



Vulnerability of the Critically Endangered leatherback turtle to fisheries bycatch in the eastern Pacific Ocean. I. A machine-learning species distribution model

Jon Lopez^{1,*}, Shane Griffiths¹, Bryan P. Wallace^{2,3,4,*}, Verónica Cáceres⁴, Luz Helena Rodríguez⁴, Marino Abrego⁵, Joanna Alfaro-Shigueto^{6,7,8}, Sandra Andraka⁹, María José Brito¹⁰, Leslie Camila Bustos¹¹, Iliá Cari¹², José Miguel Carvajal¹³, Ljubitzá Clavijo¹², Luis Cocas¹¹, Nelly de Paz¹⁴, Marco Herrera¹⁰, Jeffrey C. Mangel^{7,8}, Miguel Pérez-Huaripata¹⁵, Rotney Piedra¹⁶, Javier Antonio Quiñones Dávila¹⁵, Liliana Rendón⁹, Juan M. Rguez-Baron^{17,18}, Heriberto Santana¹⁹, Jenifer Suárez²⁰, Callie Veelenturf²¹, Rodrigo Vega¹², Patricia Zárate¹²

¹Inter-American Tropical Tuna Commission, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA

Full author addresses are given in the Appendix

ABSTRACT: The Eastern Pacific population of leatherback turtles *Dermochelys coriacea* is Critically Endangered, with incidental capture in coastal and pelagic fisheries as one of the major causes. Given the population's broad geographic range, status, and extensive overlap with fisheries throughout the region, identifying areas of high importance is essential for effective conservation and management. In this study, we created a machine-learning species distribution model trained with remotely sensed environmental data and fishery-dependent leatherback presence ($n = 1088$) and absence data ($> 500\,000$ fishing sets with no turtle observations) from industrial and small-scale fisheries that operated in the eastern Pacific Ocean between 1995 and 2020. The data were obtained through a participatory collaboration between the Inter-American Convention for the Protection and Conservation of Sea Turtles and the Inter-American Tropical Tuna Commission as well as non-governmental organizations to support the quantification of leatherback vulnerability to fisheries bycatch. A daily process was applied to predict the probability of leatherback occurrence as a function of dynamic and static environmental covariates. Coastal areas throughout the region were highlighted as important habitats, particularly highly productive feeding areas over the continental shelf of Ecuador, Peru, and offshore from Chile, and breeding areas off Mexico and Central America. Our model served as the basis to quantify leatherback vulnerability to fisheries bycatch and the potential efficacy of conservation and management measures (Griffiths & Wallace et al. 2024; Endang Species Res 53:295–326). In addition, this approach can provide a modeling framework for other data-limited vulnerable populations and species.

KEY WORDS: *Dermochelys coriacea* · Species distribution model · Probability of occurrence · Boosted regression trees · Conservation priority-setting

1. INTRODUCTION

Marine turtles spend most of their lives at sea but are tied to terrestrial habitats (i.e. nesting beaches) for re-

production. This life history strategy exposes these species to indirect (e.g. climate change and habitat loss) and direct (e.g. fishing, pollution and marine debris, egg consumption, coastal development) anthro-

*Corresponding authors: Jlopez@iattc.org

© The authors 2024. Open Access under Creative Commons by Attribution Licence. Use, distribution and reproduction are unrestricted. Authors and original publication must be credited.

pogenic threats that have caused declines in many populations (Wallace et al. 2011, Northwest Atlantic Leatherback Working Group 2019, Wibbels & Bevan 2019). Consequently, conservation efforts have increased in many regions of the world to address these threats, with varying degrees of success (e.g. Chaloupka et al. 2008, Casale & Tucker 2017, Mazaris et al. 2017).

Of particular conservation concern is the leatherback turtle *Dermochelys coriacea*, the largest (typically > 1.5 m curved carapace length) and most geographically widespread of all marine turtle species (Eckert et al. 2012). Distributed circumglobally from tropical to temperate regions and regularly occurring in coastal as well as high-seas areas, the species comprises 7 regional management units (RMUs) or subpopulations (Wallace et al. 2023). Two RMUs—the East Pacific (EP) and West Pacific—exist in the Pacific Ocean, both of which are currently classified as Critically Endangered on the IUCN Red List (Tiwari et al. 2013, Wallace et al. 2013). The EP leatherback RMU has declined by over 90% since the 1980s, largely due to unsustainable levels of incidental mortality by industrialized and artisanal fisheries, which mainly affect sub-adults and adults, as well as egg consumption by humans (Laúd OPO Network 2020).

Given that industrial and artisanal tuna (e.g. yellowfin *Thunnus albacares*, albacore *T. alalunga*, big-eye *T. obesus*, skipjack *Katsuwonus pelamis*) purse-seine and longline fisheries cover a high proportion of the species' distribution in the east Pacific Ocean (EPO) (Wallace et al. 2023), they unavoidably interact with leatherback turtles and other sea turtle species during their normal fishing operations as they target tunas, billfish, and other species that share similar epipelagic habitats (IATTC 2020). The Inter-American Tropical Tuna Commission (IATTC) is the Regional Fisheries Management Organization that is responsible for the long-term conservation and sustainable management of fisheries that target tuna and tuna-like species in the EPO. Since the Antigua Convention entered into force in 2010, which required a significantly broader ecosystem-based approach to management by the IATTC, several resolutions pertaining to the conservation and management of various sensitive bycatch species have been implemented, including sea turtles (IATTC Resolution C-07-03). However, in light of the declining population of leatherback turtles in the EPO, the IATTC implemented more stringent conservation measures in 2021 to mitigate fishery impacts on sea turtles (IATTC Resolution C-19-04).

Further, the Inter-American Convention for the Protection and Conservation of Sea Turtles (IAC) is a

binding intergovernmental treaty that provides the legal framework for states of North, Central, and South America and the Caribbean Sea to take actions to benefit sea turtles, in both nesting beaches and the Parties' territorial waters. Concerned with the critical status of leatherback turtles in the EPO, the IAC adopted Resolution CIT-COP7-2015-R2 in 2015, updated in 2022 to Resolution CIT-COP10-2020-R6, which requests IAC Parties to implement or improve measures to reduce bycatch of leatherback sea turtles in the eastern Pacific fisheries based on the best scientific information available and using recommendations from IAC Resolution CIT-COP10-2022-R7 to exercise FAO guidelines to reduce sea turtle mortality in fishing operations (FAO Fisheries Department 2009).

Assessing fisheries impacts on bycatch species is challenging due to the frequent lack of reliable biological and catch information, especially for species of little or no commercial value or in data-limited settings. Therefore, assessing bycatch species using traditional stock assessment approaches is often both cost-prohibitive and impractical, thus requiring alternative approaches. To address this issue, Griffiths et al. (2019) developed a flexible, spatially explicit, quantitative ecological risk assessment approach—Ecological Assessment of Sustainable Impacts of Fisheries (EASI-Fish)—to quantify the cumulative impacts of multiple fisheries on data-limited bycatch species, such as sea turtles.

Because the extent of areal overlap between the species and each interacting fishery is of critical importance in the EASI-Fish approach, a reliable species distribution model (SDM) is required. SDMs are built to describe the relationship between a species and environmental conditions and can predict how environmental variability may affect their distribution and habitat choice (Elith & Leathwick 2009). Although SDMs have been widely applied in predicting suitable habitats for marine species over the past few decades (Melo-Merino et al. 2020), only a few studies have focused on bycatch species that are data-poor and are of particularly high conservation importance (e.g. Sequeira et al. 2014, Abrahms et al. 2019, Lezama-Ochoa et al. 2020, Lopez et al. 2020). For example, a simple environmental envelope model was developed for NOAA's TurtleWatch tool (Howell et al. 2008, 2015) to mitigate loggerhead *Caretta caretta* and leatherback turtle bycatch by longline fisheries in the central north Pacific Ocean.

