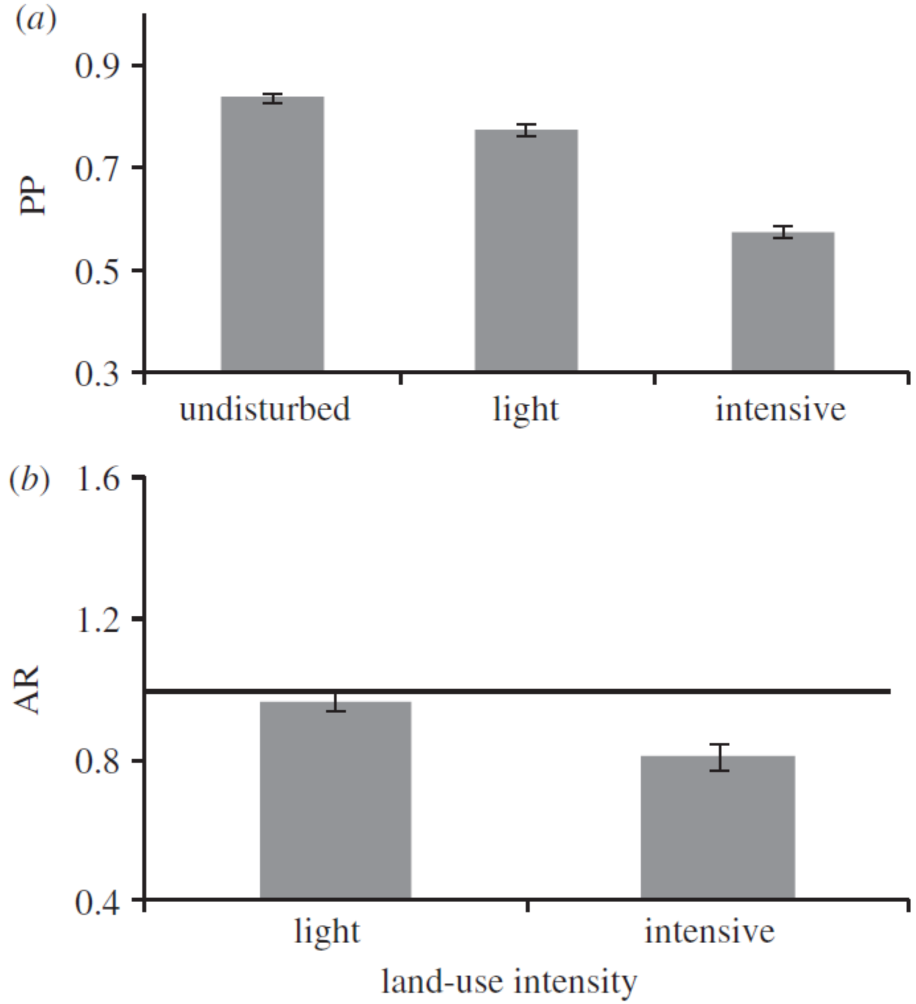
Bayesian approach useful, because of focus on prediction

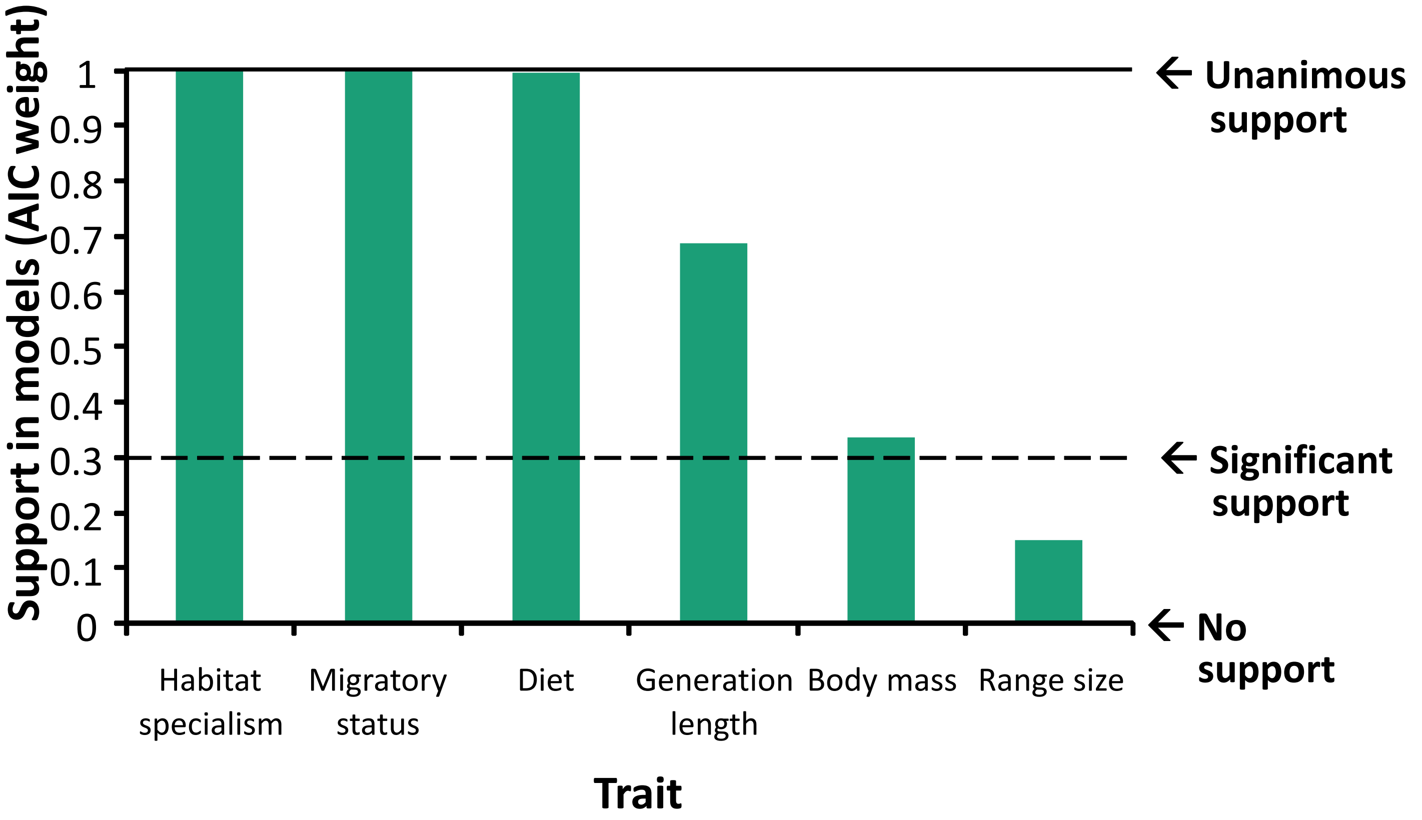




Newbold et al. (2013). *Proceedings of the Royal Society, Series B* **280**: 20122131

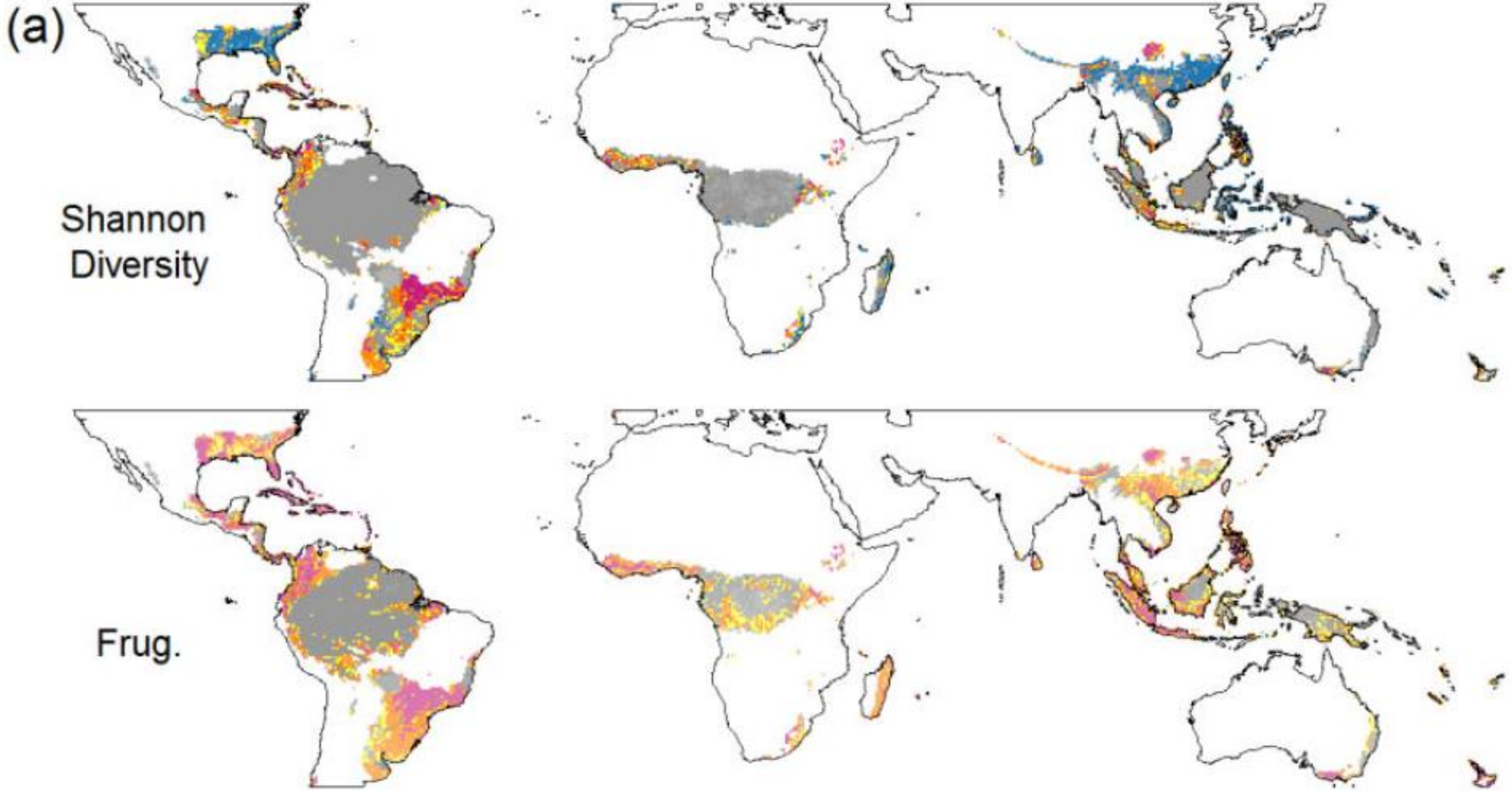
Applications of Bayesian modelling approaches: impacts of land use on bird biodiversity





Newbold et al. (2013). *Proceedings of the Royal Society, Series B* **280**: 20122131

