

MarineSDMs

Marine Species Distribution Models

Ben Best

2023-11-23

Table of contents

1	Introduction	3
1.1	Background	3
1.2	Goals	3
1.3	Motivations	4
1.4	Process	4
1.5	Contribute	4
1.6	Features	6
1	Prepare	8
2	Prepare	9
3	Prepare	10
3.1	Environmental Predictors	11
3.1.1	Physiographic	11
3.1.2	Time Varying	11
3.1.3	Depth & Time Varying	11
4	Occurrences	12
4.1	Fetch OBIS	12
4.2	Filter occurrences	12
5	Pseudo-absences	14
5.1	All background	14
5.2	Mask by FAO areas	14
5.3	Use occurrences from same Family, different species	14
6	Environment	15
6.0.1	Physiographic	15
6.0.2	Time Varying	15
6.0.3	Depth & Time Varying	15

II Model	16
7 Model	17
8 Split	18
9 Fit	19
10 Calibrate	20
11 Predict	21
12 Evaluate	22
12.1 More Resources	22
III Combine	24
13 Combine	25
14 Ensemble	26
15 Mosaic	27
16 Taxa	28
17 Indicators	29
17.1 Diversity	29
17.2 Endemism	29
17.3 Extinction Risk	29
17.4 Functional Importance	30
17.5 Habitat Forming	30
17.6 Phylogenetic Uniqueness	30
17.7 Richness	30
17.8 Sensitivity	30
17.9 Trophic Index	30
18 Software	31
18.1 R	31
18.2 Python	31
18.3 Google Earth Engine	33
19 Organize	34
19.1 Partners	34
19.2 Portal	35

19.3 Timeline	35
19.3.1 Meetings	35
19.3.2 Sequence	36
References	37
Appendices	39
A Glossary	39
IV Explorations	40
AquaMaps Downscaled	41
AquaMaps Envelope	42
SDM using predicts in R	43
OBIS Top Species by Class	44

1 Introduction

1.1 Background

The best available global distributions are presently AquaMaps (Kaschner et al. 2006; Ready et al. 2010) with supplementation by IUCN RedList range maps¹. These have been used to calculate the biodiversity within national waters (Halpern et al. 2012) as well as beyond in the high seas (Visalli et al. 2020).

1.2 Goals

This book aims to capture the overview and details of modeling species distributions in the marine environment for the purposes of advancing the status quo of global and U.S. national species distributions along the following dimensions:

1. **Space**

The current AquaMaps distributions are $1/2^\circ$ (~ 55 km at equator), whereas the best available global bathymetry is $1/240^\circ$ (< 0.5 km).

2. **Time**

The current AquaMaps distributions are based on static climatic averages over all seasons, which does not capture temporal dynamics: seasonally within a year, nor long-term climate change trends. This will necessitate sampling the environment contemporaneously with species observations before fitting the model and predicting to different environmental snapshots.

3. **Environment**

Other environmental variables besides the initial physiographic (depth) and oceanographic (temperature, chlorophyll, primary productivity and ice) may elicit an improved statistical fit, related to species' environmental niche. Some candidates include: temperature fronts, eddy kinetic energy, distance from shore, distance from shelf.

4. **Biology**

Where sufficient observations exist, additional models should be developed highlighting differences between:

¹IUCN RedList range maps: <https://www.iucnredlist.org/resources/spatial-data-download>

- **Life stage**, e.g. larval vs adult.
- **Gender** where varies, such as male sperm whales being more cosmopolitan.
- **Subpopulations** for understanding metapopulation dynamics
- **Behavior**, such as migrating, feeding or breeding.

By definition MBONMarine Biodiversity Observation Network; see MarineBON.org is a network, so this is inclusive of and meant for all participants.

1.3 Motivations

- [AquaMaps.org](#)
AquaMaps (Kaschner et al. 2006; Ready et al. 2010) represents a massive amount of work to gather parameters for >33.5K marine species, including areas to mask out.
- [OBIS.org](#)
The Ocean Biogeographic Information System (Klein et al. 2019; Grassle 2000) is the central portal for continuously added observations with extra flags for quality control, all of which makes marine SDMs possible.
- **Modeling** methods have dramatically improved over time and are ripe for fresh application. The R package `dismo` originally came out with an [SDM vignette](#) as a practical supplement to their excellent review of SDMs (Elith and Leathwick 2009) and using the Maxent algorithm (Elith et al. 2011). The raster package furthered that (`raster`[sdm](#)) and now there's [terra sdm](#). Alongside these developments has been a boon of cloud-computing, particularly Google Earth Engine (Gorelick et al. 2017; Campos et al. 2023), allowing for dense global raster processing.
- The world is quickly moving towards a future trying to conserve 30% of the oceans by 2030, so called “**30 by 30**”. In the U.S., this is [America the Beautiful](#) initiative. We need biodiversity indicators to track progress. This push for conservation is driven by increasing impacts of **climate change**, as evidenced by marine heatwaves and shifts in population distributions.

