Author(s):	Nerea Goikoetxea, Izaro Goienetxea, Jose A. Fernandes-Salvador, Nicolas Goñi, Igor Granado, Iñaki Quincoces, Leire Ibaibarriaga, Jon Ruiz, Hilario Murua, Ainhoa Caballero
Title:	Machine-learning aiding sustainable Indian Ocean tuna purse seine fishery
Year:	2024
Version:	Published version
Copyright:	The Author(s) 2024
Rights:	CC BY-NC-ND 4.0
Rights url:	https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the original version:

Nerea Goikoetxea, Izaro Goienetxea, Jose A. Fernandes-Salvador, Nicolas Goñi, Igor Granado, Iñaki Quincoces, Leire Ibaibarriaga, Jon Ruiz, Hilario Murua, Ainhoa Caballero, Machine-learning aiding sustainable Indian Ocean tuna purse seine fishery, Ecological Informatics, Volume 81, 2024, 102577, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2024.102577.

Ecological Informatics 81 (2024) 102577

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Machine-learning aiding sustainable Indian Ocean tuna purse seine fishery

Nerea Goikoetxea [a,*], Izaro Goienetxea [b], Jose A. Fernandes-Salvador [a], Nicolas Goñi [c], Igor Granado [b], Iñaki Quincoces [a], Leire Ibaibarriaga [a], Jon Ruiz [a], Hilario Murua [d], Ainhoa Caballero [b]

[a] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi Ugartea z/g, 48395 Sukarrieta, Spain
[b] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Herrera kaia, Portualdea z/g, 20110 Pasaia, Spain
[c] LUKE, Natural Resources Institute Finland, Itäinen Pitkäkatu 4A, FI-20520 Turku, Finland
[d] International Sustainable Seafood Foundation, Pittsburgh, PA 15201, United States

ARTICLE INFO

Keywords:
Sustainable fishing
Bycatch
Machine-learning
Tropical tuna
Fisheries oceanography
Species distribution models

ABSTRACT

Among the various challenges facing tropical tuna purse seine fleet are the need to reduce fuel consumption and carbon footprint, as well as minimising bycatch of vulnerable species. Tools designed for forecasting optimum tuna fishing grounds can contribute to adapting to changes in fish distribution due to climate change, by identifying the location of new suitable fishing grounds, and thus reducing the search time. While information about the high probability to find vulnerable species could result in a bycatch reduction. The present study aims at contributing to a more sustainable and cleaner fishing, i.e. catching the same amount of target tuna with less fuel consumption/emissions and lower bycatch. To achieve this, tropical tuna catches as target species, and silky shark accidental catches as bycatch species have been modelled by machine learning models in the Indian Ocean using as inputs historical catch data of these fleets and environmental data. The resulting models show an accuracy of 0.718 and 0.728 for the SKJ and YFT, being the absences (TPR = 0.996 for SKJ and 0.993 for YFT, respectively) better predicted than the high or low catches. In the case of the BET, which is not the main target species of this fleet, the accuracy is lower than that of the previous species. Regarding the silky shark, the presence/absence model provides an accuracy of 0.842. Even though the model's performance has room for improvement, the present work lays the foundations of a process for forecasting fishing grounds avoiding vulnerable species, by only using as input data forecast environmental data provided in near real time by earth observation programs. In the future these models can be improved as more input data and knowledge about the main environmental conditions influencing these species becomes available.

1. Introduction

The Food and Agriculture Organization (FAO) statistics show that marine fish catches have remained stable in the last 20 years (FAO, 2022). However, fuel consumption has increased 20%, resulting in the consequent greenhouse gas emissions (Bell et al., 2017; Parker et al., 2018). For instance, although the European fishing fleet recorded a reduction in fuel consumption between 2009 and 2019 (STECF, 2021), the increase in fuel costs over recent years is one of the main challenges to the sector, since fuel consumption represents 60–70% of the total annual cost of vessel activity (Rojon and Smith, 2014; Suuronen et al., 2012). Furthermore, fuel price increases are expected to affect every industry, including marine industries (Chrysafis et al., 2022; Roll et al., 2022). Among the fisheries, those fleets targeting highly migratory large pelagic species have one of the highest and most variable fuel consumption (Parker and Tyedmers, 2014). Since a total of 90% of the fuel consumption is dedicated to searching for tuna schools by purse seiners and reaching the fishing grounds (Basurko et al., 2022), reducing the search effort can contribute to save fuel by these fleets (Granado et al., 2021). Furthermore, there is evidence that tropical tuna habitat distribution has changed globally due to ocean warming (Erauskin-Extramiana et al., 2020, 2023) and that the shift will continue in the future (Nataniel et al., 2021). Purse seine skippers make day-to-day decisions on where to go fishing and thus they need to take adaption actions when looking for fishing grounds (Rubio et al., 2022). For instance, higher digitalization could aim at reducing fuel consumption and the time at

* Corresponding author.
E-mail address: ngoikoetxea@azti.es (N. Goikoetxea).

https://doi.org/10.1016/j.ecoinf.2024.102577
Received 25 January 2023; Received in revised form 20 February 2024; Accepted 24 March 2024
Available online 26 March 2024

sea of purse seiners as an adaptation of the fishing industry to climate change (Erauskin-Extramiana et al., 2023). In addition, early information on the probable distribution of target species could contribute to a more efficient localisation of fish schools. Indeed, Sun et al. (2022) highlighted the value of near-term ecological forecasting as a rapid and science-based decision-making tool.

Tropical tuna purse seiners use two main fishing techniques: fishing on free-swimming tuna schools, and fishing on schools gathering under Fish Aggregating Devices (FADs). In this regard, Scott and Lopez (2014) estimated that 65% of all the purse seine sets are made on floating objects, either anchored or drifting devices. The target species of the tropical tuna purse seine fleet are mainly skipjack tuna (SKJ, *Katsuwonus pelamis*), yellowfin tuna (YFT, *Thunnus albacares*) and to a lesser extent, bigeye tuna (BET, *Thunnus obesus*). For instance, in 2019, SKJ accounted for 60% of world tuna catches, and YFT for 27.5% (ISSF, 2020). These species are widespread in all tropical waters and are highly migratory. They also have different migratory patterns in different oceans (Arregui et al., 2019; Hallier and Fonteneau, 2015): they appear to be less mobile in the Atlantic Ocean, with a higher proportion of regional-scale migration, whereas in the Indian Ocean long-range displacements are a more common feature (Fonteneau and Hallier, 2015).

