



# A global systematic review of species distribution modelling approaches for cetaceans and sea turtles

E. Pasanisi<sup>a,b,\*</sup>, D.S. Pace<sup>a</sup>, A. Orasi<sup>b</sup>, M. Vitale<sup>a</sup>, A. Arcangeli<sup>b,\*</sup>

<sup>a</sup> Department of Environmental Biology, Sapienza University of Rome, 00185 Rome, Italy

<sup>b</sup> Department for Biodiversity Conservation and Monitoring, Italian Institute for Environmental Protection and Research (ISPRA), 00143 Rome, Italy

## ARTICLE INFO

### Keywords:

Cetacean  
Habitat modelling  
Predictive models  
PPM  
Sea turtles  
Species distribution

## ABSTRACT

In the last decades, interest in species distribution models (SDMs) has grown greatly. The descriptive and predictive power of correlative SDMs is highly valued to meet the high demand for filling gaps in the spatial ecology of wide-ranging and elusive species, such as cetaceans and sea turtles, living in habitats that are technically challenging to survey and where the availability of high quality, unbiased data at appropriate spatial and temporal resolution is not straightforward.

This study endeavours to offer a comprehensive global overview of recent advancements in modelling techniques within the realm of SDMs applied to cetaceans and sea turtles. Through a rigorous systematic review of 295 research papers, we identified gaps in species and geographic coverage and highlighted the underrepresentation of biotic, anthropogenic, and water column variables. Our examination revealed a diverse array of modelling approaches, showcasing a notable preference for standard regression-based or machine-learning models, such as GAMs or Maxent, with Bayesian-based models emerging and experiencing growing development.

Critical limitations and decisions in constructing and evaluating SDMs were discussed, proposing best practices for future studies. Emphasis was placed on the importance of validating models using fully independent datasets, particularly in the context of conservationist studies. This work not only sheds light on the state of the field but also serves as a valuable tool for those interested in modelling the distribution of these magnificent and enigmatic animals, as well as other cryptic species, offering insights that can guide researchers in making informed decisions in the realm of SDMs.

## 1. Background

As highly mobile, long-lived vertebrates residing at the highest levels of marine trophic webs, cetaceans and sea turtles are pivotal for marine biodiversity. They serve as indicators of ecosystem health and productivity (Katona and Whitehead, 1988), often referred to as ‘sentinels’ for ecosystem change (Aguirre and Tabor, 2004; Moore, 2008), or ‘umbrella’ species benefiting other species and the ecosystem (Mann, 2000). Many species are opportunistic feeders, interacting with the marine ecosystem in complex ways, and undertake extensive migrations between feeding and breeding grounds. Particularly for cetaceans, research suggests an active role in nutrient cycling, transporting energy, nutrients and materials horizontally and vertically through the water (Estes et al., 2016), including faecal plumes and urine (Roman et al., 2014).

The cetacean group encompasses 89 known living species divided in

two orders, *odontocete* (toothed whales) with 74 species and *mysticete* (baleen whales) with 15 species (Fordyce and Perrin, 2024). These creatures inhabit oceans globally, with mysticetes showing greater diversity around 30° S and odontocetes around the tropical coasts (Pompa et al., 2011). Sea turtles consist of seven species within the superfamily *Chelonioidae* and family *Cheloniidae*, except for the leatherback turtle (*Dermochelys coriacea*), the only member of the family *Dermochelyidae*, and they are distributed across all oceans except for the polar regions.

Despite their ecological significance, these species face increasing threats from human pressures such as urbanisation, pollution and climate change. Examples of threats faces by these species include incidental catches, entanglement in marine debris, ship collisions, habitat loss, prey depletion, and noise disturbance. Also, climate-related changes potentially affect migration patterns and ecosystem structure, impacting future conservation (Grose et al., 2020).

Cetaceans and sea turtles, as iconic and charismatic flagship taxa,

\* Corresponding authors at: Department of Environmental Biology, Sapienza University of Rome, 00185 Rome, Italy.

E-mail address: [eugenia.pasanisi@uniroma1.it](mailto:eugenia.pasanisi@uniroma1.it) (E. Pasanisi).

<https://doi.org/10.1016/j.ecoinf.2024.102700>

Received 28 March 2024; Received in revised form 20 June 2024; Accepted 20 June 2024

Available online 21 June 2024

1574-9541/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

draw public, media, political, and scientific attention. Yet, due to their highly mobile nature, the extensive time spent underwater (Derville et al., 2018) and the difficult-to-survey environment in which they live (Redfern et al., 2006), obtaining high-quality, unbiased data at appropriate spatial and temporal resolutions remains a challenge. As a consequence, most questions about their spatial ecology remain unresolved.

In recent years, habitat-based or species distribution models (SDMs) have gained popularity in addressing this lack of knowledge and unravelling the spatial ecology of these species.

These models establish links between species data and environmental factors, employing diverse methodologies. They may rely on statistically or theoretically derived response curves, aiming to best reflect the totality of species' ecological requirements, thereby revealing ecological processes and approximating the ecological niche (Guisan et al., 2017; Guisan and Thuiller, 2005; Guisan and Zimmermann, 2000).

Within the scientific literature, various terms such as 'Species Distribution,' 'Spatial Distribution,' 'Habitat Suitability,' 'Ecological Niche Modelling' and 'Resource Selection Function' are interchangeably used to describe models sharing the same quantitative foundation, although scope can slightly differ. This interchangeable use adds complexity to the understanding of existing approaches, models, and tools. While the concept of ecological niche remains central, the term "species distribution" best captures the common focus of these models. For this reason, we will refer to these models collectively as SDMs (Species Distribution Models) in this paper.

SDMs are grounded on ecological niche theory, where the ecological niche refers to the conditions for a species to survive and reproduce on its own (Grinnell, 1917) or the functional role of a species in a community (Elton, 1927). SDMs also rely on Hutchinson's concepts from 1957, distinguishing between the "fundamental niche", representing the set of abiotic environmental conditions under which a species can persist indefinitely, and the "realised niche", influenced by interactions with other species. In simpler terms, the realised niche is the overlap of conditions suitable for a species, considering abiotic conditions, the biotic environment, and accessible conditions based on movement and dispersal.

Two primary SDM approaches exist: the mechanistic approach, which considers known tolerances to environmental conditions (e.g., the temperature range in which species can survive), and the correlative, or data-driven, approach, deriving species tolerances from the data itself. Because detailed physiological response data for cetaceans and sea turtles are lacking, correlative SDMs have been the primary choice for these species. Specifically, correlative SDMs utilize field observations of species occurrence to establish mathematical relationships between the observed locations and environmental variables (Jarnевич et al., 2015). This quantifies species response in environmental space, providing an approximation of the realised niche, or a portion of it, and then transferring back this information in geographical space. In more detail, SDMs can serve various purposes: i) to develop hypotheses about the drivers of species distribution ("descriptive/ explorative"); ii) to predict the occurrence of the species/its suitable environmental conditions in unsampled locations across the study area ("predictive"); iii) to make predictions in separate areas or different periods, thus "projecting" the model in space and time (i.e., new study area, future or past climate scenario) ("projective") (Guisan et al., 2017).

Given the great need to fill gaps in the spatial ecology of highly mobile species such as cetaceans and sea turtles and to protect them from emerging anthropogenic threats, correlative SDMs represent a crucial step in improving knowledge of species ecology, conservation management and planning strategies. This array of models and approaches may fulfil both ecological and conservation research needs, providing valuable insights into the distribution and habitat preference of these iconic marine species. However, due to the complexity and dynamics of marine ecosystems and the multiple spatial and temporal

scales at which species-environment relationships occur, the choice of modelling approach and the decisions made during model development represent crucial steps that require careful evaluation. Inappropriate choices could lead to misleading results, ecological interpretations and conclusions, diverting attention from appropriate conservation actions. In this regard, providing a current overview of modelling approaches, detailing the decision made in the process, can greatly assist researchers interested in applying these methods to the study of cetaceans and sea turtles as well as to the broader field of marina fauna modelling.

To comprehensively review SDMs for the distribution and habitat studies of cetaceans and sea turtles, we conducted a systematic literature search with a specific focus on correlative SDMs, and presented the main findings here. Previous efforts to summarise the state of the art in the use of SDMs include the work of Redfern et al. (2006), which referred only to cetaceans, and the works of Robinson et al. (2017) and Melo-Merino et al. (2020) referred to the marine realm in general. This paper endeavours to provide an unbiased, up-to-date and comprehensive assessment of the current literature on this topic. We discuss the modelling process, focusing on the choices made during different passages of model implementation, including the selection of environmental variables to represent the ecological requirements of cetacean and sea turtles' species. We also identify research trends and gaps and conclude by highlighting the strengths and limitations of each modelling approach, as well as suggesting future research directions in applying SDMs to these two specific taxonomic groups.

## 2. Material and methods

### 2.1. Modelling development steps, research questions and eligibility criteria

The SDMs modelling process ideally follows five main steps (modified after Guisan et al., 2017):

**1) Conceptualisation:** identification of species, data collection and sampling design; **2) Data preparation:** identification of available data for species and selection of ecogeographical variables (EGVs); **3) Model calibration:** identification of the type of statistical model for the data and tuning; **4) Model validation:** method used for validation, selection of appropriate metrics; **5) Model results:** model final product, such as habitat suitability or density map, and its intended scope.

According to this scheme, we formulated and defined the following research questions:

- i. Which species of sea turtles and cetaceans have been considered so far?
- ii. What type of occurrence data was used, and what methods were employed in its collection?
- iii. What kind of (correlative) modelling approaches have been applied?
- iv. Which environmental variables have been used as candidates in the models?
- v. What temporal and spatial resolution was chosen for the model?
- vi. Which validation method was used and which metrics were selected?
- vii. What is the main output for which these models were implemented?
- viii. What are the strengths, weaknesses and opportunities of each modelling approach?

Based on these research questions, related eligibility criteria were then developed using the standard approach for research questions definition (i.e., Population, Intervention, Comparator - PIC framework, CEE, 2022, Table 1).

**Table 1**  
Eligibility criteria concerning question key elements following the PIC framework.

Question key elements	Eligibility criteria
Population (P): ● Point location/occurrence data of Cetacean and/or Sea turtle species	<b>Included:</b> True <u>at-sea</u> animal spatial occurrence/abundance data about at least one cetacean or sea turtle species collected in any area of the world. <b>Excluded:</b> spatial occurrences/abundance data on the terrestrial or inland waters (i.e., nests locations, riverine species); prey occurrences/abundances; stranding data; simulated occurrences.
Intervention (I) ● Application of correlative SDMs to spatial ecology	<b>Included:</b> Any type of <u>correlative SDMs</u> which associate species occurrence with environmental predictors. <b>Excluded:</b> mechanistic models; expert knowledge-based models (i.e., Aquamaps).
Comparators (C) ● Type of correlative approach implemented	<b>Included:</b> type of approaches; environmental covariates; evaluation metrics.

2.2. Search strategy

To obtain comprehensive, balanced and accurate results, we conducted a systematic literature search in two academic databases (Web of

Science and SCOPUS), curating specific search terms and synonyms and used Boolean operators and wildcards (asterisks) to account for word variants and plural forms. Additional information on search strategy are provided in the Appendix A along with Table S1 showing all the terms included in the final search. The literature search was conducted in January 2023 and considered all studies currently available in the two databases until 31 December 2022. We did not add any language filters/restrictions to the search strategy developed so that scientific literature with an abstract in English but a full text in other languages was potentially included in the search. Except for conference proceedings, we did not include grey literature (technical reports, dissertations) because we wanted to limit our search to the most relevant and advanced scientific applications on the topic, given the extensive literature already available.

2.3. Screening process

Screening was performed at three stages for efficiency: title, abstract and full text. The inclusion and exclusion criteria were aligned with those listed and described in Table 1, ensuring consistency throughout the screening process. To include abstracts and full text, all the criteria were applied and the reason for exclusion was reported. To check for inter-rater reliability in the inclusion criteria, two reviewers independently screened a subset (10%) of titles and abstracts for relevance against predetermined criteria before the screening process began. The

**Flow Diagram for Systematic Map of evidence**

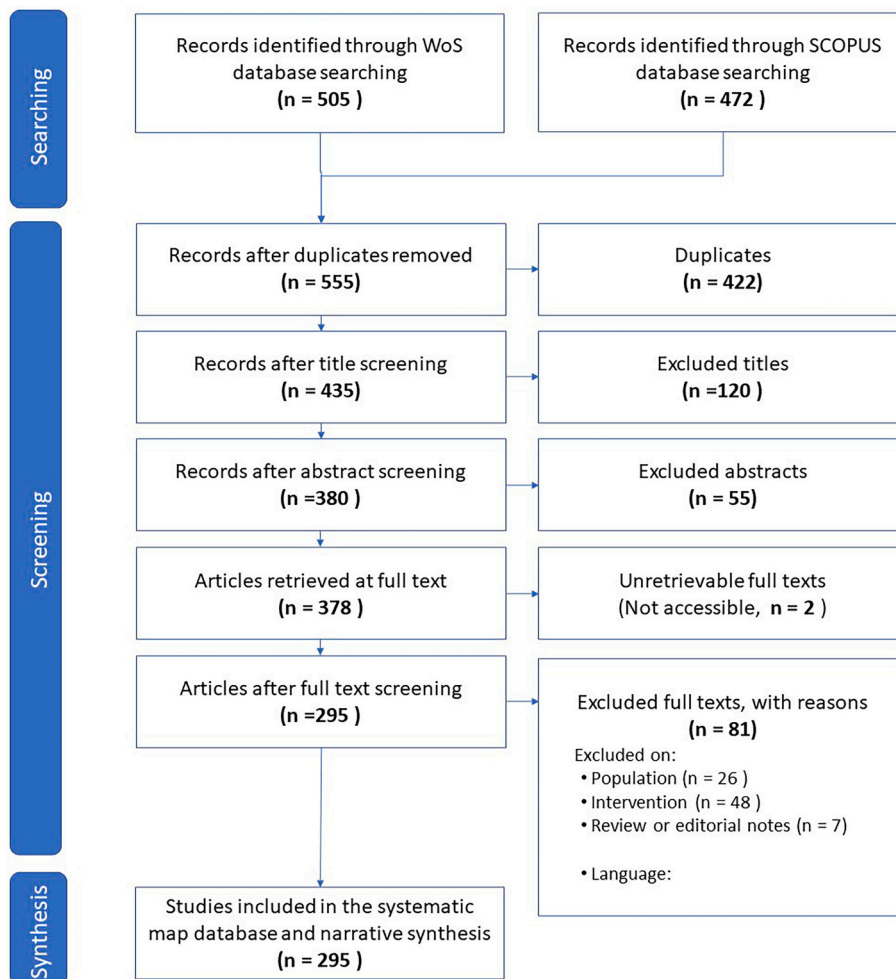


Fig. 1. Flow Diagram resuming the searching, screening and synthesis process performed systematically according to criteria listed in Table 1.

Cohen's Kappa test measured a perfect agreement ( $k = 1$ ,  $n = 100\%$ ) for both title and abstract screening. The Fig. 1 illustrates the flow diagram of the screening process.

