Received: 6 February 2025 | Revised: 28 May 2025 | Accepted: 12 July 2025

DOI: 10.1111/cobi.70154

Conservation Biology

CONTRIBUTED PAPER

Integrating expert range maps and opportunistic occurrence records of marine fish species in range estimates

Zhixin Zhang[1,2,3] | Ákos Bede-Fazekas[4,5] | Jorge García Molinos[6] | Stefano Mammola[7,8,9] | Jamie M. Kass[10,11] | Junmei Qu[1,2,3] | Julian Oeser[12] | Songxi Yuan[1,2,3] | Chongliang Zhang[13] | Jiqi Gu[14] | Liuyong Ding[15] | Qiang Lin[1,2,3]

[1]State Key Laboratory of Tropical Oceanography, Guangdong Provincial Key Laboratory of Applied Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]Marine Biodiversity and Ecological Evolution Research Center, South China Sea Institute of Oceanology, Guangzhou, China

[4]HUN-REN Centre for Ecological Research, Institute of Ecology and Botany, Vácrátót, Hungary

[5]Institute of Geography and Earth Sciences, Department of Environmental and Landscape Geography, ELTE Eötvös Loránd University, Budapest, Hungary

[6]Arctic Research Center, Hokkaido University, Sapporo, Japan

[7]Molecular Ecology Group (MEG), Water Research Institute (IRSA), National Research Council of Italy (CNR), Verbania Pallanza, Italy

[8]Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

[9]NBFC, National Biodiversity Future Center, Palermo, Italy

[10]Graduate School of Life Sciences, Tohoku University, Sendai, Japan

[11]Advanced Institute for Marine Ecosystem Change (WPI-AIMEC), Tohoku University, Sendai, Japan

[12]Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany

[13]College of Fisheries, Ocean University of China, Qingdao, China

[14]State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Life Sciences of Beijing Normal University, Beijing, China

[15]Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture (CAS), Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

Correspondence
Zhixin Zhang and Qiang Lin, State Key Laboratory of Tropical Oceanography, Guangdong Provincial Key Laboratory of Applied Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China.
Email: zxzhang@scsio.ac.cn; linqiang@scsio.ac.cn

Article impact statement: Species distribution model predictions can be improved by integrating expert range maps and opportunistic occurrence records.

Funding information
National Key Research and Development Program of China, Grant/Award Number: 2023YFC3108800; National Natural Science Foundation of China, Grant/Award Numbers: 42230409, 42425004, 42276126, 42306133, 32170535; János Bolyai Research Scholarship of the Hungarian Academy of Sciences; National Multidisciplinary Laboratory for Climate Change, Grant/Award Numbers: NKFIH-471-3/2021, RRF-2.3.1-21-2022-00014; Basic and Applied Basic

Abstract

Species distribution models (SDMs) are commonly used to estimate species' geographic distributions to inform biodiversity assessments and conservation planning. However, despite their growing popularity, range predictions of SDMs are affected by biases in opportunistic occurrence records and the lack of information on range limits. Integration of expert range maps in SDMs could help, but this strategy is still rarely used, especially for marine species. We built SDMs for 196 marine fish species with global distributions of Epinephelidae and Syngnathidae, 4 modeling algorithms, and opportunistic occurrence data. We then developed 2 types of SDM ensembles (i.e., combined predictions of multiple individual SDMs): with and without integration of expert range maps. We quantified the level of dissimilarity in range estimates between the 2 ensembles and explored the effects of taxonomic identity, geographic attributes, and conservation status on dissimilarity in model predictions. Although both types of ensembles had good predictive performance, ensembles informed by expert range maps avoided overpredictions of ranges past geographical barriers. Moreover, the dissimilarity between predictions of the 2 ensembles depended on multiple factors, including the number and extent of opportunistic occurrences, distance of occurrences to the expert range polygons, and fish family. Based on our findings, we recommend that researchers combine complementary information provided by expert range maps and opportunistic occurrences when predicting marine species distributions with SDMs.

Research Foundation of Guangdong Province, Grant/Award Number: 2024A1515010604; development fund of South China Sea Institute of Oceanology of the Chinese Academy of Sciences, Grant/Award Number: SCSIO202208

KEYWORDS
data integration, dispersal barrier, expert knowledge, species distribution model, stacked generalization

INTRODUCTION

To effectively assess and preserve biodiversity, a better understanding of species' geographic distributions is needed. Species distribution models (SDMs) associate species occurrence data with environmental variables (Elith & Leathwick, 2009; Feng et al., 2019; Guisan et al., 2017) to predict species' potential distributions and represent one of the most widely used tools in biodiversity assessments (Araújo et al., 2019) and conservation planning (Velazco et al., 2020). Researchers have proposed multiple best-practice guidelines for SDMs (e.g., Araújo et al., 2019; Feng et al., 2019; Zurell et al., 2020), and SDMs have been applied successfully in terrestrial (e.g., Thuiller et al., 2019), freshwater (e.g., Domisch et al., 2016), and marine realms (e.g., Melo-Merino et al., 2020). Despite these outstanding advances in this field, researchers highlight challenges in predicting species' geographic ranges, including sampling biases underlying species occurrence records, which hinder the ability to capture accurately species–environment associations (e.g., Hughes et al., 2021), as well as lack of consideration of dispersal capacity (e.g., Shipley et al., 2022) and biotic interactions (Wisz et al., 2013), which shapes species' range patterns. Therefore, there is a clear and urgent need to integrate multiple sources of distribution data to improve species' range predictions and thus better inform conservation practices.

With respect to modeling species' ranges across large spatial extents, 2 types of distribution data are typically used in SDM studies that have their advantages and drawbacks: opportunistic occurrence records (e.g., Assis et al., 2024; Qu et al., 2023) and expert range maps (e.g., Peng et al., 2023; Thuiller et al., 2019). Opportunistic occurrence records reflect the fine-scale distribution of species as distinct points, but they are prone to sampling biases (e.g., concentration of records in easily accessible regions and partial coverage of the entire geographic range of a species) (Hughes et al., 2021; Merow et al., 2017) and other errors (e.g., taxonomic misidentification and invalid coordinates) (Zizka et al., 2019). Expert range maps use a variety of data sources (including opportunistic occurrences) to delineate the known boundaries of species' geographic ranges as polygons based on expert knowledge (Xiao et al., 2025; Zhang et al., 2025). However, these maps describe species distributions with limited spatial accuracy and resolution (Hurlbert & Jetz, 2007) and can include areas both suitable and unsuitable within species' ranges, potentially including substantial areas of false presence (Mainali et al., 2020). Expert range maps are used to build SDMs by sampling occurrence records from them randomly (e.g., Sales et al., 2019) or systematically (e.g., Fourcade, 2016; Thuiller et al., 2019). Occurrence data derived from expert range maps do not reflect true species presences and may be, to some extent, sampled from unsuitable areas, which can result in biased species–environment associations. Additionally, mismatches in spatial resolution can occur between coarse-scale expert range maps and fine-scale environmental data, leading to spurious conclusions (Hurlbert & Jetz, 2007). Given their respective strengths and shortfalls, SDMs based on either of these 2 types of distributional data may produce biased predictions. To improve the utility of SDMs for applied conservation, complementary information from both data sources should be integrated into SDMs (Isaac et al., 2020; Jetz et al. , 2012).

Researchers have proposed different strategies to integrate expert range maps and opportunistic occurrences to improve species' range estimates (e.g., Domisch et al., 2016; Merow et al., 2017, 2022; Oeser et al., 2024). For instance, Merow et al. (2022) improved range estimates by refining model predictions with expert-derived filters, such as habitat types, physiological limits, and biotic interactions. In addition to post hoc masking of model predictions, researchers also developed novel model-based strategies that incorporate expert range maps into models trained on opportunistic occurrences (e.g., Domisch et al., 2016; Merow et al., 2017; Oeser et al., 2024). For instance, Domisch et al. (2016) applied a distance decay function to expert range maps (i.e., the farther away from the map border, the lower the weight) for 3 freshwater fish species in North America and included this variable in SDMs, which resulted in improved model performance and refined distribution predictions. More recently, Oeser et al. (2024) integrated expert range maps into SDMs based on a metalearner regression model that combines the distributional predictions of different algorithms with expert range maps. These efforts combine the complementary information provided by expert range maps and opportunistic occurrences to improve the reliability of species' range estimates.

