Rewinding a bit…… To yesterday

We have detection histories – now what?

Pr(eh = 0 0 1 0 1 1 0 0) =

(1-ε3)(1-p4) (1-ε4)p5 (1-ε5)p6 [ε6 + (1-ε6)(1-p7){ε7+(1- ε7)(1- p8)}]= H1

Likelihood (parameters|data) = $H_1^{\text{No. cases}}\ H_2^{\text{No. cases}}\ H_x^{\text{No. cases}}$

0001010010	10
0010001101	3
0101010100	4
1001000000	18
0001010111	2
0011101000	1
	3
	No. cases

Maximize likelihood function so that, under the assumed statistical model, the observed data is most probable.

https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

Likelihood (parameters|data) = $H_1^{\text{No. cases}}\ H_2^{\text{No. cases}}\ H_x^{\text{No. cases}}$

Maximize likelihood function so that, under the assumed statistical model, the observed data is most probable.

https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

Likelihood (parameters|data) = $H_1^{\text{No. cases}}\ H_2^{\text{No. cases}}\ H_x^{\text{No. cases}}$

Maximize likelihood function so that, under the assumed statistical model, the observed data is most probable.

https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

Likelihood (parameters|data) = $H_1^{\text{No. cases}}\ H_2^{\text{No. cases}}\ H_x^{\text{No. cases}}$

- **ε(constant)p(time-varying)**
- **ε(time-varying)p(sea-levels)**

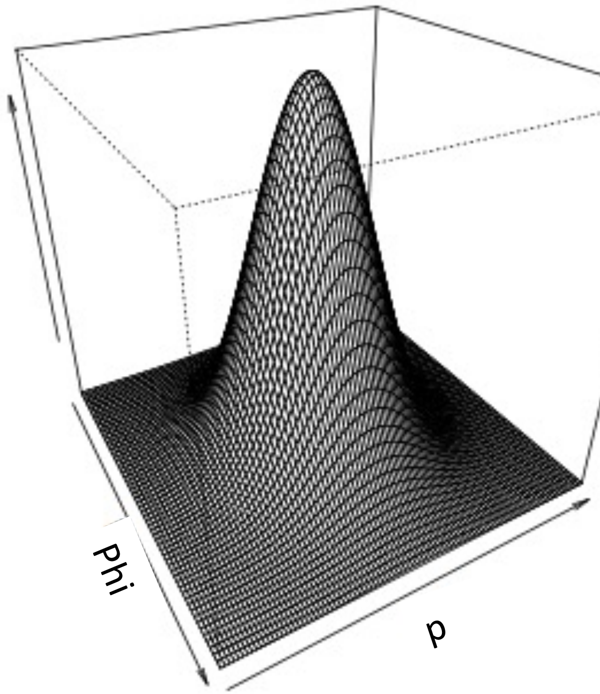
Maximize likelihood function so that, under the assumed statistical model, the observed data is most probable.

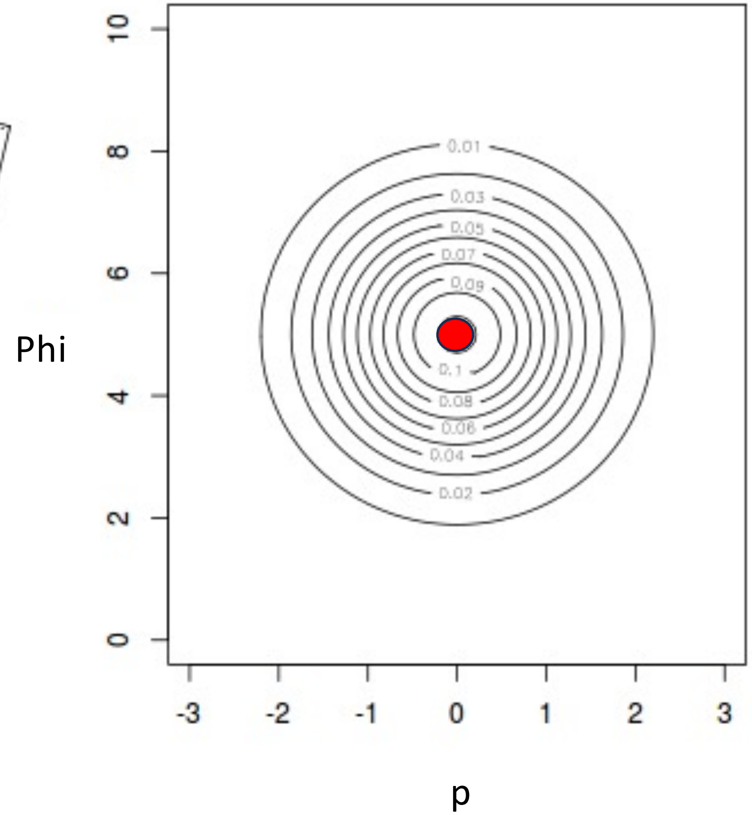
https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

Likelihood (parameters|data) = $H_1^{\text{No. cases}}\ H_2^{\text{No. cases}}\ H_x^{\text{No. cases}}$

- **ε(constant)p(time-varying)**
- **ε(time-varying)p(sea-levels)**
- **phi(constnat))p(sconstant)**

0001010010
0010001101
0101010100
1001000000
0001010111
0011101000





https://www.flutterbys.com.au/stats/tut/tut4.3.html

```
> AIC(dipper.const,dipper.phi.t, dipper.p.t, dipper.phi.t.p.t)
                           model npar rank   logLik     AIC    AICc   dAIC  AICwt
dipper.const      p~1 phi~1    2    2 -333.4188 670.838 671.068  0.000 0.7953
dipper.phi.t      p~1 phi~t    7    7 -329.8650 673.730 676.113  2.892 0.1873
dipper.p.t        p~t phi~1    7    7 -332.2401 678.480 680.863  7.642 0.0174
dipper.phi.t.p.t  p~t phi~t   12   12 -328.4751 680.950 688.379 10.112 0.0000
>
```

Higher log-likelihood value better fit to the data
Least negative (highest) is dipper.phi.t.p.t
Why is dipper.constant on “top” (the code rearranges the models with the best as the first row)

$$\mathrm{AIC} = 2k - 2\ln \widehat{L}$$

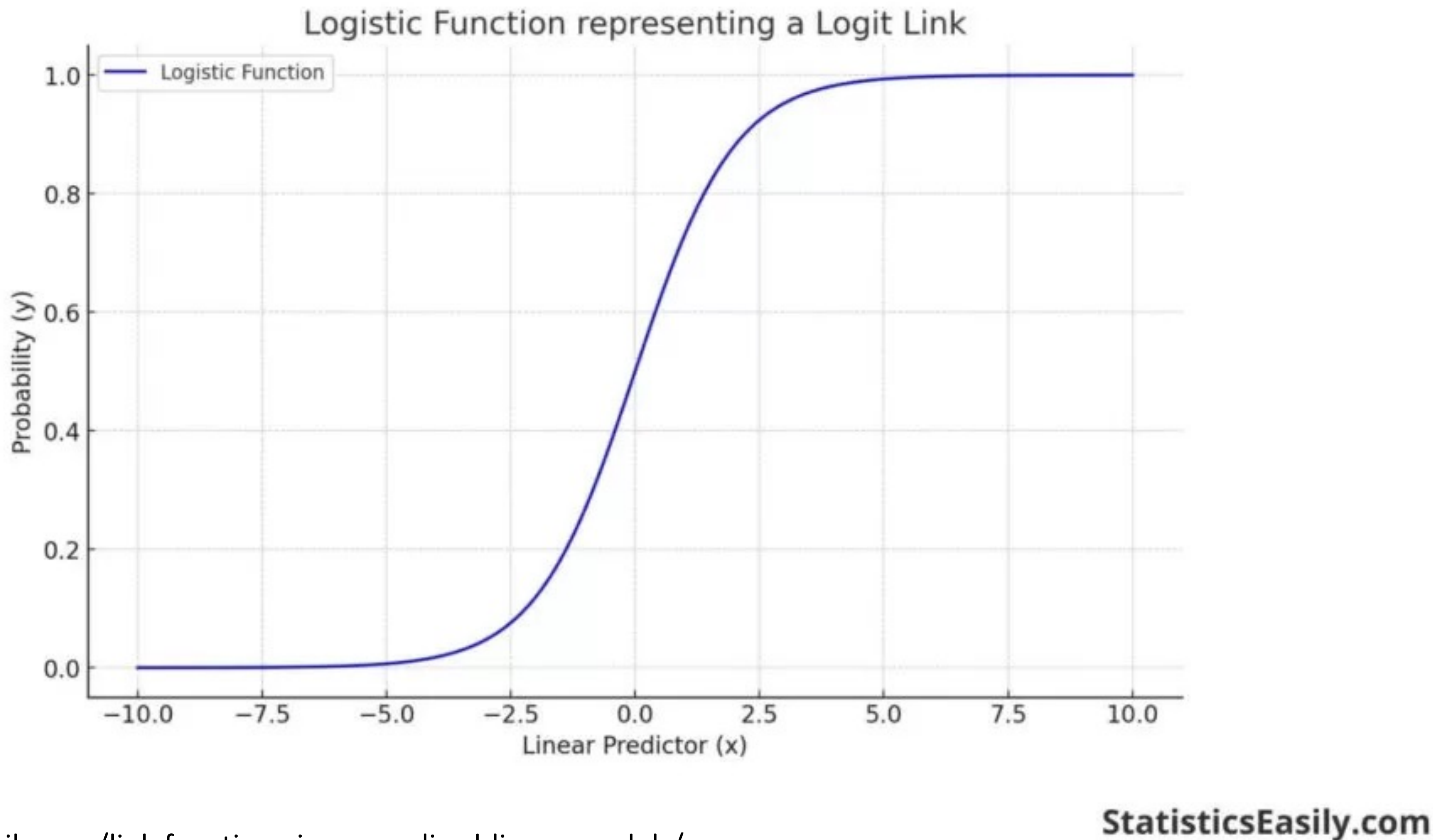
$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