As in other fisheries, tropical tuna purse seine fishery produces incidental fishing of non-targeted species (bycatch), especially when fishing on FADs (Amandè et al., 2017; Hall and Roman, 2013). Bycatches can be categorized into three groups: minor tuna and tuna-like species such as neritic tunas, other teleost fishes, and sensitive species such as sharks and rays. Among these sensitive bycatch species, the most common species associated with purse seine fishery is the silky shark (FAL, *Carcharhinus falciformis*) (Amandè et al., 2010; Clavareau et al., 2020; Dagorn et al., 2012; López et al., 2020), which is listed as a vulnerable species by the International Union for Conservation of Nature (IUCN, 2021).

To move towards more sustainable fisheries, fishing fleets are increasingly relying on the use of technological developments. This innovation was coined as 'Smart Fishery' (Honarmand Ebrahimi et al., 2021). For instance, the fishing industry is augmenting its use of Earth Observation data to characterize environmental conditions of marine areas and geolocate fishing grounds with less effort based on their high level of digitalization (McCauley et al., 2016). Hence, this can reduce search times, fuel consumption and the operating cost of fishing vessels. However, owing to the large volume and diversity of sources and the quality of recorded data, they are underused for analysis, remain intact and unstructured, and require lots of resources for real-time analysis. Big data processing techniques enhanced by machine-learning (ML) methods can increase the value of such unexploited data and, as a result, enhance their applicability. ML has already started to prove its potential in marine sciences applied to fisheries (Fernandes et al., 2010; Groba et al., 2015, 2018), to feed Fisheries Route Optimization Decision Support Systems (Granado et al., 2021; Granado et al., 2024), forecast fishing grounds (Amandè et al., 2017) and plan which tuna species to fish according to allocated quotas (ISSF, 2020). Despite these developments, the use of artificial intelligence in the fishing industry lags behind other shipping sectors, both in state-of-the-art and day-to-day applications (Agra et al., 2015; Christiansen et al., 2004; Fernandes-Salvador et al., 2022).

The objective of the present study is to predict fishing grounds to catch tropical tuna species (SKJ, YFT and BET), while simultaneously reducing the bycatch of vulnerable species (FAL) in the Indian Ocean, for thus help improve the fishing efficiency. This contribution for improving the fishing efficiency is to be achieved by applying ML to combined catches and environmental datasets. Here, efficiency is understood as fishing the same amount of target species while reducing the bycatch of silky shark and consuming less fuel by reducing the time spent searching for tuna schools.

2. Material and methods

In this study a ML-based approach has been applied to forecast the probability of high catch zones of different species of tropical tuna, by training models with environmental data and historical catch and bycatch data.

2.1. Catch and bycatch data

The target catch and bycatch data selected for the present study were collected by observers on-board the Spanish tropical tuna purse seine fleet operating in the Indian Ocean during the period 2014–2020. The sampling is part of the Spanish Fisheries Data Collection Program conducted under the EU fisheries Data Collection Framework (Regulation EU 2017/1004). In this work, only catches on FADs were studied since they account for 95% of total catches. Furthermore, fishing on FADs showed a higher presence of bycatch (e.g. silky shark) when targeting tropical tuna in the Indian Ocean (Amandè et al., 2010; Dagorn et al., 2012; López et al., 2020).

In order to characterize a fishing set, observations can be distinguished between positive (the catch is hauled on-board) and null sets (the net is deployed but no catch is hauled on-board due to different reasons, such as high shear currents, broken gear, insufficient catchability when tunas are located too deep or moving too fast). Hence, potential catch of tuna or bycatch information in weight was only available when there was a positive fishing set, while null sets have been discarded. In addition to fishing sets, the data collected by on-board observers also included information on other activities carried out by the vessels, such as searches or operations with a floating object (any natural or man-made object that can be found in the sea). These additional data have been used in this study to determine absences of tuna schools, that is, positions where the vessel sailed but the net was not deployed.

2.2. Environmental, geographical, and temporal data

Potential predictors of the distribution of the assessed species used in other mechanistic models (Fernandes et al., 2013a; Nielsen et al., 2018) were considered in the model building analysis of this study. The list of these potential predictors that include environmental, geographical, and temporal information (Table 1) was extracted for each catch position and date.

With regard to the environmental variables daily physical and biogeochemical environmental data were obtained from the European Union's Earth Observation programme Copernicus. While, sea surface temperature (SST) data was obtained from NASA, since this database provides a higher spatial resolution for SST. In the case of the biogeochemical data, some of the variables were integrated from the sea surface to different depths (e.g. 10, 20, 50… m). The reason for this approach was the vertical mobility of the species considered (Sabarros et al., 2015), which implies considering the productivity of the entire water column they mainly inhabit.

The relation between the presence of pelagic species such as tropical tunas and oceanographic processes like temperature and chlorophyll fronts have been widely studied (Fiedler and Bernard, 1987; Zainuddin et al., 2017). Primary and secondary production seems to increase in such fronts, causing the aggregation of pelagic species. Such water mass interfaces of different densities (Sund et al., 1981) are characterized by abrupt changes in temperature and/or chlorophyll concentration at minimum horizontal distances (Rivas and Pisoni, 2010). Furthermore, López et al. (2020) also highlighted the importance of including temperature and chlorophyll fronts in future studies given the presence of certain sharks in areas with frequent front occurrences. Oceanic fronts for SST and chlorophyll concentration were estimated using the Belkin and O'Reilly (2009) algorithm implemented in the grec library (Lau-Medrano, 2020) in R (R Core Team, 2021).

Table 1
Summary of the variables considered in this study.

Variable	Depth (m)	Integration depth (m)
Sea water salinity	0, 10, 20, 50, 100, 125, 150, 175, 200	
Silicate concentration	0, 10, 20, 50, 100, 125, 150, 175, 200	
Phosphate concentration	0, 10, 20, 50, 100, 125, 150, 175, 200	
Nitrate concentration	0, 10, 20, 50, 100, 125, 150, 175, 200	
Current speed*	0, 10, 20, 50, 100, 125, 150, 175, 200	
Temperature	0, 10, 20, 50, 100, 125, 150, 175, 200	
Sea level anomaly		
Mixed layer depth		
Bottom temperature		
SST		
Net primary production		0, 10, 20, 50, 100, 125, 150, 175, 200
Dissolved molecular oxygen concentration		0, 10, 20, 50, 100, 125, 150, 175, 200
Chlorophyll concentration		0, 10, 20, 50, 100, 125, 150, 175, 200
Chlorophyll fronts		
SST fronts		
Thermocline intensity		
Thermocline depth		
Temperature gradient	10, 20, 50, 100, 125, 150, 175, 200	
Bathymetry		
Latitude		
Longitude		
Month		