Despite recent progress with terrestrial systems, the integration of expert range maps and occurrence records in the marine realm remains underexplored. There are many fundamental differences that preclude the transferability of results and conclusions from terrestrial studies to marine systems, including the distinctiveness of 3-dimensional marine habitats, biodiversity patterns, taxonomic knowledge gaps, food web structures, biogeographical features (e.g., the role of barriers), and ecosystem processes (e.g., Webb, 2012; Wiens, 2023). Given this, there is a clear need to investigate the effects of integrating expert range maps on range estimates of marine species.

We integrated expert range maps and opportunistic occurrence records with SDMs to estimate the ranges of 196 globally distributed marine fish species (Appendix S1) that play important ecological roles, are of high economic value, and have abundant distribution data available in the form of opportunistic occurrence records and expert range maps. We developed ensemble models with and without the integration of expert range maps, compared dissimilarity in model predictions between the 2 types of models, and explored the drivers of this dissimilarity. Our aim was to explore the effects of data

integration on range estimates of marine fish species to better guide conservation planning.

METHODS

Input data

We retrieved opportunistic occurrence records of Epinephelidae (groupers) and Syngnathidae (seahorses, seadragons, pipehorses, and pipefishes) from Zhang et al. (2025), who compiled a distribution dataset for 364 fish species belonging to these families from different sources, including field surveys and biodiversity databases (Global Biodiversity Information Facility, Ocean Biogeographic Information System, iNaturalist, the Atlas of Living Australia, Reef Life Survey, Barcode of Life Database, and Integrated Digitized Biocollections). We excluded invalid records (e.g., records on land) and removed clustered occurrences by keeping one occurrence per 3-arcminute (~5.5 km at the equator) grid cell corresponding to the spatial resolution of our marine data layers. We retained for analysis 196 fish species (Appendix S2) that had at least 30 occurrence records (Wisz et al., 2008). We used 57,432 records in our study (Appendices S1 & S2). We obtained species-specific expert range maps from the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (https://www.iucnredlist.org, accessed 26 February 2024) and retained only polygons with high confidence of occurrence (extant category). We generated a distance layer by assigning cells inside the expert range a value of 0 and cells outside a value equal to their great-circle distance to the nearest polygon border.

Epinephelidae and Syngnathidae inhabit benthic coastal waters, so we used marine benthic data layers from the Bio-ORACLE database 3.0 (Assis et al., 2024) at a spatial resolution of 3 arcminutes to provide sufficient geographical detail without inflating computational demand. Different marine environmental factors, including temperature, salinity, ocean currents, pH, and oxygen, have critical effects on the distribution and physiology of marine fish species (e.g., Miller, 2009). Moreover, bathymetric factors, such as seabed ruggedness, have been considered when mapping spatial distributions of marine species (Champion & Coleman, 2021). We initially considered 22 marine predictors (details in Appendix S3), which we subsequently reduced to 10 after excluding predictors with high collinearity (Pearson's $|r| > 0.7$) (Dormann et al., 2013), and expert knowledge on their ecological relevance for fish species (e.g., Bosch et al., 2018). These variables were bathymetric depth, ruggedness index, annual minimum pH, annual minimum dissolved molecular oxygen, annual range of dissolved molecular oxygen, annual minimum salinity, annual maximum temperature, annual maximum current velocity, annual minimum current velocity, and annual mean current velocity (Appendix S3).

SDM development

We developed separate SDMs for each of the studied marine fish species via 4 modeling algorithms: generalized linear models (GLMs), generalized additive models (GAMs), maximum entropy (Maxent), and random forest. We calibrated SDMs within a 1000-km buffer around the minimum convex polygon of the occurrence records for each species (Huang et al., 2025; Waldock et al., 2022; Zhang et al., 2024). We randomly selected 10,000 grid cells as pseudoabsences or background data. Following best-practice recommendations for SDM studies, we tuned parameter settings for each modeling algorithm. Following Brun et al. (2020), we fitted models with intermediate parameterization complexity for GLMs (intercept, linear terms, and second order polynomials) and GAMs (3 knots with thin plate regression splines). With Maxent and random forest, we first fitted candidate models with different parameter combinations (feature class and regularization multiplier for Maxent; tree number [number of predictors evaluated at each split or mtry] and minimum node size for random forest) (details in Appendix S4). We then selected the best model based on the omission rate and area under the receiving operating characteristic curve (AUC) calculated on the validation data (Radosavljevic & Anderson, 2014; Zhang et al., 2021). From the set of candidate models, we retained 10% of the models with the lowest training omission rate (10th percentile) and then selected the best-performing model with the highest validation AUC value from that subset (Radosavljevic & Anderson, 2014; Zhang et al., 2021). When fitting random forest models, following the recommendation by Valavi et al. (2021), we handled the class imbalance between occurrence and background data with a downsampling approach in which the same number of presence and background observations was used in each tree of the forest. We fitted GLMs with the stats package 4.1.2, GAMs with the mgcv package 1.8.38 (Wood, 2011), Maxent 3.4.3 tuned with the ENMeval package 2.0.5 (Kass et al., 2021), and random forest with the randomForest package 4.6.14 (Liaw & Wiener, 2002).

For each species, we quantified predictive performance of optimal models via a 4-fold spatial cross-validation approach. In short, we divided the study area into 4 spatial folds with the get.block function in ENMeval. Folds were divided by latitude and longitude lines, and each contained approximately the same number of occurrence records. We used 3 spatial folds for model training and the remaining spatial fold for model validation. This step was repeated 4 times until all spatial folds were used for model validation. Performance metrics were averaged among folds. We transformed the continuous model predictions into binary predictions (i.e., suitable or unsuitable) with the 10% omission threshold (i.e., the 10th percentile suitability value of species occurrences) (Huang et al., 2025; Oeser et al., 2025). We calculated species richness by summing the binary distribution maps per grid cell. We reported model details following the ODMAP metadata protocol (Zurell et al., 2020) (Appendix S4). All analyses were conducted in R (R Core Team, 2025).

Stacked generalization

Following the model-based data integration approach proposed by Oeser et al. (2024), we integrated expert range maps into SDMs with stacked generalization, a model ensemble method that uses model predictions at one level as the inputs for a metalearner at a second level to improve model predictive performance (Naimi & Balzer, 2018; Wolpert, 1992). This ensemble approach has been widely applied in machine learning but has received relatively little attention in SDM studies (Oeser et al., 2024). The metalearner helps identify an optimal weighting of input algorithms, which can be advantageous compared with simpler approaches (e.g., weighting models based on AUC). In addition, the approach can be used to integrate other sources of distribution information, such as expert range maps.

Following Oeser et al. (2024), we developed 2 ensemble models that used stacked generalization: an SDM ensemble (metalearner predictors are predictions of the 4 SDM algorithms) and an expert-informed ensemble (same as the previous predictors but with the addition of a gridded surface with cell values equal to the distance to the expert range polygon) (Figure 1). Distance to the polygon is by definition 0 for points inside the polygon. To train metalearner predictors, we used out-of-sample predictions from cross-validation results for the SDM algorithms to avoid overfitting (super learner approach [Naimi & Balzer, 2018]). To fit logistic regression metalearners with GLMs, we used a bias reduction method in the brglm2 0.9 package (Kosmidis, 2023). Because we did not have true species absence data, traditional discrimination metrics, such as AUC and the true skill statistic, may be misleading as absolute measures of model performance, although we used AUC for relative comparisons during model selection to break ties in omission rate (Leroy et al., 2018). Therefore, we measured the predictive performance of SDMs with 3 other measures: continuous Boyce index (Hirzel et al., 2006), a model calibration metric designed for presence–background models (Di Cola et al., 2017); confidence in the predicted presences (Somodi et al., 2024), a novel measure focusing on the positive certain predictions distinguished from uncertain predictions; and consistency of predicted presences (Somodi et al., 2024), a novel transferability measure used to compare the confidence in the training and validation subsets.