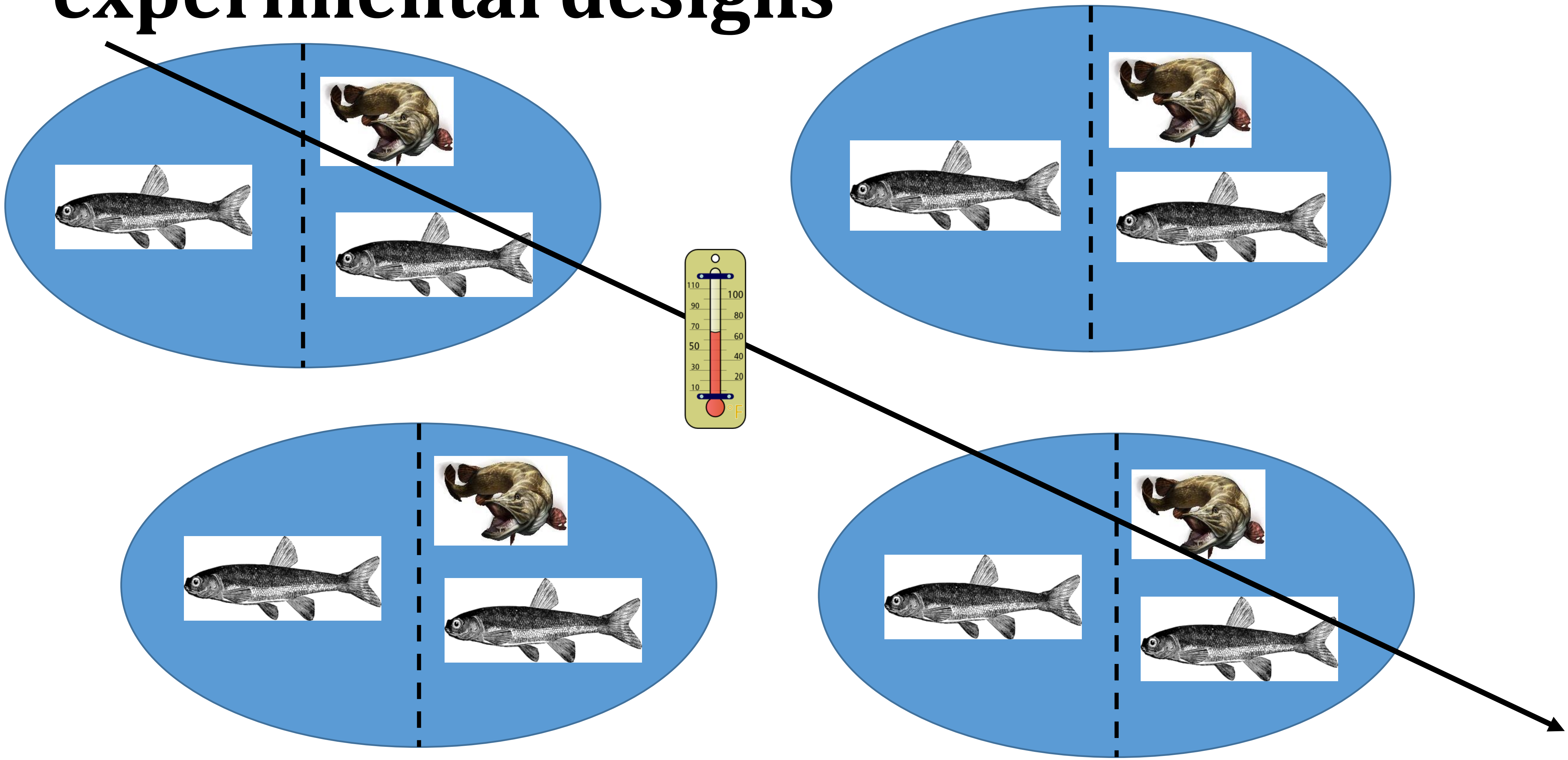
Applications of Bayesian modelling approaches: impacts of land use on bird biodiversity



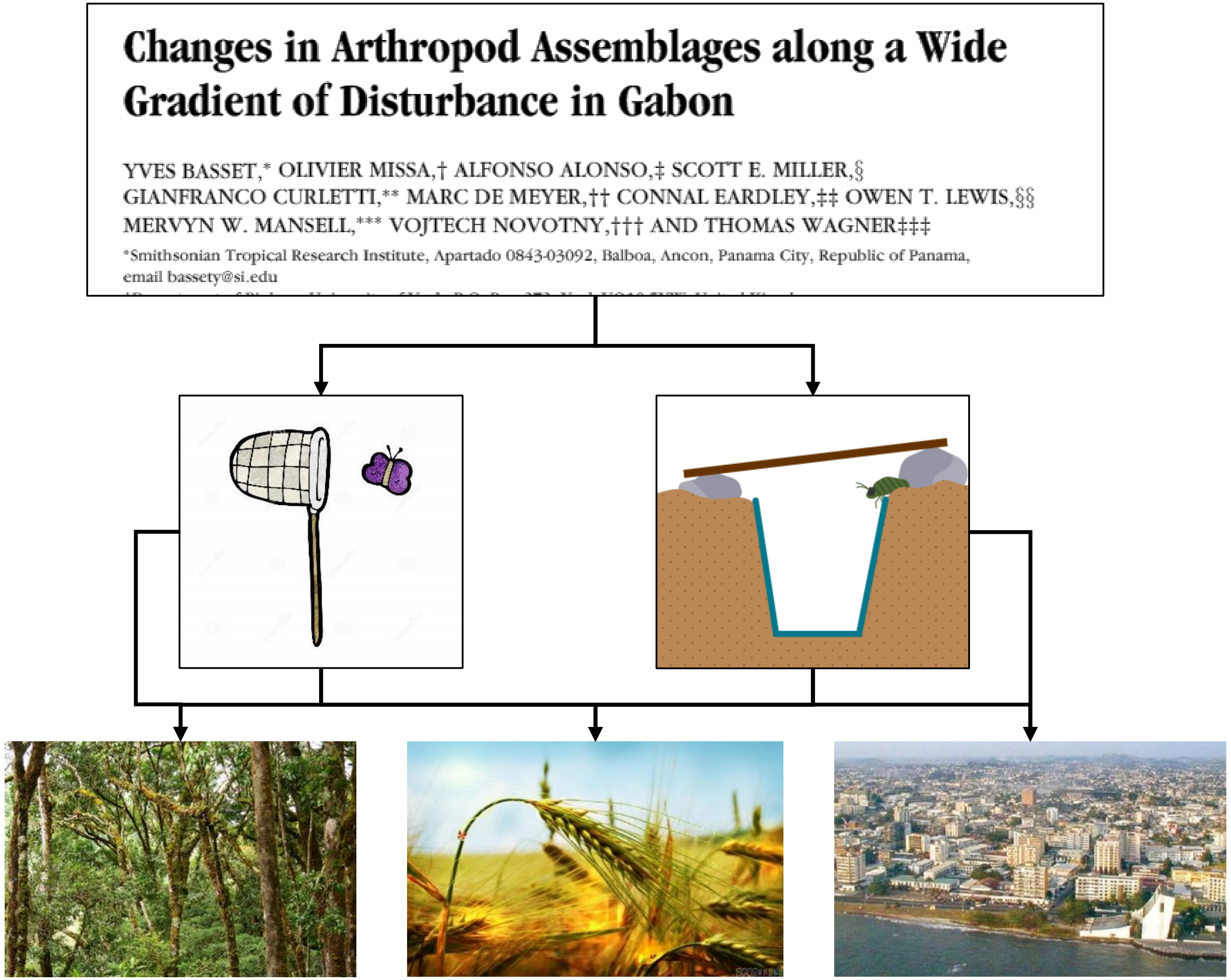
Application of models to predict historic changes in community diversity/structure

Newbold et al. (2014). *Global Ecology & Biogeography* **23**: 1073-1084

Hierarchical ecological data: nested experimental designs

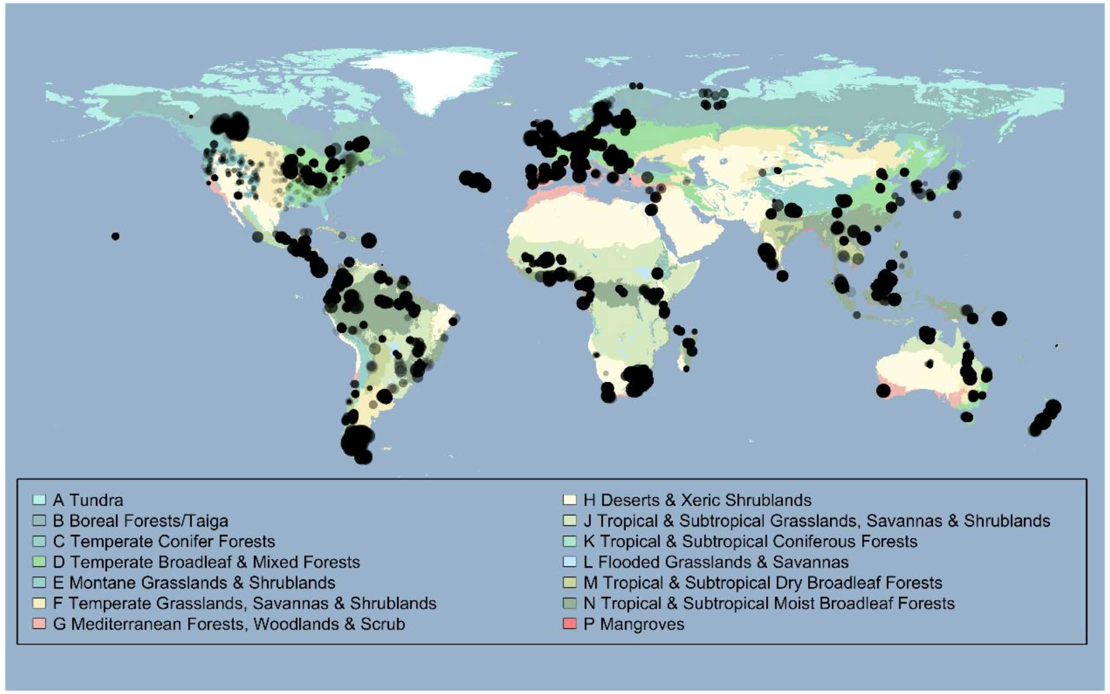


Hierarchical ecological data: synthetic studies



The PREDICTS database: hundreds of studies; many different sampling protocols

Hudson et al. (2014). *Ecology & Evolution* **4**: 4701-4735



Solutions for hierarchical data

Separate model for each level in the hierarchy – but loss of statistical power and generality

Effect of hierarchical structure in model:

$$SR_{site} \sim LandUse_{site}$$
$$SR_{site} \sim LandUse_{site} + Study_{site}$$

