Received: 16 July 2025 | Revised: 29 March 2026 | Accepted: 11 April 2026

DOI: 10.1111/cobi.70347

Conservation Biology

CONTRIBUTED PAPER

Comparative species distribution model framework for marine conservation and its application to loggerhead turtles in the Mediterranean

E. Pasanisi[1,2,3] | F. Maffucci[4] | F. Atzori[5] | M. Azzolin[6,7] | I. Campana[8,9,10] | L. Carosso[5] | A. Castelli[11] | M. Costantino[6,7] | R. Crosti[1] | L. David[12] | N. Di-Meglio[12] | M. Farina[4] | N. Fraija-Fernández[13] | O. Garcia-Garin[14,15] | M. Gazo[14] | M. Gregorietti[1] | F. Grossi[16] | S. Hochscheid[4] | R. Míguez-Lozano[13] | A. Moulins[16] | D. S. Pace[10,17,18] | M. Pecoraro[20] | E. Pignata[8] | J. A. Raga[13] | M. Roul[12] | E. Santini[1] | G. Sarà[20] | A. Scuderi[19] | R. Teti[4] | P. Tepsich[16] | M. Vitale[2] | M. Vighi[14] | A. Arcangeli[1]

[1]Department for Biodiversity Conservation and Monitoring, The Italian Institute for Environmental Protection and Research (ISPRA), Rome, Italy

[2]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy

[3]Echoes-Capitoline Centre for Marine Research–ETS, Rome, Italy

[4]Marine Turtle Research Group, Department of Marine Animal Conservation and Public Engagement, Stazione Zoologica Anton Dohrn, Napoli, Italy

[5]Capo Carbonara Marine Protected Area, Villasimius, Italy

[6]Department of Life and System Biology, University of Turin, Turin, Italy

[7]Gaia Research Institute, Turin, Italy

[8]Accademia del Leviatano APS, Maccarese, Italy

[9]Department of Ecological and Biological Sciences, Tuscia University, Viterbo, Italy

[10]National Inter-University Consortium for Marine Sciences (CoNISMa), Rome, Italy

[11]Department of Biology, University of Pisa, Pisa, Italy

[12]EOI – EcoOcéan Institut, Montpellier, France

[13]Marine Zoology Unit, Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, University of Valencia, Valencia, Spain

[14]Department of Evolutionary Biology, Ecology and Environmental Sciences, and Biodiversity Research Institute (IRBio), Faculty of Biology, University of Barcelona, Barcelona, Spain

[15]Institute of Aquatic Ecology (IEA), and Department of Environmental Sciences, Faculty of Sciences, Universitat de Girona, Girona, Spain

[16]CIMA Research Foundation, Savona, Italy

[17]Department of Biology, Tor Vergata University of Rome, Rome, Italy

[18]CNR - Institute for the Study of Anthropogenic Impacts and Sustainability in the Marine Environment - Capo Granitola Unit, Trapani, Italy

[19]Association Nereide, Tarifa, Spain

[20]Department of Earth and Marine Sciences (DiSTeM), University of Palermo, Palermo, Italy

Correspondence
Pasanisi E., Department of Environmental Biology, Sapienza University of Rome, Italy.
Email: eugenia.pasanisi@uniroma1.it

Abstract

Modeling species distributions in dynamic pelagic environments remains challenging, particularly for wide-ranging and highly mobile species when there is limited guidance on

Article impact statement: Marine conservation can be improved by identifying robust, transferable models for mapping habitat use of mobile species in dynamic offshore ecosystems.

Funding information
LIFE, Grant/Award Numbers: Conceptu Maris project, LIFE20 NAT/IT/001371

model performance. This limitation constrains the effective use of species distribution models (SDMs) in marine conservation, where robust and transferable predictions are essential for conservation planning. Using the loggerhead turtle (*Caretta caretta*) as a case study, we developed a benchmarking workflow to compare modeling strategies across spatial resolutions, environmental predictor variables, and algorithms. We evaluated five common modeling approaches—generalized linear models, generalized additive models, random forest, boosted regression trees, and MaxEnt—and used a 15-year dataset of visual surveys collected along standardized ferry-based transects across the Mediterranean Sea, with environmental predictor variables averaged across years to capture recurrent summer conditions. Model outputs were assessed for their ability to predict habitat suitability and space-use patterns in dynamic pelagic environments. At the same time, robustness and spatial transferability were evaluated using an independent, basin-scale aerial survey dataset. Our results revealed consistent patterns informing SDM selection for wide-ranging pelagic species. In offshore Mediterranean pelagic systems, increased spatial resolution improved the ecological interpretability of predicted habitat patterns, whereas an extended set of dynamic oceanographic predictors enhanced the characterization of habitat associations. Independent validation was critical for selecting robust and transferable modeling algorithms, with MaxEnt and generalized additive models providing the most reliable performance for suitability and encounter rate (ER) modeling, respectively. Combining presence-only and encounter rate models helped distinguish habitat suitability from observation intensity. Overall, this study developed an empirically tested workflow to support robust, consistent, and policy-relevant SDM applications for mobile marine species in offshore, data-limited systems. By clarifying trade-offs among modeling approaches, this framework enhanced the application of SDMs to inform spatial planning, conservation prioritization, and species distribution and habitat-use assessments within regional and international conservation frameworks.

KEYWORDS
conservation, ferry-based surveys, habitat modeling, loggerhead turtle, marine megafauna, Mediterranean Sea, spatial resolution, species distribution models

INTRODUCTION

Species distribution models (SDMs) are key tools for investigating biodiversity patterns and informing spatial conservation planning. By relating species occurrence data to environmental conditions, these models provide spatial predictions of habitat suitability and potential distribution. This information is particularly valuable for highly mobile marine taxa such as cetaceans, sea turtles, and other pelagic megafauna that occupy vast and dynamic environments for which long-term standardized data are often scarce. For these groups, SDMs offer a practical framework to identify the ecological drivers of distribution and support management across broad spatial scales. For example, SDM outputs have been used to delineate species' potentia ecological ranges and their temporal trends, identify areas of importance, and develop spatial risk maps for major anthropogenic pressures (Almpanidou et al., 2022; Arcangeli et al., 2023, 2025). These outputs can also inform indicators and assessments within policy frameworks such as the European Union Habitats Directive.

Despite their wide application, SDMs remain sensitive to key methodological choices, including input data type (e.g., presence–pseudoabsence versus density/counts), spatial resolution, and algorithm selection, all of which can strongly influence model outputs and their interpretability. Traditional statistical approaches, particularly generalized additive models (GAMs) and generalized linear models (GLMs), have long dominated SDM applications in marine contexts. However, recent literature highlights the growing potential of machine learning techniques such as MaxEnt, random forest (RF), and boosted regression trees (BRT) (Melo-Merino et al., 2020; Robinson et al., 2017), which may offer greater flexibility in capturing complex, nonlinear ecological relationships, especially in data-limited or heterogeneous environments. Another critical yet often overlooked aspect related to these approaches is external validation using independent data collected under different sampling designs and platforms. Such validation is essential for assessing model robustness and spatial transferability (Valavi et al., 2022) but remains underused in marine SDM studies.

These challenges are particularly pronounced in offshore pelagic systems, where logistical constraints limit data availability and habitat use is shaped by highly dynamic oceanographic processes (Robinson et al., 2011). Although these regions are ecologically important, they remain undersampled, and the strong spatial and temporal variability in environmental conditions further complicates modeling efforts. Although the main components of SDM development, such as the choice of spatial resolution, environmental predictors, and modeling algorithms, are well established, their empirical consequences have rarely been evaluated using long-term, standardized datasets system-

atically in offshore environments, specifically those for highly mobile marine taxa. As a result, the extent to which common modeling decisions affect their ecological interpretability, spatial transferability, and conservation relevance in pelagic systems remains poorly understood, limiting the development of consistent and transferable modeling approaches.

To address this gap, we conducted a systematic benchmarking of SDM strategies by comparing two spatial resolutions (5 vs. 10 km), two sets of environmental predictors (base vs. extended), five model algorithms (GLM, GAM, BRT, RF, and MaxEnt), and two types of response data (presence–pseudoabsence and encounter rate [ER]). We used a 15-year dataset of at-sea sightings collected along fixed ferry-based transects across the Mediterranean Sea (2008–2023), derived from multispecies monitoring programs targeting cetaceans and sea turtles. These standardized observations, compiled through the Fixed Line Transect Mediterranean Network (FLT Med Net) and LIFE Conceptu Maris (LIFE20 NAT/IT/001371) projects, provide extensive spatial and temporal coverage, enabling SDM benchmarking under basin-wide conditions. This dataset has supported previous SDM applications for pelagic species (Arcangeli et al., 2023, 2024; Grossi et al., 2025; Zampollo et al., 2022), contributing to both methodological development and ecological understanding. We focused on the summer season, modeling recurrent seasonal conditions as a baseline framework that could be extended to other periods. Model outputs were then validated using independent basin-wide data from the ACCOBAMS Survey Initiative (ASI; Panigada et al., 2024), providing a rigorous test of spatial transferability across platforms, sampling designs, and spatial domains.

As a case study, we modeled the distribution of the loggerhead turtle (*Caretta caretta*), a wide-ranging and generalist species protected under multiple international agreements. Its high mobility and flexible behavior reflect key challenges associated with modeling pelagic megafauna in offshore environments (Wallace et al., 2011, 2025). Similar challenges apply to other wide-ranging marine vertebrates, including cetaceans and large pelagic fishes, making loggerhead turtles a suitable model system for SDM benchmarking. At the same time, sea turtles exhibit structured space-use patterns, including fidelity to foraging areas and consistent migration routes, which support the identification of recurrent habitat associations despite large-scale movements (Cerritelli et al., 2022; Perez et al., 2022; Schofield et al., 2010). In the Mediterranean Sea, this wide-ranging ecology also results in extensive overlap with human activities, leading to high levels of exposure to anthropogenic pressures, particularly fisheries bycatch, and reinforcing the conservation relevance of the species (Casale, 2011; Casale et al., 2018, 2020).