Covariate modeling via link functions

- Link functions allows model to accommodate response variables that do not naturally fit a linear scale, such as probabilities between 0 and 1
- In logistic regression model (one example of a GLM Generalized Linear Model), the logit link function transforms the probability scale to an unbounded scale, where linear regression can be applied

https://statisticseasily.com/link-functions-in-generalized-linear-models/

- **logit link function is ideal for binary outcome modelling (survive or not survive, preserved or not preserved)**
- **correct link function improves model fit and accuracy**
- Identity, Probit, Log , Inverse, Sine are other link functions you might encounter.



https://statisticseasily.com/link-functions-in-generalized-linear-models/

Covariate modeling via link functions

$$\text{logit}(\varepsilon_{i,t}) = \log\left(\frac{\varepsilon_{i,t}}{1-\varepsilon_{i,t}}\right) = \beta_0 + \beta_1 x_i + \beta_2 y_t$$

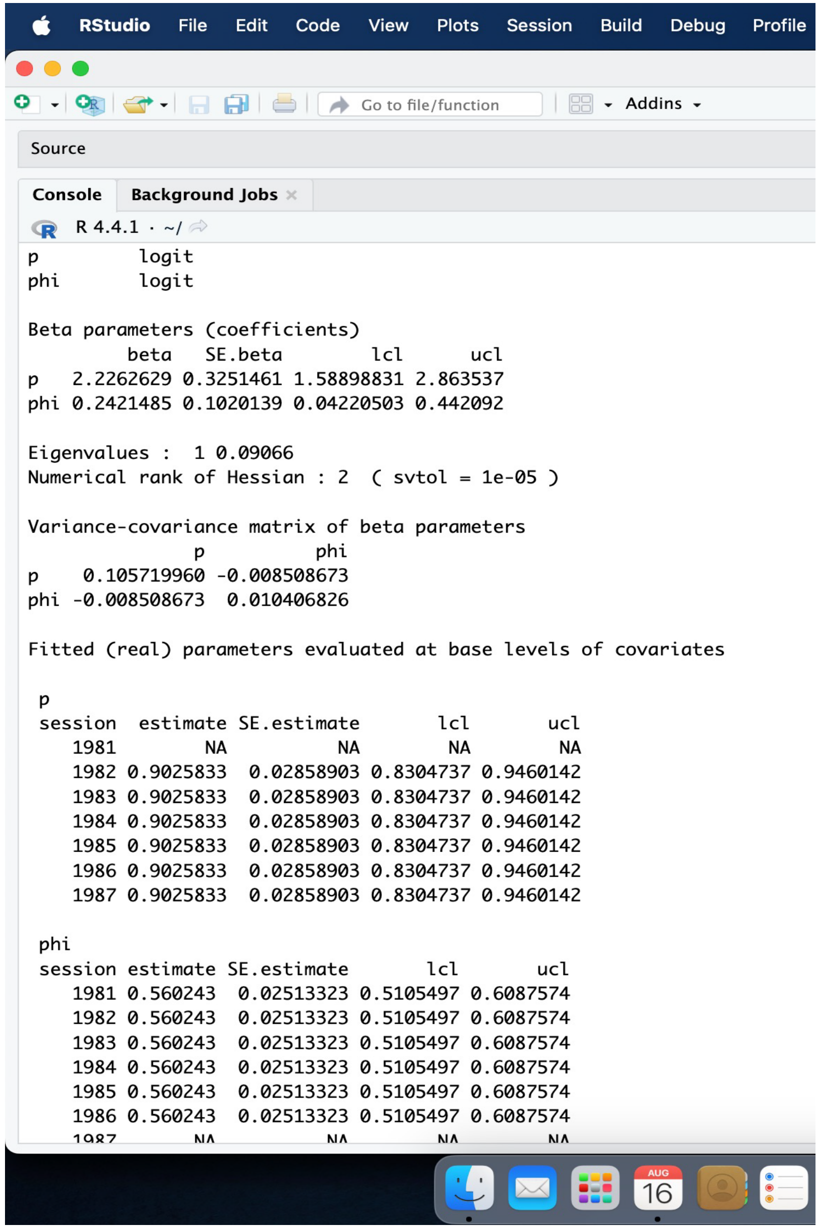
Time specific covariates

- Duration of bin
- Sea-level
- Temperature

$$\text{logit}(\varepsilon_{i,t}) = \log\left(\frac{\varepsilon_{i,t}}{1-\varepsilon_{i,t}}\right) = \beta_0$$

invlogit(2.2262629)
[1] 0.9025833

invlogit(0.2421485)
[1] 0.560243



Covariate modeling via link functions

$$\mathrm{logit}(\varepsilon_{i,t}) = \log\left(\frac{\varepsilon_{i,t}}{1-\varepsilon_{i,t}}\right) = \boxed{\beta_0} + \boxed{\beta_1} x_i + \boxed{\beta_2} y_t$$

$$\varepsilon_{i,t} = \frac{e^{\boxed{\beta_0} + \boxed{\beta_1} x_i + \boxed{\beta_2} y_t}}{1 + e^{\boxed{\beta_0} + \boxed{\beta_1} x_i + \boxed{\beta_2} y_t}}$$

A bit of everything related to CMR in paleo

Lee Hsiang Liow

Natural History Museum and Centre for Planetary Habitability, University of Oslo, Norway

16.8.2024 (Friday)