But this often introduces many parameters

Mixed effects models

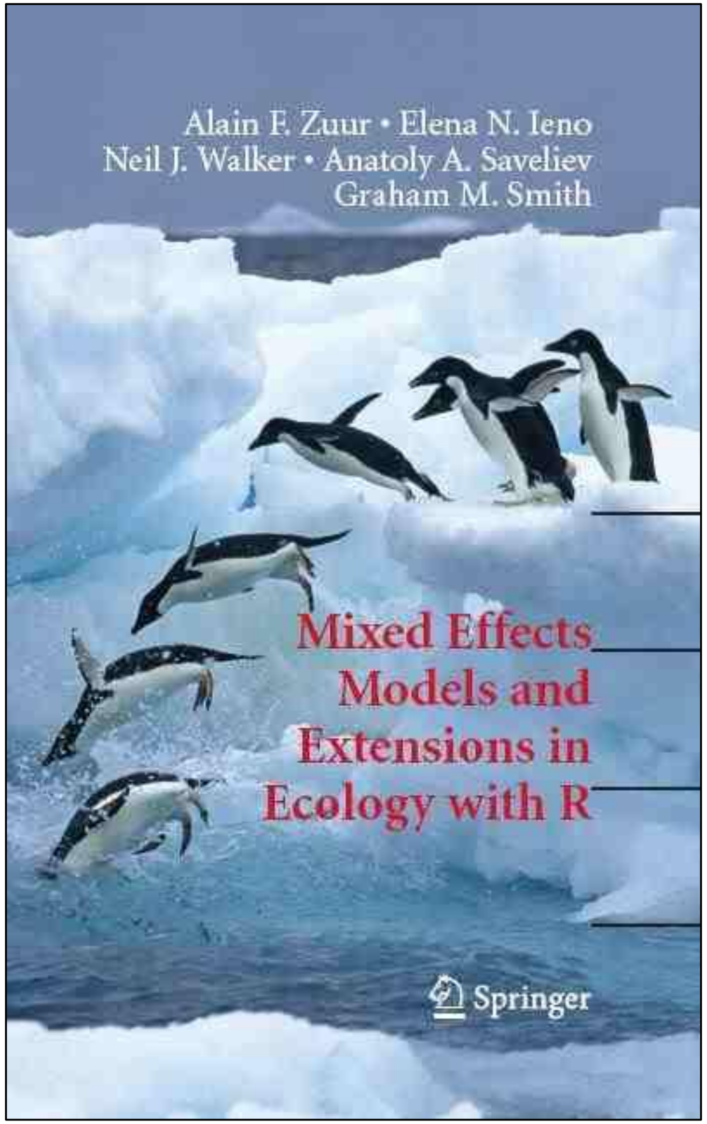
Mixed-effects models: random intercepts

Mixed-effects models composed of fixed effects and random effects

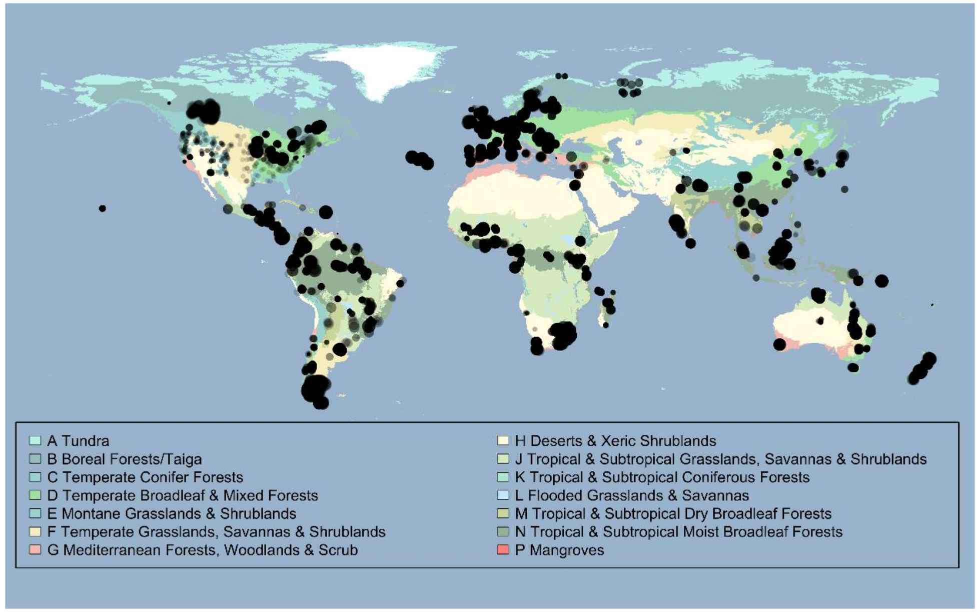
Fixed effects are those you are interested in (e.g. land use); assumed to represent a finite sample of the population

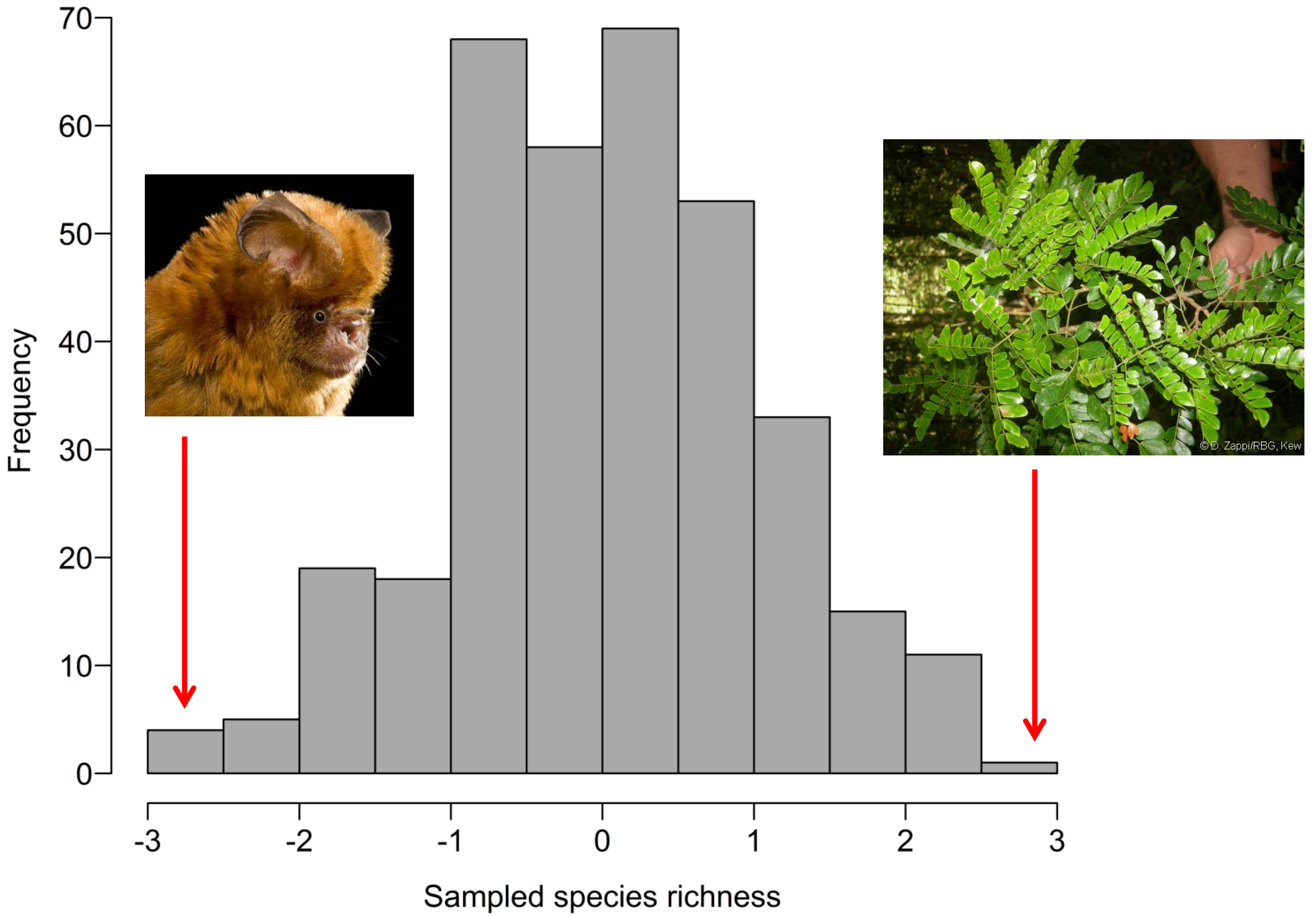
Random effects describe important variation, but parameter estimates not needed for each level (e.g. study); assumed to be representative of the super-population:

$$Study \sim N(0, \sigma_{Study})$$



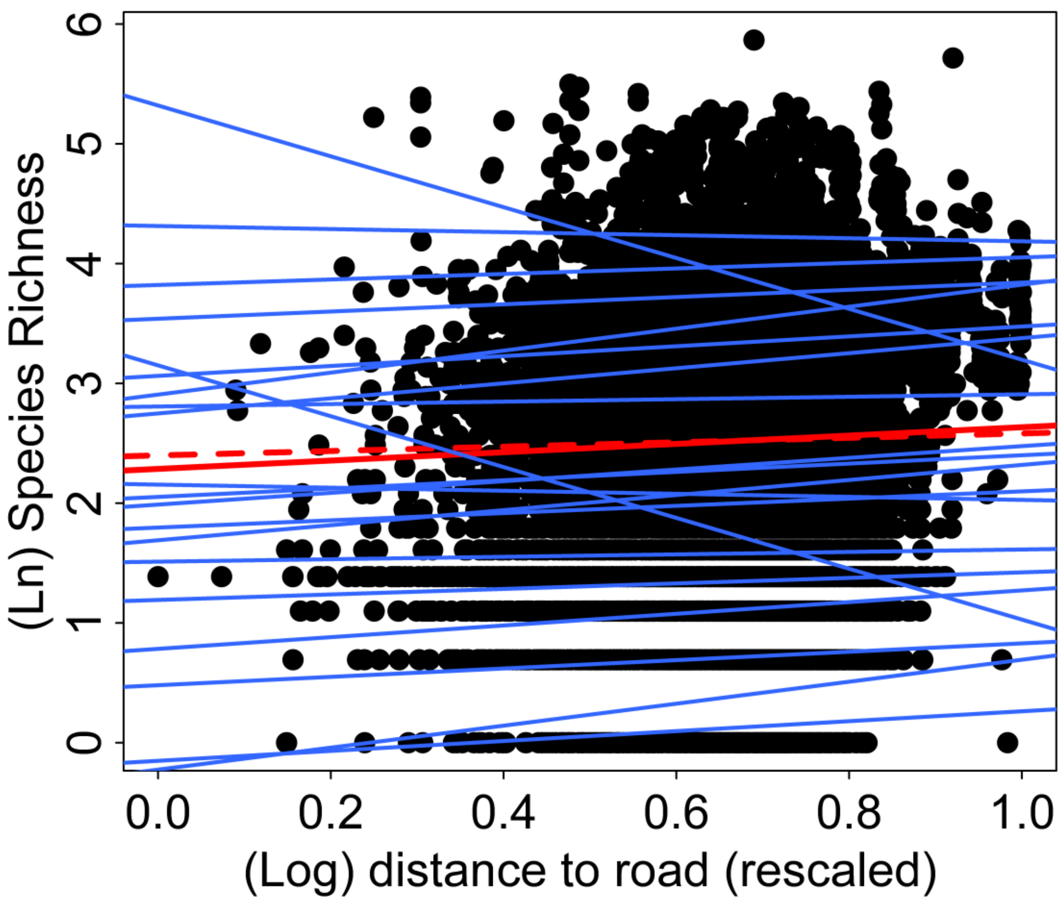
Mixed-effects models: random intercepts





Mixed-effects models: random slopes

Sometimes a relationship between two variables varies among the levels in the sampling hierarchy



In this case, we can fit random slopes:

