Habitat modelling
using marine animal telemetry

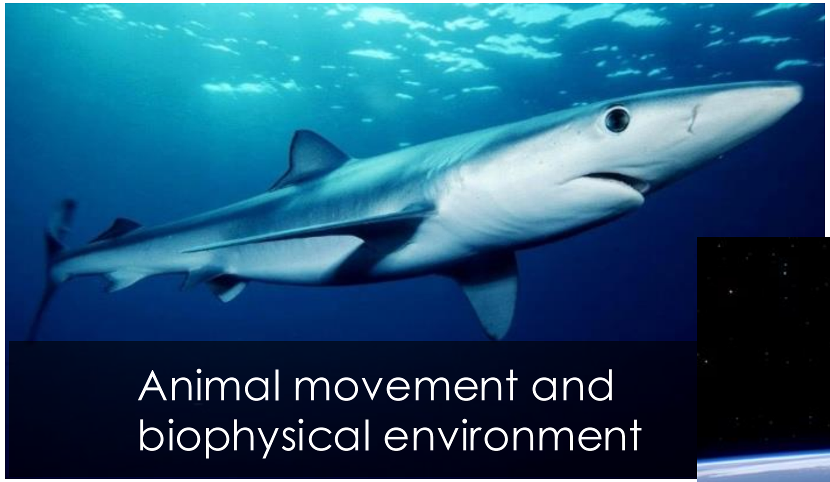
David March (david.march@uv.es)

Barcelona, 15 May 2025

Research lines

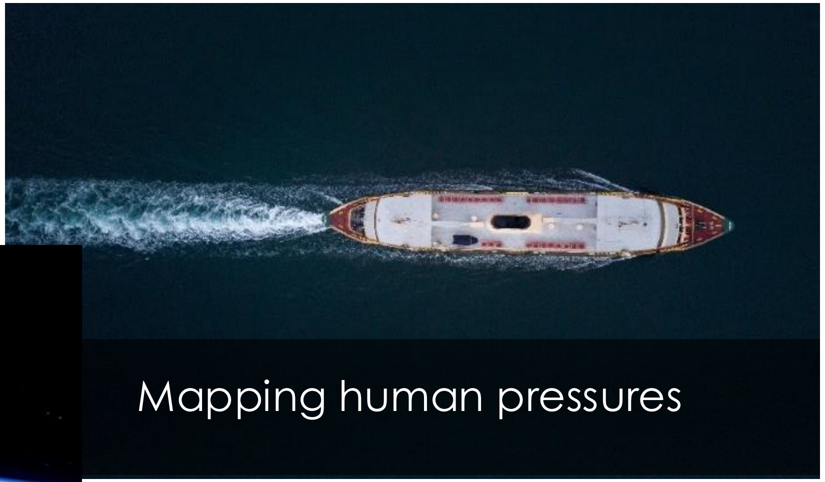
VNIVERSITAT DE VALÈNCIA

ICBiBE
Institut Universitari Cavanilles de Biodiversitat i Biologia Evolutiva

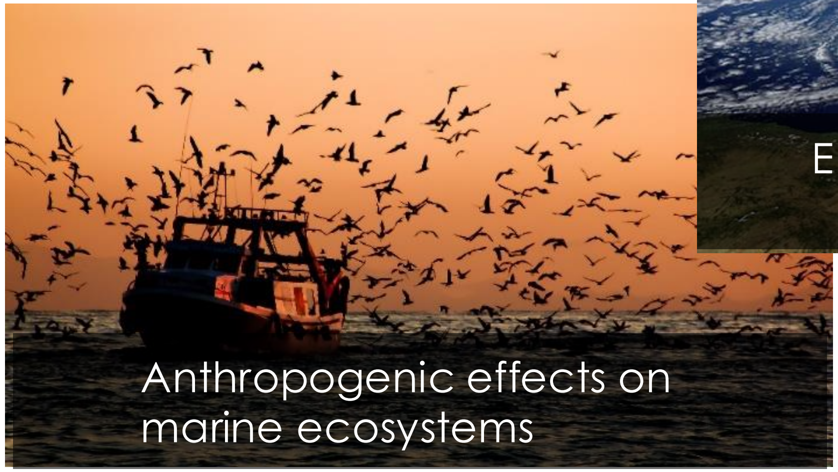


Animal movement and biophysical environment

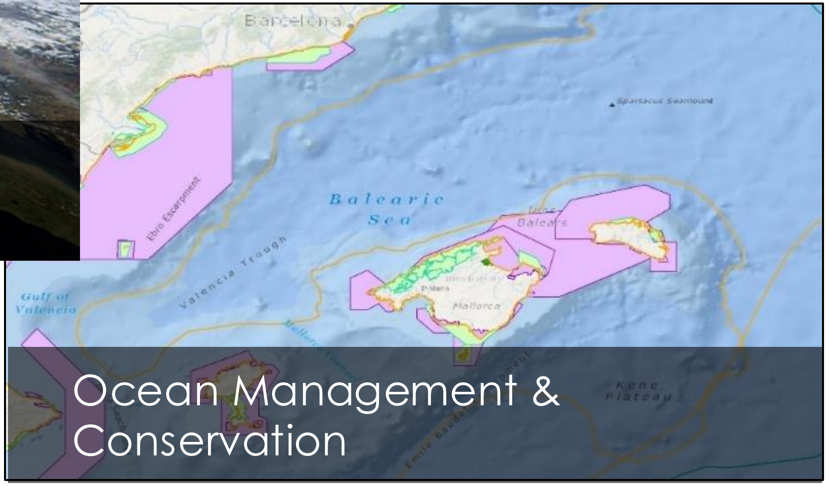
Electronic Monitoring Systems and Earth Observations



Mapping human pressures



Anthropogenic effects on marine ecosystems



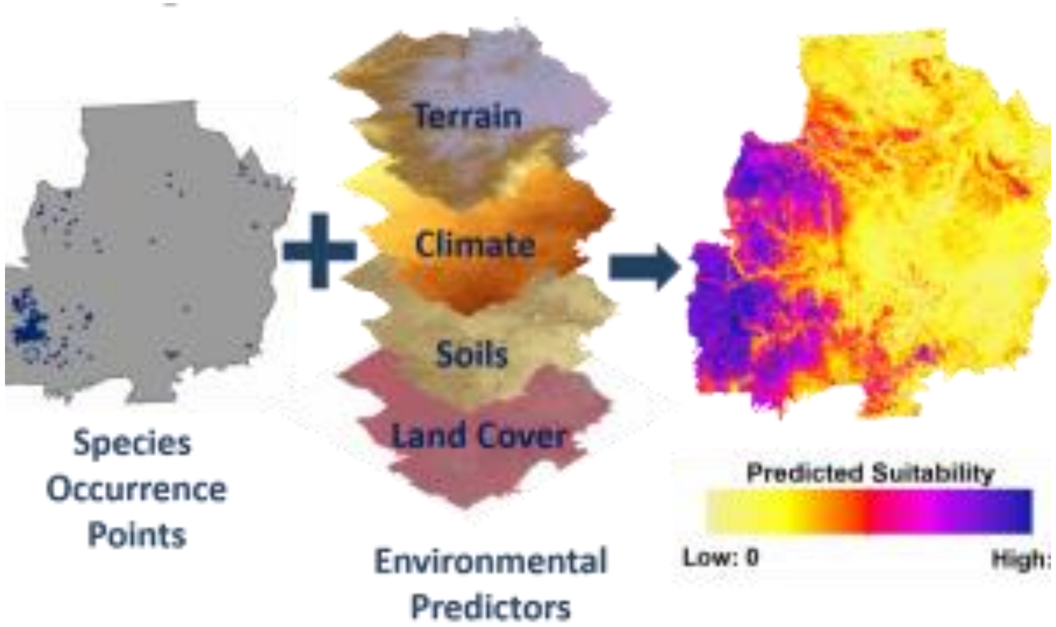
Ocean Management & Conservation

Introduce yourself :)

Outline

09:15 - 10:00	Introduction
10:00 - 11:30	Hands-on with R - Part 1 Data Preparation
11:30 - 12:00	Break
12:00 - 13:30	Hands-on with R - Part 2 Habitat modeling

What is a habitat model?

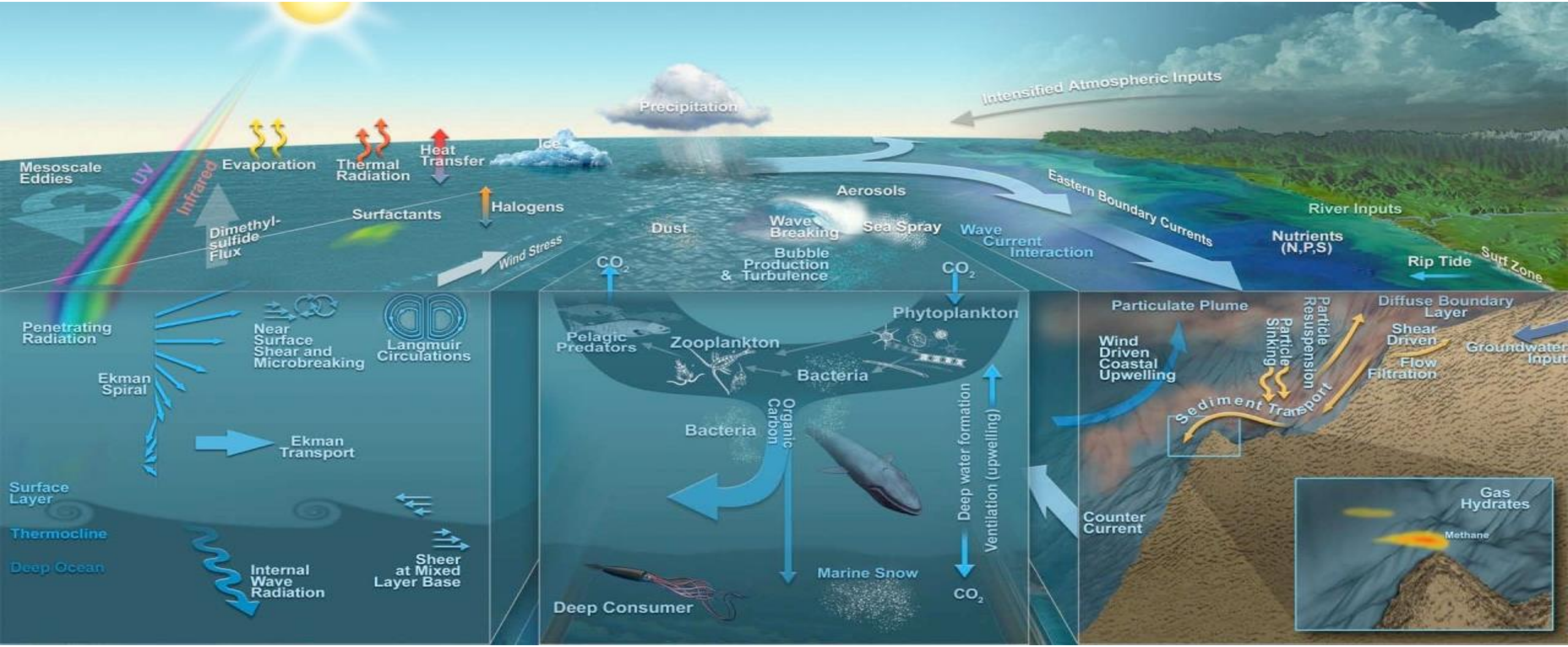


Also known as

Species distribution modelling OR Environmental niche modelling OR (Ecological) niche modelling OR Predictive habitat distribution modelling OR Climate envelope modelling

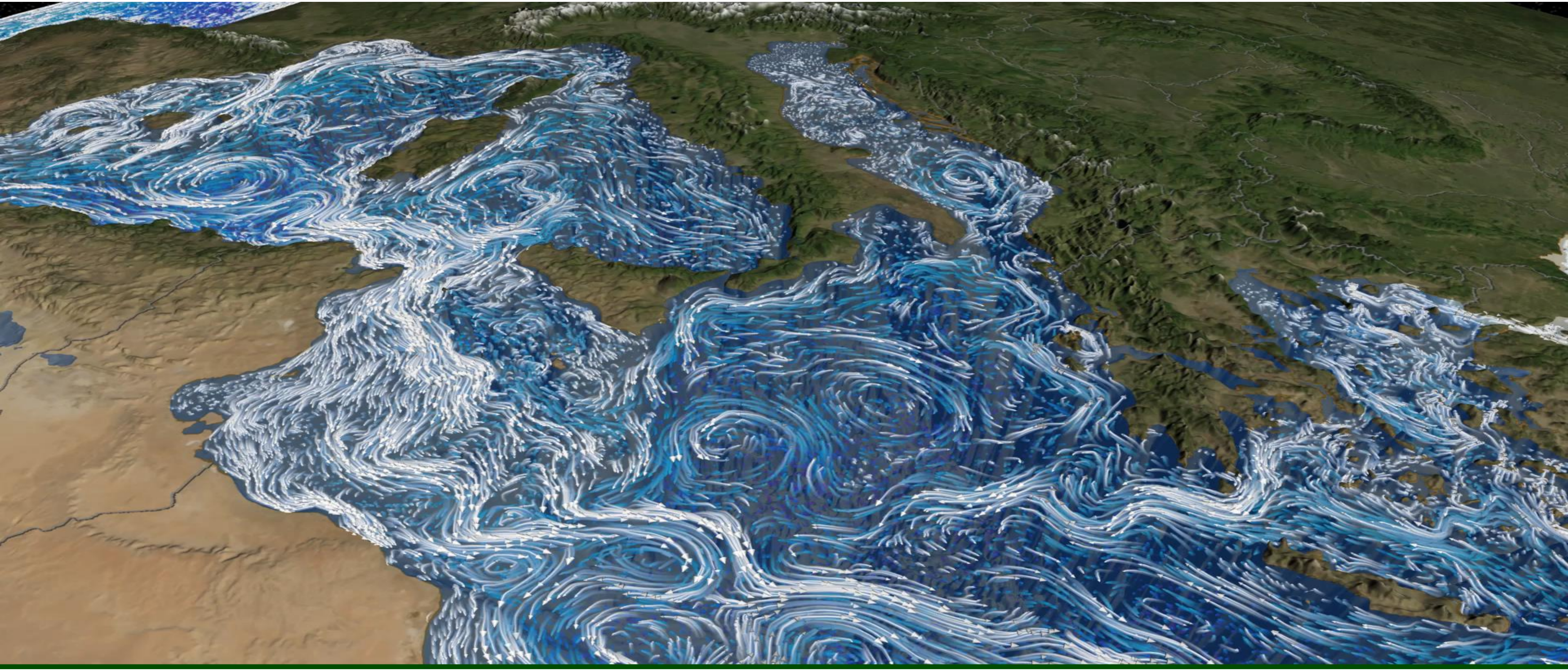
What are the particularities of marine realm for SDM?

