How predictable is extinction?

Forecasting species survival at million-year timescales

Peter D Smits, Seth Finnegan

Department of Integrative Biology, University of California – Berkeley

Foundational assertion of conservation paleobiology

By studying the past,
we can better predict the future.

What are we predicting?

Extinction is hard to predict, but is important to conservation decisions.

Predicting extinction

- A taxon with a greater than average global geographic range is likely to survive for longer than a taxon with less than average global geographic range.

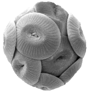
Predicting extinction

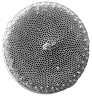
- A taxon with a greater than average global geographic range is likely to survive for longer than a taxon with less than average global geographic range.
- A taxon's global geographic range can change over time.

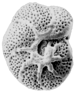
Predicting extinction

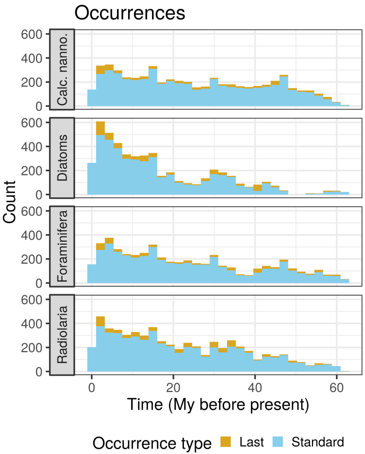
- A taxon with a greater than average global geographic range is likely to survive for longer than a taxon with less than average global geographic range.
- A taxon's global geographic range can change over time.
- What happens to extinction risk as a taxon changes geographic range? How is extinction risk impacted if that taxon's global geographic range has recently increased or decreased?

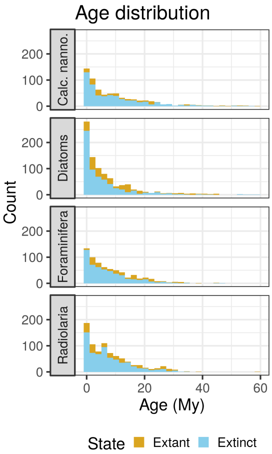
Data being analyzed: Neptune database











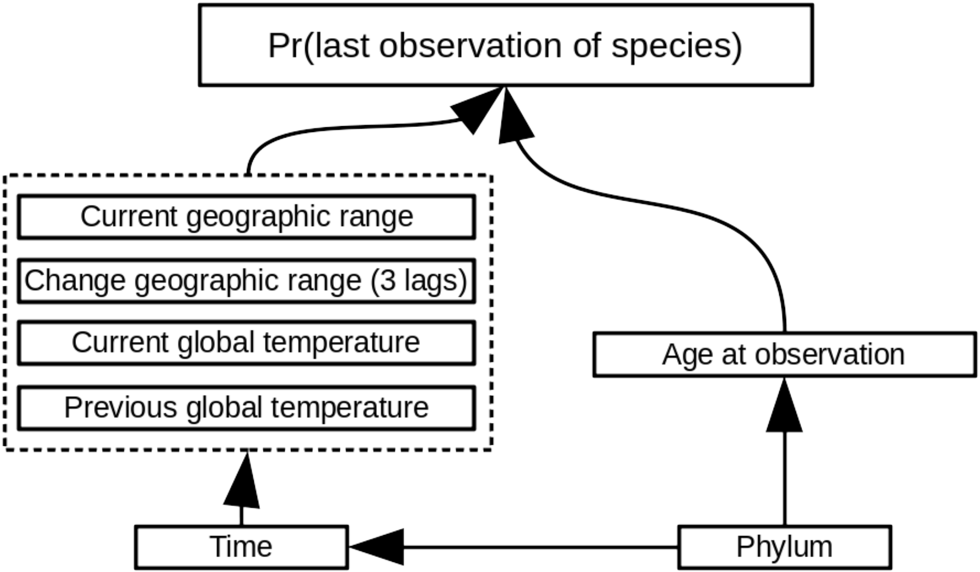
(UCL)

Global occurrences from Deep Sea Drilling Program and Ocean Drilling Project. – Lazarus. 1994. Math. Geo.; Spencer-Cervato. 1999. Palaeo. Elec.

