

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.

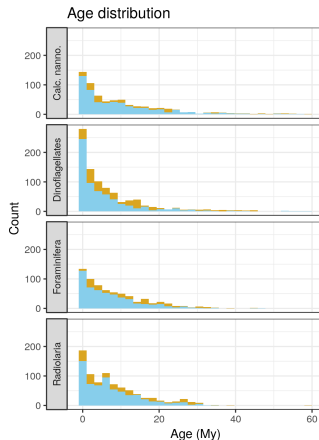
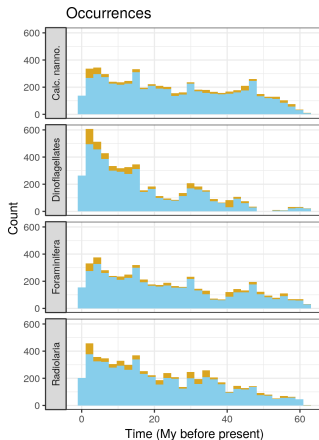
Predicting extinction

- ▶ 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)

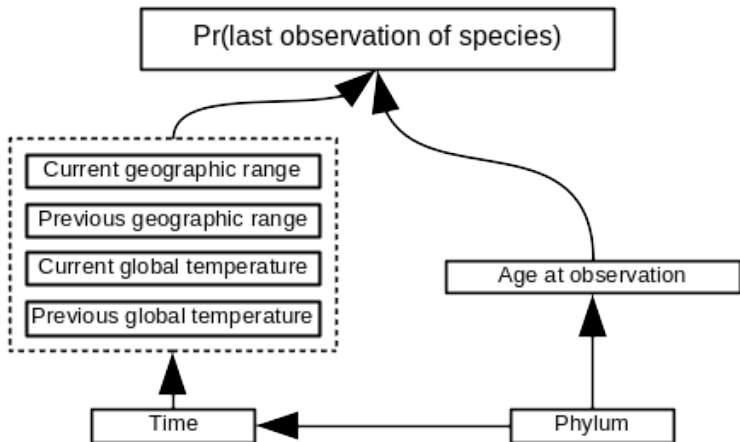


Global occurrences from Deep Sea Drilling Program and Ocean Drilling Project.
– Lazarus. 1994. Math. Geo.; Spencer-Cervato. 1999. Palaeo. Elec.

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

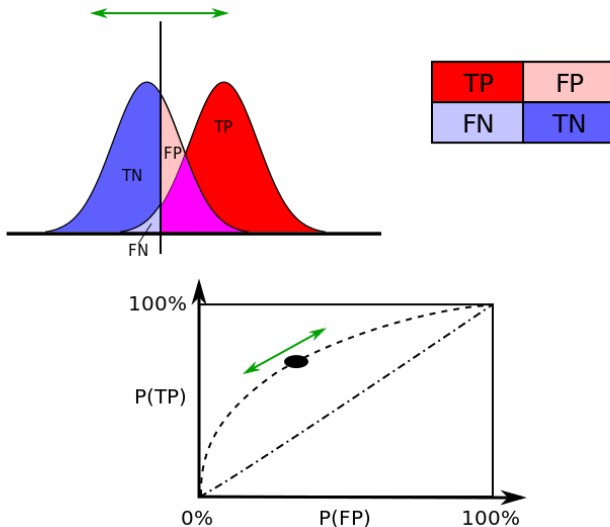


Measuring performance: confusion matrix

		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

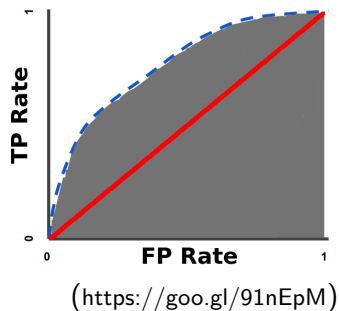
(wikimedia)

Measuring performance: Receiver Operating Characteristic



(wikimedia)