This work complements previous modeling efforts focused on low-density cetaceans with more selective habitat preferences (e.g., *Ziphius cavirostris*, *Globicephala melas*, and *Grampus griseus*; Arcangeli et al., 2024) by extending the comparison across modeling choices and taxa. Our approach also builds on recent systematic reviews of SDM applications to cetaceans and sea turtles, which highlight the diversity of modeling approaches and the lack of clear, standardized guidance for their selection and evaluation in marine systems (Pasanisi et al., 2024). Based on previous methodological studies, we expected that key modeling decisions, particularly spatial resolution, predictor selection, and algorithm choice, would influence model performance and predicted habitat suitability patterns. We further hypothesized that different modeling strategies would vary in robustness, leading to differences in performance when evaluated against independent datasets. By explicitly benchmarking common modeling decisions and combining internal and external validation, this study aimed to provide empirical evidence to support more consistent and conservation-oriented SDM applications in data-limited pelagic environments.

MATERIALS AND METHODS

Data sources, model inputs, and independent data

The occurrence data used as input for all model calibrations were collected between 2008 and 2023 across a representative portion of the Western Mediterranean Sea and the Adriatic Sea. Visual surveys were carried out onboard ferries operating along 33 fixed transect routes within the FLT Med Net program (2008–2021), the LIFE Conceptu Maris project (LIFE20 NAT/IT/001371, 2022–2026), and additional national and international monitoring initiatives (see the ACKNOWLEDGMENTS). Surveys followed standardized visual monitoring protocols for large marine fauna (Arcangeli et al., 2013, 2019, 2022, 2026). Observations were conducted under suitable wind and sea-state conditions (Beaufort ≤ 3), and records collected under poor visibility were excluded to reduce detection bias. All datasets were consolidated into a single, internally consistent database during the LIFE project. For modeling purposes, only summer observations (July–September) were retained, resulting in 2308 loggerhead turtle presence records used in the subsequent workflow.

An independent dataset used for external model validation was obtained from the ASI, a basin-wide aerial survey conducted across the Mediterranean Sea in June–September 2018 to estimate the distribution and abundance of cetaceans and other marine megafauna, including sea turtles (Panigada et al., 2024). Due to logistical constraints inherent to aerial surveys, sightings could not be reliably identified at the species level and were therefore aggregated at the family level (Cheloniidae). In the Western Mediterranean, the vast majority of Cheloniidae records are expected to correspond to loggerhead turtles, whereas in the Adriatic Sea, this assumption holds with slightly greater uncertainty; nonetheless, loggerhead turtles remained the predominant species. Records located west of the Strait of Gibraltar were excluded. To ensure temporal consistency with the modeling dataset, the ASI data were filtered to retain only Cheloniidae sightings recorded between July and September 2018, resulting in a final dataset of 2673 independent records that were used for external model validation. The spatial distribution of survey effort and sighting locations provided an overview of the structure of the modeling input data and

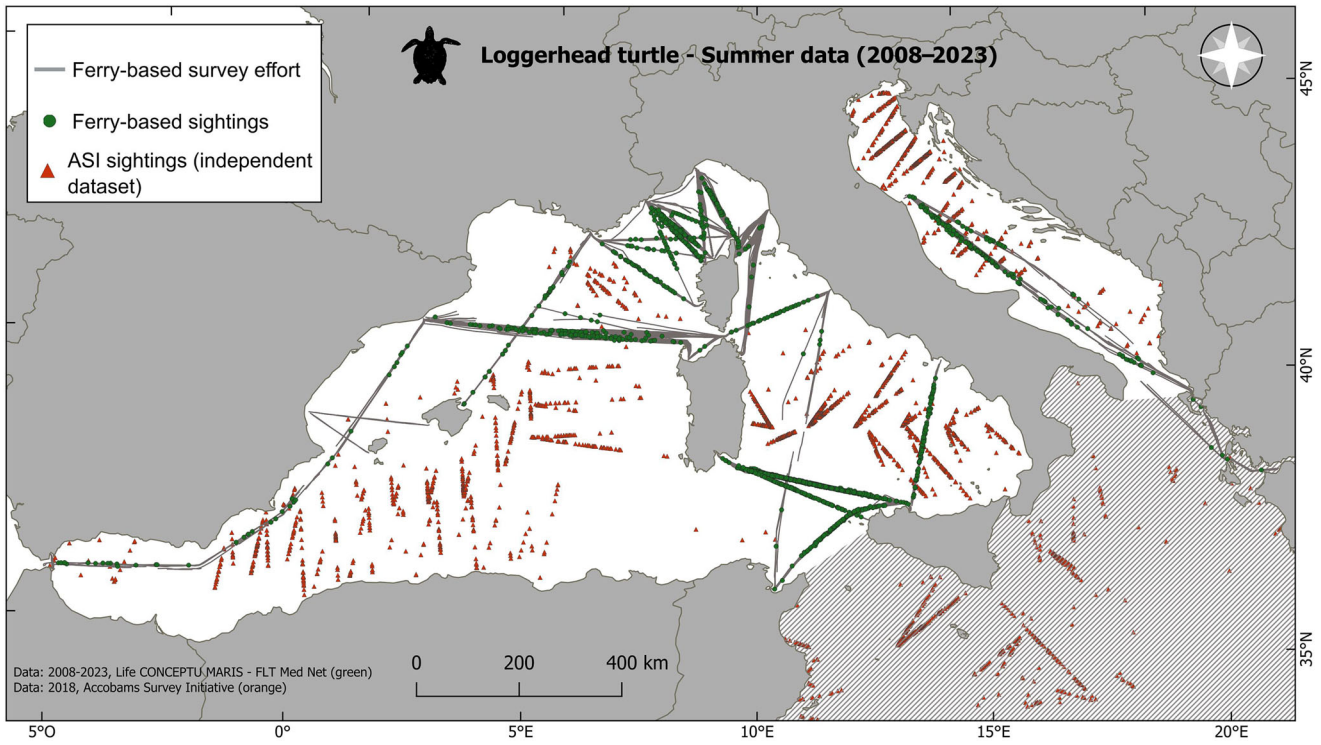

FIGURE 1 Summer loggerhead turtle sightings across the Mediterranean Sea. Gray lines show ferry-based survey effort along fixed transect routes (2008–2023), and green points indicate related sightings ($n = 2308$); orange triangles represent sightings from the independent ASI aerial survey conducted in July–September 2018 ($n = 2673$).

the spatial coverage of the datasets considered in this study, including those used for external model validation (Figure 1).

Preparation of environmental predictor variables

Environmental predictor variables were selected based on ecological relevance and spatial coverage across the study area and screened to ensure that surveyed conditions were representative of the environmental domain across the Mediterranean Sea. This screening step provided an additional verification beyond the a priori FLT survey design, which selected routes to capture major environmental gradients both within routes and at the basin scale.

Dynamic predictor variables were derived from Copernicus marine reanalysis products (2008–2022) and included sea surface temperature, salinity, chlorophyll a concentration, sea surface height, mixed layer depth, and surface ocean currents. From current velocity components, additional variables describing surface current magnitude and direction were derived, whereas thermal gradients were calculated from sea surface temperature by estimating *x*- and *y*-direction slopes and combining them into a single gradient magnitude.

Static predictor variables included bathymetry and slope (from GEBCO 2024), distance to coastline, and distances to geomorphological features such as seamounts and submarine canyons. Distances were calculated from European Marine Observation and Data Network (EMODnet) geology shapefiles, accounting for land barriers through least-cost path analysis. All raster layers were resampled to 5 × 5 km and 10 × 10 km European standard grids (ETRS-89, EPSG:3035). For dynamic predictors, monthly values were averaged over the summer period (July–September) over the full study period (2008–2022). Environmental predictor variables were thus structured to represent typical seasonal environmental conditions over the study period and to characterize recurring seasonal habitat associations (full details on data sources, native resolution, and variable derivation are provided in Appendix S1). Collinearity among predictor variables was assessed using pairwise correlation ($r > 0.7$) and variance inflation factor (VIF > 10) thresholds, above which variables were excluded. No ecological predictor variables exceeded these thresholds. Two environmental predictor variable sets were defined for model comparison. The base set (10 variables) consisted of chlorophyll a concentration, sea surface temperature, sea surface salinity, distance to canyons, distance to seamounts, distance to coastline, depth, latitude, longitude, and slope, representing the most commonly used predictors in marine SDMs. The extended set (base + five variables) contained surface thermal gradient, sea surface height, thermocline depth, and surface current magnitude and direction, which captured additional aspects of mesoscale oceanographic variability.

Benchmarking SDMs

We structured the analysis into three sequential modeling steps: spatial resolution, environmental predictor variables, and algorithm comparisons, and each was informed by the outcome of the previous one (Appendix S2).