1.4 Process

1.5 Contribute

We very much welcome your feedback, contributions and collaboration. Here are a few ways from least to most involved:

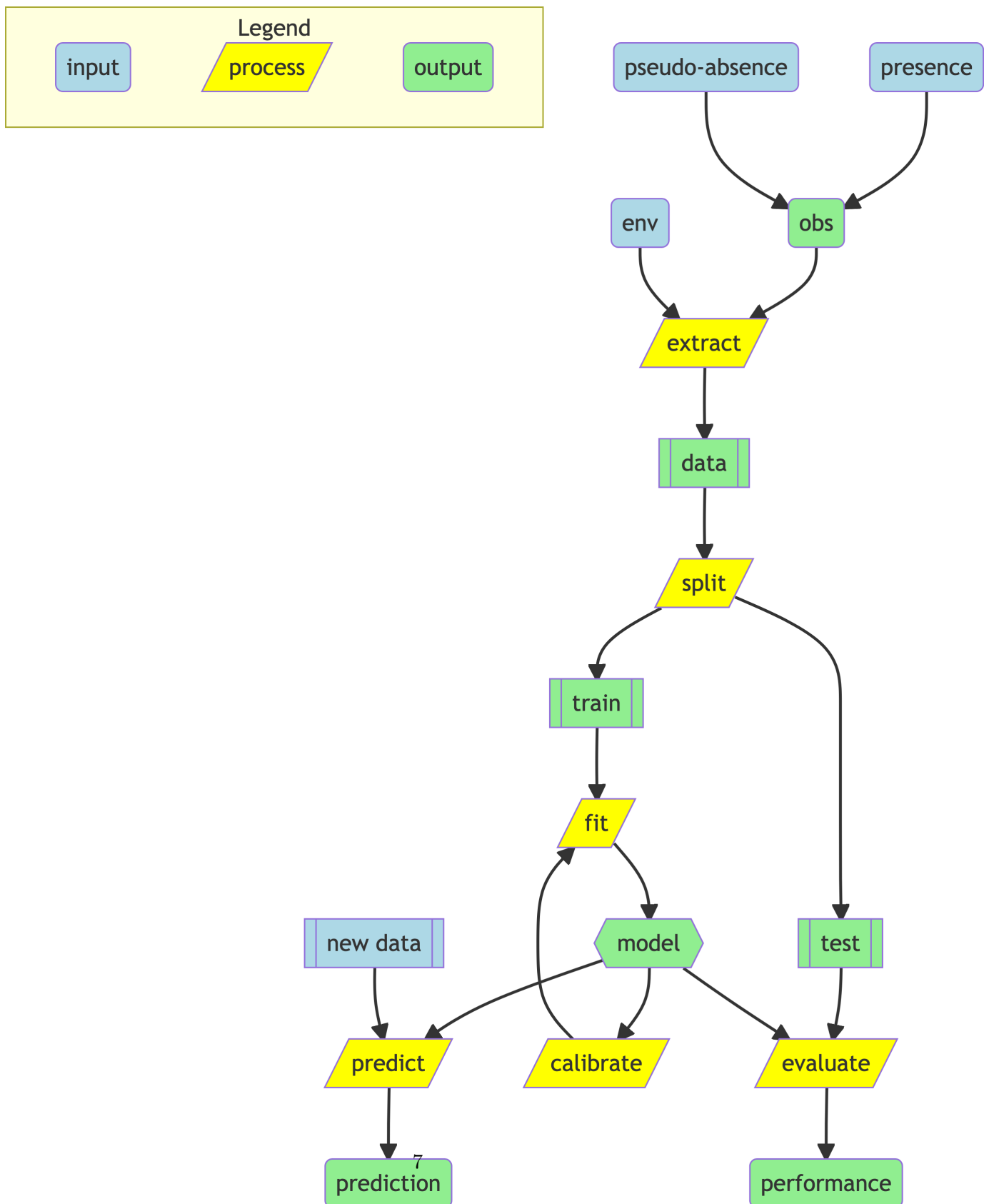




Figure 1.1: Diagram of SDM data preparation and model fitting.

1. Email Ben (ben@ecoquants.com) with any suggestions, including suggested revisions of this online book.

Note

Note that you can download this entire book as:

-  Adobe Acrobat [pdf](#) to add annotations; or
-  Microsoft Word [docx](#) to edit with Track Changes on.

These are available in the upper left navigation menu by clicking the download icon

.

2. Submit a [New Issue](#) on Github.
3. Click on “ Edit this Page” in the upper right. If you have a Github account, then you can fork this repository from owner “marinebon” to your username, edit the page(s) and submit a pull request.
4. If you are a regular contributor, you can be added to the collaborators of this repository to push changes directly (without needing a pull request).

1.6 Features

This Quarto book has a few cool features:

- Multiple formats
From the single set of source Quarto documents (*.qmd), several output formats are rendered: html, pdf, docx. This is particularly helpful when suggesting changes. It also lends itself well to being carved into manuscripts.
- Self-rendering
Github hosts the web pages (*.html), which get rendered from the source code (*.qmd) using a Github Action. So edits can be made simply through the web interface and all outputs get updated (html, pdf, docx). It also ensures the reproducibility of the document with a common setup environment.
- Mermaid diagrams
e.g., [Figure 1.1](#), [Figure 3.1](#), [Figure 7.1](#)
- Quarto document listings
- References
- Glossary