How we're analyzing the data

- Encoding the past
  - Change in geographic range between current observation and previous observation.
  - Average global temperature at time of previous observation (Mg/Ca elemental ratio).
  - Age in millions of years at time of observation.
- Explore model adequacy using posterior predictive distribution.
- Estimate out-of-sample predictive performance using k-fold cross-validation.

A conceptual model for predicting extinction

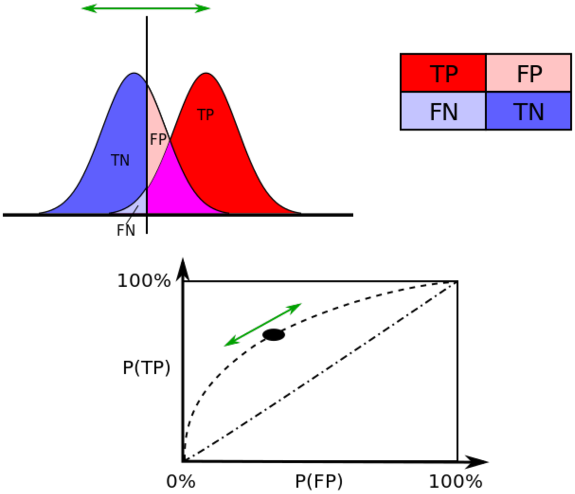


Measuring performance: confusion matrix

		True condition	
	Total population	Condition positive	Condition negative
Predicted condition	Predicted condition positive	**True positive**, Power	**False positive**, Type I error
	Predicted condition negative	**False negative**, Type II error	**True negative**

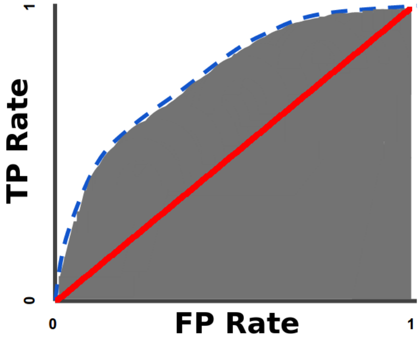
(wikimedia)

Measuring performance: Receiver Operating Characteristic



(wikimedia)

Measuring performance: AUC ROC

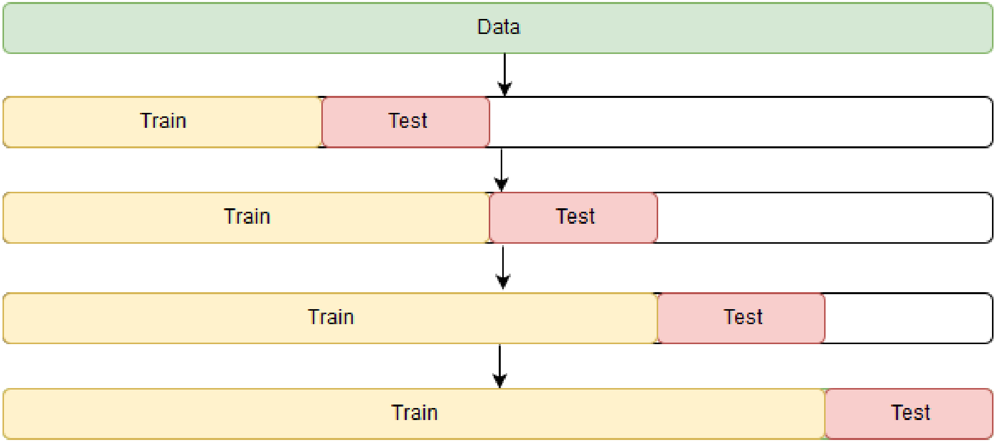


(https://goo.gl/91nEpM)

The area represents the probability of correct ranking of a random "extant"-"extinct" pair.