In addition, thermocline depth and intensity were estimated from temperature values at different depths along the water column by approaching the gradient with the ratio of their finite differences. The thermocline depth has been identified as the depth at which the maximum temperature gradient is found, and it is considered herein as a predictor of vertical limitations in the distribution of tunas that show an overall preference for shallower waters (Sabarros et al., 2015) due to their physiological needs (Reilly and Fiedler, 1994). For example, the distribution of SKJ is restricted to the water column between the ocean surface and the thermocline (warm near cooler water masses with high oxygen concentration) as a result of its limitation in thermoregulation (Druon et al., 2017). Moreover, for silky sharks, SST and bathymetry were the main predictors for describing their distribution (Lezama-Ochoa et al., 2016) and, although they make diel vertical movements, they spend most of their time in waters below the thermocline (Curnick et al., 2020). In addition to this vertical gradient in temperature (i.e. the thermocline), consideration was also given to thermal gradients between the sea surface and different depths. Finally, the mixed layer depth, as the depth until which there is a near-homogeneity in the properties of the seawater, was also considered.

Oceanic eddies are rotating water masses that influence the surrounding ecosystems by regulating the horizontal and vertical dynamics of the water column (Bakun, 2006). As such, it has been observed that the increasing opportunities for foraging driven by certain eddies lead to an aggregation of pelagic predators (Arostegui et al., 2022). Therefore, this study considers sea level anomaly as a proxy of oceanic eddies to assess the potential influence of said structures in the distribution of these predators.

Bathymetry at each catch observation was also considered. The estimation of depths was based on the ETOPO1 global relief model (Amante and Eakins, 2009) developed by the NOAA National Centers for Environmental Information (NCEI). The mean and standard deviation of the depths and the minimum and maximum depths in each half-grade rectangle in the highest-resolution ETOPO raster were estimated.

Finally, latitude and longitude, as well as the month of the year, were also included in the models to take into account the geographical and temporal influence on the catches.

2.3. Machine-learning pipeline

A pipeline has been created to build supervised classification models for each of the species considered in this work (i.e. SKJ, YFT, BET and FAL). A diagram of the building process of the models is shown in Fig. 1. Different languages (Python, R) and the graphical interface of Weka have been used for different parts of the process. The resulting independent models will be able to predict probabilities of encountering each species, based on the environmental conditions of the study area.

The first step of the process is to merge the catch data with the environmental data described in Section 2.2. To do so, the values of each environmental variable in the positions of the entries of the catch data have been extracted. Since the resolution of the catch data is higher than the resolution of the environmental data, for each position of the catch dataset, the nearest environmental data available is considered. As stated before, in addition to the environmental data, the latitude, longitude and the month of each entry in the catch data have also been considered as explanatory variables in order to add spatial and temporal information. The Python package xcube (xcube.readthedocs.io/) has been used to extract the environmental data corresponding to each point in the catch data.

Then, the catch data of each tuna species has been discretized, to differentiate between high and low captures. To do so, the quantiles of captured weights have been computed, and the median has been selected to set the threshold. Thus, taking the median as a threshold, both capture types (high and low) will have a similar number of individuals. Since a high number of absences of tuna are also available in the catch data, these absences have also been included in the data that is used to train the models. In the case of the SKJ and YFT, the number of absences was significantly higher than that of the high/low classes, therefore, a random subset of the absences has been selected, considering the number of elements of the high/low classes of each species. For both species, a model that is able to classify between high captures, low captures and absences is built. For BET, its importance is secondary compared to that of the SKJ and YFT which are the main target species. Consequently, BET absences have not been considered, and a model that only classifies high and low captures of BET is built. In the case of FAL, as it is a bycatch species that we want to avoid fishing, an absence/presence type model is built. To do so, all the FAL captures have been labelled as presences, and all the entries in the catch data where captures of tuna have happened with no FAL capture, have been labelled as absence. The number of instances of each class and species is shown in Table 2.

The second step of the process is the feature selection, in which an analysis of the environmental data is performed to select the variables that provide more information about the type of catch (high, low or absence). To select the best environmental variables, two feature selection methods have been combined; the Correlation-based Feature Subset selection algorithm (CFS; Hall, 2000) and the Symmetric Uncertainty Score (SUS; Duda et al., 2001). SUS is a nonlinear correlation metric based on the information theory that measures the dependence between a predictor and the class label. CFS seeks to identify a set of predictors that are highly correlated with the type of capture but show low correlation among them. On the first step of the feature selection process, CFS is applied to each species' dataset, with a 10-fold cross-validation (Rodriguez et al., 2009), assuming that the number of folds in which a variable is selected (F) is an indicator of its robustness for describing the type of capture. Afterwards, all the variables that have been selected in more than five folds on the first step have been studied to find correlations between them. When a correlation score of more than 0.7 is found between two variables that have been selected in the same number of folds, only the one with the highest SUS value has been

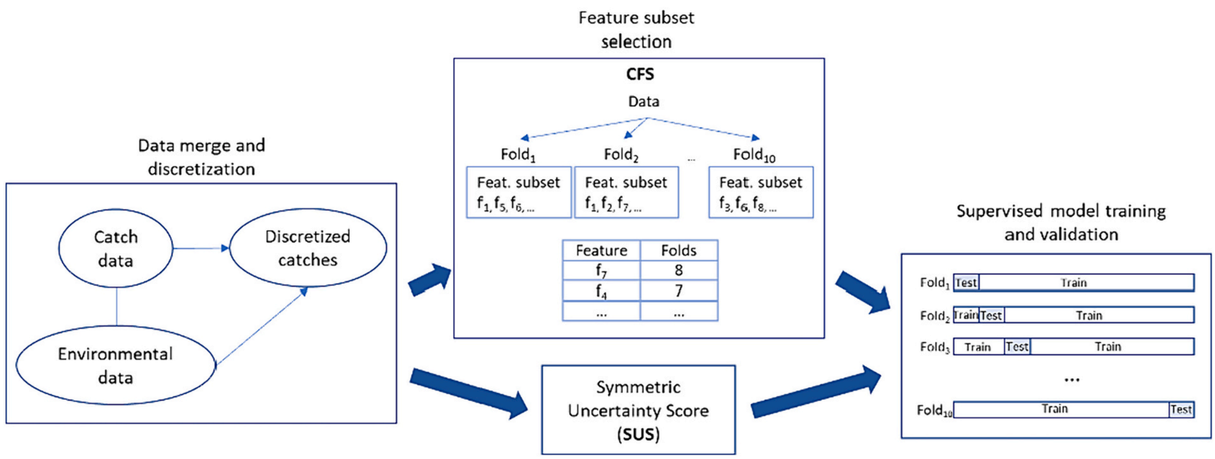

Fig. 1. Diagram of the model building process.