- Search

Part I

Prepare

2 Prepare

Prepare observations and environmental data for modeling

3 Prepare

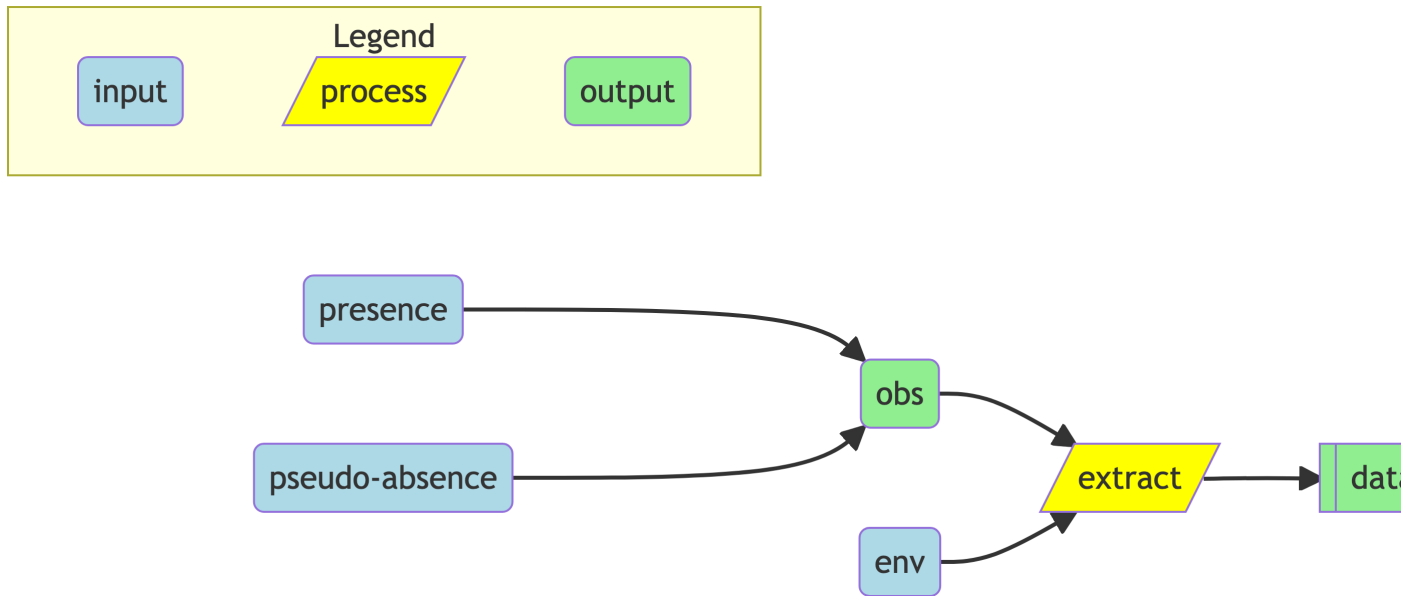


Figure 3.1: Diagram of SDM data preparation for model fitting.

- **obs**
observations: occurrences from OBIS; masked by FAO regions defined by AquaMaps (Skyttner 2020)
 - **presence**
OBIS: species occurrence
 - **absence**
OBIS not-species, but same family
- **env**
environment
- **tbl**
table of observations (presence and absence) with environmental values

3.1 Environmental Predictors

3.1.1 Physiographic

- `depth`
Bathymetric Depth
- `d2coast`
Distance to Coast
- `d2shelf`
Distance to Shelf

3.1.2 Time Varying

- `vgpm`
Vertically integrated primary Productivity model

3.1.3 Depth & Time Varying

- `temp`
Temperature, either sea-surface temperature (SST) or some modeled product from HyCOM, ROMS or Copernicus
- `salin`
Salinity

4 Occurrences

Fetch presence observations and filter for quality control

To describe:

- `robis`
- Filter based on quality flags
- Remove outliers

– [eks](#)

Tidy and Geospatial Kernel Smoothing for spatially filtering outlier observations

4.1 Fetch OBIS

4.2 Filter occurrences



Figure 4.1: Source: Kernel density estimates for tidy and geospatial data in the eks package

5 Pseudo-absences

Generate pseudo-absence or background environmental values to compare with occurrence environment

Describe various strategies for generating pseudo-absences.

- [Pseudo-absences](#) • [biomod2](#)
 - (Barbet-Massin et al. 2012)

5.1 All background

A common Maxent strategy is to feed all background points into Maxent, and then to use the resulting distribution as a null model. This is the default strategy in Maxent (Phillips et al. 2017; Phillips, Anderson, and Schapire 2006; Phillips and Dudík 2008).

5.2 Mask by FAO areas

The FAO areas applicable to species are included in the `aquamapsdata`, presumably from evaluating OBIS observations and the literature.

5.3 Use occurrences from same Family, different species

By using the same family, we can be sure that the pseudo-absences are ecologically similar to the species of interest.

6 Environment

Extract environmental predictors (static and/or dynamic) from various sources for observations (presence and pseudo-absence)

These data are also used at the prediction step.

6.0.1 Physiographic

- `depth`
Bathymetric Depth
- `d2coast`
Distance to Coast
- `d2shelf`
Distance to Shelf

6.0.2 Time Varying

- `vgpm`
Vertically integrated primary Productivity model

6.0.3 Depth & Time Varying

- `temp`
Temperature, either sea-surface temperature (SST) or some modeled product from HyCOM, ROMS or Copernicus
- `salin`
Salinity