First, to assess the influence of spatial resolution on SDM performance, we fitted GAMs at two spatial resolutions: 5 and 10 km. Presence records ($n = 2308$) were aligned with each grid and spatially aggregated (i.e., distilled to a single presence per grid cell), yielding 719 unique presences at 5 km and 460 at 10 km, to ensure consistency with temporally aggregated environmental predictor variables and prevent repeated sampling of identical environmental conditions. Pseudoabsences were sampled from surveyed grid cells with no sightings, ensuring that only areas with survey effort were included. A balanced 1:1 presence–pseudoabsence design was adopted following Barbet-Massin et al. (2012) to improve model stability and reduce class imbalance. GAMs were fitted independently at each resolution using a binomial error distribution with a logit link (0 = pseudoabsence; 1 = presence) and using the base set of environmental predictors. The significance of smooth terms was assessed using approximate chi-square tests, evaluating the null hypothesis that each predictor does not contribute to model fit, and reporting the associated p values. Model performance was evaluated using a two-stage internal validation framework consisting of a 75% stratified training set with 10-fold cross-validation and a withheld 25% test set. Performance metrics included area under the receiver operating characteristic curve (AUC; discrimination ability), accuracy (proportion of correct classifications), sensitivity (true positive rate), specificity (true negative rate), precision (positive predictive value), F1 score (harmonic mean of precision and sensitivity), and the true skill statistic (TSS; sensitivity + specificity − 1, a measure of overall classification skill). Based on the comparative performance across these metrics, subsequent modeling steps were conducted at the best-performing spatial resolution (5 km).

Second, to evaluate the influence of predictor variable complexity on SDM performance, models fit at the selected 5 km resolution were run using the two environmental predictor sets defined above (base and extended). Models followed the same spatial aggregation, pseudoabsence generation, and internal validation framework described above. This comparison allowed the assessment of how increasing environmental complexity affected model behavior while keeping spatial resolution and modeling structure constant. Based on this comparison, the extended predictor variable set was retained for subsequent modeling steps. To account for the Monte Carlo variability introduced by the 1:1 presence–pseudoabsence sampling, we performed a sensitivity analysis by repeating the modeling procedure for the selected best-performing configuration (5 km resolution with the extended predictor set) across 30 replicates using independently generated pseudoabsence sets, while keeping all other modeling parameters constant. This action was completed in preparation for the next step because variability in pseudoabsence sampling could influence model comparison and ranking.

Third, we compared five model algorithms, GLM, GAM, RF, BRT/GBM, and MaxEnt (classification only), across two modeling tasks, presence–pseudoabsence classification and ER regression. All models followed the internal 25:75 validation framework described above.

The extended predictor variable set was initially used for all algorithms. However, sensitivity tests indicated that including geographic coordinates (latitude/longitude) in flexible machine learning models (RF, BRT/GBM, and MaxEnt) generated spatial artifacts in Mediterranean-wide projections (e.g., sharp discontinuities and geographically constrained hotspots). This result likely reflected uneven survey effort across the Mediterranean, with sampling concentrated in the western and Adriatic regions and limited coverage in the eastern basin. Because the predictive performance on the withheld test set (measured by AUC for classification and root mean square error [RMSE] for regression) remained unchanged whether or not coordinates were included, latitude and longitude were excluded from final RF/BRT/MaxEnt projections to enhance transferability. In contrast, these coordinates were retained in GLM/GAM models.

For the classification models (MaxEnt, GLM, GAM, RF, and BRT/GBM), spatial aggregation and pseudoabsence generation followed the framework described at the first step. MaxEnt did not use pseudoabsences; instead, all surveyed grid cells (survey effort > 0) were used as background points sampled from the effort-based bias layer, ensuring a common sampling domain across algorithms. GLM and GAM were fit using standard implementations, whereas RF and BRT/GBM hyperparameters (e.g., mtry/maxnodes for RF and n.trees, interaction depth and learning rate for BRT) were tuned via grid search within cross-validation (Appendix S3). These parameters regulate model complexity and regularization by controlling tree number/depth and the rate and stochasticity of learning (i.e., the number of predictors sampled at each split, tree growth constraints, and subsampling/learning rate).

For the regression models (GLM, GAM, RF, and BRT), ER analyses were restricted to post-2013 data, when survey protocols were revised to ensure more consistent recording of observation effort, allowing reliable effort-standardized estimates. The 5 × 5 km grid cell represented the modeling unit, and ER was calculated as sightings per kilometer of survey effort. Only cells with ≥10 km of effort were retained, which was supported by an exploratory assessment showing stabilization of ER variance beyond this value (Appendix S4). Given strong zero inflation (~80% zero ER), GLM and GAM were fitted using a Tweedie distribution (variance power $p \approx 1.48$, log link). For the GLM and GAM calculations across all steps, model assumptions and fit were evaluated using standard diagnostic checks (including residual diagnostics and GAM smooth-term diagnostics). RF and BRT (XGBoost) were tuned via grid search, selecting configurations that minimized the RMSE (Appendix S5). Model performance was additionally summarized using mean absolute error (MAE) and R^2.

We used the selected environmental predictor set to project the final models (classification and regression) across the Mediterranean Sea. To prevent unrealistic extrapolation in Tweedie-based ER models, spatial projections and associated uncertainty layers were clipped to the observed training range (maximum = 0.2 sightings/km).

External validation was conducted using an independent dataset not involved in model calibration and derived from the ASI aerial surveys (see the "Data sources, model inputs, and independent data" section), which differed from ferry-based observations in terms of platform, sampling design, and spatial coverage and provided an explicit test of spatial transferability (i.e., the ability of models trained on one survey domain to produce reliable predictions when applied to a different spatial domain and sampling context). Predicted values were extracted at ASI sighting locations, and pseudoabsences were randomly sampled across the validation domain (entire Mediterranean basin) using a 1:1 ratio, excluding presence and missing prediction values. Multiple ASI sightings within the same 5 × 5 km cell were collapsed to a single presence for classification models; for ER models, duplicates were retained to reflect local encounter intensity. To enable validation against presence-only independent data and ensure comparability across modeling approaches, continuous model predictions were converted into binary presence–absence outputs using predefined thresholds. For classification models, thresholds were selected by maximizing the TSS for all algorithms except MaxEnt, which used the equal training sensitivity and specificity criterion, identified by Arcangeli et al. (2023) as a conservative and ecologically realistic threshold. For regression models, we used the first quartile (25th percentile) of nonzero ER values to convert the continuous ER predictions into binary presence–absence, thereby excluding very low ERs and identifying areas of consistent use. Model performance on the independent dataset was evaluated using the same discrimination metrics applied to the binary models and described in the first step. This approach allowed for a direct comparison between the internal model fit and external predictive performance. To account for the Monte Carlo variability introduced by the 1:1 presence–pseudoabsence sampling, the external validation process was repeated across 30 replicates for all classification models. Each replicate used a separately generated pseudoabsence dataset.

Across all modeling steps, model performance was evaluated using complementary diagnostic criteria, including indicators of overfitting, generalization, and transferability, in addition to the algorithm-specific tuning and regularization procedures described above. These additional diagnostics included unrealistically high internal performance (e.g., AUC = 1), internal performance gaps between training and withheld test data ($\Delta AUC_{int} = AUC_{train} - AUC_{test}$ for classification; $\Delta R^2 = {R^2}_{train} - {R^2}_{test}$ for regression), and screening of spatial projections to detect artifacts consistent with spatial overfitting (e.g., sharp discontinuities and geographically constrained hotspots). In addition, external transferability was evaluated (classification only) by comparing test set and independent validation performance ($\Delta AUC_{ext} = AUC_{test} - AUC_{ind}$).

Software and packages

All analyses were conducted in R v4.2.0. The following R packages were used throughout the modeling workflow: sf, sp, terra, and raster (spatial data preprocessing, extraction, and mapping); usdm (collinearity diagnostics using vifcor() and vstep()); caret (model training, cross-validation, and hyperparameter tuning); stats (GLMs and prediction functions); mgcv (GAMs and diagnostic checks using gam.check()); randomForest (RF classification models); ranger (RF regression models); gbm (BRT/GBM classification models); xgboost (gradient boosting models implemented via caret xgbTree); dismo with MaxEnt v3.4.x (Java implementation; MaxEnt); and DHARMa (simulation-based residual diagnostics).

RESULTS

Effects of spatial resolution and predictor variable sets on presence–pseudoabsence GAM performance

Model performance was moderately higher and more consistent at the 5 km resolution than at the 10 km resolution (Table 1; Appendix S5). At 5 km, the GAM achieved similar discrimination between training and test datasets (AUC ≈ 0.78–0.79) and explained up to ~26% of deviance, whereas models fit at 10 km showed slightly lower discrimination and explanatory power (AUC ≈ 0.71–0.77; deviance explained ~22%).

Predicted habitat suitability patterns were broadly consistent across spatial resolutions, with both configurations identifying similar core areas of high occurrence probability. However, models at the 5 km resolution captured finer spatial structures, including coastal features and localized hotspots, whereas patterns at the 10 km resolution were smoother and less spatially detailed (Appendix S6). Predictor responses were also more differentiated at 5 km than at 10 km (Appendix S7). Based on these results, the 5 km resolution was retained for subsequent analyses.

At 5 km, expanding the predictor set from the base to the extended configuration resulted in limited changes in model performance (Table 1). The deviance explained increased from 26.1% to 28.1%, whereas AUC values remained comparable between configurations. Despite similar performance metrics, predicted spatial patterns differed between models (Appendix S6b,c). The extended model produced more continuous and spatially structured suitability patterns, particularly in offshore regions and along major circulation features, whereas the base model showed more fragmented patterns.

The contribution of predictor variables to model fit, assessed using chi-square tests against the null hypothesis of no effect, differed between configurations (Figure 2; Appendix S7). Sea surface salinity and distance to coastline remained important predictors in both models. In the extended model, additional dynamic variables, including sea surface height, thermal gradients, mixed-layer properties, and surface currents, showed

TABLE 1 Comparison among generalized additive models (GAMs) tested at 5 and 10 km spatial resolutions with the base and extended predictor set.