For rarely encountered bycatch species, whether their rarity is due to fishing gear selectivity issues, non-reporting, or declining population size, there is often a small number of observations from which to develop

an SDM. Therefore, this limits the types of models available to develop an SDM that can make use of scant presence records. Machine-learning algorithms, including boosted regression trees (BRT) (Elith et al. 2006), are a relatively new suite of powerful tools that can accommodate non-linear relationships, high-dimensional large data sets, imbalanced classes, and limited species occurrences (Elith et al. 2008, Mi et al. 2017).

The goal of this study was to generate a reliable high-resolution SDM that would support the evaluation of EP leatherback vulnerability to the impacts of fisheries as well as the potential efficacy of conservation management measures (e.g. IATTC C-19-04) using the EASI-Fish approach. This study was part of a multi-year, collaborative effort between IATTC and IAC representatives, under a 2011 Memorandum of Understanding (MoU) between the 2 conventions to inform conservation and management of EP leatherback turtles in the EPO. Previous research using satellite telemetry has estimated overlaps between Atlantic leatherbacks and the potential risk of accidental interactions with fisheries (e.g. James et al. 2005, Fossette et al. 2014). Similarly, for EP leatherbacks, previous studies have developed potential methods for producing SDMs using either individual-based satellite telemetry data (Hoover et al. 2019), presence-only observation data (Degenford et al. 2021), or both (Liang et al. 2023). Furthermore, there is clear value in a detailed investigation of inter- and intra-annual leatherback distributions in relation to large-scale environmental cycles (e.g. El Niño–Southern Oscillation [ENSO] regimes, climate change) to inform adaptive management options that reflect these dynamics (e.g. Hazen et al. 2018, Willis-Norton et al. 2015, Pons et al. 2022). Leatherback movements, habitat use, and life history are known to be strongly influenced by environmental conditions, specifically how those conditions affect resource availability (Saba et al. 2007, 2008, Shillinger et al. 2008, 2010, Wallace & Saba 2009, Bailey et al. 2012, Hoover et al. 2019). Thus, predictions of leatherback occurrence in time and space at different scales would be highly informative and useful for fine-tuning conservation strategies in the EPO.

In this study, we developed a hierarchical machine-learning modeling approach that used a region-wide presence–absence data set for EP leatherback turtles and incorporated different predictive variables and modeling configurations to (1) understand the potential distribution of the species at different spatio-temporal scales, (2) identify the environmental preferences of the species, and (3) develop a final prediction map describing the most plausible distribution for the

species that will be used in a concurrent study to assess the species' vulnerability to fishery interactions. This paper describes the development of a novel SDM that was then used to quantify and mitigate the impacts of EPO fisheries on EP leatherback turtles, helps inform their vulnerability status, and guide the development of appropriate conservation and management decisions (Griffiths & Wallace et al. 2024, this volume). It was also envisaged that if the model was successful for leatherback turtles, it could also be applied to other data-limited vulnerable taxa.

2. METHODS

All data processing and analytical work was carried out using R version 3.4.3 (R Core Team 2017)).

2.1. Fisheries observer data

We used 26 years (1995–2020) of observer and log-book data collected from a variety of industrial and small-scale coastal (artisanal) fisheries (Table 1) operating in 6 countries and the high seas within the IATTC convention area—defined as the region from the Pacific coast of the Americas to 150° W between 50° S and 50° N. Our data set included observations of leatherback turtle presence as well as absence during fishing operations compiled by a participatory collaboration coordinated by the IATTC and IAC to undertake the EASI-Fish leatherback vulnerability assessment.

Nearly 90% of high-seas presence–absence data were derived from large-scale tuna longline fishing vessels (≥ 24 m, hereafter called the 'industrial longline fishery'; 484 active vessels in 2022; <https://iattc.org/en-US/management>) and the industrial tropical tuna large purse-seine fishing fleet (Class 6 with a carrying capacity of > 363 mt, 180 active vessels in 2022; <https://iattc.org/en-US/management>). The data include set-level information on leatherback turtle interactions along with location, date, and time of the observation. The distribution of industrial tuna fishing effort was mostly concentrated ($> 75\%$ of sets) between 20° N and 20° S during the study period. The data for these fleets were collected by IATTC onboard scientific observers or submitted to the IATTC by its members under Resolution C-19-08. The observer coverage rate was close to 100% for purse-seine vessels of Class 6 and approximately 5% for the industrial longline fishery.

In contrast to the industrial fisheries in the EPO, catch and effort by the numerous artisanal fleets that

Table 1. Data sources and period of coverage of data used to build the leatherback species distribution model for the East Pacific Ocean

Country	Gear	First year	Last year	Presence only	Abundance	Effort	No. of presences	No. of ind.	No. total sets	% of presences	Source
Chile	Purse-seine	2015	2019	No	No	–	3	3	4396	0.07	Observers
Chile	Industrial longline	2001	2018	No	Yes	Yes	327	365	13828	2.36	Observers
						(No. of hooks)					
Chile	Artisanal longline	2002	2018	No	Yes	Yes	59	62	1831	3.22	Observers
						(No. of hooks)					
Chile	Artisanal longline (espinel)	2010	2019	No	No	Yes	2	2	564	0.35	Observers
						(No. of hooks)					
Chile	Artisanal gillnet	2007	2019	No	Yes	No	22	24	1399	1.57	Observers
Colombia	Gillnet	2017	2018	Yes	No	No	3	3	3	–	Observers
Colombia	Longline	2018	2018	Yes	No	No	2	2	2	–	Observers
IATTC	Purse-seine	1995	2020	No	Yes	No	272	274	532857	0.05	Observers
IATTC	Longline	2013	2020	No	Yes	No	67	67	24005	0.28	Observers
Panama	Purse seine / longline / gillnet	2018	2020	Yes	No	No	10	10	10	–	Observers
Peru	Longline (surface), gillnet	2001	2019	Yes	No	–	186	186	186	–	Pro-Delphinus
Ecuador	Purse-seine	2019	2020	No	No	–	3	3	2746	0.11	Observers
Ecuador	Longline (bottom)	2017	2020	No	No	No	0	0	766	0.00	Observers
Ecuador	Longline (surface)	2019	2020	No	No	No	2	2	1667	0.12	Observers
Peru	Driftnet / gillnet	1997	2015	Yes	No	No	141	141	141	–	IMARPE/ACOREMA
Peru	Driftnet / gillnet	2013	2020	Yes	Yes	No	21	21	21	–	IMARPE (LAMBAYEQUE)
WWF (various)	Longline	2004	2009	No	Yes	Yes (Various)	20	20	7539	0.27	WWF-IATTC
Costa Rica	Longline	2005	2012	No	Yes	Yes	5	5	2602	0.19	Observers
All						(No. of hooks)					
		1995	2020				1145	1190	594563	0.19	

operate within the exclusive economic zones (EEZs) of countries in the EPO generally have very low (if any) observer coverage and are poorly documented in general. However, based on available data (i.e. collected by opportunistic or sporadic observations, and port-based interviews with fishermen), leatherback turtles have been shown to be heavily impacted by coastal artisanal gillnet and longline fisheries, particularly in foraging areas but also in migratory and reproduction areas (Frazier & Brito 1990, Alfaro-Shigueto et al. 2011, 2018, Quiñones et al. 2021). Reasonably detailed effort data for artisanal longline vessels throughout Central America was available from

IATTC's long-term research program that examined the effects of different hook types on marine turtle bycatch rates, reported in part by Andracka et al. (2013). In addition, unpublished data pertaining to leatherback turtle interaction and fishing effort information for several artisanal fisheries operating in territorial waters of 6 countries in the EPO were opportunistically compiled (Table 1).