Table 2
Number of instances of each species and class.

	SKJ	YFT	BET	FAL
HIGH	6307	5234	3395	8929
LOW	6533	5831	3447	
ABS	6300	5200		6704

kept. To apply both the CFS cross-validation and SUS, their implementation in Weka (Frank et al., 2016) has been used.

In the third step, a supervised classification model is trained with the selected variables. Four different methods, widely used in species distribution modelling, have been tried to build the classification model:

- Random Forest (Breiman, 2001): Classifier that combines the output of multiple decision trees to reach a single result.
- SMO (SVM clone) (Platt, 1998): Kernel method that creates a hyperplane where the categories are divided by a clear gap that is as wide as possible.
- Multilayer Perceptron (Haykin, 1994): A classifier that uses backpropagation to learn a multi-layer perceptron to classify instances.
- Naïve Bayes (John and Langley, 1995): Statistical learning algorithm based on Bayesian rules. Given that the value of the class is known, it assumes independence between the occurrence of feature values to predict the class.

The implementation of these methods in Weka has been used to measure their **performances** in order to choose the best one for each species, using their standard configuration with a 5 × 10-fold cross-validation. The accuracies of each method (number of correct predictions divided by the total number of predictions made by each model) have been computed and compared to select the one that is the most suitable for the classification of each species' dataset. A final model has been trained for each species, with the Caret package of R (Kuhn, 2008).

The forecast of the final distribution models for the study area can be represented in maps (as in Fig. 2). To create these maps, the environmental variables selected for each species are gathered for each point of the study area for a period of time. The trained models are used to predict the probabilities of having high catches in each point of the study area depending on the environmental conditions of each day.

3. Results

3.1. Selected environmental predictors

The selected cut-off point that discriminates between “high” and “low” catches is the highest for SKJ (14 t), followed by YFT (6 t) and BET (3 t). Considering these cut-off points, and by using the CFS method, the combinations of the variables that best discriminate among the different classes defined for each species are summarized in Table 3. The number of selected variables vary from 23 for SKJ to 10 for BET. Thermal and biogeochemical variables within the water column are the most frequently selected for the three tropical tunas and the shark species studied in this paper. With regard to the physical properties of the seawater, the importance of water temperature is highlighted as it is among the first variables selected by the CFS method for all the species at different depths. Salinity also contributes to the characterisation of SKJ, BET and FAL, albeit at both surface and subsurface depths. Regarding the biogeochemical parameters, the variables that contribute most to the characterisation of the four species studied are silicate, phosphate, and nitrate concentrations at different depths. A common predictor for the three tropical tunas is the oxygen concentration integrated at the first 50 m. In addition, surface oxygen concentration is selected for SKJ, whereas for YFT oxygen concentration just below the surface (10 m) is selected. By contrast, for BET oxygen concentration in deeper layers (down to 175 m) is highlighted. Trophic variables related to food availability and energy transfer efficiency, such as chlorophyll concentration integrated at 20 m and 50 m are selected for YFT and SKJ. Similarly, chlorophyll fronts are selected in the case of SKJ. In the case of the oceanographic processes that have been analysed, thermocline (also related with the water temperature influence) intensity has been selected for SKJ and YFT, but not for BET. For the silky shark, the mixed layer, although with a lower F value, it is among the selected variables. The sea level anomaly has been selected for SKJ, YFT and for FAL. Lastly, as an indicator of geographic distribution, longitude seems to be important only for YFT, whereas latitude is the most important predictor for FAL. Finally, the month when the catches were registered is important for the two target tropical tunas, SKJ and YFT.

3.2. Performance of the models and forecast distribution

The mean accuracies and standard deviations of the 5 × 10-fold cross-validations for each method and species are shown in Table 4. Random Forest is the method that achieves the highest accuracy for all the analysed species (0.718 for SKJ, 0.728 for YFT, 0.589 for BET and 0.842 for FAL). Consequently, this is the method that has been used to build the prediction models.

A model has been trained for each species, by using the selected variables in Table 3. Then, for each model, the probabilities of finding high and low catches and absences of each species in the study area have been predicted by using the corresponding predictors given in Table 3. The results of the validation of the trained models are shown in Table 5,