#### 2.4. Meta data coding & data synthesis

For all included publications the following data were extracted according to the five main steps of model implementation listed in 2.1. subparagraph: 1) Conceptualisation: Family, species, study area, type of data collected (e.g., visual, acoustic), methods used for data collection, sampling design, temporal range of dataset and research scope (i.e., conservation, methodological and pure/theoretical ecology); 2) Data preparation: Source of EGVs (e.g., in situ data collection or remotely-sensed/modelled data) and type of EGVs (i.e., sea water temperature, salinity, etc.); 3) Model calibration: type of data input (i.e., count data, presence/absence, presence only), seasonality (e.g., one season only, seasonal comparison, no seasonal comparisons), model grid size, model approach, (Additive models, Machine learning, Point process, Other), model type (i.e., RF, BRT, Maxent, GLMs, GAMs); 4) Model validation: strategy used for validation (internal or external evaluation), and specific approach (i.e., Data partitioning, Cross-validation, Bootstrap, Independent dataset), type of evaluation metrics used (i.e., AUC, Kappa, TSS, other); 5) Model results: map of predicted habitat suitability/probability of occurrence, map of projected habitat suitability/probability of occurrence, risk maps, or any other map. As far as possible, the information extracted was organised into standard categories (see Appendix A for greater details), then summarised in sections discussing the most important characteristics. No statistical analysis was conducted as the focus was on reporting implementation methods rather than a specific outcome and the topics included were very heterogeneous.

### 3. Results

#### 3.1. General patterns in publication

In our analysis of 295 published articles, spanning from 2001 to 2022, we observed exponential growth in publications ( $y = a(bx)$  where  $a = 0.76$  and  $b = 1.25$ ;  $R^2 = 0.88$ ,  $p$ -value  $< 0.001$ ), particularly in the last four years as shown in Fig. 2. The majority of papers focused on cetaceans, with limited and recent studies on sea turtles, published from 2013 onwards. Geographical variations emerged in research efforts, with the North Atlantic, having the highest number of publications ( $n = 82$ , 28%), followed by the North Pacific ( $n = 65$ , 22%). Interestingly, a

notable proportion ( $n = 40$ , 13%), despite its comparatively smaller size than to other marine regions, centred its focus on the Mediterranean. In contrast, the Southern and Arctic Oceans garnered less attention overall ( $n = 19$ , 6%). The same geographical patterns were consistent across taxa, although the Southern Ocean exhibited a higher prevalence of studies on mysticetes, while being underrepresented in odontocete research, as shown in Fig. 3. The Mediterranean region was specifically examined for odontocetes rather than mysticetes, as the only regular species of mysticete is the fin whale. The complete absence of sea turtle studies in polar zones, was attributed to the limited geographical range of the species, extending from tropical to subpolar regions.

#### 3.2. Conceptualisation (species, type of data, methods for data collection, design, temporal range and research scope)

A total of 70 species belonging to 11 families were studied using SDMs. Among these, 12 and 51 species belong to the suborders *Mysticete* and *Odontocete*, respectively, and 7 species to the superfamily *Chelonioidae* (sea turtles). The families *Dephinidae* and *Balaenopteridae* were by far the most studied. The number of publications grouped by Suborders or Superfamily (for sea turtles), Families and Species are shown in Fig. 4A and B. As previously anticipated, sea turtles were underrepresented compared to cetaceans, constituting only 22 out of 295 publications (7%). Most studies looked at multiple species simultaneously ( $n = 221$ , 75%), with only four studies focussing on both cetaceans and sea turtles.

When preparing SDMs, assembling an appropriate species occurrence dataset is crucial. The occurrence data used for distribution modelling were predominantly visual ( $n = 248$ , 84%), while only a minority came from the use of satellite tracking ( $n = 37$ , 13%) or acoustic recordings ( $n = 31$ , 11%). The percentages are calculated based on the total number of reviewed articles. To be correctly interpreted, it should be noted that a few studies utilised a combination of different data types, such as both visual and acoustic recordings ( $n = 18$ ) or visual and tracking data ( $n = 3$ ). In contrast, for sea turtles, most data were from satellite tracking devices ( $n = 12$ , 67%) rather than visual data ( $n = 6$ , 33%), as evidenced by Fig. 5A. The occurrence dataset, regardless of data type, mostly covered short to medium period of 2–5 years ( $n = 89$ , 30%) and 6–10 years ( $n = 77$ , 26%). A considerable proportion also covered long-term periods of 11–30 years ( $n = 64$ , 22%), but rarely spanned periods of  $> 30$  years ( $n = 20$ , 7%). For 9 studies (3%), however, no information was provided about the temporal range of the dataset as shown in Fig. 5B.

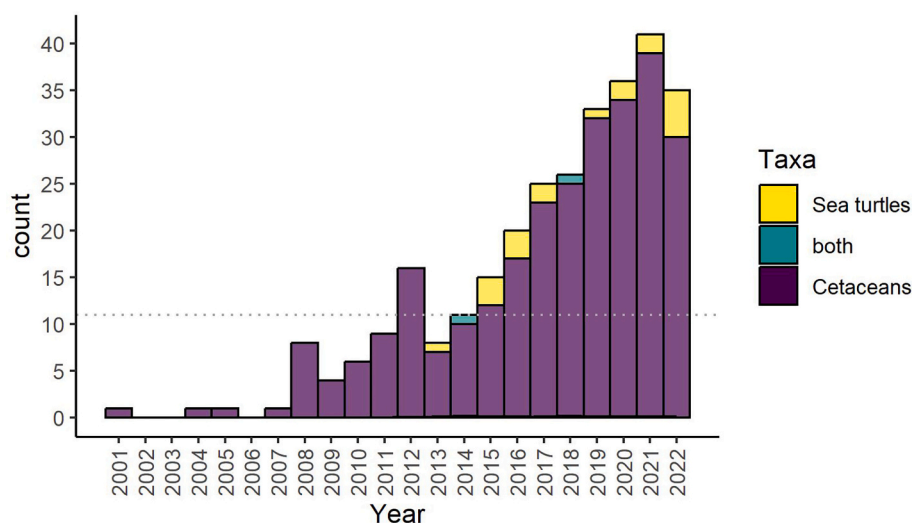
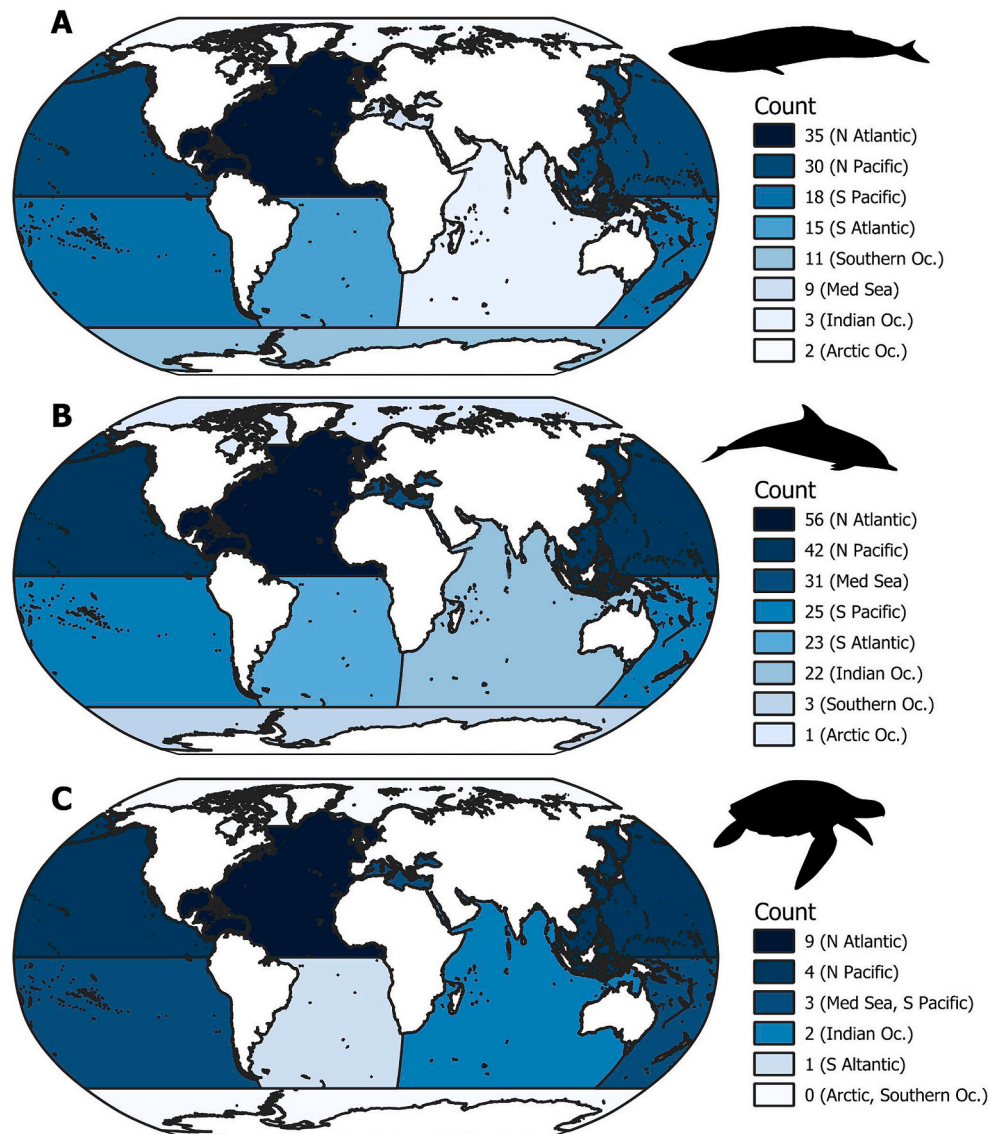


Fig. 2. Number of publications sorted by articles that focus solely on cetaceans, sea turtles, or both, in one-year intervals ranging from 2001 to 2022. The dotted grey line indicates the median value of total publications per year, which is  $n = 11$ .

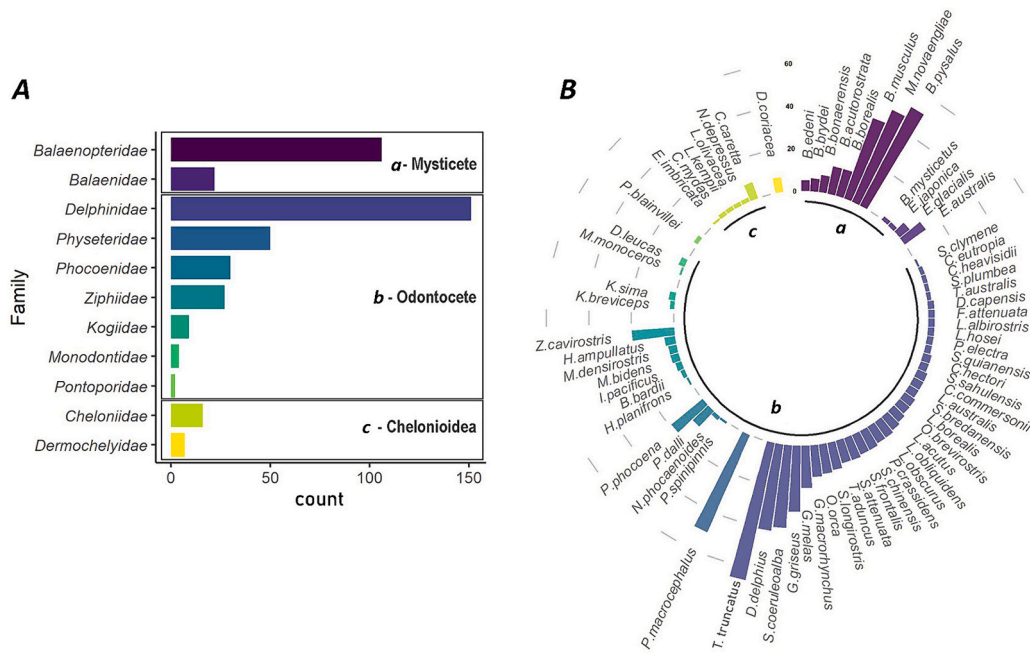


**Fig. 3.** Distribution of correlative SDMs studies worldwide applied to Mysticetes (A, total n. articles = 119) Odontocetes (B, total n. articles = 194) and sea turtles (C, total n. articles = 22) taken from 295 articles published between 2001 and 2022 and considering that some papers modelled more than one species and taxa.

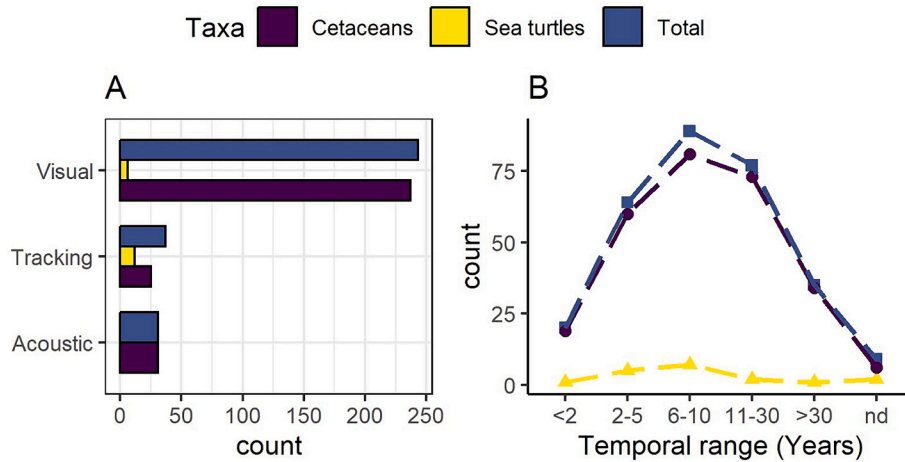
Visual data were mainly collected by the authors during in-situ dedicated surveys as evidenced in Fig. 6A ( $n = 177$ ), primarily from boats ( $n = 118$ ) and in a few cases from aircraft ( $n = 22$ ) or land ( $n = 2$ ). In the remaining cases ( $n = 24$ ), a combination of data collected from different platforms (boats and aircrafts, land and boats or all three) was used. In some studies ( $n = 9$ ), no information was provided on the observation platform used, or there was no observation platform, such as in the only study where occurrence data were collected during scuba diving ( $n = 1$ ). Regarding methods used for visual data collection within in-situ dedicated surveys, the most commonly used method was the line transect ( $n = 113$ ), conducted mostly in closing ( $n = 40$ ) or passing mode ( $n = 17$ ) or both ( $n = 7$ ), although in many cases, this specific information was not provided ( $n = 49$ ). Other methods included haphazard surveys ( $n = 31$ ), focal follows ( $n = 8$ ), point sampling ( $n = 5$ ), and strip transect methods ( $n = 4$ ), as presented in Fig. 6B. The in-situ dedicated surveys were mostly conducted along predetermined routes ( $n = 95$ ), often following an equal coverage probability design i.e., each point in the study area has the same probability of being sampled ( $n = 63$ ), as shown in Fig. 6C. In addition to dedicated surveys, occurrence data also originate from opportunistic in-situ surveys ( $n = 52$ ), conducted mainly as part of other research activities aimed at different scientific scopes ( $n$

$= 17$ ) or during whale watching ( $n = 11$ ), specific citizen science campaigns ( $n = 7$ ), fishing activities ( $n = 7$ ) or governmental coastal surveillance ( $n = 2$ ). A final proportion of data came from open/restricted access historical datasets ( $n = 53$ ), in particular from databases ( $n = 29$ , with OBIS, the Ocean Biodiversity Information System, as the most used open-access platform) or scientific and grey literature searches ( $n = 12$ ). Additionally, data were obtained from historical whaling records ( $n = 5$ ), social media ( $n = 2$ ) other sources such as scientific networks or government datasets ( $n = 8$ ), or no specified sources ( $n = 2$ ).