Complex metocean variables (aerial, surface, sub-surface)



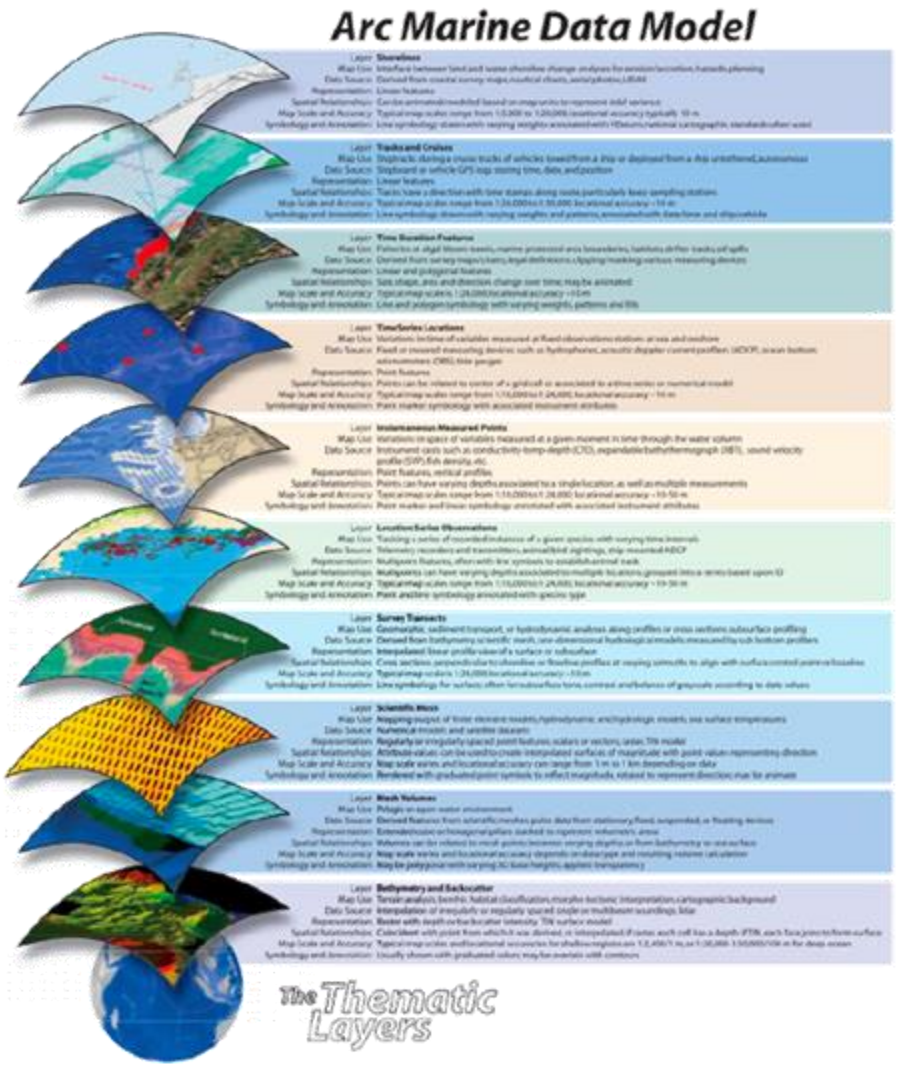
What are the particularities of marine realm for SDM?

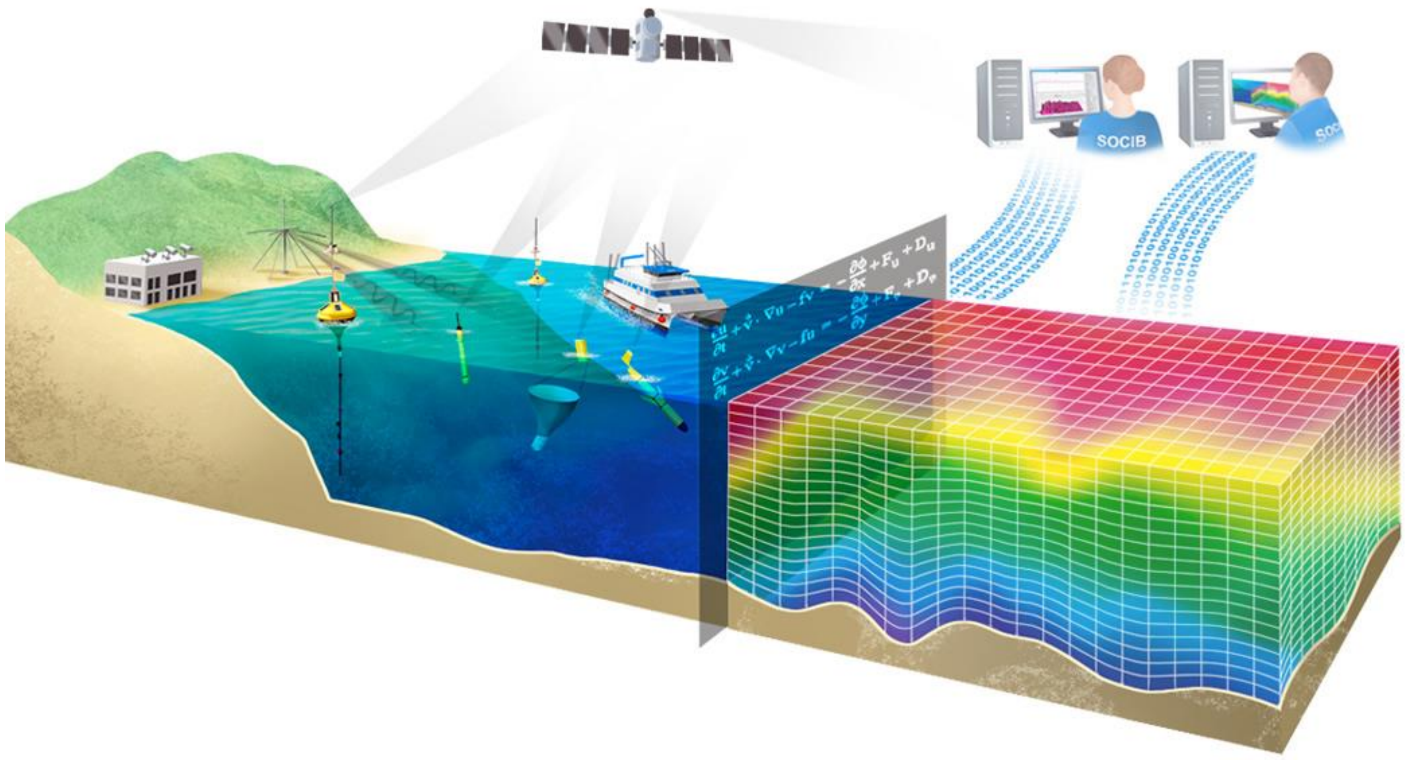
Highly dynamic environment



What are the particularities of marine realm for SDM?

Diversity of sources and resolutions

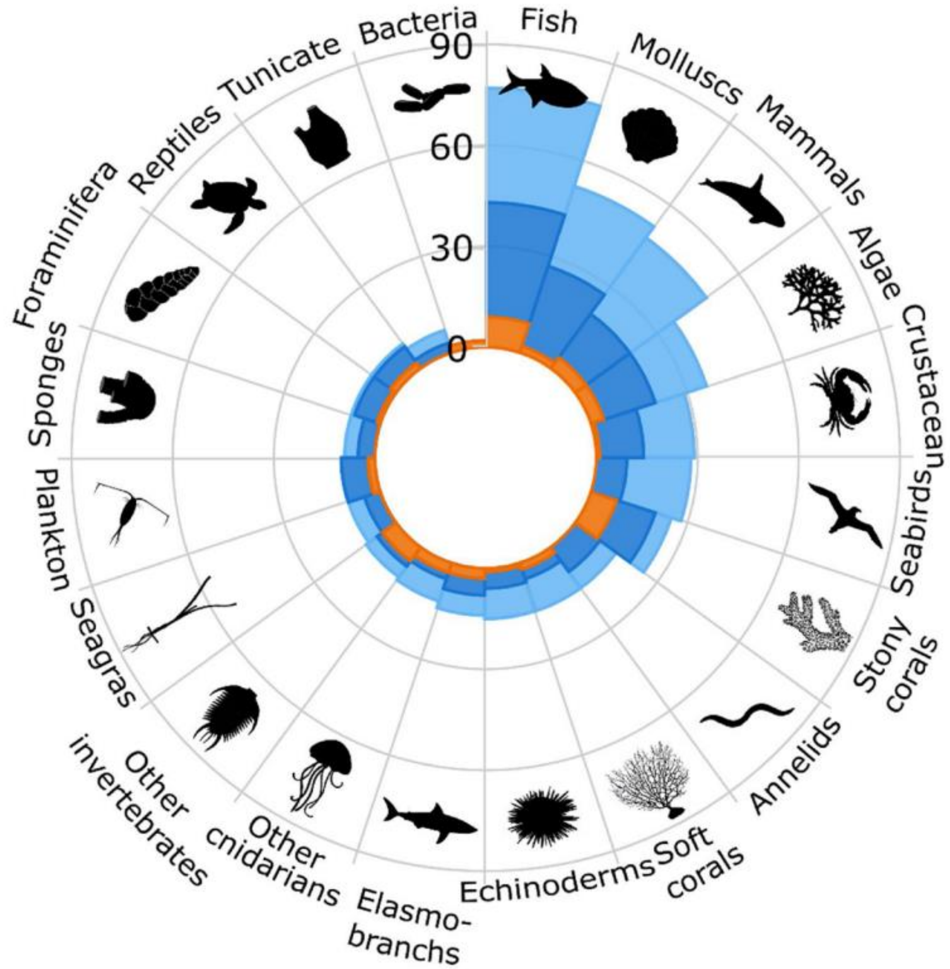




Remote sensing & in-situ observations vs. Numerical models

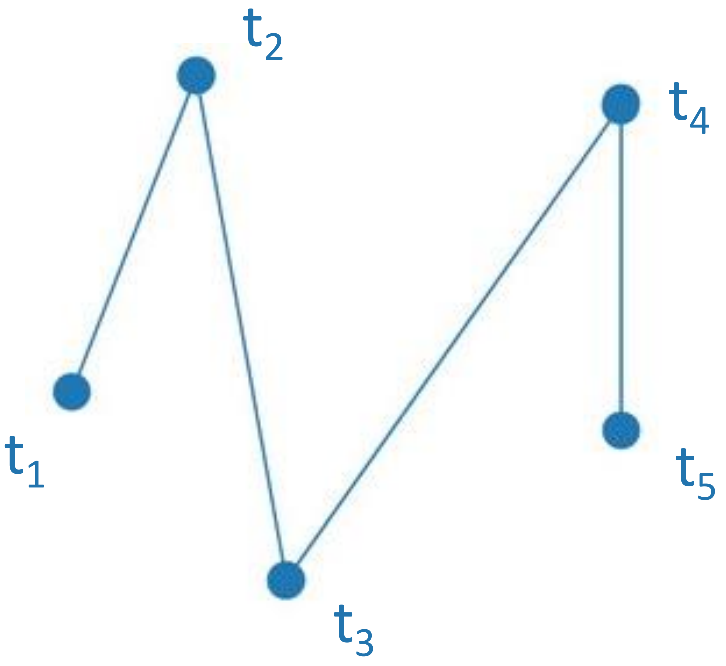
What are the particularities of marine realm for SDM?

Diversity of taxonomic groups



Melo-Merino. 2020. Ecological Modelling. https://doi.org/10.1016/j.ecolmodel.2019.108837

What are the particularities of telemetry data for SDM?



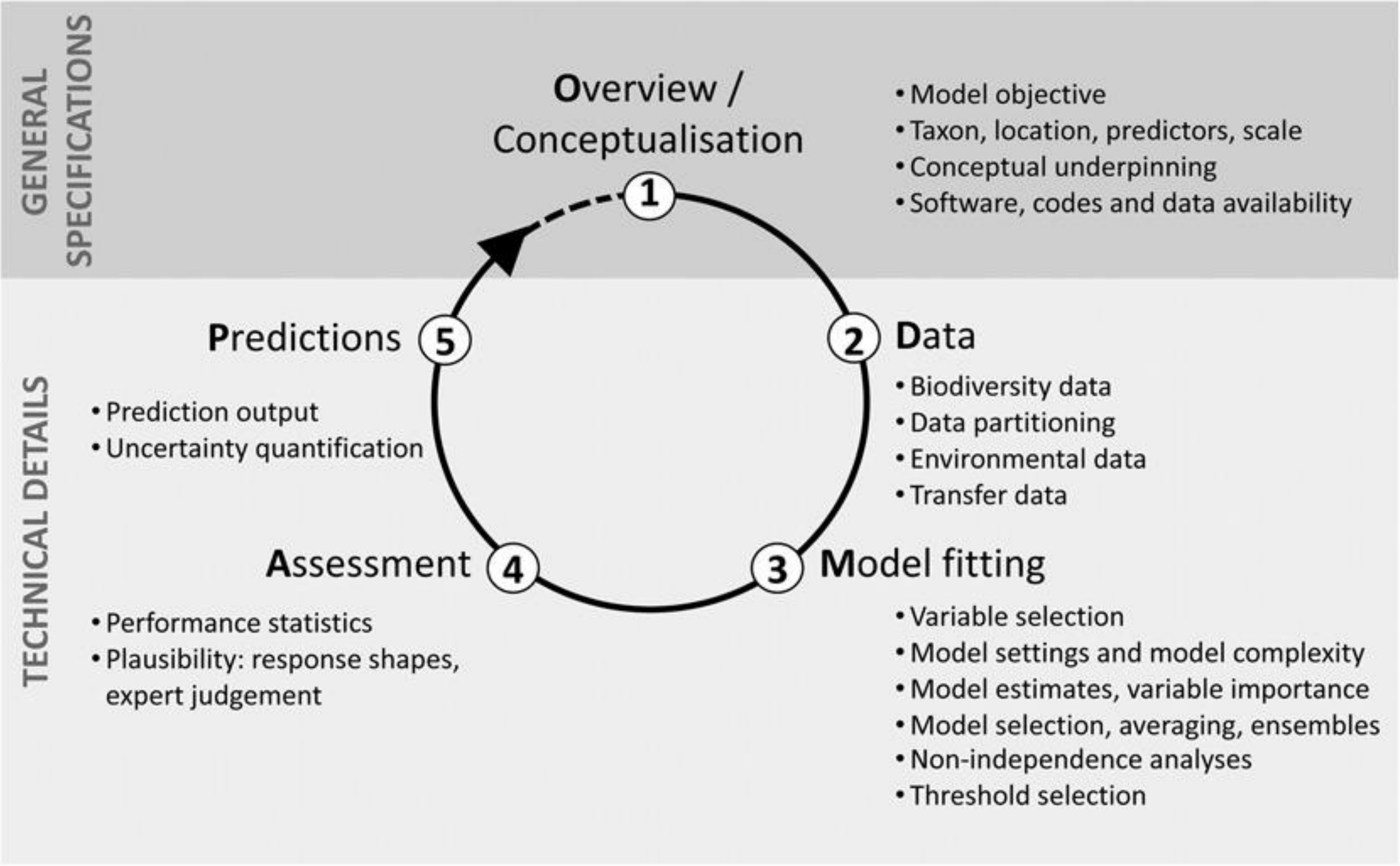
Issue: Autocorrelation
Solution: Thinning, Model structure