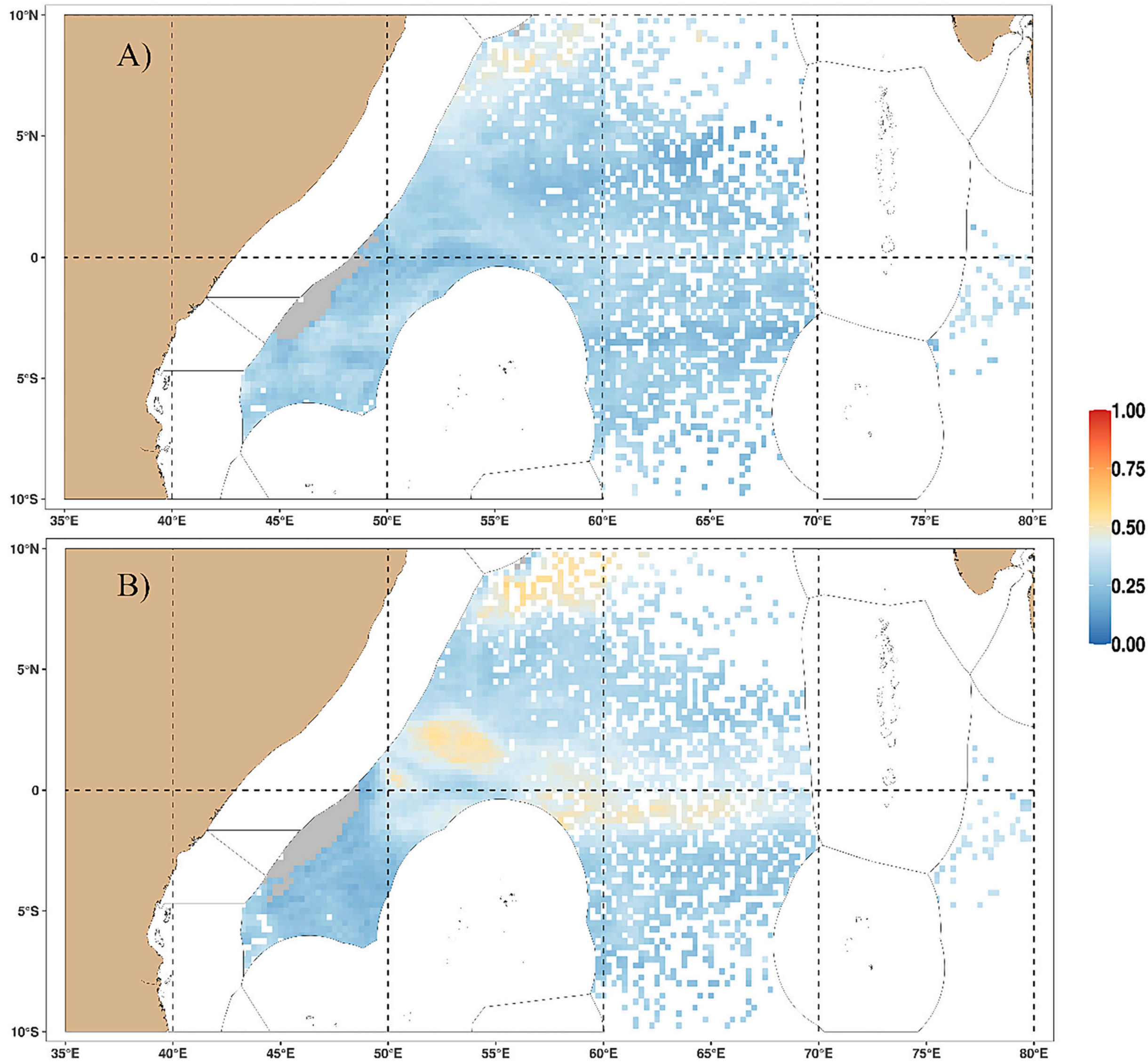

Fig. 2. Distribution forecast maps for the 21st of May 2020 in the Indian Ocean for YFT (A) and SKJ (B). The scale in the legend refers to high catch probability. The areas that have been predicted to have no captures (absences) have been painted in grey. EEZ areas are shaded in white.

where, in addition to the accuracy of each model, the true positive rate (TPR) and false positive rate (FPR) of each class are also shown. The TPR represents the proportion of actual positive cases that were correctly identified or classified as positive by the model. On the other hand, the FPR is calculated as the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events.

The species forecast models show overall promising performances, where those trained for SKJ and YFT show the highest accuracies: 0.718 and 0.728, respectively. Most of the errors made by these models may have occurred in the classification between high and low captures, while absences are very well classified with true positive rates of 0.996 and 0.993 for SKJ and YFT, respectively. For BET an accuracy of 0.589 is achieved, which is significantly lower than accuracies of the other two species. This is probably related to the fact that, since BET is a less important target than SKJ or YFT, absences have not been considered for the model, and it is the species with the lowest catch number in the input catch data. The model built to discriminate between FAL presences and absences has achieved a high accuracy (84%), which is key in a model that will be used to avoid bycatch of this species.

After validation of the models, forecasts of catch distribution have been estimated, taking as input forecast of oceanographic variables. These daily probability maps have been computed for each species. Since the training data does not cover the whole study area, only the probabilities of the positions that have been included in the training data have been included in these forecast maps. Fig. 2 shows an example of the output of these models in forecast mode, where the probability of finding high catches of SKJ and YFT for one specific day (21st May 2020) and for the study area is provided. In this particular example, overall, the probabilities of catching SKJ are higher than for YFT. The main fishing ground for YFT is found in front of the northern Somalian coast, whereas for SKJ tuna, the north and the east areas of Seychelles archipelago are also described as probable fishing areas.

4. Discussion

A careful selection of the variables has been the cornerstone of this study, since for training accurate ML models, it is essential to use variables that are good predictors of the feature that is being modelled. The variables selected in this study have been shown in previous studies to have a significant influence on the species analysed. For instance, the importance of **temperature** has been highlighted in this study and

Table 3
Variables selected for each species in the Indian Ocean based on the CFS method together with the values of the F and SUS statistics. The variables were ranked according to F (the number of folds for which each variable was selected).

SKJ			YFT			BET			FAL		
Predictor	F	SUS	Predictor	F	SUS	Predictor	F	SUS	Predictor	F	SUS
SST	10	0.0078	Oxygen concentration integrated at 10 m	10	0.0080	Silicate at 150 m	10	0.0068	Latitude	10	0.0080
Temperature at 50 m	10	0.0078	Temperature at 50 m	10	0.0079	Salinity at 50 m	10	0.0067	Salinity at 100 m	10	0.0066
Month	10	0.0075	Nitrate at 50 m	10	0.0078	Temperature at 50 m	10	0.0066	SST	10	0.0052
Temperature gradient at 200 m	10	0.0071	Surface phosphate	10	0.0075	SST	10	0.0055	Bottom temperature	10	0.0037
Chlorophyll concentration integrated at 20 m	10	0.0069	Month	10	0.0069	Nitrate at 100 m	10	0.0048	Temperature gradient at 100 m	10	0.0029
Nitrate at 50 m	10	0.0067	Nitrate at 20 m	10	0.0056	Salinity at 200 m	10	0.0047	Phosphate at 50 m	10	0.0026
Phosphate at 50 m	10	0.0064	SST	10	0.0055	Oxygen concentration integrated at 50 m	10	0.0041	Surface nitrate	10	0.0025
Nitrate at 20 m	10	0.0064	Temperature at 150 m	10	0.0051	Oxygen concentration integrated at 175 m	9	0.0094	Salinity at 200 m	9	0.0040
Temperature gradient at 150 m	10	0.0061	Phosphate at 175 m	10	0.0050	Phosphate at 10 m	7	0.0058	Silicate at 100 m	9	0.0026
Sea level anomaly	10	0.0061	Longitude	10	0.0048	Phosphate at 150 m	5	0.0070	Current speed at 150 m	9	0.0026
Phosphate at 200 m	10	0.0060	Temperature gradient at 100 m	10	0.0046				Temperature gradient at 10 m	8	0.0023
Nitrate at 200 m	10	0.0054	Temperature gradient at 50 m	10	0.0038				Mixed layer depth	8	0.0021
Oxygen concentration integrated at 100 m	10	0.0052	Temperature at 100 m	10	0.0038				Silicate at 200 m	7	0.0033
Temperature gradient at 50 m	10	0.0048	Bottom temperature	10	0.0037				Sea level anomaly	6	0.0029
Temperature gradient at 100 m	10	0.0047	Silicate at 50 m	10	0.0032						
Silicate at 50 m	10	0.0047	Chlorophyll concentration integrated at 50 m	9	0.0069						
Oxygen concentration integrated at 50 m	10	0.0042	Oxygen concentration integrated at 50 m	9	0.0055						
Salinity at 50 m	10	0.0038	Oxygen concentration integrated at 125 m	9	0.0046						
Thermocline intensity	10	0.0028	Sea level anomaly	8	0.0056						
Surface oxygen concentration	9	0.0059	Thermocline intensity	7	0.0019						
Temperature at 100 m	9	0.0051									
Silicate at 100 m	9	0.0051									
Chlorophyll fronts	7	0.0034									

Table 4
Mean accuracies and their standard deviations after the 5 × 10-fold cross-validation of each model.