Consistency measures how much the performance of the model decreases when switching from the training subset to the validation subset; a large decrease suggests overfitting, and a small decrease suggests a transferable model (Randin et al., 2006; Warren et al., 2020). We computed the continuous Boyce index with the ecospat package 3.2 (Di Cola et al., 2017) and calculated confidence and consistency metrics with the confcons package 0.3.1 (Somodi et al., 2024). We measured predictive performance of the 2 ensemble models with the same 4-fold spatial cross-validation approach described earlier. To estimate the contribution of integrating expert range maps in SDMs, we quantified the permutation importance of distance to the expert range map in the expert-informed ensemble models.

We hypothesized that expert-informed ensemble models have large effects on range estimates for species with ranges that are strongly constrained by dispersal barriers. We illustrated this with the Amazon–Orinoco Plume in the western Atlantic Ocean as an example. The huge freshwater discharge of the Amazon and Orinoco Rivers into the tropical Atlantic Ocean results in strong environmental gradients in temperature and salinity, among other variables, representing a significant biogeographic barrier to marine species dispersal (Giachini Tosetto et al., 2022; Luiz et al., 2012). We focused on 3 species with different distributional patterns around this oceanic barrier: yellowedge grouper (*Hyporthodus flavolimbatus*) and yellowmouth grouper (*Mycteroperca interstitialis*), which occur on both sides of the Amazon–Orinoco Plume, and the shortfin pipefish (*Cosmocampus elucens*), which occurs only on its northern side.

Drivers of dissimilarity in model predictions

We measured the dissimilarity level between the 2 types of ensembles with and without consideration of expert range maps and explored its potential drivers. A high degree of dissimilarity indicated that integrating expert information on range limits had a strong effect on the predicted species distribution. This should particularly be the case for species with ranges strongly constrained by non-abiotic factors (i.e., dispersal barriers or biotic interactions) that are not typically captured by traditional SDM approaches.

We considered 2 similarity indices, Schoener's D for continuous predictions and the Jaccard coefficient for binary predictions, and computed the dissimilarity by subtracting these indices from one (i.e., 1 − Schoener's D and 1 − Jaccard). To determine the drivers of dissimilarity in model predictions between the 2 ensemble types, we fitted 2 GLMs with 6 different candidate covariates: number of opportunistic occurrence records, average distance of opportunistic occurrence records to the expert range polygon (where occurrences within the polygon have a distance of zero), phylogenetic group, number of publications, extent of opportunistic occurrence records, and conservation status. We hypothesized that a higher number of occurrence records leads to more consistent model predictions and thus less dissimilarity between the 2 model predictions. For distance of opportunistic occurrence records to the expert range polygon, a larger average distance indicates a less accurate expert range map (assuming the occurrence data are accurate) and less influence of the range map on the SDM prediction. We specified fish family (Syngnathidae and Epinephelidae) as a covariate to determine whether one group has a higher dissimilarity level. Using the species' scientific names as search keywords, we extracted the number of publications for each species from Crossref (https://www.crossref.org). Given that understanding of biodiversity greatly depends on scientific interest (Mouquet et al., 2024), we hypothesized that the higher the number of publications on a species, the greater its general scientific interest (Adamo et al., 2021; Mammola et al.,

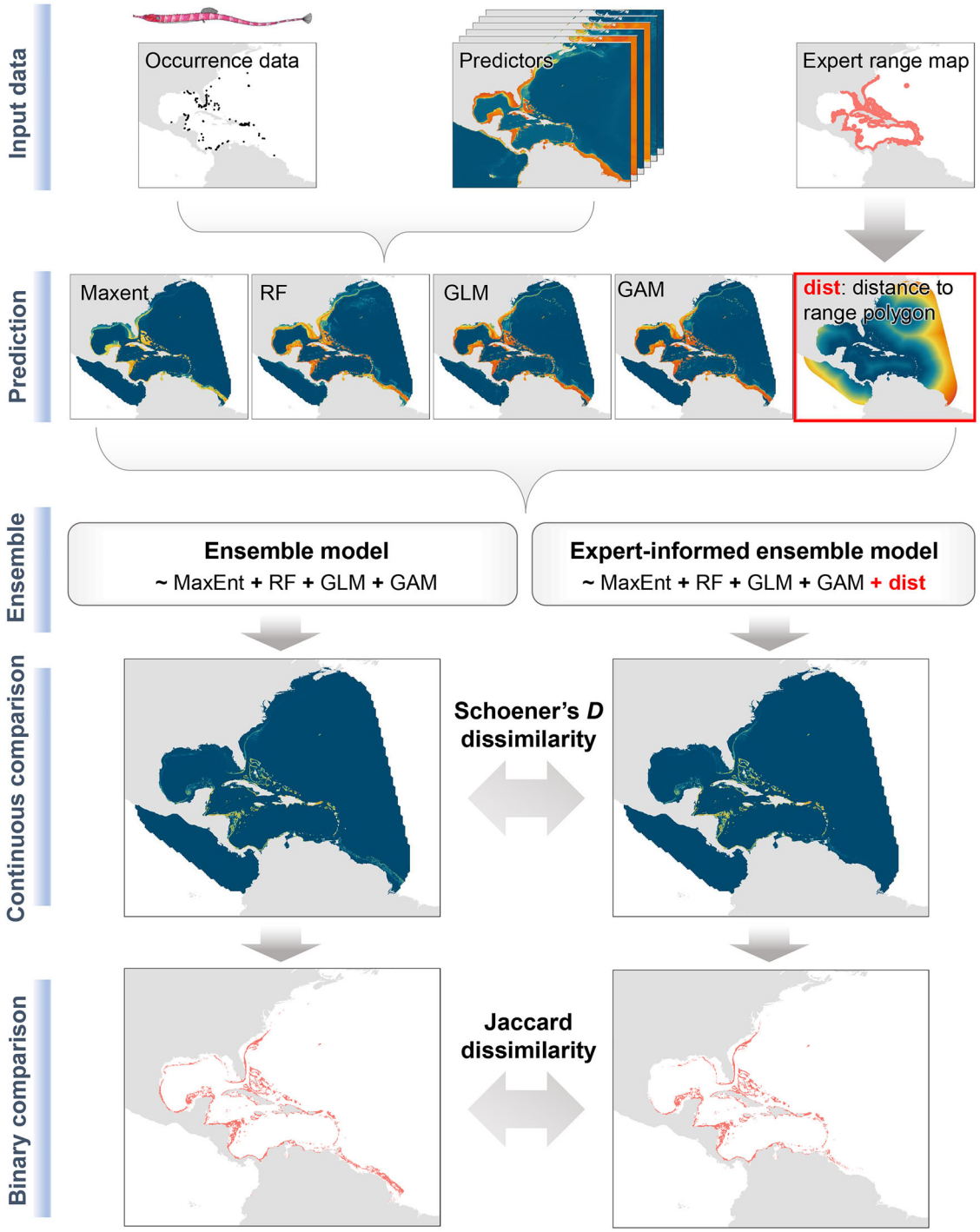

FIGURE 1 Workflow of species distribution modeling that integrates expert range maps and opportunistic occurrence records (Maxent, maximum entropy; RF, random forest; GLM, generalized linear model; GAM, generalized additive model). Shortfin pipefish (*Cosmocampus elucens*) are the example here. Expert range maps of shortfin pipefish are from the International Union for Conservation of Nature Red List of Threatened Species.