Part II

Model

7 Model

Model the distribution of a species

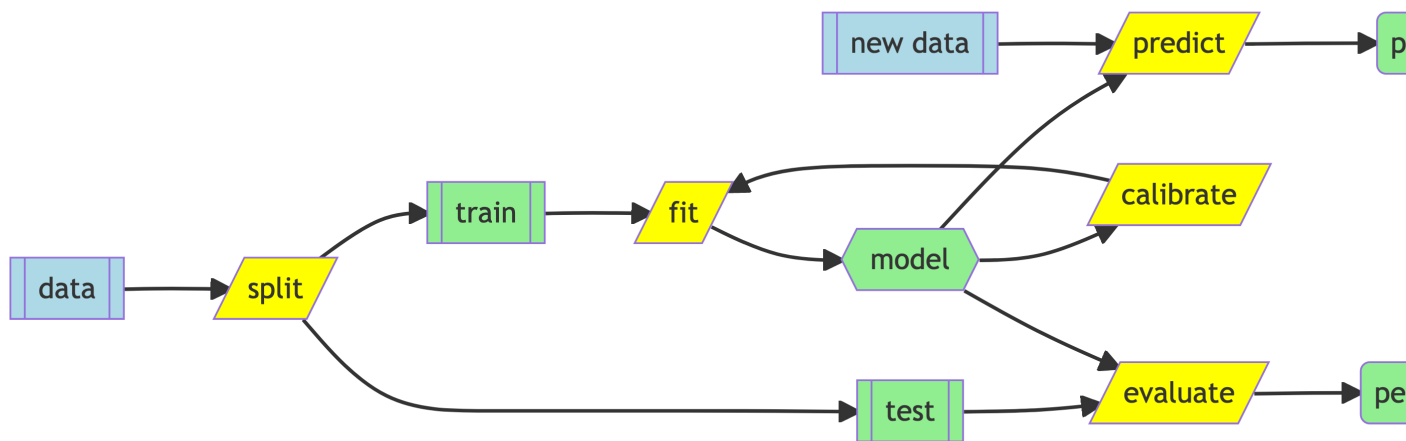


Figure 7.1: Diagram of SDM Modeling processes.

8 Split

Split data into training (to fit) and test (to evaluate prediction)

Data is often split so that ~20% of the observations (presence and absence) are set aside from the model fitting to be used for model evaluation.

The `k-fold` function is often used to split the data into k groups, and then the model is fit k times, each time using a different group as the test data and the remaining groups as the training data.

9 Fit

Fit environmental relationship distinguishing presence from absence of species

Model fitting in theory is quite complex, but quite simple in practice, with feeding the prepared data into the modeling function.

However there are MANY modeling techniques from which to choose. For instance check out 238 entries in [6 Available Models | The caret Package](#).

10 Calibrate

Calibrate model fit, i.e., model selection

The process of refining the model to only the most relevant environmental predictor terms is commonly called “Model Selection.” One of the most cited scientific paper of all time (Akaike 1974) is based on taking a most parsimonious approach to this process – the so called Akaike Information Criteria (AIC).

It is important to avoid using environmental predictors that are correlated with each other, since the effect of a predictor on the response could be the ecologically inverse, the result of explaining variance on the residuals of the other correlated predictor.

11 Predict

Predict distribution of the species with environmental relationship from fitted model

The prediction step applies the environmental relationships from the fitted model to a new set of data, typically the seascape of interest, and perhaps with some sort of temporal snapshot (e.g., climatic annual or monthly average).

12 Evaluate

Evaluate performance of the predicted model with the test data

Model evaluation uses the set aside test data from the earlier splitting to evaluate how well the model predicts the response of presence or absence. Since the test response data is binary [0,1] and the prediction from the model is continuous [0-1], a threshold needs to be applied to assign to convert the continuous response to binary. This is often performed through a Receiver Operator Characteristic (**ROC**) curve (Figure 12.1), which evaluates at each threshold the **confusion matrix** (Table 12.1).

Table 12.1: Confusion matrix to understand predicted versus observed.

		Predicted	
		0 (absence)	1 (presence)
Observed	0 (absence)	True absence	False presence
	1 (presence)	False absence	True presence

From the ROC curve, the area under the curve (**AUC**) is calculated, which is a measure of the model's ability to distinguish between presence and absence. AUC values range from 0 to 1, with 0.5 being no better than random, and 1 being perfect.

12.1 More Resources

- [Classification: ROC Curve and AUC | Machine Learning | Google for Developers](#)