Modeling component	Model/algorithm	Internal performance
PA models[a]	GAM—10 km (base)	$AUC_{train} = 0.71$; $AUC_{test} = 0.77$; dev.exp = 22%
PA models	GAM—5 km (base)	$AUC_{train} = 0.77$; $AUC_{test} = 0.79$; dev.exp = 26%
PA models	GAM—5 km (extended)	$AUC_{train} = 0.77$; $AUC_{test} = 0.78$; dev.exp = 28%

[a]Presence–pseudoabsence (PA) models were evaluated using the area under the curve for training data (AUC_{train}) and test data (AUC_{test}) and deviance explained (dev.expl.).

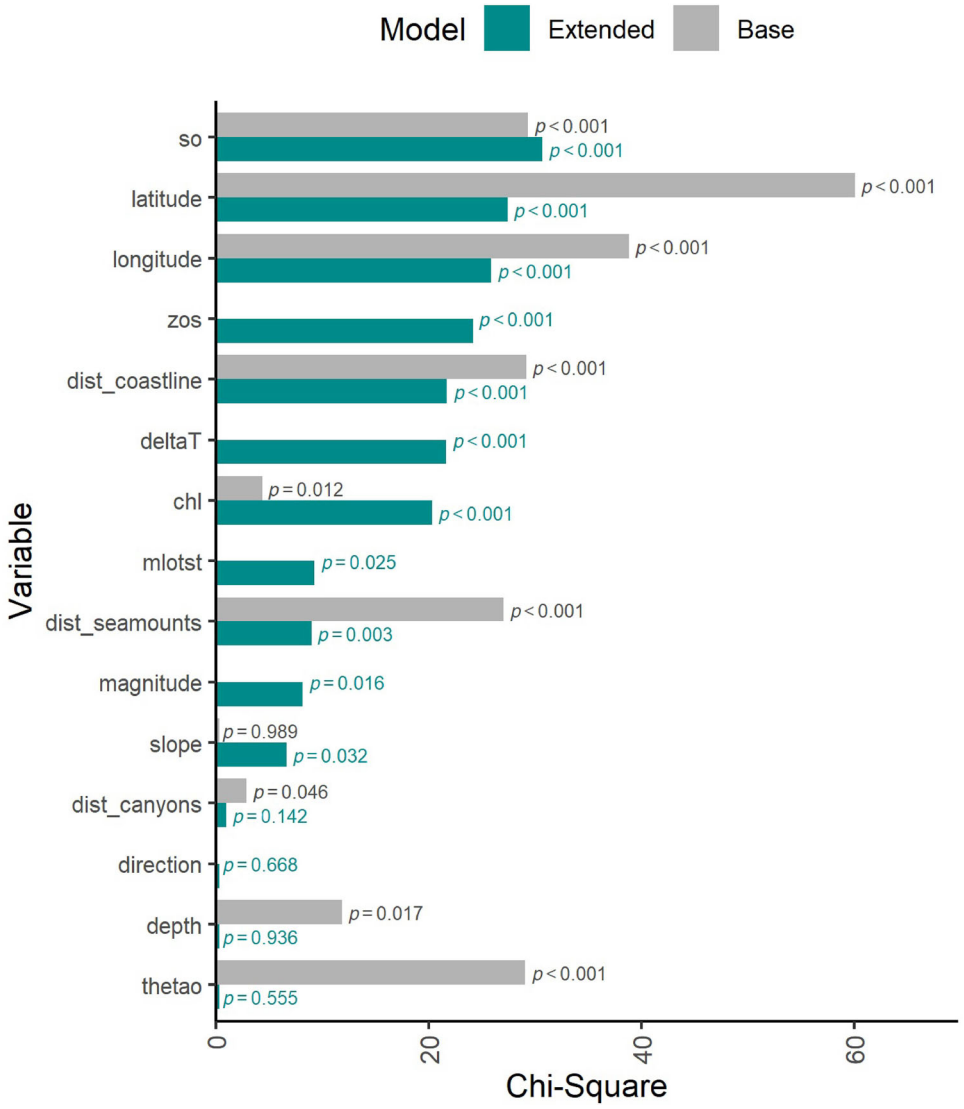

FIGURE 2 Contribution of each predictor to model fit in summer for loggerhead turtles in the Mediterranean Sea, comparing the base and extended GAMs. Bars represent approximate chi-square statistics for smooth terms. Exact *p* values from tests of each smooth term are reported to the right of the bars. so, salinity; zos, sea surface height; deltaT, surface thermal gradient; chl, chlorophyll a concentration; mlotst, mixed layer depth; magnitude, surface current magnitude; thetao, sea surface temperature; direction, current direction (north-referenced).

considerable contributions. In contrast, several bathymetric variables (e.g., depth and distance to canyons) and sea surface temperature did not show a significant effect. The relative contribution of spatial coordinates also decreased compared with that in the base model.

Response curves from the extended model indicated that habitat suitability was associated with restricted ranges of environmental conditions (Appendix S8), consistent with differences between presence and available conditions (Appendix S9). Suitability peaked at intermediate salinity values (approximately 37.2–37.8 PSU) and moderate levels of productivity and mesoscale activity. Effects of dynamic predictor variables were concentrated within narrow ranges, including sea surface height (approximately −0.42 to −0.33 m), chlorophyll a (≈0.12 mg/m^3), thermal gradients (≈0.0045°C), and current magnitude (≈0.4–0.6 m/s). Sea surface temperature did not

TABLE 2 Comparison of internal and external species distribution model (SDM) performance for presence–pseudoabsence (PA) and encounter rate (ER) models fit with the extended predictor set.

Modeling component	Model/algorithm	Internal performance[a]	External validation[b]
PA models	GAM—5 km (extended)	$AUC_{train} = 0.77$; $AUC_{test} = 0.78$; $\Delta AUC_{int} = -0.01$	$AUC_{ind} = 0.67$; $\Delta AUC_{ext} = 0.11$
PA models	GLM—5 km (extended)	$AUC_{train} = 0.67$; $AUC_{test} = 0.68$; $\Delta AUC_{int} = 0.01$	$AUC_{ind} = 0.57$; $\Delta AUC_{ext} = 0.11$
PA models	RF—5 km (extended)	$AUC_{train} = 0.77$; $AUC_{test} = 0.78$; $\Delta AUC_{int} = 0.01$	$AUC_{ind} = 0.51$; $\Delta AUC_{ext} = 0.27$
PA models	BRT– 5 km (extended)	$AUC_{train} = 0.76$; $AUC_{test} = 0.77$; $\Delta AUC_{int} = 0.01$	$AUC_{ind} = 0.54$; $\Delta AUC_{ext} = 0.23$
PA models	MaxEnt—5 km (extended)	$AUC_{train} = 0.75$; $AUC_{test} = 0.72$; $\Delta AUC_{int} = 0.03$	$AUC_{ind} = 0.71$; $\Delta AUC_{ext} = 0.01$
ER models	GLM– 5 km (extended)	$R^2_{train} = 0.15$; $R^2_{test} = 0.23$; $\Delta R^2 = -0.08$	$AUC_{ind} = 0.62$
ER models	RF—5 km (extended)	$R^2_{train} = 0.33$; $R^2_{test} = 0.44$; $\Delta R^2 = -0.11$	$AUC_{ind} = 0.61$
ER models	BRT—5 km (extended)	$R^2_{train} = 0.29$; $R^2_{test} = 0.37$; $\Delta R^2 = -0.008$	$AUC_{ind} = 0.60$
ER models	GAM—5 km (extended)	$R^2_{train} = 0.25$; $R^2_{test} = 0.38$; $\Delta R^2 = -0.13$	$AUC_{ind} = 0.74$

[a]Model performance was evaluated using the area under the curve for the training data (AUC_{train}) and test data (AUC_{test}) and for the independent dataset based on independent ACCOBAMS Survey Initiative (ASI) aerial survey validation (AUC_{ind}), and overfitting was assessed using ΔAUC_{int} ($AUC_{train} - AUC_{test}$).
[b]Transferability was quantified using ΔAUC_{ext} ($AUC_{test} - AUC_{ind}$).

retain an independent effect in the extended model (edf ≈ 0; $p = 0.55$).

Algorithm performance for presence–pseudoabsence and ER models

Algorithm performance varied significantly between internal validation and testing on the independent ASI dataset, leading to a notable shift in algorithm ranking.

For presence–pseudoabsence models, RF, BRT, and GAM achieved similarly high discrimination during internal validation, but their performances decreased when applied to independent aerial survey data (Table 2; Appendix S10). In contrast, MaxEnt demonstrated the best performance under external validation (AUC = 0.71), even though it had slightly lower internal discrimination.

For the ER models, GAMs showed consistent performance between internal and external validation and explained up to 41% of deviance. Although RF and BRT achieved higher goodness-of-fit values during model training, their performance decreased with independent data (Table 2; Appendix S11). GAMs yielded the highest external discrimination (AUC = 0.74; Table 2). GLMs consistently showed weaker performance across both modeling frameworks.

A sensitivity analysis of model performance to pseudoabsence sampling revealed moderate variability in internal testing performance (AUC SD ≈ 0.025), whereas variability in independent validation was low across all models (AUC SD ≈ 0.006–0.012), with model ranking remaining consistent across replicates. This result indicated that stochastic variation in pseudoabsence selection had a limited influence on model comparison.

Models were calibrated and evaluated at the basin scale (Figure 3a,b). Spatial uncertainty associated with model predictions highlighted areas of greater variability and reduced model confidence (Figure 3). Standard diagnostic checks for GLM and GAM models were also conducted (Appendix S12, with representative diagnostic plots shown in Appendix S13).