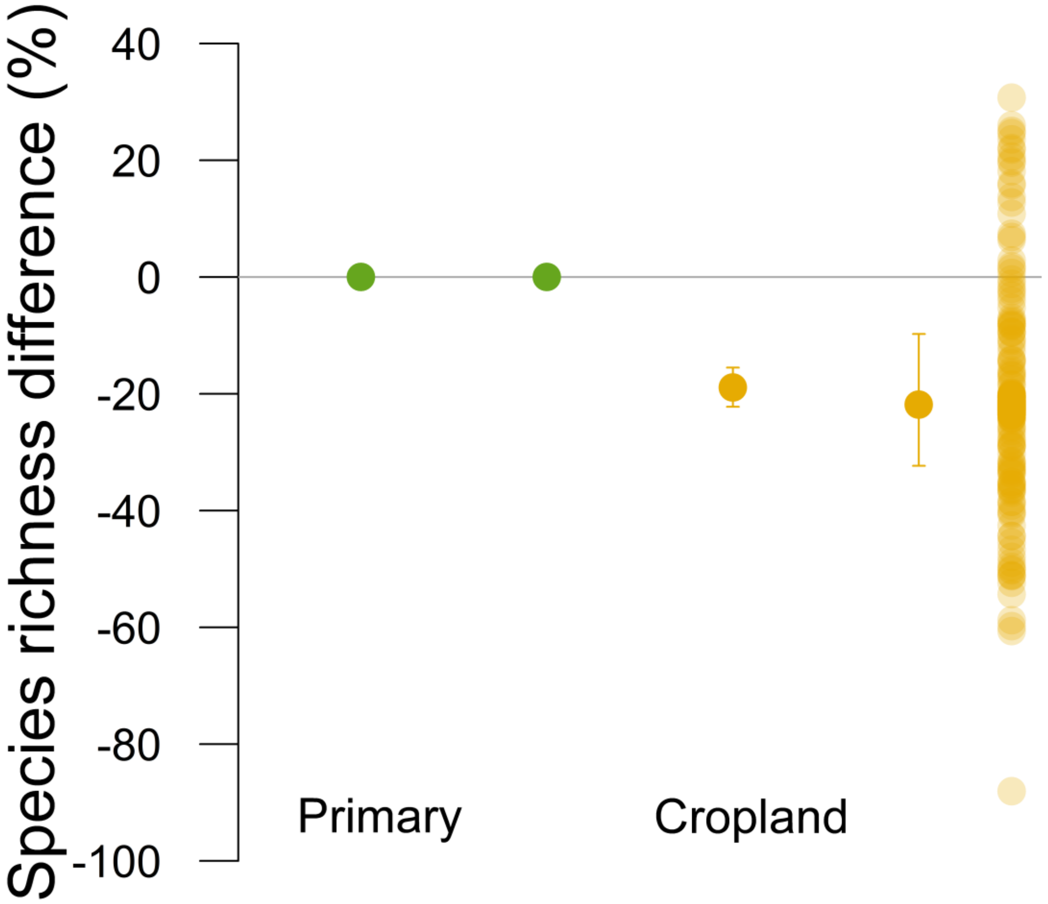
$$Slope \sim N(0, \sigma_{Slope})$$

Mixed-effects models: random slopes

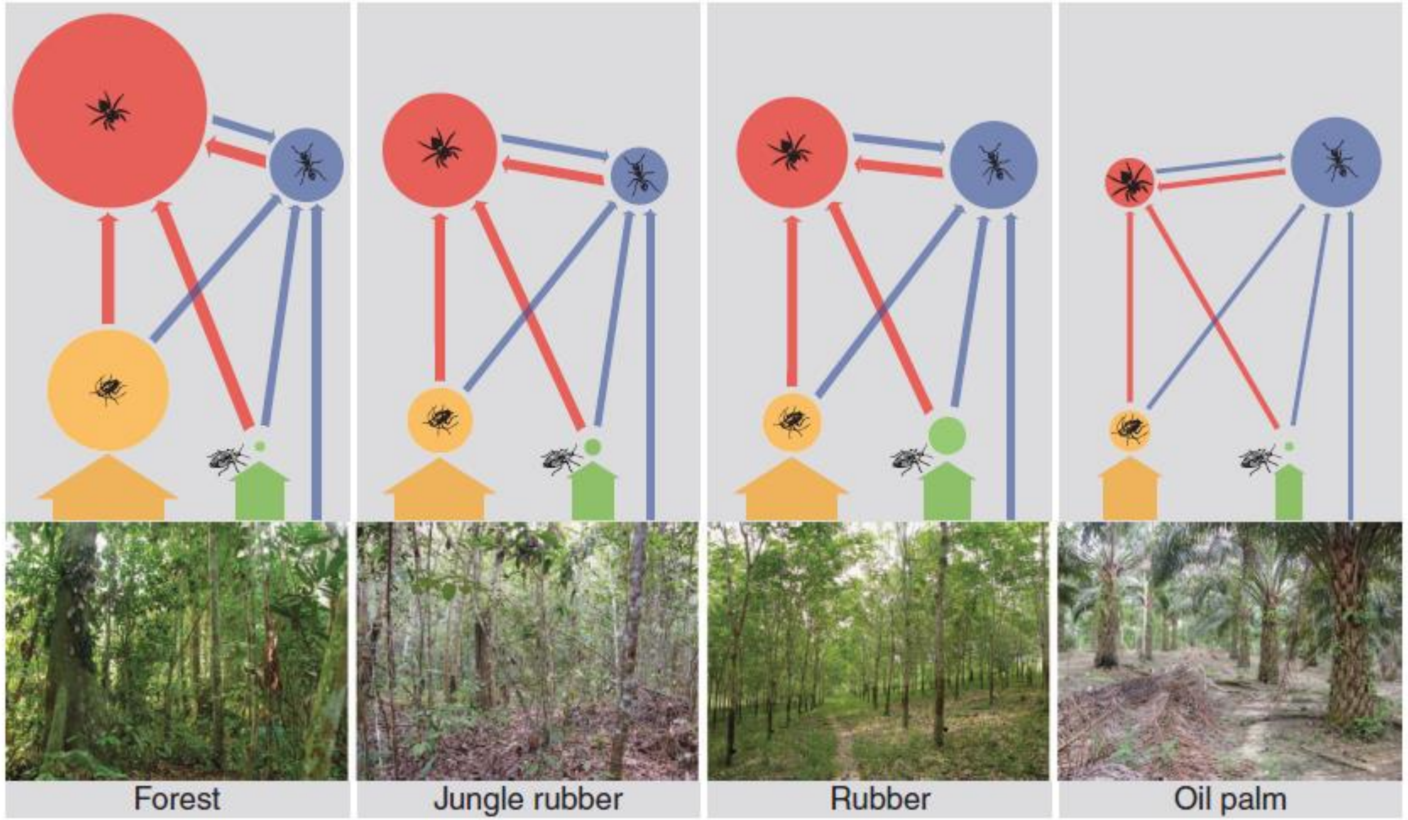
Random ‘slopes’ can also be used to describe variation in the effect of a categorical variable

This time there is one set of random ‘slopes’ for each factor level in the model

$$Slope \sim N(0, \sigma_{Slope})$$



Applications of mixed-effects models: effects of land use on biodiversity



Herbivores, Omnivores, Predators, Detritivores

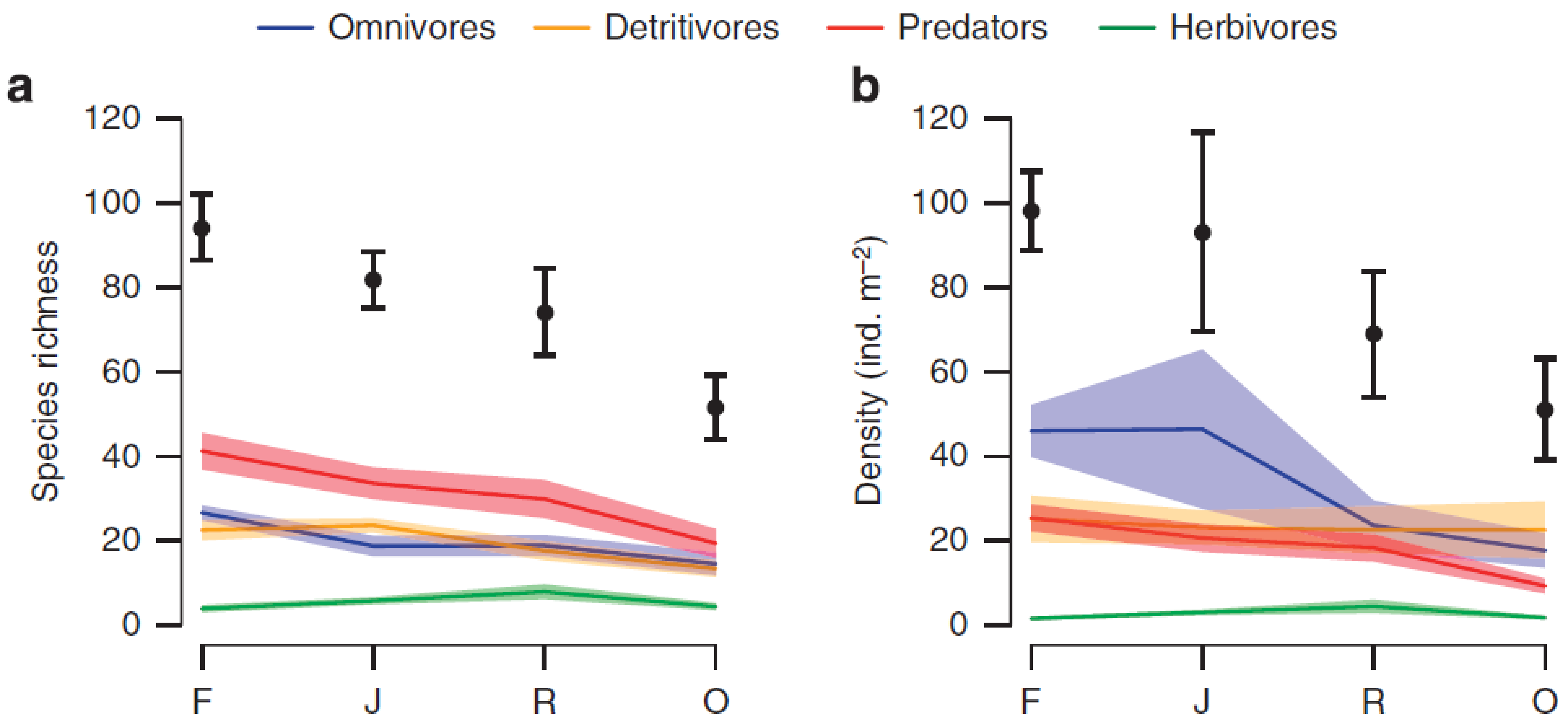
Sampled four trophic levels in four land uses in two landscapes

Random effects: landscape

Fixed effects: land use and trophic level and interaction

Barnes et al. (2014). *Nature Communications* **5**: 5351

Applications of mixed-effects models: effects of land use on biodiversity



F = Forest; J = Jungle rubber; R = Rubber plantation; O = Oil palm

Barnes et al. (2014). *Nature Communications* **5**: 5351

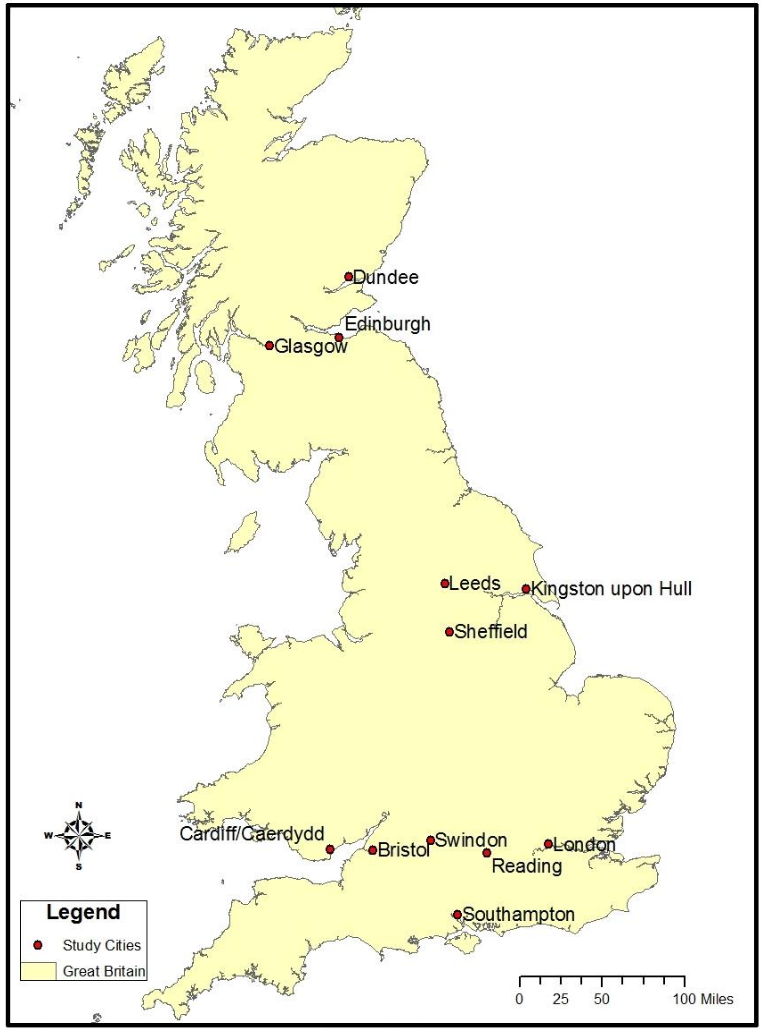
Applications of mixed-effects models: effects of land use on biodiversity