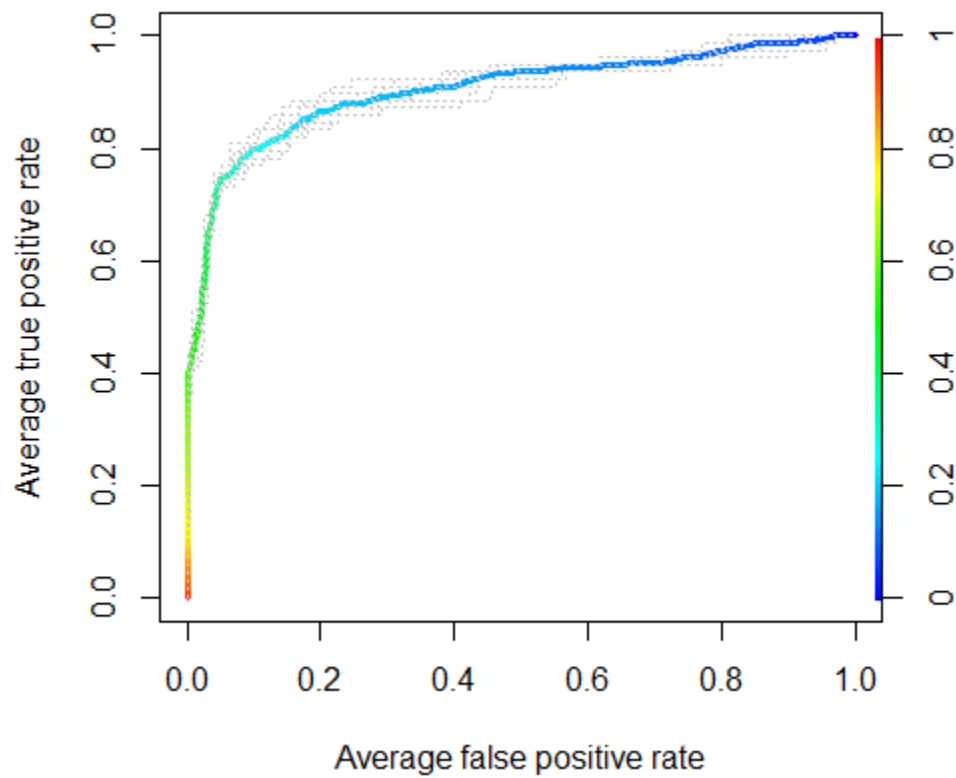


Figure 12.1: ROC curve generated by showing rates of false positive vs false negative as function of changing the threshold value (rainbow colors). Source: [ROCR: visualizing classifier performance in R](#)

Part III

Combine

13 Combine

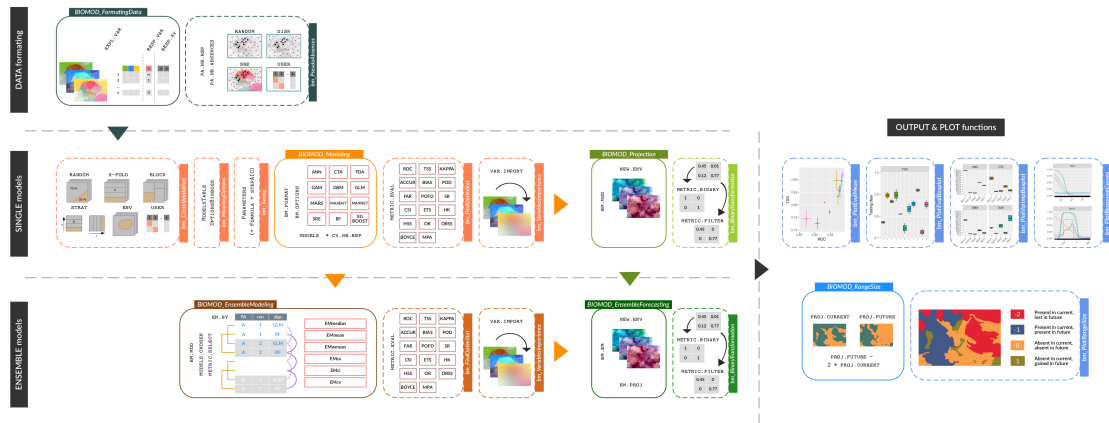
Combine SDMs from the same or multiple species

We look at combining SDMs to calculate biodiversity based on addressing questions of interest and relevance.

14 Ensemble

- biomod2

Species distribution modeling, calibration and evaluation, ensemble modeling



15 Mosaic

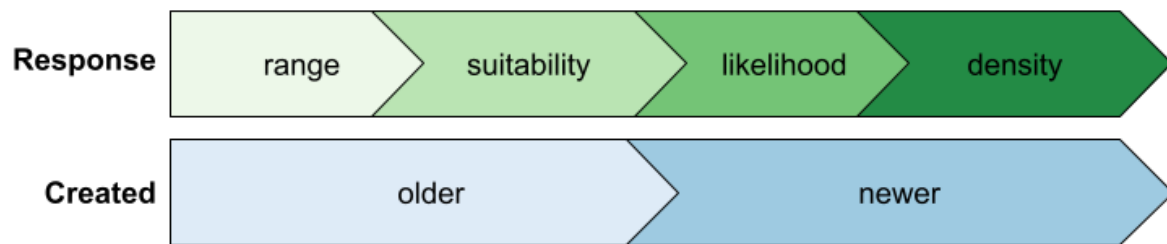


Figure 15.1: Hierarchy of preferred model outputs based on response type and age.

This is illustrated well by Figure [15.1](#).

16 Taxa

Group SDMs by taxonomy

Taxonomic groups (Tittensor et al. 2010) in the high seas (Visalli et al. 2020) were packaged with simple query statements in the draft R package [gmbi](#) (global marine biodiversity indicators).

17 Indicators

Calculate indicators of ecological or management interest beyond taxonomic groupings

17.1 Diversity

Here are the classic diversity indices from the R package `vegan`:

$$\begin{aligned} H &= - \sum_{i=1}^S p_i \log_b p_i && \text{Shannon-Weaver} \\ D_1 &= 1 - \sum_{i=1}^S p_i^2 && \text{Simpson} \\ D_2 &= \frac{1}{\sum_{i=1}^S p_i^2} && \text{inverse Simpson} \end{aligned}$$

where p_i is the proportion of species i , and S is the number of species so that $\sum_{i=1}^S p_i = 1$, and b is the base of the logarithm.

17.2 Endemism

Endemism could be measured as a function of the presence or average of the species range, given by either a global SDM converted to a binary range or using the existing IUCN range maps.

17.3 Extinction Risk

This is provided by IUCN RedList, as well as sometimes at a national level, such as Nature-Serve's [Conservation Status Ranks](#) for the U.S.

17.4 Functional Importance

17.5 Habitat Forming

Habitat forming species, such as coral, mangrove, seagrasses and kelp are especially important for biodiversity and ecosystem services.