2023). Expert-based range maps rely on the integration of various sources of information, including distributional data and accumulated knowledge on species' biology, ecology, biogeography, and evolutionary history (Zhang et al., 2025). Therefore, in principle, it is reasonable to assume that the more papers published on a species, the more complete the body of knowledge that can inform the generation of expert-based maps; thus, the higher the accuracy of the expert range map, the greater its influence on the SDM prediction. For each species, we determined the extent of its occurrence records by computing the area of a minimum convex polygon around these records. Compared with wide-ranging species, narrow-ranging species can

have lower niche breadths (Slatyer et al., 2013), making their distributions more constrained by environmental conditions and thus more predictable with SDMs trained on environmental covariates (McPherson & Jetz, 2007). Therefore, we hypothesized that narrow-ranging species have lower range prediction dissimilarity between models that consider expert range maps and those that do not. The geographical distributions of wide-ranging species are more likely to be constrained by dispersal barriers or biotic interactions, which means their relationships with the environment should be more difficult to predict with SDMs that do not benefit from expert range map information, leading to higher dissimilarity. We obtained species conservation status from the IUCN Red List of Threatened Species (https://www.iucnredlist.org, accessed 26 February 2024). We classified species as data deficient ($n = 25$) and data sufficient ($n = 171$). Data-deficient species have insufficient data on abundance or distribution, which precludes the assessment of their risk status, although the species and its biology may be well studied. Therefore, we hypothesized that expert maps of species with sufficient data better reflect their actual distribution, resulting in greater influence of expert range maps on SDM predictions and thus higher levels of dissimilarity.

For the GLMs, we standardized the continuous covariates with a z transformation (i.e., zero mean and unit variance) to facilitate model convergence. We constructed GLMs in a Bayesian framework with the brms 2.16.3 package with 4 Markov chains and 20,000 iterations per chain with 5000 warmup interactions (Bürkner, 2017). We considered a covariate to have a significant effect if its 95% credible interval did not include zero and to have a marginally significant effect if this was true for the 90% credible interval.

RESULTS

Model predictive performance

Both model types had similarly high performance in predicting withheld data. The mean continuous Boyce index values were not significantly different (2-sided paired Wilcoxon signed rank tests, $V = 8592$, $p = 0.269$): regular ensembles 0.669 (SD 0.158) and expert-informed ensembles 0.674 (0.159) (Figure 2a; Appendix S5). Both also showed similar reasonable (>0.4) confidence values ($V = 7781$, $p = 0.214$) for the predicted presences. The mean confidence measures were 0.462 (0.135) for the regular ensembles and 0.465 (0.137) for the expert-informed ensembles (Figure 2b). However, there was a significant difference in model transferability ($V = 6679$, $p < 0.001$; Figure 2c). The expert-informed ensembles were significantly more transferable, based on their higher consistency for predicted presences (−0.237 [0.125]), than the regular ensembles (−0.246 [0.123]). The permutation importance of distance to expert range polygons was significantly higher for Epinephelidae (30.0 ± 28.8%) than for Syngnathidae (17.7 ± 21.2%) (2-sided Wilcoxon signed rank tests, $V = 5919$, $p < 0.01$) (Appendix S6).

Range predictions based on different model types

Species richness predicted by the 2 types of models showed similar spatial patterns (Appendix A7); fishes of the 2 families were distributed along shallow coastal waters (Figure 3), and both types of models predicted accurately the Indo-Pacific hotspot for the 2 families (Appendix S8). Despite these similarities in overall richness patterns, integrating expert range maps resulted in significantly smaller predicted range sizes (1.6×10^6 km^2 [SD 1.8×10^6]) than those from regular SDM ensembles (1.8×10^6 km^2 [2.2×10^6]) ($V = 15{,}624$, $p < 0.001$) (Appendix S9). Compared with the regular ensemble models, predictions from expert-informed ensemble models predicted lower richness for 49.7% of nonzero richness regions (i.e., where at least one of the ensemble methods predicted >0 richness), the same richness in 32.7% of regions, and higher richness in 17.6% of regions. Where the expert-informed ensemble models predicted lower richness, the most extreme differences were in the Indo-Pacific, southern Australia, and southern Africa (Figure 3c). Expert-informed ensemble models tended to predict more realistic distributions by avoiding overprediction across geographical barriers (spillovers), such as major river plumes or land isthmuses (Figure 4).

Drivers of dissimilarity in model predictions

The regular ensembles and expert-informed ensembles produced highly similar spatial predictions (Schoener's D dissimilarity 0.126 [SD 0.101], Jaccard dissimilarity coefficient 0.186 [0.140]) (Appendix S10). Regression analyses revealed consistent results for both dissimilarity indices (Figure 5a,b). Differences were larger between predictions for species with larger occurrence extents, supporting our hypothesis. Our hypotheses regarding the distance of occurrences to expert range polygons and the number of occurrence records were also supported because these factors had a significant negative effect on the dissimilarity of the predictions (Figure 5). We found a significant effect for fish family (marginally for Schoener's D dissimilarity index), and Syngnathidae had lower dissimilarity levels than Epinephelidae (Figure 5). Conservation status and number of publications did not significantly affect either dissimilarity index (Figure 5).

DISCUSSION

We found that our integration of expert range maps and opportunistic occurrence records effectively avoided overprediction across biogeographical barriers and that the effects of expert range map integration depended on sample size, occurrence data extent, distance of occurrences to expert range polygon, and taxonomic group. Our results provide evidence of the benefits of integrating expert knowledge into species range estimates and insight into the drivers of the observed differences