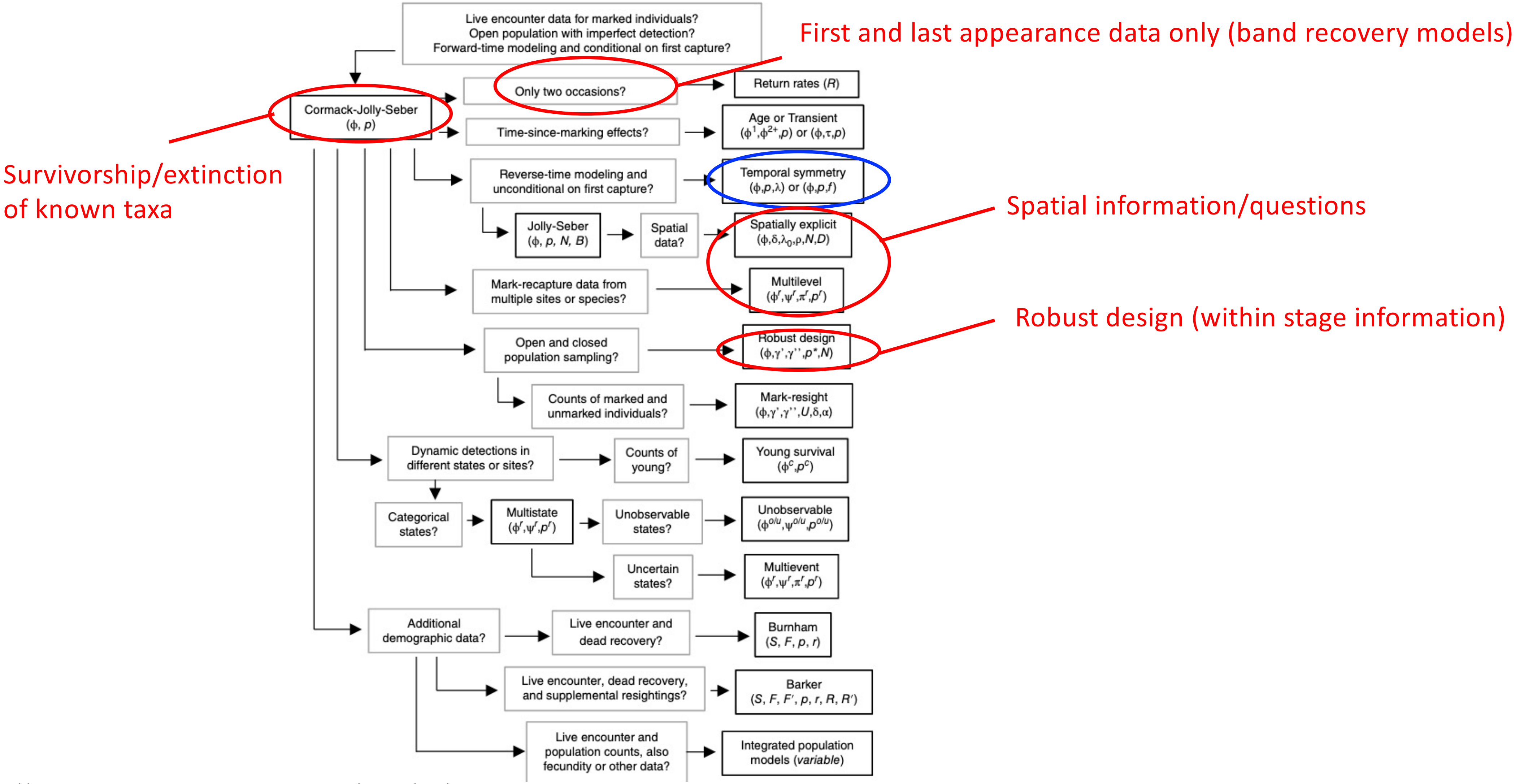
Summary

$p \neq 1$ Two types of zeros

Common tutorial questions

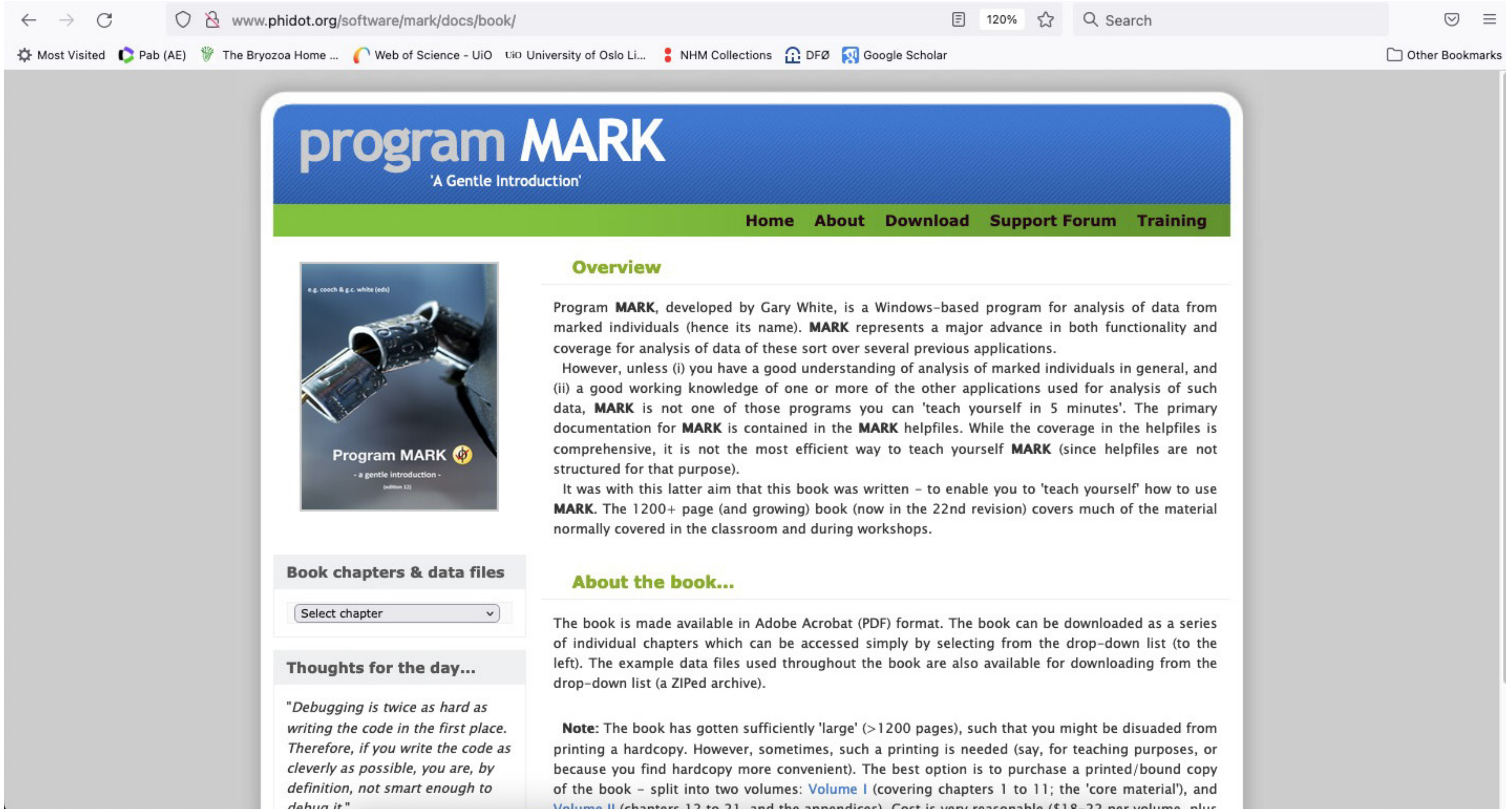
- Additive versus multiplictive models
- Model-averaging

Many models: few explored in paleo



https://bksandercock.files.wordpress.com/2020/11/sandercock2020popecolpractice.pdf

MARK is the “main” software and easy to read reference (but you must “translate”)



http://www.phidot.org/software/mark/docs/book/

Gary White

Self learning material and software

Gary White MARK

http://www.phidot.org/software/mark/

(great if you are windows user, a bit more involved if you us Mac) great to pair with Rmark (Jeff Laake)

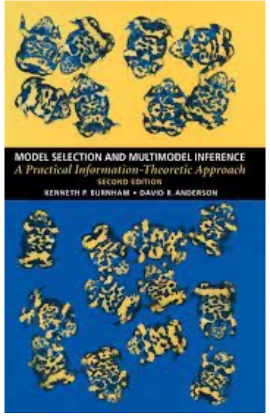
Gary White MARK book

http://www.phidot.org/software/mark/docs/book/

Michael Conroy lecture notes (bare bones description of models; useful for Mark/Rmark code and data) https://sites.google.com/site/cmrsoftware/

Burnham and Anderson 2022:

Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach



R packages for CMR

- RMark (runs MARK from R, works best with Windows)
- serc and openCR
- Marked (some overlap with openCR and MARK but this doesn’t have Pradel models)
- R2ucare (for goodness of fit tests and simulations for CMR)
- mra, Rcapture, BaSTA

Other langauges for capture recapture

Python https://www.python.org/

- https://austinrochford.com/posts/2018-01-31-capture-recapture.html
- https://pyro.ai/examples/capture_recapture.html