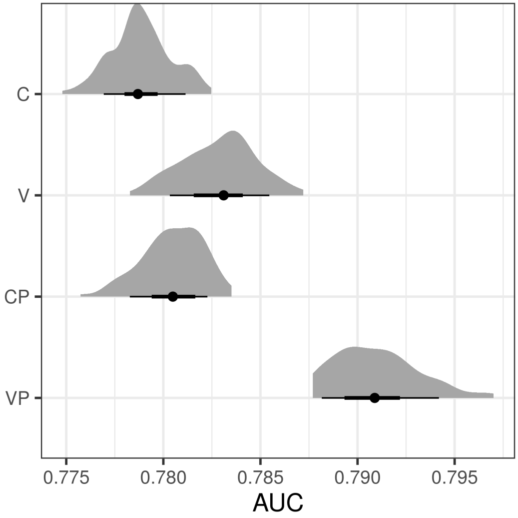
$$\mathrm{AUC} = \begin{cases} 0.5 & \text{non discrimination} \\ 0.6-0.7 & \text{poor} \\ 0.7-0.8 & \text{acceptable/fair} \\ 0.8-0.9 & \text{excellent/good} \\ >0.9 & \text{outstanding} \end{cases}$$

Measuring performance: *k*-fold cross-validation



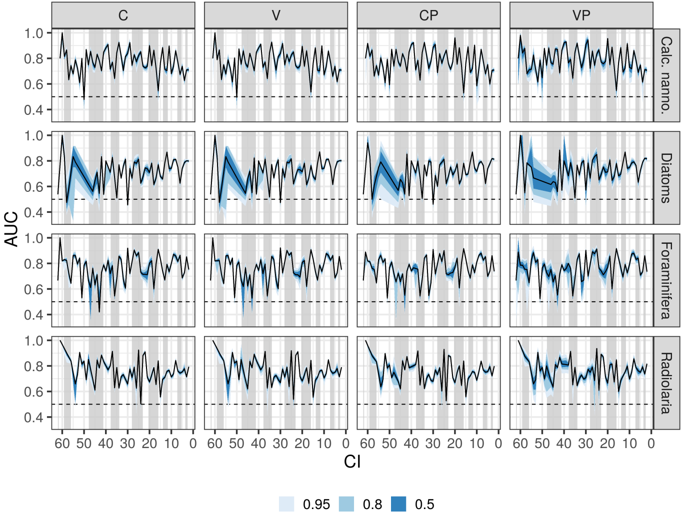
(Ken Williams, https://goo.gl/qLcfL8)