17.6 Phylogenetic Uniqueness

17.7 Richness

17.8 Sensitivity

Sensitivity to specific human activities, such as shipping or fishing. Some activities may have different stages of development, such as construction versus operation of offshore wind energy.

17.9 Trophic Index



Figure 18.1: Source: [Kernel density estimates for tidy and geospatial data in the eks package](#)

18.3 Google Earth Engine

- [XEE](#)

XEE is a new Python package for Earth Engine that provides a set of functions to facilitate the use of Earth Engine API. It is designed to be used in Jupyter notebooks and Google Colab. [Documentation](#)

19 Organize

19.1 Partners

- *Confirmed*

- **AquaMaps**

- distribution maps for over 33,500 species of fishes, marine mammals and invertebrates*

- contacts:

- * Gabriel Reygondeau (g.reygondeau@oceans.ubc.ca)

- **DisMAP**

- Distribution Mapping and Analysis Portal (NOAA)*

- contacts:

- * Melissa Karp (melissa.karp@noaa.gov)

- * Roger Griffis (roger.b.griffis@noaa.gov)

- **MBON**

- Marine Biodiversity Observation Network (NOAA, NASA)*

- contacts:

- * Ben Best

- * Tylar Murray

- * Dan Otis

- * Frank Muller-Karger

- **OBIS**

- Ocean Biodiversity Information System (UNESCO)*

- contacts:

- * Candido Principe De Souza, Silas (s.principe@unesco.org)

- * Ward Appeltans (w.appeltans@unesco.org)

- *Proposed*

- **FishGLOB**
fish biodiversity under global change (global bottom trawl)
 contacts:
 - * Malin Pinsky
- **MGEL**
Marine Geospatial Ecology Lab, Duke
 contacts:
 - * Jason Roberts
 - * Patrick Halpin
- **PMEL**
Pacific Marine Environmental Lab, NOAA
 contacts:
 - * Heather Welch
 - * Elliot Hazen

19.2 Portal

- Candidate website:
MarineSpeciesMaps.org (BDB registered domain)
 - similar to:
 - * MarineRegions.org
 spatial authority
 - * MarineSpecies.org
 taxonomic authority

19.3 Timeline

19.3.1 Meetings

- **2024-01-23**
 Marine Biodiversity Summit in DC
 (organized by Gabrielle, Emmett)
- **2024-03-21**
 SDM meeting in New Orleans
 (organized by BDB; so far AquaMaps)

19.3.2 Sequence

- 1.

References

- Akaike, H. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19 (6): 716-723.
- Barbet-Massin, Morgane, Frédéric Jiguet, Cécile Hélène Albert, and Wilfried Thuiller. 2012. "Selecting Pseudo-Absences for Species Distribution Models: How, Where and How Many?" *Methods in Ecology and Evolution* 3 (2): 327–38. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>.
- Campos, João C., Nuno Garcia, João Alírio, Salvador Arenas-Castro, Ana C. Teodoro, and Neftalí Sillero. 2023. "Ecological Niche Models Using MaxEnt in Google Earth Engine: Evaluation, Guidelines and Recommendations." *Ecological Informatics* 76 (September): 102147. <https://doi.org/10.1016/j.ecoinf.2023.102147>.
- Elith, Jane, and John R. Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 677–97. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>.
- Elith, Jane, Steven J. Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J. Yates. 2011. "A Statistical Explanation of MaxEnt for Ecologists." *Diversity and Distributions* 17 (1): 43–57. <http://dx.doi.org/10.1111/j.1472-4642.2010.00725.x>.
- Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." *Remote Sensing of Environment*, Big remotely sensed data: Tools, applications and experiences, 202 (December): 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Grassle, J. Frederick. 2000. "The Ocean Biogeographic Information System (OBIS): An on-Line, Worldwide Atlas for Accessing, Modeling and Mapping Marine Biological Data in a Multidimensional Geographic Context." *Oceanography* 13 (3): 5–7. <https://www.jstor.org/stable/43924357>.
- Halpern, Benjamin S., Catherine Longo, Darren Hardy, Karen L. McLeod, Jameal F. Samhouri, Steven K. Katona, Kristin Kleisner, et al. 2012. "An Index to Assess the Health and Benefits of the Global Ocean." *Nature*. <https://doi.org/10.1038/nature11397>.
- Kaschner, K., R. Watson, A. W. Trites, and D. Pauly. 2006. "Mapping World-Wide Distributions of Marine Mammal Species Using a Relative Environmental Suitability (RES) Model." *Marine Ecology Progress Series* 316 (July): 285–310. <https://doi.org/10.3354/meps316285>.
- Klein, Eduardo, Ward Appeltans, Pieter Provoost, Hanieh Saeedi, Abigail Benson, Lenore Bajona, Ana Carolina Peralta, and R. Sky Bristol. 2019. "OBIS Infrastructure, Lessons Learned, and Vision for the Future." *Frontiers in Marine Science* 6. <https://www.frontiersin.org/articles/10.3389/fmars.2019.00588>.