Stan https://mc-stan.org/

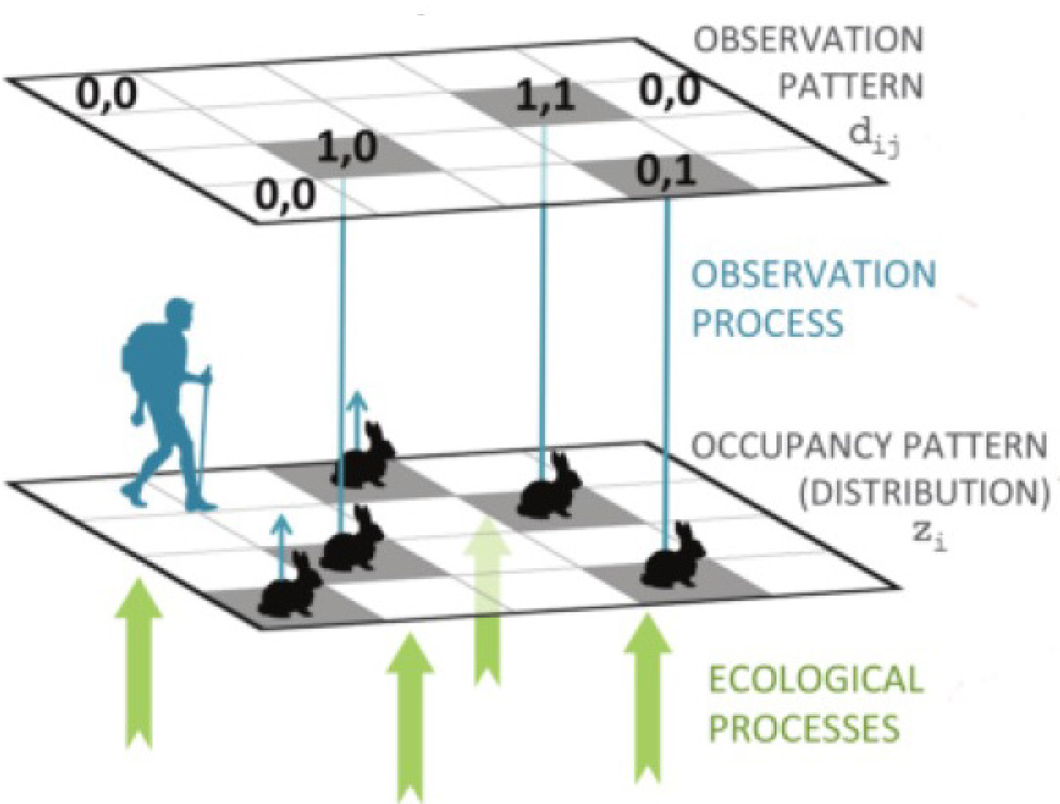
- https://mc-stan.org/docs/stan-users-guide/latent-discrete.html#mark-recapture-models

BUGS, JAGS

https://bcss.org.my/tut/bayes-with-jags-a-tutorial-for-wildlife-researchers/abundance-from-capture-recapture-data/basic-spatial-capture-recapture-models/

https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecs2.3810

Site-occupancy modeling in statistical ecology

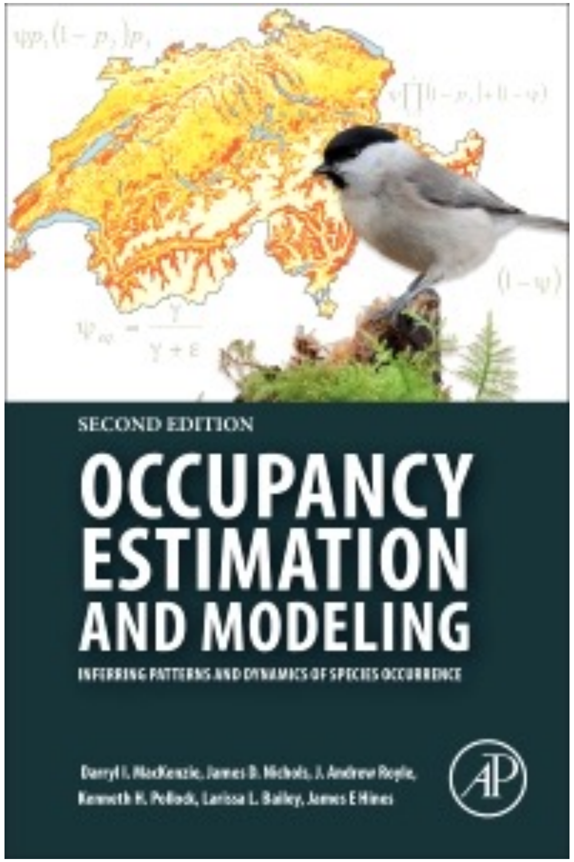


Not to be confused with
Foote, M. *et al.* (2007) Rise and fall of species occupancy in Cenozoic fossil molluscs. *Science (80-.).* 318, 1131–11341

Foote, M. (2016) On the measurement of occupancy in ecology and paleontology. *Am. Nat.* 42, 707–729

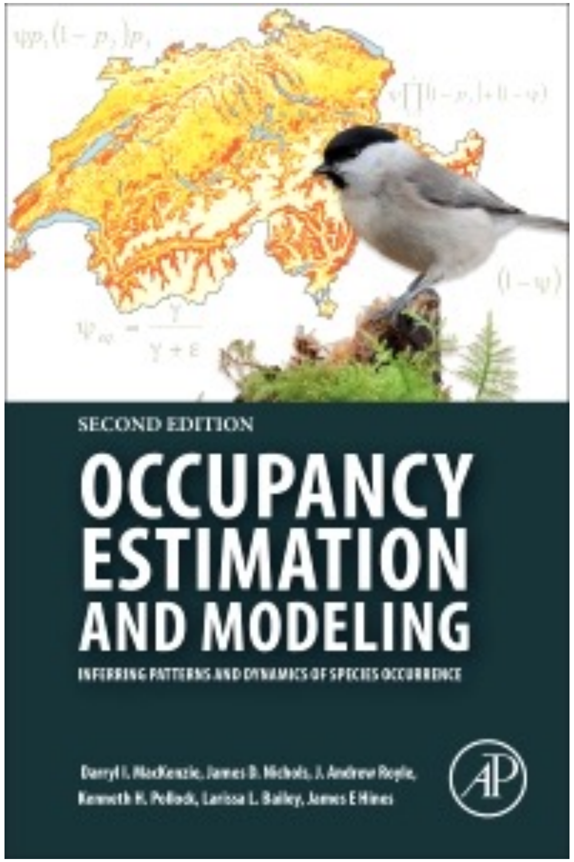
http://www.seec.uct.ac.za/single-season-occupancy-models-using-bayesian-approach