	SKJ	YFT	BET	FAL
Random Forest	0.718 (±0.76)	0.728 (±0.7)	0.589 (±1.97)	0.842 (±0.77)
SMO	0.463 (±1.18)	0.439 (±1.13)	0.535 (±1.65)	0.671 (±0.01)
Multilayer Perceptron	0.714 (±1.15)	0.708 (±1.73)	0.546 (±2.28)	0.668 (±1.66)
Naïve Bayes	0.436 (±1.16)	0.415 (±1.18)	0.553 (±1.99)	0.655 (±0.51)

according to Arrizabalaga et al. (2015), each tuna species has its own temperature preference; thus, YFT prefers higher surface temperatures (above 25 °C) than BET and SKJ (between 20 and 28 °C). In the case of silky shark, its temperature preference in the Indian Ocean is between 28 and 30 °C (Lezama-Ochoa et al., 2016). Regarding the **salinity,** it can influence the large-scale spatial distribution of tunas (Druon et al., 2017; Fromentin et al., 2014; Maury et al., 2001; Reygondeau et al., 2012; Vahabnezhad et al., 2023) and silky sharks (Lezama-Ochoa et al., 2016). Although the mechanisms through which salinity could affect the distribution of tuna and silky shark remain unclear, it could be a proxy of other underlying processes. In a study focused on the tropical Atlantic, Maury et al. (2001) suggested that low salinities induced by fluvial water supplies could indicate favourable trophic areas for juvenile tuna in the Gulf of Guinea, or that it could be related to osmotic regulation. Lopetegui-Eguren et al. (2022) suggested that the presence of a shark species in low-**nitrogen** waters could be explained by its foraging behaviour as a predator using vision for feeding. Tropical tunas, which are also visual predators, could prefer oligotrophic waters for easier localisation of prey aggregations when diving. This is again in

Table 5
Cross-validation results for each tuna species and the silky shark with accuracy (acc.), TPR and FPR values for high and low catch and absence (abs.) classes. Note that n refers to the number of observations accounted for each species model and that the standard deviation is indicated in brackets.

	SKJ (n = 19,140)		YFT (n = 16,265)		BET (n = 6847)		FAL (n = 15,338)	
Acc.	0.718 (±0.76)		0.728 (±0.7)		0.589 (±1.97)		0.842 (±0.77)	
	TPR	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**
HIGH	0.571	0.201	0.560	0.184	0.586	0.407	0.863	0172
LOW	0.594	0.21	0.641	0.217	0.593	0.414		
ABS.	0.996	0.013	0.993	0.013			0.828	0.137

accordance with their vertical distribution, and it highlights the importance of the **oxygen concentration** as a limiting factor for the three tropical tunas, as shown in the literature (Chan, 2023; Druon et al., 2017). Oxygen affects important biological processes and thus determines the spatial distribution of tunas (Bard et al., 1998; Barkley et al., 1978; Boyce et al., 2008; Brill, 1994; Stramma et al., 2012), particularly in tropical areas where it is considered a good predictor for vertical and horizontal limitations due to tunas' physiological needs (Reilly and Fiedler, 1994). The trophic variables related to food availability and energy transfer efficiency, such as **chlorophyll** concentration integrated at 20 m and 50 m (Druon et al., 2017; Fernandes et al., 2013b; Heneghan et al., 2021; López et al., 2020), were selected for YFT and SKJ, respectively in accordance with the vertical distribution of these species. In accordance with Zainuddin et al. (2017) tunas show a preferential distribution near strong chlorophyll **fronts**. In the case of the thermocline, its intensity has been selected for SKJ and YFT, but not for BET. These results could be related with their respective vertical distributions patterns in tropical areas (Bard et al., 1998): the most frequent depth of BET is mainly below the thermocline (Hampton et al., 1998) while the preferred depth of YFT and SKJ is near or above the thermocline (Schaefer et al., 2009; Song et al., 2008). Finally, the sea level, which has been also selected as a predictor for some species, has been related to the presence of tuna juveniles (López et al., 2020). The **sea level anomaly** can be considered as a proxy of the presence of mesoscale eddies. These rotating water masses influence the surrounding ecosystems by regulating the horizontal and vertical dynamics of the water column. As such, it has been observed that the increasing opportunities for foraging driven by certain eddies lead to an aggregation of pelagic predators (Arostegui et al., 2022).

Regarding the performance of the **models**, the forecast maps previously analysed for both target tropical tunas identified the area off the Somalian coast as being a probable high catch area. This is a well-known fishery area for YFT tuna, either on free school (Vahabnezhad et al., 2023) or with FADs (Fonteneau et al., 2002). Similarly, it has been described as a feeding area for migratory pelagic species such as SKJ (Druon et al., 2017) during the summer monsoon season, when the upwelling of the area reduces SST and increases biological productivity (Young et al., 2015). Moreover, the model seems to successfully predict SKJ catch areas off the south coast of Somalia (around 0°-5°N and 50°-55°E) in accordance with Orue et al. (2020), who also showed high probabilities of finding tropical tunas in this area in May. Even though the models built to forecast tropical tuna distribution in the Indian Ocean successfully predict the main catch areas of these target species, they fail to forecast other catch areas. However, they are very good at predicting areas with no presence of the studied species. This is desirable due to the high cost (fuel consumption and emissions) of sailing to anticipated fishing grounds in areas where there is no tuna or where are low biomasses (Basurko et al., 2022; Granado et al., 2024).