Issue: Pseudo-replication
Solution: Block-factor and random effects

Issue: Presence-only data
Solution: Generate pseudo-absences

Issue: Central place foragers
Solution: Model habitat accessibility

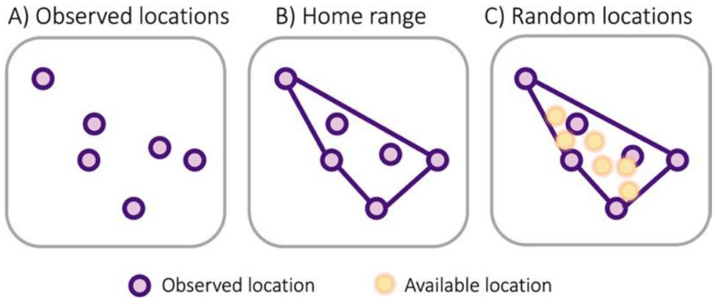
General workflow



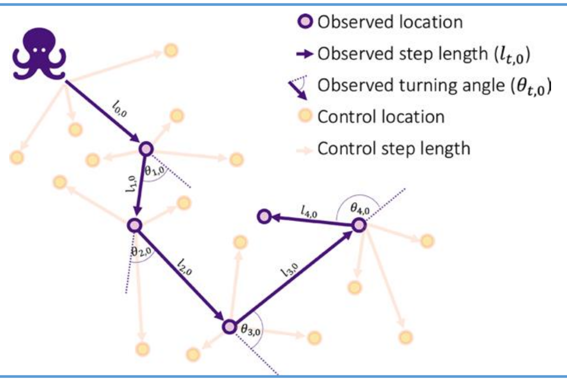
Zurell et al. 2020. Ecography. https://doi.org/10.1111/ecog.04960

Models

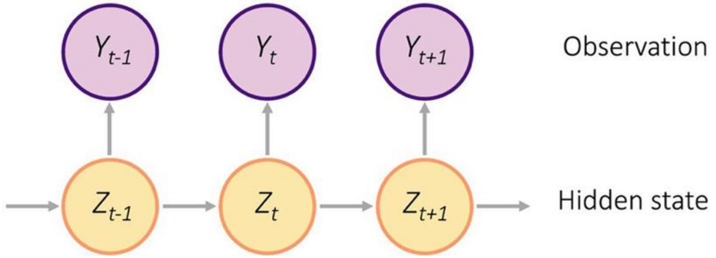
Resource Selection Function (RSF)

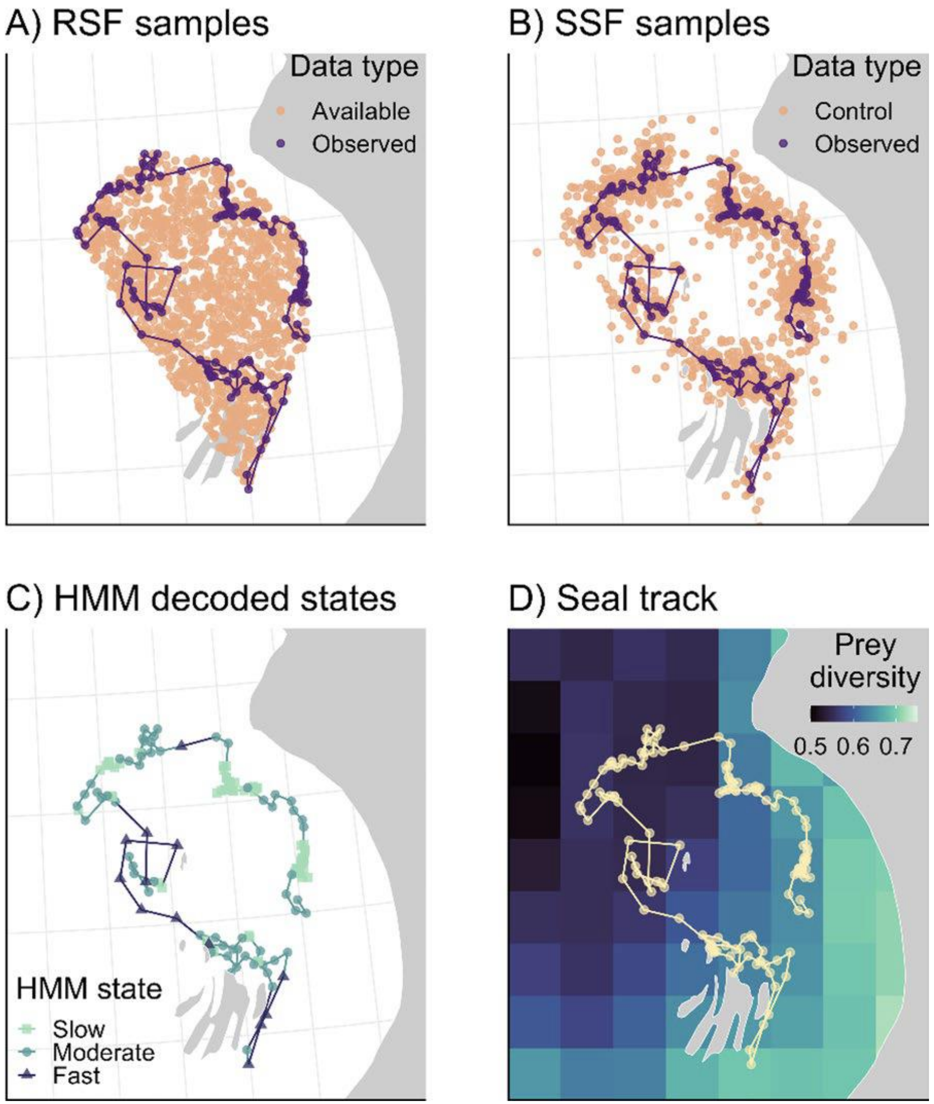


Step Selection Function (SSF)



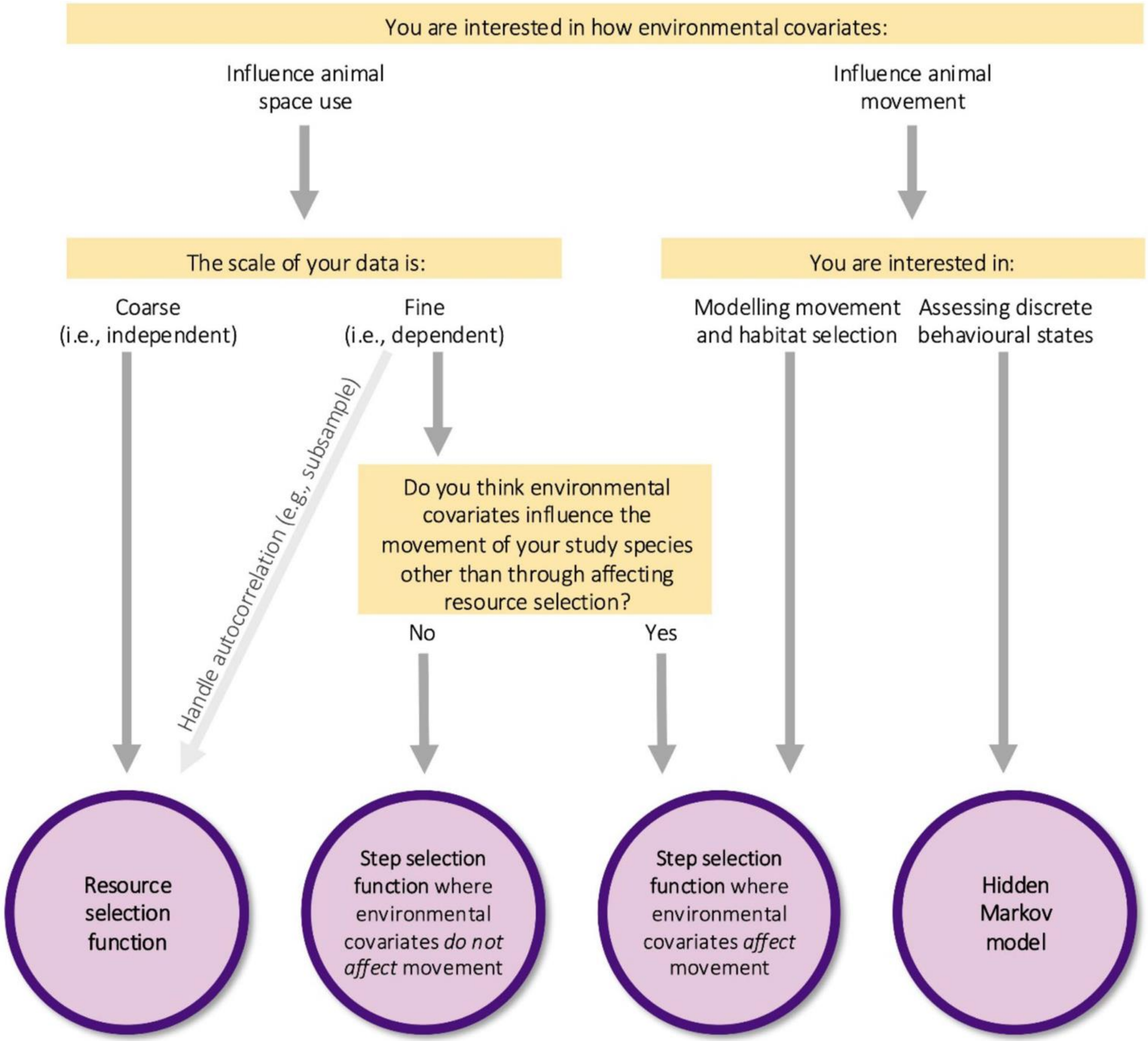
Hidden Markov Model (HMM)





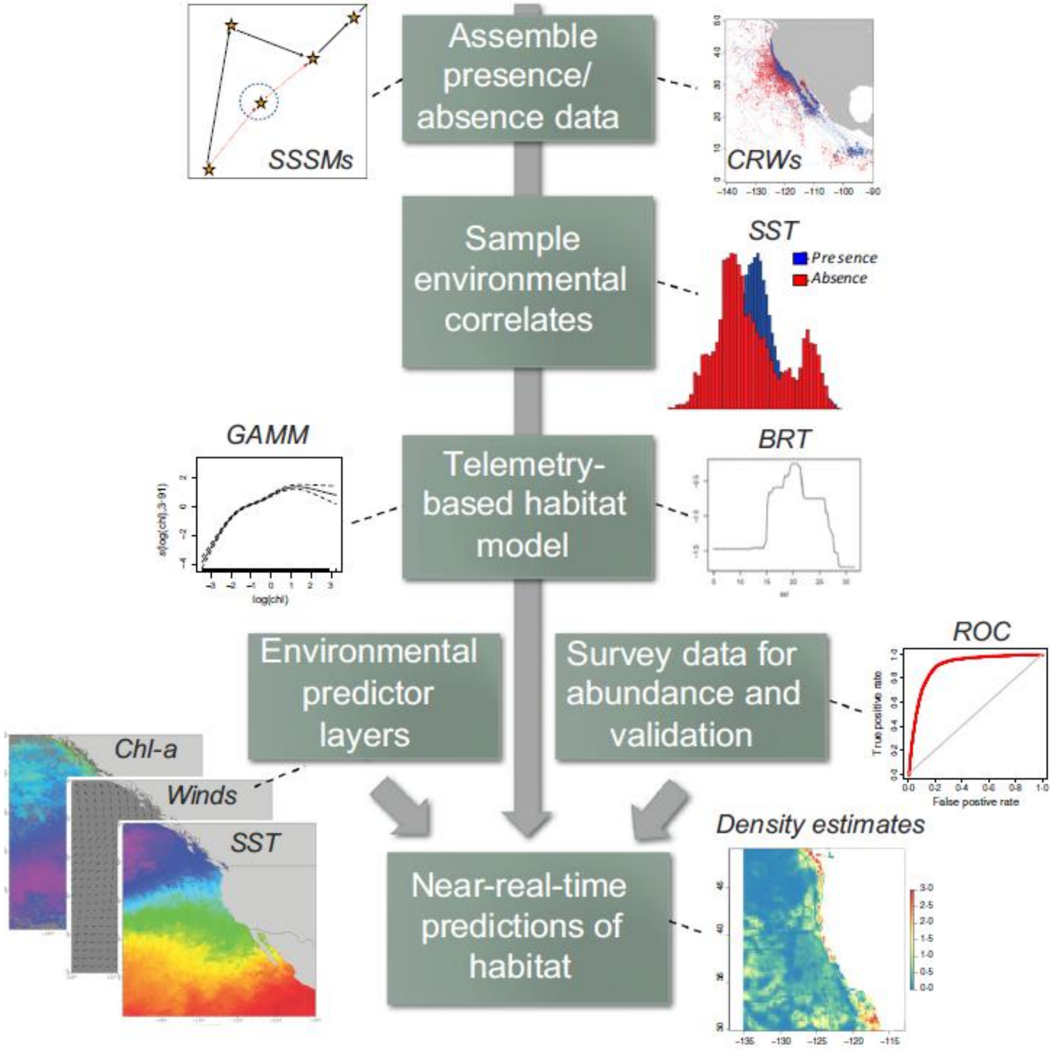
Florko et al. 2025. Mov Ecol https://doi.org/10.1186/s40462-025-00549-2