(Site)-Occupancy modeling

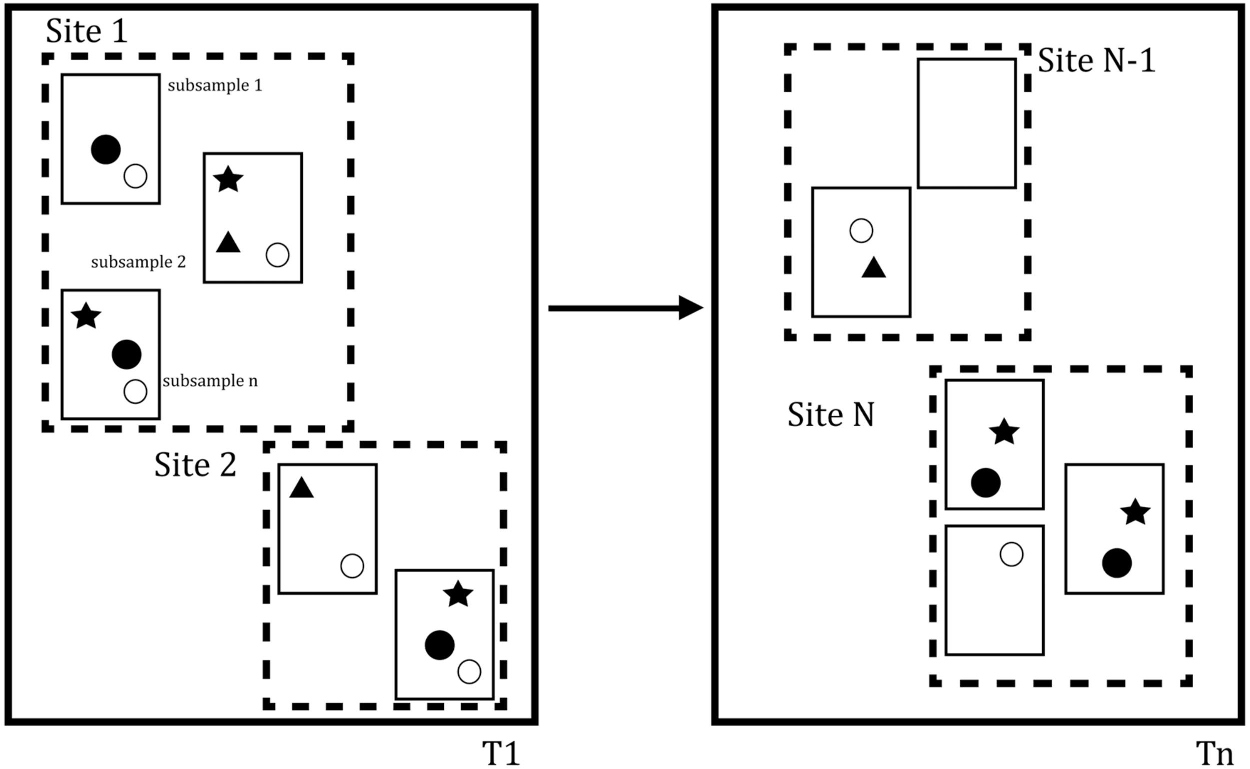


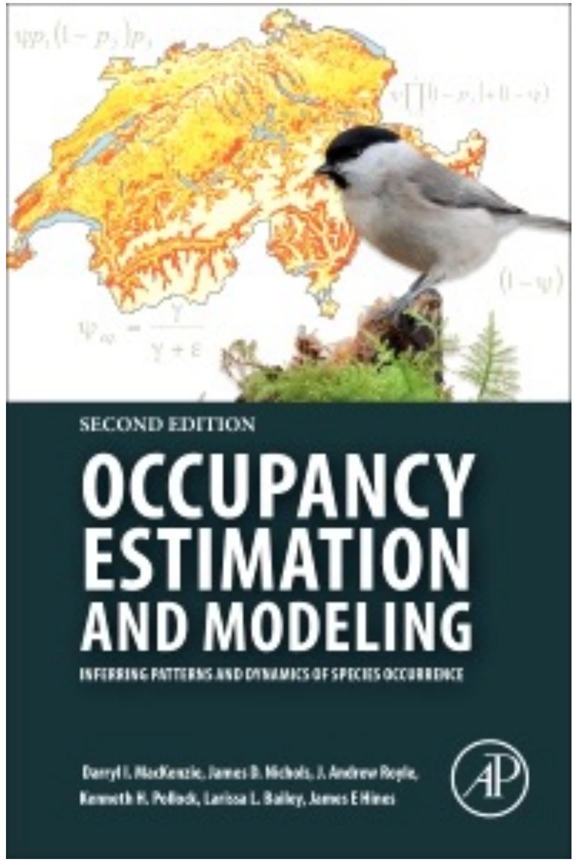
Occupancy modeling in paleo

- Liow, L.H. (2013) Simultaneous estimation of occupancy and detection probabilities: an illustration using Cincinnatian brachiopods. *Paleobiology* 39, 193–213
- Lawing, A. Michelle, et al. (2021). Occupancy models reveal regional differences in detectability and improve relative abundance estimations in fossil pollen assemblages. Quaternary Science Reviews 253: 106747. https://doi.org/10.1016/j.quascirev.2020.106747
- Reitan, T., Ergon, T., & Liow, L. H. (2022). Relative species abundance and population densities of the past: Developing multispecies occupancy models for fossil data. *Paleobiology,* 1-16. doi:10.1017/pab.2022.17
- Reitan, T., E. D. Martino, and L. H. Liow. (2024). Estimating relative species abundance using fossil data identified to different taxonomic levels. Ecography 2024:e06866.



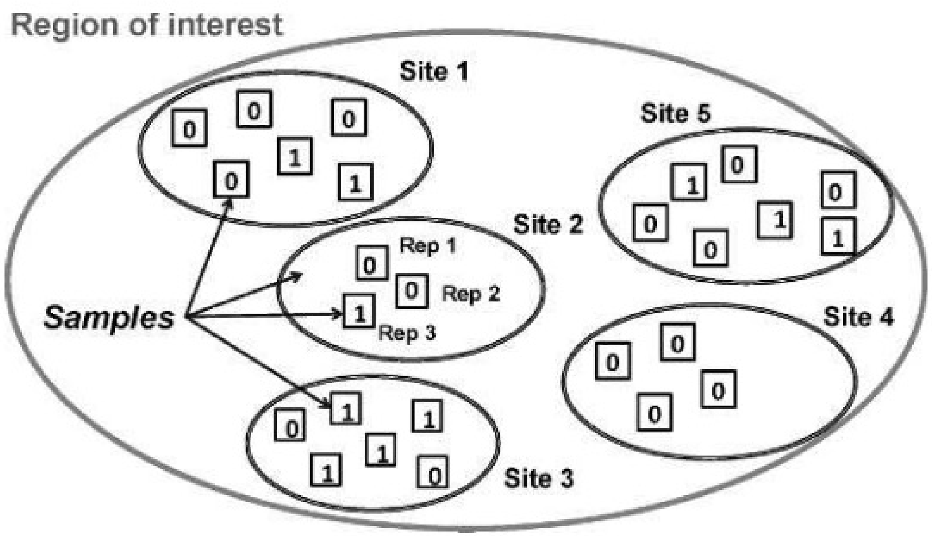
Occupancy modeling in paleo





$$Pr(h = 1010) = \varphi(1-p)^2 p^2$$

$$Pr(h = 0000) = ?????$$



$$Pr(h = 1010) = \varphi(1 - p)^2 p^2$$

$$Pr(h = 0000) = (1 - \varphi) + \varphi(1 - p)^4$$

$$L(\varphi, p | h_{1,} h_{2,\ldots} \; h_{s,}) = \prod_{i=1}^{s} \Pr(h_i)$$

$$Pr(h = 1010) = \varphi(1 - p)^2 p^2$$

$$Pr(h = 0000) = (1 - \varphi) + \varphi(1 - p)^4$$

$$L(\varphi, p | h_{1,} h_{2,\ldots} \; h_{s,}) = \prod_{i=1}^{s} \Pr(h_i)$$

R package occupancy modeling

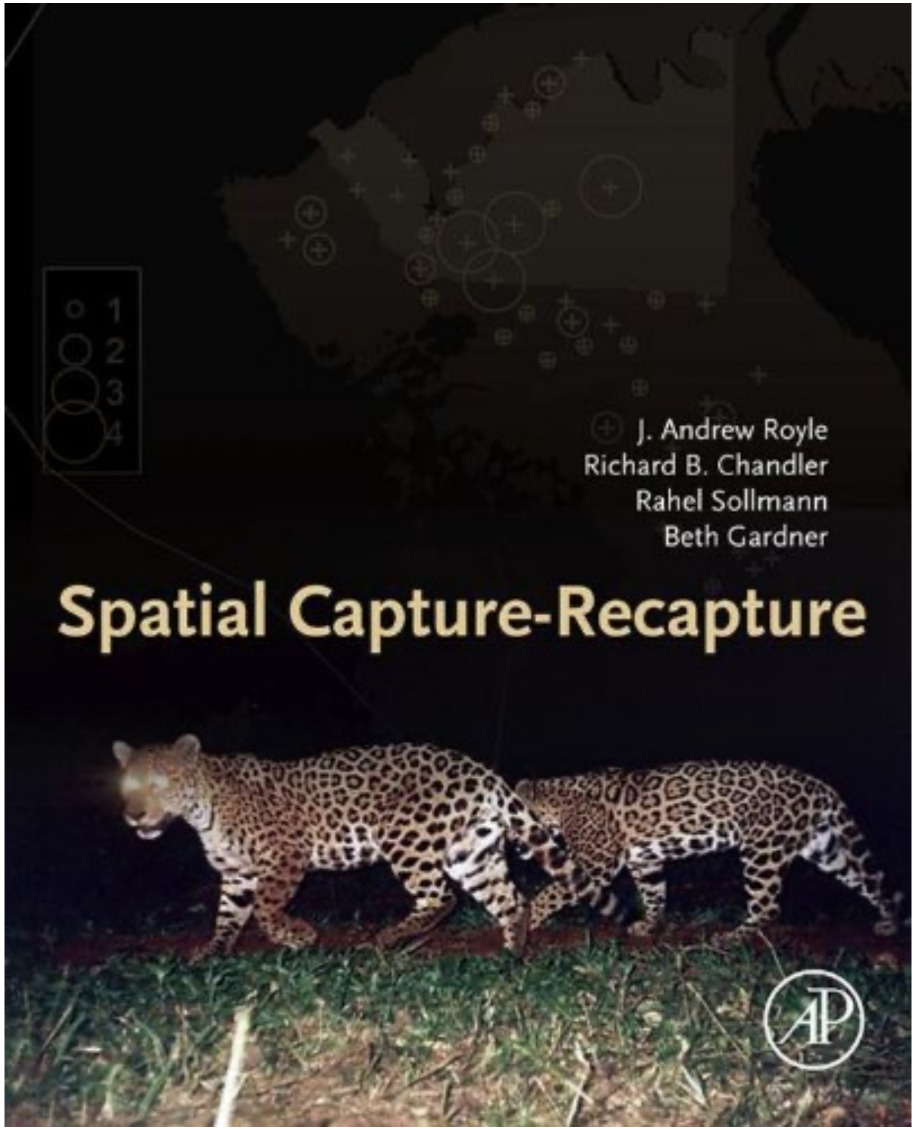
- openCR, serc
- unmarked

Some resources for occupancy modeling

- https://kevintshoemaker.github.io/NRES-746/Occupancy.html
- https://science.uct.ac.za/seec/stats-toolbox-seminars-spatial-and-species-distribution-toolboxes/single-season-occupancy-models-using-bayesian-approach