Duplicated records, data outside the EPO, and observations without reliable date and location information were removed from the data set ($n = 57$ records, ~5% of the initial presences). The final data set included 1088 leatherback records from nearly

575 000 fishing sets (i.e. 0.19% of sets containing at least one leatherback observation) from all years with available data (Table 1, Fig. 1).

2.2. Predictive variables

A total of 23 variables were included in SDMs, which included 3 spatio-temporal variables, 11 surface variables, 2 subsurface variables, and 2 static variables (Table 2). The 3 spatio-temporal variables included location and date of set, as seasonality can affect catches. Spatio-temporal variables can be confounded with environmental factors and reflect certain natural processes not captured by the environmental (i.e. surface and subsurface) variables. The majority of environmental data was sourced from daily or weekly fields of global data assimilative models (i.e. assimilated data derived from satellites and *in situ* platforms) for the IATTC convention area at 0.25° (~25 km²) resolution (available at <http://marine.copernicus.eu/> and <https://www.aviso.altimetry.fr/>). The 0.25° spatial resolution, combined with a fine temporal scale, is considered adequate for habitat modeling (Scales et al. 2017a).

The 11 surface variables chosen included sea surface temperature (SST) and its gradient (SST_grad; calculated as the change in temperature at the same pixel over a period of 7 d), salinity (Sal), sea surface height (SSH), current speed (Vel), current direction (Dir), eddy kinetic energy (EkE), finite size Lyapunov exponents (FSLE), front index (FrontIndex; estimated as a count of the front pixels in the grid cell for the 7 d window), chlorophyll *a* (chl), and chl *a* gradient (chl_grad; computed as the difference in chl *a* concentration in the same pixel over a 7 d period).

The 2 subsurface variables included temperature at 100 m depth (SST₁₀₀) and mixed layer depth (MLD). SST₁₀₀ and MLD have proven to be helpful to improve SDMs for large pelagic species (Brodie et al. 2018) and help describe the 2-dimensional (i.e. vertical and horizontal space) structure of the water column properties.

The 2 static variables included bathymetry (Depth) and the distance

to land (LandDistance). These variables were extracted from the Global Marine Environmental Data-sets (Basher et al. preprint, doi:10.5194/essd-2018-64) and MARSPEC Ocean Climate Database (Sbrocco & Barber 2013), respectively, and have been shown to be important in defining leatherback turtle habitat (e.g. Hazen et al. 2018, Robinson et al. 2016, Willis-Norton et al. 2015).

2.3. Model development

In the interest of robustness and to inform comparisons, we took a hierarchical multi-model approach, building 6 presence–absence (catch vs. zero catch per set, binary response) model configurations with each set of variables, from the simplest to the most complex models. The following model configurations were established with all presence and absence data:

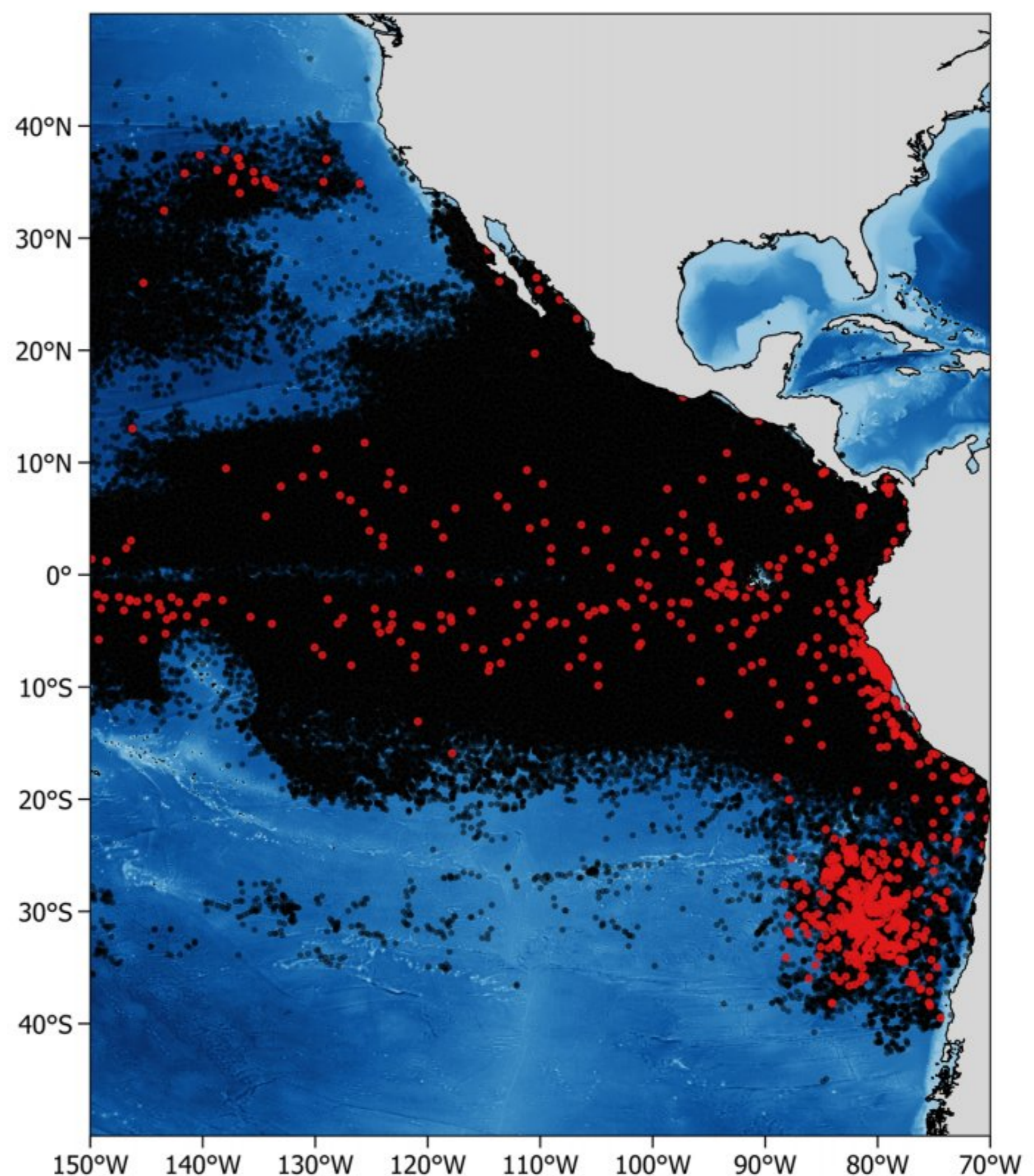


Fig. 1. Distribution of leatherback sea turtle presences (red points) and absences (black points) collected by on-board observers and logbooks for industrial longline and purse-seine fisheries and small-scale longline and surface gillnet fisheries in the Eastern Pacific Ocean between 1995 and 2020

Table 2. Comparing data sources and spatio-temporal resolutions for variables used in species distribution models. D: daily; W: weekly