The need to adapt to climate change and contribute to its mitigation was highlighted in a review by the FAO (Barange et al., 2018). The adaptation of the tuna fishing industry may be crucial, considering that changes in the distribution of fish species are already being observed (Baudron et al., 2020; Chust et al., 2019; Erauskin-Extramiana et al., 2020) and may be exacerbated in the near future (Erauskin-Extramiana et al., 2023; Lezama-Ochoa et al., 2016). The models developed following the presented **methodology** might help to improve the fishing industry's capacity for **adapting** to climate change by regularly updating species distribution forecasts based on up-to-date data. The distribution forecast maps built in this study could be a step forward to identify probable fishing areas and could be useful for tracking changes in the spatio-temporal distribution of target tuna species. This change in the spatial distribution of tunas will also affect the distribution of fishing effort. Therefore, given the direct relationship between distances travelled and fuel costs, any changes in the distance to fishing grounds would have a direct impact on the economic performance of the fishery (Chan, 2023). Indeed, despite political recognition of the need to address climate change adaptation, climate-adaptation targets in fisheries management have not been set (Bryndum-Buchholz et al., 2021).

Our approach is in line with the three main fisheries adaptation strategies identified in the literature (Galappaththi et al., 2022): coping mechanisms (e.g. change of fishing grounds), adaptive strategies (e.g. incorporation of technology); and management responses (e.g. adaptation planning). These adaptation strategies can help reduce fuel consumption per landed tonne of fish and, therefore, the carbon footprint of fisheries to reduce the impact of the fishing fleets on climate change (Palomares et al., 2021; Vinuesa et al., 2020). Indeed, most of the fuel consumption of purse seiners targeting migratory pelagic species is related to time sailing to the fishing ground or finding fish (Bastardie et al., 2022). Hence, routing methods based on target species distribution models, such as the models built in this study, appear as the most suitable approach to reduce the distance and time at sea spent searching for fishing grounds as suggested in Granado et al. (2024). However, to the best of our knowledge, there are no tuna models in the literature that differentiate between the main three tuna species and bycatch, based solely on environmental data. Thus, the proposed ML models could enable the individual consideration of the main tuna species and the predominant bycatch species. This distinction provides additional flexibility when defining routes, enabling the selection of optimal fishing grounds according to the commercial interest for each species while also mitigating bycatch.

Similarly, habitat prediction of bycatch species could reduce unwanted catches of vulnerable species such as silky sharks. In addition, this type of predictions could aid in illegal fishing detection (Watson et al., 2023). Consequently, distribution forecast maps could be used as efficient spatial management strategies when planning fishing operations. This would maintain similar yields for target species while simultaneously reducing interactions with vulnerable species. Indeed, the efficacy of spatial management strategies remains a priority research area in tRFMOs (Hilborn et al., 2022; Kaplan et al., 2014; Lopetegui-Eguren et al., 2022; Tolotti et al., 2015). Having models for different target and bycatch species could help build a real-time decision-making system that bears allocated quotas in mind. The reduction of fuel consumption due to less time spent searching for fishing grounds would help mitigate climate change caused by the fishing industry and reduce operating costs (Granado et al., 2021). Nevertheless, it must be considered that this kind of models learn from historical catch data, and if the data used to train them is not significant enough, the models will not be able to make good predictions. Considering this fact, performing periodic updates of the models with new catch data is desired to improve their performance.

Nowadays, an increasing number of ML techniques are being applied in marine ecology (Rubbens et al., 2023) that could be interesting to be tested in future studies. Likewise, exploration of spatio-temporal modelling (e.g. geographical RF, eXtreme Gradient Boosting or Gausian Process boosting) (Georganos and Kalogirou, 2022; Li, 2022; Sigrist, 2022), and performing spatial and temporal cross validation could help to better understand the predictive performance of the models in different fishing dates and positions. Further research on recent approaches to interpret model predictions could contribute to better understand the spatio-temporal changes and mechanisms of the studied species (Lundberg and Lee, 2017). While, exploring interpretable ML tools, such as Shapley additive feature explanations (SHAP) for assessing the predictor marginal effects (Aria et al., 2021; Lundberg and Lee, 2017) could be useful to evaluate the importance of selected environmental variables on the distribution of each species. Finally, a step forward in this research topic would be to evaluate whether the inclusion of the proposed models when planning a fishing trip, result in the reduction of searching effort and bycatch, ultimately defining more efficient routes. This can be achieved by comparing the efficiency of suggested routes provided by some routing algorithm that use the proposed models with i) those that do not utilize any ML model outputs (historical routes recorded by vessel monitoring systems); ii) those that consider tuna distribution without distinguishing between species as suggested in Granado et al. (2024).

CRediT authorship contribution statement

Nerea Goikoetxea: Methodology, Supervision, Writing – original draft, Writing – review & editing. **Izaro Goienetxea:** Methodology, Software, Validation, Writing – review & editing. **Jose A. Fernandes-Salvador:** Conceptualization, Methodology, Project administration, Writing – review & editing. **Nicolas Goñi:** Methodology, Software, Validation, Writing – review & editing. **Igor Granado:** Methodology, Software, Writing – review & editing. **Iñaki Quincoces:** Methodology, Software, Validation. **Leire Ibaibarriaga:** Methodology, Software, Validation. **Jon Ruiz:** Methodology, Software, Validation. **Hilario Murua:** Conceptualization, Writing – review & editing. **Ainhoa Caballero:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that may have influenced the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 869342 (SusTunTech). This study has been conducted using EU Copernicus Marine Service (https://marine.copernicus.eu/) and JPL MUR MEaSUREs Project (https://podaac.jpl.nasa.gov/MEaSUREs-MUR) environmental data and tuna catches compiled under the compulsory compliance European Regulation. The authors are grateful to the reviewer's valuable comments, which improved the manuscript. We also thank Doctors Marina Santurtun, Josu Santiago and Xabier Irigoien for reviewing the first draft of the manuscript. Igor Granado has benefited from the IKERTALENT grant of the Department of Economic Development and Infrastructures of the Basque Government. This paper is contribution No. 1215 from AZTI, Marine Research, Basque Research and Technology Alliance (BRTA).

References

Agra, A., Christiansen, M., Delgado, A., Hvattum, L.M., 2015. A maritime inventory routing problem with stochastic sailing and port times. Comput. Oper. Res. 61, 18–30. https://doi.org/10.1016/j.cor.2015.01.008.

Amandè, M.J., Ariz, J., Chassot, E., Delgado de Molina, A., Gaertner, D., Murua, H., Pianet, R., Chavance, P., 2010. Bycatch of the European purse seine tuna fishery in the Atlantic Ocean for the 2003-2007 period. Aquat. Living Resour. 23 (4), 353–362.

Amandè, M.J., N'Cho, A.J., Kouakou, N.D., N'Cho, C.M., Koffi, K.F., Kouadio, A.N.C., Dewals, P., Restrepo, V., 2017. Utilization and trade of faux poisson landed in Abidjan. Collect Vol Sci Pap ICCAT 73 (2), 749–754.