Spatial capture recapture

- Combining the best of capture recapture and occupancy
- Also implemented in serc, openCR, unmarked



Brief comparisons of CMR with other approaches (paleo context)

HISTORY OF ESTIMATION TAXONOMIC RICHNESS AND DIVERSIFICATION RATES

Nichols & Pollock 1983 Estimating taxonomic diversity, extinction rates, and speciation rates from fossil data using capture-recapture models. *Paleobiology* 9, 150–163

Foote & Raup 1996 Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*

Foote 1999/2001 (Boundary crossers method)

Alroy et al. 2001 (sampling standardization)

Connolly and Miller papers 2001-2 using CMR (Connolly is an ecologist)

Foote 2003 (few people use this) – CMR-like, but accounts for orgination and extinction within time interval (but see robust design)

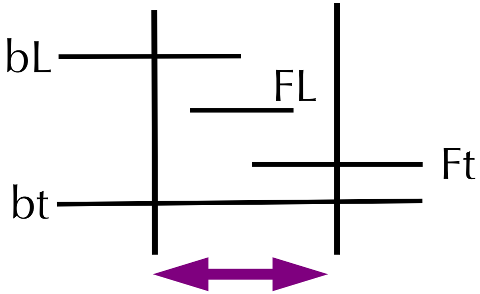
(2009) My own first capture recapture paleo-paper – I met Nichols in 2006; short course paper with Nichols

Silvestro, Schinitzler & Liow Syst bio 2014 Pyrate model paper (not the software)

Warnock et al. 2020 RevBayes (starting from birth death models but dropping the “relationships”)

Comparisons with other approaches

Foote's per capita origination and extinction rates (boundary crossers)



$$\hat{p} = \ln\left(\frac{N_t}{N_{bt}}\right) / \Delta t$$

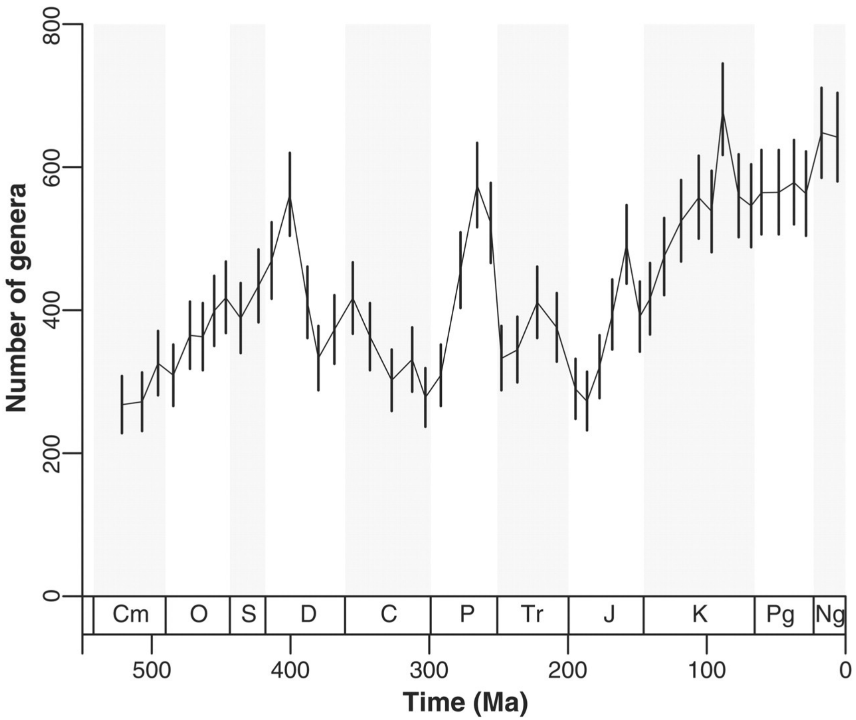
$$\hat{q} = \ln\left(\frac{N_b}{N_{bt}}\right) / \Delta t$$

- Easy to use
- Lose information of "singletons" (FL class information not used)
- Hard to compare different models

Foote, M. 2000. Origination and extinction components of taxonomic diversity: general problems. **Paleobiology 26:74-102.**

Comparisons with other approaches

Sampling standardization approaches (including SQS)



- Easy to use
- Assume that even sample or quorums will allow unbiased relative change to be estimated
- Ad hoc rather than modelling approach (preservation is not modelled although it is part of the process)
- Hard to compare different models

Comparisons with other approaches

CMR-like approaches in paleo independently developed

Foote 2003 Journal of Geology

Alroy “three-timer” and related methods

- Some hard, some easy to use, not easy to understand
- Hard to compare different models
- (very) Special cases of CMR

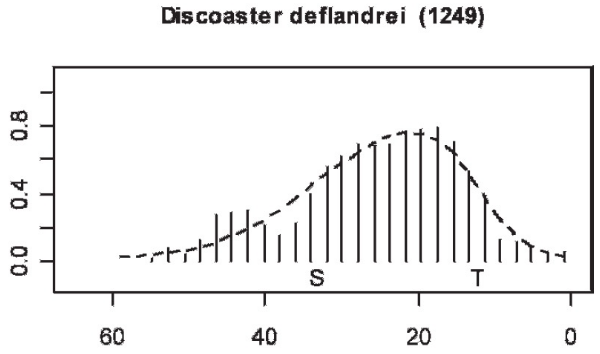
Comparisons with other approaches

PyRate (Silvestro et al. Sys Bio 2014)

Key preservation assumption (different) is the shape of species observations (beta distribution based on “hat”)

Smooths out temporal information

Conditioned on at least one observation per taxon (like the CMR models conditioned on first observation)



RevBayes

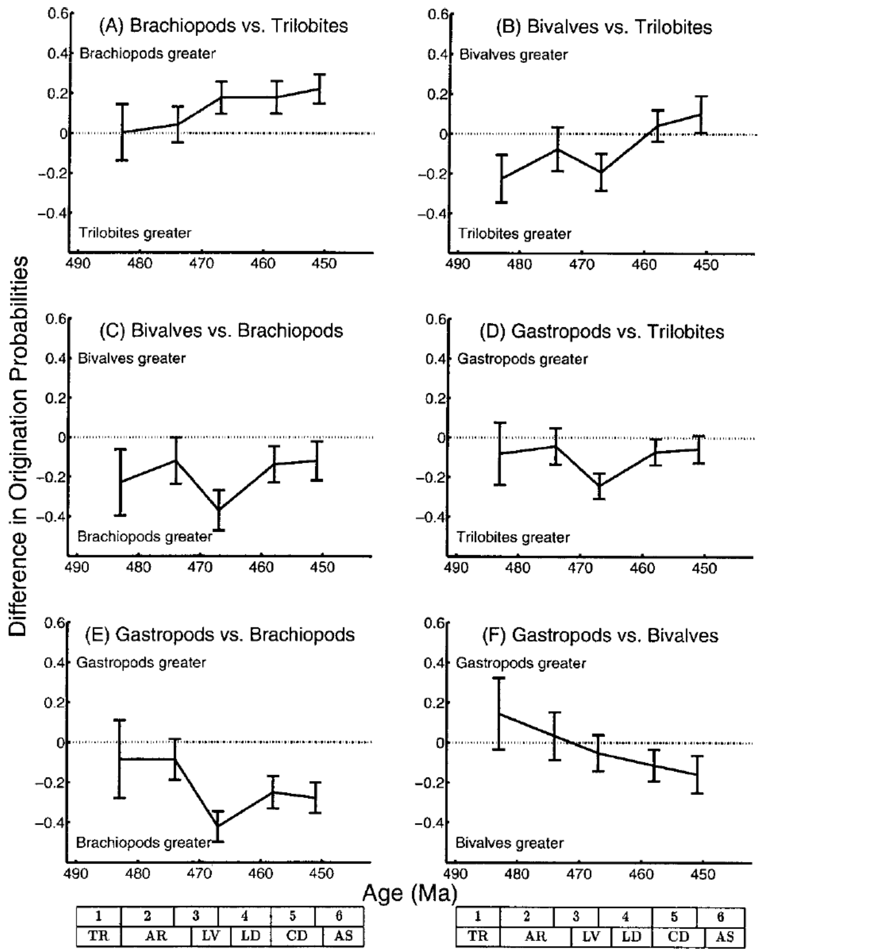
https://revbayes.github.io/tutorials/fbd_range/

Rachel’s lecture!

