A Systematic Review of Marine-Based Species Distribution Models (SDMs) with Recommendations for Best Practice

Néstor M. Robinson[1,2], Wendy A. Nelson[1,2], Mark J. Costello[3], Judy E. Sutherland[2] and Carolyn J. Lundquist[1,4]*

[1] School of Biological Sciences, University of Auckland, Auckland, New Zealand, [2] National Institute of Water and Atmospheric Research, Wellington, New Zealand, [3] Institute of Marine Sciences, University of Auckland, Auckland, New Zealand, [4] National Institute of Water and Atmospheric Research, Hamilton, New Zealand

In the marine environment Species Distribution Models (SDMs) have been used in hundreds of papers for predicting the present and future geographic range and environmental niche of species. We have analyzed ways in which SDMs are being applied to marine species in order to recommend best practice in future studies. This systematic review was registered as a protocol on the Open Science Framework: https://osf.io/tngs6/. The literature reviewed (236 papers) was published between 1992 and July 2016. The number of papers significantly increased through time ($R^2 = 0.92$, $p < 0.05$). The studies were predominantly carried out in the Temperate Northern Atlantic (45%) followed by studies of global scale (11%) and studies in Temperate Australasia (10%). The majority of studies reviewed focused on theoretical ecology (37%) including investigations of biological invasions by non-native organisms, conservation planning (19%), and climate change predictions (17%). Most of the studies were published in ecological, multidisciplinary, or biodiversity conservation journals. Most of the studies (94%) failed to report the amount of uncertainty derived from data deficiencies and model parameters. Best practice recommendations are proposed here to ensure that novice and advanced SDM users can (a) understand the main elements of SDMs, (b) reproduce standard methods and analysis, and (c) identify potential limitations with their data. We suggest that in the future, studies of marine SDMs should report on key features of the approaches employed, data deficiencies, the selection of the best explanatory model, and the approach taken to validate the SDM results. In addition, based on the literature reviewed, we suggest that future marine SDMs should account for uncertainty levels as part of the modeling process.

Keywords: best practice, coastal, distributional patterns, marine, model validation, predictive models, SDMs

INTRODUCTION

Knowing the distributions of species is important for environmental management. However, it is difficult to know where individuals of every species are at any one time, except perhaps for some well-researched, and highly endangered megafauna or rare plants. Thus, Species Distribution models (SDMs) are used to predict a species' geographic and environmental range, typically incorporating both seasonal and temporal variability.

In the field of Species Distribution Modeling (SDM) the concept of niche is a central element in the process of modeling the distribution of a species. Since its early development in the ecological literature (Grinnell, 1917; Elton, 1927; Hutchinson, 1957) the concept has been used as a surrogate to describe the habitat and environment of a given species. However, some authors argued that these concepts have different meaning in the subject of SDMs. For instance, Chase and Leibold (2003) have reviewed the historic context of the niche concept and other studies (Kearney, 2006) have presented an integrated way of using those terms in the context of SDMs. To avoid confusion, in this systematic review we follow the niche concept presented by Chase and Leibold (2003) in which the niche of a species is defined as "the joint description of the zero net growth isocline (ZNGI) of an organism along with the impact vectors on that ZNGI in the multivariate space defined by the set of environmental factors that are present." In SDMs, the correlative model correlates the presence or abundance of a species with spatial habitat data. As such, these models map the probability of occurrence of a species across a landscape. For instance, Do Amaral et al. (2015) modeled the distribution patterns of *Stenella* dolphins across the southwestern Atlantic Ocean using averages of climatic variables such as salinity, chlorophyll *a*, and sea surface temperature. A second approach, known as the mechanistic model, determines the links between a species' environment and its fitness, and then maps the species' fitness consequences onto a landscape. For instance, Melle et al. (2014) estimated the distribution and abundance of a copepod species (*Calanus finmarchicus*) across its North Atlantic habitat. In this mechanistic model, the distribution of the species was estimated based on the demography, dormancy, egg production, and mortality data. The third approach, known as hybrid model, consists of integration of both correlative and mechanistic models. For instance, Fordham et al. (2013) modeled the spatial explicit abundance patterns of commercially harvested blacklip (*Haliotis rubra*) and greenlip abalone (*Haliotis laevigata*) using an ecological niche-modeling approach based on environmental predictors, and a niche-population model that accounted for demographic processes and physiological responses to climate-related factors.

In the marine environment, a variety of SDMs are now being used to address a range of research goals. Most studies have focused on designing conservation strategies, assessing the impacts of climate change, the spread of invasive species, and understanding the relationships between marine organisms and their physico-chemical environment (Báez et al., 2010; Gormley et al., 2015; Cheung et al., 2016). In addition, SDMs have been coupled with other popular methods such as connectivity analysis (Esselman and Allan, 2011) as a method for prioritizing conservation areas.

A previous literature review of marine SDMs (Robinson et al., 2011) considered the conceptual and practical issues related to terrestrial and marine SDMs, and highlighted the relevant challenges for improving marine applications. That literature review focused on discussing the influence of environmental predictors (i.e., dispersal, species interactions, ontogenetic shifts, and aggregation) on model predictions. Other studies aimed to improve performance of SDMs by incorporating a multi-model ensemble approach. For instance, studies conducted recently (Jones et al., 2012; Jones and Cheung, 2015; Anderson et al., 2016; Scales et al., 2016) have demonstrated that assembling multiple SDMs provides a framework to incorporate relevant model features such as uncertainty and agreement levels between model outputs. Some of the issues that are particularly relevant for marine SDMs are a consequence of data deficiencies in spatial or temporal sampling and biological data collection (Robinson et al., 2011; Costello et al., 2015; Chaudhary et al., 2016). For instance, spatio-temporal bias in global satellite-derived sea surface temperature (SST) measurements may result from many factors including variation in cloud density, water vapor and aerosol concentrations, and the lack of *in situ* data used in tuning the SST retrieval algorithms (Zhang et al., 2004). In other cases, spatial sampling biases in occurrence data arise from higher sampling effort in sites closer to the coast and in shallow waters (Robinson et al., 2011). In the context of SDMs the estimation of model error or uncertainty involves quantifying not only model mistakes and faults, but also the statistical concept of variation (Barry and Elith, 2006). Overall, detection of such model errors is important for SDMs, in particular for studies that aim to manage endangered species, assess the potential risk of biological invasion, and estimate the impacts of climate change. However, many studies fail to give sufficient estimations of model error and uncertainty. Ideally, SDMs should be based on standard sampling protocols and rigorous data quality control checks. Instead, many, if not most SDMs use data collected under methods which result in presence-only data (Pacifici et al., 2017). While in many studies, this is a pragmatic solution, building models using these data can violate assumptions of some of the models (Yackulic et al., 2013). Therefore, best practice recommendations for the construction of marine SDMs are necessary. Here we review the ways in which SDMs are being applied to marine species in order to recommend best practice in future studies. Key features that were analyzed from each publication are organized around two main pillars: (1) the general features of the paper, and (2) the parameters used in each SDM. Recommended practices are proposed here to ensure that novice and advanced SDM users can (a) understand the main elements of SDMs, (b) reproduce standard methods and analysis, and (c) identify potential limitations with their data.