DISCUSSION

SDM performance and transferability

Our results provided empirical evidence on how spatial resolution, environmental predictor variable selection, and algorithm choice jointly shaped model performance for offshore pelagic systems and on where data limitations and environmental variability pose additional challenges. Using a standardized 15-year dataset from ferry-based surveys across the Mediterranean Sea, we evaluated how common SDM choices affected both predictive performance and ecological interpretability.

Taken together, our findings showed that in offshore Mediterranean pelagic systems, increased spatial resolution improved ecological interpretability even when gains in discrimination were modest; dynamic oceanographic predictor variables were essential for capturing habitat variability; independent validation was critical for identifying robust and transferable modeling algorithms; and combining presence-only and ER models helped distinguish general habitat suitability from observation intensity, providing complementary information on use patterns.

Spatial resolution emerged as a key determinant of SDM outcomes, influencing both the spatial structure and ecological interpretability of model predictions. Although the 5 × 5 km and 10 × 10 km models performed similarly for standard accuracy metrics, the smaller grid yielded more spatially coherent and ecologically interpretable predictions, especially along bathymetric and coastal gradients. This pattern is consistent with previous evidence showing that resolution choice in presence-only and presence–pseudoabsence models can influence habitat inference and downstream conservation decisions, even when overall model accuracy appears stable (Elith &

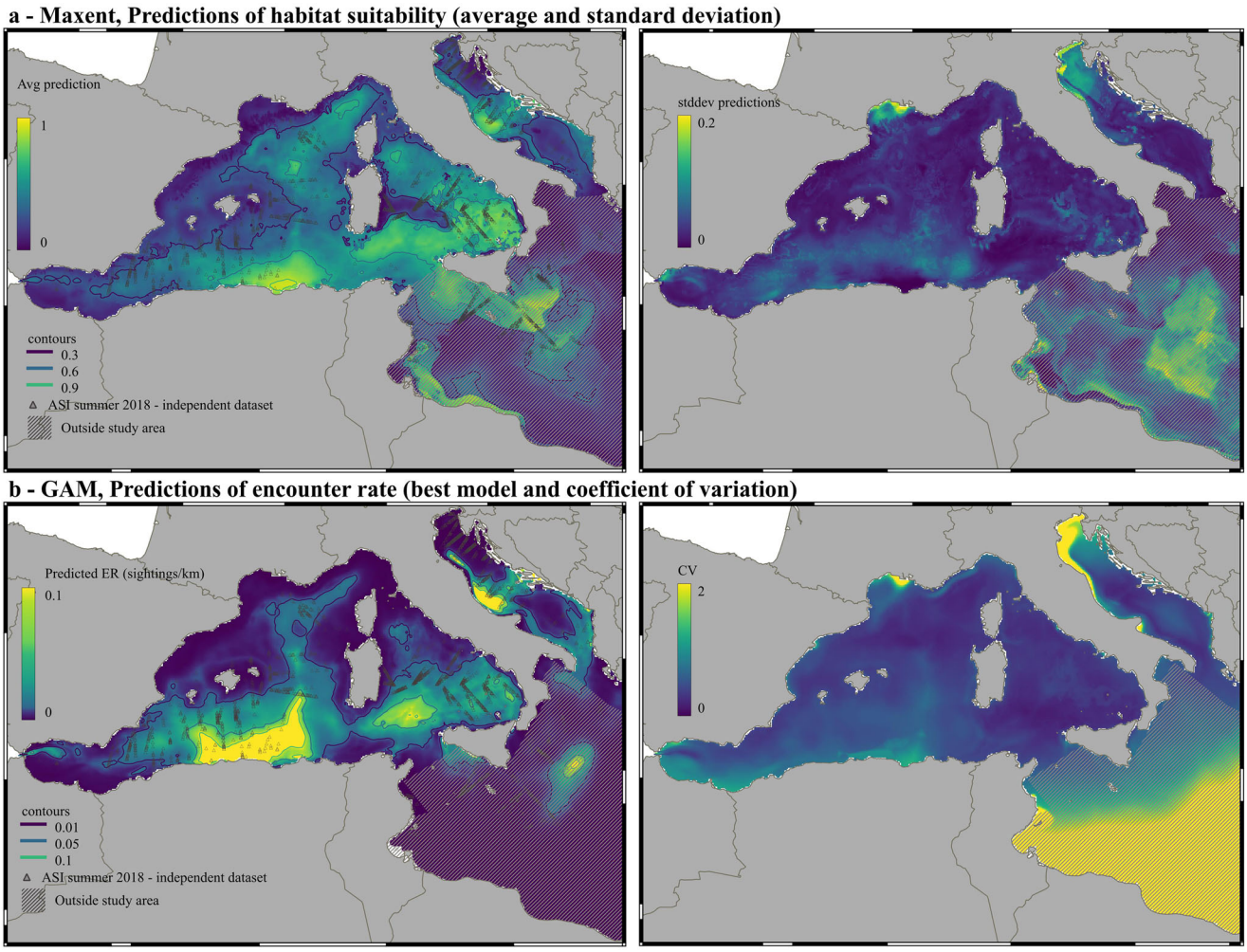

FIGURE 3 Summer predictions of loggerhead turtle (a) habitat suitability (MaxEnt) and (b) encounter rate (GAM) in the Mediterranean Sea, showing mean predictions and associated uncertainty. Predictions are displayed in full color only where survey effort was available; black triangles indicate independent ASI sightings (2018). Contour lines indicate the binary thresholds used for external validation (MaxEnt: 0.30, equal training sensitivity and specificity; GAM ER: 0.01 sightings/km).

Leathwick, 2009). By retaining environmental heterogeneity and better matching the native resolution of oceanographic predictor variables, the 5 km resolution enhanced the representation of mesoscale features (e.g., fronts, eddies, and topographic discontinuities) known to structure marine megafauna distributions (Hays et al., 2003; Schofield et al., 2013). This pattern was further reinforced by the effort-informed pseudoabsence strategy, which constrained background sampling to surveyed grid cells, thus maintaining ecological interpretability. Although pseudoabsences were subsampled to maintain a 1:1 ratio with presences, smaller grid cells increased the pool of candidate pseudoabsence locations. Although even finer spatial resolutions may capture additional small-scale processes, a 5 km grid represents a practical compromise among ecological interpretability, predictor resolution, and sample size constraints in offshore systems. From a conservation perspective, resolution-driven differences in predicted habitat patterns are consequential: coarser models may smooth over localized hotspots, whereas finer predictions better match the spatial scales at which measures such as marine protected areas and spatial mitigation actions are implemented. This scenario reinforces the view that SDM selection should not only rely on statistical performance metrics but also consider ecological validity and conservation relevance (Guisan & Thuiller, 2005). For marine megafauna, spatial resolution should therefore be treated as a strategic modeling choice rather than a purely technical one (Fossette et al., 2014).

Beyond spatial resolution, predictor variable selection strongly influenced model interpretability and ecological relevance. Although simpler models are often favored in conservation applications because of their transparency and lower risk of overfitting (Burnham & Anderson, 2002; Zuur et al., 2009), they may overlook key drivers in systems governed by multiscale and nonlinear processes (Elith & Leathwick, 2009; Merow et al., 2014). In our case, restricting models to a limited set of static and dynamic variables risked underrepresenting ecological mechanisms in dynamic offshore environments, where mesoscale variability plays a central role. Incorporating ecologically justified, more complex oceanographic predictors allowed transient habitat features relevant to foraging and movement

to be represented explicitly, rather than inferred indirectly from spatial proxies. Variables reflecting hydrographic structure, productivity, and mesoscale circulation (e.g., salinity, chlorophyll a, sea surface height, mixed layer depth, current intensity, and thermal gradients) emerged as informative predictors, consistent with evidence that marine megafauna aggregate along frontal systems and retention zones (Luschi et al., 2013; Polovina et al., 2000; Revelles et al., 2007; Scales et al., 2017). As dynamic predictor variables were included, the explanatory role of spatial coordinates declined, suggesting that environmental heterogeneity was captured directly rather than through geographic surrogates. Static features such as distance to coast, canyons, and seamounts nevertheless remained influential, reflecting their indirect role in structuring prey aggregation and foraging opportunities (Hays et al., 2014; Revelles et al., 2007; Schofield et al., 2006). The feasibility of fitting an extended predictor variable was enabled by the size, spatial coverage, and temporal consistency of our dataset, allowing moderately complex and ecologically informed models to be calibrated without compromising stability. In data-poor systems, simpler calculations may remain more appropriate (Wenger & Olden, 2012; Yates et al., 2018). Overall, these results suggested that when collinearity and overfitting are adequately controlled (De Marco & Nóbrega, 2018; Dormann et al., 2013) and supported by independent validation, combining static and dynamic predictors can yield SDMs that are both ecologically realistic and operationally relevant for conservation planning.