Choosing a model



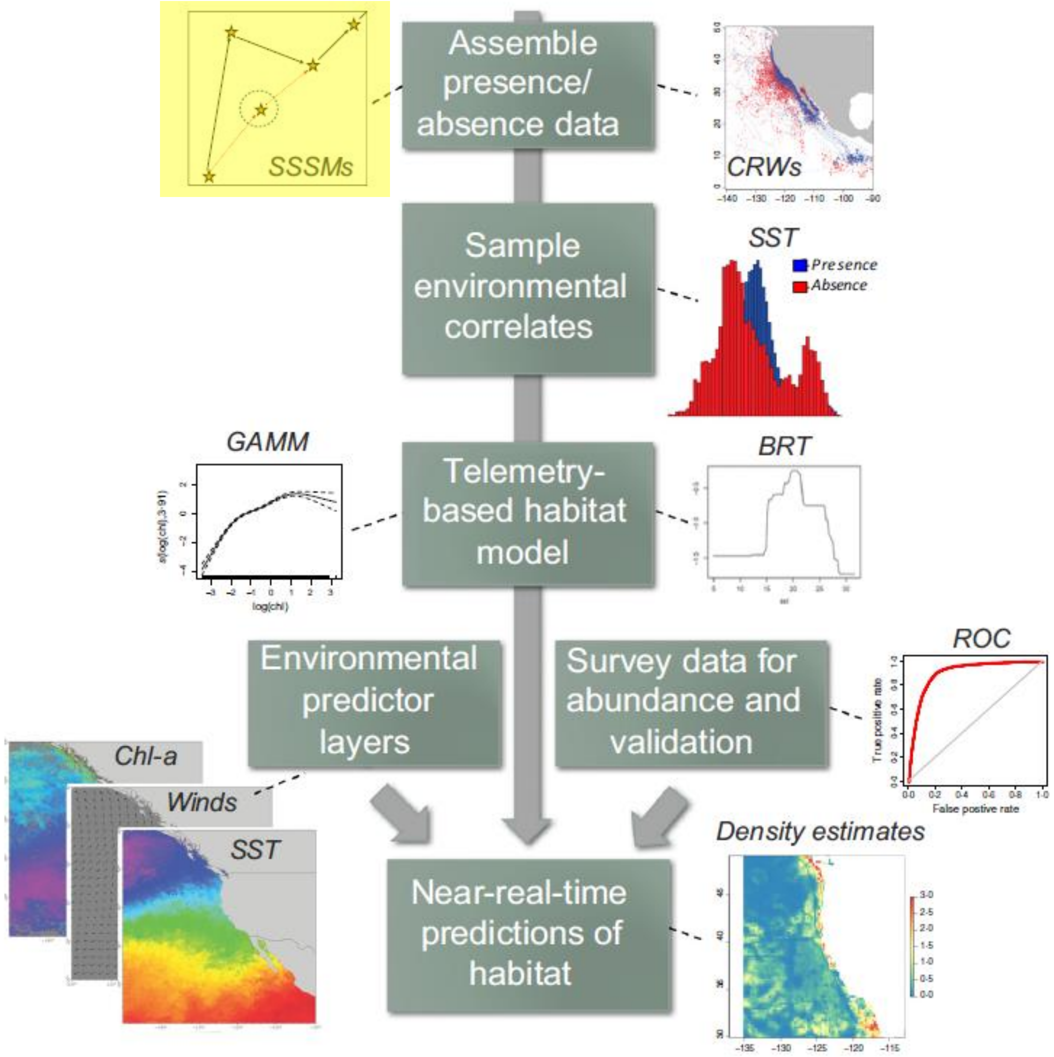
Florko et al. 2025. Mov Ecol https://doi.org/10.1186/s40462-025-00549-2

Workflow for marine telemetry data



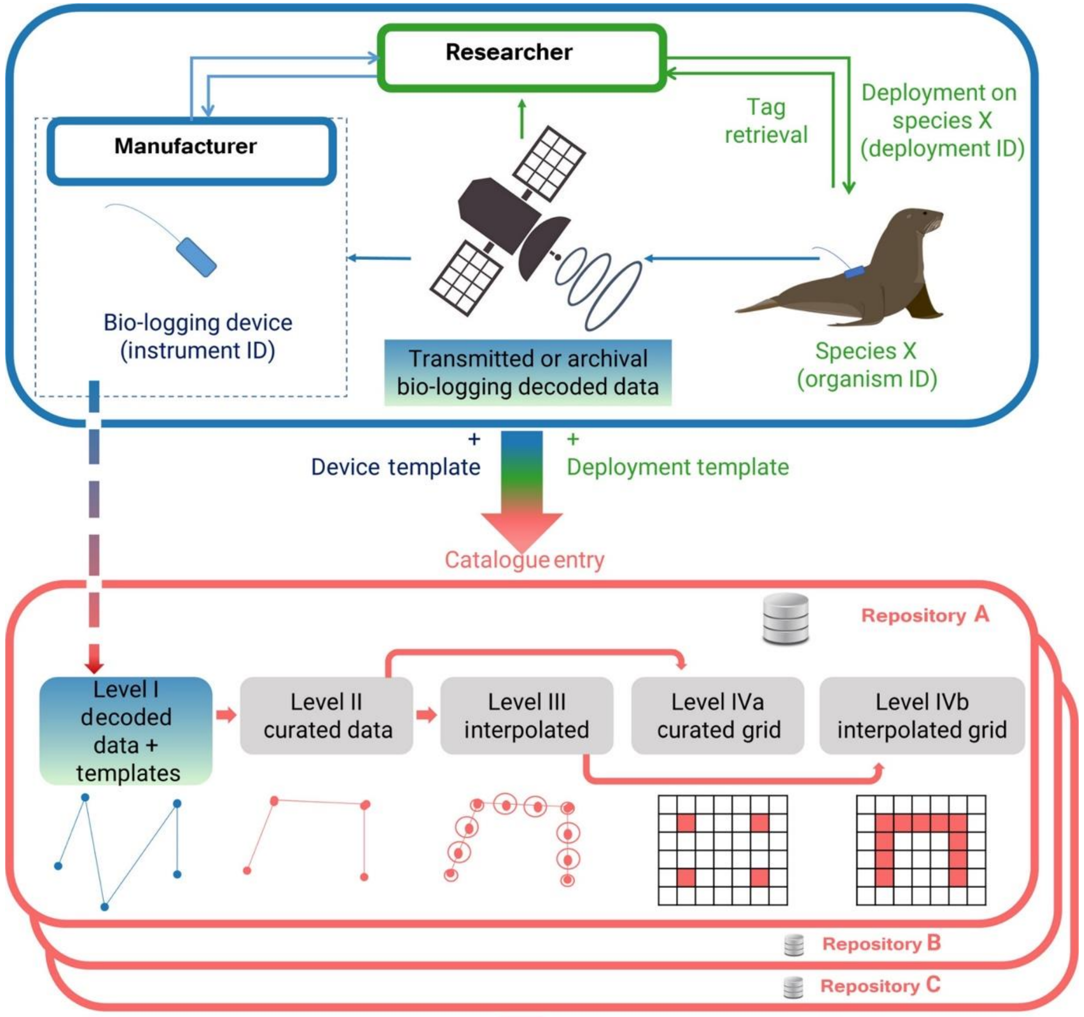
Hazen et al. 2016. JAE https://doi.org/10.1111/1365-2664.12820

Workflow for marine telemetry data



Hazen et al. 2016. JAE https://doi.org/10.1111/1365-2664.12820

Location data processing



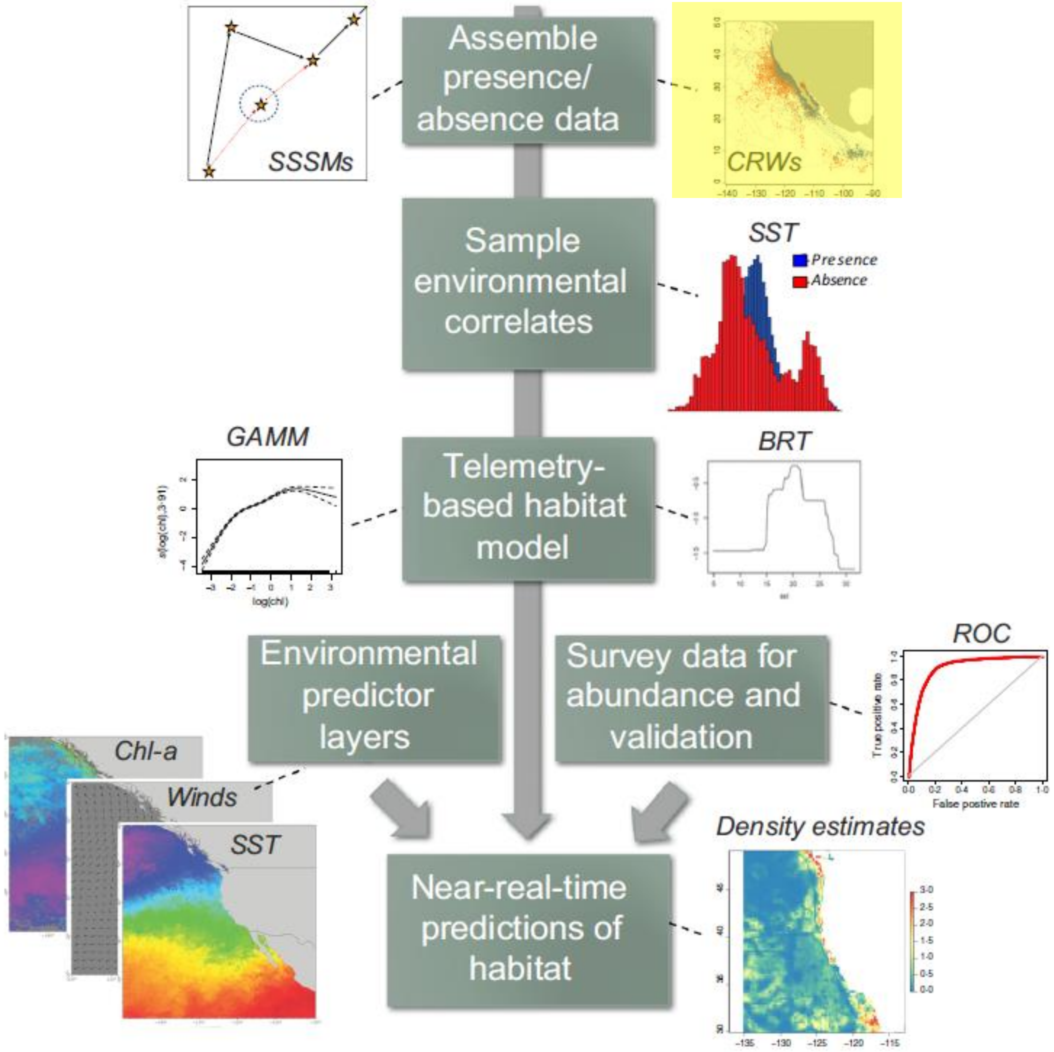
Data curation
- Near-duplicates
- Unrealistic speeds

Regularization
- Interpolation
- State-space models

Gidding
- Rasterization of observations

Sequeira et al. 2021. MEE. https://doi.org/10.1111/2041-210X.13593

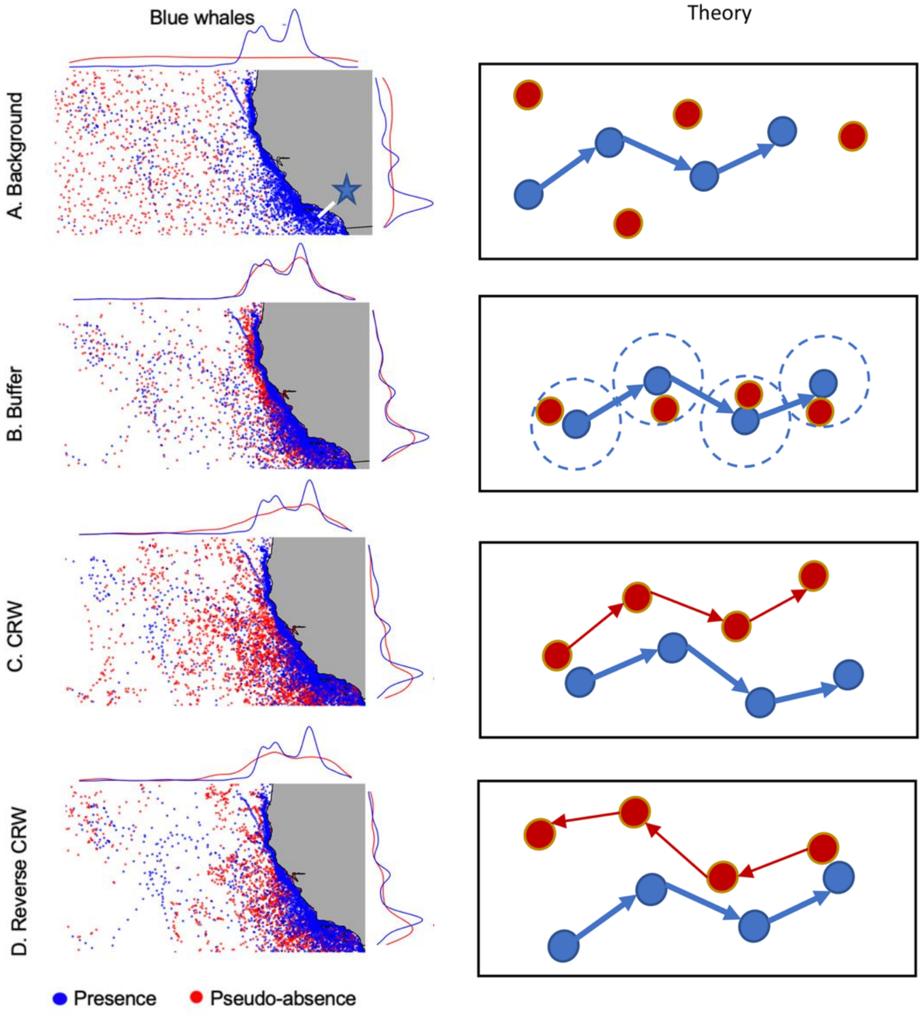
Workflow for marine telemetry data



Hazen et al. 2016. JAE https://doi.org/10.1111/1365-2664.12820

Pseudo-absences

How to generate them?



Background sampling

Sampling within a *buffer* zone around presence locations

Correlated random walks beginning at the tag release location

Reverse correlated random walks beginning at the last tag location

Hazen et al. 2021. Movement Ecology. https://doi.org/10.1186/s40462-021-00240-2

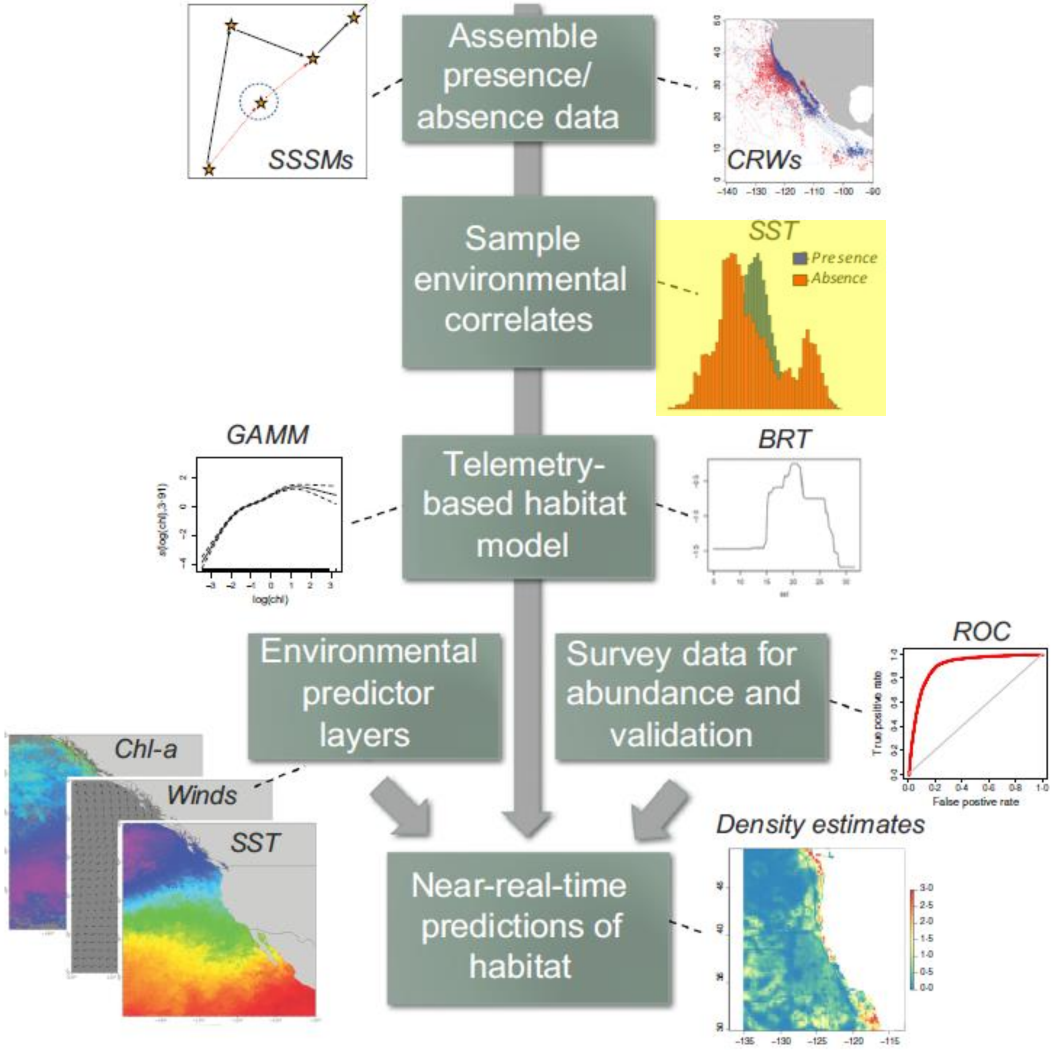
Pseudo-absences

How many?

	Number of pseudo-absences
GLM, GAM	10 000 PA or a minimum of 10 runs with 1000 PA with an equal weight for presences and absences
MARS	**A minimum of 10 runs with 100 PA**
MDA	**A minimum of 10 runs with 100 PA with an equal weight for presences and absences**
CTA, BRT, RF	**Same as number of presences, 10 runs when less than 1000 PA with an equal weight for presences and absences**

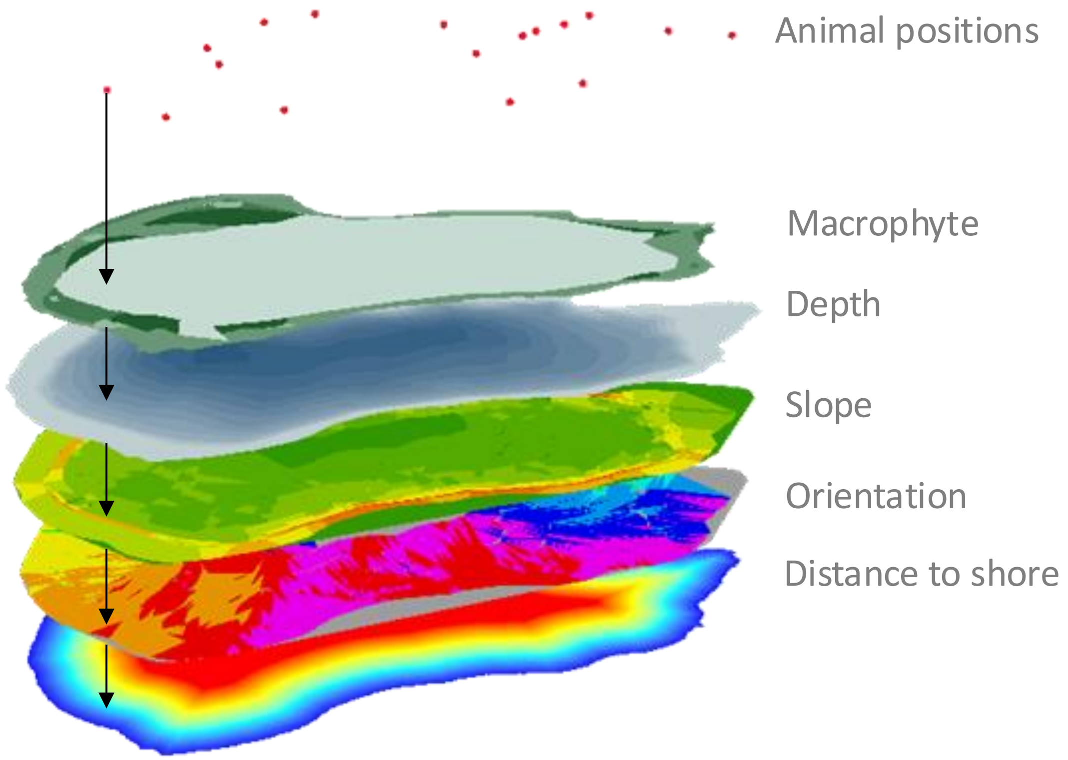
Barbet-Massin et al. 2012. MEE. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Workflow for marine telemetry data



Hazen et al. 2016. JAE https://doi.org/10.1111/1365-2664.12820

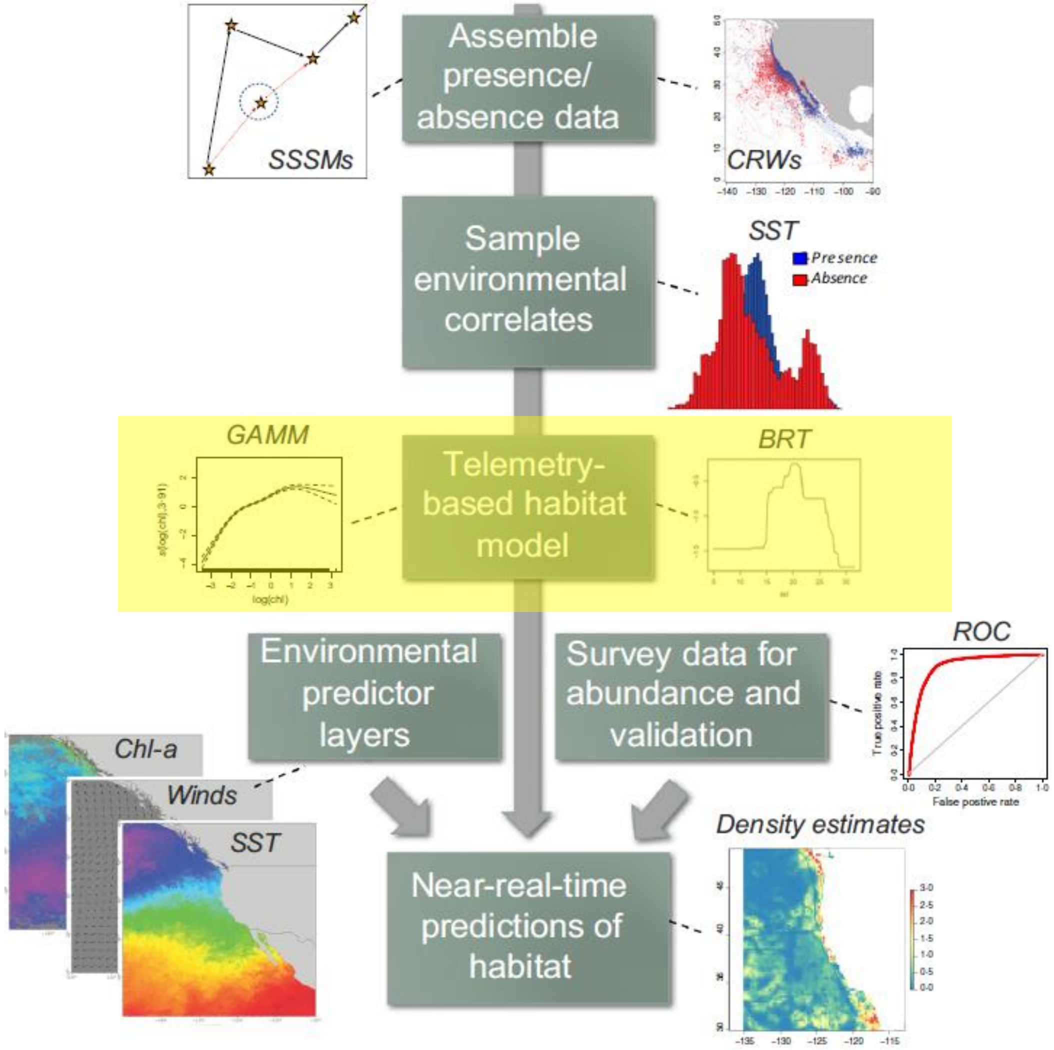
Sample environmental correlates



Sample environmental correlates

- Bathymetry: GEBCO (Global), EMODnet (EU Seas)
- Benthic habitats: EUSeamap
- Oceanographic data: Copernicus Marine Service
- Other datasets
  - Eg. Seamounts: Yesson et al. (2020) https://doi.pangaea.de/10.1594/PANGAEA.921688

Workflow for marine telemetry data



Hazen et al. 2016. JAE https://doi.org/10.1111/1365-2664.12820

Species distribution modelling in marine environments

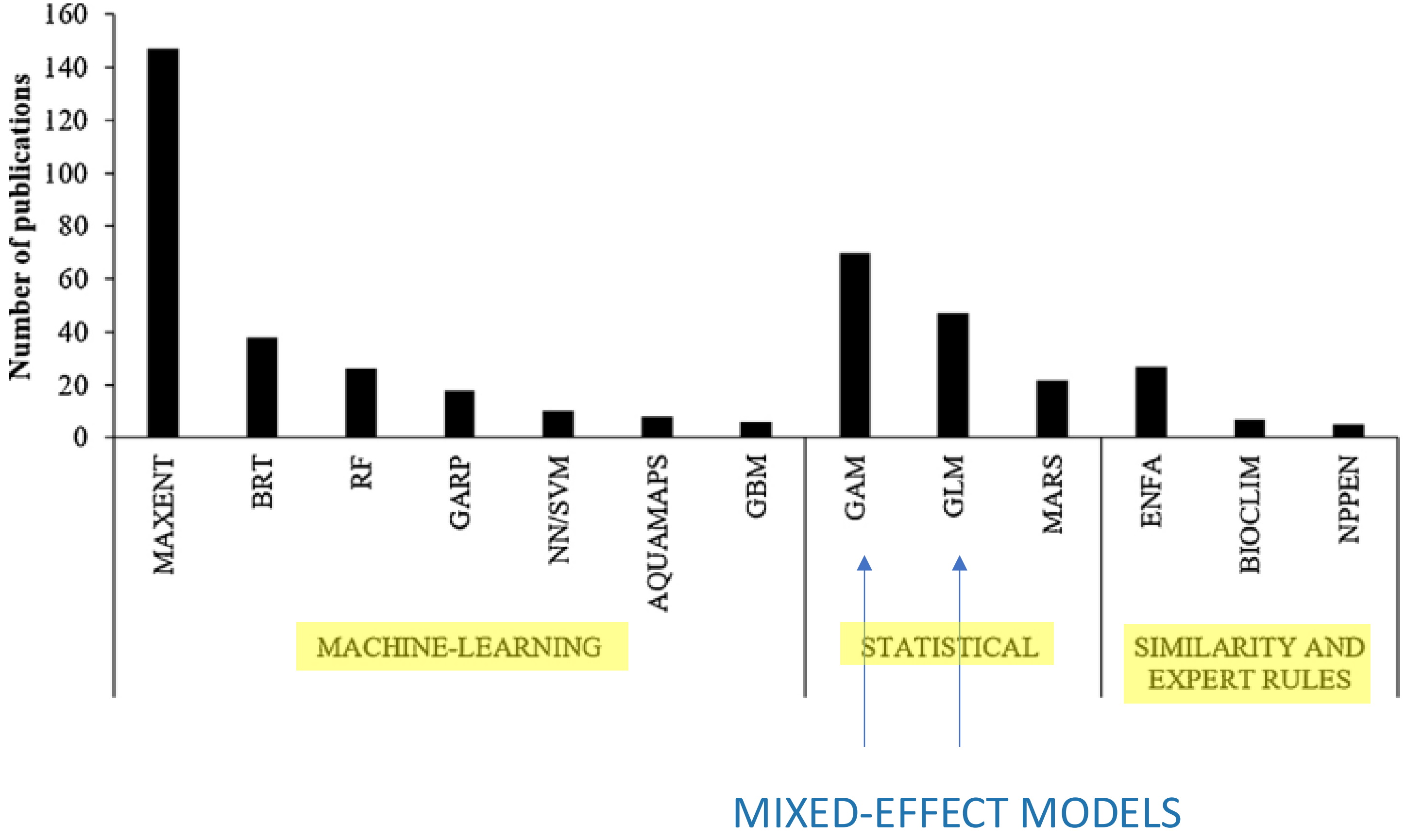


Fig. 8. Tendency in the use of ENM and SDM algorithms. We present only the most frequently used algoritms. MAXENT = Maximum Entropy; BRT = Boosted Regression Trees; RF = Random Forest; GARP = Genetic Algorithm for Rule Set Production; NN/SVM = Neural Networks/Support Vector Machine; GBM = Generalized Boosting Models; GAM = Generalised Additive Model; GLM = Generalized Linear Model; MARS = Multivariate Adaptive Regresion Splines; ENFA = Ecological Niche Factor Analysis; NPPEN = Non Parametric Probabilistic Ecological Niche.

Melo-Merino. 2020. Ecological Modelling. https://doi.org/10.1016/j.ecolmodel.2019.108837

Boosted Regression Trees

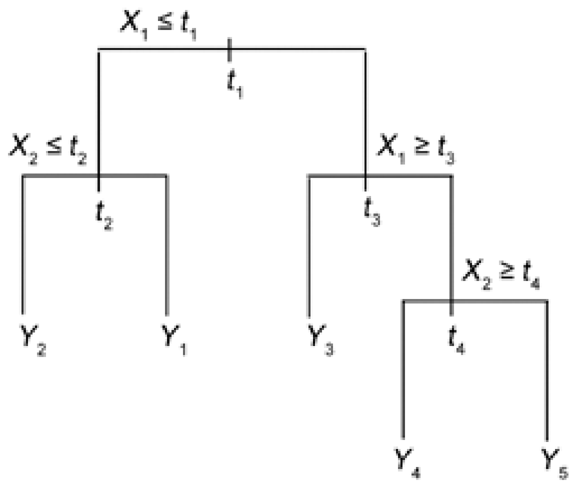
Journal of Animal Ecology
BRITISH ECOLOGICAL SOCIETY

Free Access

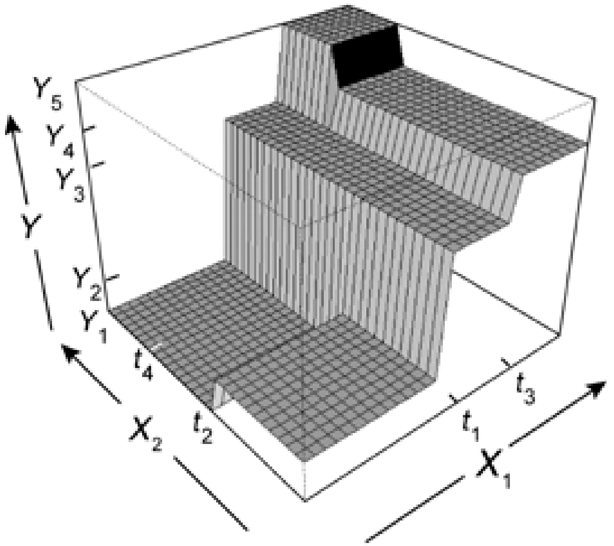
A working guide to boosted regression trees

J. Elith, J. R. Leathwick, T. Hastie

First published: 08 April 2008 | https://doi.org/10.1111/j.1365-2656.2008.01390.x |
Citations: 3,279



Decision tree algorithms + Boosting methods



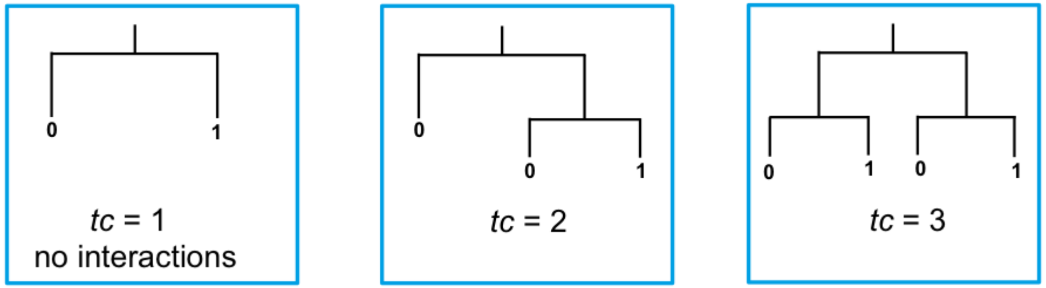
- Can be used with a variety of response types (binomial, gaussian, poisson)
- Stochastic, which improves predictive performance
- The best fit is automatically detected by the algorithm
- Model represents the effect of each predictor after accounting for the effects of other predictors
- Robust to missing values and outliers

Going bayesian: Bayesian additive regression trees (BART)

Boosted Regression Trees

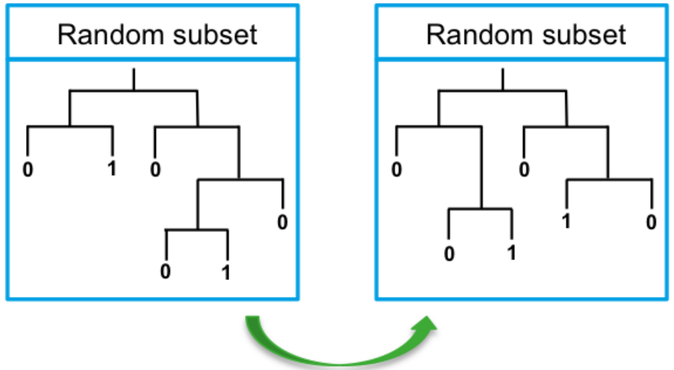
Tree complexity (*tc*)

Number of splits in each tree



Learning rate (*lr*)

Contribution of each tree to the growing model



lr: contribution to growing model
small value = many trees

Bag fraction

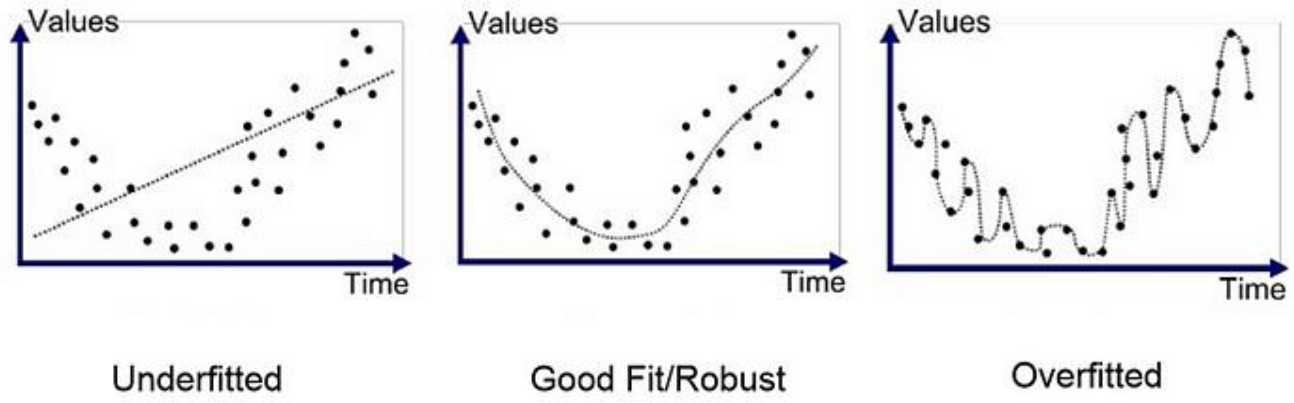
Proportion of data randomly selected at each iteration

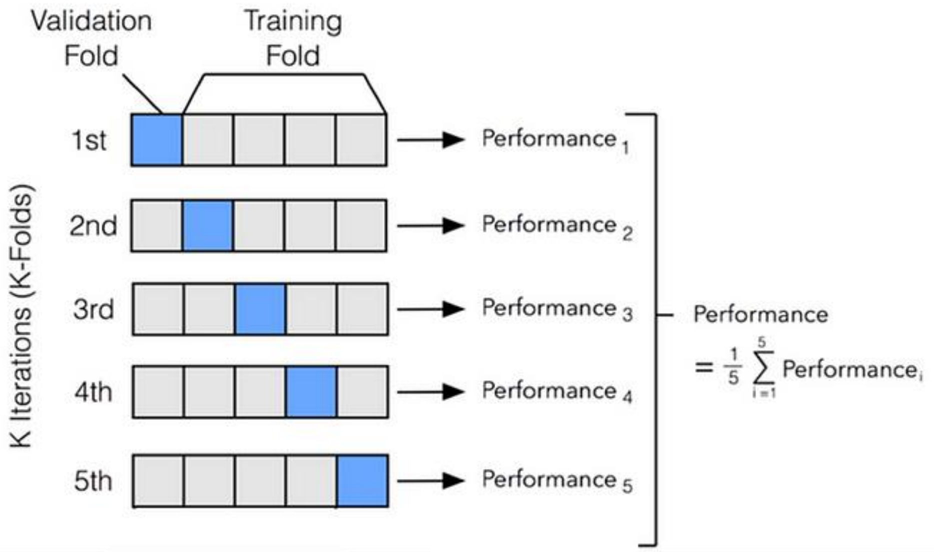
Number of trees

Boosting iterations (at least 1,000)

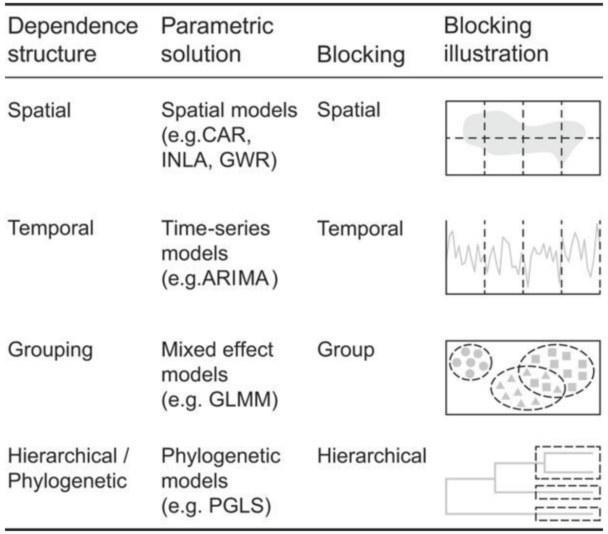
Boosted Regression Trees

Cross-validation





Boosted Regression Trees

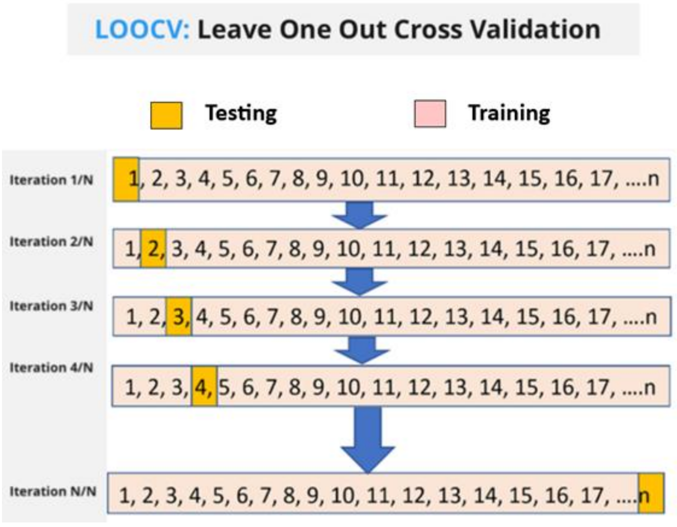


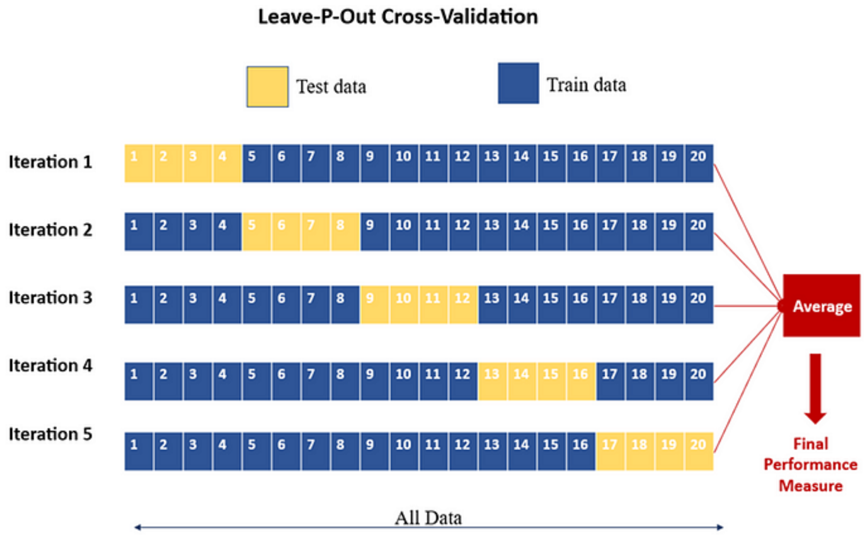
Dependence structure	Parametric solution	Blocking	Blocking illustration
Spatial	Spatial models (e.g.CAR, INLA, GWR)	Spatial	
Temporal	Time-series models (e.g.ARIMA)	Temporal	
Grouping	Mixed effect models (e.g. GLMM)	Group	
Hierarchical / Phylogenetic	Phylogenetic models (e.g. PGLS)	Hierarchical	

Cross-validation

Block-factor in RSF models

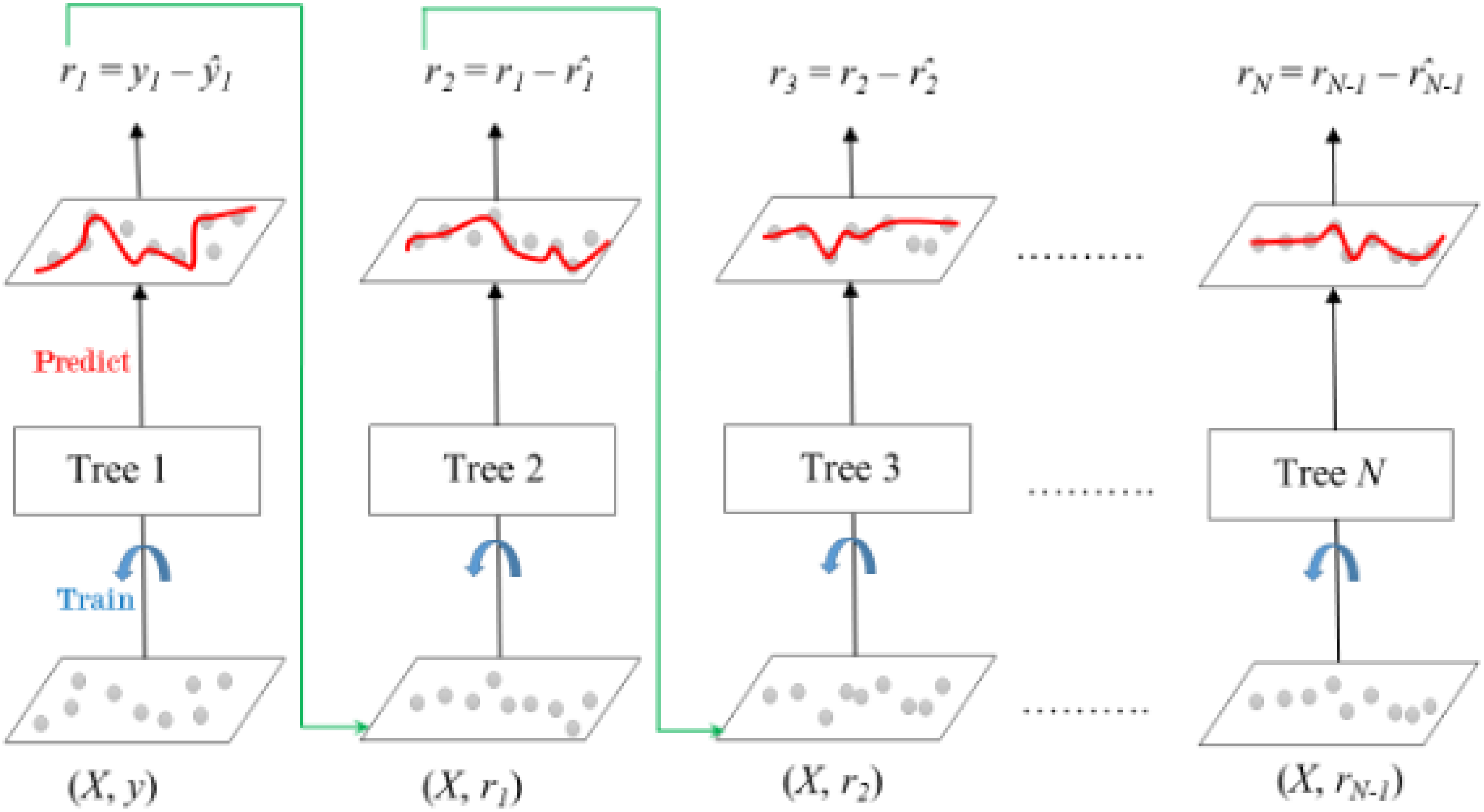
Individual effects: use block-factor (here, ID number refers to animal ID)