RevBayes

Bayesian phylogenetic inference using probabilistic graphical models and an interpreted language

Some paleo papers using CMR



E 4. Model-averaged estimates of between-class differences in genus origination probabilities over time ob-

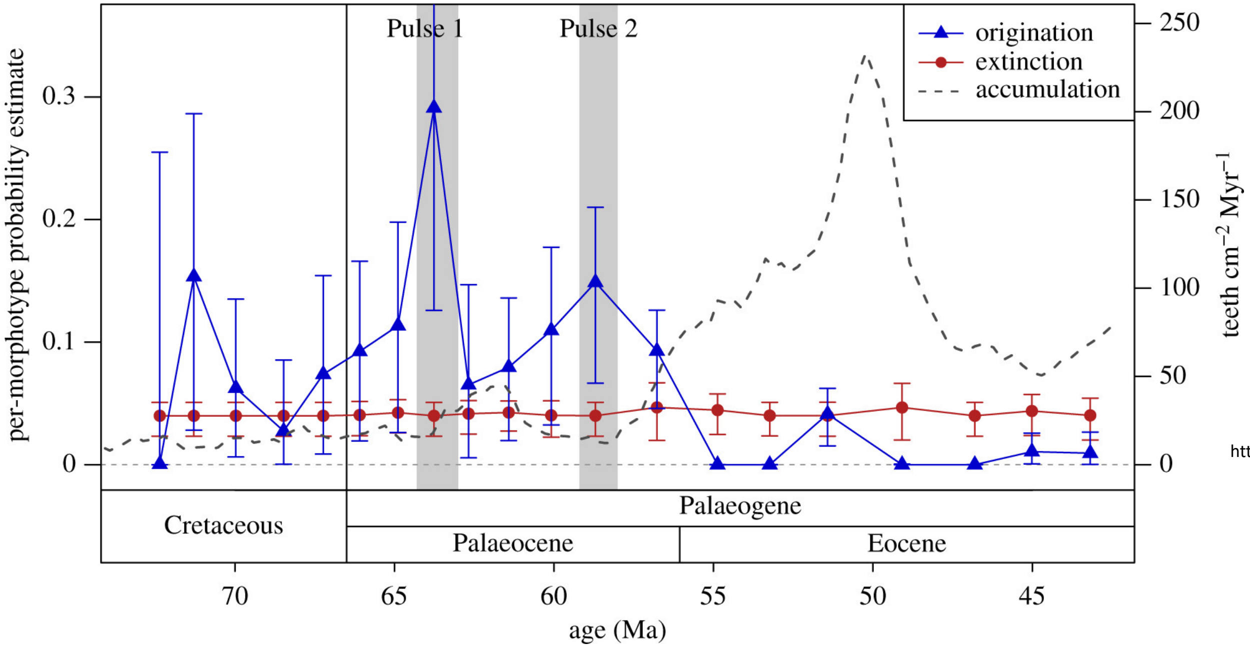
Nichols, J.D. and Pollock, K.H. (1983) Estimating taxonomic diversity, extinction rates, and speciation rates from fossil data using capture-recapture models. *Paleobiology* 9, 150–163

Connolly, S.R. and Miller, A.I. (2001) Joint estimation of sampling and turnover rates from fossil databases: Capture-Mark-Recapture methods revisited. *Paleobiology* 27, 751–767

Connolly, S.R. and Miller, A.I. (2001) Global Ordovician faunal transitions in the marine benthos: proximate causes. *Paleobiology* 27, 779–795

Connolly, S.R. and Miller, A.I. (2002) Global Ordovician faunal transitions in the marine benthos: ultimate causes. *Paleobiology* 28, 26–40

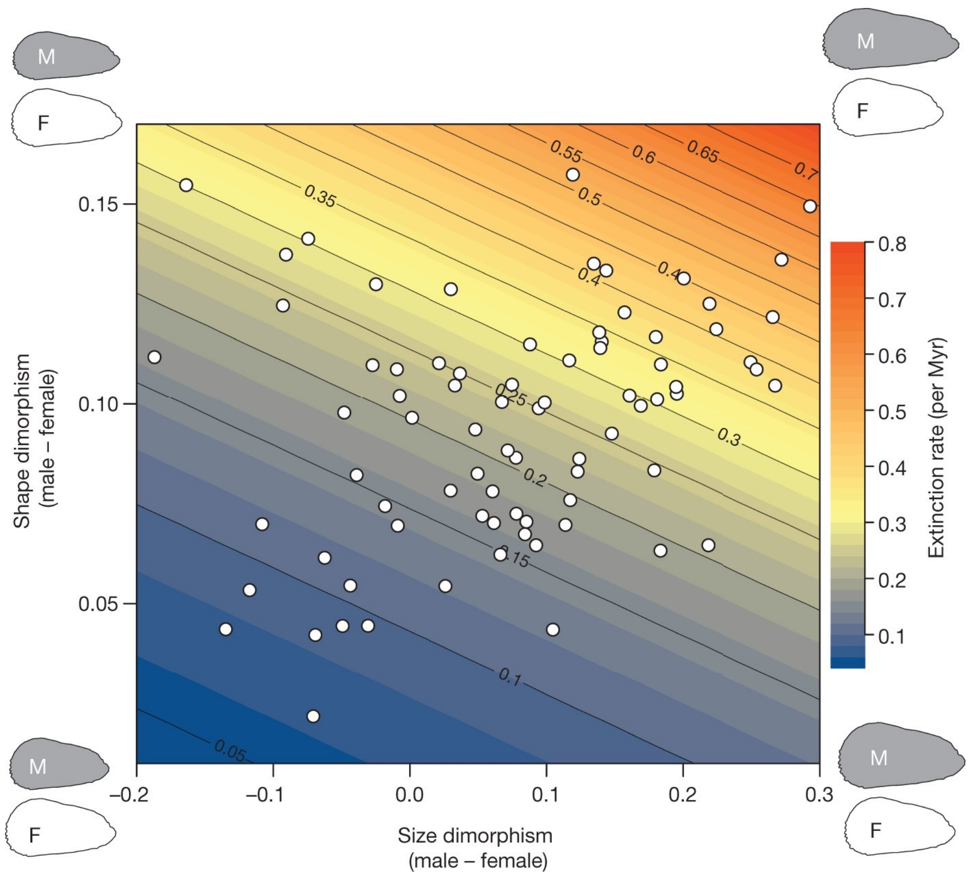
Some paleo papers using CMR

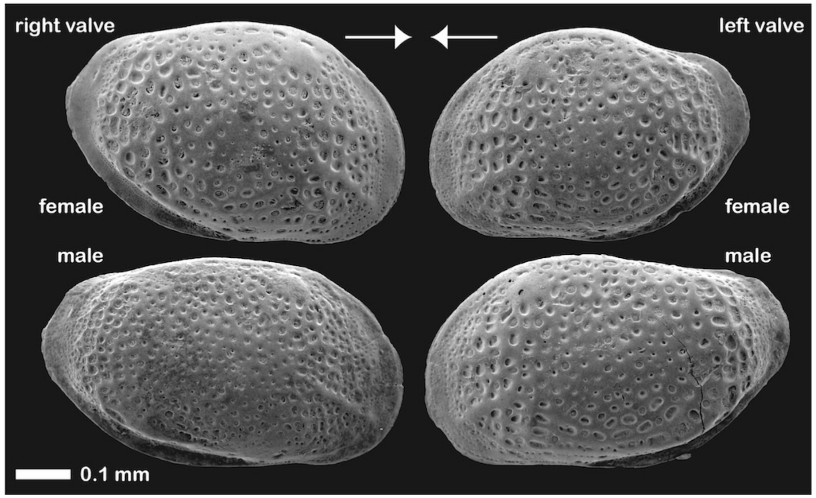


https://news.harvard.edu/gazette/story/2018/11/tiny-teeth-tell-the-story-of-two-fish-species-rapid-evolution/