In addition to spatial resolution and predictor variable selection, algorithm choice emerged as a key determinant of SDM robustness and transferability when performance was assessed using independent data. Although all tested algorithms achieved acceptable internal performance, clear differences emerged under external validation, highlighting the limitations of relying solely on internal validation procedures, such as cross-validation or bootstrapping, to assess generalization (Araújo & Guisan, 2006; Robinson et al., 2017). Under independent evaluation, MaxEnt and GAMs emerged as the most robust options for presence–pseudoabsence and ER modeling, respectively. The strong performance of MaxEnt likely reflects its flexible regularization framework and its implementation with an effort-based bias layer for background-point selection, which helped account for uneven sampling effort in a spatially heterogeneous survey domain, a key advantage in marine systems characterized by uneven survey coverage (Elith et al., 2011; Fourcade et al., 2014; Merow et al., 2013; Phillips et al., 2006). For ER models, GAM fit to effort-standardized data provided a favorable balance between flexibility and parsimony, capturing nonlinear ecological responses while limiting overfitting through penalized smoothers (Pedersen et al., 2019; Wood, 2017). Explained deviance values for the binary GAMs were moderate, whereas the Tweedie GAM achieved higher explanatory power (~40%), reflecting the different response structures and modeling objectives. In contrast, tree-based methods showed reduced transferability despite high internal performance, consistent with evidence that complex algorithms may overfit spatial structure rather than capture generalizable ecological relationships, particularly for species with diffuse habitat associations (Bahn & McGill, 2013). We found that including latitude and longitude in flexible ML algorithms increased the risk of spatial artifacts and reduced interpretability, likely due to geographically clustered sampling effort. Excluding coordinates improved the interpretability of basin-wide projections without affecting internal predictive performance, consistent with concerns about geolocation predictors in spatial ML applications (Meyer et al., 2019). The GLMs underperformed across both modeling tasks, reflecting their limited ability to represent nonlinear and scale-dependent ecological processes (Austin, 2002).

Importantly, algorithm performance remained context dependent. The relative success of MaxEnt and GAMs in this study likely reflected favorable conditions, including extensive spatial coverage, consistent long-term sampling, and a predictor set with low collinearity and high ecological relevance. Under different data constraints, ecological contexts, or sampling designs, algorithm rankings may have differed (Araújo & Guisan, 2006; Robinson et al., 2017). However, the discrepancy between internal and external validation highlighted the need for independent validation for evaluating SDM transferability, a practice that is still not widely adopted in marine SDM applications (Arcangeli et al., 2024; Pasanisi et al., 2024; Robinson et al., 2017). Together, these results provided a framework for interpreting the ecological patterns identified below.

Ecological findings

Our predictions described offshore pelagic habitat use and were primarily based on observations of late juvenile and subadult loggerhead turtles, which accounted for most records in the ferry-based survey dataset. These life stages are characterized by high mobility and ecological flexibility, and they typically occupy offshore foraging and migratory habitats (Casale et al., 2008; Chimienti et al., 2020; Hochscheid et al., 2010, 2019). Coastal and nearshore environments were not represented because they are governed by different environmental dynamics and life-stage-specific behaviors.

Within this context, the consistent identification of offshore areas of high suitability or recurrent use across modeling approaches, together with their confirmation through independent validation, supported their interpretation as functionally important habitats during the summer season. Several offshore regions, including the southern Tyrrhenian Sea, the Sardinian–Sicilian channels, the Adriatic Sea, and the Algerian Basin, were consistently highlighted across modeling approaches, and these results are consistent with those in previous studies (Almpanidou et al., 2022; Arcangeli et al., 2019; Chimienti et al., 2020; Di Matteo et al., 2022; Luschi et al., 2018; Mingozzi et al., 2016; Zampollo et al., 2022; Zbinden et al., 2008). These areas are associated with hydrographic configurations combining productivity, salinity, and mesoscale circulation features (e.g., sea surface height anomalies and intermediate current intensity), conditions that promote prey aggregation and retention. This interpretation is consistent with satellite tracking studies reporting increased loggerhead turtle foraging activity along eddies, frontal systems, and retention zones (Revelles et al., 2007).

In this context, the Algerian Basin cautiously emerged as a relevant summer turtle habitat despite the absence of direct survey coverage because this pattern was supported by independent validation and is consistent with previous evidence highlighting the importance of this area for loggerhead turtles (Cardona et al., 2024; Revelles et al., 2007). This finding highlights the value of integrating model predictions with independent validation to identify offshore areas of ecological importance, although future studies, including through direct survey efforts, should assess whether the relevance of this area is associated with particular life stages and remains stable across seasons.

Although ER models did not explicitly account for detectability bias, the use of standardized surveys along fixed routes reduced variability in observation conditions, allowing relative ERs to serve as a complementary proxy for space-use intensity. Interpreting suitability and ER models together therefore provided a more nuanced view of offshore habitat use, distinguishing areas of concentrated use from regions with more diffuse utilization (e.g., the Liguro–Provençal Basin).

Overall, the consistency between spatial patterns and underlying hydrographic drivers supported the interpretation that offshore habitat use is primarily driven by dynamic oceanographic processes rather than static geographic features. This finding is consistent with the role of mesoscale circulation features in structuring feeding opportunities and habitat selection in immature loggerhead turtles (Revelles et al., 2007), as well as with ecological evidence identifying adult oceanic foraging grounds in the western Mediterranean (Cardona et al., 2024). More broadly, the recurrence of these dynamic offshore areas suggests that they may represent functionally important regions for other large marine vertebrates. Although species-specific responses may differ, such areas likely support habitat use across multiple taxa, highlighting their relevance for integrated, multispecies conservation planning in dynamic marine environments.

Conservation implications

Building on modeling comparisons and ecological patterns, this study illustrated how a structured and context-aware modeling workflow, based on standardized and effort-corrected offshore surveys, can be used to generate robust and operationally relevant SDMs in dynamic pelagic systems, where probabilistic sampling is rarely feasible and data limitations are common. By explicitly integrating survey effort, spatial resolution, and ecologically meaningful predictors, this framework produced spatially explicit outputs that can support conservation decision-making beyond coastal waters. Importantly, the spatial patterns identified by this framework are broadly consistent with existing and emerging spatial protection and management initiatives in the Mediterranean, suggesting that model outputs can complement and refine ongoing conservation planning efforts rather than propose entirely new priorities. Model outputs can be used to delineate ecological ranges, inform indicators of distributional change, and contribute to basin-scale assessments of exposure to key anthropogenic pressures (e.g., marine traffic, bycatch, underwater noise, and climate-driven habitat shifts; Almpanidou et al., 2022; Chatzimentor et al., 2024; Mancino et al., 2022; Martino et al., 2021; Pace et al., 2022; Pasanisi et al., 2022; Stock et al., 2020) within existing Mediterranean conservation and monitoring frameworks (e.g., European Union Habitats Directive, Marine Strategy Framework Directive, and United Nations Environment Programme/Mediterranean Action Plan Integrated Monitoring and Assessment Programme).

Beyond the specific case of loggerhead turtles, the modeling framework developed here is transferable to other wide-ranging marine taxa and regions where comparable monitoring programs are in place. Observation schemes based on regular ferry or cargo routes can generate the consistent large-scale datasets required to apply this approach across ocean basins, as demonstrated by existing initiatives operating under analogous principles (e.g., ORCA in the North Atlantic and CETUS in Macaronesia; Campana & Vighi, 2021; Correia et al., 2015; ORCA, 2025). This scenario highlights the potential to apply the study framework more broadly to support cross-regional assessments of mobile marine species.

AUTHOR CONTRIBUTIONS

E.P. and A.A. contributed to conceptualization, methodology, and overall design of the study. A.A. provided scientific leadership, project coordination, funding acquisition, and supervision within the framework of the Life Conceptu Maris project. E.P. performed the formal data analysis and wrote the original draft of the manuscript. D.S.P. and M.V. provided scientific supervision within the PhD program Evolutionary and Environmental Biology at Sapienza University. All other authors contributed substantially to the long-term data collection, data harmonisation, and curation for their respective survey routes and monitoring programs. All authors contributed to reviewing and editing the manuscript and approved the final version.

ACKNOWLEDGMENTS

We gratefully acknowledge the collaboration of the ferry companies that facilitated onboard data collection, including Grimaldi Lines, Corsica Sardinia Ferries, Grandi Navi Veloci, Minoan Lines, Tirrenia, and Balearia. We also thank all institutions, observers, and researchers involved. We are particularly grateful to ACCOBAMS for providing the independent validation dataset. We also sincerely thank the reviewers for their constructive and insightful methodological and conceptual comments, which substantially improved the quality and robustness of this study. This study was supported by the LIFE Conceptu Maris project (LIFE20 NAT/IT/001371, 2022–2026) and FLT Med Net, both coordinated by ISPRA, which enabled long-term data collection from 2008 onward. Additional support was obtained from national and international projects on Mediterranean marine megafauna monitoring, including SICOMAR PLUS INTERREG Italy–France, funded by the Sanctuaire Pelagos through the Parc National de Port-Cros (Convention nos. 14–028, 12–036, and 11-011-83400); ACCOBAMS-funded ferry monitoring off Tunisia; and the

Monitoraggio dei cetacei nel Mar Tirreno centrale project, supported by Fondazione Cariciv.

Open access publishing facilitated by Universita degli Studi di Roma La Sapienza, as part of the Wiley - CRUI-CARE agreement.

DATA AVAILABILITY STATEMENT

To ensure transparency and reproducibility, all scripts used for data processing, modeling, and analysis, together with a subset of species data, are archived on Zenodo and are available at https://doi.org/10.5281/zenodo.19321489.