Variable	Unit	Spatial resolution	Temporal resolution	Source
Spatio-temporal				
Latitude	Decimal degree	GPS	–	–
Longitude	Decimal degree	GPS	–	–
Day of the year	Calendar day	–	D	Processed
Environmental surface				
Sea surface temperature	°C	1/4°	D	CMEMS
Sea surface temperature gradient	°C	1/4°	D	Processed
Salinity	g kg ⁻¹	1/4°	D	CMEMS
Sea surface height	m	1/4°	D	CMEMS
Current speed	m s ⁻¹	1/4°	D	Processed
Current direction	Degrees	1/4°	D	Processed
Eddy kinetic energy	cm ² s ⁻²	1/4°	D	Processed
Finite size Lyapunov exponents	d ⁻¹	1/25°	D	AVISO
Front index	Count of front pixels per cell for the 7 d window	1/4°	W	Processed
Chlorophyll	mg m ⁻³	1/4°	W	CMEMS
Chlorophyll gradient	mg m ⁻³	1/4°	W	Processed
Environmental subsurface				
Temperature at 100 m	°C	1/4°	D	CMEMS
Mixed layer depth		1/4°	D	CMEMS
Static				
Bathymetry		1/12°	–	GMED
Distance to coast		1/120°	–	MARSPEC (Sbrocco & Barber 2013)

(1) spatio-temporal, (2) surface, (3) subsurface, (4) environmental (surface + subsurface), (5) static, and (6) full (environmental and static) (Fig. 2). Because the full model had the best performance metrics (see Table 3), subsequent models using different proportions of presence to absence data were only established with this set of variables (full; environmental and static variables) (see Section 2.3.3 for details).

2.3.1. Model building

BRTs are a flexible classification and regression algorithm based on machine-learning principles (Elith et al. 2006, De'ath 2007). Consequently, some of the caveats of more commonly used techniques, such as generalized linear models (GLMMs) or generalized additive mixed models (GAMM), are not applicable. BRTs have the particular advantage of being tolerant of missing values, outliers, correlation, collinearity, non-independence, and allowing for the inclusion of irrelevant predictors (Leathwick et al. 2006). BRTs are also designed to accommodate non-linear relationships, large high-dimensional data sets, imbalanced classes, and limited species occurrences (Elith et al.

2008, Mi et al. 2017). While GLMMs and GAMMs seek to fit the most parsimonious model to a data set, BRTs combine stochastic predictions of many simple models (i.e. many shallow classification trees) to maximize robustness and predictive performance to reduce associated error (Scales et al. 2017b). Accordingly, we fitted BRTs with all available sets of covariates. In the past, authors also fitted GAMM and random forest (RF) models to presence–absence data for other species to compare and better understand consistency and interpretation between algorithms (e.g. Lopez et al. 2019). In these cases, BRTs performed better than GAMMs and had very similar performance to the RFs. As such, we decided to use BRTs to build all the models in this study, which were implemented using the R package 'dismo' (Hijmans et al. 2017).

In fitting BRTs, we adapted the protocols outlined by Elith et al. (2008), Scales et al. (2017b), Brodie et al. (2018), and Hazen et al. (2018). Presence–absence models were built with a binomial (Bernoulli) distribution. We used a tree complexity of 3, a bag fraction of 0.7, and conducted sensitivity analyses on learning rate ('shrinkage') for each model set, aiming for at least 1000 trees in final model configurations. The sensitivity runs determined 0.01 as the learning rate to be used in all the models, except for the model with

the same number of presences and absences (i.e. 50:50 presence–absence ratio; see Section 2.3.3), where a value of 0.005 was used. Tree complexity refers to the number of nodes in a tree, which constrains the maximum size of each of the regression trees that together make up a boosted regression tree model. By controlling the number of nodes or branches, tree complexity also sets the maximum number of interactions between predictor variables that are possible (i.e. 3 in this case, as 2- and perhaps 3-way interactions among variables may be important, but higher orders are unnecessary in fisheries contexts; Soykan et al. 2014). Bag fraction refers to the proportion of the data that is randomly used for model building at each step, which usually ranges between 0.6 and 0.75 (Elith et al. 2008). The stochasticity that this step provides to the model-building process improves model performance (Soykan et al. 2014).

The potential for model simplification was evaluated with the function 'gbm.simplify'. Simplified models were fitted by re-running models without those variables that gave no evidence of improving predictive performance. Deviance explained, variable importance, and interactions between variables were also estimated for all models using the function 'gbm.interactions'. Each of these configuration settings and the performance procedures are described in detail by Elith & Leathwick (2017), Elith et al. (2008), Hazen et al. (2018), Scales et al. (2017b), and Soykan et al. (2014).

2.3.2. Model validation

Patterns derived from SDMs — particularly those produced by machine-learning techniques — are susceptible to bias if the original data is biased (see Lee-Yaw et al. 2022), which may be the case with our highly imbalanced data set from different sources (note that the leatherback turtle is not a target species but a bycatch species for all fisheries included in this study; hence, the data is expected to be less influenced by fishing strategies or other variables that can affect data collection and bias). In an attempt to address this issue, we conducted extensive model validation exercises. Independent data sets such as leatherback turtle space use derived from satellite tracking data (e.g. Shillinger et al. 2008, Bailey et al. 2012) were not made available for our analysis; such tests of how well SDM predictions agree with independent estimates of distributions are rare (Lee-Yaw et al. 2022). However, we employed multiple validation approaches as well as an examination of uncertainty or confidence around predictions by using subsets of our data set. We acknowledge that such internal validation approaches do not explicitly address biased predictions resulting from biased underlying data sets. Two cross-validation methods were used to evaluate the reliability and the predictive performance of the final models: *k*-fold cross-validation (main method) and a hold-out cross-validation (additional complementary method). These methods consist of using independent data sets for model

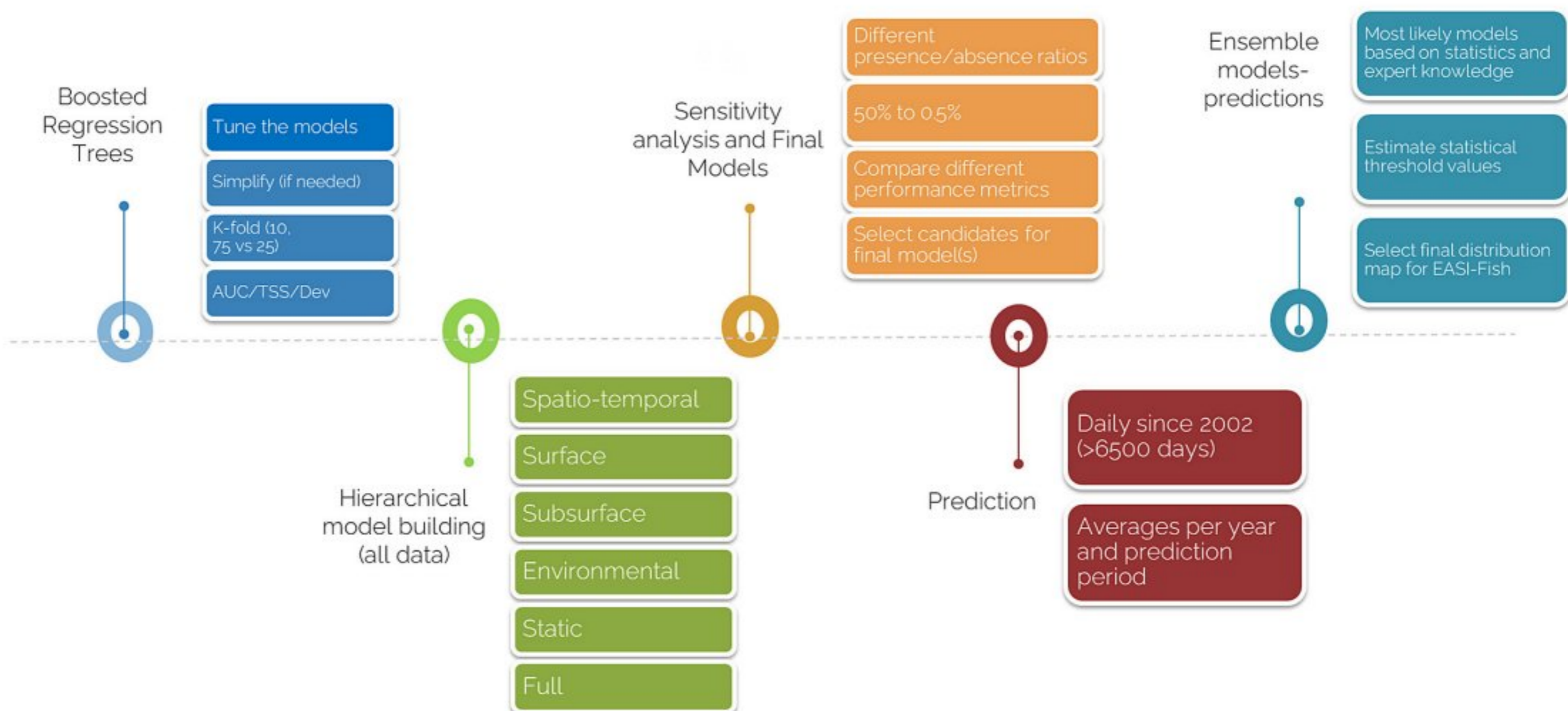


Fig. 2. Infographic description of the methods (sequential from left to right) used to develop the species distribution models. See Section 2.3 for details

building (i.e. the training data) and model validation (i.e. the test data), where data are partitioned into k equally sized segments or folds through random resampling (k -fold cross-validation) or are intentionally partitioned into different segments based on spatial-temporal properties (hold-out cross-validation). Model performance is assessed by successively removing each subset, rebuilding the model on the retained data, and predicting on the omitted data (Elith & Leathwick 2009). In this study, a $k = 4$ partitioning method was used, meaning that 75% of the observations were used for model building, and the other 25% were used for model cross-validation. In the hold-out cross-validation, there is no random shuffling of observations. Instead, data is partitioned into blocks of interest for model training, and testing is done with the remaining contrasting block of data (e.g. Becker et al. 2020). As such, data from neutral years was used to train the model and test on El Niño–La Niña data. Similarly, data was also partitioned into inshore and offshore blocks, where inshore data was used to train the model and testing offshore, and vice versa (2 different limits were explored to separate inshore and offshore data; 90° W and 100° W). Both k -fold and hold-out cross-validations avoid the overlap between training data and test data, yielding a more accurate estimate of the generalization performance of the algorithm (Villarino et al. 2015).

The predictive power and the stability (i.e. uncertainty or confidence) of the model were assessed by computing a set of diagnostic metrics. The mean area under the receiver-operating curve (AUC) (Hanley & McNeil 1982) and the mean true skill statistic (TSS) (Allouche et al. 2006) were calculated for each iteration from each confusion matrix to evaluate the predictive performance of the models, and the coefficient of variation (CV) of the predictions was used to evaluate their stability (e.g. Wang et al. 2019, Montoya-Jiménez et al. 2022, Borokini et al. 2023, Roberts et al. 2023). The AUC provides a single measure of overall model accuracy that is threshold-independent, with an AUC value of 0.5 indicating that the prediction is as good as random, whereas $AUC = 1$ indicates perfect prediction (Fielding & Bell 1997). AUC has been extensively used in SDMs and measures the ability of the model to correctly predict where a species is present or absent (Elith et al. 2006). An AUC value of >0.75 is considered to have good predictive power and is acceptable for conservation planning (Pearce & Ferrier 2000). TSS is an alternative measure of model accuracy that is threshold-dependent and not affected by the size of the validation set,

and it is an appropriate evaluative tool in cases where model predictions are formulated as presence–absence maps (Allouche et al. 2006). TSS is on a scale from -1 to $+1$, with 0 representing no predictive skill, and is calculated from the confusion matrix outputs as sensitivity plus specificity minus 1 (i.e. $TSS = \text{sensitivity} + \text{specificity} - 1$). Threshold independent and dependent statistics, such as AUC and TSS, respectively, should be used in combination when evaluating the predictive power of an SDM (Pearson et al. 2006). The model uncertainty or confidence in the resulting probability estimates was quantified using the CV, which measures the percentage of variation around the arithmetic mean of a series (i.e. the ratio of the standard error to the predicted value per observation) and was mapped to describe uncertainty alongside the predictions.

2.3.3. Sensitivity analyses

The model utilizing all data (hereafter called the 'original model') contained 573 883 observations (1088 presences; 0.19%). To determine the effect of using different proportions of presence on model performance, the relationship between the response variable and the covariates (i.e. environmental and static), as well as the generated predictions, a multi-model approach was conducted using 10 data sets, each having a different presence–absence ratio. The presence–absence ratio in the data used to build the final models was incrementally decreased from 50 to 0.5%. In each case, all presence observations were included with a variable number of randomly selected absences. For example, the 50:50 model included 1088 presences and 1088 absences while the 0.5:99.5 model included 1088 presences and 216 512 absences (see Table 4). All models (hereafter called 'final models') were run using all environmental and static variables, as per the full model, and followed the same model building and validation procedures mentioned above.

2.4. Predictions

2.4.1. Daily predictions

Daily predictions of the probability of occurrence of leatherback turtles across the IATTC convention area were conducted for 2002–2020 (i.e. 6935 daily predictions). A series of time-matched environmental data fields (both surface and subsurface as well as static

variables) were used to generate daily predictions based on the 11 final models and their best number of trees using function 'predict' in the R package 'raster' (Hijmans & van Etten 2015). Therefore, 11 different predictions were computed for each day of the time series to inform consistency and interpretation, and to visualize the effect of accounting for different proportions of presences on the predictions. The spatial resolution of the predictive surface was set to the lowest common resolution of environmental data fields (0.25°).

2.4.2. Prediction averaging and ensemble

Daily predictions ($n = 6935$) were averaged for 2002–2020 for each of the 11 final models, at the full regional scale averaged across the study period (Figs. S1 & S2 in the Supplement at www.int-res.com/articles/suppl/n053p271_supp.pdf), quarterly (Fig. S3 in the Supplement), and by ENSO phase (i.e. neutral, La Niña, El Niño; Figs. S4 & S5 in the Supplement). Different ENSO regimes were defined following NOAA's Earth System Research Laboratories Oceanic Niño Index, where index values of +0.5 or higher indicate El Niño and values of -0.5 or lower indicate La Niña phases (<https://www.ncei.noaa.gov/access/monitoring/enso/sst#oni>). Visual inspection of predictions and exploration of performance metrics and the relationships between the response variable and covariates suggested 2 groups of similar models: (1) 6 models with a presence–absence ratio ranging from 50 to 10% and (2) 5 models with a presence–absence ratio ranging from 5 to 0.19%. We concluded that the models of the first group best reflected the full suite of existing leatherback turtle distribution data (e.g. Shillinger et al. 2008, Quiñones et al. 2021) and SDMs (Hoover et al. 2019, Degenford et al. 2021, Liang et al. 2023) in the region. Therefore, an ensemble model was created using the average predictions from all models having a presence–absence ratio of 50 to 10%. Similarly, the model with a 25% presence–absence ratio was also identified by both expert opinion and performance metrics as a plausible model (hereafter called the 'reference model'). Therefore, an ensemble model was developed using the average predictions from the upper (33:66), lower (20:80), and intermediate (i.e. reference) (25:75) models. Generating the 2 ensemble predictions allowed comparisons between candidates and the exploration of the potential effects of model selection on the final prediction (Fig. S6 in the Supplement).

2.4.3. Probability-of-occupancy threshold

The first stage of EASI-Fish in estimating the vulnerability of a species to fisheries impacts is to determine the number of grid cells in which the species is considered to be present and which also contain fishing effort. Because SDMs estimate the probability of a species to occur in each grid cell, a probability-of-occupancy threshold value (ψ) must be used to define whether the species is present or absent in each cell. However, the spatial extent of the species' distribution increases and decreases with lower and higher ψ values, respectively, and thus influences the proportion of the species' population that is exposed to fishing. To account for this uncertainty, 3 ψ values (0.1, 0.2, and 0.3) were applied to each 0.25° cell, based on statistically determined thresholds and expert evaluation of prediction maps. This range was statistically determined by overlaying the distribution of predicted probability of presence with that of predicted probability of absence. The ψ value where these 2 distributions intersected was selected to define the most probable species distribution, and upper and lower bounds were selected by visual inspection of the 2 distributions on either side of the intersection point (Lopez et al. 2020).

Given the Critically Endangered status of EP leatherbacks, we selected relatively low ψ values to conservatively include areas where experts considered leatherbacks likely to occur, even if in relatively low numbers and for limited periods of time, based on documented patterns of spatio-temporal habitat use (Shillinger et al. 2008, 2011, Donoso & Dutton 2010, Bailey et al. 2012, Quiñones et al. 2021) and previously published SDM maps (Hoover et al. 2019, Degenford et al. 2021). This was important to ensure that EASI-Fish would be precautionary in its calculations of fishery impacts on leatherbacks throughout their distribution and across industrial and small-scale fisheries known to interact with the species (see Griffiths & Wallace et al. 2024).

3. RESULTS

3.1. Model performance

Models that included all data and environmental and static variables (e.g. Model 6–full) performed better according to the diagnostic measures we used (deviance explained, AUC, and TSS) (Table 3). In general, complex models (e.g. Models 4 and 6) had better performance than simpler models including

sets of variables individually (e.g. Models 1, 2, 3, and 5). These models explained between 19.71 and 40.57% of the deviance in the data, had AUC values between 0.79 and 0.94, and had TSS values that ranged between 0.51 and 0.76 (Table 3). The number of trees created by these models ranged between 3600 and 7000. The comparison in model performance led us to recommend the use of the full model configuration (Model 6—full, environmental and static variables) for further exploration on the sensitivity analysis.

Models including different proportions of presence ratios with environmental and static variables (i.e. Model 6—full) showed similar but also different performance under the diagnostic measures that we used (Tables 4 & S1 in the Supplement). In general, models with balanced or slightly imbalanced data sets (i.e. models 50:50, 33:66, 25:75, 20:80, 15:85, and 10:90) had better performance than models that used highly imbalanced data sets (i.e. models 5:95, 2.5:97.5, 1:99, 0.5:99.5, and original—0.19%). These models explained between 40.57 and 61.54% of the deviance in the data, had AUC values between 0.92 and 0.96, and had TSS values that ranged between 0.76 and 0.81 (Table 4). Models built for the hold-out additional complementary cross-validation also performed well (Table S1). In particular, models trained without data pertaining to specific periods and then tested against them (e.g. El Niño, La Niña) showed high AUC, TSS, and deviance explained values (AUC ~0.9–0.85, TSS ~0.8–0.60, 50–60% deviance explained) (Table S1). Similarly, models trained with inshore or offshore data (i.e. limits at 90W and 100W explored) and tested against the remaining data showed reasonable performance metrics (AUC ~0.7–0.6, TSS ~0.3–0.2, ~30–60% deviance explained) (Table S1). Based on these diagnostic measures, the model using a 25% presence–absence ratio was identified as the reference model.

Table 3. Hierarchical model performance metrics with all data (n = 573 889). Fit: fitted model; Simp: simplified model; ST: spatio-temporal variables; SUR: environmental surface variables; SUB: environmental subsurface variables; ENV: environmental surface and subsurface variables; STA: static variables; FULL: all environmental variables and static variables. lr: learning rate; n.trees: number of trees generated by model; Dev: deviance; AUC: area under the curve; TSS: true skill statistics; Drop: Dropped variables after model simplification; EkE: eddy kinetic energy

		lr	n.trees	Dev	AUC	TSS	Drop
1	Fit	0.01	4500	38.64	0.89	0.66	
ST	Simp	—	—	—	—	—	No drop
2	Fit	0.01	5150	32.28	—	—	
SUR	Simp	—	5150	32.40	0.92	0.71	EkE
3	Fit	0.01	3600	23.62	0.88	0.65	
SUB	Simp	—	—	—	—	—	No drop
4	Fit	0.01	5600	38.14	0.94	0.74	
ENV	Simp	—	—	—	—	—	No drop
5	Fit	0.01	4400	19.71	0.79	0.51	
STA	Simp	—	—	—	—	—	No drop
6	Fit	0.01	7000	40.57	0.94	0.76	
FULL	Simp	—	—	—	—	—	No drop

Table 4. Performance metrics of the models conducted during the sensitivity analyses for full models (all environmental and static variables) with different presence–absence ratios (e.g. 50:50 presence to absence ratio = 50% presence data and 50% absence data in the model). Fit: fitted model; Simp: simplified model; Drop: Dropped variables after model simplification

		lr	n.trees	Dev	AUC	TSS	Drop
50:50	Fit	0.005	2900	58.10	—	—	
n = 2176	Simp	0.005	3650	59.37	0.92	0.71	7 variables
33:66	Fit	0.01	2050	59.16	—	—	
n = 3264	Simp	0.01	2650	61.54	0.94	0.76	5 variables
25:75	Fit	0.01	2700	59.70	—	—	
n = 4352	Simp	0.01	3250	60.42	0.96	0.81	7 variables
20:80	Fit	0.01	3200	60.29	—	—	
n = 5440	Simp	0.01	3350	59.14	0.94	0.77	6 variables
15:85	Fit	0.01	3950	60.21	—	—	
n = 7250	Simp	0.01	3650	58.96	0.94	0.76	3 variables
10:90	Fit	0.01	5050	60.93	—	—	
n = 10880	Simp	0.01	4800	60.16	0.95	0.78	2 variables
5:95	Fit	0.01	5300	56.36	0.95	0.78	
n = 21760	Simp	—	—	—	—	—	No drop
2.5:97.5	Fit	0.01	6400	54.67	—	—	
n = 43520	Simp	0.01	5650	52.08	0.95	0.77	7 variables
1:99	Fit	0.01	6650	49.24	0.94	0.78	
n = 108800	Simp	—	—	—	—	—	No drop
0.5:99.5	Fit	0.01	6100	45.23	0.94	0.76	
n = 217600	Simp	—	—	—	—	—	No drop
0.19:99.81	Fit	0.01	7000	40.57	0.94	0.76	
n = 573889	Simp	—	—	—	—	—	No drop

3.2. Drivers of leatherback turtle presence

An examination of the relationships between species and the environmental and static variables showed a range of interesting patterns for each of the

11 final models, based on variable importance analysis (Fig. 3) and partial dependence plots (Fig. S7). These indicators suggested 2 groups of similar models: (1) models with a presence–absence ratio ranging from 50 to 10% and (2) models with a presence–absence ratio ranging from 5 to 0.19%. The first group showed higher variable importance with SST, SST₁₀₀, chl, MLD, FrontIndex, chl_grad and Dir, whereas the second group showed higher values for Depth, Sal, LandDistance, FSLE and SSH.

After simplification of the reference model, 8 variables were included in the final model, for which relative variable importance was 3.3–49.9%. EKE, FSLE, LandDistance, SSH, Vel, SST_grad, and FrontIndex were dropped from the final model as they did not improve predictive performance. With the exception of chl_grad (3.3%), all variables contributed more than 5%: SST (49.9%), Depth (12.5%), MLD (9%), SST₁₀₀ (7.5%), Dir (6.7%), chl (6.2%), and Sal (5%) (Fig. S7).

The model identified higher probabilities of leatherback turtle presence around SST values of 16–20°C and in waters shallower than 1000 m. MLD values of < 100 m were associated with higher probabilities of leatherback turtle presence, whereas SST₁₀₀ showed a positive relationship. Currents with southwest direction showed an evident negative relationship with leatherback turtle presence. A positive relationship was also observed between the leatherback turtle presence probability and CHL, while the opposite relationship was observed for Sal (i.e. salinities higher than 30 PSU). Similarly, the model showed higher probabilities of leatherback turtle presence at positive chl_grad values.

3.3. Predictions

Final models were used to predict species habitat suitability in the convention area for 2002–2020

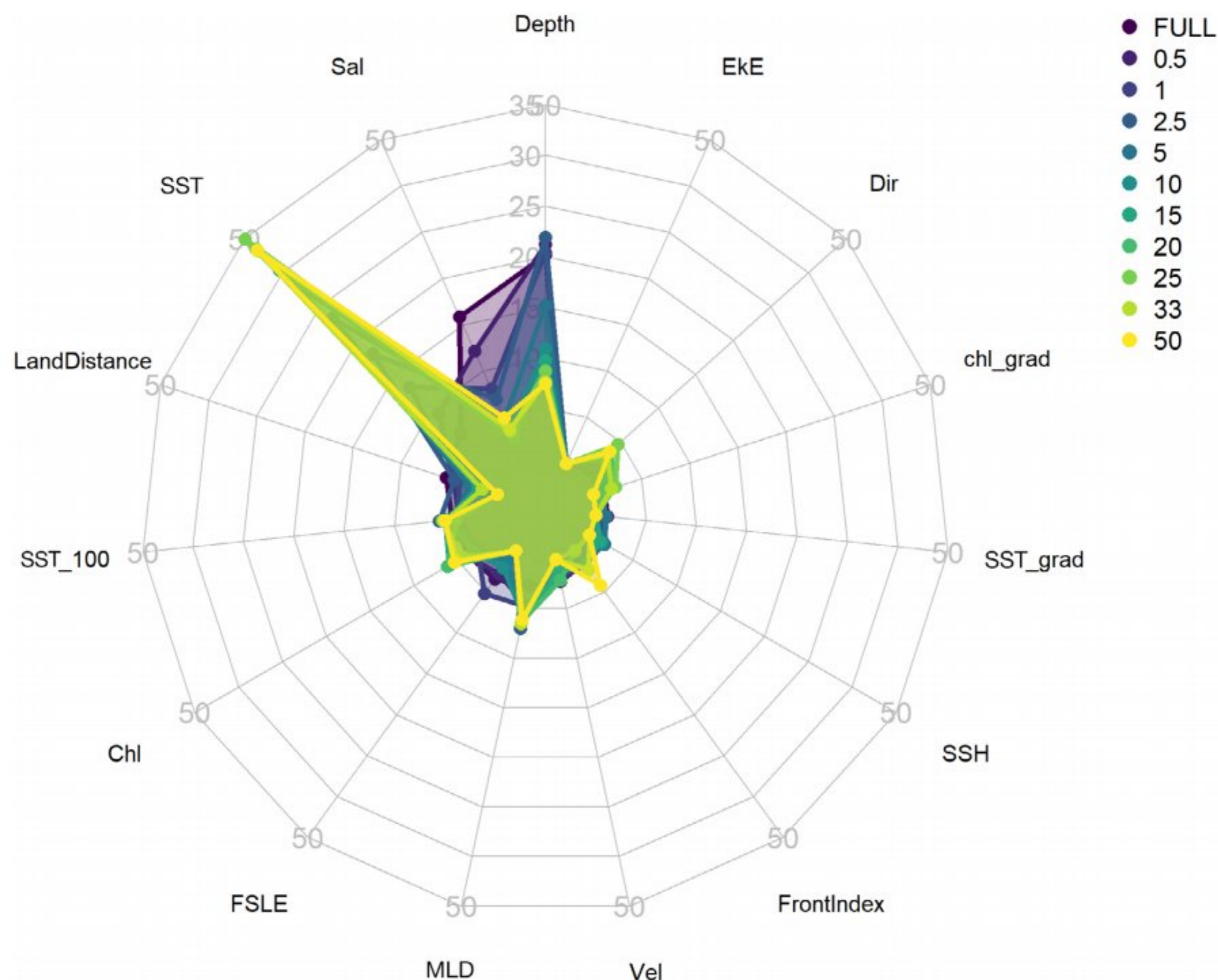


Fig. 3. Relative variable importance in the series of species distribution models generated using different proportions of presence versus absence data; i.e. the ratio of presences to absences used to build the final model (e.g. 50 refers to a 50:50 presence to absence ratio; 33 refers to a 33:66 presence to absence ratio, etc.). Variables included (see Section 2.3 for more details): sea surface temperature (SST) and its gradient within the same pixel over a 7 d period (SST_grad), salinity (Sal), sea surface height (SSH), current speed (Vel), current direction (Dir), eddy kinetic energy (EKE), finite size Lyapunov exponents (FSLE), front index (FrontIndex), chlorophyll *a* (chl), chlorophyll *a* gradient (chl_grad), temperature at 100 m depth (SST₁₀₀), mixed layer depth (MLD), bathymetry (Depth), and the distance to land (LandDistance)

(Figs. S1 & S2), as well as by quarter (Fig. S3), and ENSO phase (Figs. S4 & S5). Predictions revealed spatial differences among models, with, in general, higher probabilities and lower CV of leatherback turtles predicted by models with presence–absence ratios ranging from 50 to 10%. The predictions of this group of models highlighted several areas expected to support higher residence or occurrence by leatherback turtles (Fig. 4). These include coastal areas near nesting beaches in Mexico and Central America, nearshore foraging areas in southern latitudes (Quiñones et al. 2021), and some high-seas areas through which leatherbacks transit and in which they are presumed to forage (e.g. Bailey et al. 2012). These areas also showed relatively low CV values (Figs. 5 & S2). In particular, the continental shelf and adjacent high-seas areas within South American EEZs (specifically from southern Ecuador to northern Chile) as well as the higher latitude subtropical convergence zone extending from south-central Chile clearly and consistently supported higher probabilities of leatherback turtle presence, with relatively low associated CV values (Figs. 5 & S2). These patterns were also reflected in maps of the ensemble predictions and after applying the probability of occupancy thresholds mentioned above (Fig. 4).