- Phillips, Steven J., Robert P. Anderson, Miroslav Dudík, Robert E. Schapire, and Mary E. Blair. 2017. “Opening the Black Box: An Open-Source Release of Maxent.” *Ecography* 40 (7): 887–93. <https://doi.org/10.1111/ecog.03049>.
- Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. “Maximum Entropy Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3): 231–59. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Phillips, Steven J., and Miroslav Dudík. 2008. “Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation.” *Ecography* 31 (2): 161–75. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>.
- Ready, Jonathan, Kristin Kaschner, Andy B. South, Paul D. Eastwood, Tony Rees, Josephine Rius, Eli Agbayani, Sven Kullander, and Rainer Froese. 2010. “Predicting the Distributions of Marine Organisms at the Global Scale.” *Ecological Modelling* 221 (3): 467–78. <https://doi.org/10.1016/j.ecolmodel.2009.10.025>.
- Skyttner, Markus. 2020. “Aquamapsdata: Curated Data from AquaMaps.org.” <https://github.com/raquamaps/aquamapsdata>.
- Tittensor, Derek P., Camilo Mora, Walter Jetz, Heike K. Lotze, Daniel Ricard, Edward Vanden Berghe, and Boris Worm. 2010. “Global Patterns and Predictors of Marine Biodiversity Across Taxa.” *Nature* 466 (7310): 1098–1101. <https://doi.org/10.1038/nature09329>.
- Visalli, Morgan E., Benjamin D. Best, Reniel B. Cabral, William W. L. Cheung, Nichola A. Clark, Cristina Garilao, Kristin Kaschner, et al. 2020. “Data-Driven Approach for Highlighting Priority Areas for Protection in Marine Areas Beyond National Jurisdiction.” *Marine Policy*, March, 103927. <https://doi.org/10.1016/j.marpol.2020.103927>.

A Glossary

```
glossary::glossary_table(as_kable=F) |>  
  knitr::kable("pipe", escape = F, row.names = F)
```

term	definition
MBON	Marine Biodiversity Observation Network; see MarineBON.org

Part IV

Explorations

AquaMaps Downscaled

Downscale AquaMaps from $1/2^\circ$ to GEBCO $1/240^\circ$ using Google Earth Engine and Shiny.

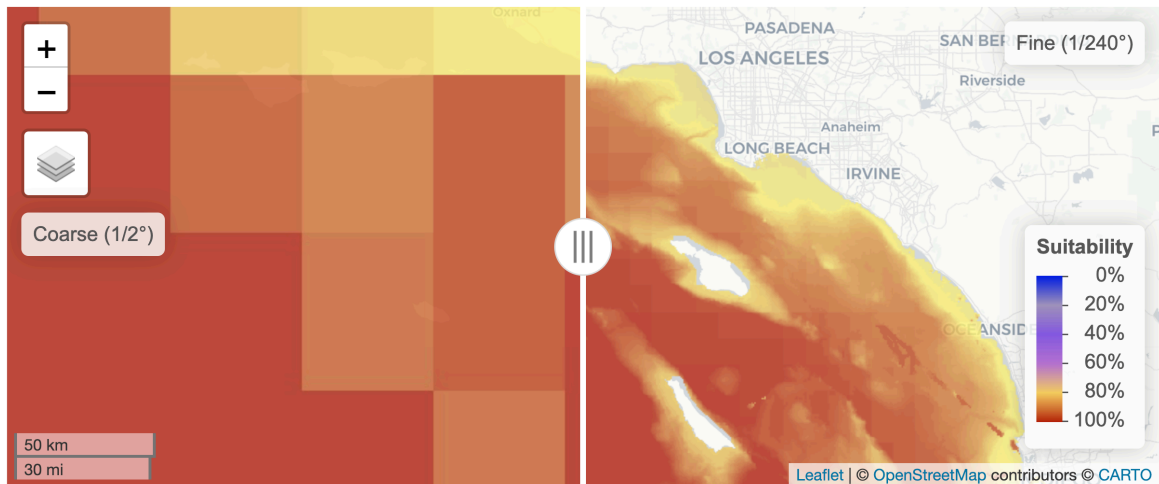
- [website](#)
- [code](#)