METHODS

Literature Search

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement as a guide (Moher et al., 2015); see Supplementary Material S1. The bibliographic search was performed using the SciVerse Scopus (https://www.scopus.com), ISI Web of Science (https://webofknowledge.com), and Google Scholar (https://scholar.google.com) databases. Papers published between 1950 and the cut-off date 26 July 2016 with the terms "Species Distribution Models" and "marine" or "coastal" in the title, keywords, or abstract were included. Gray literature, non-english publications, and papers reporting terrestrial SDMs were excluded from our search. This search strategy resulted in 375 papers. The process of selecting papers to include in our review

started with a screening of abstracts (**Figure 1**). Articles were excluded (18%, $n = 67$) if they: (i) mentioned the keyword SDM for justification or discussion without implementing a SDM as part of the study; and (ii) were purely conceptual and had not included a real case study (i.e., review, perspective, and comment papers). Thus, only the papers that reported applications of SDMs in the marine environment including the intertidal, pelagic, and deep ocean, or those whose content was unclear based on reading the abstract alone were retained for the second step of the analysis. A further 72 papers were excluded upon further investigation as they were only addressing the distributions of coastal terrestrial ($n = 65$) or freshwater ($n = 7$) species. In total, 236 publications were retained for the quantitative analysis. The bibliographic details of all references retained for this study are available in Supplementary Material S2.

Data Analysis

Features and parameters of each study were categorized (**Table 1**) and compiled into a database of marine SDM publications. When

necessary, appendices and supplementary materials were also inspected. In order to characterize the principal elements of marine SDMs (**Table 2**) we integrated the collated information into a step-by-step process for SDM construction. We identified the details that were reported in the 236 reviewed publications and provided a descriptive summary of the essential details that need to be reported in published marine SDMs. We divided the methodological framework into six steps, starting from clarifying the taxa and goals of the research, through to data collection and manipulation, model implementation and calibration, finishing with model validation. Key decision points and feedback loops were identified throughout.

RESULTS

Synthesized Findings
Annual Trend of Publications

The reviewed literature was published from 1992 to July 2016, and included at least one paper per year except for the years 1993
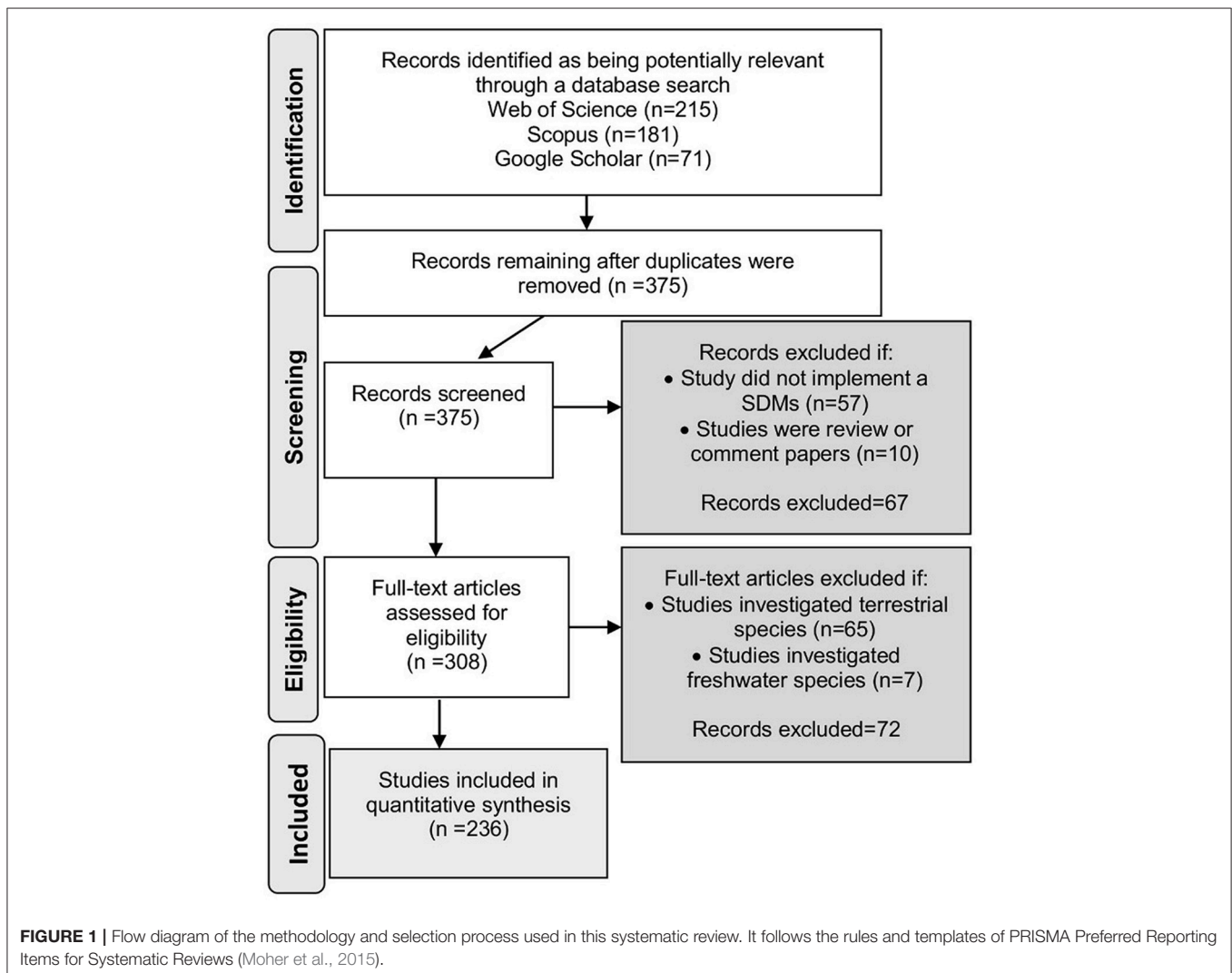

FIGURE 1 | Flow diagram of the methodology and selection process used in this systematic review. It follows the rules and templates of PRISMA Preferred Reporting Items for Systematic Reviews (Moher et al., 2015).

TABLE 1 | Parameters used to categorize marine SDMs included within the database.