ORCID

E. Pasanisi https://orcid.org/0000-0001-9575-6748
F. Maffucci https://orcid.org/0000-0003-3078-6510
F. Atzori https://orcid.org/0000-0002-0279-5885
M. Azzolin https://orcid.org/0000-0002-4007-5010
I. Campana https://orcid.org/0000-0003-0801-0927
L. Carosso https://orcid.org/0009-0008-4950-8688
R. Crosti https://orcid.org/0000-0001-5149-2525
L. David https://orcid.org/0000-0003-3892-4420
M. Farina https://orcid.org/0009-0007-0291-4147
N. Fraija-Fernández https://orcid.org/0000-0001-6341-4276
O. Garcia-Garin https://orcid.org/0000-0002-0754-564X
M. Gazo https://orcid.org/0000-0003-1159-322X
M. Gregorietti https://orcid.org/0000-0002-7581-2818
F. Grossi https://orcid.org/0000-0003-1949-2802
S. Hochscheid https://orcid.org/0000-0002-8103-2236
R. Míguez-Lozano https://orcid.org/0000-0001-8716-3360
A. Moulins https://orcid.org/0000-0003-0843-1166
D. S. Pace https://orcid.org/0000-0001-5121-7080
J. A. Raga https://orcid.org/0000-0003-2687-2539
E. Santini https://orcid.org/0009-0004-7090-2648
G. Sarà https://orcid.org/0000-0002-7658-5274
A. Scuderi https://orcid.org/0000-0001-6915-7813
P. Tepsich https://orcid.org/0000-0003-0055-3740
M. Vitale https://orcid.org/0000-0002-3652-7029
M. Vighi https://orcid.org/0000-0003-3422-9041
A. Arcangeli https://orcid.org/0000-0003-2156-9987

REFERENCES

Almpanidou, V., Tsapalou, V., Chatzimentor, A., Cardona, L., Claro, F., Hostetter, P., Kaska, Y., Liu, W., Mansui, J., Miliou, A., Pietroluongo, G., Sacchi, J., Sezgin, Ç., Sözbilen, D., & Mazaris, A. D. (2022). Foraging grounds of adult loggerhead sea turtles across the Mediterranean Sea: Key sites and hotspots of risk. *Biodiversity and Conservation*, *31*(1), 143–160.

Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, *33*(10), 1677–1688.

Arcangeli, A., Atzori, F., Azzolin, M., Babey, L., Campana, I., Carosso, L., Crosti, R., Garcia-Garin, O., Gregorietti, M., Orasi, A., Scuderi, A., Tepsich, P., Vighi, M., & David, L. (2023). Testing indicators for trend assessment of range and habitat of low-density cetacean species in the Mediterranean Sea. *Frontiers in Marine Science*, *10*, Article 1116829.

Arcangeli, A., Azzolin, M., David, L., Farina, M., Fraija-Fernández, N., Gregorietti, M., Maffucci, F., Mancuso, F. P., Pasanisi, E., Pupillo, G., Rota, A., Roul, M., Sarà, G., Tepsich, P., Valsecchi, E., Belhadjer, A., Di-Méglio, N., Favaro, L., Giovannini, A., … Vighi, M. (2025). Identification of important offshore CEPTU areas and risk areas in Western Mediterranean (WMED) and Adriatic and Ionian (ADRION) marine regions. LIFE CONCEPTU MARIS CONservation of CEtaceans and Pelagic sea TUrtles in Med: Managing Actions for their Recovery In Sustainability – Deliverable C1. May 2025. https://www.lifeconceptu.eu/documenti-di-progetto/

Arcangeli, A., Azzolin, M., Babey, L., David, L., Garcia-Garin, O., Moulins, A., Rosso, M., Scuderi, A., Tepsich, P., Vighi, M., & Orasi, A. (2024). Looking for reliable species distribution models for low-density cetacean species: Compared effectiveness of SDMs for *G. griseus*, *G. melas*, *Z. cavirostris* in the Mediterranean Sea based on long-term fixed-transect data. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *34*(3), Article e4115.

Arcangeli, A., Azzolin, M., David, L., Mancuso, F. P., Tepsich, P., Valsecchi, E., Atzori, F., Carosso, L., Crosti, R., Giacoma, C., Gregorietti, M., Grossi, F., Maffucci, F., Roul, M., Orasi, A., Pasanisi, E., Santini, E., Favaro, L., Fraija, N., … Sarà, G. (2022). *First draft protocol data collection & analysis*. LIFE CONCEPTU MARIS: CONservation of CEtaceans and Pelagic sea Turtles in Med: Managing Actions for their Recovery In Sustainability—Deliverable A2. https://www.lifeconceptu.eu/wp-content/uploads/2022/12/1_Data-Collection-Data-analysis.pdf

Arcangeli, A., Babey, L., Azzolin, M., Campana, I., Carosso, L., Costantino, M., Crisan, M.-M., Crosti, R., David, L., Di Stefano, C., Di-Meglio, N., Dudley, R., Falconi, M., Farina, M., Fraija-Fernández, N., Garcia Garin, O., Gregorietti, M., Madonia, A., Correia, A. M., … Zanderink, F. (2026). Handbook: Monitoring marine macro-fauna using ferries/large vessels as multidisciplinary research platforms. ISPRA Manuali e linee guida 215/2026, 70 pp, Roma.

Arcangeli, A., Maffucci, F., Atzori, F., Azzolin, M., Campana, I., Carosso, L., Crosti, R., Frau, F., David, L., Di-Méglio, N., Roul, M., Gregorietti, M., Mazzucato, V., Pellegrino, G., Giacoletti, A., Paraboschi, M., Zampollo, A., de Lucia, G., & Hochscheid, S. (2019). Turtles on the trash track: Loggerhead turtles exposed to floating plastic in the Mediterranean Sea. *Endangered Species Research*, *40*, 107–121.

Arcangeli, A., Marini, L., & Crosti, R. (2013). Changes in cetacean presence, relative abundance and distribution over 20 years along a trans-regional fixed line transect in the Central Tyrrhenian Sea. *Marine Ecology*, *34*(1), 112–121.

Austin, M. P. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling*, *157*(2-3), 101–118.

Bahn, V., & McGill, B. J. (2013). Testing the predictive performance of distribution models. *Oikos*, *122*(3), 321–331.

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many?. *Methods in Ecology and Evolution*, *3*(2), 327–338.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Campana, I., & Vighi, M. (2021). *Monitoring large marine vertebrates along fixed transects from ferries and cargo vessels: A state of the art* (Report number: 2020/17). IMPEL.

Cardona, L., Abalo-Morla, S., Cani, A., Feliu-Tena, B., Izaguirre, N., Tomás, J., & Belda, E. J. (2024). Identifying the foraging grounds of the new loggerhead turtle nesters in the western Mediterranean. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *34*(1), Article e4059. https://doi.org/10.1002/aqc.4059

Casale, P. (2011). Sea turtle by-catch in the Mediterranean. *Fish and Fisheries*, *12*(3), 299–316.

Casale, P., Abbate, G., Freggi, D., Conte, N., Oliverio, M., & Argano, R. (2008). Foraging ecology of loggerhead sea turtles (*Caretta caretta*) in the central Mediterranean Sea: Evidence for a relaxed life history model. *Marine Ecology Progress Series*, *372*, 265–276.

Casale, P., Affronte, M., Insacco, G., Freggi, D., Vallini, C., Pino d'Astore, P., Basso, R., Paolillo, G., Abbate, G., & Argano, R. (2018). Sea turtle strandings reveal high anthropogenic mortality in Italian waters. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *28*(3), 467–481.

Casale, P., Hochscheid, S., Kaska, Y., & Panagopoulou, A. (2020). Sea turtles in the Mediterranean region: MTSG annual regional report 2020.

Cerritelli, G., Casale, P., Sözbilen, D., Hochscheid, S., Luschi, P., & Kaska, Y. (2022). Multidirectional migrations from a major nesting area in Turkey support the widespread distribution of foraging sites for loggerhead turtles in the Mediterranean. *Marine Ecology Progress Series*, *683*, 169–177.

Chatzimentor, A., Doxa, A., Butenschön, M., Kristiansen, T., Peck, M. A., Katsanevakis, S., & Mazaris, A. D. (2024). Diving into warming oceans: Assessing 3D climatically suitable foraging areas of loggerhead sea turtles under climate change. *Journal for Nature Conservation*, *79*, Article 126620.

Chimienti, M., Blasi, M. F., & Hochscheid, S. (2020). Movement patterns of large juvenile loggerhead turtles in the Mediterranean Sea: Ontogenetic space use in a small ocean basin. *Ecology and Evolution*, *10*(14), 6978–6992.

Correia, A. M., Tepsich, P., Rosso, M., Caldeira, R., & Sousa-Pinto, I. (2015). Cetacean occurrence and spatial distribution: Habitat modelling for offshore waters in the Portuguese EEZ (NE Atlantic). *Journal of Marine Systems*, *143*, 73–85. https://doi.org/10.1016/j.jmarsys.2014.10.016

De Marco, P., & Nóbrega, C. C. (2018). Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS ONE*, *13*(9), Article e0202403.

DiMatteo, A., Cañadas, A., Roberts, J., Sparks, L., Panigada, S., Boisseau, O., Moscrop, A., Fortuna, C. M., Lauriano, G., Holcer, D., Peltier, H., Ridoux, V., Raga, J. A., Tomás, J., Broderick, A. C., Godley, B. J., Haywood, J., March, D., Snape, R., … Hochscheid, S. (2022). Basin-wide estimates of loggerhead turtle abundance in the Mediterranean Sea derived from line transect surveys. *Frontiers in Marine Science*, *9*, Article 930412.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., 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*(1), 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.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57.

Fossette, S., Witt, M. J., Miller, P., Nalovic, M. A., Albareda, D., Almeida, A. P., Broderick, A. C., Chacón-Chaverri, D., Coyne, M. S., Domingo, A., Eckert, S., Evans, D., Fallabrino, A., Ferraroli, S., Formia, A., Giffoni, B., Hays, G. C., Hughes, G., Kelle, L., … Godley, B. J. (2014). Pan-Atlantic analysis of the overlap of a highly migratory species, the leatherback turtle, with pelagic longline fisheries. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1780), Article 20133065.

Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS ONE*, *9*(5), Article e97122.

Grossi, F., Hazen, E. L., Leo, G. D., David, L., Di-Méglio, N., Arcangeli, A., Pasanisi, E., Campana, I., Paraboschi, M., Castelli, A., Rosso, M., Moulins, A., & Tepsich, P. (2025). Evaluating three modelling frameworks for assessing changes in fin whale distribution in the Mediterranean Sea. *Ecology and Evolution*, *15*(3), Article e71007.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009.