Sampled pollinator communities in 36 sites, in or around 12 UK cities

Sampled abundance and species richness

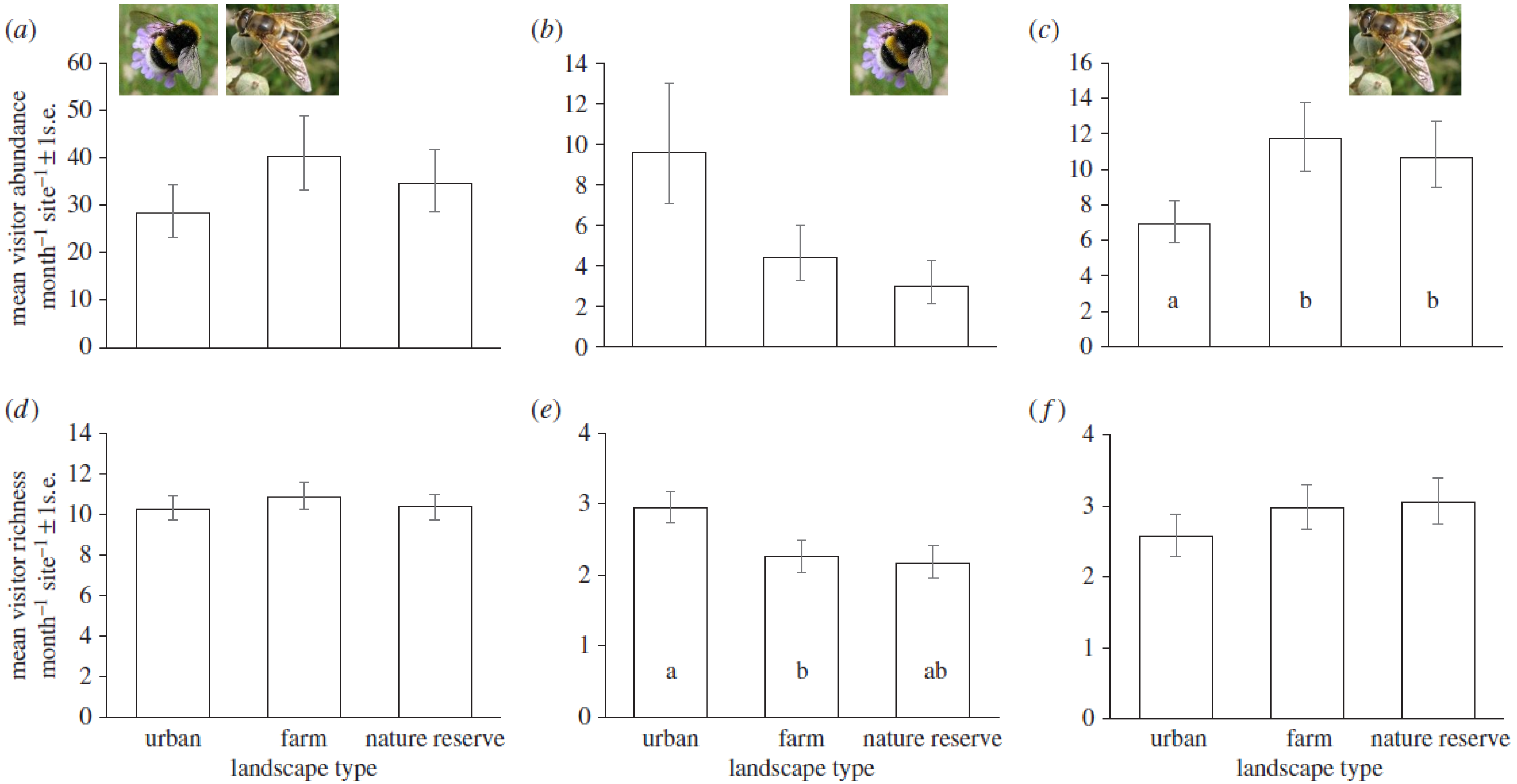
Fixed effects: land use (urban, agriculture and nature reserve), month, flower abundance, woodland cover

Random effects: Site nested within city



Baldock et al. (2015). *Proceedings of the Royal Society, Series B* **282**: 20142849

Applications of mixed-effects models: effects of land use on biodiversity



Baldock et al. (2015). *Proceedings of the Royal Society, Series B* **282**: 20142849

Applications of mixed-effects models: effects of land use on biodiversity

The Value of Primary, Secondary, and Plantation Forests for a Neotropical Herpetofauna

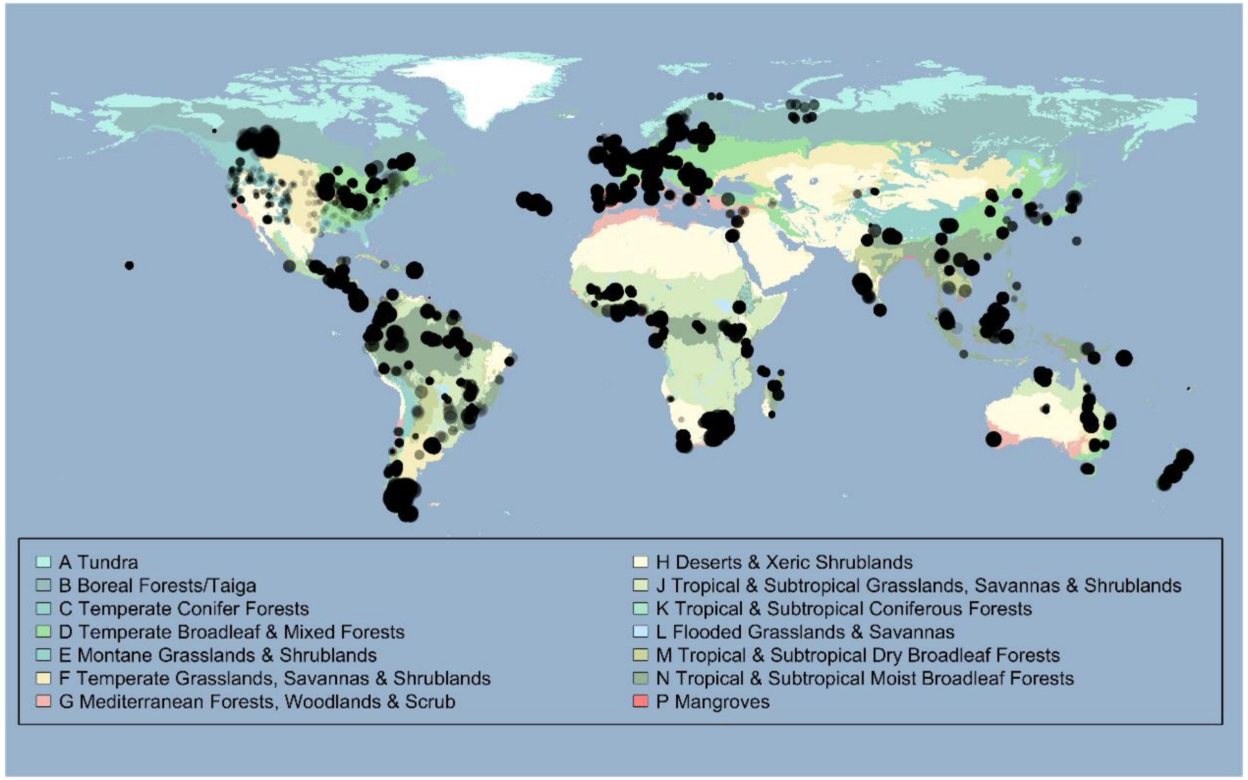
TOBY A. GARDNER,*‡ MARCO ANTÔNIO RIBEIRO-JÚNIOR,† JOS BARLOW,*† TERESA CRISTINA SAUER ÁVILA-PIRES,† MARINUS S. HOOGMOED,† AND CARLOS A. PERES*

*School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, United Kingdom
†Museu Paraense Emilio Goeldi (MPEG)/CZO and Programa de Posgraduação em Zoologia UFPA-MPEG, Caixa Postal 399, 66017-970, Belém, PA, Brazil

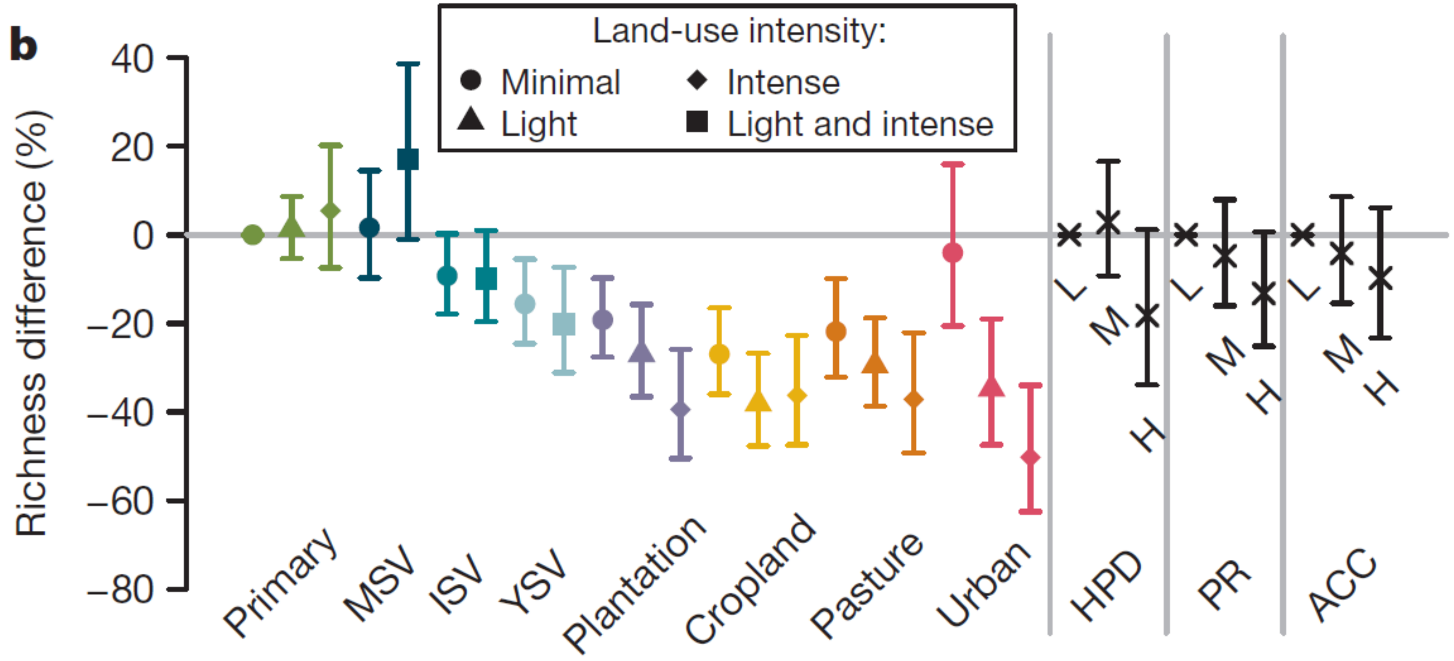
Abstract: *forest land*
options for
ondary, an
four comp
leaf-litter a
forest types
structure I

APPENDIX. Species of amphibian and lizard caught by standardized trapping methods in the Jari landscape, northeastern Brazilian Amazonia.