Category	Definition
(1) General features of the paper	
Year of publication	Year of publication as it appears in the final print
Reference	Author-year citation style (e.g., Anderson et al., 2016)
Scale of the study	Spatial scale of the study: regional or global
Biogeographic region	Relevant marine biogeographic location of the study
Species	Scientific name of the species included in the analysis
Taxonomic group	Taxonomic group of the species included in the analysis (e.g., fish, corals, macroalgae, etc.)
Research goal	Aim of the study (e.g., conservation planning, impact assessment)
Approach	Type of methodological approach followed: correlative, mechanistic, or hybrid
Journal	Name of the journal were the study was published
Subject area	Research area of the journal (e.g., Ecology, Biodiversity, etc.)
(2) Key features of each SDMs	
Statistical algorithm(s)	Name of the algorithm(s) used in the study (e.g., GLMs, GAMs, Maxent, etc.)
Model selection	Name of the model(s) selection procedures implemented (e.g., cross-validation, stepwise selection)
Model validation	Name of the model(s) validation procedures implemented (e.g., goodness-of-fit, regression)
Uncertainty measure(s)	Name of the uncertainty measures estimated (e.g., model agreement, prediction similarity)

TABLE 2 | Principal elements of marine SDMs.

Category	Definition
Selection of taxa and main study goal	Taxonomic group and main purpose of the study
Data selection	Selection of type of data for SDMs construction (e.g., occurrence data, functional traits)
GIS database manipulation	Downloading data and transforming data layers
Model selection	Implementation of statistical procedures to select the best set of functional predictor variables
Model implementation	Implementation of numerical models (e.g., GAMs, GLMs, BRT, etc.) to predict spatial patterns
Model validation	Implementation of performance measures (e.g., AUC, specificity, etc.) to estimate the validity of model outputs
Error and uncertainty estimation	Quantification of error and uncertainty derived from data deficiencies (e.g., biased samples) and model misspecification (e.g., fitting the model with inadequate response functions)
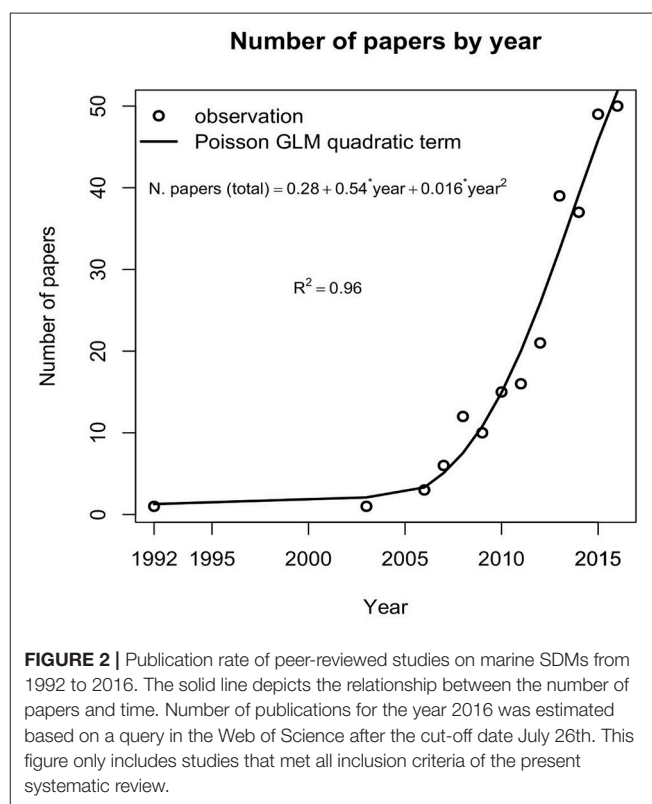
FIGURE 2 | Publication rate of peer-reviewed studies on marine SDMs from 1992 to 2016. The solid line depicts the relationship between the number of papers and time. Number of publications for the year 2016 was estimated based on a query in the Web of Science after the cut-off date July 26th. This figure only includes studies that met all inclusion criteria of the present systematic review.

to 2002. The number of published papers significantly increased through time after the year 2003 (GLM [family = "poisson"]: $R^2 = 0.92$, $p < 0.05$). The use of SDMs in the marine environment has increased since their first implementation in 1992, reaching an annual rate of 49 publications in 2015 (**Figure 2**). The occurrence of only 11 papers before 2008 indicates that the field of marine SDMs is relatively new. From 2008 to 2011 the average rate of publication was 13 papers per year. Thereafter the publication rose to 34 papers per year on average. Correlative models were the most common approach (83%) used by the studies included in our sample, followed by hybrid (9%), and mechanistic models (8%).

Geographic Distribution

The marine SDMs covered all 12 coastal realms (**Figure 3**). A total of 111 publications (47%) focused on the temperate Northern Atlantic, compared to temperate Australasia (9%), temperate Northern Pacific (8%), and the Southern Ocean (6%). Whilst 35% ($n = 82$ papers) of the studies investigated marine global distributions, nationally focused studies were more common in the United States of America, Australia, and the United Kingdom ($n = 29$, 25, and 20 publications respectively). In contrast, fewer studies were located in Spain, Italy, Canada, Portugal, and New Zealand ($n =$ between 4 and 9 papers

each). The remaining studies were located in 25 other countries (Supplementary Material S3).

Discipline

A strong representation of marine SDM publications was observed in the fields of general biology and ecology (as categorized in the ISI Web of Science). Most of the studies were published in ecological, multidisciplinary, and biodiversity conservation journals ($n = 81$, 39, 38 papers respectively;
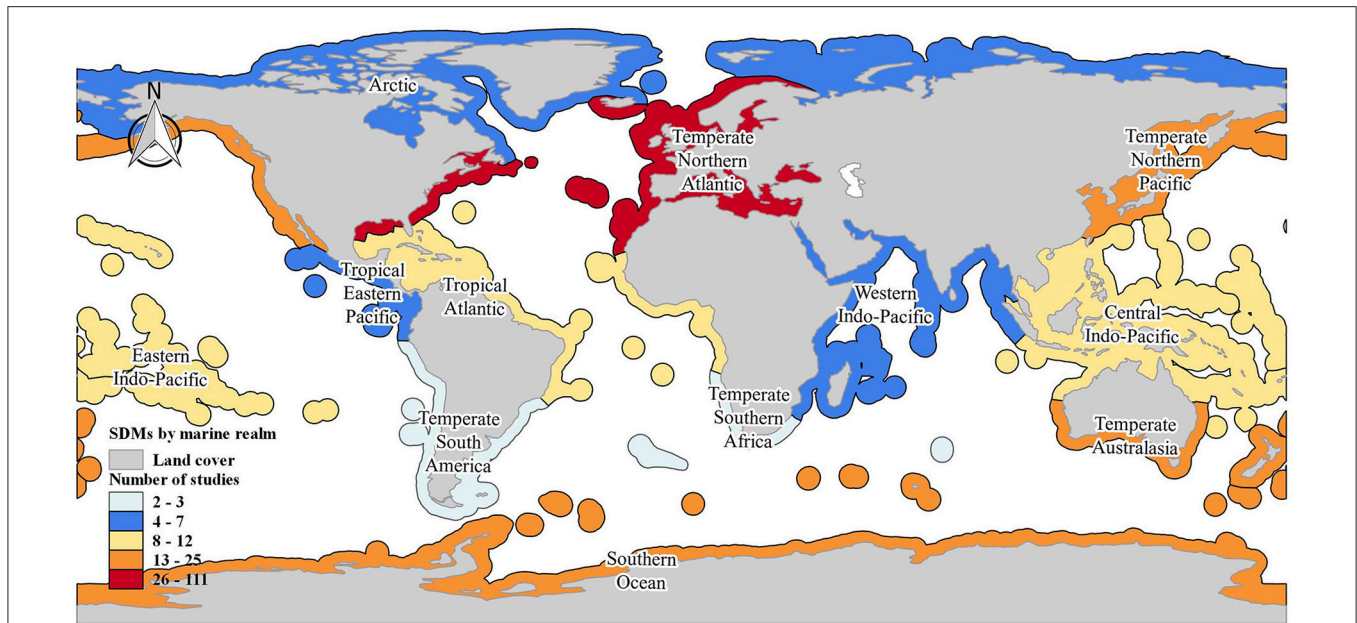
FIGURE 3 | Map illustrating the geographic focus of the marine SDMs studies. The color gradient represents the number of studies conducted per marine realms. The realm classification follows the marine ecoregions of the world (Spalding et al., 2007).