While these spatial patterns were generally consistent by quarter (Figs. 6 & S3) and ENSO phase (Figs. 7 & S4), the relative area and distribution of predicted habitat varied at these scales. Predicted leatherback habitat in the Northern Hemisphere, particularly off Mexico and Central America, was most available during Q1 and reduced in other quarters, roughly opposite the trends for predicted habitat off South America (Fig. 6). In addition, predicted leatherback habitat generally increased during the La Niña phase relative to neutral and El Niño phases, particularly along the equator and in the Humboldt Current region (Figs. 7, 8, S4 & S5). However, there were some areas where predicted habitat increased during the El Niño phase, such as the high-latitude region between 25° and 40° S (Figs. 7, 8, S4 & S5).

4. DISCUSSION

Understanding the spatial distributions of marine species is becoming increasingly important as international and regional oceans and fisheries management instruments evolve further toward ecosystem-based approaches to conservation and management of marine resources (Aburto et al. 2012, Kirkfeldt 2019). In a fisheries context, having a reliable prediction of a spe-

cies' distribution that can be compared with that of fishing effort can allow managers to determine the extent of overlap and implement conservation and management measures (CMMs), such as spatial closures, that can reduce fishery interactions and fishing mortality and thus ensure long-term population sustainability.

4.1. Model performance and relevant features

SDMs have traditionally required a large number of species presence locations at a fine resolution to be useful for fisheries management in jurisdictions that typically span spatial scales of hundreds or thousands of kilometers. Furthermore, because the largest quantity of data is usually collected for species of high commercial importance (i.e. target species), the application of SDMs has primarily been constrained to these species (Melo-Merino et al. 2020). Unfortunately, reliable catch or even simple occurrence data is often lacking for bycatch species for a variety of reasons, such as being of lower economic importance, the absence of observer programs and policies mandating fishers to record all species interactions in catch logbooks, policies requiring rapid release and no-retention of species, the typically low frequency of interactions of these non-target species, and issues pertaining to species identification. As the present study showed, the incidence of observations of the Critically Endangered EP leatherback turtle population in EPO fisheries was extremely low—1088 occurrences in over half a million observations. Regardless of whether the low frequency of interactions is due to a naturally low density of leatherback turtles in the EPO, rarity of the species following its documented population decline (Laúd OPO Network 2020), or low gear selectivity by EPO fisheries, the data available are insufficient to undertake conventional stock assessments. Therefore, simpler ecological risk assessment approaches that are designed to assess the vulnerability of data-limited species (e.g. EASI-Fish) are highly reliant on estimating the 3-dimensional overlap between the species and fisheries, and thus reliable SDMs.

Although BRTs have been applied to several marine species, they are typically used with larger numbers of occurrences than we had available for this study. Furthermore, validation of our model predictions using a truly 'independent' data set (e.g. satellite telemetry data) was not possible because such data were not made available for our project. This required a deep exploration of specific model diagnostics and sensitivity analyses to determine the utility of BRTs in such data-limited settings. The present study showed the im-

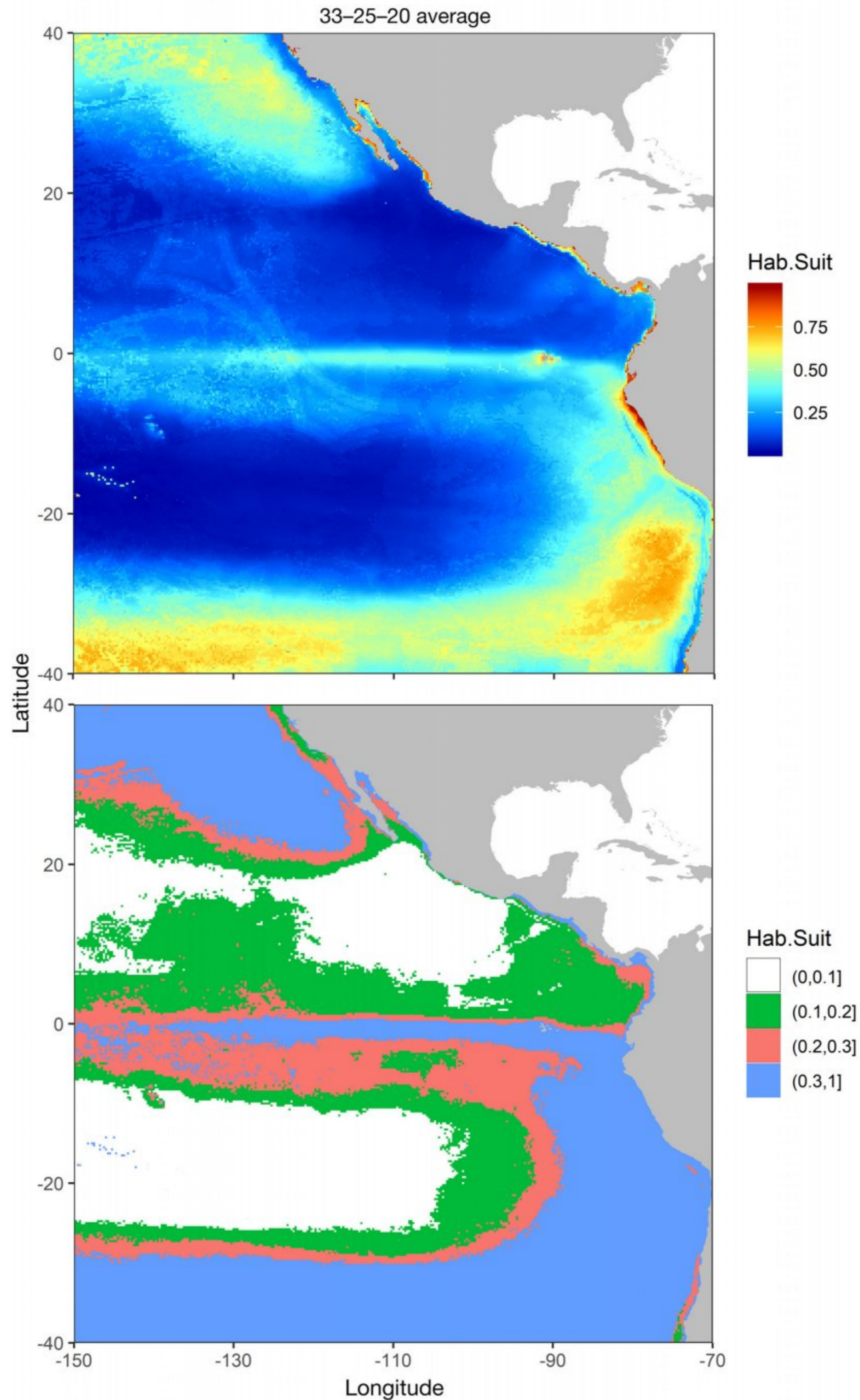


Fig. 4. (a) Average predictions from an ensemble of species distribution models (SDMs) where the ratio of presence to absences were 33–25–20 for 2002–2020; warmer colors: higher probabilities. (b) The final appearance of the SDM predictions using 3 threshold values (minimum = 0.3, blue; most probable = 0.2, red; maximum = 0.1, green) upon which the predicted probability of presence is used to create binary values of species presence. For example, at a threshold of 0.2, predicted probabilities of presence above and below 0.2 are predicted to be absence and presence records, respectively