Species	Code for Figures 3 & 4	Family	Eucalyptus	Secondary Forest	Primary Forest	Total number captured	Microhabitat
Amphibia							
Atelopus spumarius	V	Bufonidae			1	1	Leaf litter
Bufo guttatus	S	Bufonidae	4	2	2	8	Leaf litter
Bufo margaritifer	L	Bufonidae		56	5	61	Leaf litter
Bufo marinus	K	Bufonidae	9	29	5	43	Leaf litter
Bufo sp.	B	Bufonidae	3	51	74	128	Leaf litter
Colostethus sp.	D	Dendrobatidae		1	30	31	Leaf litter
Dendrobates tinctorius	J	Dendrobatidae			6	6	Leaf litter
Epipedobates femoralis	E	Dendrobatidae		5	24	29	Leaf litter
Epipedobates hahneli	C	Dendrobatidae		9	55	64	Leaf litter
Adenomera sp.	A	Leptodactylidae	697	194	265	1156	Leaf litter
Eleutherodactylus chiastonotus	G	Leptodactylidae			8	8	Leaf litter
Eleutherodactylus marmoratus	R	Leptodactylidae			2	2	Leaf litter
Eleutherodactylus zeuctotylus	Q	Leptodactylidae		1	2	3	Leaf litter
Leptodactylus knudseni	N	Leptodactylidae	1	32	3	36	Leaf litter



Applications of mixed-effects models: effects of land use on biodiversity



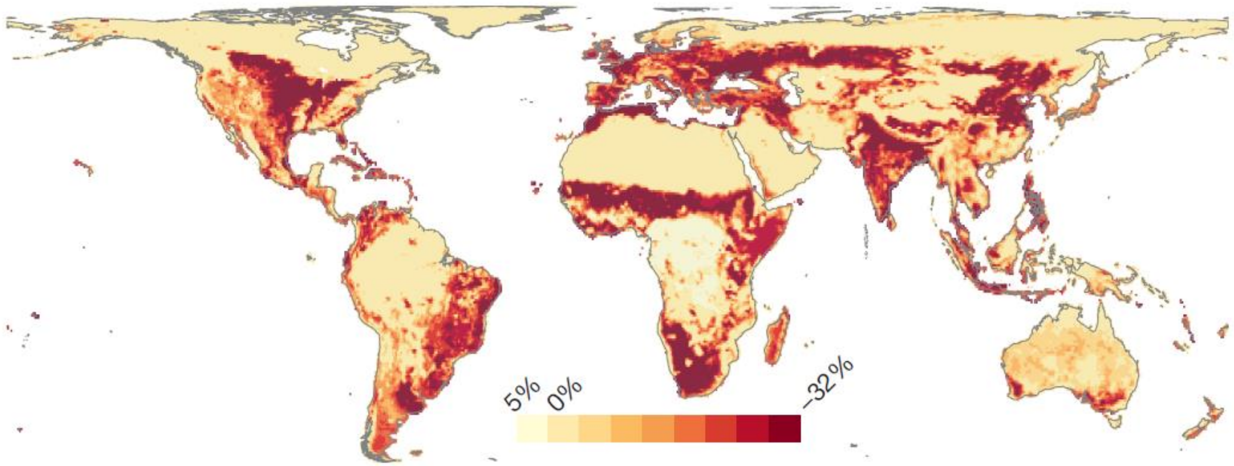
Response variables: various measures of local biodiversity

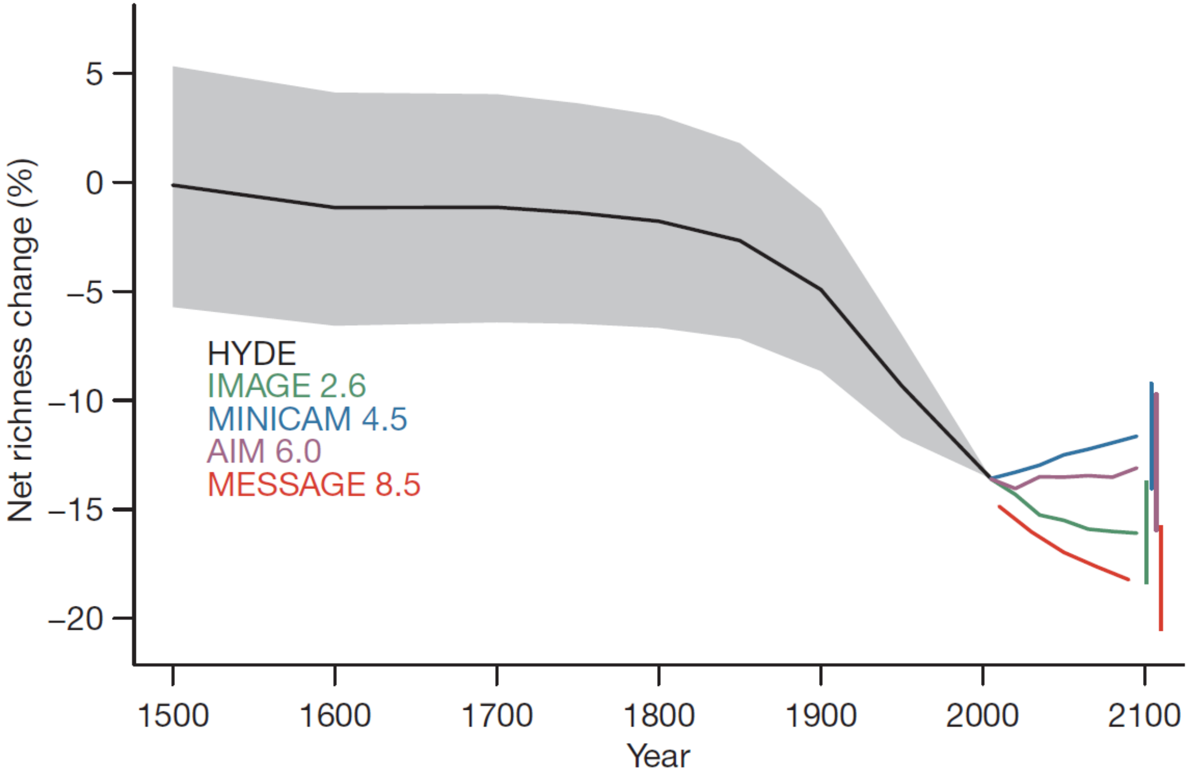
Fixed effects: land use, human population density (HPD), proximity to roads (PR), accessibility to towns/cities (ACC)

Random effects: study identity (differences in sampling), spatial structure of sampling within studies

Newbold et al. (2015). *Nature* **520**: 45-50

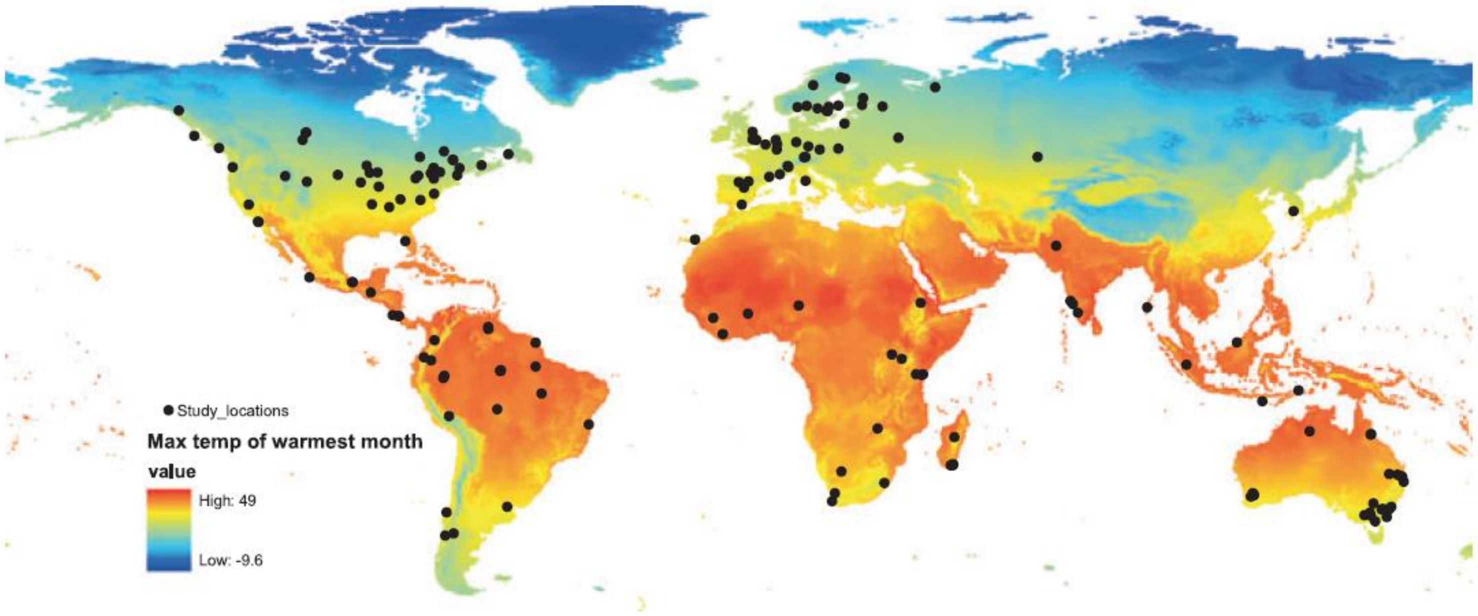
Applications of mixed-effects models: effects of land use on biodiversity





Newbold et al. (2015). *Nature* **520**: 45-50

Applications of mixed-effects models: effects of climate on responses of biodiversity to land use



Data on effects of habitat loss on biodiversity from published papers

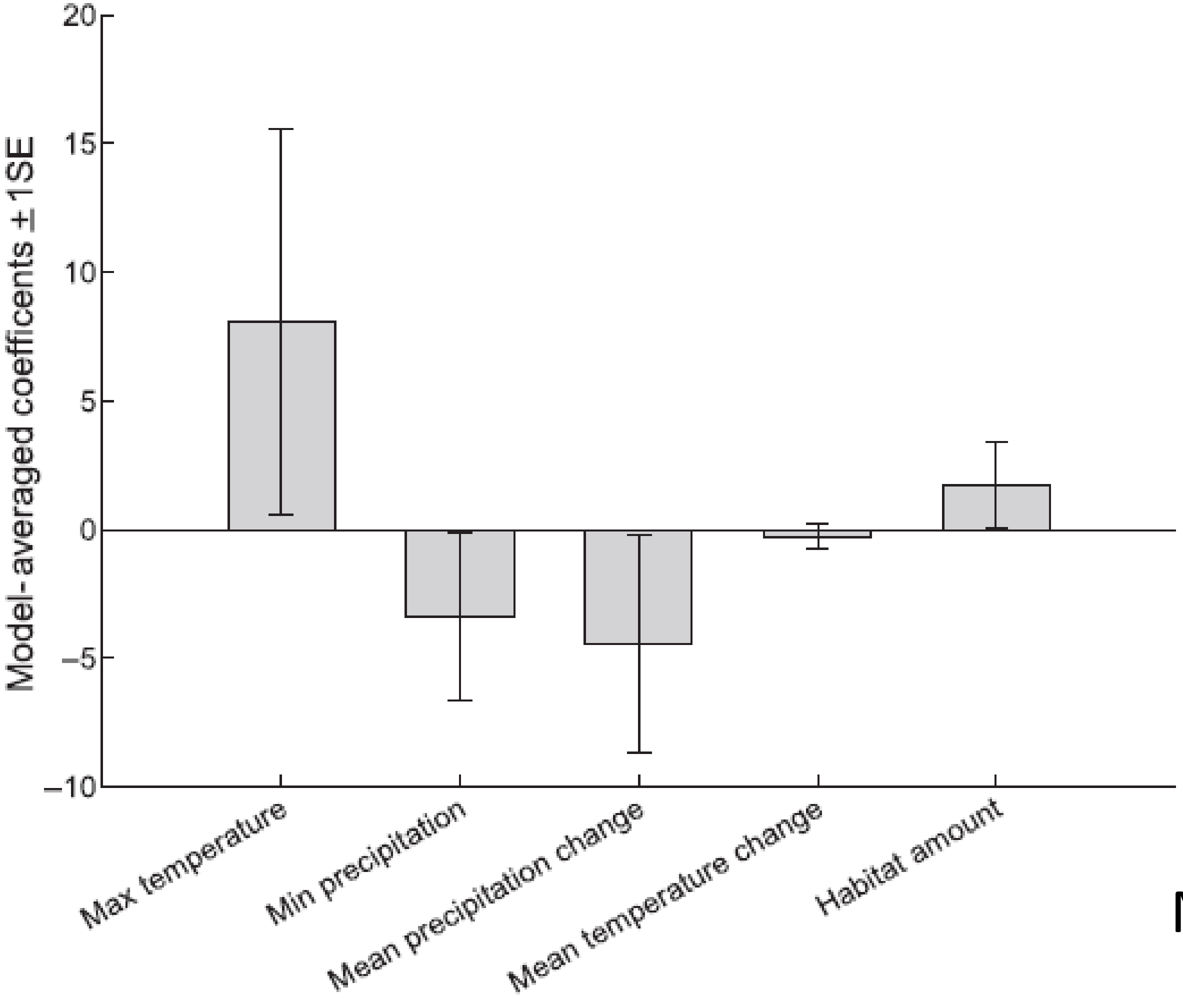
Response variable: negative effect of habitat loss or not