Figure 4). Other disciplines such as paleontology, and oceanography comprised less than 10% of our reviewed literature.

Taxonomic Group

A quarter of the marine SDMs used marine fish distributions as features for model prediction (Supplementary Material S4). Other well-studied taxa were mammals, macroalgae, seabirds, and corals ($n = 27, 26, 22$, and 16 publications respectively). In contrast, marine taxa that have been the focus of a single study were bryozoans, platyhelminthes, protozoans, turtles, and sea urchins. Taxa not represented in any SDM study included marine fungi, viruses, nematodes, and other echinoderms such as asteroids.

Study Goals

In our sample 12 categories of research goals were distinguished (**Figure 5**) Most studies were conducted to answer theoretical questions ($n = 89$), design conservation planning strategies ($n = 46$), estimate climate change impacts ($n = 42$), and evaluate model implementations ($n = 40$). A single publication of each of the following topics was also represented in the literature analyzed: development of modeling approaches, niche evolution, and spatial resolution caveats. In our systematic review, papers that modeled invasive or introduced species were allocated to one of the 12 research goals, and we did not include a specific goal on invasive species. Although the number of publications ($n = 17$ papers) indicates that SDMs are being underutilized for invasive marine species, it is important to mention that biological invasions warrant special attention in the context of SDMs. The first application of SDMs to marine invaders was published in

2006 (Inglis et al., 2006). Since then, other publications presented new tools for forecasting invasion patterns (Johnston and Purkis, 2012) and other research efforts have focused on improving the transferability of SDMs for introduced species (Verbruggen et al., 2013).

Selection Procedures

In our sample, a total of 43 parameter selection procedures were implemented (Supplementary information S5). Cross-validation, step-wise, and multi-model selection procedures were the most popular ($n = 61, 35$, and 31 publications respectively). In contrast, 12% of the model selection procedures were implemented only once (e.g., average predictive comparisons, boosting methods, cross-variograms, and Monte Carlo simulations). We observed that 7% of the publications in our sample had not explicitly implemented a parameter selection procedure.

Model Validation Measures

Model validation involves providing a quantitative measure of model "performance" and/or "accuracy," precision, specificity, and sensitivity. We found that 29 validation methods had been used (Supplementary Material S6). Threshold independent measurement were used in 43% of studies, and cross-validation methods in 12%. Only 9% ($n = 22$) of the studies tested the results of SDMs against independent datasets. Validation methods that were used in only one publication were: probability distributions, residual plots, Spearman rank correlations, and weight average consensus. We found that 12% ($n = 30$) of the studies had not explicitly implemented a validation method to provide a measure of model performance.
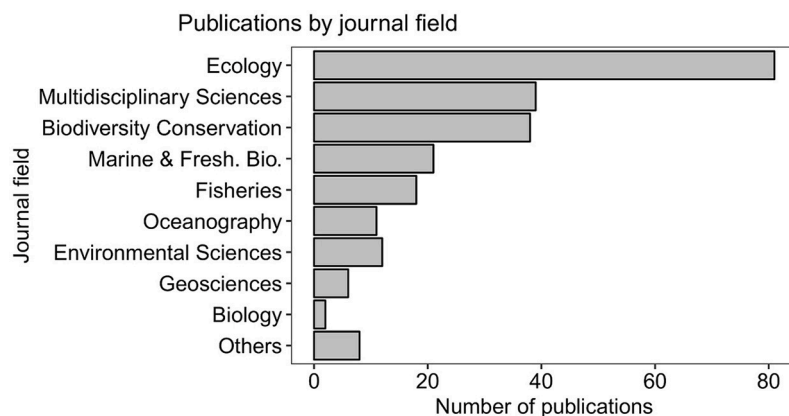
FIGURE 4 | Trends in various disciplines. Proportion of marine SDMs studies (*N* = 236) published within several disciplines using categories as in the Journal of Citation Reports (https://jcr.incites.thomsonreuters.com/). Others category includes the following journal fields: behavioral sciences, biochemistry and molecular biology, operations research, paleontology, parasitology, plant science, remote sensing, and zoology.
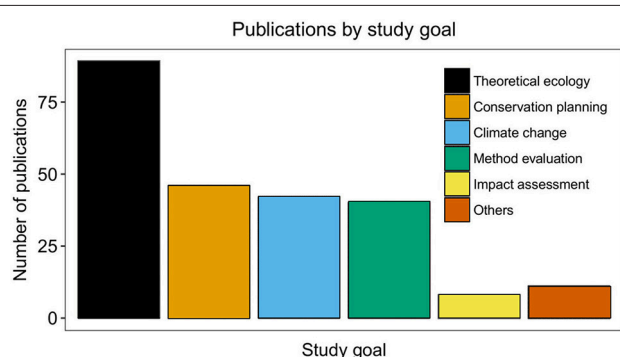

FIGURE 5 | Trends in various research goals. Proportion of marine SDMs studies (*N* = 236) and their main research focus. Others category includes the following study goals: Paleoenvironmental reconstruction, phylogeography, software development, development of modeling approaches, future applications, niche evolution, and spatial resolution.

TABLE 3 | Summary of uncertainty measures reported in marine SDMs.