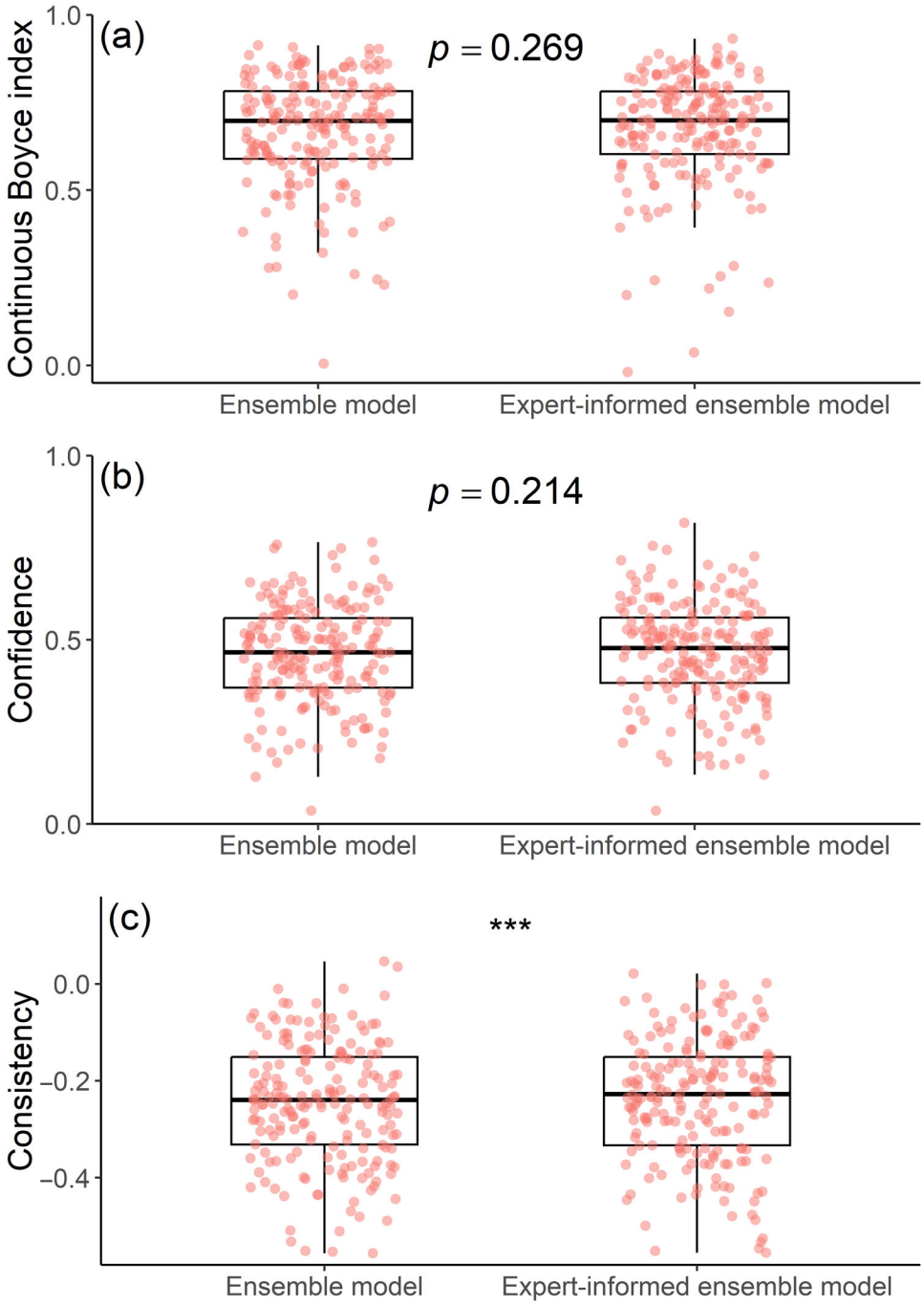

FIGURE 2 Predictive performance of the 2 types of ensemble species distribution models used in the study of integration of expert range maps and occurrence records: (a) continuous Boyce index, (b) confidence, and (c) consistency (horizontal lines, median values; box ends, interquartile range; whiskers, maximum and minimum values no further than 1.5 × interquartile range from the box end; ***$p < 0.001$).

in predictions between regular SDMs and those that integrate expert maps.

Data integration in biodiversity assessments

Constrained by the quality and quantity of distributional and environmental data, SDM studies in the marine realm lag far behind those in the terrestrial realm (Robinson et al., 2011). However, relative to terrestrial species, the geographic distributions of marine species should be more amenable to modeling because their geographical range boundaries more closely align with their thermal tolerances (Sunday et al., 2012). As a result, the marine realm represents an ideal arena to test data integration in SDMs with the purpose of addressing overpredictions of ranges that fail to capture species' dispersal capabilities and

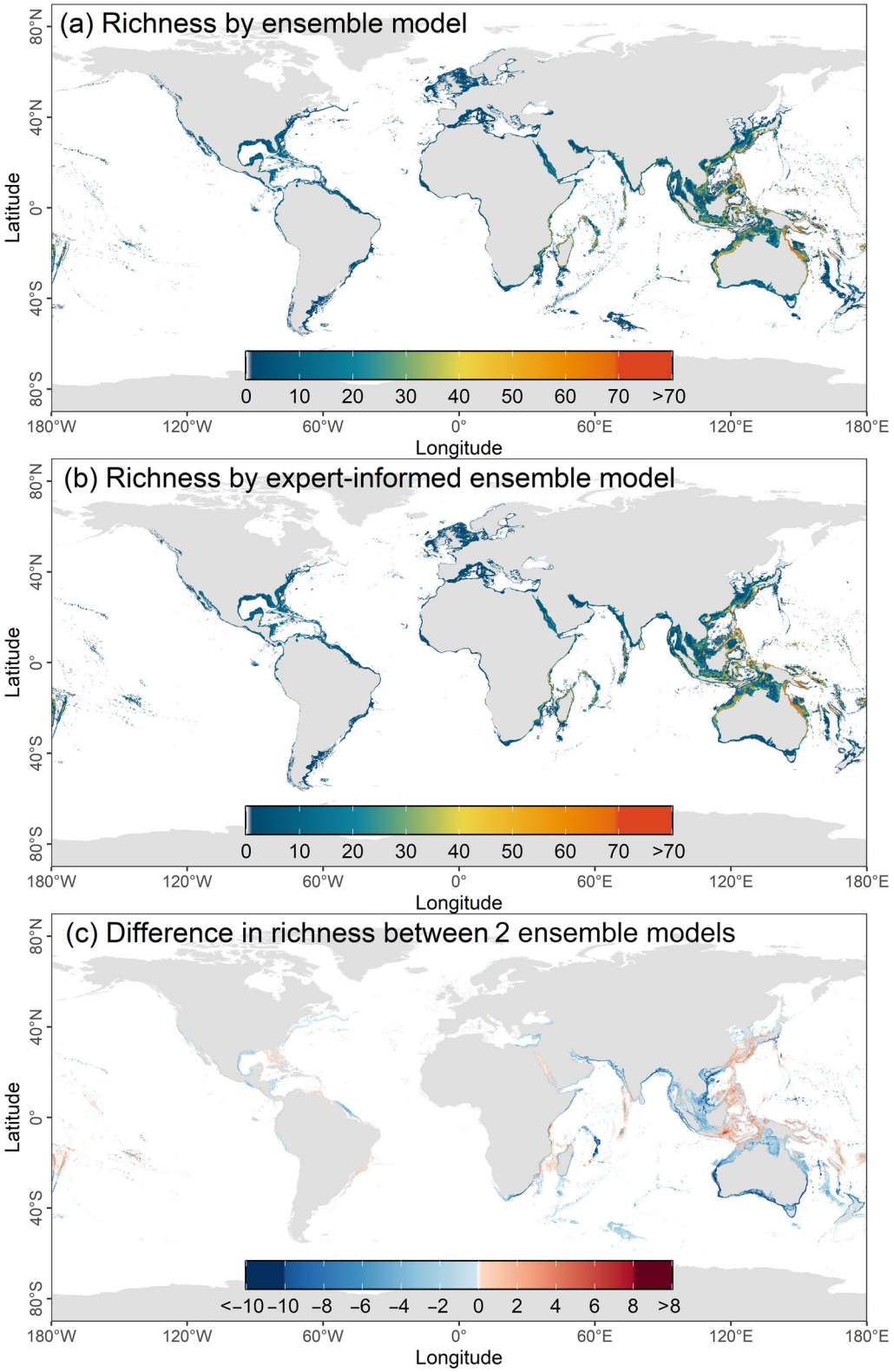

FIGURE 3 Spatial patterns of fish richness predicted by (a) ensemble models and (b) expert-informed ensemble models and (c) difference in richness between the models in (a) and (b).