Fixed effects: climate, climate change, habitat amount

Random effects: study identity, taxonomic group, habitat and land use

Mantyka-Pringle et al. (2012). *Global Change Biology* **18**: 1239-1252

Applications of mixed-effects models: effects of climate on responses of biodiversity to land use



Negative effects of habitat loss more likely in places that are hot, dry and that have got drier recently

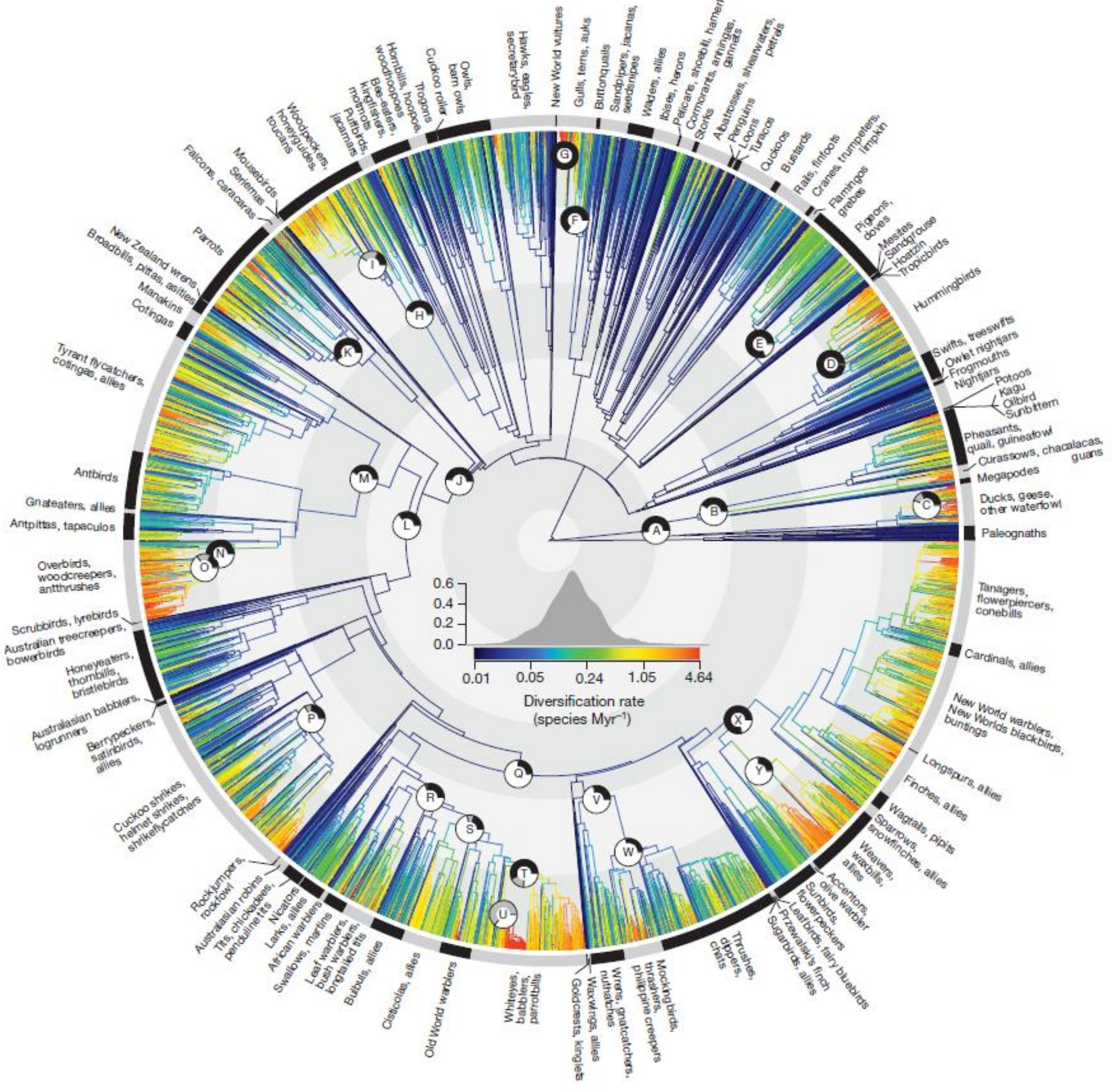
Mantyka-Pringle et al. (2012). *Global Change Biology* **18**: 1239-1252

Applications of mixed-effects models: accounting for phylogenetic non-randomness

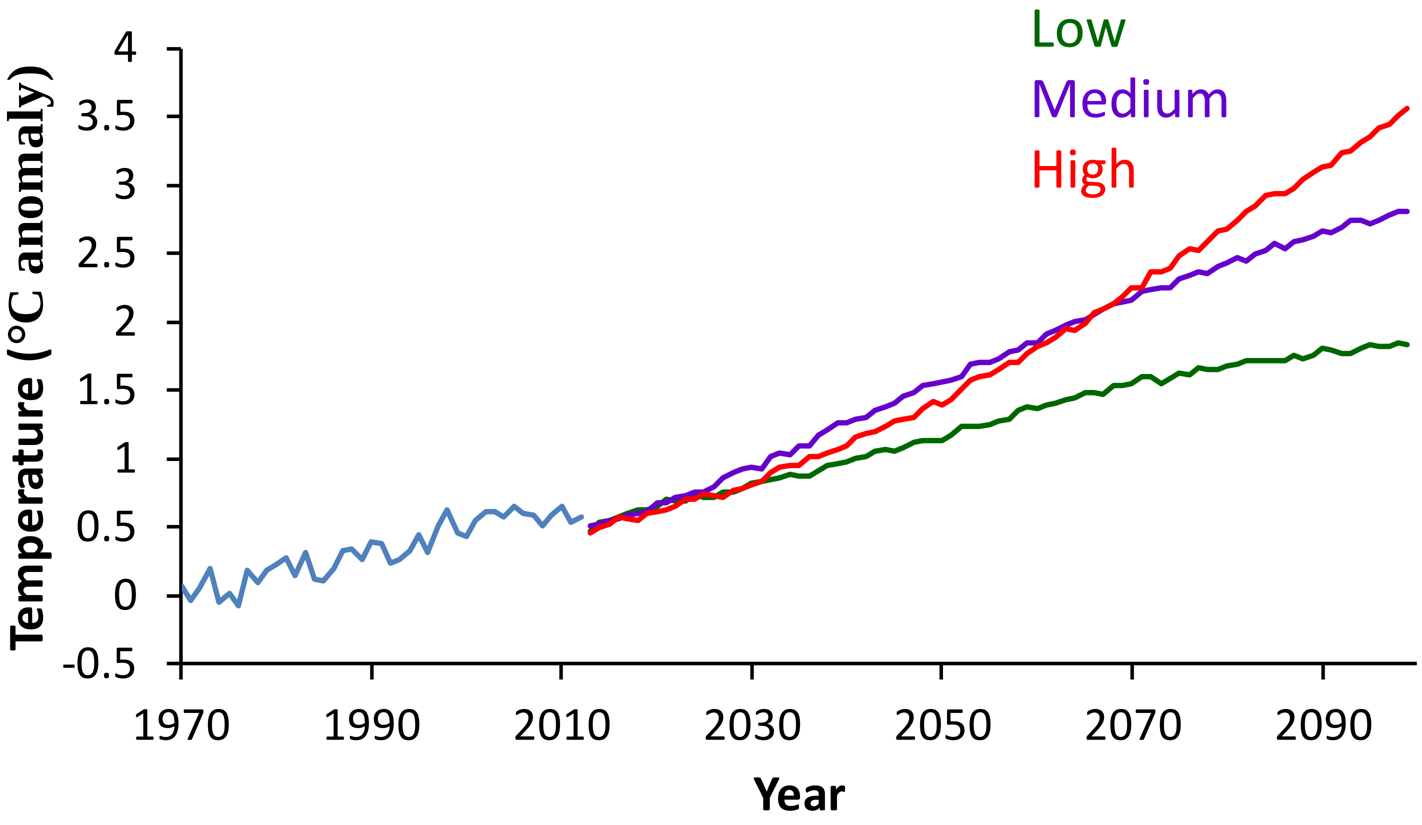
Species-level measures are often phylogenetically non-random

If ignored, can bias results of statistical models

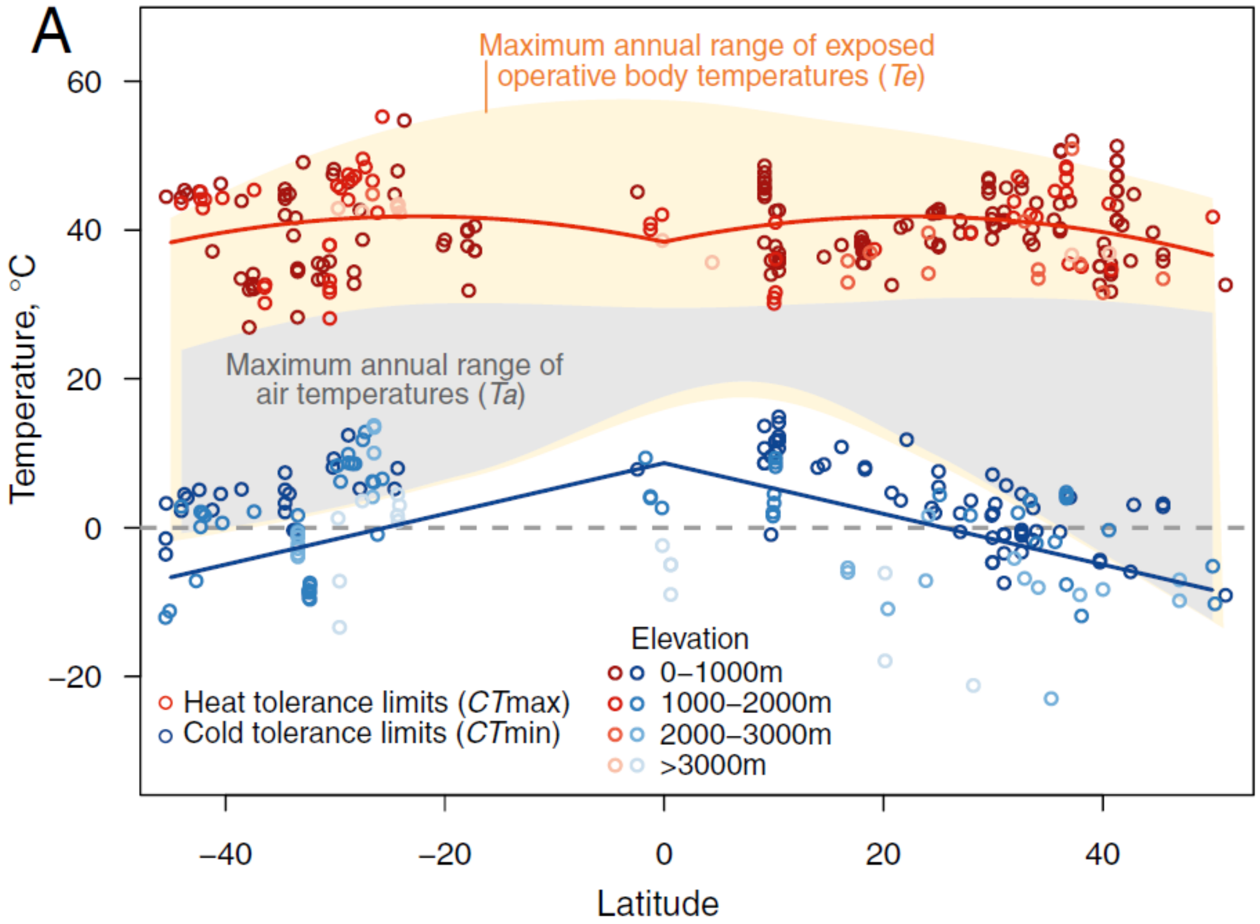
There are several statistical approaches to account for phylogeny, but one is nested random effects in mixed-effects models