Category	Frequency
Not reported	224
Multi-model comparisons	6
Prediction similarity	2
Bounding box method	1
Predictive power	1
Multivariate similarity surfaces (MESS) and most dissimilar variables (MoD)	1

Error and Uncertainty Estimation

Only 12 papers (**Table 3**) explicitly took into account some error or uncertainty measures. However, 94% (*n* = 224 papers) of the studies did not model error or uncertainty in modeled distributions.

Association between Categories

The degree of statistical association between the categories was assessed by checking the residuals of chi-squared test (Yates, 1934), whereby residuals >2 or < −2 indicate significant positive or negative associations, respectively (Agresti, 2007). The strongest association observed was between the categories algorithm and invasive status ($X^2 = 90.88$, $p < 0.01$, V = 0.67). Ensemble models and Maxent were common algorithms for invasive species. In addition, a significant though less strong association was found between the categories taxa and study goal ($X^2 = 846.78$, $p < 0.01$, V = 0.47), with a significant over-representation of theoretical ecology studies on fish. Although strong association between some categories was observed, the association between most of the categories was weak. For instance, the association between categories realm and statistical algorithm was $X^2 = 417.16$, $p = 0.68$, V = 0.38 and realm vs. study goal was $X^2 = 145.16$, $p = 0.19$, V = 0.38. Representation of each combination of study goals and year of publication by marine realm can be seen in **Figure 6**. Trellis plot shows overrepresentation of conservation planning, climate change, and theoretical ecology studies in the Temperate Northern Atlantic. In contrast, studies on niche evolution have only been conducted at a global scale.

Framework to Guide Future Applications

We observed that the SDMs process can be divided into six general steps. Commonly, published marine SDMs reported information regarding the following steps. Step 1: SDMs users selected some taxa to study and a main research goal. Step 2: selection of data for model predictions and evaluation of data quality (e.g., accuracy of georeferenced records and sampling bias). Step 3: download and manipulation of environmental layers inspecting important attributes such as layer resolution, layer projection, prediction collinearity, etc. Step 4: selection of a statistical algorithm, accordingly to the data available. Commonly, SDMs relied on a single algorithm (e.g., Maxent, Aquamaps, BIOCLIM, etc.). However, in some cases, results of

model predictions were compared based on ensemble models. Step 5: integration of data for model calibration, including measurements that reflect model fitness (e.g., goodness-of-fit and correlation scores). Step 6: validation of model results, involving mapping the predictions in geographic space and then providing measures that reflect prediction accuracy (e.g., precision, specificity, AUC, etc.). For a description of the terms and acronyms used in this review see Supplementary Materials S7, S8. Individual analysis of each publications included in the review can be seen in Supplementary Material S9.

DISCUSSION

Summary of Main Findings

Previous reviews of SDM applications have (i) discussed challenges including improvements of methods for modeling presence only data (Elith and Leathwick, 2009), (ii) signaled the areas of research that need to be improved in SDMs (Guisan and Thuiller, 2005), and (iii) considered conceptual and practical issues associated with terrestrial SDMs applied to a range of marine organisms (Robinson et al., 2011). Although the aim and extent of previous reviews differ from this study, similar insights have been drawn after conducting the systematic review. Several issues associated with the use of SDMs in the marine environment were identified from this study. First, recent research has documented the effectiveness of multi-model ensemble approaches, e.g., Jones and Cheung (2015) on reducing model uncertainty and providing alternative methods to study geographical patterns at global scale, in contrast to earlier studies using single models. Second, only 6% of the published marine SDMs provided estimates of error or uncertainty measures, as defined by Barry and Elith (2006). Predictions of species distributions are fundamental for marine spatial data planning. However, when resource and conservation strategies are designed based on SDMs, including unacknowledged or unrecognized errors and high levels of uncertainty, these strategies can produce counterproductive effects in the marine environment. Therefore, a critical step in the modeling process is understanding the interplay between the ecological processes that drive the realized species distribution, and the input data used to represent and model it. Our quantitative analysis of the research conducted to date using SDMs in marine contexts has allowed us to identify commonalities across research disciplines and to identify the components of the models that have been frequently overlooked by previous studies. We have developed a framework to guide future applications (**Table 4**, **Figure 7**). This framework also helps to identify the elements of models that warrant more attention in future research.

The methodological framework is divided into six steps, starting from clarifying the taxa and goals of the research, through to data collection and manipulation, model implementation and calibration, finishing with model validation. Key decision points and feedback loops are identified throughout. The framework shows a range of options that are available at the various stages of model construction and implementation, so that the quality of data and environmental layers as well as model fit and prediction accuracy are explicitly addressed.

TABLE 4 | A checklist of model features that need to be explicitly reported in published marine SDMs.

(a) General features that should be explicitly reported in the publication

Scientific name of the species under study validated and taxonomic identification guide used cited

Main research goal of the study

Geographic location of the study

Type of methodological approach followed: correlative, mechanistic, or hybrid

(b) Questions that should be answered regarding data deficiencies

What functional ecological predictors (GIS layers) are missing?

What is the kind and extent of errors in functionally relevant predictors?

Is the sample size adequate?

What is the geographical extent of sampling effort?

How can the lack of absence records be alleviated?

(c) Questions that should be answered regarding model parameters

What is the relationship between the estimated and fitted models?

How does probability of presence vary with the environmental variables?

How do interactions between data and model errors affect model robustness?

How do the selection of the "best explanatory model" vary among different model selection procedures?

(d) Questions that should be answered regarding model validation

What are the results of evaluating model performance with different validation methods?

What are the results of extrapolating model outputs in space and time?

What are the results of testing the model predictions against independent datasets?

Do model results from different algorithms agree or disagree across geographical space?