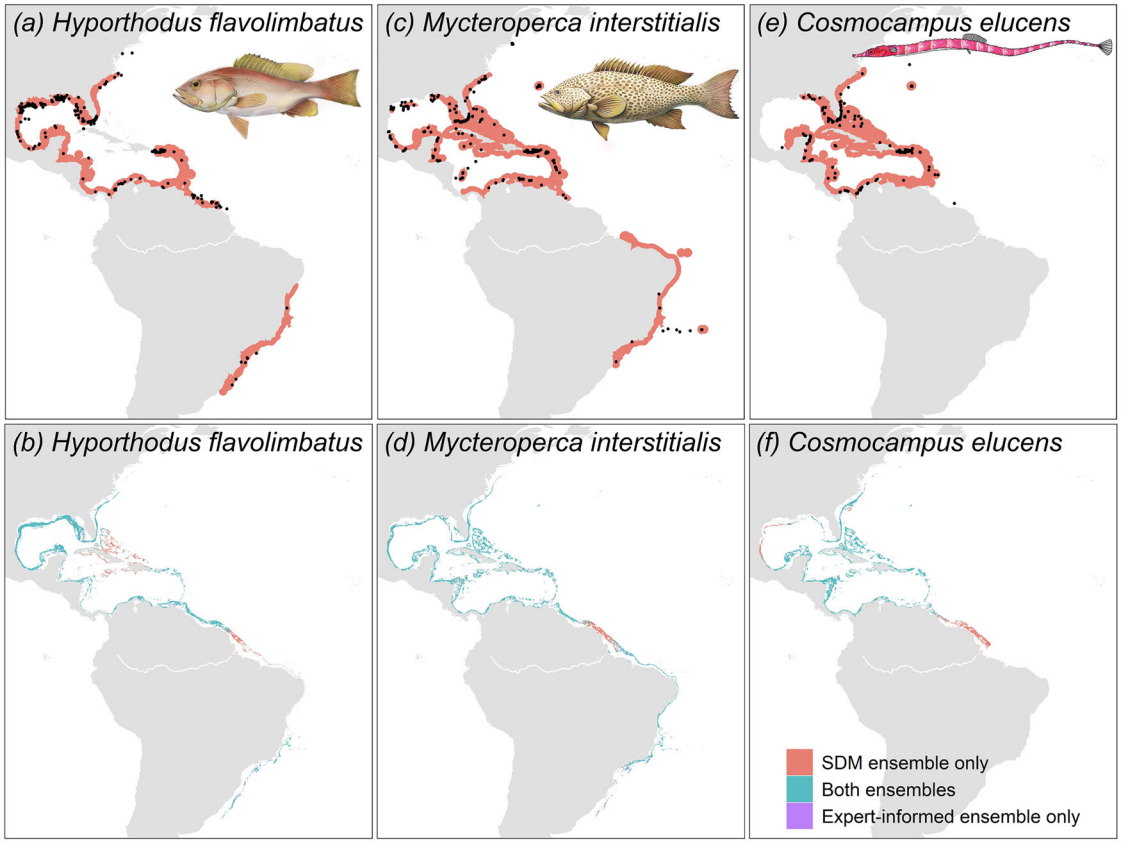

FIGURE 4 Effect of the significant oceanic barrier of the Amazon–Orinoco Plume in the tropical Atlantic Ocean on the distribution ranges of (a) *Hyporthodus flavolimbatus*, (c) *Mycteroperca interstitialis*, and (e) *Cosmocampus elucens* based on opportunistic occurrence records (points) and International Union for Conservation of Nature expert range polygons (red shading areas) of and range predictions based on species distribution models for (b) *Hyporthodus flavolimbatus*, (d) *Mycteroperca interstitialis*, and (f) *Cosmocampus elucens* (white line, Amazon River; SDM, species distribution model). High-quality figures are available on Figshare (https://figshare.com/s/eea6a4b3f664d6cad5ce).

limits imposed by biotic interactions. For instance, Zhang et al. (2024) built SDMs incorporating physiological knowledge for the Japanese sea cucumber to make range predictions with improved accuracy; yet, their models overestimated the distribution of the species south of the Yangtze Estuary, which represents a significant known dispersal barrier for the species. Our results showed that the expert range map integration approach proposed by Oeser et al. (2024) can address this issue effectively (Figure 4). Overall, we found that estimates of marine species' realized ranges can be improved by integrating opportunistic occurrences and expert range maps. Such improved estimates could inform conservation planning and management (e.g., expansion of protected areas or prioritizing species and areas for conservation) (Velazco et al., 2020).

Although our results revealed clear merits of integrating expert range maps in SDM ensembles, the continuous Boyce index and the confidence in predicted presences did not show any significant difference in the predictive performance of the 2 ensembles (Figure 2a,b). Both metrics focus on predicted presences (Hirzel et al., 2006; Somodi et al., 2024), which is a favorable feature for presence–background SDMs (e.g., Di Cola et al., 2017; Leroy et al., 2018). However, the main advantage of integrating expert range maps is to improve predictions of absences (i.e., excluding suitable but uninhabited areas for different reasons, such as presence of geographical barriers), which are not considered in these metrics. In contrast, the consistency metric, which compared the predictive performance in the training and evaluation subsets to measure the model's capacity to generalize the occurrence–environment relationship (i.e., transferability), demonstrated that the transferability of the expert-informed ensemble models was significantly better (Figure 2c).

Factors influencing data integration

Differences between marine fish SDMs that do and do not account for expert range maps (Figure 5) were associated with the number and the extent of opportunistic occurrences, the average distance of occurrences to expert range polygons, and fish family. These results confirmed our hypothesis that this dissimilarity is negatively associated with the number of occurrence records. Because performance of SDMs is closely associated with sample size (Wisz et al., 2008) and given that the distributions of species with ample occurrence information is more complete, range estimates of data-rich species tended to be reliable whether or not they accounted for expert range maps. In addition, our results indicated that species with large occurrence

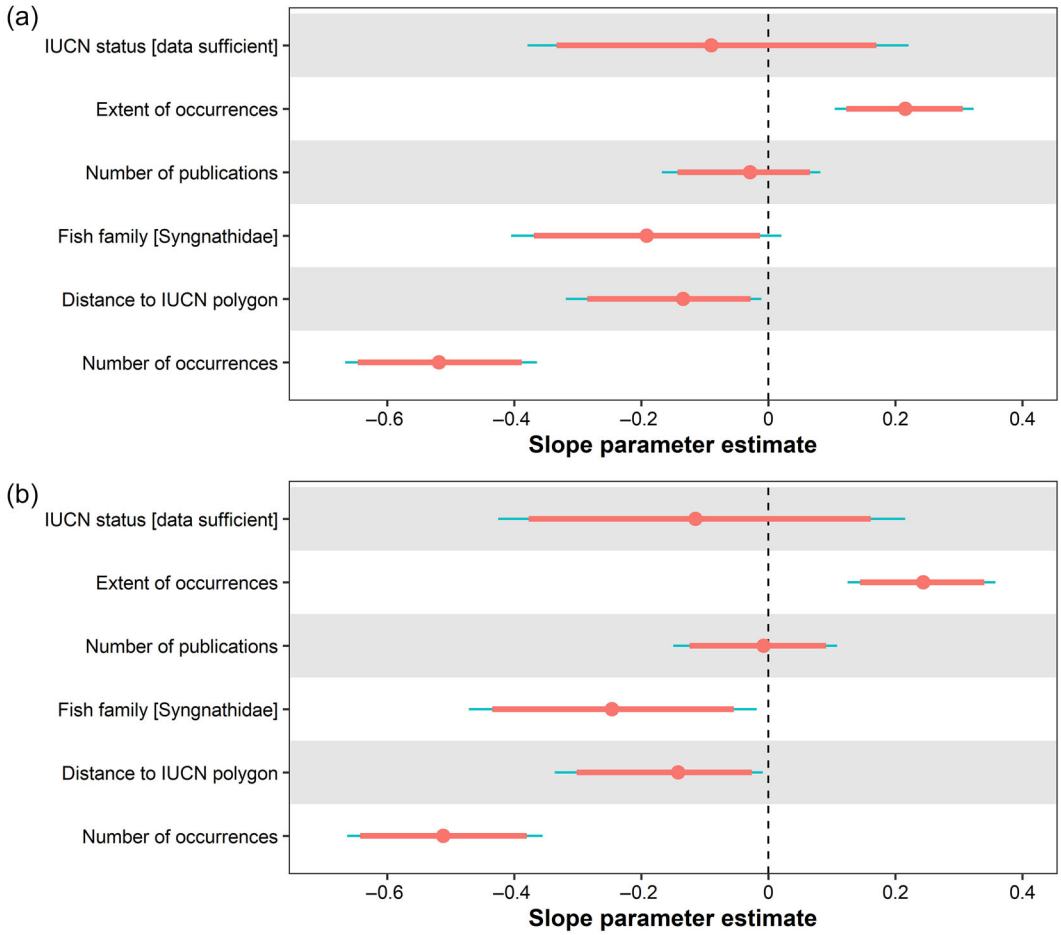

FIGURE 5 Factors driving the difference in predictions of marine fish habitat between ensemble models and expert-informed ensemble models as measured by (a) Schoener's *D* dissimilarity index for continuous predictions and (b) Jaccard dissimilarity coefficient for binary predictions (red lines, 90% credible intervals; blue lines, 95% credible intervals; IUCN, International Union for Conservation of Nature).

data extents showed high dissimilarity between the ensemble models. Meyer et al. (2016) found that the degree of occurrence data range coverage for widely distributed species is lower than that for species with restricted distributions. Thus, data missing for wide-ranging species are more likely to lead to incomplete niche estimates by SDMs, which could explain our results.

The expert-informed ensemble models identified the critical role of the distance data layer (Appendix S6); dissimilarity between the ensemble models was high when the average distance of occurrences to expert range polygons was low (Figure 5), which increased the weight of the expert map. A deeper understanding of this relationship requires further investigation, but our results highlight that the benefits of integrating expert range maps into the ensembles are best exploited when the expert range map and opportunistic occurrences are in broad spatial agreement.