In-sample predictive performance, full dataset

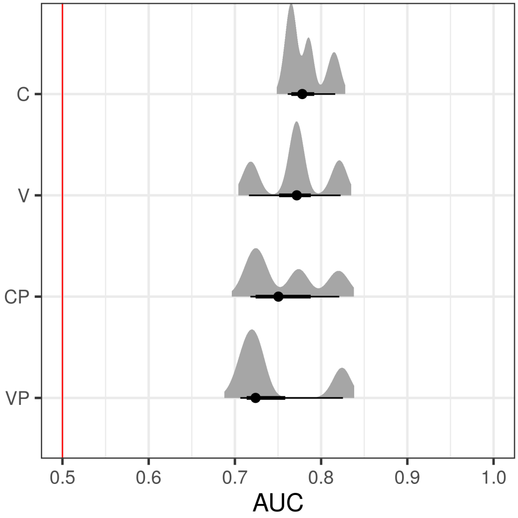


AUC $=$ 0.7-0.8 acceptable/fair

In-sample predictive performance, by time and taxa

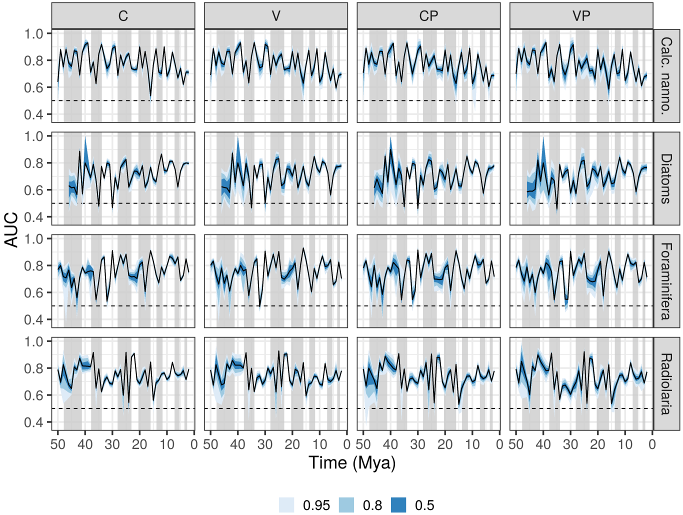


Cross-validation results, full dataset



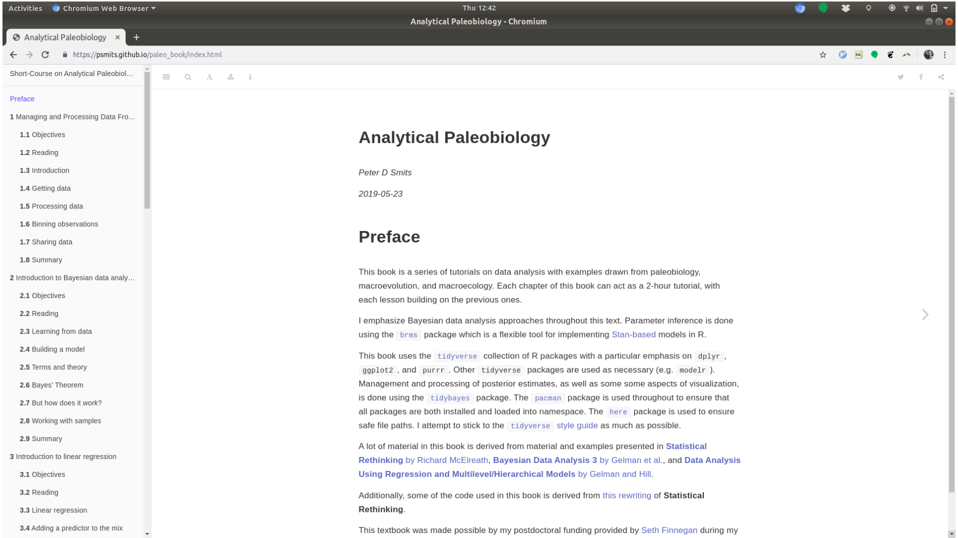
AUC = 0.7-0.8 acceptable/fair

Cross-validation results, by time and taxa



Summary

- The past matters. . .
  - Our best supported model includes either our historical covariates or allows all effects to vary over time.
- But not that much. . .
  - Models only average/fair expected out-of-sample performance.
- Allowing effects to vary over time is probably preferable to historical covariates – measures and accounts for variation which is important when predicting extinction in novel environments.
- Mechanisms behind changes to geographic range operate at sub-million year scales. Perhaps their effects are weak/masked at million (or greater) year scales.



Analytical Paleobiology

Peter D Smits

2019-05-23

Preface

This book is a series of tutorials on data analysis with examples drawn from paleobiology, macroevolution, and macroecology. Each chapter of this book can act as a 2-hour tutorial, with each lesson building on the previous ones.

I emphasize Bayesian data analysis approaches throughout this text. Parameter inference is done using the `brms` package which is a flexible tool for implementing Stan-based models in R.

This book uses the `tidyverse` collection of R packages with a particular emphasis on `dplyr`, `ggplot2`, and `purrr`. Other `tidyverse` packages are used as necessary (e.g. `modelr`). Management and processing of posterior estimates, as well as some some aspects of visualization, is done using the `tidybayes` package. The `pacman` package is used throughout to ensure that all packages are both installed and loaded into namespace. The `here` package is used to ensure safe file paths. I attempt to stick to the `tidyverse` style guide as much as possible.

A lot of material in this book is derived from material and examples presented in **Statistical Rethinking** by Richard McElreath, **Bayesian Data Analysis 3** by Gelman et al., and **Data Analysis Using Regression and Multilevel/Hierarchical Models** by Gelman and Hill.

Additionally, some of the code used in this book is derived from this rewriting of **Statistical Rethinking**.

This textbook was made possible by my postdoctoral funding provided by Seth Finnegan during my

https://psmits.github.io/paleo_book/index.html

Acknowledgements

- UC Berkeley
  - Adiel Klompmaker
  - Emily Orzechowski
  - Larry Taylor
  - Sara Kahanamoku
  - Josh Zimmt
- University of Oslo
  - Franziska Franeck
- Neptune DB/MFN - Berlin
  - David Lazarus

GitHub

psmits.github.io/ **trident**

@PeterDSmits

Neptune SANDBOX Berlin