The selection of taxa in Step 1 will be driven by the study goal. SDMs have been applied on taxa ranging from bacteria through to vertebrates, and from single taxa to multiple taxa. In Step 2 decisions have to be made about what types of data are to be used—what data are available and most directly related to the questions you are asking, are there sufficient data to provide for a robust analysis? At this stage, it is important to evaluate the quality of input data. Previous reviews on SDMs indicate that progress in the field of SDM has been hindered by the lack of linkages between SDMs practice and ecological theory (Elith and Leathwick, 2009). In addition to that, we observed that is not a common practice for SDM studies to report data deficiencies such as inaccuracy of georeferenced records, sampling bias, lack of latitude and longitude data in occurrence records, and mismatch of biological vs. environmental data. Here we recommend that future studies need to acknowledge and report the data deficiencies mentioned above. After selecting the input data for modeling, Step 3 addresses the manipulation of the environmental layers. While there is no general guide for data manipulation, minimum control checks need to be carried out including inspecting the relevance of environmental predictors, layer resolution, layer
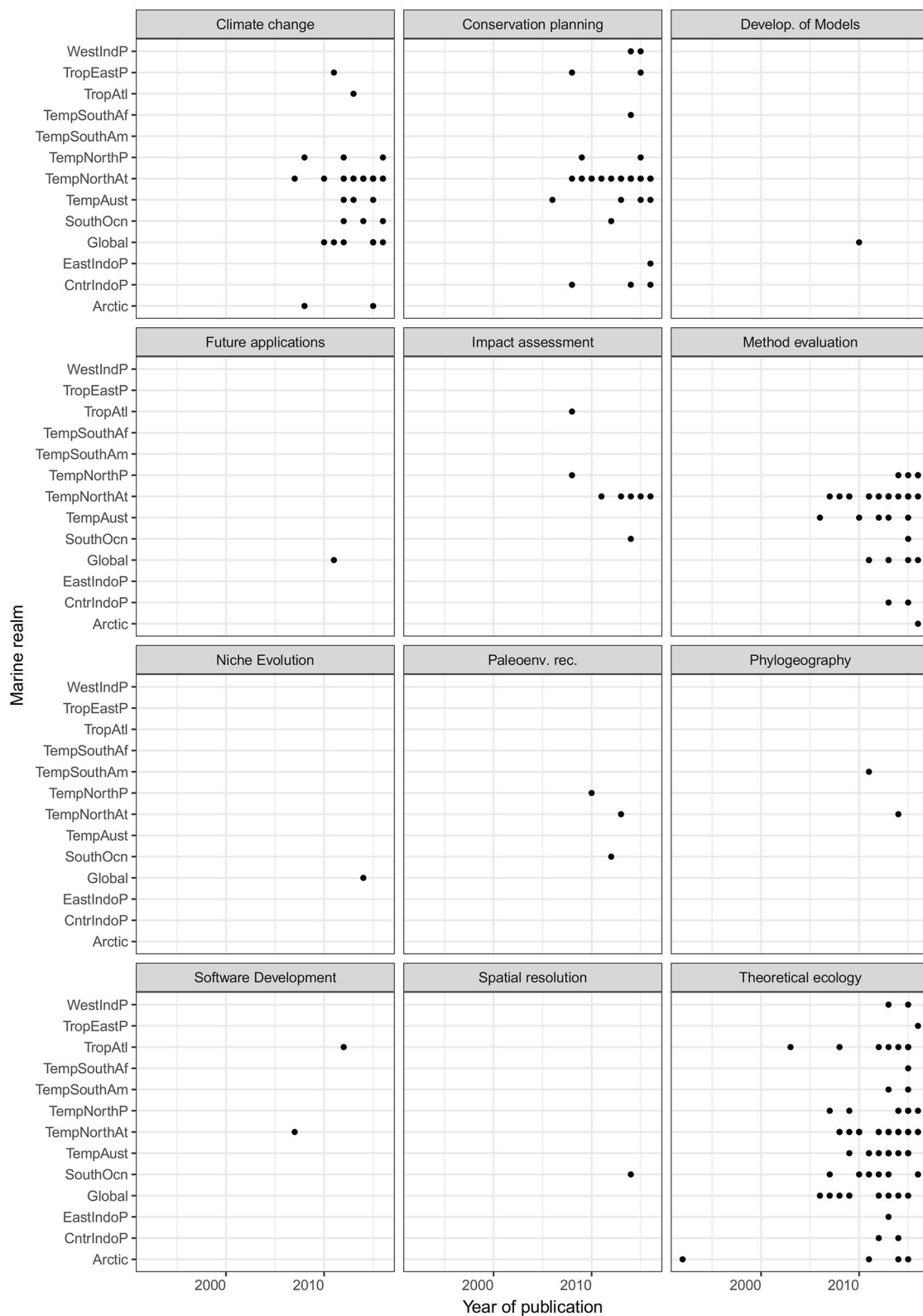
FIGURE 6 | Trellis plot of studies classified by marine realm and each combination of study goal and year of publication. Black dots indicate the occurrence of publication per each combination. Abbreviation codes are as follow. WestIndP, West Indo-Pacific; TropEastP, Tropical Eastern Pacific; TropAtl, Tropical Atlantic; TempSouthAf, Temperate Southern Africa; TempSouthAm, Temperate South America; TempNorthP, Temperate Northern Pacific; TempNorthAt, Temperate Northern Atlantic; TempAust, Temperate Australasia; SouthOcn, Southern Ocean; EastIndoP, Eastern Indo-Pacific; CntrIndoP, Central Indo-Pacific.
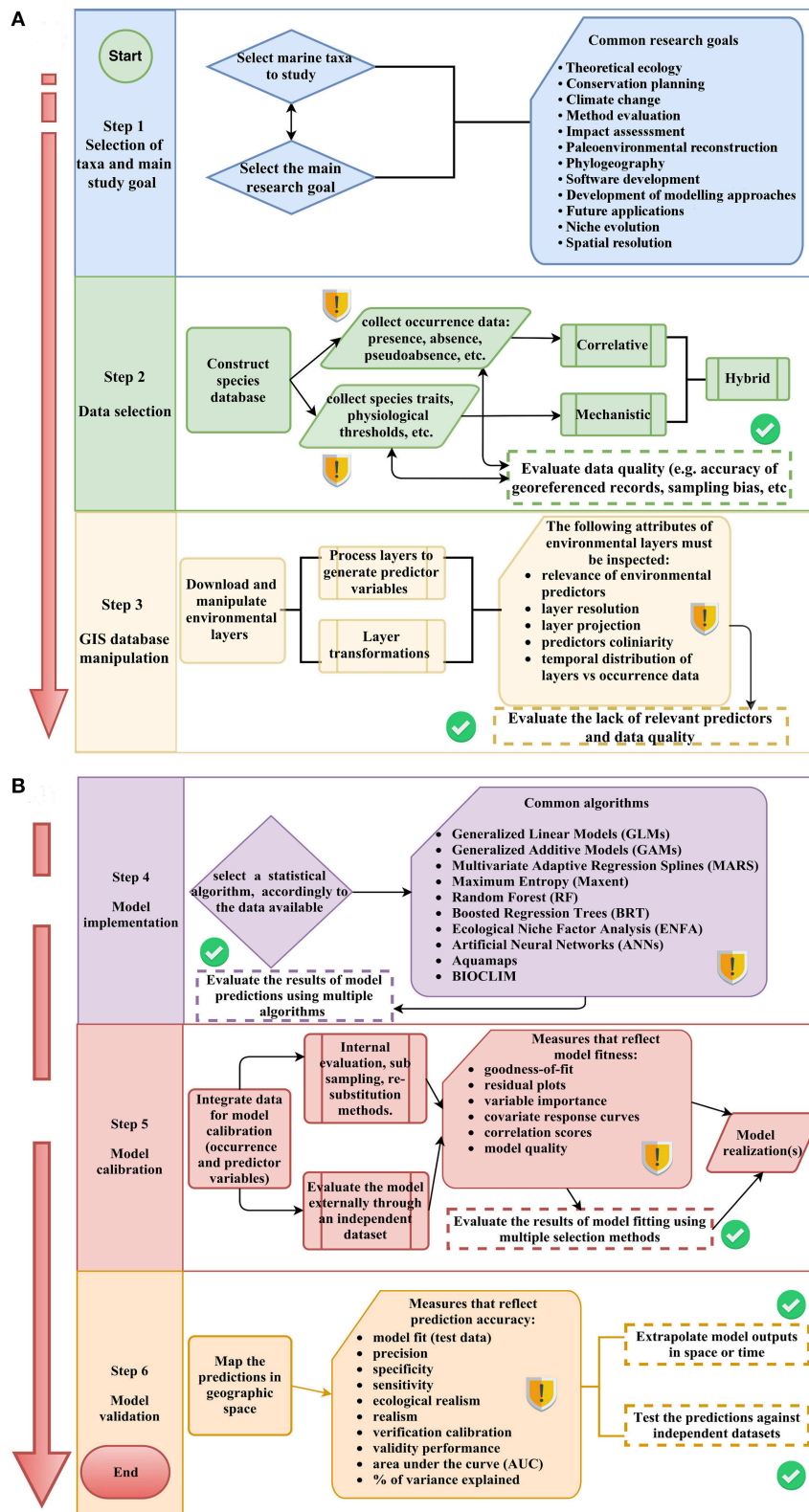
FIGURE 7 | Flow diagram of the marine SDMs construction process. **(A)** Steps in the process involving data collection and manipulation. **(B)** Steps in the process involving model calibration, implementation and validation. Warning signs indicate that the step in the process requires data quality control checks and dashed rectangles give suggestion for assessing model uncertainties and evaluating model performance.