Epinephelidae species showed a higher level of dissimilarity between the ensemble models than Syngnathidae. Compared with Syngnathidae species, Epinephelidae species exhibit a stronger dispersal ability and utilize a broader variety of habitats (Zhang et al., 2025). This is partially supported by the significantly higher number of occurrences (2-sided Wilcoxon signed rank tests, $V = 6316$, $p < 0.001$) and larger occurrence data extent ($V = 6010$, $p < 0.01$) of species in Epinephelidae relative to Syngnathidae (Appendix S11). We attributed the existing difference in dissimilarity in model predictions between fish families to the differences in the number and extent of occurrences and to the differences in movement abilities and habitat preferences. Another reason for this difference might be the variation in accuracy of expert range maps among fish families. Oeser et al. (2024) found that when an expert range map is inaccurate (i.e., a large proportion of occurrences are outside the map polygon), the data integration method has a minor effect on distribution predictions. Given the vital role of the quality of expert range maps in data integration and the diverse range of data and bodies of knowledge informing these maps, we recommend that map providers include standardized metadata detailing the type of data used and the process followed to generate these maps. Such information will help researchers assess the quality and accuracy of expert range maps and help guide decisions about their reliability and use (Zhang et al., 2025).

Further perspectives

Driven by climate change, species are rapidly shifting their geographical distributions (Lenoir & Svenning, 2015; Lenoir et al., 2020). Researchers frequently use SDMs to project the future distribution of biodiversity (e.g., García Molinos et al., 2016), providing critical information on species vulnerability under climate change (e.g., Peng et al., 2023) and informing proactive management policies (e.g., Velazco et al., 2020). Although data integration can improve estimates of current realized ranges, new approaches are needed to create future predictions that are compatible with expert-informed ensembles, for example, by adding dispersal buffers around the current realized range to identify habitats that could potentially be occupied under future climate change.

We found that range estimates from SDMs with opportunistic occurrences of marine fish species benefit from integrating expert range maps and that they can be especially useful to help avoid overprediction past oceanic barriers. The benefits, however, depend on species attributes, such as range size and occurrence information. In general, we recommend the expert range map integration approach to improve the reliability of range estimates for marine species, but we acknowledge that more research should be done on different taxonomic groups to determine the universality of these findings.

AUTHOR CONTRIBUTIONS

Zhixin Zhang, Qiang Lin, Ákos Bede-Fazekas, and Jorge García Molinos conceived the idea. Zhixin Zhang, Junmei Qu, Songxi Yuan, and Qiang Lin collected and cleaned species distribution data. Zhixin Zhang performed data analyses with critical suggestions from Ákos Bede-Fazekas, Jorge García Molinos, Stefano Mammola, Jamie M. Kass, and Julian Oeser. Zhixin Zhang, Ákos Bede-Fazekas, and Qiang Lin led the writing of the manuscript with helpful suggestions from all authors. All authors approved the final manuscript.

ACKNOWLEDGMENTS

This work was financially supported by the National Key Research and Development Program of China (2023YFC3108800), the National Natural Science Foundation of China (42230409, 42425004, 42276126, 42306133, 32170535), the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, the National Multidisciplinary Laboratory for Climate Change (NKFIH-471-3/2021, RRF-2.3.1-21-2022-00014), the Guangdong Basic and Applied Basic Research Foundation (2024A1515010604), and the development fund of South China Sea Institute of Oceanology of the Chinese Academy of Sciences (SCSIO202208). We sincerely thank T. Kümmerle (Humboldt University Berlin, Germany) for his kind help with the methodology.

DATA AVAILABILITY STATEMENT

Species occurrence records, expert range maps, and high-quality figures in the main text are available from Figshare (https://figshare.com/s/eea6a4b3f664d6cad5ce). Marine predictors used for model development are available from the Bio-ORACLE version 3.0 database (https://www.bio-oracle.org).

ORCID

Zhixin Zhang https://orcid.org/0000-0002-3457-2934

REFERENCES

Adamo, M., Chialva, M., Calevo, J., Bertoni, F., Dixon, K., & Mammola, S. (2021). Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers. *Nature Plants*, *7*, 574–578.

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*, Article eaat4858.

Assis, J., Fernández Bejarano, S. J., Salazar, V. W., Schepers, L., Gouvêa, L., Fragkopoulou, E., Leclercq, F., Vanhoorne, B., Tyberghein, L., Serrão, E. A., Verbruggen, H., & De Clerck, O. (2024). Bio-ORACLE v3. 0. Pushing marine data layers to the CMIP6 Earth System Models of climate change research. *Global Ecology and Biogeography*, *33*, Article e13813.

Bosch, S., Tyberghein, L., Deneudt, K., Hernandez, F., & De Clerck, O. (2018). In search of relevant predictors for marine species distribution modelling using the MarineSPEED benchmark dataset. *Diversity and Distributions*, *24*, 144–157.

Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R. O., Wang, Z., & Zimmermann, N. E. (2020). Model complexity affects species distribution projections under climate change. *Journal of Biogeography*, *47*, 130–142.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Champion, C., & Coleman, M. A. (2021). Seascape topography slows predicted range shifts in fish under climate change. *Limnology and Oceanography Letters*, *6*, 143–153.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., d'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N., & Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, *40*, 774–787.

Domisch, S., Wilson, A. M., & Jetz, W. (2016). Model-based integration of observed and expert-based information for assessing the geographic and environmental distribution of freshwater species. *Ecography*, *39*, 1078–1088.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*, 27–46.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677–697.

Feng, X., Park, D. S., Walker, C., Peterson, A. T., Merow, C., & Papeş, M. (2019). A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology & Evolution*, *3*, 1382–1395.

Fourcade, Y. (2016). Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. *Ecological Informatics*, *36*, 8–14.

García Molinos, J., Halpern, B. S., Schoeman, D. S., Brown, C. J., Kiessling, W., Moore, P. J., Pandolfi, J. M., Poloczanska, E. S., Richardson, A. J., & Burrows, M. T. (2016). Climate velocity and the future global redistribution of marine biodiversity. *Nature Climate Change*, *6*, 83–88.

Giachini Tosetto, E., Bertrand, A., Neumann-Leitão, S., & Nogueira Júnior, M. (2022). The Amazon River plume, a barrier to animal dispersal in the Western Tropical Atlantic. *Scientific Reports*, *12*, Article 537.

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*, 142–152.

Huang, H., Zhang, Z., Bede-Fazekas, Á., Mammola, S., Gu, J., Zhou, J., Qu, J., & Lin, Q. (2025). Cross-validation matters in species distribution models: A case study with goatfish species. *Ecography*, *2025*, Article e07354.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*, 1259–1269.

Hurlbert, A. H., & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 13384–13389.

Isaac, N. J., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G., & O'Hara, R. B. (2020). Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, *35*, 56–67.

Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology & Evolution*, *27*, 151–159.

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., Soley-Guardia, M., & Anderson, R. P. (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods in Ecology and Evolution*, *12*, 1602–1608.

Kosmidis, I. (2023). *brglm2: Bias reduction in generalized linear models*. R package version 0.9. https://CRAN.R-project.org/package=brglm2

Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Murienne, J., & Grenouillet, G. (2020). Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution*, *4*, 1044–1059.

Lenoir, J., & Svenning, J. C. (2015). Climate-related range shifts—A global multidimensional synthesis and new research directions. *Ecography*, *38*, 15–28.

Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018). Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, *45*, 1994–2002.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Luiz, O. J., Madin, J. S., Robertson, D. R., Rocha, L. A., Wirtz, P., & Floeter, S. R. (2012). Ecological traits influencing range expansion across large oceanic dispersal barriers: Insights from tropical Atlantic reef fishes. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1033–1040.

Mainali, K., Hefley, T., Ries, L., & Fagan, W. F. (2020). Matching expert range maps with species distribution model predictions. *Conservation Biology*, *34*, 1292–1304.