Climate change is a rapidly growing threat to biodiversity



Applications of mixed-effects models: climate change and thermal safety margins



Data on the thermal tolerance limits of amphibians, reptiles and insects

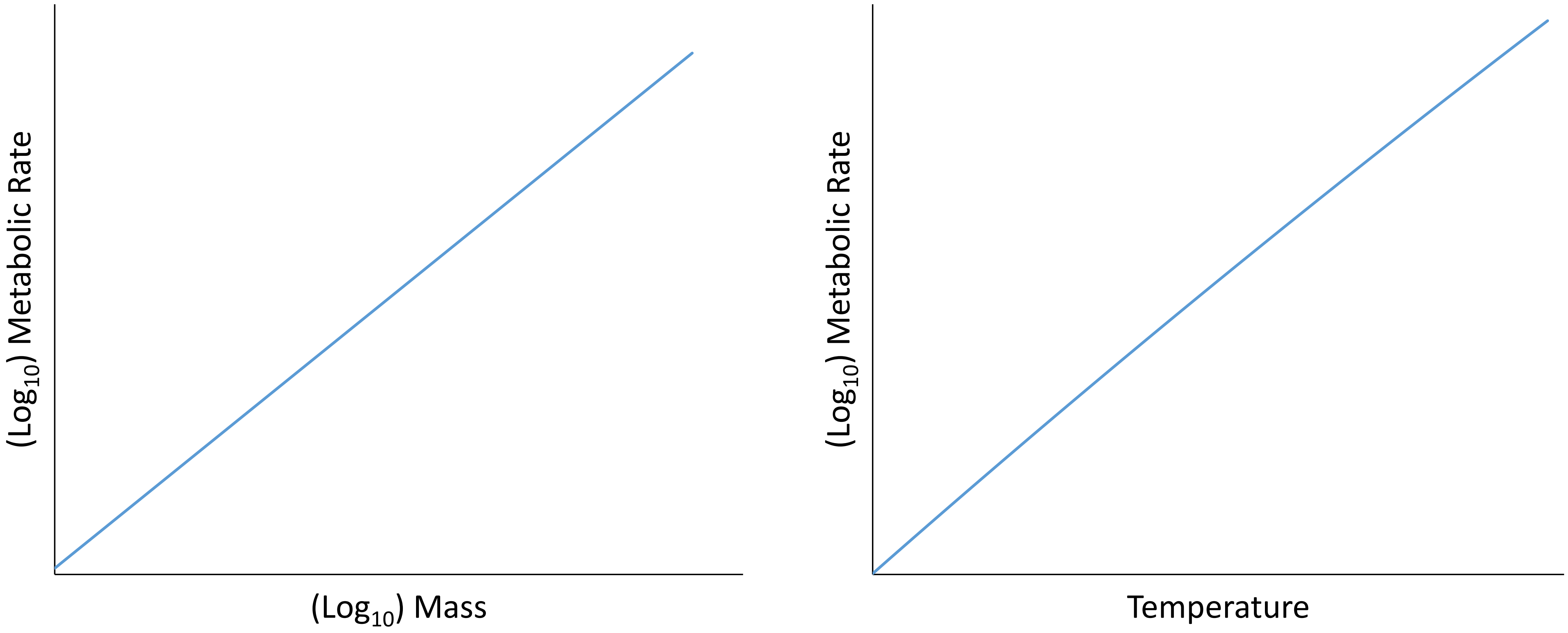
Response variable: thermal cold and heat tolerance

Fixed effects: latitude and elevation

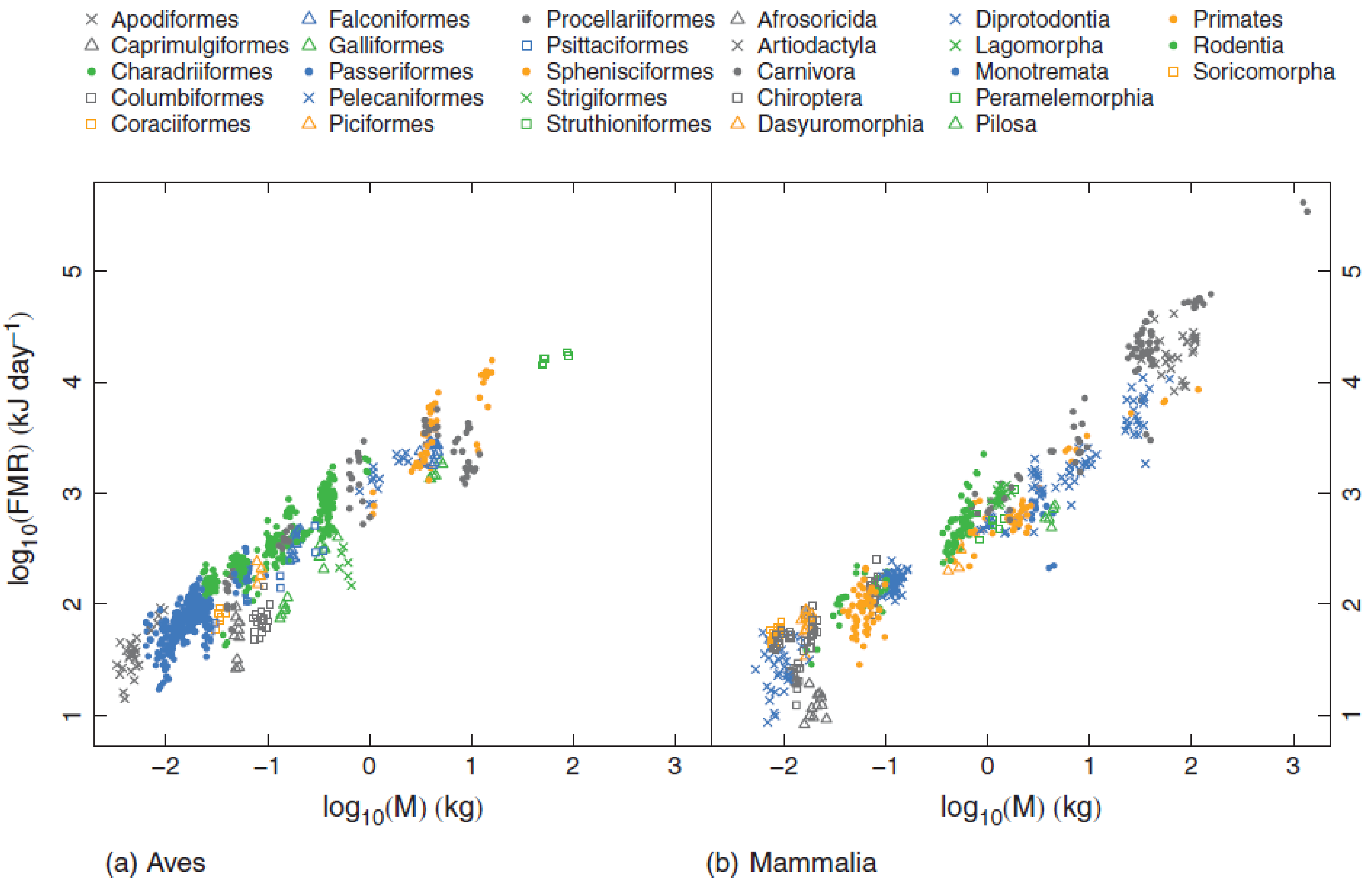
Random effects: hierarchical taxonomic terms, cold-tolerance metric

Sunday et al. (2014). *PNAS* **111**: 5610-5615

Metabolic rates: a key process underlying organismal processes and ecology



Applications of mixed-effects models: scaling of metabolic rate with body mass



Data on field metabolic rates of birds and mammals

Response variable: field metabolic rates

Fixed effects: Body mass, mammals vs. birds

Random effects: hierarchical taxonomic terms (random intercepts and slopes)

Hudson et al. (2013). *Journal of Animal Ecology* **82**: 1009-1020

Summary: Statistical models

Ecological data often have a non-straightforward structure

There is a trend toward more synthetic analyses, using data from multiple studies, to generalize patterns more broadly

This trend exacerbates the difficulties around data structure

But there are statistical approaches that can deal with these complications

Reading list

Baldock et al. (2015). Where is the UK's pollinator biodiversity? The importance of urban areas for flower-visiting insects. *Proceedings of the Royal Society of London, Series B* **282**: 20142849.

Ball et al. (2015). Body size determines functional responses of ground beetle interactions. *Basic & Applied Ecology* **16**: 621-628.

Barnes et al. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nature Communications* **5**: 5351.

Beale et al. (2013). Protected area networks and savannah bird biodiversity in the face of climate change and land degradation. *Ecology Letters* **16**: 1061-1068.

Bolker (2008). *Ecological Models and Data in R*. Princeton University Press.

Ellison (2004). Bayesian inference in ecology. *Ecology Letters* **7**: 509-520.

Evans (2012). Modelling ecological systems in a changing world. *Philosophical Transactions of the Royal Society, Series B* **367**: 181-190.

Evans et al. (2013). Do simple models lead to generality in ecology? *Trends in Ecology & Evolution* **28**: 578-583.

Hudson et al. (2013). The relationship between body mass and field metabolic rate among individual birds and mammals. *Journal of Animal Ecology* **82**: 1009-1020.

Reading list

Johnson & Omland (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19**: 101-108.

MacKenzie et al. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**: 2248-2255.

Mantyka-Pringle et al. (2012). Interactions between climate and habitat loss effects on biodiversity: a systematic review and meta-analysis. *Global Change Biology* **18**: 1239-1252.

Newbold et al. (2010). Testing the accuracy of species distribution models using species records from a new survey. *Oikos* **119**: 1326-1334.

Newbold et al. (2013). Ecological traits affect the response of tropical forest bird species to land-use intensity. *Proceedings of the Royal Society, Series B* **280**: 20122131.

Newbold et al. (2014). Functional traits, land-use change and the structure of present and future bird communities in tropical forests. *Global Ecology & Biogeography* **23**: 1073-1084.

Newbold et al. (2015). Global effects of land use on local terrestrial biodiversity. *Nature* **520**: 45-50.

Reading list

Sunday et al. (2014). Thermal-safety margins and the necessity of thermoregulatory behavior across latitude and elevation. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 5610-5615.

Woodcock et al. (2016). Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nature Communications* **7**: 12459.

Zuur et al. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer.