ICES Journal of Marine Science, 2024, Vol. 00, Issue 0, 1–22
https://doi.org/10.1093/icesjms/fsae043
Received: 12 July 2023; **revised:** 11 March 2024; **accepted:** 12 March 2024
Original Article

Modeling bycatch abundance in tropical tuna purse seine fisheries on floating objects using the Δ method

Agathe Dumont[1,2], Antoine Duparc[2,3], Philippe S. Sabarros[2,3], David M. Kaplan[2,3,*]

[1]PHIM Plant Health Institute, CIRAD, INRAE, Institut Agro, IRD, University of Montpellier, Montpellier 34980, France
[2]MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, 34203 Sète, France
[3]Institut de Recherche pour le Développement (IRD), Av. Jean Monnet, CS3017, 34203 Sète Cédex, France

*Corresponding author. Institut de Recherche pour le Développement (IRD), UMR MARBEC, ave. Jean Monnet, CS3017, 34203 Sète Cédex, France. E-mail: david.kaplan@ird.fr

Abstract

Bycatch rates are essential to estimating fishery impacts and making management decisions, but data on bycatch are often limited. Tropical tuna purse seine (PS) fisheries catch numerous bycatch species, including vulnerable silky sharks. Even if bycatch proportion is relatively low, impacts on pelagic ecosystems may be important due to the large size of these fisheries. Partial observer coverage of bycatch is a major impediment to assessing impacts. Here we develop a generic Δ modeling approach for predicting catch of four major bycatch species, including silky sharks, in floating object-associated fishing sets of the French Indian Ocean PS fleet from 2011 to 2018 based on logbook and observer data. Cross-validation and variable selection are used to identify optimal models consisting of a random forest model for presence–absence and a negative binomial general-additive model for abundance when present. Though models explain small to moderate amounts of variance (5–15%), they outperform a simpler approach commonly used for reporting, and they allow us to estimate total annual bycatch for the four species with robust estimates of uncertainty. Interestingly, uncertainty relative to mean catch is lower for top predators than forage species, consistent with these species having similar behavior and ecological niches to tunas.

Keywords: prediction intervals; random forest; general additive model (GAM); silky shark (*Carcharhinus falciformis*); dolphinfish (*Coryphaena hippurus*); rainbow runner (*Elagatis bipinnulata*); rough triggerfish (*Canthidermis maculata*)

Introduction

Purse seine fishing is a fishing technique that consists of targeting and catching entire fish schools in surface waters by encircling them with a fishing net called a seine. In the tropical oceans, purse seine (PS) is used extensively to target tropical tunas, specifically skipjack tuna (*Katsuwonus pelamis*), yellowfin tuna (*Thunnus albacares*), and bigeye tuna (*Thunnus obesus*) (Kaplan et al. 2014). Tropical tuna fisheries are extremely important economically and in terms of protein provision for human consumption (Amandè et al. 2010, Bell et al. 2019), catching annually over 2 million tons of target tunas worldwide (Pons et al. 2023). In addition to target species, tropical tuna fisheries catch non-target species collectively referred to as bycatch. These species can be discarded at sea, retained to be sold on local markets, or consumed on board (Amandè et al. 2010). In the Indian Ocean, purse seiners have the obligation to retain bycatch on board since 2018 (IOTC Resolutions 17/04 and 19/05), with the exception of species concerned by non-retention measures (e.g. turtles, cetaceans, rays, most sharks, etc.). Even if the ratio of bycatch to target tuna catch is relatively low for tropical tuna PS fisheries (roughly 5−10%; Kaplan et al. 2014), the impact on pelagic populations and ecosystems can be important, specifically for long-lived species with low reproductive rates (e.g. sharks and rays; Dulvy et al. 2014, Pacoureau et al. 2021). Efforts to reduce bycatch are mandated for sensitive or protected species, or species that are essential to the proper functioning of the marine ecosystem (FAO 2020). However, impact studies for bycatch species are often complicated by a lack of bycatch data recorded in fisheries logbooks, which typically only include data on target species.

For this purpose, observers are deployed on board fishing vessels to provide data on bycatch (specific composition, numbers, and size samples) and discards of tuna. In the case of the French PS fleets in the Indian and Atlantic Oceans, 10–20% of fishing trips are covered by observers financed by the European Union Data Collection Framework (DCF), with a complementary observation program supported by industry (Goujon et al. 2017) increasing total observer coverage to ∼50% of fishing sets in the Indian Ocean and nearly 100% of fishing sets in the Atlantic Ocean. Though these partial data provide essential information on bycatch rates and species compositions, accurate estimates of *total* PS mortality of specific bycatch species, including uncertainties, are essential to understanding mortality rates and causes, and identifying population trends and overfishing. Similar issues of partial observer coverage are impediments to accurately quantifying the full impact of fishing on bycatch species in numerous tuna and non-tuna fisheries worldwide. In addition, understanding the extent to which bycatch is predictable based on fishing, spatio-temporal, and environmental covariates is essential to developing effective spatial management (Pons et al. 2022). Therefore, the objective of this study is to develop robust statistical models for predicting the catches of bycatch species as a function of various covariates, including the quantity of target tuna caught, the time and location of fishing

activities, and associated environmental conditions. These models will allow us to both estimate total bycatch mortality for the French Indian Ocean PS fleet and help us understand the links between target species, bycatch species, and the environment, improving fishery-based population status indices and enhancing management capacity for these understudied species.

One persistent challenge for modeling bycatch of relatively rare species is the large number of absences observed in the data (e.g. in the Indian Ocean, silky sharks, the principal shark species caught by PS, are absent from approximately one quarter of fishing sets), which typically prohibits the use of simple statistical models that do not explicitly account for both bycatch presence–absence (i.e. ones and zeros) and abundance when present (i.e. number of individuals caught per set given that at least one individual was caught in the set; hereafter referred to simply as "abundance" when there is no risk of confusion). Moreover, presence–absence and abundance may be partially driven by different biological processes (Stock et al. 2020). Therefore, it seems reasonable to model these two processes separately, combining model results in the end to develop a final prediction of bycatch. This type of modeling approach is known as a Δ model, and it has been used successfully for modeling catch per unit effort of non-target species (Ortiz and Arocha 2004) and target species (Shono 2008). It has the advantage over standard statistical distributions (e.g. normal, log-normal, and Poisson), zero-inflated distributions (e.g. the zero-inflated Poisson distribution), and distributions that permit integrated modeling of presence–absence and abundance data (e.g. Tweedie distribution) in that it allows for greater flexibility and a more pragmatic approach by separately modeling of each aspect of bycatch abundance. For example, a random forest model, a machine learning modeling approach often used for binomial response variables, can be used to estimate presence–absence of a bycatch species, whereas a general additive model (GAM), a non-linear, semi-parametric approach to modeling continuous variables, can be used for abundance when present. In addition to providing maximum flexibility when estimating each aspect of bycatch, this approach allows us to examine results from the individual models separately, as well as in combination.

Below, we develop a series of Δ models for four major bycatch species (silky shark, *Carcharhinus falciformis*; common dolphinfish, *Coryphaena hippurus*; rainbow runner, and *Elagatis bipinnulata*; rough triggerfish, *Canthidermis maculata*) of the French tropical tuna PS fishery in the Indian Ocean based on logbook and observer data. First, we identify optimal modeling approaches and model formulations for presence–absence and abundance of these bycatch species using a multistep approach based on multiple modeling frameworks, multiple distribution families for residuals, predictor variable selection, and cross-validation of results. Next, we combine Δ model outputs with observations to estimate total annual bycatch of the four species, including uncertainties in those predictions. Finally, we compare our results to a simpler bycatch-to-landings ratio approach currently extensively used for bycatch reporting to regional fisheries management organizations. We end with a discussion of results and their implications for fisheries management. Code for carrying out all analyses is made freely available (https://github.com/Agathedumont/modeling-bycatch-abundance-delta-method) so that the approach can be applied worldwide in other fisheries.

Materials and methods

Fishing activity data

The primary data used in this study are catch data for target tunas from logbooks (i.e. mandatory declarations) and bycatch species from observer data from the Indian Ocean French tropical tuna PS fishery between 2011 and 2018. The Exploited Tropical Pelagic Ecosystems Observatory (Ob7) of the French National Research Institute for Sustainable Development (IRD) collects and manages fishery and observer data from the French tropical tuna PS fleet in the Atlantic and Indian Oceans. Though the approach we develop could be applied in the Atlantic and elsewhere, we choose to consider only data from the Indian Ocean, for which ∼34% of all fishing sets have observer coverage (∼37% if only floating object fishing sets are taken into account). Data on fishing set vessel, time, location, school type, and catch in tons of the three major target tunas (skipjack tuna, yellowfin tuna, and bigeye tuna) were obtained from captain logbooks data corrected for bias in species composition based on port sampling using the T3 procedure (Pianet et al. 2000). Data on bycatch per species in number of individuals for each fishing set were obtained from onboard observers as part of two PS observer programs: the European Union-funded "Data Collection Framework" (DCF; Reg 2017/1004 and 2016/1251) and the "Observateur Commun Unique Permanent" (OCUP) program initiated and funded by the PS fishing industry (Goujon et al. 2017). Each line of the data corresponds to a fishing set, i.e. a single fishing activity involving encircling a tuna school with the seine and bringing the fish caught within the seine on board. The data extraction initially had 21 240 sets, of which 7188 sets had observer data.

Two fishing modes are commonly used in the Indian Ocean tropical tuna PS fishery (Kaplan et al. 2014). "Free-swimming school (FSC)" sets correspond to catch on tuna schools not associated with any object in the water and detected by crew members during a fishing trip, either visually or via use of technologies, such as bird radars. FSC sets are generally composed of monospecific individuals in which bycatch species are rare (Kaplan et al. 2014). "Floating object" (FOB) sets consist of fishing on tuna schools associated with natural objects at sea (e.g. logs, palm branches, or debris from human activities, such as marine cargo) or artificial floating objects specifically deployed by PS vessels to aggregate tunas, referred to as drifting fish aggregating devices (dFADs; Davies et al. 2014), which today represent the vast majority of floating objects (Maufroy et al. 2016).

FOB sets represent >80% of recent PS catch in the Indian Ocean (Floch et al. 2021) and have roughly three times the bycatch per set of FSC sets, including some emblematic species susceptible to overfishing, such as silky sharks, that are rarely caught in FSC sets (Kaplan et al. 2014). Therefore, we limited our analyses to FOB sets. After limiting data to FOB sets, our data set consisted of 16 281 sets distributed throughout the PS fishing zone in the Indian Ocean (Fig. 1), of which 6046 sets have observer coverage.

As bycatch data contain well over 100 species or species groups, many of which are very rare, we focused our study on four of the more common and emblematic bycatch species (Ruiz et al. 2018): silky shark (*Carcharhinus falciformis*), rough triggerfish (*Canthidermis maculata*), rainbow runner (*Elagatis bipinnulata*), and common dolphinfish (*Coryphaena hippurus*). Top predators such as silky sharks and common

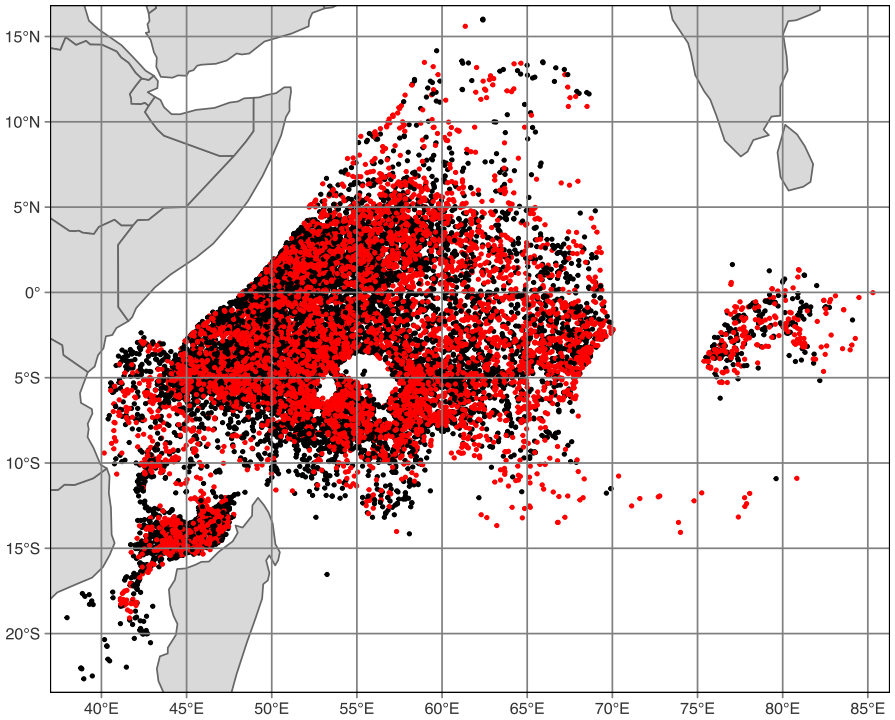

Figure 1. Map of the spatial distribution of FOB fishing sets in our dataset from the French PS fleet in the Indian Ocean between 2011 and 2018.Dark, black dots indicate sets without observer coverage, whereas lighter, red dots indicate sets with observer coverage.

dolphinfish are typically found as individuals or in small groups in PS sets, whereas rainbow runner and rough triggerfish are often caught as part of large aggregations. Silky shark is generally considered to be the PS bycatch species of most concern due to it being caught in relatively large numbers in FOB sets and due to its life history traits common to many sharks of high longevity and low reproductive rates (Dulvy et al. 2014). As such, though models are developed for all four bycatch species, certain results are only presented in the main text of the article for silky shark, with results for the other species being presented in the online supplementary materials.

Observer bycatch data separates bycatch observations into three categories: live release, dead release, and retained (i.e. landed). Given our objective of estimating mortality rates and a number of indications that sharks released live by PS vessels have high post-release mortality (e.g. Hutchinson 2015, Onandia et al. 2021), for the sake of this paper, we decided to sum all three categories into a total "bycatch" mortality observation for each species for sets with observer coverage. Inclusion of live-release individuals will have a tendency to overestimate true mortality rates.

Environmental data

As we wished to understand how bycatch rates vary as a function of spatial, temporal, and environmental covariates, we first identified a set of environmental variables that potentially impact bycatch in tropical tuna PS fisheries. Sea surface temperature (SST), salinity (SSS), and depth-integrated net primary production (NPP) were identified as general environmental variables potentially important for the distribution and abundance of pelagic species (e.g. Mannocci et al. 2020). Surface mixed layer depth (MLD) was also considered as it can limit the vertical distribution of species, modifying the accessibility of tuna schools to surface fisheries, among other effects (Arrizabalaga et al. 2015).

Environmental variables were obtained from Copernicus (the Earth observation component of the European Union's Space program) at different spatial scales and attributed to fishing observations by associating each point observation with Copernicus data of the containing grid cell and closest timestamp. Daily sea surface potential temperature and salinity, and surface mixed layer depth were obtained from the Copernicus—Global Ocean Physics Reanalysis based on the current available real-time global forecasting CMEMS system, having a $0.083^\circ \times 0.083^\circ$ resolution and covering the time period 01/01/1993–31/12/2019 (https://doi.org/10.48670/moi-00021, 01/09/2022, date last accessed). Daily primary production was obtained from the Copernicus-PISCES biogeochemical global hindcast having a $0.25^\circ \times 0.25^\circ$ horizontal resolution with daily means available from −5500 m to 0 m depth covering the time period 01/01/1993–31/12/2020 (https://doi.org/10.48670/moi-00019, 01/09/2022, date last accessed). Net primary production was calculated as the sum of primary production from the sea surface to 200 m depth.

Modeling bycatch rates with a Δ method

A conceptual overview of our modeling approach is presented in Fig. 2. In initial exploratory analyses of bycatch data, we noted that zeros (absences) are prevalent in the data at levels that are not consistent with standard statistical distributions. For example, 1346 (22%) of the total 6046 FOB sets with observer coverage included no silky shark bycatch (Fig. 3; Table 1). As such, we chose to develop Δ models to estimate total bycatch. The Δ method breaks catch down into two separate models, one for presence–absence and the other for abundance when present (i.e. the number of individuals of a bycatch

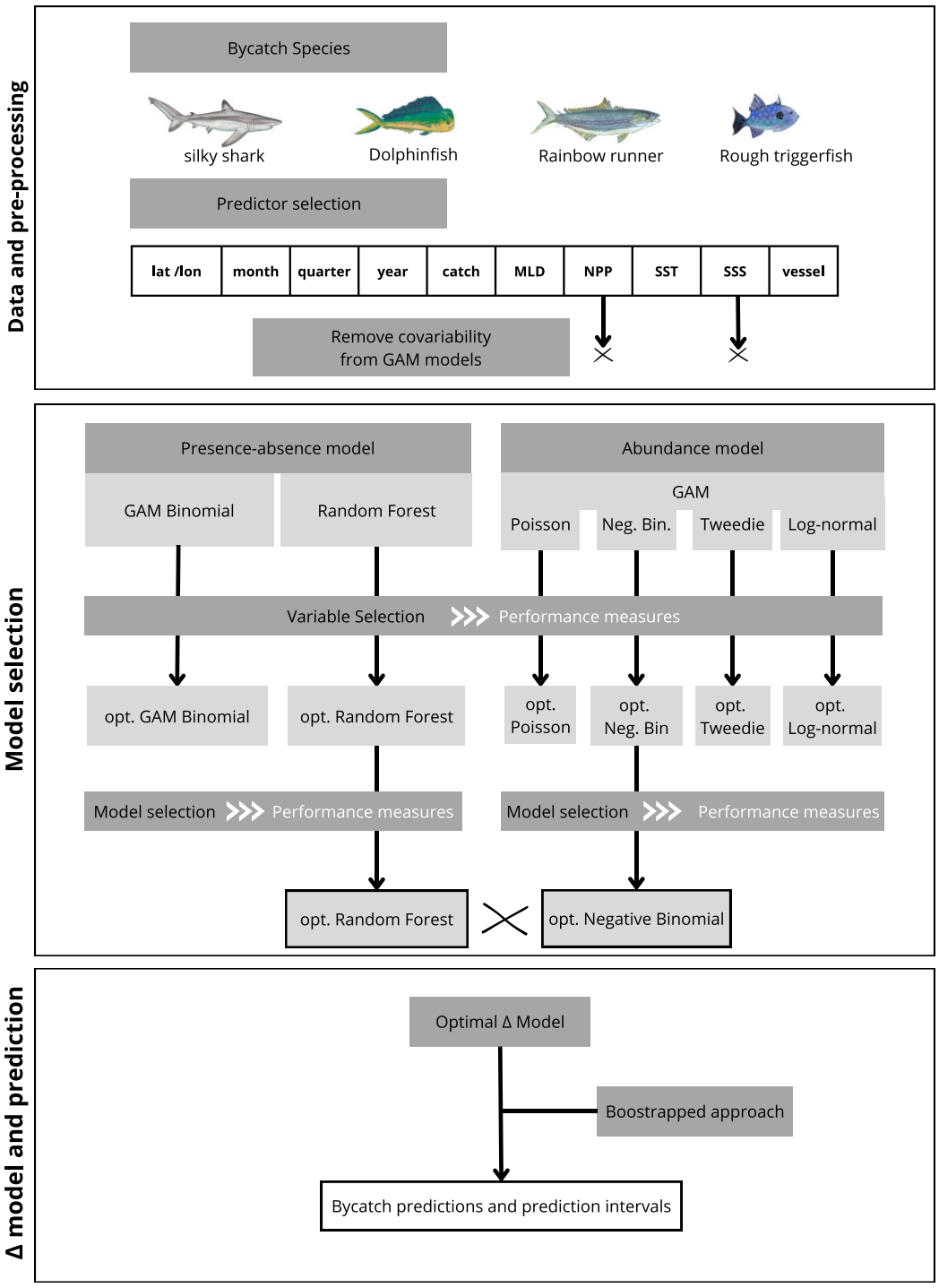

Figure 2. A conceptual overview of our approach to modeling bycatch rates of the four target species. The approach is generic, but some of the information presented, such as the modeling approaches and distribution families used to construct our final "optimal" Δ models used to predict bycatch rates represent results specific to this study. Fish illustrations ©Léa Ortega.

species caught in the fishing set given that at least one individual was caught; Ortiz and Arocha 2004). Total bycatch in a given set is estimated as the product of predictions from these two sub-models. This methodology has the advantages of allowing for different model formalisms and structures in the two components, thereby permitting greater flexibility than other approaches to dealing with large numbers of zero observations, such as zero-inflated models and integrated likelihood models.

Given the potential for non-linearity of the relationships between spatial, temporal, and environmental predictors and response variables (i.e. presence–absence or number of individuals of a bycatch species per set), we developed models that explicitly permit non-linearity. When modeling presence–absence of a bycatch species, sets for which any non-zero number of the species were recorded by observers were assigned a one, and all other sets were assigned a zero. We compared two overall modeling frameworks to model presence–absence: (i)

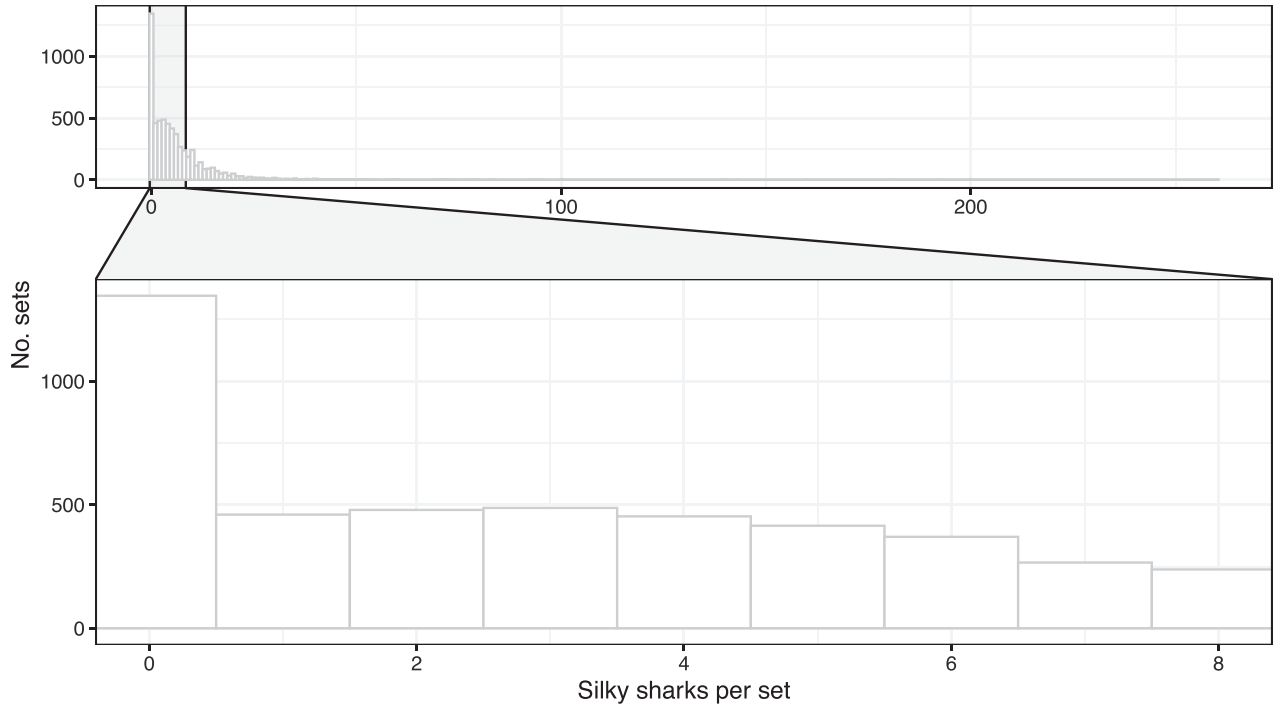

Figure 3. Histogram of silky shark bycatch per fishing set from observer data of the Indian Ocean French tropical tuna PS fishery on FOBs between 2011 and 2018.

Table 1. Prevalence of each of the 4 bycatch species in our dataset of 16 281 fishing sets, of which 6046 sets were covered by observers.

	No. of sets	Prevalence (%)	No. of individuals	Mean no./Set	Mean no./Set when present	CV[a]
Silky shark	4700	77.7	39 635	6.56	8.43	1.59
Rough triggerfish	4604	76.1	836 083	138.29	181.60	2.19
Rainbow runner	4346	71.9	313 810	51.90	72.21	4.74
Common dolphinfish	4213	69.7	115 547	19.11	27.43	1.80

[a] CV = coefficient of variability calculated as the ratio between the standard deviation of the number of individuals caught over all sets and the mean number of individuals per set.

general additive models (GAMs; Wood 2017) with a binomial distribution for the response variable, and (ii) random forest (RF) models (Breiman 2001). Details of both approaches are given below in Sections 2.3.2 and 2.3.3.

When modeling abundance when present for a bycatch species, the number of individuals observed (or a direct function thereof) was used as the response variable and all sets for which the species was not observed were removed. For abundance, only GAMs were developed, but several candidate statistical distributions for the response variable were explored: Poisson, negative binomial, Tweedie, and log-normal.

Predictors variables

Spatial, temporal, environmental, and vessel identifier variables were included as covariates in models (Table 2). Spatio-temporal factors may have an effect on bycatch as species abundance and catchability vary on seasonal, inter-annual, and regional scales (Kaplan et al. 2014). To take this into account, models included longitude and latitude, as well as the month and year, as explanatory variables. Month and year were treated as continuous variables, and month was modeled as a circular variable in GAMs. Both direct and interaction terms were included in GAMs when examining the impacts of longitude and latitude, and of year and month [i.e. in the notation of the `mgcv` R package, both "s()" and "ti()" terms were included in models]. The effects of longitude and latitude in GAMs were stratified by quarter of the year. Unfortunately, for random forest, it is not possible to treat month as a circular variable. Nevertheless, we chose to keep month as a continuous variable, as this allows random forest models to associate (or not) temporally adjacent months (other than December and January) in an ordered fashion.

In addition to these spatio-temporal variables and the environmental variables described in Section 2.2, the total catch of target tunas (i.e. all three major tropical tuna species summed together) in tons per fishing set was also included in our models, as it has been extensively used as a predictor of PS bycatch (Amandè et al. 2010). Finally, we included the potential for persistent differences among vessels in terms of bycatch species composition via the inclusion of a categorical vessel identifier in models as a fixed effect so as to be able to predict bycatch per set for specific vessels.

Models including all the predictor variables listed above will be referred to as "full" models, whereas models post-variable selection to reduce model complexity and/or decrease problems of model overfitting will be referred to as "optimal" models (see Section 2.3.4 for the variable selection

Table 2. Description of potential predictors of bycatch per FOB fishing set.

Full name	Abbreviation[a]	Units	Type	Mean	95% Interval
Latitude / Longitude	*lat / lon*	Decimal degrees	Continuous	–	–
Month	*month*	[1:12]	Continuous	–	–
Quarter	*quarter*	[1:4]	Categorical	–	–
Year	*year*	[2011:2018]	Continuous	–	–
Trop. tuna catch	*catch*	Tons	Continuous	27.27	[1.98; 103.69]
Mixed layer depth	*MLD*	Meters	Continuous	18.62	[10.53; 47.89]
Net primary production	*NPP*	mg/d/m^3	Continuous	179.33	[48.06; 597.05]
Sea surface temperature	*SST*	°C	Continuous	28.48	[25.47; 30.64]
Sea surface salinity	*SSS*	Practical salinity unit	Continuous	35.29	[34.47; 36.05]
Vessel identifier	*vessel*	–	Categorical	–	–

[a]The "Abbreviation" column gives the short name for each variable that will be used in text, figures, and tables. The "95% Interval" and "Mean" columns refer to the central 95% of the data and the mean of the data over FOB fishing sets with observer coverage in our dataset.

methodology used). Before fitting models, cross-correlation analysis among predictors was carried out to understand potential for variable redundancy in model formulations and, if need be, remove any redundant variables (see Results section).

GAMs for presence–absence and abundance of bycatch

In GAMs, the effects of predictor variables on the response variable are modeled as the sum of a certain number of non-linear, smoothing spline functions. Various classes of smoothing splines exist, including thin plate regression splines, which we used for modeling the effect of single continuous variables, and cyclic cubic regression splines, which we used for the circular variable month. Tensor product regression splines were used for modeling two-dimensional variability in lon–lat and year–month space, with cubic regression splines being the basis for these tensor products for all dimensions but month, for which cyclic cubic regression splines were used. GAMs were run using the R package `mgcv` version 1.9.1 (Wood 2017) in R version 4.3.3 (29/02/2024) (R Core Team 2023); see Online Supplementary Materials for more details and specific package versions used for all analyses.

The maximum number of basis smoothing spline functions (k) to be considered when estimating the effect of a predictor variable is typically specified in GAMs (regularization is used during the model fitting process to only include a subset of this maximum number of splines in the final model predictions). For the spatial effect, estimations were stratified by quarter and a maximum of 13 splines were used for each dimension, following the example of Wain et al. (2021) studying the same geographic region and similar data. For the month and year effects, 12 circular cubic splines and 8 thin plate regression splines were used, respectively, these values being the maximum possible values in each case. For other predictor variables, default values set in the `mgcv` R package for the maximum number of splines were used (typically $k = 10$; Wood 2017).

To assess model fit, QQ-plots of residual quantiles with respect to the theoretical quantiles were examined, along with plots of residuals as a function of predicted values and examination of model performance indicators developed below (Section 2.3.5).

When modeling presence–absence of a given bycatch species, a binomial distribution was assumed for the response variable with the standard "logit" link function. When modeling abundance when present, several distributions were tested (Table 3): the Poisson distribution, which is theoretically suitable for count data; the negative binomial, which accounts for positive definite data, like the Poisson distribution, but allows for overdispersion; the Tweedie distribution, which accounts for both overdispersion and potentially excessive numbers of small observations; and a normal distribution for the log-transformed catch data. Given that the abundance data included no zeros, the number of individuals per set minus one was used as the response variable for the Poisson, negative binomial, and Tweedie distributions (as these distributions start at zero).

Random forest models for presence–absence of bycatch

RF models were also developed as an alternative approach to modeling presence–absence of bycatch species. RF models are a machine learning algorithm based on decision trees where each node of a tree represents a cutoff threshold on one of the predictor variables, with the cutoff selected to maximize correct classification of observations (Breiman 2001). RF models use a large number of such trees, weighting them in the final model prediction according to their predictive capacity based on the results of internal cross-validation. As predictor variables can occur multiple times in each tree and the ensemble average of many trees is used in making predictions, the method implicitly allows for highly non-linear effects and interactions among predictors. RF models were run using the R package `randomForest` version 4.7.1.1 (Liaw and Wiener 2002).

Two free parameters in RF models, the total number of decision trees to use and the number of predictor variables to randomly select and test for predictive capacity at each node, were fixed based on analyses of the full RF model using the `tuneRF` function of the `randomForest` package (Liaw and Wiener 2002).

Model selection

Predictor variable selection was used to identify the most parsimonious models that accurately represent the data. First, several distribution families for residuals were tested for "abundance when present" GAMs with all the predictor variables included (i.e. the "full model"): Poisson, Tweedie, negative binomial, and log-normal. The selection of the best distribution family to use was based on examination of standard diagnostic plots for GAMs (e.g. QQ-plots and plots of residuals versus predicted values), as well as the predictive performance (e.g. deviance explained) of the models, in both cases considering primarily results for silky shark given its importance as a bycatch species.

Table 3. Summary of the different GAM distributions of residuals used to model presence–absence and abundance when present of silky shark bycatch per fishing set.

Distribution	Model type	Data type	Transformation	Link function
Poisson	Abundance	No. of individuals	N-1	log
Negative binomial	Abundance	No. of individuals	N-1	log
Tweedie	Abundance	No. of individuals	N-1	log
Log-normal	Abundance	No. of individuals	log(N)	identity
Binomial	Presence–absence	0 = absence / 1 = presence	N	logit

In the "Model type" column, "Abundance" indicates that abundance when present was modeled. The "Transformation" column indicates the transformation that was applied to the response variable before running the model, with "N" indicating that no transformation was applied.

Second, variable selection was carried out for both GAMs and RF models to choose the best set of predictor variables for presence–absence and abundance. The overarching principle of parsimony was used to select the simplest model with fewest variables and computational complexity that accurately describes the data, thereby reducing potential for model overfitting (Gregorutti et al. 2017).

For GAMs, the model using the set of all explanatory variables (i.e. the full GAM) was run. Then backwards selection using *P*-values was carried out by sequentially dropping the single term with the highest non-significant *P*-value from the model and re-fitting, until all terms were significant (Wood 2017). This approach was used as stepwise regression based on the Akaike Information Criterion (AIC) would have been too time-consuming to run due to the large number of complex models that would need to be estimated and because model selection based on *P*-values has been shown to produce roughly equivalent results to that based on AIC (Murtaugh 2014). When a smooth term was dropped, it was replaced with the corresponding non-smoothed term [e.g. "s(year)" would be replaced with "year"] to test for significance in the absence of the smooth before proceeding to remove other terms from the model.

For the RF model, automatic variable selection using the R package `VSURF` version 1.2.0 (Genuer et al. 2022) was used, constructing the nested collection of RF models and selecting the variables involved in the model leading to the smallest out-of-bag estimated error percentage (OOB error). The "variable selection procedure for prediction" procedure of the `VSURF` package was used with the parameters selected at the tuning step (i.e. the number of trees and number of variables examined per node determined by the `tuneRF` function). The `VSURF` method has been tested on several data sets, which showed that it was reliable and versatile (Genuer et al. 2015).

The models resulting from these selection processes are referred to as "optimal GAMs" (for presence–absence or abundance) and "optimal RF models" (for presence–absence).

Model performance statistics

In order to compare GAMs and RF models for species presence–absence, as well as to compare GAM models across residual distributions and to assess combined Δ models, a set of standard model performance statistics was developed and calculated in the same way for all comparable models. For presence–absence models, the percentage of well-classified observations was calculated and compared across models. An observation was considered to be well classified if the model estimated probability of presence (absence) was ≥ 0.5 (<0.5) when the true observation was a presence (absence).

Goodness-of-fit metrics root mean square error (RMSE) and mean absolute error (MAE) were calculated for all presence–absence, abundance, and Δ models as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(prediction_i - observation_i)^2}{N}} \text{ and}$$

$$MAE = \frac{\sum_{i=1}^{N}|prediction_i - observation_i|}{N},$$

where N is the total number of observations, and $prediction_i$ and $observation_i$ are the model predicted and observed values for the i^{th} observation.

For abundance when present and combined Δ models, we also calculated the fraction of the deviance explained by the model. Though GAMs report the deviance explained, we chose not to use these model-reported values to avoid issues with whether or not they are calculated in response or link space and to use a single approach for both GAMs and Δ models. Fraction of deviance explained was calculated as:

$$\frac{deviance_{total} - deviance_{residual}}{deviance_{total}}, \quad (1)$$

where

$$deviance_{residual} = \sum_{i=1}^{N}(prediction_i - observation_i)^2$$

and

$$deviance_{total} = \sum_{i=1}^{N}(\overline{observation} - observation_i)^2,$$

with $\overline{observation}$ being the mean of all observations (for a given species and excluding or including zeros for abundance when present versus Δ models, respectively).

In order to assess the robustness of performance statistics and potential for model overfitting, 10-fold cross-validation was used to develop model assessments of predictive capacity that are independent of the training data used to construct models. Our data set was broken into 10 randomly selected equal parts, using 9 of them to train the model and 1 of them to validate the model. Model performance was measured for each validation fold in terms of RMSE, MAE, the percentage of well classified observations for presence–absence models, and the deviance explained for abundance and Δ models. The average of the 10 performance measures (from using each of the 10-folds as the validation data set) was used as the overall performance of the model in cross-validation (Berrar 2018). Model performance measures calculated as the average of these 10 values are prefixed with "v-" to distinguish them from performance statistics based on the model trained with

the entire data set. For example, "v-RMSE" indicates the average of the 10 RMSE calculated from each of the 10 folds being used as the validation data set. To assure comparability of these cross-validation measures across models, the same random folds were used for all models based on equivalent numbers of observations (i.e. a single set of folds was used for presence–absence and Δ models, and one set of folds was used for each of the four bycatch species when running abundance when present models).

Model prediciton intervals

As one of our objectives is to be able to predict bycatch for fishing sets lacking observer data, we needed to assess our confidence in model predictions, i.e. not the certainty in the expectation value for a given set of explanatory variable values (i.e. the confidence interval), but rather the certainty in new predictions. These "prediction intervals" can be evaluated directly for general linear models (GLMs) using functionality already integrated into standard R functions for GLMs. However, for GAMs and combined Δ models, standard software tools for calculating prediction intervals are lacking. Instead, we decided to use a bootstrap approach, following Andersen (2022).

First, the Δ model expected mean bycatch prediction of a set, ($\widehat{\mu}_C$), was estimated as the product of mean predictions from the presence–absence and abundance components of the model $\widehat{\mu}_C = \widehat{\pi} \times \widehat{\eta}$. $\widehat{\pi} = Pr(presence)$ is the expected mean presence probability and $\widehat{\eta}$ is the expected abundance when present, both having been back-transformed as necessary to response variable space (e.g. following Fletcher 2008; for log-normal models). Using these mean expected values, we calculated performance statistics of the overall Δ model (i.e. RMSE, MAE, and deviance explained) using the formulas given above.

Next, we randomly generated 10^4 sets of potential coefficients of the GAMs for both the GAM presence–absence and GAM abundance models. This was done using the Cholesky Trick, following the approach of Andersen (2022). Simply put, this approach uses the optimal coefficients estimated by the model and the standard errors of these coefficients to estimate a distribution for each coefficient from which potential coefficients are randomly drawn.

Then, for each of these 10^4 sets of potential model coefficients, model predictions in link-function space, including (via the standard errors as described above) uncertainty in parameters were generated. These were then back-transformed into response units using the appropriate inverse link function. Final predictions were generated by randomly drawing from the appropriate model distribution family (e.g. binomial for presence–absence), including, for two parameter distributions such as the negative binomial, the residual variance of each of the GAMs (i.e. the scale parameter). This final step integrates, via the model scale parameter, unexplained variance in observations into predictions. This produced 10^4 simulated predictions, $\widehat{\eta}^*$, for the abundance GAM and 10^4 $\widehat{\pi}^*$ for the presence–absence GAM, for each fishing set in the dataset.

RF models of presence–absence have no equivalent for the uncertainty in model parameter estimates and, therefore, lack parameter estimates with standard errors. Instead, for each fishing set in the dataset, we drew 10^4 values of 0 or 1 from a single binomial distribution with probability of presence determined by the prediction of the RF model, thereby obtaining 10^4 predictions of presence–absence, $\widehat{\pi}^*$, for each fishing set that include fundamental uncertainty in whether a species is present due to the binomial probability of presence, but not uncertainty in the model-estimated binomial probability of presence parameter itself.

The last step is to combine the simulated predictions of the GAM for abundance when present and the GAM or RF model for presence–absence by calculating $\widehat{\mu}_C^* = \widehat{\pi}^* \times \widehat{\eta}^*$. These predictions include both uncertainty in model estimates of parameters (for GAMs) and uncertainty due to variance not explained by the model via random selection of values from the model distribution family, including where appropriate the model-estimated scale parameter. We then obtain a simulated distribution of predicted bycatch based on the 10^4 values of $\widehat{\mu}_C^*$ for each fishing set in the dataset, from which a 95% prediction interval is calculated as the 2.5% and 97.5% quantiles of the simulated data.

Tests for residual spatio-temporal autocorrelation

We assessed whether or not spatial, temporal, and spatio-temporal autocorrelation were adequately accounted for in model formulations by calculating variograms and autocorrelation functions (ACFs) on optimal Δ model residuals following the general approach of Legendre and Legendre (2012). Spatial variograms were estimated using the R package `geoR` version 1.9.3 (Ribeiro et al. 2023), temporal autocorrelation functions were calculated using the `acf` function of R, and spatio-temporal variograms were estimated using the R package `gstat` version 2.1.1 (Gräler et al. 2016). As potentially multiple fishing sets with observer data existed for a given day, when calculating temporal ACFs, observations were randomly selected 100 times, for each day generating 100 estimates for the ACF, which were summarized to characterize temporal autocorrelation. A small number of observations (<0.1%) with identical spatial positions were randomly removed when calculating spatial or spatio-temporal variograms. Results, presented in Supplementary Figs S15–S18, indicated an absence of significant unexplained autocorrelation, so we decided that the addition of spatial and/or temporal covariance terms to models was unnecessary.

Prediction performance of Δ models at different scales

The predictions of the optimized Δ models for each of the four bycatch species were aggregated in time and space in order to examine our predictive capacity at different and larger spatio-temporal scales that may be of interest for fisheries management of bycatch. The temporal scales of aggregation were: month (i.e. aggregating across all fishing sets for a specific month of a specific year), 2 months (bi-month), 3 months (quarter), 6 months (half year), year, and the entire time period (all time). The spatial scales of aggregation were: 1° × 1° and 5° × 5° spatial grid cells, and the entire western Indian Ocean fishing area.

The Δ model prediction for bycatch in a stratum S, $\widehat{\mu}_{C_S}$, is calculated as:

$$\widehat{\mu}_{C_S} = \sum_{i=1}^{J_S} \widehat{\mu}_{C_i},$$

with J_S the number of observations in the stratum and $\widehat{\mu}_{C_i}$ the predicted bycatch for fishing set i in the stratum S.

Prediction performance statistics were calculated for aggregated predictions as described above. For example, RMSE for a given level of spatio-temporal aggregation was calculated as:

$$RMSE = \sqrt{\frac{\sum_S \left(\widehat{\mu}_{C_S} - \sum_{i=1}^{J_S} observation_i\right)^2}{N_S}},$$

where, in this case, N_S is the number of strata S at the given level of aggregation. To compare performance statistics across multiple levels of aggregation, we calculate the normalized RMSE and MAE statistics (denoted "$RMSE_{norm}$" and "MAE_{norm}") for each level of aggregation by dividing the RMSE and MAE by the average aggregated bycatch prediction. For example, for $RMSE_{norm}$:

$$RMSE_{norm} = \frac{RMSE}{\frac{\sum_S \widehat{\mu}_{C_S}}{N_S}}.$$

The same 10-fold cross-validation procedure that was used to assess uncertainty in model predictions for individual fishing sets was also used on the aggregated data (i.e. the folds were based on the original fishing set data, not the aggregates). For each fold, we calculated the $RMSE_{norm}$ and MAE_{norm} as described above, and then these 10 results were averaged together to calculate the final error estimates (i.e. v-$RMSE_{norm}$ and v-MAE_{norm}).

Prediction intervals for aggregate predictions were calculated in a similar way. The procedure described above was used to develop 10^4 predictions for each fishing set, which were then aggregated by strata to produce 10^4 predictions of bycatch in each strata, the 2.5–97.5% quantiles of which were used to define the prediction interval.

Δ model performance comparison for different bycatch species

We compared Δ model results for our four bycatch species (silky shark, rough triggerfish, rainbow runner, and common dolphinfish) to test whether predictive capacity differs among different species with different life history traits. Optimal models (i.e. models including predictors optimized for each species) for both presence–absence and abundance when present were used to develop Δ model predictions for each species. The normalized RMSE and MAE (i.e. $RMSE_{norm}$ and MAE_{norm}), with and without 10-fold cross-validation, were used as a common indicator of prediction error across species with different levels of mean prevalence and abundance.

Prediction of total bycatch

The total bycatch of silky sharks for the French PS fleet in the Indian Ocean between 2011 and 2018 was calculated for a given year as the sum of observed bycatch for those fishing sets with observer data and model predictions of bycatch for those fishing sets without observer data. These estimates were calculated using the optimal Δ model for the corresponding species. In order to understand spatial patterns of predicted total bycatch, we also calculated bycatch per set and prediction uncertainty in bycatch per set on $5^\circ \times 5^\circ$ grid cells for the entire study time period. Prediction uncertainty was calculated as the difference between the 97.5% and 2.5% quantiles of the predicted bycatch in each grid cell (uncertainty being assumed to be zero for sets with observer data; Valliant et al. 2000).

When making these predictions, one issue is that six French vessels had no observer coverage of fishing activities over the study period, which represents ~26% of the fleet (this absence generally being due to space constraints placed on certain vessels by the need to carry military personnel to prevent pirate attacks). This prohibited prediction for fishing sets by these vessels using models including vessel identifier as an explanatory variable. We therefore developed Δ models without vessel identifier as an explanatory variable specifically for predicting bycatch of these vessels. Three different sets of models were developed without the vessel identifier predictor: (i) "full w/o vessel": corresponding to models including all predictors except *vessel*; (ii) "optimal w/o vessel": corresponding to models including the optimal (RF or GAM) predictors with *vessel* removed after model selection; and (iii) "no vessel optimal": corresponding to models where the *vessel* predictor was removed prior to predictor variable selection. Performance statistics of these three sets of models were compared to assess which approach was best for predicting bycatch for these vessels without observer data.

Final total bycatch predictions were therefore the combination of observed bycatch for sets with observer data, predicted bycatch by optimal Δ models for sets without observer data for vessels with observer data, and predicted bycatch by Δ models without *vessel* as a predictor for vessels with no observer data in the Indian Ocean over the study time period.

Comparison with the bycatch-over-landings ratio approach

A bycatch-over-landings extrapolation approach is currently used to estimate total PS bycatch by fleet for reporting to the Indian Ocean Tuna Commission (IOTC) and other tuna RFMOs. The approach is based on first stratifying the catch by quarter and by year. For each combination of quarter and year, the ratio of total bycatch of a given species to total catch of target tunas is calculated, summing over all fishing sets i for which observer data is available: $ratio = \frac{\sum_i observation_i}{\sum_i T_i}$ with T_i the catch of target tunas in set i. Given this ratio, the expected bycatch of the species in a given set i is calculated as the product of the ratio and the total catch of target tunas in the set $\widehat{\mu}_{C_i} = ratio \times T_i$.

We compared the Δ model and the bycatch-over-landings ratio approach to estimating bycatch using the standard error estimates described above for the Δ model (i.e. 10-fold estimation of RMSE, MAE, and percent deviance explained).

The total annual bycatch of the four study species was also estimated by this approach, using the bycatch-over-landings extrapolation exclusively for sets without observer coverage, and compared with the estimation from the optimal Δ model.

Results

We begin by presenting cross-correlations among predictor variables, followed by results from both the presence–absence and abundance when present components of the Δ model. This is followed by examinations of the certainty of model estimates at different spatio-temporal scales of aggregation. Finally, we present overall predictions of total bycatch of the four bycatch species for the French fleet in the Indian Ocean

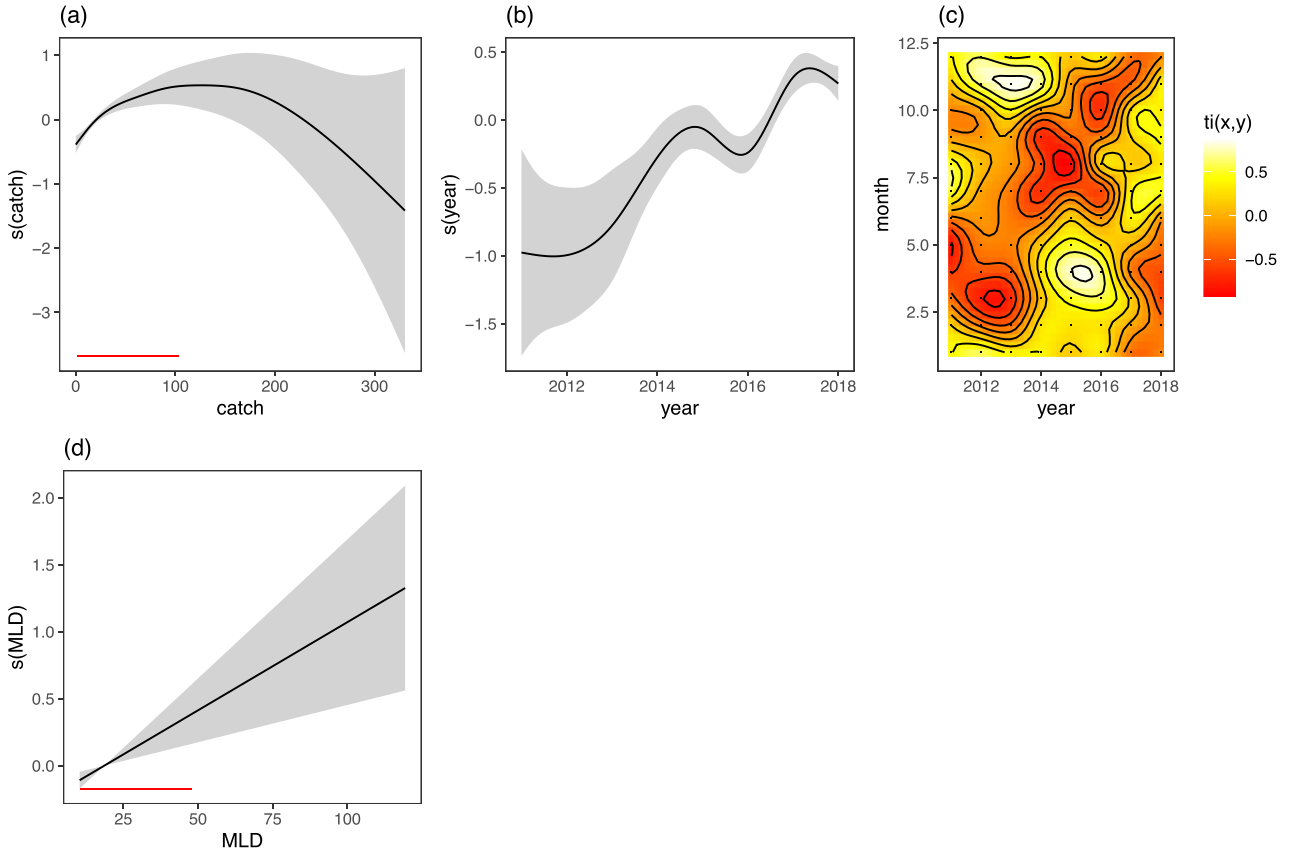

Figure 4. Marginal effects of non-spatial predictors with smooths on the presence of silky sharks in floating object fishing sets of the French PS fleet in the Indian Ocean. The effects were estimated with the optimal GAM for presence–absence and are listed in the following order: (a) s(catch); (b) s(year); (c) ti(year,month); and (d) s(MLD). In all panels except the 2D interaction plot (c), the horizontal axis indicates the value of the predictor variable and the vertical axis indicates the marginal effect of the predictor on presence–absence in the scale of the linear predictor, with positive values indicating an increase in probability of presence. Solid, black curves indicate the best estimate of the mean effect, whereas gray areas indicate 95% confidence intervals around the mean effect. The horizontal red lines at the bottom of panels (a) and (d) represent the central 95% of the distribution of the corresponding predictor variable. For the 2D interaction plot in year–month (c), the effect is plotted as a raster plot with yellow colors indicating a positive effect on probability of presence.

and compare model results with predictions from the simpler bycatch-over-landings ratio approach regularly used by scientists and managers to estimate bycatch in data-limited situations.

Covariability of predictor variables

In order to understand predictor redundancy in model formulations, we analyzed cross-correlations between predictor variables. Non-negligible correlations were found between multiple pairs of predictor variables, with correlations >0.5, for latitude and SSS, as well as for SST and NPP. All other correlations are <0.5. We therefore decided to remove SSS and NPP from GAM models to enhance model stability and interpretability. We chose to retain latitude as latitude is modeled in GAMs jointly with longitude making it difficult to remove, and to retain SST as this variable is thought to have a more immediate impact on tropical tuna behavior and physiology and is also more readily available from both satellite data and oceanographic model outputs.

All predictor variables, including SSS and NPP, were included in full RF models as RF models do not require that predictor variables be uncorrelated.

Presence–absence models of bycatch

Binomial GAM models

In general, optimal models for the four study species are fairly similar. The optimal model for rainbow runner is:

$$\begin{aligned}\Pr(\text{Rainbow runner}) &\sim s(catch) + s(lon, by = quarter) + s(lat, by = quarter) \\ &+ ti(lon, lat, by = quarter) + s(month) + s(year) \\ &+ ti(year, month) + s(MLD) + s(SST) + vessel\end{aligned}$$

The model for rough triggerfish is identical except for exclusion of the term *s(catch)*, whereas the model for common dolphinfish is identical except for the exclusion of the term *s(SST)*. The model for silky shark differs from that for rainbow runner by the exclusion of the terms *s(lon,by=quarter)*, *s(month)*, and *s(SST)*. The lon–lat by quarter and year–month interaction terms were retained for all species, as was the smooth on SST, and the smooth on catch was retained for three of the four species.

The marginal effects of predictors on probability of presence varied across species, but there are several somewhat consistent trends (Figs 4 and 5; Supplementary Figs S1–S6).

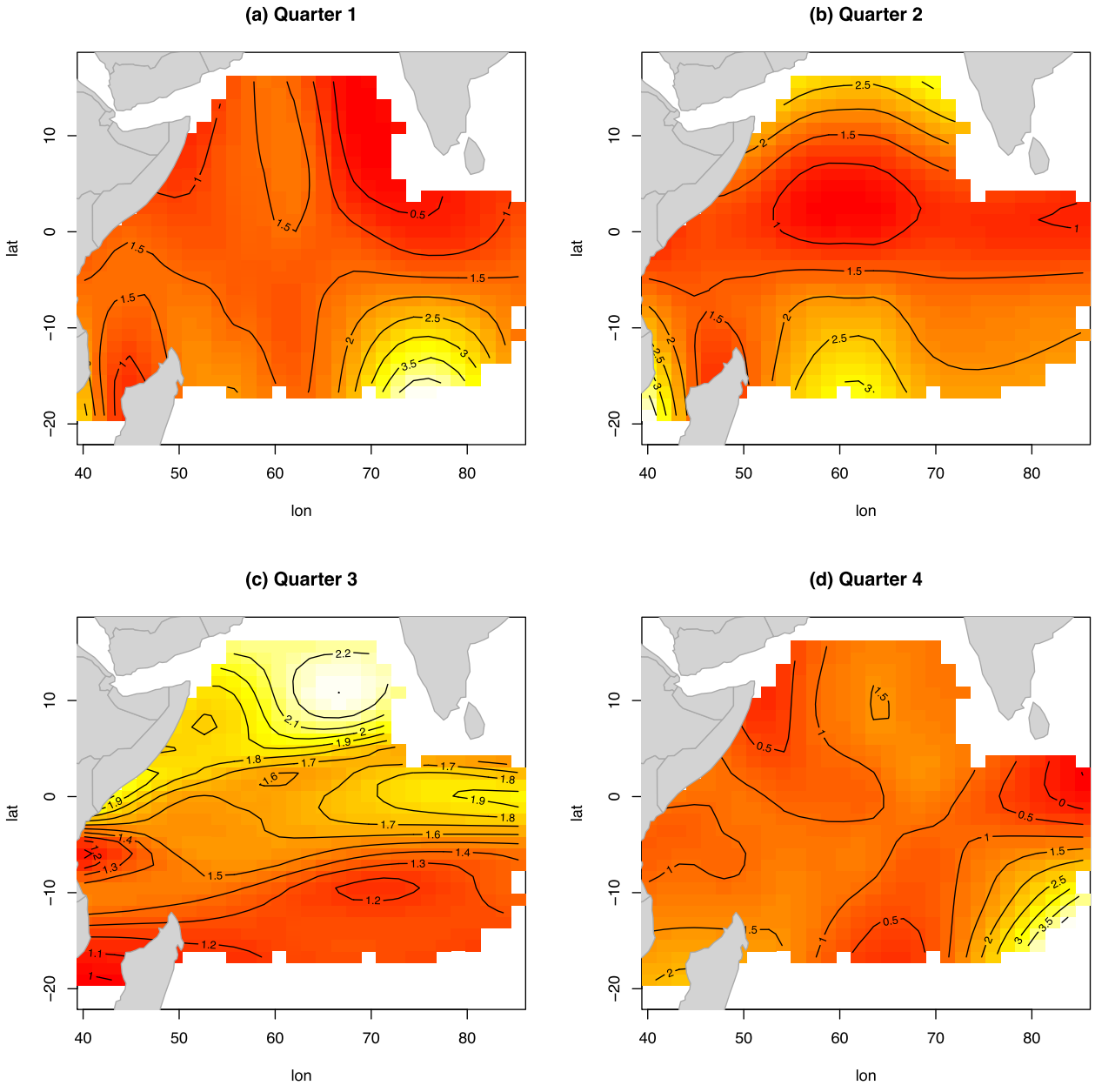

Figure 5. Marginal effects of spatial predictors by quarter on the presence of silky sharks in floating object fishing sets of the French PS fleet in the Indian Ocean. The effects were estimated for the optimal GAM for presence–absence using the `mgcv::vis.gam` function with all predictors other than lon, lat, quarter, and month fixed at their median values. Month was set to be the central month of the corresponding quarter (e.g. for quarter 1, month was set to 2). Lighter, yellow colors indicate a positive effect on probability of presence and numbers on contour lines indicate the marginal effect of spatial variables on the scale of the linear predictor.

For silky shark, rainbow runner, and, to a lesser degree, common dolphinfish, the effect of target tuna catch for small to medium set sizes shows an increasing trend. Target tuna catch had a particularly strong impact on presence of silky sharks (Fig. 4a), for which the trend is increasing more or less linearly up to set sizes of ∼150 t (note that mean set size in our data is 27 t and only 0.8% of sets are >150 t), suggesting that model results are unlikely to be reliable above this limit. For all species, increasing MLD increases probability of presence for the range of values typically encountered in the data, and for species with *s(month)* included in the model, probability of presence was higher in the second half of the year than in the first half of the year. The marginal effects of other predictors and interactions are more variable across species (and for spatial effects, quarters) and difficult to interpret, though there does appear to be a consistent peak in bycatch presence in 2012–2014 for all species except silky shark.

Random forest

For silky shark, the number of variables randomly tested at each node for the full RF model (i.e. the model including all predictors) was estimated by `tuneRF` to be three, and plots of the error as a function of the number of trees used (not shown) indicate that errors stabilize >200 trees. To be certain that RF model results were stable, the number of trees in the forest was fixed at 400, somewhat higher than this minimum.

Table 4. Performance comparison of presence–absence models of the four bycatch species in floating object fishing sets of the French tropical tuna PS fleet in the Indian Ocean.

Species	Model	RMSE	MAE	well class. (%)	v-RMSE	v-MAE	v-well class. (%)	AIC	OOB err. (%)
Silky shark	Full GAM	0.39	0.31	78.3	0.40	0.32	77.6	5985.2	–
Silky shark	Optimal GAM	0.39	0.31	78.3	0.40	0.32	77.5	5985.1	–
Silky shark	Full RF model	0.15	0.12	100.0	0.40	0.31	78.2	–	21.9
Silky shark	Optimal RF model	0.15	0.11	100.0	0.40	0.31	78.3	–	21.8
Rough triggerfish	Full GAM	0.40	0.32	77.6	0.41	0.33	77.0	6122.1	–
Rough triggerfish	Optimal GAM	0.40	0.32	77.5	0.41	0.33	76.8	6123.9	–
Rough triggerfish	Full RF model	0.15	0.11	100.0	0.40	0.31	77.7	–	22.5
Rough triggerfish	Optimal RF model	0.19	0.13	98.4	0.42	0.29	75.5	–	24.7
Rainbow runner	Full GAM	0.42	0.36	73.4	0.44	0.38	72.0	6712.1	–
Rainbow runner	Optimal GAM	0.42	0.36	73.4	0.44	0.38	72.0	6712.1	–
Rainbow runner	Full RF model	0.16	0.13	100.0	0.42	0.35	73.9	–	25.8
Rainbow runner	Optimal RF model	0.16	0.13	99.9	0.43	0.34	73.4	–	25.9
Common dolphinfish	Full GAM	0.43	0.37	72.9	0.44	0.38	71.2	6753.9	–
Common dolphinfish	Optimal GAM	0.43	0.37	72.9	0.44	0.38	71.4	6753.3	–
Common dolphinfish	Full RF model	0.16	0.13	100.0	0.43	0.36	73.1	–	27.2
Common dolphinfish	Optimal RF model	0.16	0.13	99.9	0.43	0.35	72.7	–	27.2

Results for the full and optimal GAM and RF models are presented. Data were limited to sets with observers on board between 2011 and 2018. Statistics preceded with "v-" (e.g. v-RMSE, v-MAE) were computed as the average over results from 10-fold cross-validation.

As shown below, the variable selection process using VSURF (Genuer et al. 2015) yielded optimal RF models that were more variable in composition than the variable selection process for GAMs:

$$\begin{aligned}\Pr(\text{Silky shark}) \sim\ & catch + lon + lat + month \\ & + year + NPP + SST \\ & + SSS + vessel\end{aligned}$$

$$\Pr(\text{Rough triggerfish}) \sim lat + month + year + vessel$$

$$\begin{aligned}\Pr(\text{Rainbow runner}) \sim\ & lon + lat + month + year \\ & + NPP + SST + SSS + vessel\end{aligned}$$

$$\begin{aligned}\Pr(\text{Common dolphinfish}) \sim\ & lon + lat + month + year \\ & + SSS + vessel\end{aligned}$$

All models retained at least *lat*, *month*, *year*, and *vessel* as predictors, with *lon* and *SSS* also being selected for three out of four species. Surprisingly, catch of target tunas (*catch*) was only selected as a predictor for silky shark presence and *MLD* was never selected, despite both being identified as important variables in GAMs of presence–absence.

The performance of these RF models was compared both with full RF models and with RF models using the optimal set of predictors identified for GAM presence–absence models and generally found to be superior to both (e.g. in terms of lower OOB error or higher percent well classified in cross-validation), though differences were small and performance of the optimal RF model for rough triggerfish presence was somewhat lower than that of the full model (Supplementary Table S2). For example, for silky shark, the optimal RF model has an OOB error of 21.8%, whereas RF models with all predictors and the same predictor variables as the optimal GAM for presence–absence had higher OOB errors of 21.88–22.68%, respectively (Supplementary Table S2).

When vessel identifier is removed as a predictor for RF models (in order to predict bycatch for vessels without observer data over the study period), the VSURF variable selection procedure identified models that were identical to optimal RF presence–absence models with *vessel* removed, except for common dolphinfish, for which *NPP* and *SST* were added as predictors (Supplementary Table S3). These models had consistently worse performance statistics than optimal RF models including vessel identifiers (see next section).

Performance comparison for presence–absence models

Overall, optimal models determined by variable selection have similar performance statistics to full models, and RF models marginally outperform GAMs for presence–absence (e.g. in terms of well-classification rate in cross-validation; Table 4). RF models show signs of overfitting. For example, the optimal RF model for silky shark presence–absence has a well-classified rate of 100% in training that decreases to 78.3% in cross-validation, whereas the optimal GAM has a well-classified rate of 78.3% that becomes 77.5% in cross-validation, a much smaller change. Nevertheless, optimal RF models outperform optimal GAMs in predictive power, with lower RMSEs and MAEs and higher well-classified rates in cross-validation for all species except rough triggerfish, for which results are mixed with some performance statistics being higher for RF models, whereas others are higher for GAMs. Based on these results, we decided to use the optimal RF models for all further presence–absence and combined Δ model analyses (including rough triggerfish for consistency with the other species; Fig. 2).

When vessel identifier is removed as a predictor from presence to absence models both before and after variable selection, there is a noticeable drop in performance statistics for both GAMs (Supplementary Table S1) and RF models (Supplementary Table S2). In general, performance of no vessel optimal RF models (i.e. optimal RF models for which vessel identifier was removed before variable selection) is equivalent to or superior to other models not including vessel identifier as a predictor. As such, we chose to use no vessel optimal RF models for presence–absence predictions for vessels lacking observer data.

Table 5. Predictive performance results of the full GAM for abundance when present of bycatch per floating object fishing set for each of the four study species for the French tropical tuna PS fleet in the Indian Ocean, 2011–2018 as a function of the distribution family assumed for model residuals.

Species	Family	RMSE	MAE	Dev. expl. (%)	v-RMSE	v-MAE	v-Dev. expl. (%)
Silky shark	Poisson	9.06	5.05	34.2	11.5	6.02	−12.4
Silky shark	Negative binomial	10.3	5.39	14.2	10.7	5.63	4.1
Silky shark	Tweedie	10.1	5.35	17.9	10.7	5.66	3.9
Silky shark	Log-normal	10.6	5.44	10.1	10.6	5.60	6.0
Rough triggerfish	Poisson	260	136	40.1	∞	∞	−∞
Rough triggerfish	Negative binomial	310	150	15.0	318	156	6.1
Rough triggerfish	Tweedie	307	149	16.4	317	157	6.3
Rough triggerfish	Log-normal	321	165	8.9	322	170	3.3
Rainbow runner	Poisson	199	55.8	52.1	∞	∞	−∞
Rainbow runner	Negative binomial	278	60.3	6.8	189	63.2	3.1
Rainbow runner	Tweedie	276	60.5	8.2	191	63.8	−0.6
Rainbow runner	Log-normal	285	63.4	1.7	189	65.5	4.3
Common dolphinfish	Poisson	28.6	17.5	44.4	∞	∞	−∞
Common dolphinfish	Negative binomial	34.2	19.4	20.6	35.3	20.2	13.9
Common dolphinfish	Tweedie	33.5	19.2	23.8	35.7	20.3	10.8
Common dolphinfish	Log-normal	35	20.2	17.0	36.3	21.0	7.8

RMSE, MAE, and deviance explained are the values computed with all data, whereas v-RMSE, v-MAE, and v-deviance explained were computed as the average of results from 10-fold cross-validation. Deviance explained was calculated using (1). Any values outside of the range of ±1000 are indicated as infinite.

Table 6. Performance of the full and optimal negative binomial GAMs for abundance when present of the four bycatch species on floating object fishing sets of the French tropical tuna PS fleet in the Indian Ocean.

Species	Model	RMSE	MAE	Dev. expl. (%)	v-RMSE	v-MAE	v-Dev. expl. (%)	AIC
Silky shark	Full abund. GAM	10.34	5.39	14.2	10.66	5.63	4.1	28070.3
Silky shark	Optimal abund. GAM	10.33	5.38	14.4	10.68	5.63	3.7	28076.0
Rough triggerfish	Full abund. GAM	309.81	149.54	15.0	317.57	156.43	6.1	55563.9
Rough triggerfish	Optimal abund. GAM	309.71	149.53	15.1	317.36	156.46	6.3	55563.2
Rainbow runner	Full abund. GAM	277.92	60.34	6.8	189.25	63.23	3.1	44151.6
Rainbow runner	Optimal abund. GAM	277.88	60.39	6.8	189.11	63.23	3.4	44148.7
Common dolphinfish	Full abund. GAM	34.21	19.38	20.6	35.29	20.22	13.9	35142.9
Common dolphinfish	Optimal abund. GAM	34.21	19.38	20.6	35.29	20.22	13.9	35142.9

Data were limited to sets with observers on board between 2011 and 2018. RMSE, MAE, and deviance explained are the values computed with all data. v-RMSE, v-MAE, and v-deviance explained were computed with 10-fold cross-validation. Deviance explained was calculated using (1).

Abundance when present models

Whereas the Poisson distribution produces "abundance when present" GAMs with the lowest RMSE and MAE and the highest deviance explained in training, the negative binomial, Tweedie, and log-normal distributions have much better performance statistics in cross-validation (Table 5). The Poisson distribution produces highly erratic predictions on novel data, leading to poor and extremely erratic performance statistics in cross-validation (e.g. note the infinite values in Table 5). Among the other three families, the deviance explained in training is highest for the Tweedie distribution, but in cross-validation, both the log-normal and the negative binomial explain more of the deviance, with the negative binomial having equal or lower cross-validation RMSE and MAE statistics for all but silky shark. Quantile-quantile diagnostic plots also indicated that the theoretical and observed residual quantiles adequately fit the expected diagonal line for the negative binomial, log-normal, and, to a lesser degree, Tweedie distributions, whereas large discrepancies with respect to the expected relationship were observed for the Poisson model (e.g. Supplementary Fig. S7). Graphs of residuals versus predicted values showed no strong trends for any of the distributions (e.g. Supplementary Fig. S8). Based on these results, we decided that the negative binomial distribution is on balance the best distribution for all further predictions of bycatch abundance when present (Fig. 2).

Variable selection only excluded relatively few predictors from negative binomial models of bycatch abundance when present, consistently retaining tuna catch, spatial predictors and their interactions, year, the year-month interaction, and vessel identifier, and differing only in the environmental variables retained. For example, the optimal abundance GAM for common dolphinfish is:

$$\begin{aligned}\text{No. of common dolphinfish} - 1 \\ \sim s(catch) + s(lon, by = quarter) + s(lat, by = quarter) \\ + ti(lon, lat, by = quarter) + s(month) + s(year) \\ + ti(year, month) + s(MLD) + s(SST) + vessel\end{aligned}$$

The optimal model for silky shark is identical except that it does not include the smooth on SST, whereas the models for rough triggerfish and rainbow runner include no environmental predictors. The similarity of full and optimal abundance GAMs is reflected in the close performance of the two (Table 6), though optimal models have equivalent or better performance indicators (e.g. equal or higher deviance explained)

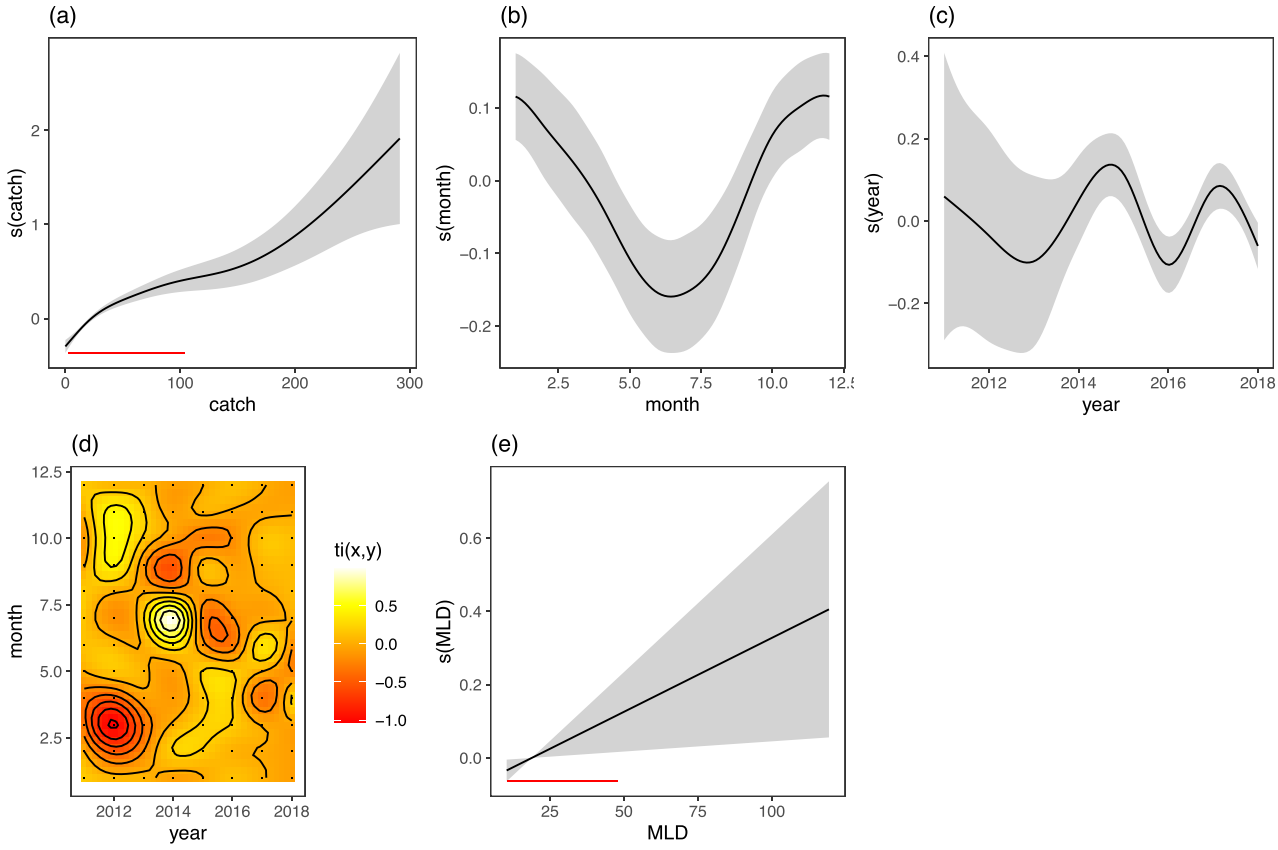

Figure 6. Marginal effects of non-spatial predictors with smooths on abundance when present of silky shark in floating objects fishing sets of the French PS fleet in the Indian Ocean. The effects were estimated with the optimal abundance GAM and are listed in the following order: (a) s(catch); (b) s(month); (c) s(year); (d) ti(year, month); and (e) s(MLD). See the caption of Fig. 4 for a description of the format with which information is presented in each panel.

under 10-fold cross-validation for all species except silky shark.

Deviances explained of the full and optimal negative models are also quite low, ranging from 3.1 to 13.9% in cross-validation, with values being particularly low for the forage species rainbow runner. The low deviance explained is reflected in error rates that are of approximately the same order of magnitude as the mean abundances when present (Table 1). For example, for silky shark, mean abundance when present is 8.4 individuals per set (Table 1), whereas the MAE of the optimal model in cross-validation is 5.63 (Table 6).

The marginal effect of smoothed predictors on bycatch abundance shows relatively few consistencies across species (Figs 6 and 7; Supplementary Figs S9–S14). Bycatch abundance does consistently increase with target tuna catch, though the strength of this increase varies considerably with species, being particularly notable for silky shark. Spatial patterns of bycatch abundance for the two top predators, silky shark (Fig. 7) and common dolphinship (Supplementary Fig. S14), generally were highest to the north of the study region, whereas those for forage species, rough triggerfish (Supplementary Fig. S12), and rainbow runner (Supplementary Fig. S13), were more variable and difficult to interpret. Year has an oscillating pattern for all species except rough triggerfish, which has a decreasing trend (Supplementary Fig. S12). Month has a peak positive impact on abundance of forage species rough triggerfish and rainbow runner in August–September (Supplementary Figs S12–S13), whereas the month effect for silky shark peaks in December–January (Fig. 7) and that for common dolphinfish is more complex, but also has a peak in the final quarter of the year (Supplementary Fig. S14). The effects of environmental predictors were generally either highly uncertain or showed no strong trend over the typical range of the predictor.

No vessel optimal models were identical in form (i.e. the terms included in Supplementary Table S4) to optimal models with vessel identifier removed as a predictor and had deviance explained in cross-validation equivalent to or superior to that of full models without vessel as a predictor (Supplementary Table S5). To be consistent with presence–absence model formulations, we decided to use no-vessel optimal models for making bycatch predictions for those vessels for which observer data is not available.

Δ models of bycatch abundance

Overall predictions of bycatch abundance were computed by multiplying the optimal RF model for presence–absence and the optimal negative binomial GAM for abundance when present for a given species. Deviance explained under cross-validation by these optimal Δ models ranges from 4.8 to 14.9%. Deviance explained is particularly low for rainbow runner, with similar amounts of deviance explained for rough triggerfish and silky shark and considerably higher deviance explained for common dolphinfish (Table 7). These low to moderately explained deviances are reflected in normalized performance indices, indicating that MAE is of the same order of magnitude as the mean catch per set, with values somewhat

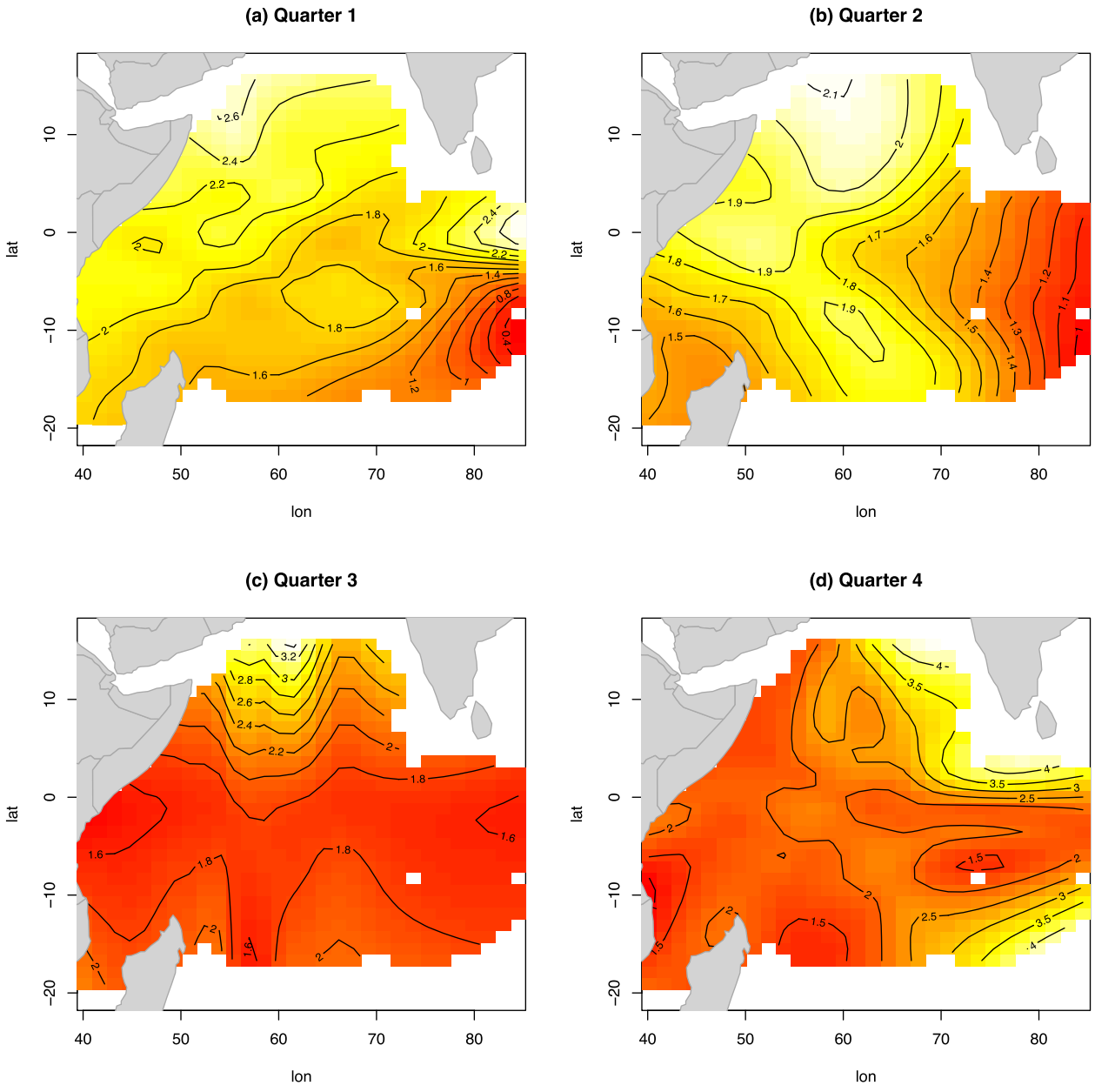

Figure 7. Marginal effects of spatial predictors by quarter on the abundance when present of silky shark in floating objects fishing sets of the French PS fleet in the Indian Ocean. The effects were estimated with the optimal GAM for abundance using the `mgcv::vis.gam` function, with all predictors other than lon, lat, quarter, and month fixed at their median values. Month was set to be the central month of the corresponding quarter (e.g. for quarter 1, month was set to 2). Lighter, yellow colors indicate a positive effect on abundance and numbers on contour lines indicate the marginal effect of spatial variables on the scale of the linear predictor.

higher for the two forage species, rough triggerfish and rainbow runner, than those for the two top predators, silky shark and common dolphinfish (Table 8). For example, the optimal Δ model for silky shark predicted a MAE (in cross-validation) of 5.28 sharks per fishing set, comparable to, but somewhat smaller than, the mean abundance of silky sharks per fishing set of 6.56 sharks (Table 1), the ratio of the two being 0.84. Corresponding values for rough triggerfish, e.g. are an MAE of 135.11 individuals per set and a mean abundance of 138.29 individuals per set (Table 1), producing a ratio of 1.03.

Patterns of differences between species in predictive power follow patterns of cross-correlations in bycatch abundance for the four species (Supplementary Fig. S19). The two teleost prey species, rough triggerfish and rainbow runner, have much stronger cross-correlations between each other, both for presence–absence and for total abundance than either of the two top predators, silky shark and common dolphinfish, do with each other or with the two prey species.

Prediction performance of Δ models at different scales

As expected, v-RMSE$_{norm}$ and v-MAE$_{norm}$ of predictions decrease with increasing spatial or temporal scales of aggregation (Fig. 8 and Supplementary Fig. S20). However, the decreases are more important in the spatial dimension than in the

Table 7. Performance of the Δ model for bycatch of the four study species per floating object fishing set for the French tropical tuna PS fishery in the Indian Ocean, 2011–2018.

Species	Model	RMSE	MAE	Dev. expl. (%)	v-RMSE	v-MAE	v-Dev. expl. (%)
Silky shark	Full model	9.21	4.49	22.3	9.78	5.28	7.9
Silky shark	Optimal model	9.19	4.49	22.6	9.78	5.28	7.8
Rough triggerfish	Full model	272.09	119.23	19.5	280.93	134.15	8.1
Rough triggerfish	Optimal model	272.50	120.45	19.3	281.86	135.11	7.4
Rainbow runner	Full model	236.05	45.96	8.1	162.35	53.02	3.9
Rainbow runner	Optimal model	236.03	45.93	8.1	162.01	52.93	4.8
Common dolphinfish	Full model	28.94	14.60	29.4	31.50	17.42	15.1
Common dolphinfish	Optimal model	28.90	14.55	29.6	31.53	17.36	14.9

Data were limited to fishing sets with observer's data. Presence–absence was modeled using RF models, whereas abundance when present was modeled using negative binomial GAMs. RMSE and MAE are the values computed with all data, whereas v-RMSE and v-MAE were computed with 10-fold cross-validation.

Table 8. Normalized performance measures for the optimal Δ models for the four study species.

Species	$RMSE_{norm}$	MAE_{norm}	v-$RMSE_{norm}$	v-MAE_{norm}
Silky shark	1.44	0.70	1.55	0.84
Rough triggerfish	2.06	0.91	2.13	1.03
Rainbow runner	4.91	0.95	3.43	1.12
Common dolphinfish	1.57	0.79	1.72	0.95

Presence–absence was modeled using RF models, whereas abundance when present was modeled using negative binomial GAMs. $RMSE_{norm}$ and MAE_{norm} are the coefficients of variation of MAE and RMSE, respectively, i.e. they have been normalized by the predicted mean number of individuals of the given species per set. v-statistics were computed with 10-fold cross-validation.

temporal dimension. For example, for silky shark, the RMSE relative to the overall mean decreases from 1.48 to 0.58 when going from aggregating over 1° × 1° cells by month to aggregating over the entire Indian Ocean by month, a decrease of 60.7%. In comparison, when going from aggregating over 1° × 1° cells by month to aggregating over the entire time series by 1° × 1° cells, the RMSE decreases by only 24.2%. This pattern of spatial aggregation producing a more important reduction in uncertainty as compared to temporal aggregation holds true for all species, though the absolute levels of uncertainty are again somewhat higher for the two prey species than they are for the two top predators (Supplementary Fig. S20). Overall, normalized RMSE at the annual-ocean level of aggregation is satisfactory, ranging from 0.18 to 0.42 for the four species (Supplementary Fig. S20).

Estimation of total bytcatch using Δ models

Spatial (Figs 9 and 10) and temporal (Fig. 11) predictions of total bycatch for the four study species were calculated using optimal RF-negative binomial Δ models (using no vessel-optimized models for sets carried out by vessels without any observer data over the study time period). Bycatch per set of the two top predators, silky shark and common dolphinfish, is highest in the northern part of fishing domain and, to a lesser degree, in the equatorial part of the fishing domain (Fig. 9). Spatial patterns of bycatch per set of the two schooling—prey species, rough triggerfish, and rainbow runner—are less similar and more variable, with both being more equatorial than the two predators, and both showing indications of a bimodal distribution with peaks north and south of the equator. The spatial coherence of bycatch of the two top predators is surprising given the low cross-correlation between bycatch of the two species (Supplementary Fig. S19), but this is consistent with the relatively minor fractions of total fishing effort in the northern part of the fishing domain (Floch et al. 2021), as evidenced by the high relative uncertainty in bycatch predictions for this area (Fig. 10).

Temporal trends in predicted total annual bycatch show increasing trends for silky shark and rainbow runner, a variable and potentially increasing trend for common dolphinfish, and a decreasing trend for rough triggerfish. Increasing trends in bycatch are consistent with the increase by the French fleet in the number and fraction of fishing sets carried out on FOBs since 2010 (Floch et al. 2021).

Comparaison with the bycatch-over-landings ratio approach

The bycatch-over-landings ratio approach currently used for reporting total bycatch to the IOTC and other RFMOs has consistently worse performance statistics (Table 9) than optimal Δ log-normal models (Table 7). In particular, the approach has negative explained deviances for all species, indicating it did not fit the data despite the increasing relationship between landings and bycatch observed for all four species (e.g. Fig. 6).

Despite this poor fit, trends in total annual bycatch predicted by the ratio approach are overall similar to those of optimal Δ models (Fig. 11), though there were some large discrepancies between the two approaches for rainbow runners.

Discussion

We developed a Δ modeling approach to estimate bycatch rates of species caught in tropical tuna PS fisheries from a combination of logbook and observer data that accounts for both high prevalence of zeros (absences) in the data and large fluctuations in abundance when present. Overall, our models for the French Indian Ocean PS fleet explain (in cross-validation) between 5 and 15% of the variability in the number of bycatch individuals per set for four common and emblematic PS bycatch species. Though much of the variance is not explained by our models, highlighting the highly stochastic nature of PS bycatch, models outperform the commonly used bycatch-over-landings ratio approach in terms of standard performance measures (e.g. RMSE, MAE, and deviance explained). Furthermore, predictive capacity improves considerably when bycatch is aggregated over spatial strata (e.g. to calculate total bycatch per year), and the method provides a robust mechanism for calculating both confidence and prediction intervals for bycatch estimates. Overall, these results

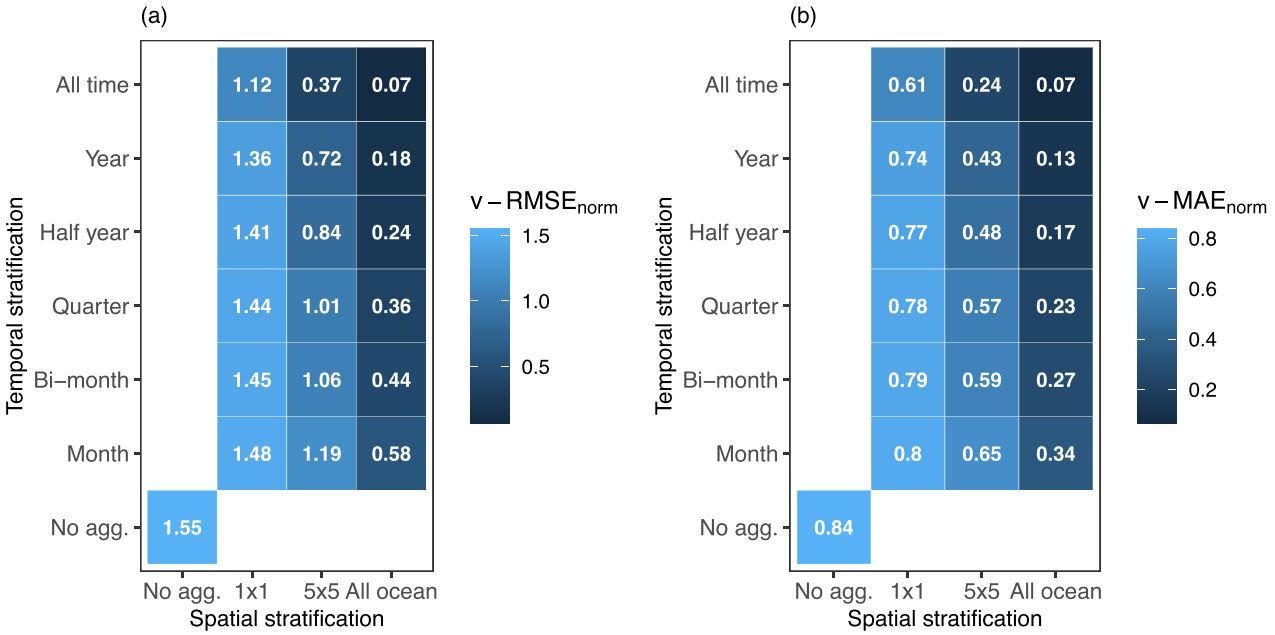

Figure 8. Prediction performance statistics of the optimal Δ model for silky sharks at different scales of aggregation. Presence–absence was modeled using RF models, whereas abundance when present was modeled using negative binomial GAMs. Panel (a) presents v-RMSE$_{norm}$, i.e. the mean over the 10-folds of the ratio of prediction RMSE to the mean prediction, whereas panel (b) presents v-MAE$_{norm}$, i.e. the mean over the 10-folds of the ratio of prediction MAE to the mean prediction. Note that results for "all ocean" and "all time" are comparing a single predicted value to a single observed total bycatch, so that a v-RMSE$_{norm}$ or v-MAE$_{norm}$ of 0.07, e.g. indicates that the observed bycatch differs from the predicted by ±7%.

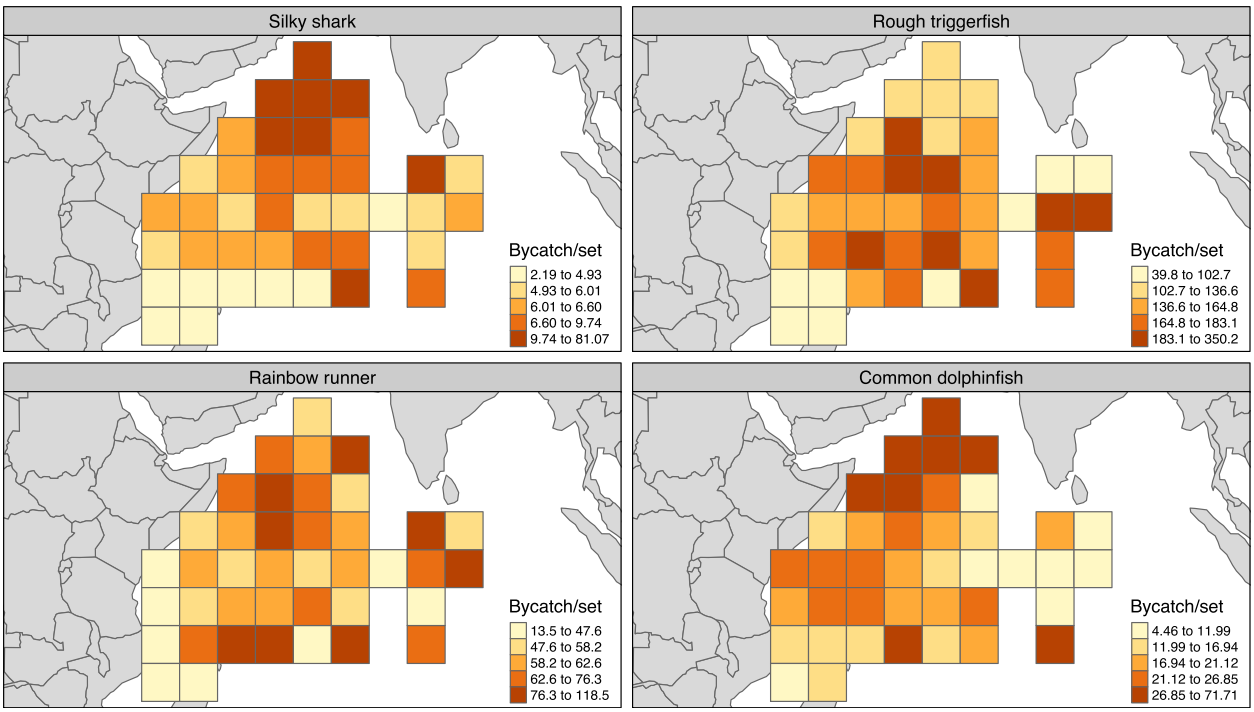

Figure 9. Spatial distribution of average bycatch per set of the four study species for the French PS fleet in the Indian Ocean between 2011 and 2018. Bycatch per set was calculated as the sum of observed bycatch for those fishing sets with observer data and model predictions of bycatch for those fishing sets without observer data, divided by the total number of sets for each 5° × 5° grid cell. Model predictions were calculated using the optimal Δ model (using "no vessel optimal models" for fishing activities by vessels without any observer data over the study time period). Shading indicates the number of individuals caught per set, with the shading scheme based on quantiles of the data for each of the four species and with darker shapes indicating more individuals per set. .

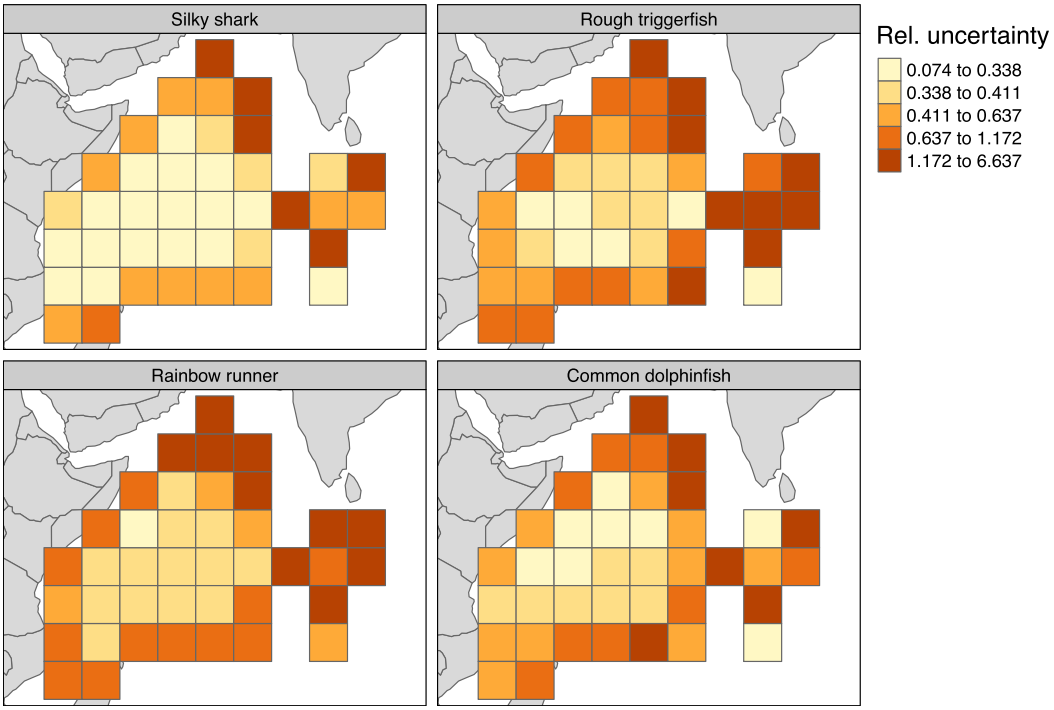

Figure 10. Spatial distribution of relative uncertainty in bycatch per set of the four study species for the French PS fleet in the Indian Ocean between 2011 and 2018. Bycatch per set was calculated as the sum of observed bycatch for those fishing sets with observer data and model predictions of bycatch for those fishing sets without observer data, divided by the total number of sets for each 5° × 5° grid cell. Model estimates were calculated using the optimal Δ model (using "no vessel optimal models" for fishing activities by vessels without any observer data over the study time period). Relative uncertainty was measured as the ratio between the 95% prediction interval for total bycatch and the total estimated bycatch in each grid cell. Shading indicates relative uncertainty with the shading scheme based on quantiles of the data across all four species (i.e. the same shading scheme is used for all species) and darker shades indicating greater uncertainty..

provide the best available estimates of bycatch rates that can be used for scientific and management assessments of the impacts of tropical tuna PS fishing on pelagic bycatch species.

Predictive capacity for top predators, silky sharks and common dolphinfish, is higher than that of schooling, forage species rough triggerfish and rainbow runner (Table 8). Spatial patterns of bycatch per set are also more consistent for the top predators than for the two forage species (Fig. 9). These differences in predictive capacity are due to a combination of reduced catch variability relative to the mean catch per set for the two top predators (the CV column of Table 1) and differences in the predictive capacity of GAMs for abundance when present (Table 6), with RF model performance of presence–absence being similar across species (Table 4). This pattern is somewhat surprising given the considerably higher prevalence of the forage species relative to the top predators (by a factor of ~2–15; Table 1), which would naively suggest these species are subject to lower stochasticity than the top predators. Nevertheless, it is consistent with the overall trophic level and ecological niche of silky sharks and common dolphinfish being closer to that of predatory target tuna species, leading to a stronger link between tunas and the bycatch of these species and thereby better fit of these species to the model. This hypothesis is supported by observations that tropical tunas and silky sharks have similar temporal patterns of presence at FOBs, but very different patterns to those of the two forage species (Forget et al. 2015). Nevertheless, positive relationships between tuna catch and bycatch abundance were noted for all four species (Fig. 6a, Supplementary Figs S9a, S10a, and S11a), and higher error in count data for highly prevalent bycatch species could also be a factor. Overall, results suggest that, while top predators may be somewhat easier to predict than forage species, management strategies designed to minimize bycatch of these species while maintaining catch of target tunas (e.g. Pons et al. 2022) are likely to be relatively ineffective due to the link between target species and non-target species and the overall high unexplained variance in models.

One simple, but important, observation of our model selection procedure is the importance of cross-validation for identifying and quantifying model overfitting. For example, cross-validation and model diagnostic plots were essential to identifying the weaknesses of Poisson distribution models of abundance when present for predicting bycatch, thereby leading us to choose to use negative binomial models. Though the Poisson model for silky shark abundance had the highest nominal performance, in cross-validation performance was very poor, indicative of overfitting and poor predictive capacity. Similarly, cross-validation was essential to estimating the true predictive capacity of RF models of bycatch presence–absence, which have a nominal well-classified rate at or close to 100%, but a cross-validation well-classified rate close to 75% (Table 4). Though cross-validation is of course commonplace in biological sciences, it is not systematically carried out, and its value for identifying overfitting is not always recognized (Yates et al. 2023).

It is difficult to interpret the variables retained in optimal models and, in particular, the retained environmental

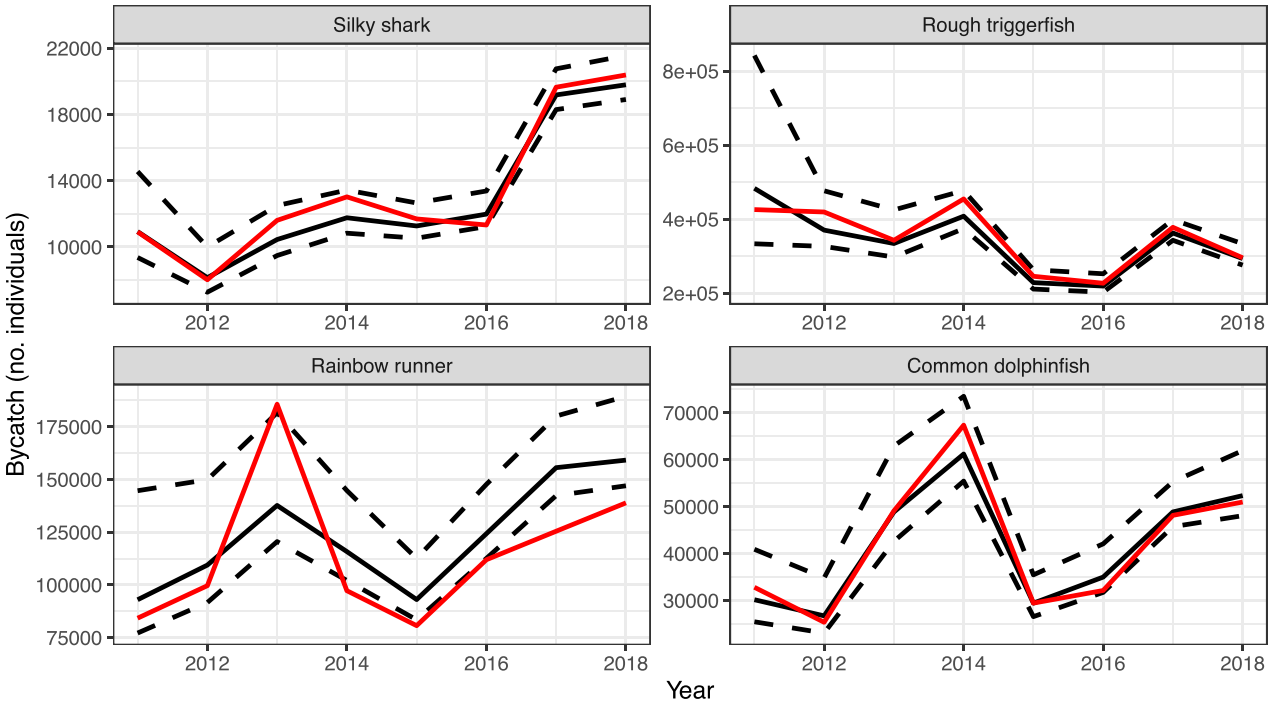

Figure 11. Estimated annual total bycatch of the four study species for the French tropical tuna PS fleet in the Indian Ocean between 2011 and 2018. Totals are the sum of the observed bycatch for sets with observer data and model predictions of bycatch for sets without observer data. Model predictions for the ligher, red, solid curve are derived from the ratio model currently used for IOTC reporting and catch assessments, whereas the darker, black, solid curve uses estimates from the optimal Δ model for each species (using "no vessel optimal models" for fishing activities by vessels without any observer data over the study time period). The two dashed curves indicate the lower and upper bounds of the 95% confidence interval for estimates from the optimal Δ model.

Table 9. Performance statistics for the bycatch-over-landings ratio approach currently used for reporting total bycatch predictions to the IOTC and other RFMOs.

Model	Species	RMSE	MAE	Dev. expl. (%)	v-RMSE	v-MAE	v-Dev. expl. (%)
Ratio	Silky shark	11.55	6.40	−22.3	11.49	6.46	−30.9
Ratio	Rough triggerfish	331.56	162.62	−19.5	329.45	164.47	−28.8
Ratio	Rainbow runner	247.65	62.84	−1.2	183.49	63.82	−45.0
Ratio	Common dolphinfish	38.10	21.32	−22.3	38.62	21.59	−28.9

predictors in terms of biological drivers of species abundance and catchability due to variability between species in retained variables and between presence–absence and abundance when present models, as well as due to the presence of spatial and temporal predictors that undoubtedly mask some of the environmental effects. This difficulty is also likely in part due to the inherent difficulty of directly associating abiotic and low-trophic-level biotic environmental predictors with behavior and abundance of higher-trophic-level pelagic species (e.g. Mondal et al. 2022), as well as the strong stochasticity of pelagic ecosystems (Paiva et al. 2013). Nevertheless, mixed layer depth (MLD) was included in all optimal GAMs of presence–absence and in optimal GAMs of abundance when present for the two top predator species, and a deeper MLD was consistently found to increase presence of bycatch species. A link between bycatch and MLD is logical given that MLD has long been assumed to be a driver of the vertical distribution of pelagic species (e.g. Kitagawa et al. 2000) and therefore their catchability in surface fisheries, though typically one would assume that a shallower MLD increases catchability of target species. We hypothesize that this seeming contradiction could be due to the fact that we separately control for catch of target tunas in our models, thereby removing the impact of overall catchability. Furthermore, we only use data in this paper from "successful" fishing operations during which tuna were caught, so the general effect of shallow MLD facilitating catch may not be reflected in our (positive) catch data.

The results of this work have numerous consequences for management of PS bycatch species. First and foremost, the Δ models developed here provide the most accurate available approach to estimating bycatch per set for the assessed species. Whereas the bycatch-over-landings ratio approach typically used for reporting bycatch levels to RFMOs consistently explained none of the variability around the mean in bycatch levels per set, our models explained small to modest amounts of the variability. Unexplained variance at the scale of an individual set remains non-negligible, but uncertainty in annual total bycatch, an important result for management organizations, was estimated with a RMSE of 0.18–0.42, which is acceptable for many management objectives. Above all, the approach provides a consistent and theoretically sound method of estimating prediction uncertainty, i.e. essential to identi-

fying significant trends in bycatch levels. Finally, though the objective of this work is to predict bycatch levels, the models developed here can be adapted (with some modification) to generate standardized abundance indices for evaluating the status of often data-poor bycatch species, such as silky sharks and other pelagic sharks (e.g. Kaplan and Tolotti 2023).

Our results also have important consequences for spatial management measures designed to reduce bycatch levels while minimizing impact of catch of target tunas. Previous work on spatial management of tropical tuna fisheries has found that static marine protected areas (MPAs) are likely an ineffective approach to management of Indian Ocean tropical tuna fisheries (Kaplan et al. 2014), whereas dynamic MPAs that target areas characterized by high bycatch and low catch of target species could be an effective approach to bycatch reduction for many large-scale industrial fisheries (Pons et al. 2022). Expected benefits of dynamic MPAs for tropical tuna PS fisheries varied, with non-negligible benefit estimated for the eastern Pacific Ocean, but little expected benefit for the Atlantic and Indian Oceans due to high covariability between catch and bycatch (Pons et al. 2022). Our results, indicating a positive relationship between target tuna catch and bycatch, support these conclusions. Furthermore, the value of dynamic MPAs is predicated on our capacity to predict future bycatch patterns at the level of individual space-time strata (e.g. $1^\circ \times 1^\circ$-month grid cells). The considerable unexplained variance in our models, even when aggregating in space and time, indicates that these types of models will likely be an ineffective basis for designing effective dynamic MPAs that depend on high accuracy in predictions to achieve significant reductions in bycatch.

Future avenues for improving predictions from statistical models of bycatch species include examining alternative model formulations and considering new predictor variables. Among the prior are using zero-inflated distributions and hierarchical models, both of which can in principle model presence–absence and abundance in a coherent framework. We initially modeled bycatch using zero-inflated GAMs, but we found that available zero-inflated distributions lacked sufficient flexibility to optimally capture our data and/or were not available for GAMs. In addition, splitting presence–absence and abundance into two models allowed us to understand each process individually, thereby providing additional insight. Though Bayesian hierarchical models are a powerful framework for modeling data driven by multiple processes, there are no fundamental reasons to believe that such a parametric approach would significantly improve predictive capacity over highly non-linear, "semi"-parametric approaches such as GAMs. Furthermore, our own work and numerous other studies have found that machine-learning approaches, such as RF models, are generally more effective at modeling multinomial data than parametric approaches (including both GAMs and Bayesian models). For these reasons, we consider that our Δ models using machine learning for modeling presence–absence and non-linear, parametric models for abundance represent the best available compromise among the constraints of predictive capacity, flexibility, and interpretability (e.g. as might be lacking in an approach based entirely on machine learning).

Among new predictor variables that could be used to model bycatch rates, we did not take into account information from previous nearby fishing operations that could be informative of local bycatch levels. This type of information has been used to develop "move-on rules" for benthic and demersal fisheries (Dunn et al. 2015), but this possibility has not to our knowledge been explored for tropical tuna PS fisheries. Though an interesting avenue to explore, our models have a relatively large unexplained variance despite including both spatial and temporal predictors, alone and in interaction, suggesting that including even more fine-scale spatial and temporal information may have difficulty overcoming the large stochasticity of pelagic ecosystems. Perhaps more promising, our models only consider environmental predictors at the time and location of individual fishing sets, whereas, given the time needed to transmit abiotic and low-tropic-level biotic signals to higher tropic levels, space-time lags between environmental predictors and bycatch levels are to be expected (Lehodey et al. 2010). One promising statistical approach to assessing this possibility is via deep learning methodologies capable of analyzing large amounts of data to identify cryptic connections that might otherwise be difficult to identify (e.g. Sokolova et al. 2021).

Despite these avenues for future developments, our modeling approach currently represents the best available approach to predicting total bycatch in tropical tuna fisheries with incomplete observer coverage. As such, we strongly encourage that this methodology be used for estimating bycatch levels for management and scientific purposes.

Acknowledgements

We thank ORTHONGEL for making industry-collected observer data via the OCUP observer program available for our analyses and the IRD-Ob7 pelagic observatory of the MARBEC laboratory for tropical tuna logbook and observer data management and preparation, particularly Laurent Floch. We thank the handling editor and three anonymous reviewers for their constructive comments that greatly improved the quality of the article.

Supplementary data

Supplementary material is available at the *ICES Journal of Marine Science* online version of the manuscript.

Conflict of interest: The authors declare no competing interests.

Funding

This work was funded by the Ob7 Exploited Tropical Pelagic Ecosystems Observatory of the IRD via support from the the European Data Collection Framework (DCF, Reg 199/2008 and 665/2008) and the French "Direction des Pêches Maritimes et de l'Aquaculture" (DPMA), as well as by the Horizon Europe project REDUCE (101135583). Data collection was financially supported by the EU via the Data Collection Framework (DCF, Reg 199/2008 and 665/2008), and by ORTHONGEL in the framework of their OCUP observer program.

Data availability and reproducibility

The fine-scale fishing activity, catch and bycatch data used in this article are confidential in nature and cannot be publicly shared. As such, requests for data access should be addressed directly to the IRD's Exploited Tropical Pelagic Ecosystems

Observatory (Ob7; https://www.ob7.ird.fr/en/) using the following e-mail address: adm-dblp@ird.fr.

Code for carrying out all analyses is available in open access at the following address: https://github.com/Agathedumont/modeling-bycatch-abundance-delta-method.

References

Amandè MJ, Ariz J, Chassot E *et al*. Bycatch of the European purse seine tuna fishery in the Atlantic Ocean for the 2003–2007 period. *Aquat Living Resour* 2010;**23**:353–62.

Andersen MM. Prediction intervals for Generalized Additive Models (GAMs). 2022. https://mikl.dk/post/2019-prediction-intervals-for-gam/ (23 March 2024, date last accessed).

Arrizabalaga H, Dufour F, Kell L *et al*. Global habitat preferences of commercially valuable tuna. *Deep Sea Res II Topic Stud Oceanogr* 2015;**113**:102–12.

Bell JD, Sharp MK, Havice E *et al*. Realising the food security benefits of canned fish for Pacific Island countries. *Mar Policy* 2019;**100**:183–91.

Berrar D. Cross-validation. In: *Reference Module in Life Sciences*. Amsterdam: Elsevier, 2018.

Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.

Davies T, Mees C, Milner-Gulland E. The past, present and future use of drifting fish aggregating devices (FADs) in the Indian Ocean. *Mar Policy* 2014;**45**:163–70.

Dulvy NK, Fowler SL, Musick JA *et al*. Extinction risk and conservation of the world's sharks and rays. *eLife* 2014;**3**:e00590. https://doi.org/10.7554/eLife.00590

Dunn DC, Maxwell SM, Boustany AM *et al*. Dynamic ocean management increases the efficiency and efficacy of fisheries management. *Proc Natl Acad Sci* 2015;**113**:668–73. https://doi.org/doi/10.1073/pnas.1513626113

FAO. Report of the expert meeting to develop technical guidelines to reduce bycatch of marine mammals in capture fisheries. Rome: FAO, 2020. http://www.fao.org/documents/card/en/c/ca7620en (6 September 2022, date last accessed).

Fletcher D. Confidence intervals for the mean of the delta-lognormal distribution. *Environ Ecol Stat* 2008;**15**:175–89.

Floch L, Marsac F, Fily T *et al*. Statistics of the French purse seine fishing fleet targeting tropical tuna in the Indian Ocean (1981–2020). IOTC-2021-WPDCS17-21. Cape Town: Indian Ocean Tuna Commission Working Party on Data Collection (WPDCS), 2021. https://www.iotc.org/WPDCS/17/21 (22 November 2021, date last accessed).

Forget FG, Capello M, Filmalter JD *et al*. Behaviour and vulnerability of target and non-target species at drifting fish aggregating devices (FADs) in the tropical tuna purse seine fishery determined by acoustic telemetry. *Can J Fish Aquat Sci* 2015;**72**:1398–405. https://doi.org/10.1139/cjfas-2014-0458

Genuer R, Poggi JM, Tuleau-Malot C. VSURF: an R package for variable selection using random forests. *R J* 2015;**7**:19.

Genuer R, Poggi JM, Tuleau-Malot C. VSURF: variable selection using random forests. 2022. https://github.com/robingenuer/VSURF(27March 2024, date last accessed)

Goujon M, Maufroy A, Ra RS *et al*. Collecting data on board French tropical tuna purse seiners with common observers: results of Orthongel's voluntary observer program OCUP in the Atlantic Ocean (2013−2017). *Collect Vol Sci Pap ICCAT* 2017;**74**:3784–805.

Gräler B, Pebesma E, Heuvelink G. Spatio-temporal interpolation using gstat. *R J* 2016;**8**:204–18.

Gregorutti B, Michel B, Saint-Pierre P. Correlation and variable importance in random forests. *Stat Comput* 2017;**27**:659–78. https://doi.org/10.1007/s11222-016-9646-1

Hutchinson MR, Itano DG, Muir JA *et al*. Post-release survival of juvenile silky sharks captured in a tropical tuna purse seine fishery. *Mar Ecol Prog Ser* 2015;**521**:143–54. https://doi.org/10.3354/meps11073

Kaplan DM, Chassot E, Amandé JM *et al*. Spatial management of Indian Ocean tropical tuna fisheries: potential and perspectives. *ICES J Mar Sci J du Conseil* 2014; **71**: 1728–49.

Kaplan DM, Tolotti MT. Silky shark abundance index based on CPUE standardization of French Indian Ocean tropical tuna purse seine observer bycatch data. IOTC-2023-WPEB19-34_Rev1. Réunion: IOTC 19th Working Party on Ecosystems and Bycatch, 2023.

Kitagawa T, Nakata H, Kimura S *et al*. Effect of ambient temperature on the vertical distribution and movement of Pacific bluefin tuna *Thunnus thynnus* orientalis. *Mar Ecol Prog Ser* 2000;**206**:251–60.

Legendre P, Legendre L. *Numerical Ecology*. Amsterdam: Elsevier, 2012.

Lehodey P, Murtugudde R, Senina I. Bridging the gap from ocean models to population dynamics of large marine predators: a model of mid-trophic functional groups. *Prog Oceanogr* 2010;**84**:69–84.

Liaw A, Wiener M. Classification and regression by randomforest. *R News* 2002;**2**:18–22.

Mannocci L, Forget F, Tolotti MT *et al*. Predicting bycatch hotspots in tropical tuna purse seine fisheries at the basin scale. *Glob Ecol Conserv* 2020;**24**:e01393.

Maufroy A, Kaplan D, Bez N *et al*. Massive increase in the use of drifting fish aggregating devices (dFADs) by tropical tuna purse seine fisheries in the Atlantic and Indian Oceans. *ICES J Mar Sci J du Conseil* 2016;**74**:fsw175. https://doi.org/10.1093/icesjms/fsw175

Mondal S, Lan YC, Lee MA *et al*. Detecting the feeding habitat zone of albacore tuna (*Thunnus alalunga*) in the southern Indian Ocean using multisatellite remote sensing data. *J Mar Sci Technol* 2022;**29**:794–807.

Murtaugh PA. In defense of *P* values. *Ecology* 2014;**95**:611–17.

Onandia I, Grande M, Galaz JM *et al*. New assessment on accidentally captured silky shark post-release survival in the Indian Ocean tuna purse seine fishery. IOTC-2021-WPEB17 (DP)-13. Cape Town: IOTC Working Party on Ecosystems & Biodiversity, 2021.

Ortiz M, Arocha F. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. *Fish Res* 2004;**70**:275–97. https://doi.org/10.1016/j.fishres.2004.08.028

Pacoureau N, Rigby CL, Kyne PM *et al*. Half a century of global decline in oceanic sharks and rays. *Nature* 2021;**589**:567–71.

Paiva VH, Geraldes P, Ramirez I *et al*. Overcoming difficult times: the behavioural resilience of a marine predator when facing environmental stochasticity. *Mar Ecol Prog Ser* 2013;**486**:277–88.

Pianet R, Pallares P, Petit C. New sampling and data processing strategy for estimating the composition of catches by species and sizes in the european purse seine tropical tuna fisheries. 2000. https://aquadocs.org/handle/1834/971 (10 February 2023, date last accessed).

Pons M, Kaplan DM, Moreno G *et al*. Benefits, concerns, and solutions of fishing for tunas with drifting fish aggregation devices. *Fish Fish* 2023;**24**:979–1002.

Pons M, Watson JT, Ovando D *et al*. Trade-offs between bycatch and target catches in static versus dynamic fishery closures. *Proc Natl Acad Sci* 2022;**119**:e2114508119. https://doi.org/10.1073/pnas.2114508119

R Core. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2023. https://www.R-project.org/

Ribeiro PJ, Diggle P, Christensen O *et al*. geoR: analysis of geostatistical data. 2023. http://www.leg.ufpr.br/geoR/(27 March 2024, date last accessed)

Ruiz J, Abascal FJ, Bach P, *et al*. Bycatch of the European, and associated flag, purse-seine tuna fishery in the Indian Ocean for the period

2008-2017. IOTC-2018-WPEB14-15. Cape Town: IOTC Working Party on Ecosystems and Bycatch (WPEB), 2018.

Shono H. Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fish Res* 2008;**93**:154–62.

Sokolova M, Mompó Alepuz A, Thompson F *et al.* A deep learning approach to assist sustainability of demersal trawling operations. *Sustainability* 2021;**13**:12362.

Stock BC, Ward EJ, Eguchi T *et al.* Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. *Can J Fish Aquat Sci* 2020;**77**:146–63.

Valliant R, Dorfman AH, Royall RM. Finite population sampling and inference: a prediction approach. New York, NY: John Wiley & Sons, Inc., 2000.

Wain G, Guéry L, Kaplan DM *et al.* Quantifying the increase in fishing efficiency due to the use of drifting FADs equipped with echosounders in tropical tuna purse seine fisheries. *ICES J Mar Sci* 2021;**78**:235–45. https://doi.org/10.1093/icesjms/fsaa216

Wood SN. Generalized Additive Models: An Introduction with R. Boca Raton, FL: CRC Press/Taylor & Francis Group, 2017.

Yates LA, Aandahl Z, Richards SA *et al.* Cross validation for model selection: a review with examples from ecology. *Ecol Monogr* 2023;**93**:e1557. https://doi.org/10.1002/ecm.1557

Handling Editor: Ruben Roa-Ureta