Regarding acoustic data ( $n = 31$ , 10%) included in the model, they were mostly from in-situ dedicated surveys ( $n = 24$ , 77%), using towed hydrophone arrays ( $n = 13$ , 54%) or autonomous recorders fixed at certain stations ( $n = 9$ , 37%), or, in the case of only two studies, using acoustic buoys, as depicted in Fig. 6D. Methods followed both predetermined ( $n = 10$ , design-based 42%) or no predetermined ( $n = 9$ , not design-based, 37%) routes/stations, while in 5 studies (21%) the information on sampling design was not provided. In a few cases only, data were taken during opportunistic in-situ surveys performed during fishing activities or other research activities ( $n = 3$ , 10%) or from open/restricted access historical datasets and literature searches ( $n = 6$ , 19%).



**Fig. 4.** Number of studied groups sorted by Family (A) and circular barplot with the number of published articles (count) for species ordered within each family group (B). In both figures, each colour corresponds to a different Family; Capital letters stand for suborder/superfamily groups a = Mysticete, b = Odontocete, c = Chelonioida.



**Fig. 5.** Number of published articles (count) using visual, tracking and acoustic data, sorted by taxa or in total (A) and the count of different temporal categories covered by the dataset (B).

In the literature reviewed, models are mainly used to study the ecology of species ( $n = 180$ , 61%) or to address conservation issues ( $n = 175$ , 59%). A minority of the articles dealt with methodological aspects of model implementation or the comparison between different modelling approaches ( $n = 70$ , 24%). A small proportion of the articles that addressed the conservation objective were aimed at assessing the risk from anthropogenic impacts ( $n = 35$ , 12%), particularly ship strikes ( $n = 23$ ), fisheries/bycatch ( $n = 35$ ), anthropogenic noise ( $n = 4$ ) and marine litter ( $n = 3$ ).

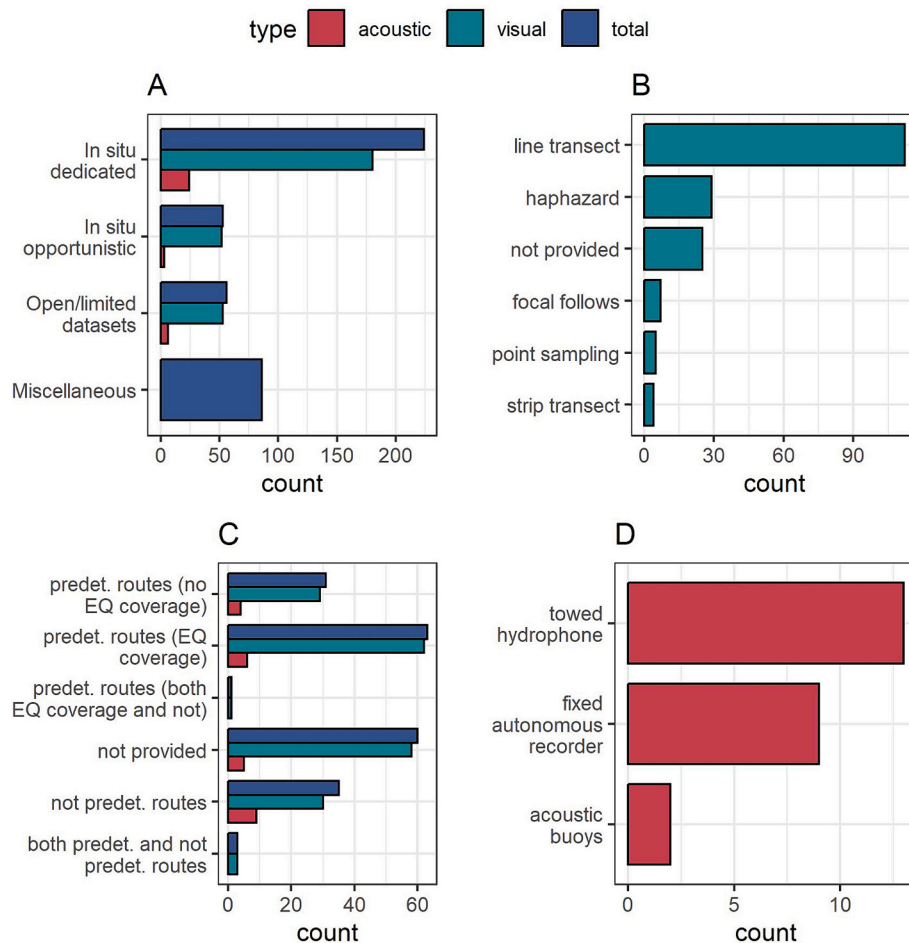
### 3.3. Data preparation (source and type of ecogeographical variables)

A variety of EGVs were employed to characterise the distribution of cetaceans and sea turtles. These variables were primarily derived from remotely sensed or modelled environmental data ( $n = 225$ , 76%), or a combination of remotely sensed and in-situ data ( $n = 48$ , 16%).

Conversely, a minority of studies ( $n = 15$ , 5%) exclusively used in-situ data. In a small number of cases ( $n = 7$ , 2%), the source of the data was not specified.

The EGVs encompass a diverse array of variables, including topographic features of the seabed ( $n = 266$ , 90% of articles), physical ( $n = 245$ , 82%) and chemical properties of the water column ( $n = 185$ , 63%) as well as geographical factors ( $n = 171$ , 60%). Environmental, anthropogenic, and climate factors were incorporated in a comparatively smaller percentage of the reviewed studies ( $n = 39$ ,  $n = 23$ ,  $n = 20$ , 13, 8 and 7%, respectively, Fig. 7a). Generally, the categories of EGVs used exhibited similar patterns among cetaceans and sea turtles as evidenced in Fig. S1.

Among the EGVs, seawater temperature and bathymetry were the most commonly used, featuring in 271 (92%) and 260 (88%) of the reviewed articles, respectively. Notably, temperature measurements were primarily related to surface temperatures, with only a few articles



**Fig. 6.** Number of published articles (count) categorised by visual, acoustic and total data for the data source (A), methods used for visual dedicated surveys (B), the sampling design employed in dedicated surveys, whether along predetermined routes or not, and whether following an equal coverage design (EQ coverage) or not (C), and the type of acoustic devices used in case of dedicated acoustic surveys (D).

explicitly considering bottom water temperature ( $n = 9$ , 3%) or temperatures at different depths ( $n = 14$ , 5%). The top ten most frequently used EGVs are illustrated in Fig. 7b and included bottom slope ( $n = 168$ , 57%), chlorophyll-a concentration ( $n = 147$ , 50%), distance to the shore ( $n = 135$ , 46%), water salinity ( $n = 57$ , 19%), sea surface height ( $n = 57$ , 19%), distance to bathymetric contours ( $n = 47$ , 15%), geographical coordinates ( $n = 44$ , 15%) and bottom aspect ( $n = 38$ , 13%). The complete list of EGVs and their frequency of use, sorted by *taxa*, can be found in Table S2.

The EGVs used predominantly represented the uppermost layers of the water column, with fewer studies incorporating variables reflecting the seabed and deeper layers of the water column. Specifically, only 15% ( $n = 44$ ) of studies considered variables including the third dimension in the model such as water temperature, current velocity and salinity at different depths and/or at the bottom.

### 3.4. Model calibration (occurrence data, spatial resolution, modelling approach, model type)

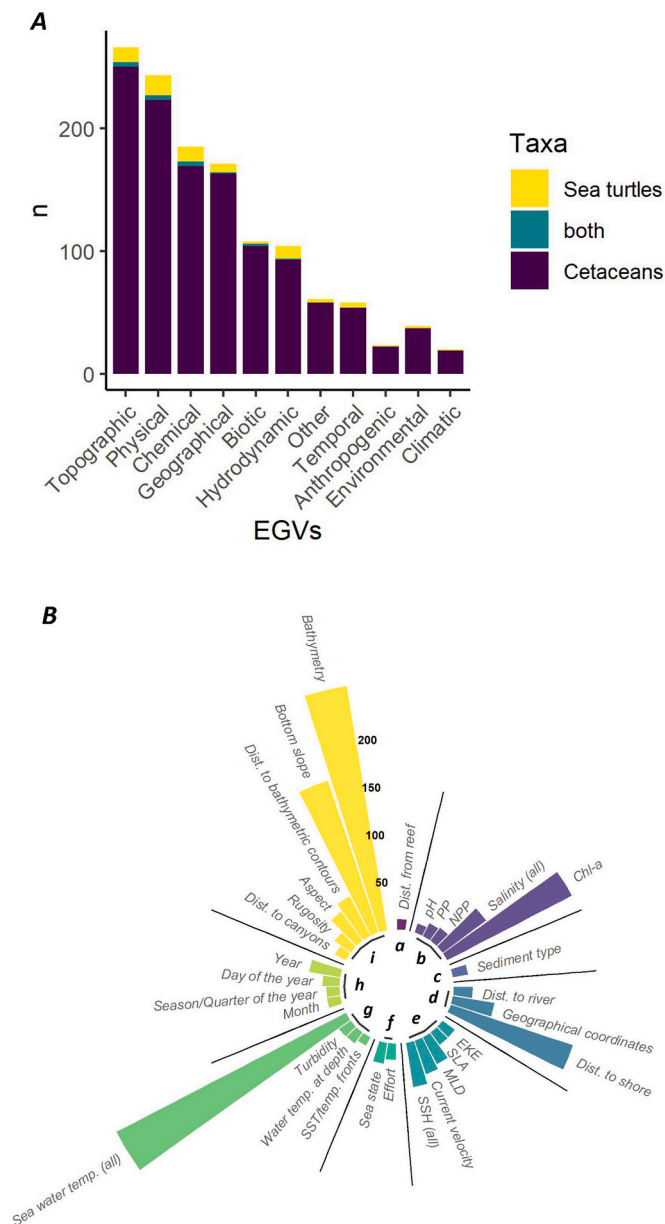
The predominant category of occurrence data used for modelling was “presence-pseudo absence/background data” ( $n = 125$ , 42%). This category involves the observed presence of animals and artificially generated or selected absence data, commonly referred to as background data or pseudo-absence. The second category “Count/abundances”, includes discrete and continuous response variables such as counts of animals (0, 1, 2, 3, etc.) and species distribution abundances, which are used more or less equally. In contrast, “presence only”

(without background/pseudo absences generated), and “presence-absence” were utilised to a lesser extent, as shown in Fig. 8.

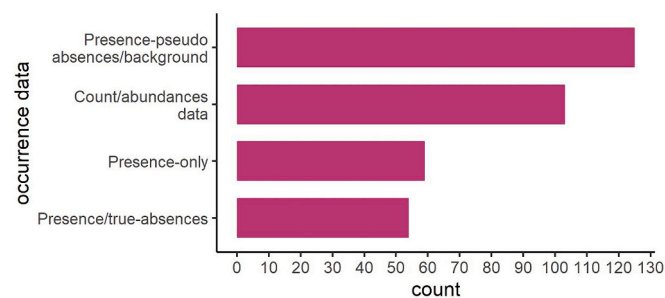
The most commonly employed spatial resolutions for grid creation fell into the “medium” (1–10 km) classes ( $n = 133$ , 51%), followed by the “fine” (100 m - 1 km) resolutions ( $n = 66$ , 25%). The other classes “very fine” (<100 m), “coarse” (10–50 km) and “very coarse” (>50 km) constituted a small proportion of the sample ( $n = 12$ ,  $n = 38$  and  $n = 7$ , corresponding to 5, 15 and 3%, respectively). The three most frequently used grid sizes included the 1 km grid ( $n = 40$ , 15%), followed by 10 km ( $n = 32$ , 12%) and 5 km ( $n = 23$ , 9%). In some articles, however, no information was provided on the spatial grid used ( $n = 35$ , 13%). We excluded from this results articles using modelling solely for descriptive purposes and considered only those producing spatial prediction/projection output ( $n = 160$ , 61% above all reviewed articles).

A minority of cases included in their modelling a comparison between different seasons ( $n = 83$ , 28%), while most referred to one season only ( $n = 105$ , 35%) or did not consider seasonality at all ( $n = 107$ , 36%).

Regarding the modelling approaches, our results indicate that the most popular modelling approach is the regression-based approach (RB,  $n = 183$ , 62%), especially generalised additive models (GAM,  $n = 145$ , 50%) and generalised linear models (GLM,  $n = 60$ , 20%). As evidenced in Fig. 8B, following closely are machine learning methods (ML), with Maxent being the most commonly used ( $N = 87$ ; 30%). Other well-represented machine learning methods include Random Forest (RF,  $n = 27$ , 9%) and Boosted Regression Trees (BRT,  $n = 22$ , 7%). As shown in Fig. 9A, the use of the RB approach to modelling dates back to 2001,



**Fig. 7.** Studies (count) grouped for different EGVs categories and taxa (A) and most frequently used EGVs (only counts equal or >10 articles are shown) (B), lowercase letters within the centre of the circular plot represent EGVs categories (a = biotic/ecological, b = chemical, c = environmental, d = geographical, e = hydrodynamic, f = other, g = physical, h = temporal, i = topographic).



**Fig. 8.** Different categories of occurrence data with the corresponding number of published articles (count).

while the first study used machine learning in 2007 with Classification and Regression Tree (CART), Discriminant Analysis and then with the integration of Maxent in 2010. Over the years, RB and ML approaches experienced a proliferation in usage, encompassing various innovative ML for both boosting and classification techniques that gained popularity later on (e.g. RF, BRT, GBM, ANN), or distance-based and envelope models (DIST-ENV) in 2001 and, more recently since 2018, the models based on point-processes (PPMs). Additionally, the ensemble modelling approach, where the results of different models are averaged (ENS,  $n = 34$ , 11%), has also increased since its first use in 2013.

**3.5. Model validation (internal/external validation strategy and metrics)**

Various strategies were employed for the internal/external validation of the model with a predominant use of internal strategies ( $n = 174$ , 59%). The most common internal strategy was “cross-validation” (CV), utilised in 84% of cases ( $n = 147$ ). Among the CV methods, “k-fold” was most frequently used ( $n = 82$ , 56%), followed by “bootstrap” ( $n = 34$ , 23%), “leave-one-out CV” ( $n = 19$ , 15%), “repeated split sampling” ( $n = 5$ , 3%) or “not stated/other CV” (including stratified monte carlo CV and spatial block CV,  $n = 13$ , 9%). Internal validation was also performed using “simple sample splitting” ( $n = 16$ , 5%) or “generalised CV” ( $n = 12$ , 4%), which does not require data partitioning into training and test datasets. Instead, it utilises a mathematical formula that considers the trajectory of fitting errors while varying the model’s hyperparameters (Hastie et al., 2009). Regarding external validation strategies, only a minority of studies ( $n = 39$ , 13%) employed a fully independent dataset for model validation, either in addition to internal validation strategies ( $n = 22$ ), or as the sole validation method ( $n = 17$ ). Notably, a considerable number of studies did not use an explicit validation method to measure model performance or did not clearly state this in the text ( $n = 94$ , 32%) (category “not provided” in Fig. 10).

A diverse set of validation metrics emerged from the analysis of studies performing model predictions ( $n = 260$ ), with the most common displayed in Fig. S2. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC curve) was the predominant choice ( $n = 144$ , 55%). However, in many cases, the metrics used were not clearly stated or provided in the text ( $n = 59$ , 23%). Approximately half of the studies ( $n = 104$ , 40%) considered more than one validation metric simultaneously to assess different aspects of model performance. Multiple software tools were employed for modelling, with a focus on R and its associated packages, and listed in Table S3.

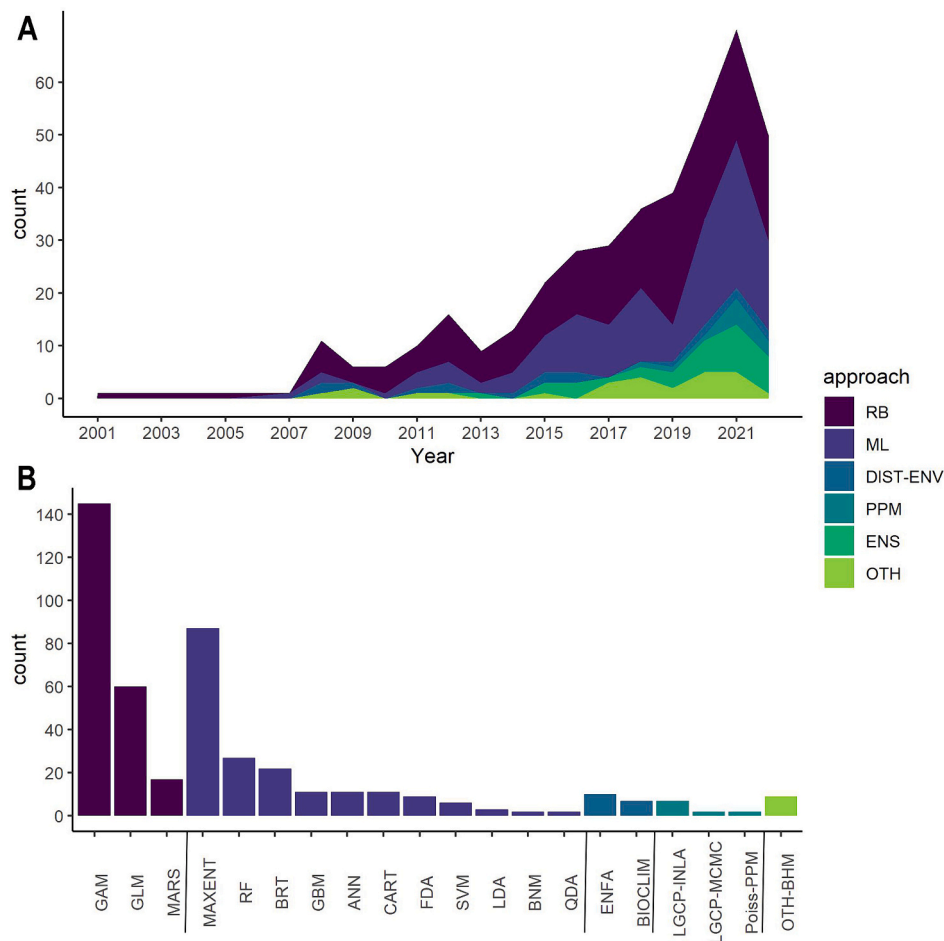
**3.6. Model results**

The majority of model results consisted of species spatial predictions ( $n = 241$ , 82%), as opposed to species spatial projections (predictions outside the study area or time used to train the model,  $n = 35$ , 12%). In the remaining cases, no spatial predictions or projections were made, and modelling was conducted solely for descriptive purposes ( $n = 35$ , 12%).

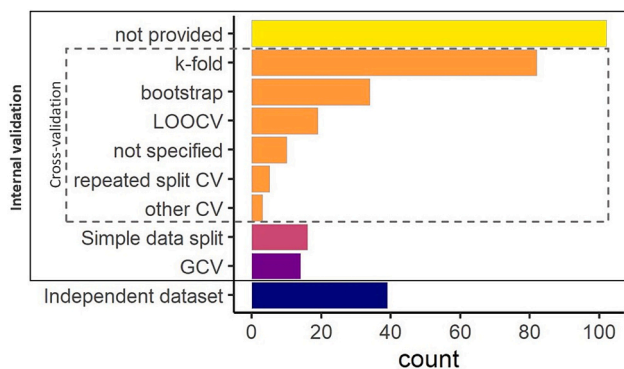
As shown in Fig. S3, spatial predictions were predominantly used to generate maps of habitat suitability or species potential distribution ( $n = 189$ , 64%) and, to a lesser extent, to predict species densities ( $n = 61$ , 21%). Similarly, when spatial projections were produced, the majority focused on generating maps of habitat suitability/potential distribution ( $n = 28$ , 9%), with a minority producing maps of projected densities ( $n = 7$ , 2%). Projections were primarily made through time ( $n = 23$ ) rather than through space ( $n = 5$ ). Lastly, a limited number of studies utilised the results of spatial modelling to produce risk assessment maps ( $n = 22$ , 7%).

Comprehensive list of included articles and the extracted information is accessible as a csv file in the Appendix B.





**Fig. 9.** Modelling approaches used in the examined articles (count) over the years (A): RB = Regression-based, ML = Machine-learning, DIST-ENV = distance based/envelope, PPM = Point Process Models, ENS = Ensemble approach, OTH=Other approaches. Specific models implemented in reviewed literature (B): GAM = Generalised Additive models; GLM = Generalised linear models; MARS = Multivariate Adaptive Regression Splines; MAXENT = Maximum Entropy model; RF = Random Forest; BRT = Boosted Regression Trees; GBM = Generalised Boosted Models; ANN = Artificial Neural Networks; CART = Classification and Regression Trees; FDA = Flexible Discriminant Analysis; SVM = Support Vector Machine; LDA = Linear Discriminant Analysis; BNM = Bipartite Network modelling; QDA = Quadratic Discriminant Analysis; ENFA = Ecological Niche Factor Analysis; BIOCLIM = BIOCLIM; LGCP-INLA = Log-gaussian Cox Process via Integrated Nested Laplace Approximation; LGCP-MCMC = Log-gaussian Cox Process via Markov Chain Monte Carlo; OTH-BHM = Other Hierarchical Bayesian models.



**Fig. 10.** Strategies and methods used for model validation with the number of published articles (count); CV=Cross-validation; LOOCV = Leave-one-out CV; GCV = Generalised CV.

#### 4. Discussion

Our analysis suggests that Species Distribution Models (SDMs) applied to cetaceans and sea turtles have seen a remarkable increase in recent years, proving to be valuable tools to compensate for the lack of

data on highly mobile species that move in dynamic and difficult-to-survey environments. The multitude of SDM studies focusing on these species has significantly advanced our understanding of the ecology of cetaceans and sea turtles in different areas of the world (e.g., Carman et al., 2019; Correia et al., 2019; Monsarrat et al., 2015). Moreover, some studies have proven instrumental in conservation efforts by identifying areas of ecological importance and particularly vulnerable to anthropogenic activities in different world oceans (e.g., Hunt et al., 2020; Passadore et al., 2018; Trew et al., 2019; Zanardo et al., 2017). Several studies also played a crucial role in discussing methodological aspects that significantly contributed to the understanding and advancement of model implementation for these species (i.e., Bennington et al., 2020; Derville et al., 2018; Scales et al., 2017b; Virgili et al., 2018). However, our review identified several gaps. While the application of SDMs for cetaceans has significantly expanded, their utilisation for sea turtles remains limited, relying mainly on satellite tag data, which only represent population habitat preferences with large sample sizes (Sequeira et al., 2019). Even within cetaceans studies, there is a predominant focus on the same few species (e.g. bottlenose, short-beaked and Risso's dolphins and fin, sperm, humpback, blue, pilot whales), with the majority of studies concentrated in the northern hemisphere as observed for sea turtles as well. Furthermore, despite the diverse range of emerging modelling methods, the same few methods

have predominantly been applied. Finally, a multitude of decisions that can be taken during the model development has emerged, without clear guidance. These gaps highlight untapped potential to expand SDM studies into different geographical areas, encompassing less studied species, and exploring emerging modelling methods. Urgency is emphasised for modelling clarification and standardisation to ensure consistency and reliability in outcomes.

#### 4.1. Issue with data collection and the challenge of defining “true” absences

The first crucial step in the modelling process involves assembling an appropriate species occurrence dataset which will influence all the following modelling steps, including the selection of proper modelling techniques (Guisan et al., 2017). Based on our results, this was predominantly achieved through dedicated effort-based field surveys, and to a slightly lesser extent, from opportunistic field surveys (e.g., whale watching, other research activities, or fishing) or the extraction of information from existing databases that gather data from diverse sources such as social media records, whaling, and museum data. The choice of data type, source, and collection method significantly impacts the availability and reliability of “absence” data, a critical aspect in modelling. While presence data are more easily available, obtaining reliable absence data in the marine environment for highly mobile species is challenging (Fernandez et al., 2017), as evidenced by the prevalent reliance on presence-only data in our reviewed articles. For visual methods, for example, sampling an area at a specific time might imply absence, but this inference may not accurately represent the true absence of a species. Factors such as availability bias (animals are submerged), influenced by species dive duration and the proportion of time spent at the surface, and perception bias (animals are at the surface but not detected), affected by animal behaviour, group size and survey conditions (sea state, visibility), contribute to the intricacies of interpreting absence. These challenges pose significant obstacles to subsequent modelling processes (Fernandez et al., 2017; Guisan et al., 2017), especially if the scope is estimating species abundances (Hammond et al., 2021). At the same time, the presence of reliable absence data is crucial in modelling, theoretically enhancing the modelled relationships between species occurrence and environmental factors, thereby facilitating more accurate assessments (Brotons et al., 2004; Wisz and Guisan, 2009).

In light of these considerations, dedicated surveys should be planned to minimise biases and accurately come as close as possible to the “true” absence, whenever feasible. One of the standard methods for dedicated surveys in our reviewed papers was the line-transect based on distance sampling, mostly used for collecting visual data and incorporating detectability functions in the models to correct associated detection and availability biases (Buckland et al., 2001). In addition to the methodology employed in collecting data, the sampling design and the degree of “representativeness” of the sample are equally pivotal factors to be considered. This can be accomplished by improving the uniformity of sampling probabilities across the study area, thereby ensuring that survey efforts comprehensively cover diverse environments (Bao et al., 2019; Thomas et al., 2010), mirroring the approach adopted in numerous reviewed papers. This precaution is necessary also to prevent bias towards easily accessible habitats, well-known utilisation areas, or other types of errors that usually occur in the case of haphazard/preferential surveys. This does not imply that data from non-uniform sampling cannot be integrated into a model but emphasises that such data come with a sampling bias that needs to be carefully considered in both the development of the model and its subsequent interpretation.

In addition to visual methods, although limited, some studies focused on cetaceans have successfully integrated acoustic data collected through passive methods, such as towed hydrophone arrays (e.g., Pace et al., 2018; Pirota et al., 2020, or albeit not included in reviewed articles Boisseau et al., 2024) or recorders placed at fixed stations (e.g.,

Frasier et al., 2021) into SDMs. These methods offer several advantages in studying especially cryptic, deep-diving species, such as sperm whales, beaked whales, and Risso's dolphins, which spend little time at the surface. In contrast to visual sightings, acoustic methods facilitate the detection of animals regardless of light and weather conditions (Marques et al., 2013), effectively mitigating both perception and availability biases for abundance estimation. Furthermore, advancements in automated techniques for detection and classification of cetacean signals have significantly improved the speed and accuracy of signal analysis (Usman et al., 2020; Usman and Versfeld, 2022). The limited number of SDMs studies based on acoustics might arise from the necessity of specialised and often expensive platforms and instruments, constraining the temporal and geographical range of their application and the amount of data. Furthermore, relying solely on acoustic data raises challenges as the instrument may fail to recognize the animal based on its position and depth, or given that the animal does not vocalise continuously, and it is challenging to accurately determine the number of individuals in an acoustically detected group (Lerebourg et al., 2023; Marques et al., 2013). As a consequence, also relying solely on acoustic data may fall short of accurately identifying true absences. A promising avenue for future modelling studies involves the exploration of a combined use of both visual and acoustic data, maximising the strengths of each method and overcoming their respective limitations, as done in the 6% of analysed studies (e.g., Frasier et al., 2021).

Despite the prevalence of dedicated effort-based field surveys which could theoretically capture both species presence and absence, the considerable uncertainty regarding the true absence of species has led modelling studies included in our review to predominantly treat data prevalently as presence-only or, as a common strategy to overcome the lack of true absences, researchers often opted for the artificial selection of absences, commonly known as pseudo-absences or background points, to enhance model performance. Diverse strategies have been employed to identify points more likely to represent true absences, or at the very least, to avoid including presence points. These strategies include creating a buffer zone around the survey track and selecting locations distant from the point of presence (e.g. Bennington et al., 2020; Chatzimentor et al., 2021; Ham et al., 2021), or using cells without sightings but with the highest survey effort (e.g. Arcangeli et al., 2016; Passadore et al., 2018; Sprogis et al., 2018). Another approach involves using the presence of non-target but related species in the same dataset as the absence (known as the “target method”, e.g. García et al., 2022; Bonneville et al., 2021; DiMatteo et al., 2022). In other cases, absences were generated into regions of lowest suitability predicted from another model (e.g., Abrahms et al., 2019). For the studies employing satellite tracking data, where we know with certainty only the presence of the animal, absences were in some cases simulated through state space models, such as continuous time-correlated random walk (e.g., Pérez-Jorge et al., 2020; Reisinger et al., 2022; Scales et al., 2017a).

It's worth noting that the terms ‘pseudo-absences’ and ‘background points’ are often used interchangeably, although distinctions exist. Specifically, ‘background points’ refer to randomly generated data that characterise the study area, encompassing potential locations where species sightings may occur (as defined by Phillips et al., 2009). On the other hand, ‘pseudo-absences’ represent data designed to mimic absence by selecting locations with minimised occurrence probabilities based on species ecology (Fernandez et al., 2022). Nevertheless, due to the improper use of the terms, we found it necessary to consider them collectively in this work. Considering the inherent complexity and the discussed challenges related to absence data in the ecological modelling of these species, there is a critical need to consolidate and summarise various approaches used in handling absence data and to advocate for consistent usage of terms in future studies.

Although to a lesser extent compared to dedicated surveys, the data included in SDM modelling also comes from large databases or citizen science social media records. These data are typically collected in an unstandardised manner. However, in contrast to data from dedicated

surveys, they are abundant and have become increasingly available, thereby constituting promising data sources. Simultaneously, they exhibit poor quality, lack comprehensive metadata, and often originate from various sources, necessitating careful consideration due to several biases. These biases encompass uncertainty in species identification (resulting from the absence of expert-collected data), improper design, bias towards most accessible areas, low or unknown sample location accuracy, incomplete or uneven spatial coverage of the true distribution of a species, and the potential presence of spatial autocorrelation in sample locations (Guisan et al., 2017). All these biases must be evaluated and addressed, and a cautious interpretation of these data becomes imperative (Barve et al., 2011).

#### 4.2. The importance of including biotic, abiotic and anthropogenic variables

Another integral aspect of SDMs involves the careful selection of environmental variables that are both spatially and temporally relevant to the site records of target species. In our comprehensive review, we extracted over 250 different candidate variables, predominantly derived from remotely sensed or modelled data rather than collected in the field. The extensive array of predictors utilised stems from the wealth of spatial environmental data services and platforms currently available (main platforms used are reported in Table S4). Regardless of the extensive set of available environmental variables, their selection demands careful consideration to ensure the accuracy and effectiveness of SDMs, acknowledging that animal movement represents a complex interplay of both abiotic and biotic factors, as eloquently articulated by Hutchinson, 1957.

The distribution of cetaceans and sea turtles is largely shaped by the availability and abundance of prey. However, our findings indicate that models predominantly relied on abiotic variables (e.g., water temperature, bathymetry, slope, chlorophyll-a, salinity, sea surface height), which indirectly relate to prey abundance, rather than incorporating direct variables associated with prey. This is due to the challenges of collecting prey data in the dynamic marine environment, especially at the relevant temporal and spatial scale where predators actively forage (Benoit-Bird and Au, 2003; Hyrenbach et al., 2000). However, even when prey data are available, challenges may arise from the temporal or spatial lags between predator and prey data, as well as between physical and biological processes in general (Redfern et al., 2006; Torres et al., 2008). As a consequence, researchers often use abiotic variables as proxies for prey availability due to their ease of sampling (e.g., Abrahms et al., 2019; Becker et al., 2016; Escobar-Flores et al., 2013; Hazen et al., 2017; Palacios et al., 2013; Palacios et al., 2019), potentially leading to an incomplete understanding of the complex biotic dynamics in the marine environment (Bennington et al., 2020). Including biotic data, would improve the performance of models for different cetaceans species at the local scale (Barlow et al., 2020; Bennington et al., 2020) and recent findings for terrestrial species also highlighted the significant role of biotic variables at the macro scale (Cosentino et al., 2023), challenging the historical belief that their influence is limited to ecological processes at locale scale only (Guisan and Thuiller, 2005; Wisz et al., 2013).

To enhance accuracy, future modelling efforts should combine biotic and abiotic variables. This requires improving simultaneously data collection in prey and predators, facilitated by innovative technologies such as hydroacoustic backscatter (Barlow et al., 2020; Širović and Hildebrand, 2011), underwater video measurements (Bennington et al., 2020), eDNA (Valsecchi et al., 2021; Zhang et al., 2023), and stable isotope sampling (McCormack et al., 2019). These advancements will help to realistically represent the spatial and temporal variability of both prey and predators at the proper scale.

In addition to predation, other intricate biotic relationships such as association, competition, mutualism, and facilitation, may shape species' spatial patterns (Wisz et al., 2013) but these were rarely

investigated in our reviewed studies, except for one study using a novel modelling framework, known as Joint Species Distribution Models (JSDM, Wilkinson et al., 2021). JSDMs seamlessly integrate into the classical SDM framework, allowing the exploration of correlation patterns across multiple taxa and their environmental responses; e.g. Astarloa et al. (2019) found a positive association between whales and seabirds in the Bay of Biscay during feeding, likely due to mutualistic associations. The scarcity of similar studies underscores the imperative for further exploration to unveil complex ecological processes related to cetaceans and sea turtles, such as predation, competition or mutualism with other taxa.

The limited inclusion of environmental characteristics representing conditions at different depths, such as temperature or salinity sampled/ modelled at various bathymetric layers, and the reliance on data from surface layers (a two-dimensional approach) indicated that most studies overlooked the three-dimensionality of the seascape. This oversight may lead to an oversimplification of species responses, especially for cetacean species such as deep-divers (Guerra et al., 2022; Virgili et al., 2022), which routinely use different depths for foraging. In the future, enhancing SDMs to incorporate both vertical and horizontal preference will better capture the full spectrum of habitat utilisation by both cetacean and sea turtles species.

Despite the potential importance of human activities as drivers of distribution, their inclusion in current models is rare, highlighting another area that deserves more attention in future research. Notably, vessel traffic, fishing and aquaculture activities emerged from reviewed studies as factors influencing the spatial behaviour of our target species in various ways. For example, bottlenose dolphins are potentially influenced by the proximity of port complexes, fishing and shipping routes given the improved prey availability (Maricato et al., 2022), or attracted to areas close to aquaculture or fish farms, in response to likely higher nutrients and related increase in prey abundance (Bearzi et al., 2016; Bonizzoni et al., 2019; Passadore et al., 2018). Fish farms, in particular, may attract wild fish by providing structure, a refuge from predators, and food resources with influences extending beyond the immediate vicinity of the farmed area (Machias et al., 2005). For fin whales, a significant decrease in call detections with increasing shipping noise emerged in offshore Irish waters (Ramesh et al., 2021) suggesting that this factor may have a role in shaping fin whale distribution. Recent availability of marine soundscape maps, including anthropogenic noise (e.g., Ho et al., 2024), could indeed be useful to consider the influence of this factor in SDMs analysis. For sea turtles, a study in Honduras by Wright et al., 2022, the only one to consider anthropogenic variables in their distribution modelling, traffic appeared to have no effect or clear causal relationship with hawksbill sea turtle distribution. However, additional research is essential in other regions and different sea turtle species, to fully understand this and other anthropogenic interactions.

#### 4.3. Temporal and spatial modelling resolutions

The spatial extent of the study area and cell resolution used for spatial analysis may significantly impact model results (Guisan and Thuiller, 2005; Levin, 1992; Redfern et al., 2006). This influence arises from the spatial dependencies in species-habitat relationships (Redfern et al., 2008) and the wide-ranging spatial behaviour of most species. In cases where the study area is relatively small, spatial sexual segregation might hinder data, resulting in modelling outcomes that reflect the habitat selected by one type of social aggregation (Pace et al., 2018). Alternatively, species could select different habitats depending on the local conditions, making a geographically localised dataset unable to capture the full variability of the species preferences (Azzolin et al., 2020). Also, the choice of cell resolution for modelling is not straightforward. While fine to medium resolutions (1–10 km), prevalent in the reviewed studies, may appear advantageous for management, they can create a false sense of precision (Baines and Weir, 2020) and fail to represent essential long-range ecological processes (Baines and Weir,

2020; García et al., 2018) or introduce issues with data spatial autocorrelation. Simultaneously, too coarse resolutions may miss meso- and submesoscale scale oceanographic variability, such as locally and temporally enhanced favourable conditions (García et al., 2018).

Temporal extent and resolution are other relevant factors to consider in the marine environment (Fernandez et al., 2017; Mannocci et al., 2017; Scales et al., 2017b), strongly affecting model results, particularly for species influenced by variables with higher dynamism (i.e., blue, fin, sei whales, and various dolphin species). Long-term data, based on how long the species lives, may more effectively capture the environmental realised niche of a species and be more suitable for projecting it into the future (Guisan et al., 2017). While seasonality significantly influences cetaceans (e.g., Arcangeli et al., 2017) and sea turtles distribution, it is worth noting that several reviewed studies gathered data exclusively during a single season. This limitation introduces potential bias, as the observed patterns may be influenced by seasonal biological or ecological constraints specific to the species (Arcangeli et al., 2024). Concerning temporal resolution, different environmental variables can show significant variation over a range of timescale. While climatological time scale (seasonal or annual) might produce better distribution estimates of cetaceans (Mannocci et al., 2014), some studies conversely found that finer temporal resolution (daily, weekly, or at maximum monthly) produce better results (Becker et al., 2012, 2014; Forney et al., 2015; Scales et al., 2017b). Scales et al. (2017b) also suggested that predicting on the finest temporal scale and averaging those predictions, rather than averaging environmental data, may enhance model predictive capacity. In summary, the selection of an adequate temporal and spatial extent and resolution can be a complex issue and researchers must be aware of the scales of ecological processes and their relevance to the study's objectives. This awareness should guide the sampling design, environmental data preparation, modelling approach, and ultimately, the breadth of ecological conclusions drawn from the study.

#### 4.4. The “best” modelling approach

In the domain of scrutinised studies about species distribution modelling (SDM) for cetaceans and sea turtles, a plethora of methodologies has emerged, leading to continual refinements in modelling techniques. Among the diverse approaches, regression-based methods were the earliest applied and, particularly the Generalised Additive Model (GAM), have predominantly emerged, with their applicability increasing over time. Although a preceding review identified Maxent as the most favoured model used for the marine realm at large (Melo-Merino et al., 2020), the overall preference for GAM can be attributed to the robust statistical theory on which regression-based model in general, the ease of interpretation and the greater flexibility in modelling different response variables types (binary, discrete, continuous) and suitability to analyse non-linear complex relationships, not requiring postulating a shape for the response curve from a specific parametric function (non-parametric) such as in GLM (Guisan et al., 2017). Furthermore, GAMs demonstrate effectiveness in handling temporal variation in distribution. On the other hand, Maxent proved versatile and potent tool for the marine realm, as it necessitates relatively modest information, rendering it an efficient choice for modelling different species of cetaceans (e.g., Azzolin et al., 2018; Azzolin et al., 2020; Fernandez et al., 2018; Friedlaender et al., 2011; Gregoriotti et al., 2021) and sea turtles (e.g., Zampollo et al., 2022). Maxent exhibits robust predictive accuracy even when dealing with diminutive or unequal sample sizes as for the case of elusive species or deep divers (Arcangeli et al., 2023; Arcangeli et al., 2024; Pace et al., 2018). Its appeal is further heightened by the model's flexibility in construction, facilitated by the user-friendly interface of the software or developed R packages, enabling users to customise the model according to specific needs and available information, incorporating “features classes” and a “regularisation multiplier” to find a proper trade-off between flexibility and overfitting (Muscarella et al., 2014). The advantages of Maxent accounts for its widespread use,

ranking as the second most applied model in our review, following GAM. Nonetheless, it is crucial to acknowledge certain limitations associated with both GAM and Maxent applications. GAMs, along with other regression-based models, assume additive relationships between predictors, limiting their ability to capture complex, nonlinear and intricate interactions. On the other side, Maxent, specifically designed for presence-only data, necessitates different strategies to improve the representativeness of the study area and address sampling bias, such as selecting background points or generating pseudo-absences. Furthermore, careful regularisation and parameter tuning are essential for Maxent, as default settings may lead to suboptimal model performance (Radosavljevic and Anderson, 2014).

Among new promising approaches, Point Process models (PPMs) have emerged and have been increasingly applied to model presence-only data for cetaceans and sea turtles. Unlike standard regression methods, PPMs primarily focus on estimating and modelling the intensity function of the spatial location of observations, capturing spatial variability and providing a measure of uncertainty (Renner et al., 2015). PPMs have demonstrated proficiency in addressing main challenges linked to presence-only data, such as uncertainties related to irregularly sampled marine data, influenced by factors like weather conditions, site accessibility, and specific research objectives. PPMs prove particularly valuable when researchers need to integrate data from different sources to construct robust modelling (Martino et al., 2021; Pace et al., 2022). Additionally, PPMs may incorporate effort information to mitigate biases arising from variations in sampling intensity or detection probabilities. Poisson-PPM models represent the simplest type, assuming a deterministic intensity function and independencies of spatial events. In contrast, Bayesian-based PPMs, like Log-Gaussian Cox Process Models (LGCP-PPMs), consider intensity as a stochastic process with a Gaussian distribution and incorporates spatial dependencies in the process, which is particularly relevant for modelling animals with grouping tendencies or specific avoidance behaviours such as cetaceans. Indeed, LGCP-PPMs have found successful application in the study of cetaceans, particularly in integrating multisource data, encompassing both acoustic and visual information and including detection functions and accessibility explanations to correct detection and sampling bias (Martino et al., 2021; Pace et al., 2022). On the other hand, the implementation of LGCP-PPMs, being based on a Bayesian approach where prior distributions need to be specified for model parameters to express uncertainty, is not straightforward, is case-specific, may be computationally costly, and requires a solid understanding of Bayesian statistics. Main advantages and disadvantages associated with models and approaches, are reported in Table S5.

Finally, ensemble modelling is a popular approach that has gained traction over time, as it offers several inherent advantages. This approach is especially useful when the primary objective is to predict species distribution rather than merely describing species-environment relationships (Renner et al., 2015). Ensemble methods rely on a combination of different modelling techniques instead of depending on a single algorithm. They leverage the strengths of diverse models, mitigating their weaknesses, thereby enhancing overall prediction accuracy and robustness.

#### 4.5. The cruciality of independence between training and testing datasets

Although a model can never be absolutely “true,” the core of modelling is to discover the best-fitted model that closely approaches conceptual truth (Anderson and Burnham, 2002). This critical evaluation is undertaken during the validation phase, an essential step in the modelling process. Model processing indeed involves both calibration and validation phases, where presence sites are divided into two independent samples: the “training” dataset, used to build the model, and the “testing” dataset, used to assess the model's efficiency with a subset of data not employed for calibration (Guisan et al., 2017; Phillips et al., 2006).

Among the various validation strategies that exist, differing in how data for validation are selected, our findings have indicated a predominant reliance on ‘internal’ validation strategies, where the data used for validation are selected from the same dataset used for constructing the model, particularly cross-validation strategies such as “k-fold”. This is in contrast to ‘external’ validation methods, where the model is evaluated with fully independent data. Internal validation practices, relying on the same data from the same area and period, introduce potential biases in estimating model performance due to a lack of full spatial and temporal independence. According to Araujo et al. (2005), internal validation with non-independent data always leads to an overoptimistic assessment of model performance. Therefore, the use of external validation methods becomes crucial, a step that was unfortunately taken in only a minority of the reviewed articles. Testing model accuracy with a fully independent dataset is considered the most robust method for assessing SDMs (Araujo et al., 2005; Guisan et al., 2017), and is essential especially for models intended for management applications to ensure SDMs become integral to decision-making processes. The constrained use of independent datasets most likely arises from the challenges and cost implications associated with obtaining external data, especially considering the need for diverse approaches across various areas and periods to ensure true independence. A promising opportunity lies in taking data from standardised citizen science programs for validation, as suggested by Matutini et al. (2020). However, as suggested by the same authors, caution is advised in prefiltering this data type to ensure its reliability. Main advantages and disadvantages of each model validation strategy are reported in Table S6.

In our review of model validation metrics, half of the examined studies relied on a single metric, and some even omitted the specification of the metric used, posing a challenge to achieving a comprehensive overview of metrics utilised. This indicates a need for improvement in metric selection and clarity. Aligning with the recommendation of Guisan et al. (2017), best practice involves incorporating multiple validation metrics when assessing model performance to ensure the integration of different aspects of the model's predictive capabilities: i.e., the extent to which a model correctly predicts the conditional probability of presence (calibration) and the ability to distinguish between occupied and unoccupied sites (discrimination). It is interesting to note that our research has shown a greater emphasis on discrimination metrics, such as AUC-ROC, or those derived from two-way contingency tables comparing presence-absence observation to primary predictions (TSS, Cohen's Kappa, Sensitivity, Specificity, Accuracy and Precision). In contrast, there appears to be comparatively less emphasis on calibration metrics like Brier score. Of course, when choosing metrics, it is essential to consider the type of response variable being modelled, whether qualitative, quantitative, or semi-quantitative (as illustrated in 15.1 of Guisan et al., 2017). While Table S7 is not exhaustive in terms of all utilised metrics in research for cetaceans and sea turtles, it can provide valuable assistance in understanding most commonly used metrics.

#### 4.6. Main issues in projecting in time and space

The projection of spatial predictions across different areas and time periods, encompassing both historical and future scenarios, remains a challenging endeavour with few studies having undertaken the process so far. This is despite the significant potential for projection modelling to address research priorities for sea turtles and cetaceans, such as understanding the impact of climate change on the location and extent of their suitable habitats both at sea (Reisinger et al., 2022; Torres et al., 2013) and, for sea turtles, at nesting sites as well (Mancino et al., 2022; Mancino et al., 2023), or addressing other conservation-related issues such as dynamic spatial planning (Becker et al., 2012; Barlow and Torres, 2021). The limited application of the projection modelling arises from the complex nature of the process, requiring careful consideration in both execution and interpretation due to various interconnected issues. Some of these challenges are due to the highly dynamic nature of

cetaceans and sea turtles, as well as their variability in response to environmental changes, which force models deviating from the theoretical assumption of a stable niche over time and species being in equilibrium with climate (Guisan et al., 2017). In reality, marine species may undergo range shifts, adapt to new conditions, colonise new suitable areas or exhibit niche shifts in response to changing environments (e.g., Arcangeli et al., 2023; Mancino et al., 2022), and neglecting to account for these dynamics can result in inaccurate projections. For instance, when projecting into the future, a long-term or historical dataset may more effectively capture the environmental realised niche of a species, potentially resulting in more accurate projections (Guisan et al., 2017) compared to shorter-term data. Also, simplifying the model by including fewer but relevant variables could enhance results compared to incorporating too many variables and related uncertainties. Another significant obstacle to projecting even well-fitted SDMs to future scenarios is the lack of data availability for the most ecologically meaningful predictors (Austin and Van Niel, 2011) at the proper spatial resolution, along with uncertainties associated with projected climate variables (Stoklosa et al., 2015), which may propagate into the projections.

All the gaps emerged from the analysis of reviewed articles along with related opportunities in Table 2.

#### 4.7. Emerging trends in SDMs: insight from most recent literature

Findings from the latest 55 publications on SDMs between January 2023 and May 2024, using the same methodology, confirmed the diversity of models and approaches already identified by our previous

**Table 2**  
Analysis of gaps and opportunities identified at each stage of the modelling process.

Model development	Gaps	Opportunities
1. Conceptualisation	<ul style="list-style-type: none"> <li>Existing studies predominantly concentrate on cetaceans, particularly Delphinidae and Balaenopteridae;</li> <li>Many studies focused on one season only;</li> <li>Few studies using data collected through passive acoustic surveys.</li> </ul>	<ul style="list-style-type: none"> <li>Extend modelling efforts to include less-studied species in less-studied areas;</li> <li>Incorporate additional seasons to include habitat relationships that are not seasonal dependent;</li> <li>Enhance non-invasive methods (i.e., passive acoustic surveys).</li> </ul>
2. Data preparation	<ul style="list-style-type: none"> <li>Lack of consideration for biotic, anthropogenic and water-column variables.</li> </ul>	<ul style="list-style-type: none"> <li>Incorporate a more comprehensive set of variables describing all relevant ecological and anthropic processes, including third dimension of seascape;</li> <li>Creating an expert approved catalogue of environmental datasets.</li> </ul>
3. Model calibration	<ul style="list-style-type: none"> <li>Lack of clarity in generating pseudo-absences/background points.</li> </ul>	<ul style="list-style-type: none"> <li>Provide a summary detailing strategy to appropriately account for absence.</li> </ul>
4. Model validation	<ul style="list-style-type: none"> <li>Prevalence of internal validation strategies;</li> <li>Lack of clarity on how model was validated.</li> </ul>	<ul style="list-style-type: none"> <li>Incorporating external strategies in the validation process using data from different source, including citizen science datasets.</li> </ul>
5. Model results	<ul style="list-style-type: none"> <li>Limited studies projecting spatially and temporally, despite the significance of the projection process for these species.</li> </ul>	<ul style="list-style-type: none"> <li>Improve testing of projection methodologies;</li> <li>Utilize long-term/historical data to project into future scenarios.</li> </ul>

analysis. However, while regression-based models remain prevalent ( $n = 23$ ), there was an increased use of machine learning techniques ( $n = 22$ ), particularly Maxent, and, although limited, deep learning strategies (e.g., neural networks, [Cazau et al., 2023](#)), likely suggesting a shift towards machine learning methodologies ([Maglietta et al., 2023](#)). Another notable trend was the increased incorporation of anthropogenic variables in modelling to assess the influence of human activity on species habitat use, especially related to fishing and boating activities ( $n = 7$  studies; e.g., [Díaz López and Methion, 2024](#)).

Additionally, there was discernible integration of predictive models within marine spatial planning and dynamic ocean management (e.g., [Welch et al., 2024](#)). In particular, dynamic management is an emerging strategy, involving continuously updating the management of marine resources to adapt to changing environmental conditions, wildlife-human interactions, socioeconomic factors, and management priorities. This innovative approach entails identifying dynamic focal marine areas and developing flexible strategies in space and time, positioning as the future of marine resource management. The list of papers can be found as csv file in Appendix C.

## 5. Criticalities and best-practices for future studies

As highlighted by the results from this review, despite regression-based models having been unequivocally the most utilised in the last decades, there is an observable trend towards the increasingly widespread adoption of diverse machine learning approaches. This is attributed to their superior capacity to handle complexity, flexibility, prediction ability and the expanding computational capabilities that now enable the execution of models that were previously considered computationally troublesome. Bayesian statistical approaches are also beginning to be used for their ability to account for uncertainty and facilitate data integration from diverse sources. However, the statistical intricacy still constrains the flexibility and widespread use of these new tools. All these new approaches revealed particularly valuable for species with limited available data and considerable unknowns such as cetaceans and sea turtles.

Despite the proliferation of available methodologies, there is no universally recognised ‘best’ SDM approach. Instead, the most suitable SDM model is the one that can be effectively adapted to the unique characteristics and goals of each study. The choice of model type and the degree of complexity should be carefully considered about the type, quality and quantity of available data (response and predictors variables), the sampling design, the distribution of effort, the temporal and geographical scale, the characteristics of the species being modelled, the complexity of the model depending on the specific research questions, the intended application and the computational resources available ([Arcangeli et al., 2024](#); [Robinson et al., 2017](#)).

Further, the analysis of scientific literature revealed a notable ambiguity in the use of modelling terminology and in the decision-making process at each stage. These decisions seem not to be thoroughly assessed based on the specific case but rather appear to be inspired by previous works, resulting in a replication of modelling choices. Additionally, we have noted that highly prolific research groups tend to replicate these modelling decisions across different cases, indicating a preference for a familiar approach, potentially overlooking considerations more tied to the specific case. Researchers should maintain an awareness of the inherent limitations of their chosen model and not become overly attached to a singular approach. Rather, the ability to critically evaluate and consider alternative methodologies when necessary is paramount, especially in the field of conservation biology. This stresses the importance of enhancing awareness regarding the implications of modelling choices. The modelling landscape remains complex, primarily due to its diversity, with no clear guide for systematic approaches owing to the inherent intricacies of studying the marine environment.

A notable obstacle hindering the advancement of a standardised

modelling framework is that, in most instances, the selection of models is primarily influenced by the characteristics of the available data rather than the specific attributes of the species under investigation or the research objectives. This tendency, as thoroughly discussed, arises from the logistical challenges associated with observing species and collecting data in open seas which determine the scarcity of reliable data, especially species “absence” data, influencing all aspects of model processing and results. Consequently, in the marine environment, it is the type and quality of the available data that determine the selection of the model and the other steps, rather than the other way around, following a convoluted workflow which is necessarily case-specific. Therefore, the availability of presence-only data and the lack of ‘true absences’ can be addressed using modelling methods proficient in handling this type of data, such as machine learning techniques, or selecting/randomly generating pseudo-absences, or incorporating uncertainties associated with data collection within a Bayesian-framework, such as in some PPMs. In contrast, if presence- “true” absence or count/abundance data are available they preferentially may lead to the application of traditional statistical models such as regression and to a lesser extent, machine learning along with a few other approaches. If count data or abundance estimates are based on well-designed dedicated surveys, accounting for perception and availability bias, they could also be used to predict abundances, or if certain assumptions are met, even project them in different areas or periods. Also, the temporal and spatial extent of the available data yields diverse results: when data include sufficient temporal resolution, the choice of models and variables capable of capturing seasonal, monthly, or even daily dynamics becomes viable, thereby improving the detail of the model and providing information on species dynamics. Moreover, the availability of long-term data facilitates the study of species dynamics over time, enabling the formulation of future projections. Finally, the selection of the validation strategy is largely determined by the study’s objectives: internal validation may be suitable for descriptive purposes or preliminary results, while in the context of informing conservation planning, it is highly recommended to independently validate the data using external sources. A condensed overview of the model-building process here described is summarised in [Fig. 11](#).

The effort undertaken within this review represents the initial step towards establishing more effective and standardised methodologies and guidance for modelling the distribution of these highly dynamic species. This standardisation process is particularly relevant within the context of conservation and long-term data collection programs, which consistently gather new data. Finally, it will facilitate us to deepen our understanding of these magnificent and mysterious animals and to apply this knowledge towards their conservation within a dynamic ocean management framework, particularly in response to rapid environmental changes.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2024.102700>.

## Fundings

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## CRediT authorship contribution statement

**E. Pasanisi:** Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **D.S. Pace:** Writing – review & editing, Methodology, Conceptualization. **A. Orasi:** Writing – review & editing. **M. Vitale:** Writing – review & editing. **A. Arcangeli:** Writing – review & editing, Supervision, Methodology, Conceptualization.

## Declaration of competing interest

All authors declare that there are no conflicts of interest.

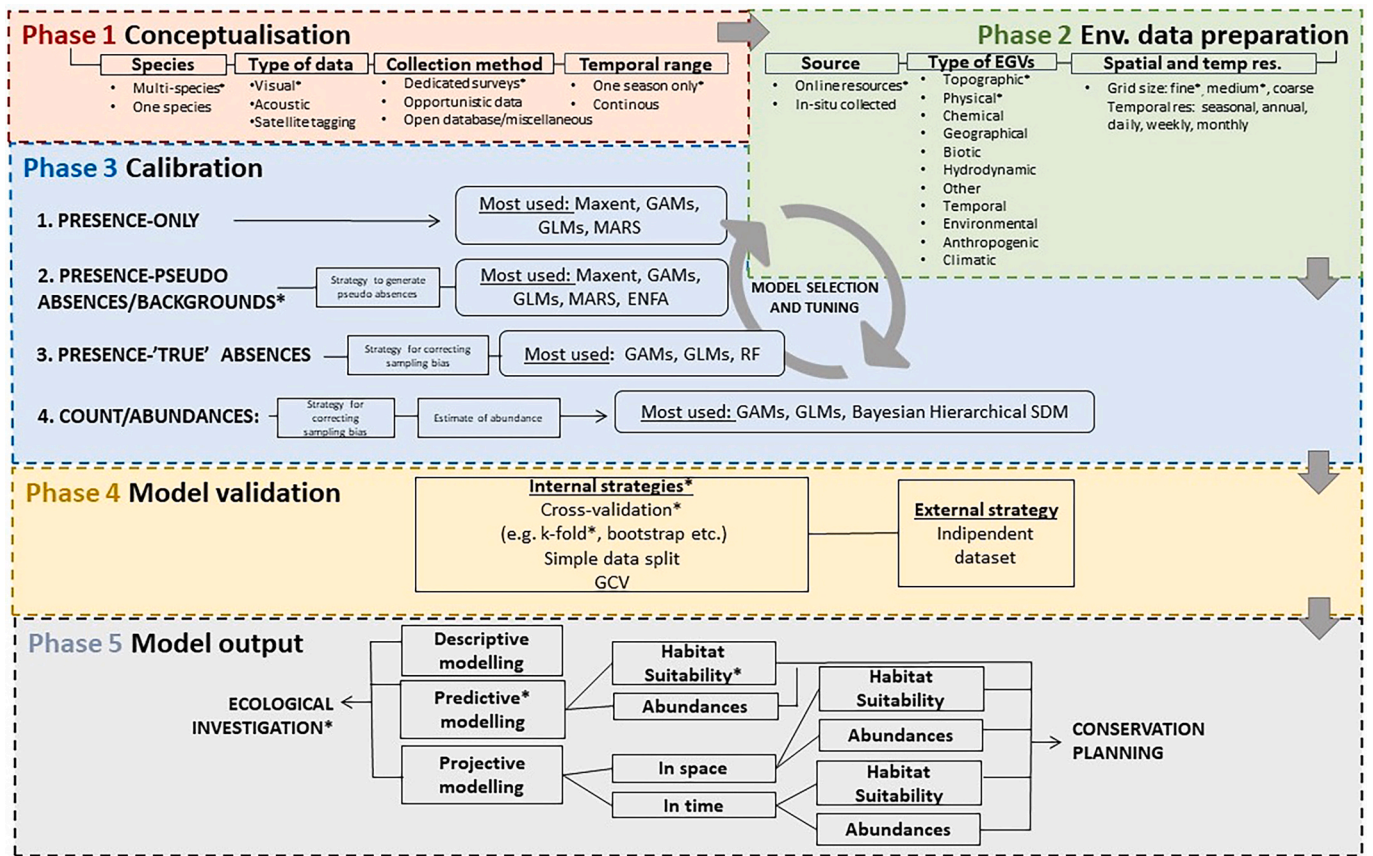


Fig. 11. Condensed overview of the model-building process for cetaceans and sea turtles, outlining its phases and potential options. Emphasis is placed on widely used models categorised by input variable types, with asterisks indicating the most prevalent choices.

## Data availability

Data will be made available on request.

## Acknowledgement

The authors express their gratitude to Dr. Giulia Pedrazzi and Dr. Martina Gregoriotti for their valuable contributions during the screening phase. Additionally, special thanks are extended to Dr. Chiara Mancino for her insightful advice regarding specific sections of the discussion. The authors also thank the editor and the reviewers for their valuable comments, which significantly improved the quality of the manuscript.

## References

- Abrahms, B., Welch, H., Brodie, S., Jacox, M.G., Becker, E.A., Bograd, S.J., Hazen, E.L., 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. *Divers. Distrib.* 25 (8), 1182–1193. <https://doi.org/10.1111/ddi.12940>.
- Aguirre, A.A., Tabor, G.M., 2004. Introduction: marine vertebrates as sentinels of marine ecosystem health. *EcoHealth* 1, 236–238. <https://doi.org/10.1007/s10393-004-0091-9>.
- Anderson, D.R., Burnham, K.P., 2002. Avoiding pitfalls when using information-theoretic methods. *J. Wildl. Manag.* 66, 912–918. <https://doi.org/10.2307/3803155>.
- Araujo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species–climate impact models under climate change. *Glob. Chang. Biol.* 11 (9), 1504–1513. <https://doi.org/10.1111/j.1365-2486.2005.001000.x>.
- Arcangeli, A., Campana, I., Marini, L., MacLeod, C.D., 2016. Long-term presence and habitat use of Cuvier's beaked whale (*Ziphius cavirostris*) in the Central Tyrrhenian Sea. *Mar. Ecol.* 37 (2), 269–282.
- Arcangeli, A., Campana, I., Bologna, M.A., 2017. Influence of seasonality on cetacean diversity, abundance, distribution and habitat use in the western Mediterranean Sea: implications for conservation. *Aquat. Conserv.: Mar. Freshw.* 27 (5), 995–1010. <https://doi.org/10.1002/aqc.2758>.
- Arcangeli, A., Atzori, F., Azzolin, M., Babey, L., Campana, I., Carosso, L., David, L., 2023. Testing indicators for trend assessment of range and habitat of low-density cetacean

- species in the Mediterranean Sea. *Front. Mar. Sci.* 10, 1116829. <https://doi.org/10.1111/maec.12272>.
- Arcangeli, A., Azzolin, M., Babey, L., David, L., Garcia-Garin, O., Moulins, A., Orasi, A., 2024. Looking for reliable species distribution models for low-density cetacean species: compared effectiveness of SDMs for *G. griseus*, *G. melas*, *Z. cavirostris* in the Mediterranean Sea based on long-term fixed-transect data. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 34 (3), e4115 <https://doi.org/10.1002/aqc.4115>.
- Astarloa, A., Louzao, M., Boyra, G., Martinez, U., Rubio, A., Irigoien, Chust, G., 2019. Identifying main interactions in marine predator–prey networks of the Bay of Biscay. *ICES J. Mar. Sci.* 76 (7), 2247–2259. <https://doi.org/10.1093/icesjms/fsz140>.
- Austin, M.P., Van Niel, K.P., 2011. Improving species distribution models for climate change studies: variable selection and scale. *J. Biogeogr.* 38 (1), 1–8. <https://doi.org/10.1111/j.1365-2699.2010.02416.x>.
- Azzolin, M., Saintingan, S., Zampollo, A., Carlucci, R., Maglietta, R., Cipriano, G., Giacomina, C., 2018. Multi source data analysis for improving striped dolphin distribution modelling at a regional scale. In: 2018 IEEE International Workshop on Metrology for the Sea. Learning to Measure Sea Health Parameters, IEEE MetroSea, pp. 145–149. <https://doi.org/10.1109/MetroSea.2018.8657895>.
- Azzolin, M., Arcangeli, A., Cipriano, G., Crosti, R., Maglietta, R., Pietrolungo, G., Carlucci, R., 2020. Spatial distribution modelling of striped dolphin (*Stenella coeruleoalba*) at different geographical scales within the EU Adriatic and Ionian Sea region, Central-Eastern Mediterranean Sea. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 30 (6), 1194–1207. <https://doi.org/10.1002/aqc.3314>.
- Baines, M., Weir, C.R., 2020. Predicting suitable coastal habitat for sei whales, southern right whales and dolphins around the Falkland Islands. *PLoS One* 15 (12), e0244068. <https://doi.org/10.1371/journal.pone.0244068>.
- Bao, M., Wang, X., Liu, W., Chen, H.L., Li, Y., Wu, F., Huang, S.L., 2019. Habitat protection actions for coastal dolphins in a disturbed environment with explicit information gaps. *OCEAN COAST MANAGE.* 169, 147–156. <https://doi.org/10.1016/j.ocecoaman.2018.12.017>.
- Barlow, D.R., Torres, L.G., 2021. Planning ahead: dynamic models forecast blue whale distribution with applications for spatial management. *J. Appl. Ecol.* 58 (11), 2493–2504. <https://doi.org/10.1111/1365-2664.13992>.
- Barlow, D.R., Bernard, K.S., Escobar-Flores, P., Palacios, D.M., Torres, L.G., 2020. Links in the trophic chain: modeling functional relationships between in situ oceanography, krill, and blue whale distribution under different oceanographic regimes. *Mar. Ecol. Prog. Ser.* 642, 207–225. <https://doi.org/10.3354/meps133339>.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Model.* 222 (11), 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>.

- Bearzi, G., Bonizzoni, S., Santostasi, N.L., Furey, N.B., Eddy, L., Valavanis, V.D., Gimenez, O., 2016. Dolphins in a scaled-down Mediterranean: the Gulf of Corinth's odontocetes. *Adv. Mar. Biol.* 75, 297–331. <https://doi.org/10.1016/bs.amb.2016.07.003>.
- Becker, E.A., Foley, D.G., Forney, K.A., Barlow, J., Redfern, J.V., Gentemann, C.L., 2012. Forecasting cetacean abundance patterns to enhance management decisions. *Endanger. Species Res.* 16 (2), 97–112. <https://doi.org/10.3354/esr00390>.
- Becker, E.A., Forney, K.A., Foley, D.G., Smith, R.C., Moore, T.J., Barlow, J., 2014. Predicting seasonal density patterns of California cetaceans based on habitat models. *Endanger. Species Res.* 23 (1), 1–22. <https://doi.org/10.3354/esr00548>.
- Becker, E.A., Forney, K.A., Fiedler, P.C., Barlow, J., Chivers, S.J., Edwards, C.A., Redfern, J.V., 2016. Moving towards dynamic ocean management: how well do modeled ocean products predict species distributions? *Remote Sens.* 8 (2), 149. <https://doi.org/10.3390/rs8020149>.
- Bennington, S., Rayment, W., Dawson, S., 2020. Putting prey into the picture: improvements to species distribution models for bottlenose dolphins in doubtful sound, New Zealand. *Mar. Ecol. Prog. Ser.* 653, 191–204. <https://doi.org/10.3354/meps13492>.
- Benoit-Bird, K.J., Au, W.W.L., 2003. Prey dynamics affect foraging by a pelagic predator (*Stenella longirostris*) over a range of spatial and temporal scales. *Behav. Ecol. Sociobiol.* 53, 364–373. <https://doi.org/10.1007/s00265-003-0585-4>.
- Boisseau, O., Reid, J., Ryan, C., Moscrop, A., McLanaghan, R., Panigada, S., 2024. Acoustic estimates of sperm whale abundance in the Mediterranean Sea as part of the ACCOBAMS survey initiative. *Front. Mar. Sci.* 11, 1164026. <https://doi.org/10.3389/fmars.2024.1164026>.
- Bonizzoni, S., Furey, N.B., Santostasi, N.L., Eddy, L., Valavanis, V.D., Bearzi, G., 2019. Modelling dolphin distribution within an important marine mammal area in Greece to support spatial management planning. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 29 (10), 1665–1680. <https://doi.org/10.1002/aqc.3148>.
- Bonneville, C.D., Derville, S., Luksenburg, J.A., Oremus, M., Garrigue, C., 2021. Social structure, habitat use and injuries of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) reveal isolated, coastal, and threatened communities in the South Pacific. *Front. Mar. Sci.* 8, 52. <https://doi.org/10.3389/fmars.2021.606975>.
- Brottons, L., Thuiller, W., Araújo, M.B., Hirzel, A.H., 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27 (4), 437–448. <https://doi.org/10.1111/j.0906-7590.2004.03764.x>.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford university press, London.
- Carman, V.G., Piola, A., O'Brien, T.D., Tormosov, D.D., Acha, E.M., 2019. Circumpolar frontal systems as potential feeding grounds of Southern Right whales. *Prog. Oceanogr.* 176, 102123. <https://doi.org/10.1016/j.pcean.2019.102123>.
- Cazau, D., Duc, Nguyen Hong, P., Druon, J.N., et al., 2023. Multimodal deep learning for cetacean distribution modeling of fin whales (*Balaenoptera physalus*) in the western Mediterranean Sea. *Mach. Learn.* 112, 2003–2024. <https://doi.org/10.1007/s10994-021-06029-z>.
- CEE, Collaboration for Environmental Evidence. 2022. Guidelines and Standards for Evidence Synthesis in Environmental Management. Version 5.1 (AS Pullin, GK Frampton, B Livoreil & G Petrokofsky) [www.environmentalevidence.org/information-for-authors](http://www.environmentalevidence.org/information-for-authors). [December 2022].
- Chatzimontor, A., Almpandou, V., Doxa, A., Dimitriadis, C., Mazaris, A.D., 2021. Projected redistribution of sea turtle foraging areas reveals important sites for conservation. *Clim. Change Ecol.* 2, 100038. <https://doi.org/10.1016/j.ecochg.2021.100038>.
- Correia, A.M., Gil, Á., Valente, R., Rosso, M., Pierce, G.J., Sousa-Pinto, I., 2019. Distribution and habitat modelling of common dolphins (*Delphinus delphis*) in the eastern North Atlantic. *J. Mar. Biol. Assoc. U. K.* 99 (6), 1443–1457. <https://doi.org/10.1017/S0025315419000249>.
- Cosentino, F., Seamark, E.C.J., Van Cakenberghe, V., Maiorano, L., 2023. Not only climate: the importance of biotic interactions in shaping species distributions at macro scales. *Nat. Ecol. Evol.* 13 (3), e9855. <https://doi.org/10.1002/eece3.9855>.
- Derville, S., Torres, L.G., Iovan, C., Garrigue, C., 2018. Finding the right fit: comparative cetacean distribution models using multiple data sources and statistical approaches. *Divers. Distrib.* 24 (11), 1657–1673. <https://doi.org/10.1111/ddi.12782>.
- Díaz López, B., Methion, S., 2024. Habitat use by Iberian harbour porpoises: ecological and human factors. *Mar. Biol.* 171, 113. <https://doi.org/10.1007/s00227-024-04438-x>.
- DiMatteo, A., Cañadas, A., Roberts, J., Sparks, L., Panigada, S., Boisseau, O., Hochscheid, S., 2022. Basin-wide estimates of loggerhead turtle abundance in the Mediterranean Sea derived from line transect surveys. *Front. Mar. Sci.* 9, 930412. <https://doi.org/10.3389/fmars.2022.930412>.
- Eltou, C.S., 1927. *Animal Ecology*. Macmillan, New York.
- Escobar-Flores, P., O'Driscoll, R.L., Montgomery, J.C., 2013. Acoustic characterization of pelagic fish distribution across the South Pacific Ocean. *Mar. Ecol. Prog. Ser.* 490, 169–183. <https://doi.org/10.3354/meps10435>.
- Estes, J.A., Heithaus, M., McCauley, D.J., Rasher, D.B., Worm, B., 2016. Megafaunal impacts on structure and function of ocean ecosystems. *Annu. Rev. Environ. Resour.* 41, 83–116. <https://doi.org/10.1146/annurev-environ-110615-085622>.
- Fernandez, M., Yesson, C., Gannier, A., Miller, P.I., Azevedo, J.M., 2017. The importance of temporal resolution for niche modelling in dynamic marine environments. *J. Biogeogr.* 44 (12), 2816–2827. <https://doi.org/10.1111/jbi.13080>.
- Fernandez, M., Yesson, C., Gannier, A., Miller, P.I., Azevedo, J.M.N., 2018. A matter of timing: how temporal scale selection influences cetacean ecological niche modelling. *Mar. Ecol. Prog. Ser.* 595, 217–231. <https://doi.org/10.3354/meps12551>.
- Fernandez, M., Sillero, N., Yesson, C., 2022. To be or not to be: the role of absences in niche modelling for highly mobile species in dynamic marine environments. *Ecol. Model.* 471, 110040. <https://doi.org/10.1016/j.ecolmodel.2022.110040>.
- Fordyce, E., Perrin, W.F., 2024. World Cetacea Database. Accessed at <https://www.marinespecies.org/cetacea> on 2024-05-10.
- Forney, K.A., Becker, E.A., Foley, D.G., Barlow, J., Oleson, E.M., 2015. Habitat-based models of cetacean density and distribution in the central North Pacific. *Endanger. Species Res.* 27 (1), 1–20. <https://doi.org/10.3354/esr00632>.
- Fraser, K.E., Garrison, L.P., Soldevilla, M.S., Wiggins, S.M., Hildebrand, J.A., 2021. Cetacean distribution models based on visual and passive acoustic data. *Sci. Rep.* 11 (1), 8240. <https://doi.org/10.1038/s41598-021-87577-1>.
- Friedlaender, A.S., Johnston, D.W., Fraser, W.R., Burns, J., Costa, D.P., 2011. Ecological niche modeling of sympatric killer predators around Marguerite Bay, Western Antarctic Peninsula. *Deep-Sea Res. II Top. Stud. Oceanogr.* 58 (13–16), 1729–1740. <https://doi.org/10.1016/j.dsr2.2010.11.018>.
- García, L.G., Pierce, G.J., Autret, E., Torres-Palenzuela, J.M., 2018. Multi-scale habitat preference analyses for Azorean blue whales. *PLoS One* 13 (9), e0201786. <https://doi.org/10.1371/journal.pone.0201786>.
- García, L.G., Pierce, G.J., Autret, E., Torres-Palenzuela, J.M., 2022. Alongside but separate: sympatric baleen whales choose different habitat conditions in São Miguel, Azores. *Deep-Sea Res. I Oceanogr. Res. Pap.* 184, 103766. <https://doi.org/10.1016/j.dsr.2022.103766>.
- Gregoriotti, M., Atzori, F., Carosso, L., Frau, F., Pellegrino, G., Sara, G., Arcangeli, A., 2021. Cetacean presence and distribution in the Central Mediterranean Sea and potential risks deriving from plastic pollution. *Mar. Pollut. Bull.* 173, 112943. <https://doi.org/10.1016/j.marpolbul.2021.112943>.
- Grinnell, J., 1917. The niche-relationships of the California thrasher. *Auk* 34 (4), 427–433. <https://doi.org/10.2307/4072271>.
- Grose, S.O., Pendleton, L., Leathers, A., Cornish, A., Waitai, S., 2020. Climate change will re-draw the map for marine megafauna and the people who depend on them. *Front. Mar. Sci.* 7. <https://doi.org/10.3389/fmars.2020.00547>.
- Guerra, M., Dawson, S.M., Somerford, T.R., Slooten, E., Rayment, W.J., 2022. Fine-scale habitat use of foraging sperm whales is driven by seafloor topography and water column structure. *Mar. Mamm. Sci.* 38 (2), 626–652. <https://doi.org/10.1111/mms.12881>.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8 (9), 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135 (2–3), 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9).
- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press, Cambridge.
- Ham, G.S., Lahaye, E., Rosso, M., Moulins, A., Hines, E., Tepsich, P., 2021. Predicting summer fin whale distribution in the Pelagos sanctuary (North-Western Mediterranean Sea) to identify dynamic whale–vessel collision risk areas. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 31 (8), 2257–2277. <https://doi.org/10.1002/aqc.3614>.
- Hammond, P.S., Francis, T.B., Heinemann, D., Long, K.J., Moore, J.E., Punt, A.E., Zerbini, A.N., 2021. Estimating the abundance of marine mammal populations. *Front. Mar. Sci.* 8, 1316. <https://doi.org/10.3389/fmars.2021.735770>.
- Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. Springer, New York, pp. 1–758.
- Hazen, E.L., Palacios, D.M., Forney, K.A., Howell, E.A., Becker, E., Hoover, A.L., Bailey, H., 2017. WhaleWatch: a dynamic management tool for predicting blue whale density in the California Current. *J. Appl. Ecol.* 54 (5), 1415–1428. <https://doi.org/10.1111/1365-2664.12820>.
- Ho, Y.W., Lin, T.H., Akamatsu, T., Karczmarski, L., 2024. Fine-scale spatial variability of marine ecotone environment corresponds with habitat utilization of Indo-Pacific humpback dolphins in Hong Kong waters. *Ecol. Indic.* 158, 111228. <https://doi.org/10.1016/j.ecolind.2023.111228>.
- Hunt, T.N., Allen, S.J., Bejder, L., Parra, G.J., 2020. Identifying priority habitat for conservation and management of Australian humpback dolphins within a marine protected area. *Sci. Rep.* 10 (1), 14366. <https://doi.org/10.1038/s41598-020-69863-6>.
- Hutchinson, G.E., 1957. Concluding remarks. In: *Cold Spring Harbor symposia on Quantitative Biology*, vol. 22. Cold Spring Harbor Laboratory Press, New York, pp. 415–427.
- Hyrenbach, K.D., Forney, K.A., Dayton, P.K., 2000. Marine protected areas and ocean basin management. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 10 (6), 437–458.
- Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morisette, J.T., Holcombe, T.R., 2015. Caveats for correlative species distribution modeling. *Ecol. Inform.* 29, 6–15. <https://doi.org/10.1016/j.ecoinf.2015.06.007>.
- Katona, S., Whitehead, H., 1988. Are cetaceans ecologically important? *Oceanogr. Mar. Biol. Annu. Rev.* 26, 553–568.
- Lerebourg, L., Saboul, D., Cléménçon, M., Coquart, J.B., 2023. Prediction of Marathon performance using artificial intelligence. *Int. J. Sports Med.* 44 (05), 352–360. <https://doi.org/10.1055/a-1993-2371>.
- Levin, S.A., 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecol.* 73 (6), 1943–1967. <https://doi.org/10.2307/1941447>.
- Machias, A., Karakassis, I., Somarakis, S., Giannoulaki, M., Papadopoulou, K.N., Smith, C., 2005. The response of demersal fish communities to the presence of fish farms. *Mar. Ecol. Prog. Ser.* 288, 241–250. <https://doi.org/10.3354/meps288241>.
- Maglietta, R., Saccotelli, L., Fanizza, C., et al., 2023. Environmental variables and machine learning models to predict cetacean abundance in the Central-Eastern