Mammola, S., Adamo, M., Antić, D., Calevo, J., Cancellario, T., Cardoso, P., Chamberlain, D., Chialva, M., Durucan, F., Fontaneto, D., Goncalves, D., Martínez, A., Santini, L., Rubio-Lopez, I., Sousa, R., Villegas-Rios, D., Verdes, A., & Correia, R. A. (2023). Drivers of species knowledge across the Tree of Life. *eLife*, *12*, Article RP88251.

McPherson, J. M., & Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography*, *30*, 135–151.

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. *Ecological Modelling*, *415*, Article 108837.

Merow, C., Wilson, A. M., & Jetz, W. (2017). Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography*, *26*, 243–258.

Merow, C., Galante, P. J., Kass, J. M., Aiello-Lammens, M. E., Babich Morrow, C., Gerstner, B. E., Grisales-Betancur, V., Moore, A. C., Noguera-Urbano, E. A., Pinilla-Buitrago, G. E., Velásquez-Tibatá, J., Anderson, R. P., & Blair, M. E. (2022). Operationalizing expert knowledge in species' range estimates using diverse data types. *Frontiers of Biogeography*, *14*, Article e53589.

Meyer, C., Jetz, W., Guralnick, R. P., Fritz, S. A., & Kreft, H. (2016). Range geometry and socio-economics dominate species-level biases in occurrence information. *Global Ecology and Biogeography*, *25*, 1181–1193.

Miller, B. S. (2009). *Early life history of marine fishes* (Vol. 364). University of California Press.

Mouquet, N., Langlois, J., Casajus, N., Auber, A., Flandrin, U., Guilhaumon, F., Loiseau, N., McLean, M., Receveur, A., Stuart Smith, R. D., & Mouillot, D. (2024). Low human interest for the most at-risk reef fishes worldwide. *Science Advances*, *10*, Article eadj9510.

Naimi, A. I., & Balzer, L. B. (2018). Stacked generalization: An introduction to super learning. *European Journal of Epidemiology*, *33*, 459–464.

Oeser, J., Zurell, D., Mayer, F., Çoraman, E., Toshkova, N., Deleva, S., Natradze, I., Benda, P., Ghazaryan, A., Irmak, S., Hasanov, N., Guliyeva, G., Gritsina, M., & Kuemmerle, T. (2024). The best of two worlds: Using stacked generalisation for integrating expert range maps in species distribution models. *Global Ecology and Biogeography*, *33*, Article e13911.

Oeser, J., Zurell, D., Mayer, F., Çoraman, E., Toshkova, N., Deleva, S., Natradze, I., Benda, P., Dietz, C., Georgiakakis, P., Levin, E., Dolev, A., Dundarova, H., Ghazaryan, A., Irmak, S., Hasanov, N., Guliyeva, G., Gritsina, M., Bukhnikashvili, A., & Kuemmerle, T. (2025). Variations of environmental niche breadth, range sizes and geographic exclusion with bat species richness. *Journal of Biogeography*, *52*, Article e15125.

Peng, S., Shrestha, N., Luo, Y., Li, Y., Cai, H., Qin, H., Ma, K., & Wang, Z. (2023). Incorporating global change reveals extinction risk beyond the current Red List. *Current Biology*, *33*, 3669–3678.

Qu, M., Zhang, Y., Gao, Z., Zhang, Z., Liu, Y., Wan, S., Wang, X., Yu, H., Zhang, H., Liu, Y., Schneider, R., Meyer, A., & Lin, Q. (2023). The genetic basis of the leafy seadragon's unique camouflage morphology and avenues for its efficient conservation derived from habitat modeling. *Science China Life Sciences*, *66*, 1213–1230.

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, *41*, 629–643.

Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography*, *33*, 1689–1703.

Robinson, L. M., Elith, J., Hobday, A. J., Pearson, R. G., Kendall, B. E., Possingham, H. P., & Richardson, A. J. (2011). Pushing the limits in marine species distribution modelling: Lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, *20*, 789–802.

Sales, L. P., Ribeiro, B. R., Pires, M. M., Chapman, C. A., & Loyola, R. (2019). Recalculating route: Dispersal constraints will drive the redistribution of Amazon primates in the Anthropocene. *Ecography*, *42*, 1789–1801.

Shipley, B. R., Bach, R., Do, Y., Strathearn, H., McGuire, J. L., & Dilkina, B. (2022). megaSDM: Integrating dispersal and time-step analyses into species distribution models. *Ecography*, *2022*, Article e05450.

Slatyer, R. A., Hirst, M., & Sexton, J. P. (2013). Niche breadth predicts geographical range size: A general ecological pattern. *Ecology Letters*, *16*, 1104–1114.

Somodi, I., Bede-Fazekas, Á., Botta-Dukát, Z., & Molnár, Z. (2024). Confidence and consistency in discrimination: A new family of evaluation metrics for potential distribution models. *Ecological Modelling*, *491*, Article 110667.

Sunday, J. M., Bates, A. E., & Dulvy, N. K. (2012). Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, *2*, 686–690.

Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N., & Zimmermann, N. E. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, *10*, Article 1446.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, *44*, 1731–1742.

Velazco, S. J. E., Ribeiro, B. R., Laureto, L. M. O., & Júnior, P. D. M. (2020). Overprediction of species distribution models in conservation planning: A still neglected issue with strong effects. *Biological Conservation*, *252*, Article 108822.

Waldock, C., Stuart-Smith, R. D., Albouy, C., Cheung, W. W., Edgar, G. J., Mouillot, D., Tjiputra, J., & Pellissier, L. (2022). A quantitative review of abundance-based species distribution models. *Ecography*, *2022*, Article e05694.

Warren, D. L., Matzke, N. J., & Iglesias, T. L. (2020). Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography*, *47*, 167–180.

Webb, T. J. (2012). Marine and terrestrial ecology: Unifying concepts, revealing differences. *Trends in Ecology & Evolution*, *27*, 535–541.

Wiens, J. J. (2023). How many species are there on Earth? Progress and problems. *PLoS Biology*, *21*, Article e3002388.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*, 763–773.

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., Forchhammer, M. C., Grytnes, J., Guisan, A., Heikkinen, R. K., Høye, T. T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M., Normand, S., Öckinger, E., Schmidt, N. M., … Svenning, J. C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, *88*, 15–30.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*, 241–259.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*, 3–36.

Xiao, B., Yuan, S., Bede-Fazekas, Á., Zhou, J., Song, X., Lin, Q., Cui, L., & Zhang, Z. (2025). Improving distribution prediction by integrating expert range maps and opportunistic occurrences: Evidence from Japanese sea cucumber. *Ecology and Evolution*, *15*, Article e71747.

Zhang, Z., Kass, J. M., Mammola, S., Koizumi, I., Li, X., Tanaka, K., Ikeda, K., Suzuki, T., Yokota, M., & Usio, N. (2021). Lineage-level distribution models lead to more realistic climate change predictions for a threatened crayfish. *Diversity and Distributions*, *27*, 684–695.

Zhang, Z., Zhou, J., García Molinos, J., Mammola, S., Bede-Fazekas, Á., Feng, X., Kitazawa, D., Assis, J., Qiu, T., & Lin, Q. (2024). Incorporating physiological knowledge into correlative species distribution models minimizes bias introduced by the choice of calibration area. *Marine Life Science & Technology*, *6*, 349–362.

Zhang, Z., Kass, J. M., Bede-Fazekas, Á., Mammola, S., Qu, J., García Molinos, J., Gu, J., Huang, H., Qu, M., Yue, Y., Qin, G., & Lin, Q. (2025). Substantial differences in predictions of marine species distribution models based on expert maps and opportunistic occurrences. *Conservation Biology*, *39*, Article e70015.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*, 744–751.

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, *43*, 1261–1277.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang, Z., Bede-Fazekas, Á., Molinos, J. G., Mammola, S., Kass, J. M., Qu, J., Oeser, J., Yuan, S., Zhang, C., Gu, J., Ding, L., & Lin, Q. (2026). Integrating expert range maps and opportunistic occurrence records of marine fish species in range estimates. *Conservation Biology*, *40*, e70154. https://doi.org/10.1111/cobi.70154