Sibert, E. *et al.* (2018) Two pulses of morphological diversification in Pacific pelagic fishes following the Cretaceous - Palaeogene mass extinction. *Proc. R. Soc. B-BIOLOGICAL Sci.* 285,

Some paleo papers using CMR





Ozawa (2012)

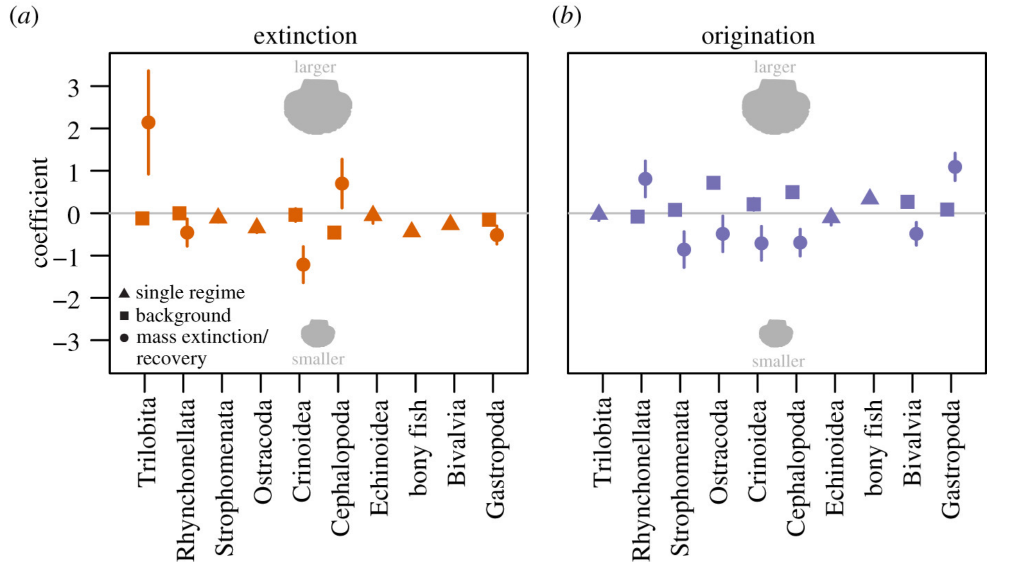
Martins, M.J.F. *et al.* (2018) High male sexual investment as a driver of extinction in fossil ostracods. *Nature*

Some paleo papers using CMR

Payne, J., & Heim, N. (2020). Body size, sampling completeness, and extinction risk in the marine fossil record. *Paleobiology, 46*(1), 23-40. doi:10.1017/pab.2019.43

Pedro M. Monarrez, Noel A. Heim and Jonathan L. Payne 2021
Mass extinctions alter extinction and origination dynamics with respect to body size
Proc B

“CMR analysis of the fossil record reveals a against the sampling of smaller-bodied genera within classes “



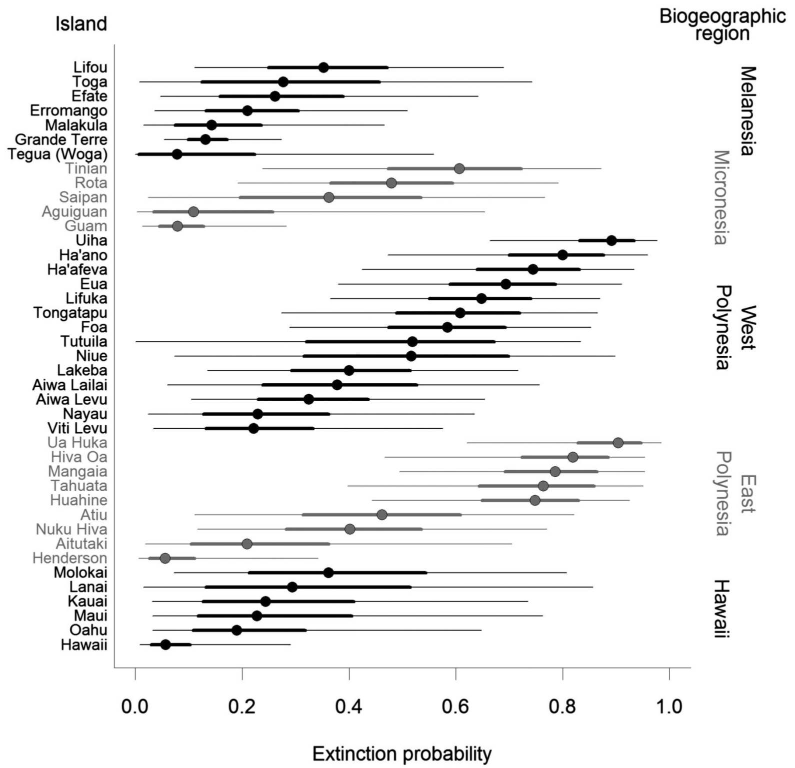
Some paleo papers using CMR

Magnitude and variation of prehistoric bird extinctions in the Pacific 2013

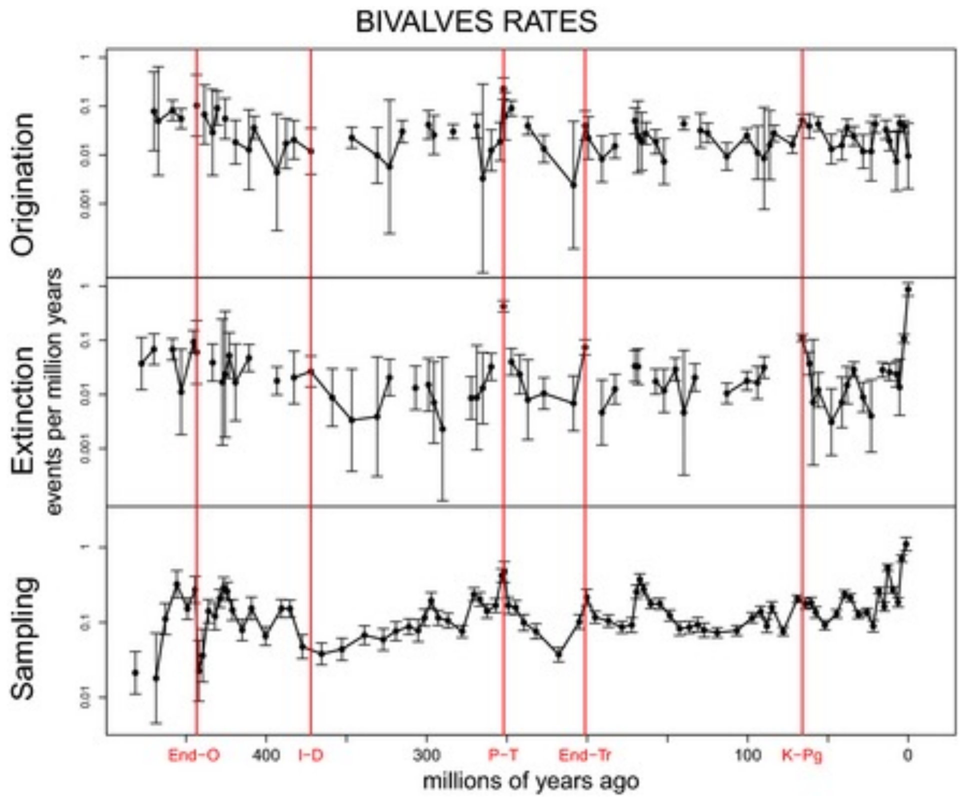
Richard P. Duncan richard.duncan@canberra.edu.au, Alison G. Boyer, and Tim M. Blackburn

PNAS

"We use a Bayesian mark-recapture approach to model gaps in the fossil record and to quantify losses of nonpasserine landbirds on 41 Pacific islands."



Some paleo papers using CMR



Liow, L.H. *et al.* (2015) Ecological interactions on macroevolutionary time scales: clams and brachiopods are more than ships that pass in the night. *Ecol. Lett.* 18, 1030–1039