DRAFT
AquaMaps
Downscaling

Species

Balaenoptera musculus

Parameters

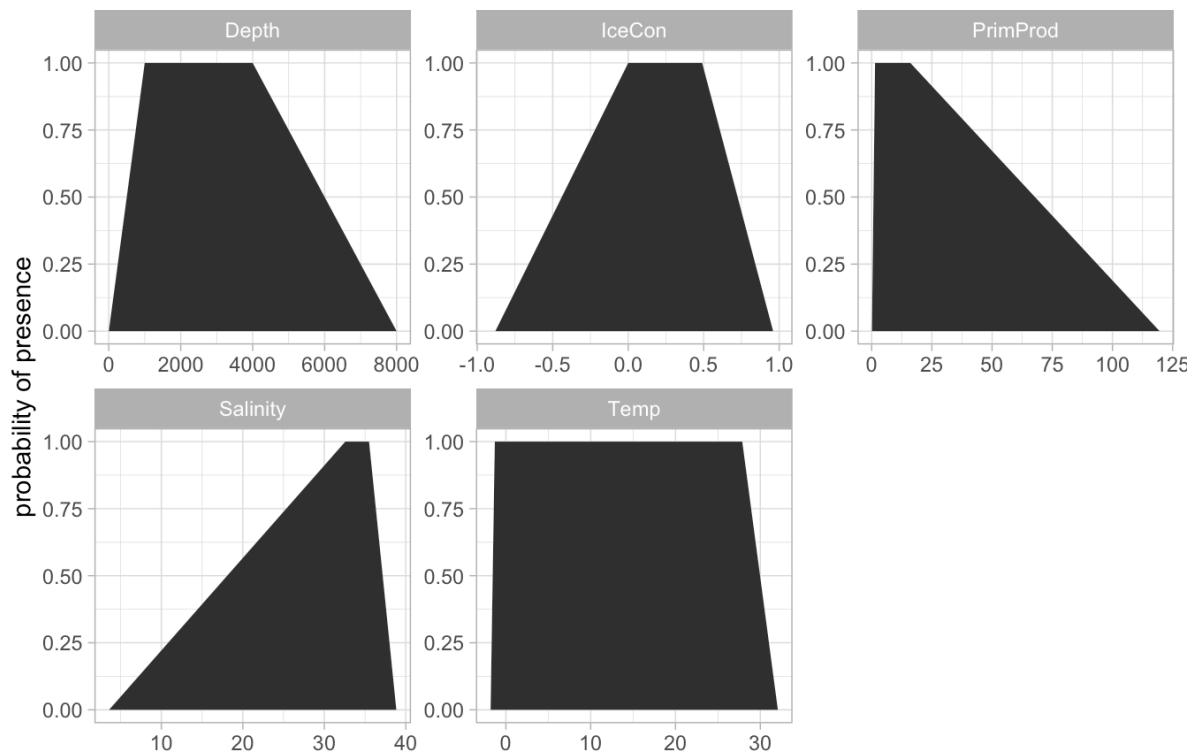


AquaMaps Envelope

Extract and plot AquaMaps environmental envelope, ramp rasters, using R.

- [website](#)
- [code](#)

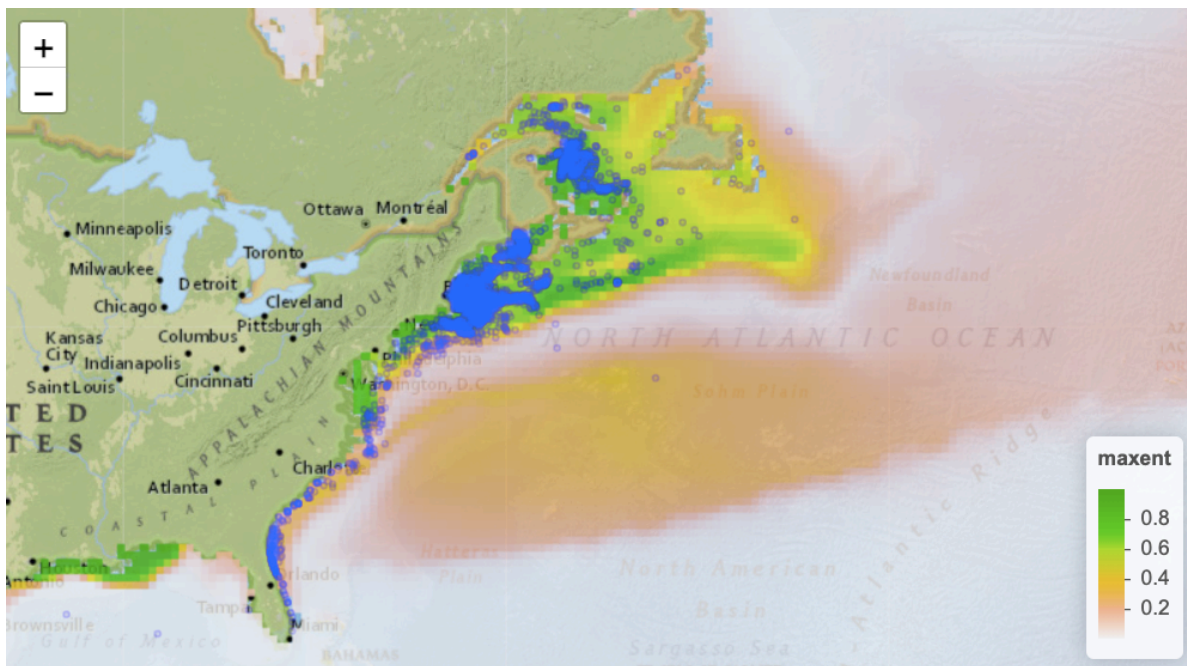
blue whale
environmental envelope



SDM using predicts in R

Predict species distribution of N Atlantic right whale using OBIS occurrences and **predicts** package in R.

- [website](#)
- [code](#)



OBIS Top Species by Class

Extract the species with the most numerous observations by unique Class from the OBIS parquet archive in R.

- [website](#)
- [code](#)

	phylum	class	scientificName	AphiaID	date_min	date_max	r
76	Chordata	Aves	Larus fuscus	137142	1758-07-02	2021-12-31	2
88	Chordata	Teleostei	Clupea pallasii	151159	1867-07-21	2022-09-06	1
85	Chordata	Mammalia	Mirounga leonina	231413	1758-07-02	2019-12-02	1
19	Arthropoda	Malacostraca	Pandalus jordani	515469	1937-10-28	2022-09-13	
160	Mollusca	Cephalopoda	Loligo reynaudii	220316	1960-04-16	2006-05-11	
81	Chordata	Elasmobranchii	Carcharhinus melanopterus	105795	1824-07-01	2021-09-28	
15	Arthropoda	Copepoda	Calanus finmarchicus	104464	1872-09-14	2020-12-14	
90	Chordata		Chelonia mydas	137206	1758-07-02	2023-06-14	
169	Myzozoa	Dinophyceae	Tripos fusus	840626	1834-07-02	2022-09-08	
187	Ochrophyta	Phaeophyceae	Macrocystis pyrifera	232231	1885-12-01	2020-10-16	