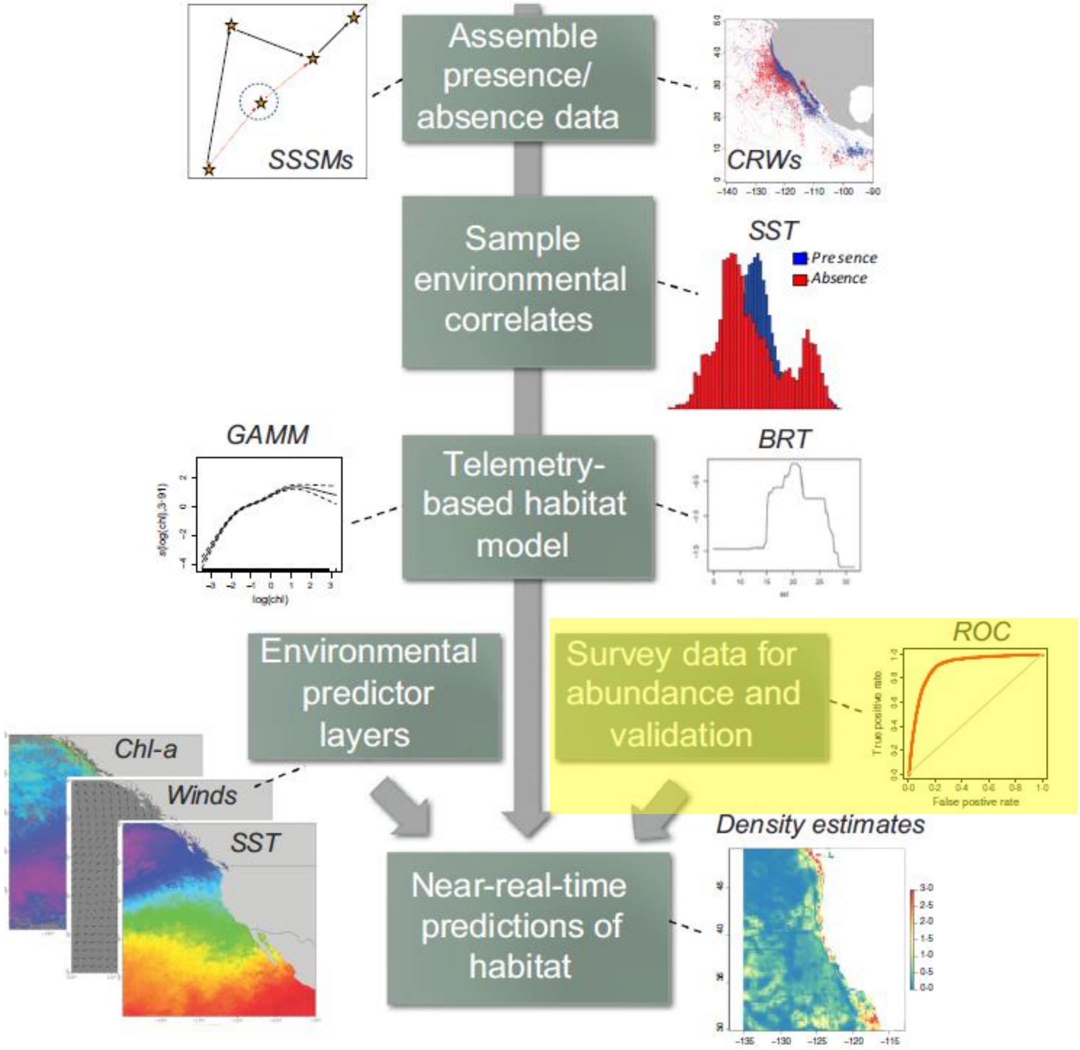


Roberts et al. 2016. Ecography. https://doi.org/10.1111/ecog.02881

Boosted Regression Trees

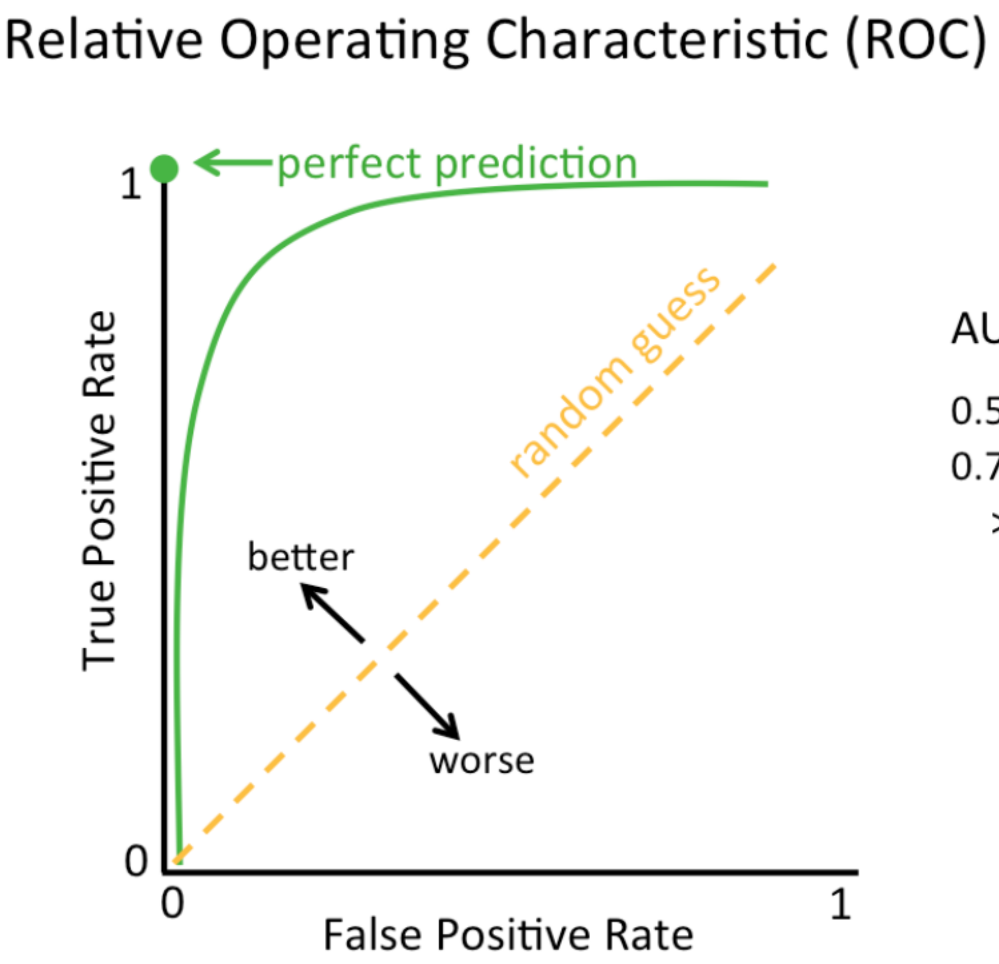


Workflow for marine telemetry data



Hazen et al. 2016. JAE https://doi.org/10.1111/1365-2664.12820

Model evaluation: Area under the curve (AUC)



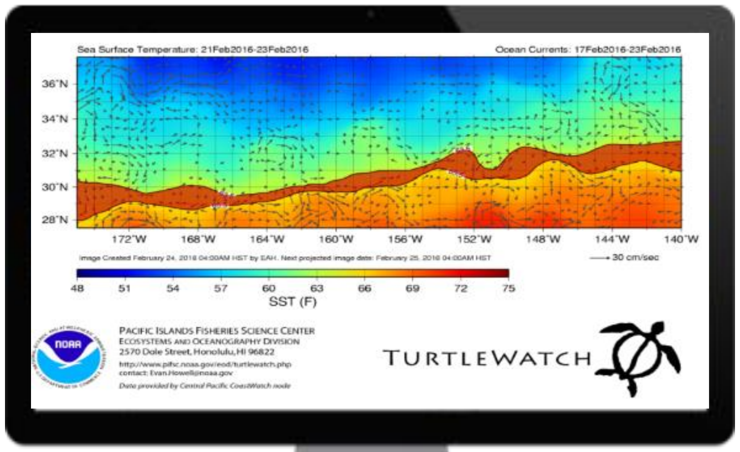
AUC = area under the curve

0.5-0.7 = poor model performance

0.7-0.9 = moderate

> 0.9 = excellent

Why habitat models are used for?



- Assess effect of environmental drivers to animal movement and distribution

- Identify areas that should be prioritised for conservation

- Evaluating the potential of an invasive species to settle in particular areas

- Combined with future projections of changes of the natural environment, to predict how biodiversity will be affected by impacts such as climate change.

Applications

nature

Explore content ⌄ **About the journal** ⌄ **Publish with us** ⌄

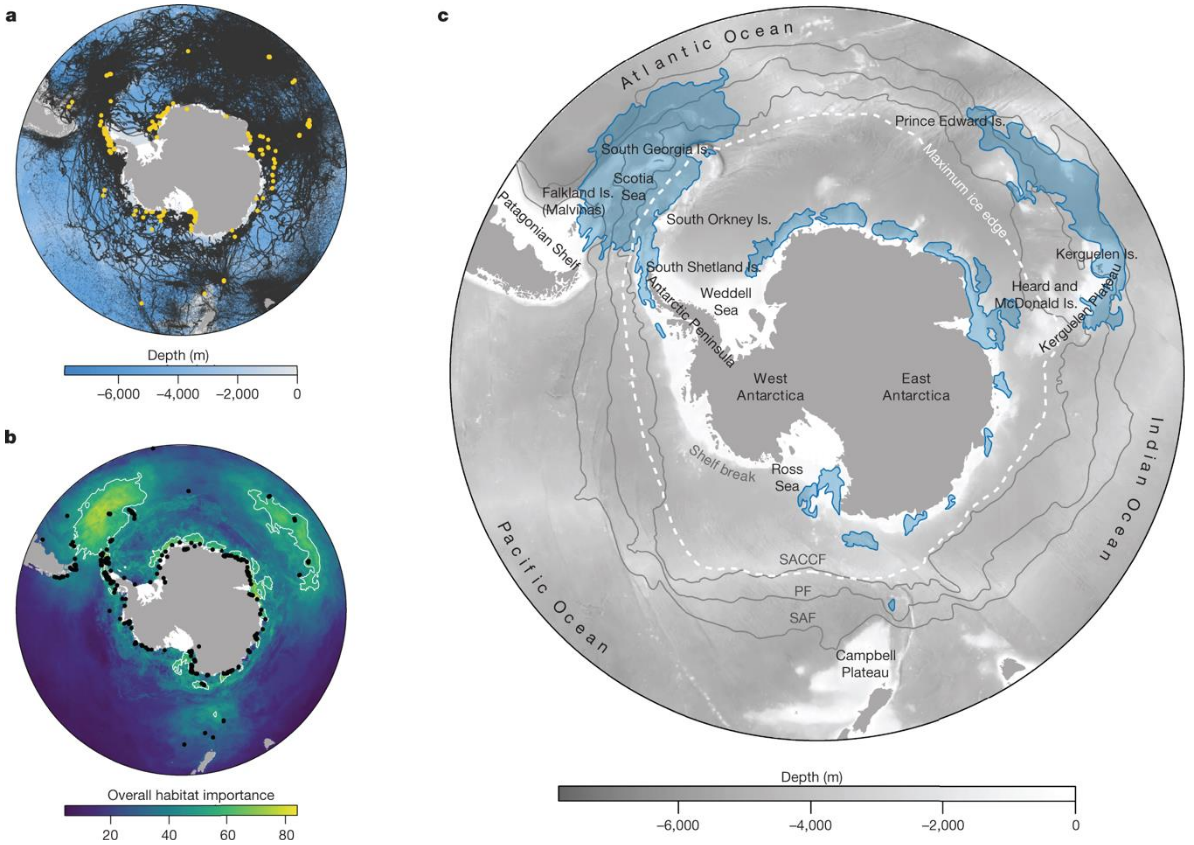
nature › articles › article

Article | Published: 18 March 2020

Tracking of marine predators to protect Southern Ocean ecosystems

Mark A. Hindell ✉, Ryan R. Reisinger, Yan Ropert-Coudert, Luis A. Hückstädt, Philip N. Trathan, Horst Bornemann, Jean-Benoît Charrassin, Steven L. Chown, Daniel P. Costa, Bruno Danis, Mary-Anne Lea, David Thompson, Leigh G. Torres, Anton P. Van de Putte, Rachael Alderman, Virginia Andrews-Goff, Ben Arthur, Grant Ballard, John Bengtson, Marthán N. Bester, Arnoldus Schytte Blix, Lars Boehme, Charles-André Bost, Peter Boveng, ... Ben Raymond + Show authors

Nature **580**, 87–92 (2020) | Cite this article



Applications

RESEARCH ARTICLE

MARINE CONSERVATION

Ship collision risk threatens whales across the world's oceans

Anna C. Nisi[1]*, Heather Welch[2,3], Stephanie Brodie[4], Callie Leiphardt[5], Rachel Rhodes[5], Elliott L. Hazen[3], Jessica V. Redfern[6], Trevor A. Branch[7], Andre S. Barreto[8], John Calambokidis[9], Tyler Clavelle[10], Lauren Dares[11], Asha de Vos[12], Shane Gero[13], Jennifer A. Jackson[14], Robert D. Kenney[15], David Kroodsma[10], Russell Leaper[16], Douglas J. McCauley[5], Sue E. Moore[1], Ekaterina Ovsyanikova[17], Simone Panigada[18], Chloe V. Robinson[11], Tim White[10], Jono Wilson[19], Briana Abrahms[1]

After the near-complete cessation of commercial whaling, ship collisions have emerged as a primary threat to large whales, but knowledge of collision risk is lacking across most of the world's oceans. We compiled a dataset of 435,000 whale locations to generate global distribution models for four globally ranging species. We then combined >35 billion positions from 176,000 ships to produce a global estimate of whale-ship collision risk. Shipping occurs across 92% of whale ranges, and <7% of risk hotspots contain management strategies to reduce collisions. Full coverage of hotspots could be achieved by expanding management over only 2.6% of the ocean's surface. These inferences support the continued recovery of large whales against the backdrop of a rapidly growing shipping industry.