projection, predictor collinearity, and temporal distribution of layers vs. occurrence data. Inspecting and reporting possible data caveats is relevant for avoiding possible bias when building SDMs. In addition, if a relevant environmental proxy is not available (e.g., bottom sea temperature for intertidal species) the lack of data should be acknowledged within the published paper. The selection of the statistical algorithm(s) to be used in the modeling, represented in Step 4, commonly is one of the most difficult decisions in the process of constructing SDMs. Guidance on the most appropriate approach can be drawn from previously published studies on similar datasets, and is likely to require intensive exploration of algorithms that suit the available environmental data. While earlier research tended to focus on single approaches (Swartzman et al., 1992; Wiley et al., 2003), and some studies (Valle et al., 2013; Meißner et al., 2014) suggest that one model performs better than others, we recommend the implementation of multi-model ensemble techniques that account for the degree of similarity and/or variance between model outputs. By employing multiple approaches, it is possible to compare the outputs using different algorithms (Jones and Cheung, 2015).

Step 5 focuses on the calibration of models. In order to find the best explanatory model, published marine SDMs have evaluated the input data using measurements that reflect model fitness (e.g., goodness-of-fit, residual plots, variable importance, and covariate response curves). While these measurements give an estimation of model fitness, evaluating the fit of data with single models can result in selecting a model that poorly represents the real environment. Therefore, we recommend adopting a multi-model selection process that accounts for the degree of uncertainty between model selection results e.g., Pérez-Jorge et al. (2015).

In the framework presented here, at Step 6 the results of the models are validated, that is, outputs examined to see how well the model performs. While many studies validate models using a single or few validation scores (e.g., AUC/ROC curve, p-value, %Dev. Expl., etc.), we recommend the assessment of model accuracy based on multiple validation scores (e.g., Zucchetta et al., 2016). We also recommend extrapolation of model outputs in space or time and testing model results against independent datasets not used as input for the models.

Conducting a systematic review allowed us to identify the model features that need to be explicitly reported in published marine SDMs (**Table 4**) and the elements of models that warrant more attention in future research. The checklist presented in this review provides a series of general features and questions that should be answered regarding aspects of the modeling process. The general features of the papers reviewed here, include information of the study such as the aim, taxonomic group, geographic location, and type of methodological approach implemented (correlative, mechanistic, or hybrid). Whereas, questions about the steps in the modeling process include

questions that should be answered about data deficiencies, ways to select the best explanatory model and ways to validate the results from SDMs. In addition, the flow diagram (**Figure 7**) provided in this review, (i) characterize the steps in the modeling process (Step 1–6), (ii) indicate steps in the process that require data quality control checks, and (iii) provide feedback loops that give suggestions for assessing model uncertainties and evaluating model performances.

CONCLUSIONS

There is worldwide interest in identifying tools for effective marine biodiversity management. SDMs show great potential for application in a wide range of situations. However, the output of the models will be most valuable when estimates of the model performance are provided, enabling increased confidence in the predictions they provide. In order for model results to be compared, e.g., between regions, over time, organism groups, and for replication, the adoption of a consistent framework that includes multi-model approaches and clear expression of errors and uncertainties is critical. The framework we present here is designed to assist future practitioners in the effective application of marine SDMs.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2017.00421/full#supplementary-material

REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* New York, NY: Wiley-Interscience.

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Clark, M. R., Mormede, S., Davies, A. J., et al. (2016). Field validation of habitat suitability models for vulnerable marine ecosystems in the South Pacific Ocean: implications for the use of broad-scale models in fisheries management.

Ocean Coast. Manage. 120, 110–126. doi: 10.1016/j.ocecoaman.2015.11.025

Báez, J., Olivero, J., Peteiro, C., Ferri-Yáñez, F., Garcia-Soto, C., and Real, R. (2010). Macro-environmental modelling of the current distribution of *Undaria pinnatifida* (Laminariales, Ochrophyta) in northern Iberia. *Biol. Invas.* 12, 2131–2139. doi: 10.1007/s10530-009-9614-1

Barry, S., and Elith, J. (2006). Error and uncertainty in habitat models. *J. App. Ecol.* 43, 413–423. doi: 10.1111/j.1365-2664.2006.01136.x

Chase, J. M., and Leibold, M. A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches.* Chicago, IL: University of Chicago Press.