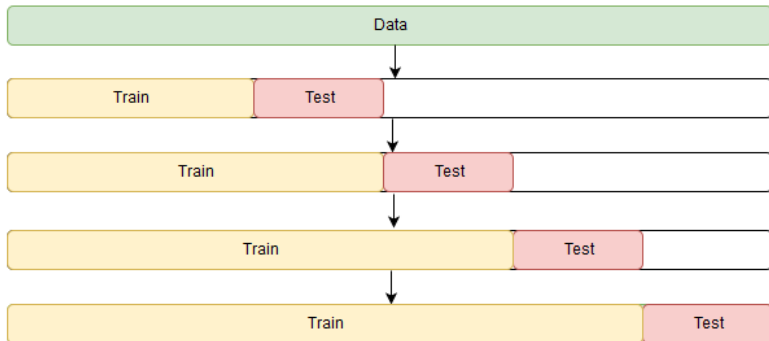
Measuring performance: AUC ROC



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

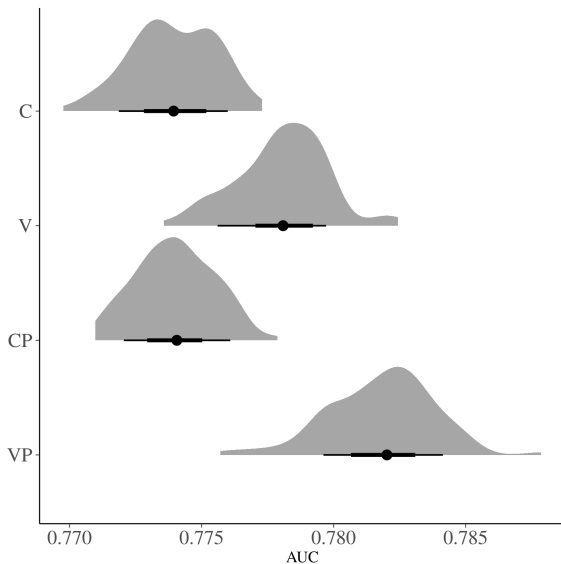
$$\text{AUC} = \begin{cases} 0.5 & \text{non discrimination} \\ 0.6 - 0.7 & \text{poor} \\ 0.7 - 0.8 & \text{acceptable/fair} \\ 0.8 - 0.9 & \text{excellent/good} \\ > 0.9 & \text{outstanding} \end{cases}$$

Measuring performance: k -fold cross-validation



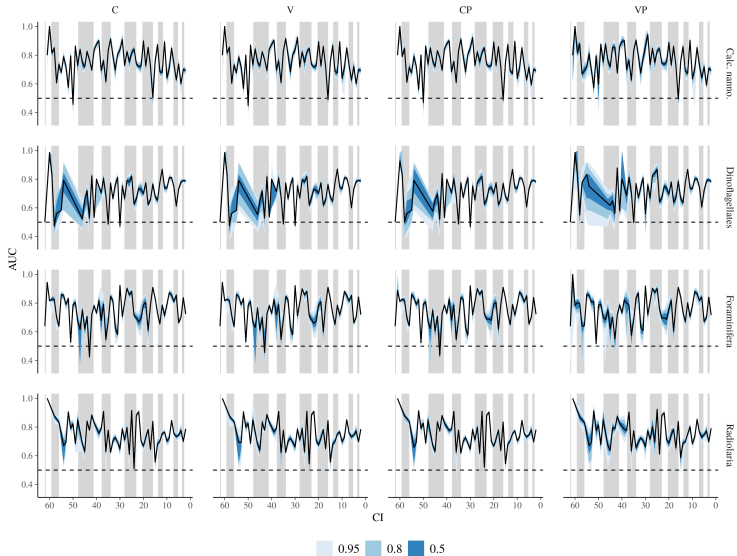
(Ken Williams, <https://goo.gl/qLcfL8>)

In-sample predictive performance, full dataset

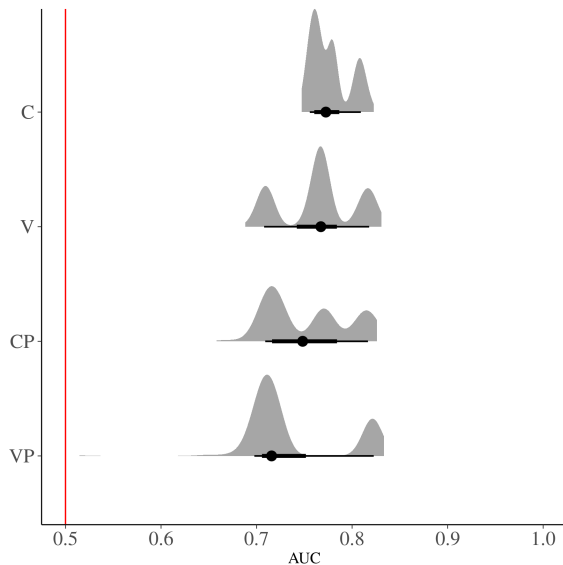


AUC = 0.7-0.8 acceptable/fair

In-sample predictive performance, by time and taxa

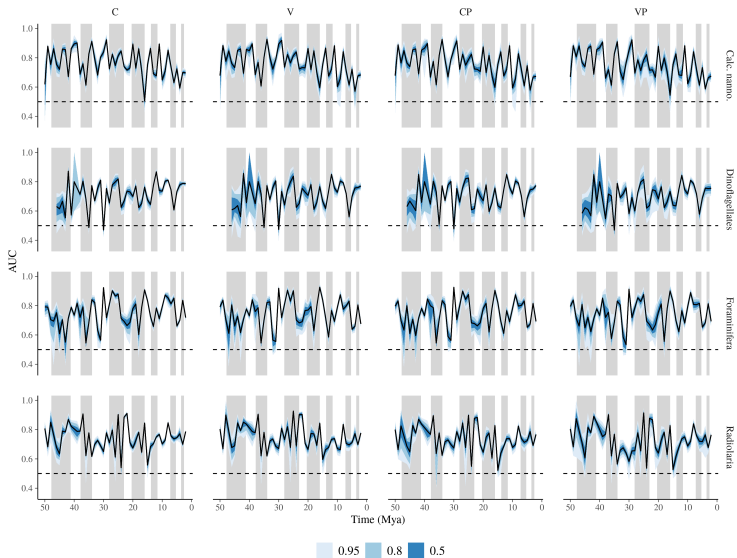


Cross-validation results, full dataset



AUC = 0.7-0.8 acceptable/fair

Cross-validation results, by time and taxa



Summary

- ▶ The past matters. . .
 - ▶ Our best supported model includes our historical covariates and allows all effects to vary over time.
- ▶ But not that much. . .
 - ▶ Models only average/fair expected out-of-sample performance.
- ▶ Mechanisms behind changes to geographic range operate at sub-million year scales. Perhaps their effects are weak/masked at million (or greater) year scales.

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