Hays, G. C., Broderick, A. C., Godley, B. J., Luschi, P., & Nichols, W. J. (2003). Satellite telemetry suggests high levels of fishing-induced mortality in marine turtles. *Marine Ecology Progress Series*, *262*, 305–309.

Hays, G. C., Christensen, A., Fossette, S., Schofield, G., Talbot, J., & Mariani, P. (2014). Route optimisation and solving Zermelo's navigation problem during long distance migration in cross flows. *Ecology Letters*, *17*(2), 137–143.

Hochscheid, S., Aksissou, M., Arapis, T., Benabdi, M., Boura, L., Broderick, A., Cardona, L., Carreras, C., Claro, F., Demetropoulos, A., Fuller, W. J., Jribi, I., Kaska, Y., Levy, Y., Maffucci, F., Margaritoulis, D., Mifsud, C., Panagopoulou, A., Sacchi, J., … Rees, A. (2019). Sea turtles of the Mediterranean Sea. SWOT Report, 14, 20–29.

Hochscheid, S., Bentivegna, F., Hamza, A., & Hays, G. C. (2010). When surfacers do not dive: Multiple significance of extended surface times in marine turtles. *Journal of Experimental Biology*, *213*(8), 1328–1337.

Luschi, P., Mencacci, R., Cerritelli, G., Papetti, L., & Hochscheid, S. (2018). Large-scale movements in the oceanic environment identify important foraging areas for loggerheads in central Mediterranean Sea. *Marine Biology*, *165*(1), Article 4.

Luschi, P., Mencacci, R., Vallini, C., Ligas, A., Lambardi, P., & Benvenuti, S. (2013). Long-term tracking of adult loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. *Journal of Herpetology*, *47*(2), 227–231.

Mancino, C., Canestrelli, D., & Maiorano, L. (2022). Going west: Range expansion for loggerhead sea turtles in the Mediterranean Sea under climate change. *Global Ecology and Conservation*, *38*, Article e02264.

Martino, S., Pace, D. S., Moro, S., Casoli, E., Ventura, D., Frachea, A., Silvestri, M., Arcangeli, A., Giacomini, G., Ardizzone, G., & Jona Lasinio, G. (2021). Integration of presence-only data from several data sources. A case study on dolphins' spatial distribution. *Ecography*, *44*(10), 1533–1543.

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., Smith, M. J., Edwards, T. C., Jr., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R., Zimmermann, N. E., & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, *37*(12), 1267–1281.

Merow, C., Smith, M. J., & Silander, J. A., Jr. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069.

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications–Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*, Article 108815.

Mingozzi, T., Mencacci, R., Cerritelli, G., Giunchi, D., & Luschi, P. (2016). Living between widely separated areas: Long-term monitoring of Mediterranean loggerhead turtles sheds light on cryptic aspects of females spatial ecology. *Journal of Experimental Marine Biology and Ecology*, *485*, 8–17.

ORCA. (2025). *The state of Cetaceans*. Author.

Pace, D. S., Panunzi, G., Arcangeli, A., Moro, S., Jona-Lasinio, G., & Martino, S. (2022). Seasonal distribution of an opportunistic apex predator (*Tursiops truncatus*) in marine coastal habitats of the Western Mediterranean Sea. *Frontiers in Marine Science*, *9*, Article 939692.

Panigada, S., Pierantonio, N., Araújo, H., David, L., Di-Méglio, N., Dorémus, G., Gonzalvo, J., Holcer, D., Laran, S., Lauriano, G., Paiu, R.-M., Perri, M., Popov, D., Ridoux, V., Vázquez, J. A., & Cañadas, A. (2024). The ACCOBAMS Survey Initiative: The first synoptic assessment of cetacean abundance in the Mediterranean Sea through aerial surveys. *Frontiers in Marine Science*, *10*, Article 1270513. https://doi.org/10.3389/fmars.2023.1270513

Pasanisi, E., Chimienti, M., Blasi, M. F., Maffucci, F., & Hochscheid, S. (2022). Ocean highways in the Western Mediterranean: Which are the areas with increased exposure to maritime traffic for loggerhead turtles? *Frontiers in Marine Science*, *9*, Article 924532.

Pasanisi, E., Pace, D. S., Orasi, A., Vitale, M., & Arcangeli, A. (2024). A global systematic review of species distribution modelling approaches for cetaceans and sea turtles. *Ecological Informatics*, *82*, Article 102700.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, Article e6876.

Perez, M. A., Limpus, C. J., Hofmeister, K., Shimada, T., Strydom, A., Webster, E., & Hamann, M. (2022). Satellite tagging and flipper tag recoveries reveal migration patterns and foraging distribution of loggerhead sea turtles (*Caretta caretta*) from eastern Australia. *Marine Biology*, *169*(6), Article 80.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3-4), 231–259.

Polovina, J. J., Kobayashi, D. R., Parker, D. M., Seki, M. P., & Balazs, G. H. (2000). Turtles on the edge: Movement of loggerhead turtles (*Caretta caretta*) along oceanic fronts, spanning longline fishing grounds in the central North Pacific, 1997–1998. *Fisheries Oceanography*, *9*, 71–82.

Revelles, M., Isern-Fontanet, J., Cardona, L., San Félix, M., Carreras, C., & Aguilar, A. (2007). Mesoscale eddies, surface circulation and the scale of habitat selection by immature loggerhead sea turtles. *Journal of Experimental Marine Biology and Ecology*, *347*, 41–57.

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*(6), 789–802.

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E., & Lundquist, C. J. (2017). A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Frontiers in Marine Science*, *4*, 421.

Scales, K. L., Hazen, E. L., Jacox, M. G., Edwards, C. A., Boustany, A. M., Oliver, M. J., & Bograd, S. J. (2017). Scale of inference: On the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. *Ecography*, *40*(1), 210–220.

Schofield, G., Hobson, V. J., Fossette, S., Lilley, M. K., Katselidis, K. A., & Hays, G. C. (2010). Biodiversity Research: Fidelity to foraging sites, consistency of migration routes and habitat modulation of home range by sea turtles. *Diversity and Distributions*, *16*(5), 840–853.

Schofield, G., Katselidis, K. A., Dimopoulos, P., Pantis, J. D., & Hays, G. C. (2006). Behaviour analysis of the loggerhead sea turtle *Caretta caretta* from direct in-water observation. *Endangered Species Research*, *2*, 71–79.

Schofield, G., Scott, R., Dimadi, A., Fossette, S., Katselidis, K. A., Koutsoubas, D., Lilley, M. K. S., Pantis, J. D., Karagouni, A. D., & Hays, G. C. (2013). Evidence-based marine protected area planning for a highly mobile endangered marine vertebrate. *Biological Conservation*, *161*, 101–109.

Stock, B. C., Ward, E. J., Eguchi, T., Jannot, J. E., Thorson, J. T., Feist, B. E., & Semmens, B. X. (2020). Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. *Canadian Journal of Fisheries and Aquatic Sciences*, *77*(1), 146–163.

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs*, *92*(1), Article e01486.

Wallace, B., Bandimere, A., Abreu-Grobois, F., Acosta, H., Akiti, J., Akomedi, M., Alfaro-Shigueto, J., Allen, C., Angenda, D., Ayissi, I., Azanza Ricardo, J., Barrientos-Muñoz, K., Barrios-Garrido, H., Bjorndal, K., Bretón Vargas, E., Broderick, A., Calderón Peña, R., Carreras, C., Ceriani, S., … Mast, R. (2025). Updated global conservation status and priorities for marine turtles. *Endangered Species Research*, *56*, 247–276.

Wallace, B. P., DiMatteo, A. D., Bolten, A. B., Chaloupka, M. Y., Hutchinson, B. J., Abreu-Grobois, F. A., Mortimer, J. A., Seminoff, J. A., Amorocho, D., Bjorndal, K. A., Bourjea, J., Bowen, B. W., Briseño Dueñas, R., Casale, P., Choudhury, B. C., Costa, A., Dutton, P. H., Fallabrino, A., Finkbeiner, E. M., … Mast, R. B. (2011). Global conservation priorities for marine turtles. *PLoS ONE*, *6*(9), Article e24510.

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, *3*(2), 260–267.

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., … Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, *33*(10), 790–802.

Zampollo, A., Arcangeli, A., Costantino, M., Mancino, C., Crosti, R., Pietroluongo, G., Giacoma, C., & Azzolin, M. (2022). Seasonal niche and spatial distribution modelling of the loggerhead (*Caretta caretta*) in the Adriatic and Ionian seas. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *32*(7), 1141–1155.

Zbinden, J. A., Aebischer, A., Margaritoulis, D., & Arlettaz, R. (2008). Important areas at sea for adult loggerhead sea turtles in the Mediterranean Sea: Satellite tracking corroborates findings from potentially biased sources. *Marine Biology*, *153*(5), 899–906.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Zero-truncated and zero-inflated models for count data. In *Mixed effects models and extensions in ecology with R* (pp. 261–293). Springer.

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: Pasanisi, E., Maffucci, F., Atzori, F., Azzolin, M., Campana, I., Carosso, L., Castelli, A., Costantino, M., Crosti, R., David, L., Di-Meglio, N., Farina, M., Fraija-Fernández, N., Garcia-Garin, O., Gazo, M., Gregorietti, M., Grossi, F., Hochscheid, S., Míguez-Lozano, R., … Arcangeli, A. (2026). Comparative species distribution model framework for marine conservation and its application to loggerhead turtles in the Mediterranean. *Conservation Biology*, e70347. https://doi.org/10.1111/cobi.70347