Chaudhary, C., Saeedi, H., and Costello, M. J. (2016). Bimodality of latitudinal gradients in marine species richness. *Trends Ecol. Evol.* 31, 670–676. doi: 10.1016/j.tree.2016.06.001

Cheung, W. W., Jones, M. C., Reygondeau, G., Stock, C. A., Lam, V. W., and Frölicher, T. L. (2016). Structural uncertainty in projecting global fisheries catches under climate change. *Ecol. Modell.* 325, 57–66. doi: 10.1016/j.ecolmodel.2015.12.018

Costello, M. J., Vanhoorne, B., and Appeltans, W. (2015). Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conserv. Biol.* 29, 1094–1099. doi: 10.1111/cobi.12496

Do Amaral, K. B., Alvares, D. J., Heinzelmann, L., Borges-Martins, M., Siciliano, S., and Moreno, I. B. (2015). Ecological niche modeling of *Stenella* dolphins (Cetartiodactyla: Delphinidae) in the Southwestern Atlantic Ocean. *J. Exp. Mar. Biol. Ecol.* 472, 166–179. doi: 10.1016/j.jembe.2015.07.013

Elith, J., and Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697. doi: 10.1146/annurev.ecolsys.110308.120159

Elton, C. S. (1927). *Animal Ecology.* Chicago, IL: University of Chicago Press.

Esselman, P. C., and Allan, J. D. (2011). Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of Northeastern Mesoamerica. *Freshw. Biol.* 56, 71–88. doi: 10.1111/j.1365-2427.2010.02417.x

Fordham, D. A., Mellin, C., Russell, B. D., Akçakaya, R. H., Bradshaw, C. J., Aiello-Lammens, M. E., et al. (2013). Population dynamics can be more important than physiological limits for determining range shifts under climate change. *Glob. Chang. Biol.* 19, 3224–3237. doi: 10.1111/gcb.12289

Gormley, K. S. G., Hull, A. D., Porter, J. S., Bell, M. C., and Sanderson, W. G. (2015). Adaptive management, international co-operation and planning for marine conservation hotspots in a changing climate. *Mar. Policy* 53, 54–66. doi: 10.1016/j.marpol.2014.11.017

Grinnell, J. (1917). The niche-relationships of the California thrasher. *Auk* 34, 427–433 doi: 10.2307/4072271

Guisan, A., and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009. doi: 10.1111/j.1461-0248.2005.00792.x

Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* 22, 415–427.

Inglis, G. J., Hurren, H., Oldman, J., and Haskew, R. (2006). Using habitat suitability index and particle dispersion models for early detection of marine invaders. *Ecol. Appl.* 16, 1377–1390. doi: 10.1890/1051-0761(2006)016[1377:UHSIAP]2.0.CO;2

Johnston, M. W., and Purkis, S. J. (2012). Invasionsoft: a web-enabled tool for invasive species colonization predictions. *Aquat. Invas.* 7, 405–417. doi: 10.3391/ai.2012.7.3.011

Jones, M. C., and Cheung, W. W. L. (2015). Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES J. Mar. Sci.* 72, 741–752. doi: 10.1093/icesjms/fsu172

Jones, M. C., Dye, S. R., Pinnegar, J. K., Warren, R., and Cheung, W. W. L. (2012). Modelling commercial fish distributions: prediction and assessment using different approaches. *Ecol. Modell.* 225, 133–145. doi: 10.1016/j.ecolmodel.2011.11.003

Kearney, M. (2006). Habitat, environment and niche: what are we modelling? *Oikos* 115, 186–191. doi: 10.1111/j.2006.0030-1299.14908.x

Meißner, K., Fiorentino, D., Schnurr, S., Arbizu, P. M., Huettmann, F., Holst, S., et al. (2014). Distribution of benthic marine invertebrates at northern latitudes - an evaluation applying multi-algorithm species distribution models. *J. Sea Res.* 85, 241–254. doi: 10.1016/j.seares.2013.05.007

Melle, W., Runge, J., Head, E., Plourde, S., Castellani, C., Licandro, P., et al. (2014). The North Atlantic Ocean as habitat for *Calanus finmarchicus*: Environmental factors and life history traits. *Prog. Oceanogr.* 129, 244–284. doi: 10.1016/j.pocean.2014.04.026

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4, 1–9. doi: 10.1186/2046-4053-4-1

Pacifici, K., Reich, B. J., Miller, D. A., Gardner, B., Stauffer, G., Singh, S., et al. (2017). Integrating multiple data sources in species distribution modeling: a framework for data fusion. *Ecology* 98, 840–850. doi: 10.1002/ecy.1710

Pérez-Jorge, S., Pereira, T., Corne, C., Wijtten, Z., Omar, M., Katello, J., et al. (2015). Can static habitat protection encompass critical areas for highly mobile marine top predators? insights from coastal East Africa. *PLoS ONE* 10:e0133265. doi: 10.1371/journal.pone.0133265

Robinson, L. M., Elith, J., Hobday, A. J., Pearson, R. G., Kendall, B. E., Possingham, H. P., et al. (2011). Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Glob. Ecol. Biogeogr.* 20, 789–802. doi: 10.1111/j.1466-8238.2010.00636.x

Scales, K. L., Miller, P. I., Ingram, S. N., Hazen, E. L., Bograd, S. J., and Phillips, R. A. (2016). Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. *Divers. Distrib.* 22, 212–224. doi: 10.1111/ddi.12389

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M., et al. (2007). Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* 57, 573–583. doi: 10.1641/B570707

Swartzman, G., Huang, C., and Kaluzny, S. (1992). Spatial analysis of Bering Sea groundfish survey data using generalized additive models. *Can. J. Fish. Aquat. Sci.* 49, 1366–1378. doi: 10.1139/f92-152

Valle, M., van Katwijk, M. M., de Jong, D. J., Bouma, T. J., Schipper, A. M., Chust, G., et al. (2013). Comparing the performance of species distribution models of *Zostera marina*: implications for conservation. *J. Sea Res.* 83, 56–64. doi: 10.1016/j.seares.2013.03.002

Verbruggen, H., Tyberghein, L., Belton, G. S., Mineur, F., Jueterbock, A., Hoarau, G., et al. (2013). Improving transferability of introduced species' distribution models: new tools to forecast the spread of a highly invasive seaweed. *PLoS ONE* 8:e68337. doi: 10.1371/journal.pone.0068337

Wiley, E. O., McNyset, K. M., Peterson, A. T., Robins, C. R., and Stewart, A. M. (2003). Niche modeling perspective on geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography* 16, 120–127. doi: 10.5670/oceanog.2003.42

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., et al. (2013). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4, 236–243. doi: 10.1111/2041-210x.12004

Yates, F. (1934). Contingency tables involving small numbers and the X^2 test. *Suppl. J. R. Stat. Soc.* 1, 217–235. doi: 10.2307/2983604

Zhang, H.-M., Reynolds, R. W., and Smith, T. M. (2004). Bias characteristics in the AVHRR sea surface temperature. *Geophys. Res. Lett.* 31:L01307. doi: 10.1029/2003GL018804

Zucchetta, M., Venier, C., Taji, M., Mangin, A., and Pastres, R. (2016). Modelling the spatial distribution of the seagrass *Posidonia oceanica* along the north African coast: implications for the assessment of good environmental status. *Ecol. Indic.* 61, 1011–1023. doi: 10.1016/j.ecolind.2015.10.059

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.