- Mediterranean Sea. *Sci. Rep.* 13, 2600. <https://doi.org/10.1038/s41598-023-29681-y>.
- Mancino, C., Canestrelli, D., Maiorano, L., 2022. Going west: range expansion for loggerhead sea turtles in the Mediterranean Sea under climate change. *Glob. Ecol. Conserv.* 38, e02264 <https://doi.org/10.1016/j.gecco.2022.e02264>.
- Mancino, C., Hochscheid, S., Maiorano, L., 2023. Increase of nesting habitat suitability for green turtles in a warming Mediterranean Sea. *Sci. Rep.* 13 (1), 19906. <https://doi.org/10.1038/s41598-023-46958-4>.
- Mann, J. (Ed.), 2000. *Cetacean Societies: Field Studies of Dolphins and Whales*. University of Chicago Press, Chicago.
- Mannocci, L., Laran, S., Monestiez, P., Dorémus, G., Van Canneyt, O., Watremez, P., Ridoux, V., 2014. Predicting top predator habitats in the Southwest Indian Ocean. *Ecography* 37 (3), 261–278. <https://doi.org/10.1111/j.1600-0587.2013.00317.x>.
- Mannocci, L., Boustany, A.M., Roberts, J.J., Palacios, D.M., Dunn, D.C., Halpin, P.N., Winship, A.J., 2017. Temporal resolutions in species distribution models of highly mobile marine animals: recommendations for ecologists and managers. *Divers. Distrib.* 23 (10), 1098–1109. <https://doi.org/10.1111/ddi.12609>.
- Maricato, G., Tardin, R., Lodi, L., Wedekin, L.L., Daura-Jorge, F.G., Maciel, I., Alves, M.A.S., 2022. Identifying suitable areas for common bottlenose dolphin in anthropized waters. *Mar. Biol.* 169 (9), 110. <https://doi.org/10.1007/s00227-022-04095-y>.
- Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., Tyack, P.L., 2013. Estimating animal population density using passive acoustics. *Biol. Rev.* 88 (2), 287–309. <https://doi.org/10.1111/brv.12001>.
- Martino, S., Pace, D.S., Moro, S., Casoli, E., Ventura, D., Frachea, A., Jona Lasinio, G., 2021. Integration of presence-only data from several sources: a case study on dolphins' spatial distribution. *Ecography* 44 (10), 1533–1543. <https://doi.org/10.1111/ecog.05843>.
- Matutini, F., Baudry, J., Pain, G., 2020. How citizen science could improve species distribution models and their independent assessment for conservation. *Nat. Ecol. Evol.* 53, 1–9. <https://doi.org/10.1002/ece3.7210>.
- McCormack, S.A., Trebilco, R., Melbourne-Thomas, J., Blanchard, J.L., Fulton, E.A., Constable, A., 2019. Using stable isotope data to advance marine food web modelling. *Rev. Fish Biol. Fish.* 29, 277–296. <https://doi.org/10.1007/s11160-019-09552-4>.
- Melo-Merino, S.M., Reyes-Bonilla, H., Lira-Noriega, A., 2020. Ecological niche models and species distribution models in marine environments: a literature review and spatial analysis of evidence. *Ecol. Model.* 415, 108837 <https://doi.org/10.1016/j.ecolmodel.2019.108837>.
- Monsarrat, S., Pennino, M.G., Smith, T.D., Reeves, R.R., Meynard, C.N., Kaplan, D.M., Rodrigues, A.S., 2015. Historical summer distribution of the endangered North Atlantic right whale (*Eubalaena glacialis*): a hypothesis based on environmental preferences of a congeneric species. *Divers. Distrib.* 21 (8), 925–937. <https://doi.org/10.1111/ddi.12314>.
- Moore, S.E., 2008. Marine mammals as ecosystem sentinels. *J. Mammal.* 89, 534–540. <https://doi.org/10.1644/07-MAMM-S-312R1.1>.
- Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. ENM eval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.* 5 (11), 1198–1205. <https://doi.org/10.1111/2041-210X.12261>.
- Pace, D.S., Arcangeli, A., Mussi, B., Vivaldi, C., Ledon, C., Lagorio, S., Ardizzone, G., 2018. Habitat suitability modeling in different sperm whale social groups. *J. Wildl. Manag.* 82 (5), 1062–1073. <https://doi.org/10.1002/jwmg.21453>.
- Pace, D.S., Panunzi, G., Arcangeli, A., Moro, S., Jona-Lasinio, G., Martino, S., 2022. Seasonal distribution of an opportunistic apex predator (*Tursiops truncatus*) in marine coastal habitats of the Western Mediterranean Sea. *Front. Mar. Sci.* 9, 939692 <https://doi.org/10.3389/fmars.2022.939692>.
- Palacios, D.M., Baumgartner, M.F., Laidre, K.L., Gregr, E.J., 2013. Beyond correlation: integrating environmentally and behaviourally mediated processes in models of marine mammal distributions. *Endanger. Species Res.* 22 (3), 191–203. <https://doi.org/10.3354/esr00558>.
- Palacios, D.M., Bailey, H., Becker, E.A., Bograd, S.J., DeAngelis, M.L., Forney, K.A., Mate, B.R., 2019. Ecological correlates of blue whale movement behavior and its predictability in the California Current Ecosystem during the summer-fall feeding season. *Movement Ecol.* 7, 1–21. <https://doi.org/10.1186/s40462-019-0164-6>.
- Passadore, C., Möller, L.M., Diaz-Aguirre, F., Parra, G.J., 2018. Modelling dolphin distribution to inform future spatial conservation decisions in a marine protected area. *Sci. Rep.* 8 (1), 15659 <https://doi.org/10.1038/s41598-018-34095-2>.
- Pérez-Jorge, S., Tobaña, M., Prieto, R., Vandepierre, F., Calmettes, B., Lehodey, P., Silva, M.A., 2020. Environmental drivers of large-scale movements of baleen whales in the mid-North Atlantic Ocean. *Divers. Distrib.* 26 (6), 683–698. <https://doi.org/10.1111/ddi.13038>.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190 (3–4), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19 (1), 181–197. <https://doi.org/10.1890/07-2153.1>.
- Pirotta, E., Brotons, J.M., Cerdà, M., Bakkers, S., Rendell, L.E., 2020. Multi-scale analysis reveals changing distribution patterns and the influence of social structure on the habitat use of an endangered marine predator, the sperm whale *Physeter macrocephalus* in the Western Mediterranean Sea. *Deep-Sea Res. I Oceanogr. Res. Pap.* 155, 103169 <https://doi.org/10.1016/j.dsr.2019.103169>.
- Pompa, S., Ehrlich, P.R., Ceballos, G., 2011. Global distribution and conservation of marine mammals. *Proc. Natl. Acad. Sci.* 108 (33), 13600–13605. <https://doi.org/10.1073/pnas.1101525108>.
- Radosavljevic, A., Anderson, R.P., 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. *J. Biogeogr.* 41 (4), 629–643. <https://doi.org/10.1111/jbi.12227>.
- Ramesh, K., Berrow, S., Meade, R., O'Brien, J., 2021. Habitat modelling on the potential impacts of shipping noise on fin whales (*Balaenoptera physalus*) in offshore Irish waters off the porcupine ridge. *J. Mar. Sci. Eng.* 9 (11), 1207. <https://doi.org/10.3390/jmse9111207>.
- Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Werner, F., 2006. Techniques for cetacean-habitat modeling. *Mar. Ecol. Prog. Ser.* 310, 271–295. <https://doi.org/10.3354/meps310271>.
- Redfern, J.V., Barlow, J., Ballance, L.T., Gerrodette, T., Becker, E.A., 2008. Absence of scale dependence in dolphin-habitat models for the eastern tropical Pacific Ocean. *Mar. Ecol. Prog. Ser.* 363, 1–14. <https://doi.org/10.3354/meps07495>.
- Reisinger, R.R., Corney, S., Raymond, B., Lombard, A.T., Bester, M.N., Crawford, R.J., Pistorius, P.A., 2022. Habitat model forecasts suggest potential redistribution of marine predators in the southern Indian Ocean. *Divers. Distrib.* 28 (1), 142–159. <https://doi.org/10.1111/ddi.13447>.
- Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Warton, D.I., 2015. Point process models for presence-only analysis. *Methods Ecol. Evol.* 6 (4), 366–379. <https://doi.org/10.1111/2041-210X.12352>.
- Robinson, Néstor M., Nelson, Wendy A., Costello, Mark J., Sutherland, Judy E., Lundquist, Carolyn J., 2017. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Front. Mar. Sci.* 4 <https://doi.org/10.3389/fmars.2017.00421>.
- Roman, J., Estes, J.A., Morissette, L., Smith, C., Costa, D., McCarthy, J., Smetacek, V., 2014. Whales as marine ecosystem engineers. *Front. Ecol. Environ.* 12 (7), 377–385. <https://doi.org/10.1890/130220>.
- Scales, K.L., Schorr, G.S., Hazen, E.L., Bograd, S.J., Miller, P.I., Andrews, R.D., Falcone, E.A., 2017a. Should I stay or should I go? Modelling year-round habitat suitability and drivers of residency for fin whales in the California current. *Divers. Distrib.* 23 (10), 1204–1215. <https://doi.org/10.1111/ddi.12611>.
- Scales, K.L., Hazen, E.L., Jacox, M.G., Edwards, C.A., Boustany, A.M., Oliver, M.J., Bograd, S.J., 2017b. Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. *Ecography* 40 (1), 210–220. <https://doi.org/10.1111/ecog.02272>.
- Sequeira, A.M., Heupel, M.R., Lea, M.A., Eguíluz, V.M., Duarte, C.M., Meekan, M.G., Hays, G.C., 2019. The importance of sample size in marine megafauna tagging studies. *Ecol. Appl.* 29 (6), e01947 <https://doi.org/10.1002/eap.1947>.
- Širović, A., Hildebrand, J.A., 2011. Using passive acoustics to model blue whale habitat off the Western Antarctic Peninsula. *Deep-Sea Res. II Top. Stud. Oceanogr.* 58 (13–16), 1719–1728. <https://doi.org/10.1016/j.dsr2.2010.08.019>.
- Sprogs, K.R., Christiansen, F., Raudino, H.C., Kobryn, H.T., Wells, R.S., Bejder, L., 2018. Sex-specific differences in the seasonal habitat use of a coastal dolphin population. *Biodivers. Conserv.* 27 (14), 3637–3656. <https://doi.org/10.1007/s10531-018-1618-7>.
- Stoklosa, J., Daly, C., Foster, S.D., Ashcroft, M.B., Warton, D.I., 2015. A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods Ecol. Evol.* 6 (4), 412–423. <https://doi.org/10.1111/2041-210X.12217>.
- Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Burnham, K.P., 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. *J. Appl. Ecol.* 47 (1), 5–14. <https://doi.org/10.1111/j.1365-2664.2009.01737.x>.
- Torres, L.G., Read, A.J., Halpin, P., 2008. Fine-scale habitat modeling of a top marine predator: do prey data improve predictive capacity? *Ecol. Appl.* 18, 1702–1717. <https://doi.org/10.1890/07-1455.1>.
- Torres, L.G., Smith, T.D., Sutton, P., MacDiarmid, A., Bannister, J., Miyashita, T., 2013. From exploitation to conservation: habitat models using whaling data predict distribution patterns and threat exposure of an endangered whale. *Divers. Distrib.* 19 (9), 1138–1152. <https://doi.org/10.1111/ddi.12069>.
- Trew, B.T., Grantham, H.S., Barrientos, C., Collins, T., Doherty, P.D., Formia, A., Metcalfe, K., 2019. Using cumulative impact mapping to prioritize marine conservation efforts in Equatorial Guinea. *Front. Mar. Sci.* 6, 717. <https://doi.org/10.3389/fmars.2019.00717>.
- Usman, A.M., Versfeld, D.J., 2022. Detection of baleen whale species using kernel dynamic mode decomposition-based feature extraction with a hidden Markov model. *Ecol. Inform.* 71, 101766 <https://doi.org/10.1016/j.ecoinf.2022.101766>.
- Usman, A.M., Ogundile, O.O., Versfeld, D.J., 2020. Review of automatic detection and classification techniques for cetacean vocalization. *Ieee Access.* 8, 105181–105206. <https://doi.org/10.1109/ACCESS.2020.3000477>.
- Valsecchi, E., Arcangeli, A., Lombardi, R., Boyse, E., Carr, I.M., Galli, P., Goodman, S.J., 2021. Ferries and environmental DNA: underway sampling from commercial vessels provides new opportunities for systematic genetic surveys of marine biodiversity. *Front. Mar. Sci.* 8, 704786 <https://doi.org/10.3389/fmars.2021.704786>.
- Virgili, A., Authier, M., Monestiez, P., Ridoux, V., 2018. How many sightings to model rare marine species distributions. *PLoS One* 13 (3), e0193231. <https://doi.org/10.1371/journal.pone.0193231>.
- Virgili, A., Teillard, V., Dorémus, G., Dunn, T.E., Laran, S., Lewis, M., Ridoux, V., 2022. Deep ocean drivers better explain habitat preferences of sperm whales *Physeter macrocephalus* than beaked whales in the Bay of Biscay. *Sci. Rep.* 12 (1), 9620 <https://doi.org/10.1038/s41598-022-13546-x>.
- Welch, H., Liu, O.R., Riekkola, L., Abrahms, B., Hazen, E.L., Samhouri, J.F., 2024. Selection of planning unit size in dynamic management strategies to reduce

- human-wildlife conflict. *Conserv. Biol.* e14201 <https://doi.org/10.1111/cobi.14201>.
- Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R., McCarthy, M.A., 2021. Defining and evaluating predictions of joint species distribution models. *Methods Ecol. Evol.* 12 (3), 394–404. <https://doi.org/10.1111/2041-210X.13518>.
- Wisz, M.S., Guisan, A., 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecol.* 9, 1–13. <https://doi.org/10.1186/1472-6785-9-8>.
- Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Svenning, J.C., 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol. Rev.* 88 (1), 15–30. <https://doi.org/10.1111/j.1469-185X.2012.00235.x>.
- Wright, M.K., Pompe, L.R., Mishra, D.R., Baumbach, D.S., Salinas, L., Dunbar, S.G., 2022. Hawksbill presence and habitat suitability of a marine reserve in Honduras. *Ocean Coast. Manag.* 225, 106204 <https://doi.org/10.1016/j.ocecoaman.2022.106204>.
- Zampollo, A., Arcangeli, A., Costantino, M., Mancino, C., Crosti, R., Pietroluongo, G., Azzolin, M., 2022. Seasonal niche and spatial distribution modelling of the loggerhead (*Caretta caretta*) in the Adriatic and Ionian seas. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 32 (7), 1141–1155. <https://doi.org/10.1002/aqc.3815>.
- Zanardo, N., Parra, G.J., Passadore, C., Möller, L.M., 2017. Ensemble modelling of southern Australian bottlenose dolphin *Tursiops* sp. distribution reveals important habitats and their potential ecological function. *Mar. Ecol. Prog. Ser.* 569, 253–266. <https://doi.org/10.3354/meps12091>.
- Zhang, S., Cao, Y., Chen, B., Jiang, P., Fang, L., Li, H., Li, M., 2023. Assessing the potential use of environmental DNA for multifaceted genetic monitoring of cetaceans: example of a wandering whale in a highly disturbed bay area. *Ecol. Indic.* 148, 110125 <https://doi.org/10.1016/j.ecolind.2023.110125>.