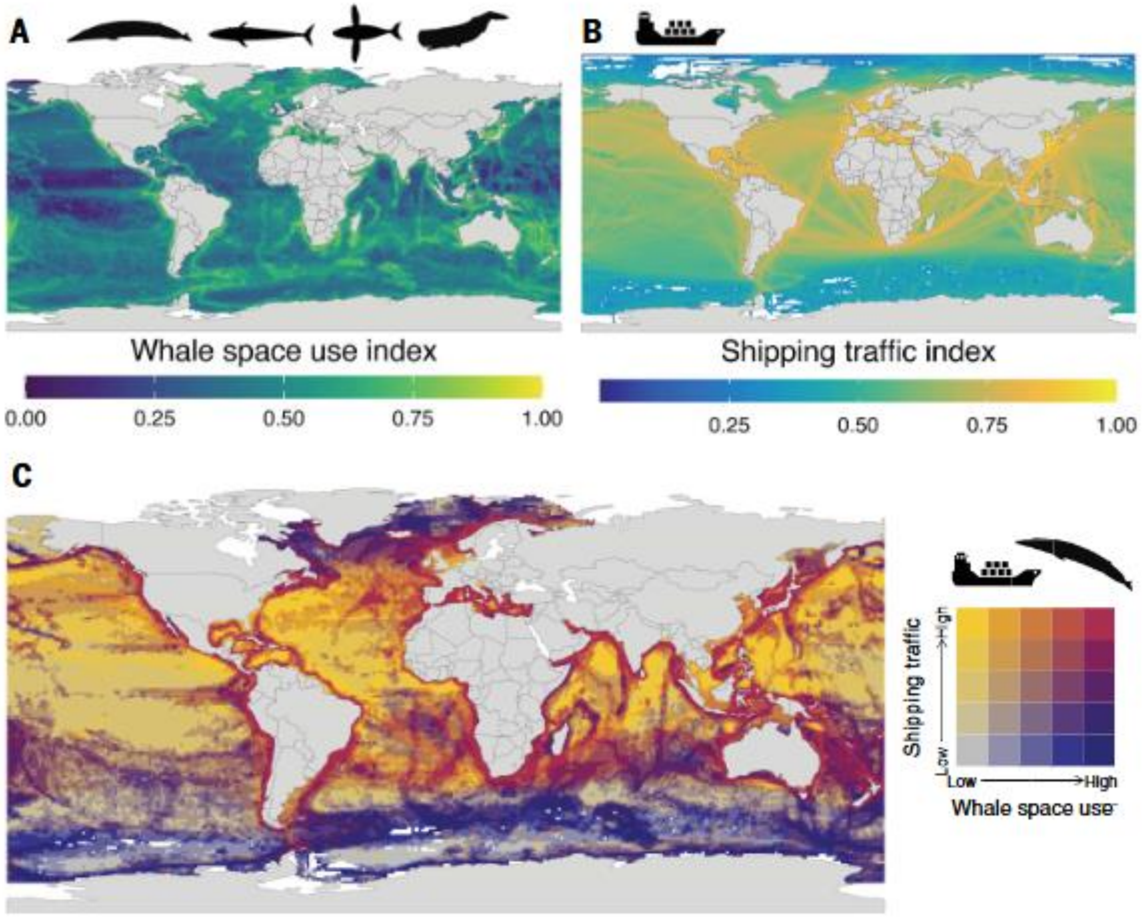


Fig. 1. Spatial overlap between whales and shipping traffic. (**A**) Average annual whale space use across blue, fin, humpback, and sperm whales. (**B**) Global marine shipping traffic for large (>300 gross tons) vessels, from AIS data from 2017 to 2022. The shipping traffic index weights shipping density by vessel speed on a log-scale, standardized between 0 and 1. (**C**) Bivariate map showing the intensity of both whale space use and shipping traffic in each 1° by 1° grid cell.

Whale observations

- Sources: GBIF, OBIS, MoveBank, …
- Temporal range: 1960-2020
- Data types:
  - Survey data (presence-absence)
  - Opportunistic sightings (presence-only)
  - Tagging (presence-only)
  - Whaling records (presence-only)
- Fit integrated species distribution models using INLA and inlabru packages

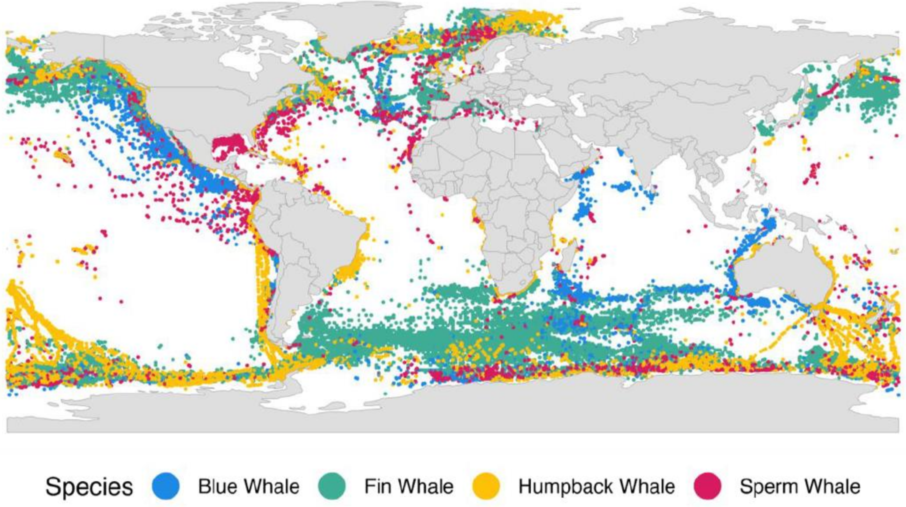
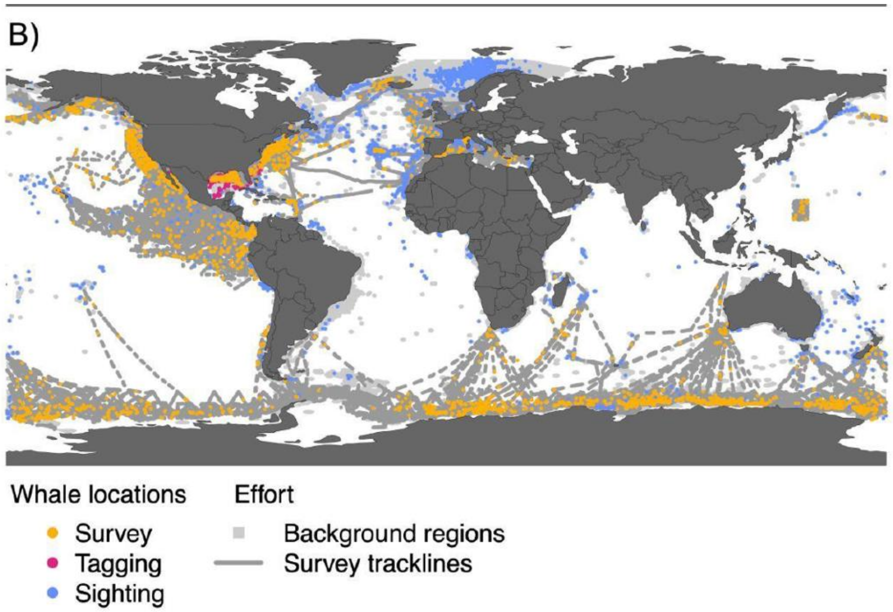


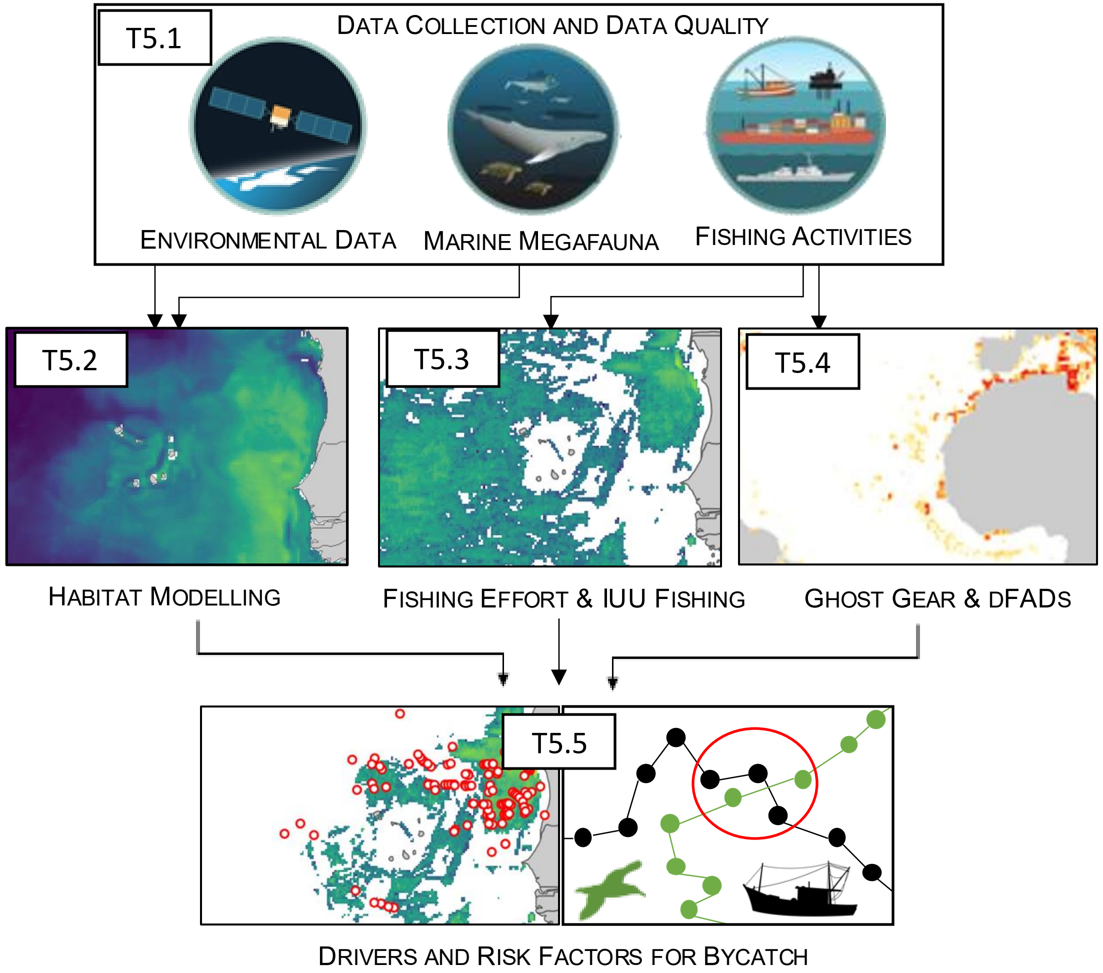
Figure S1. Whale location data. Location data for blue, fin, humpback, and sperm whales from 1960-2020.



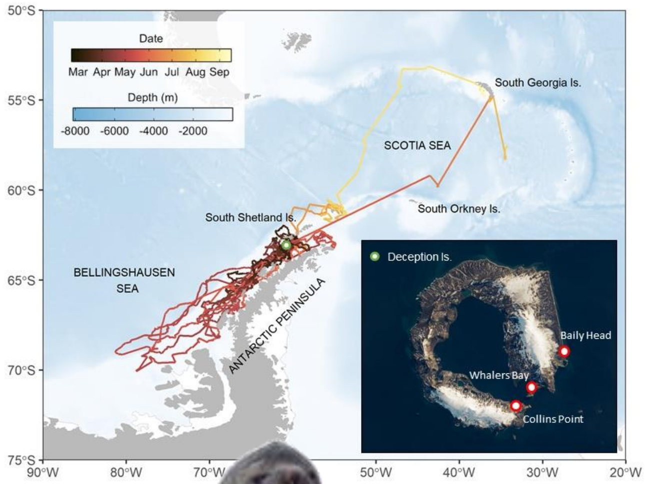
Absence and background data

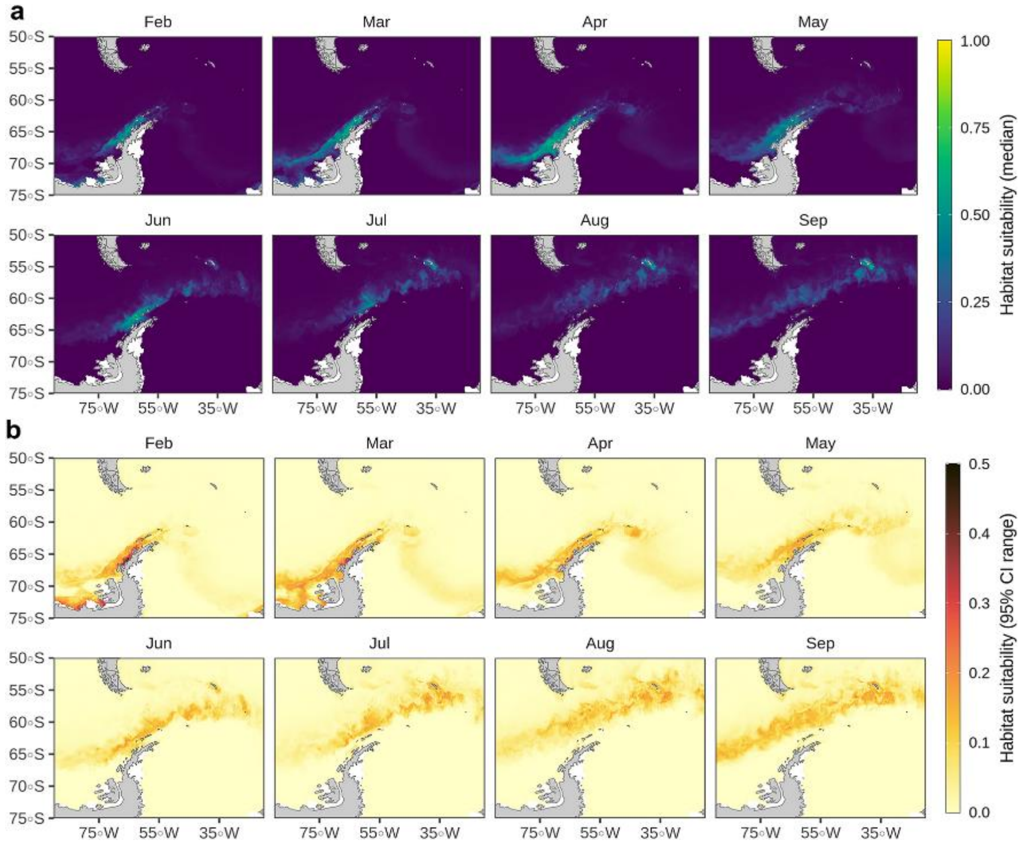
- Presence-absence data: random sub-sampling
- Presence-only: target-group approach for background locations
  - Opportunistic sightings:
    - 100-km buffer around presence of other spp
    - Background locations within buffered región
  - Tagging data
    - Subsample track (1 loc/day)
    - MCP
    - Random sampling
- Ratio 1:1
- Regional approach to account for sub-populations

Application within REDUCE



SDM on mobile and pelagic species





- Argos satellite tracking
- SSM to regularize data
- Pseudo-absences
- Env data post-processing
- BRT
- Block factor CV (LOOCV)
- Model evaluation AUC
- Uncertainty (Bootstrap)
- Accesibility model
- Github repo

Antarctic fur seals
(*Arctocephalus gazella*)
Boosted Regression Trees
March et al. 2021 Sci Rep. https://doi.org/10.1038/s41598-021-01700-w

Hands-on with R

1. Uncompress “AnimalMovementSDM-RWithData.zip”.
2. Set your working directory in the unzipped folder
3. Open HTML for part 1 and part 2
4. Install all required packages