Amante, C., Eakins, B.W., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedimientos, fuentes de datos y análisis. Memorando técnico de NOAA NESDIS NGDC-24. Centro Nacional de Datos Geofísicos, NOAA. https://doi.org/10.7289/V5C8276M[26/10/2021].

Aria, M., Cuccurullo, C., Gnasso, A., 2021. A comparison among interpretative proposals for Random Forests. Machine Learn. Appl. 6, 100094.

Arostegui, M.C., Gaube, P., Woodworth-Jefcoats, P.A., Kobayashi, D.R., Braun, C.D., 2022. Anticyclonic eddies aggregate pelagic predators in a subtropical gyre. Nat 609 (7927), 535–540.

Arregui, I., Goñi, N., Chifflet, M., 2019. Migration patterns of yellowfin, skipjack and bigeye tunas in the tropical Atlantic, based on recent tagging and recapture data. Collect. Vol. Sci. Pap. ICCAT 76 (6), 903–950.

Arrizabalaga, H., Dufour, F., Kell, L., Merino, G., Ibaibarriaga, L., Chust, G., Irigoien, X., Santiago, J., Murua, H., Fraile, I., Chifflet, M., Goikoetxea, N., Sagarminaga, Y., Aumont, O., Bopp, L., Herrera, M., Fromentin, J.M., Bonhomeau, S., 2015. Global habitat preferences of commercially valuable tuna. Deep-Sea Res. II Top. Stud. Oceanogr. 113, 102–112. https://doi.org/10.1016/j.dsr2.2014.07.001.

Bakun, A., 2006. Fronts and eddies as key structures in the habitat of marine fish larvae: opportunity, adaptive response and competitive advantage. Sci Mar. 70 (S2), 105–122.

Barange, M., Bahri, T., Beveridge, M.C., Cochrane, K.L., Funge-Smith, S., Poulain, F., 2018. Impacts of climate change on fisheries and aquaculture: synthesis of current knowledge, adaptation and mitigation options. FAO Fisheries and Aquaculture Technical Paper No. 627. Rome, FAO, 628 pp.

Bard, F.X., Bach, P., Josse, E., 1998. Habitat, ecophysiologie des thons: Quoi de neuf depuis 15 ans? ICCAT Col Vol Sci Pap. 50, 319–342.

Barkley, R.A., Neill, W.H., Gooding, R.M., 1978. Skipjack tuna, Katsuwomuns pelamis, habitat based on temperature and oxygen requirements. Fish. Bull. 76 (3), 653–662.

Bastardie, F., Feary, D.A., Brunel, T., Kell, L.T., Döring, R., Metz, S., Eiqaard, O.R., Basurko, O.C., Bartolino, V., Bentley, J., Berges, B., Bossier, S., Brooks, M.E., Caballero, A., Citores, L., Daskalov, G., Depestele, J., Gabiña, G., Aranda, M., Hamon, K.G., Hidalgo, M., Katsanevakis, S., Kempf, A., Kühn, B., Nielsen, J.R., Püts, M., Taylor, M., Triantaphyllidis, G., Tsagarakis, K., Urtizberea, A., van Hoof, L., van Vlasselaer, J., 2022. Ten lessons on the resilience of the EU common fisheries policy towards climate change and fuel efficiency – a call for adaptive, flexible and well-informed fisheries management. Front. Mar. Sci. 9, 947150 https://doi.org/10.3389/fmars.2022.947150.

Basurko, O.C., Gabiña, G., López, J., Granado, I., Murua, H., Fernandes, J.A., Krug, I., Ruiz, J., Uriondo, Z., 2022. Fuel consumption of free-swimming school versus FAD strategies in tropical tuna purse seine fishing. Fish. Res. 245 https://doi.org/10.1016/j.fishres.2021.106139.

Baudron, A.R., Brunel, T., Blanchet, M.A., Hidalgo, M., Chust, G., Brown, E.J., Kleisner, K.M., Millar, C., MacKenzie, B.R., Nikolioudakis, N., Fernandes, J.A., Fernandes, P.G., 2020. Changing fish distributions challenge the effective management of European fisheries. Ecography 43 (4), 494–505. https://doi.org/10.1111/ecog.04864.

Belkin, I.M., O'Reilly, J.E., 2009. An algorithm for oceanic front detection in chlorophyll and SST satellite imagery. J. Mar. Syst. 78 (3), 319–326. https://doi.org/10.1016/j.jmarsys.2008.11.018.

Bell, J.D., Watson, R.A., Ye, Y., 2017. Global fishing capacity and fishing effort from 1950 to 2012. Fish Fish. 18 (3), 489–505.

Boyce, D.G., Tittensor, D.P., Worm, B., 2008. Effects of temperature on global patterns of tuna and billfish richness. Mar. Ecol. Prog. Ser. 355, 267–276.

Breiman, L., 2001. Random Forests. Mach. Learn. 45 (1), 5–32.

Brill, R.W., 1994. A review of temperature and oxygen tolerance studies of tunas pertinent to fisheries oceanography, movement models and stock assessments. Fish. Oceanogr. 3 (3), 204–216. https://doi.org/10.1111/j.1365-2419.1994.tb00098.x.

Bryndum-Buchholz, A., Tittensor, D.P., Lotze, H.K., 2021. The status of climate change adaptation in fisheries management: policy, legislation and implementation. Fish Fish. 22 (6), 1248–1273. https://doi.org/10.1111/faf.12586.

Chan, H.L., 2023. How climate change and climate variability affected trip distance of a commercial fishery. PLOS Clim. 2 (2) https://doi.org/10.1371/journal.pclm.0000143 e0000143.

Christiansen, M., Fagerholt, K., Ronen, D., 2004. Ship routing and scheduling: status and perspectives. Transplant. Sci. 38 (1), 1–18. https://doi.org/10.1287/trsc.1030.0036.

Chrysafis, K.A., Theotokas, I.N., Lagoudis, I.N., 2022. Managing fuel price variability for ship operations through contracts using fuzzy TOPSIS. Res. Transp. Bus. Manag. 100778 https://doi.org/10.1016/j.rtbm.2021.100778.

Chust, G., Goikoetxea, N., Ibaibarriaga, L., Sagarminaga, Y., Arregui, I., Fontán, A., Irigoien, X., Arrizabalaga, H., 2019. Earlier migration and distribution changes of albacore in the Northeast Atlantic. Fish. Oceanogr. 28 (5), 505–516. https://doi.org/10.1111/fog.12427.

Clavareau, L., Sabarros, P.S., Escalle, L., Bach, P., Abascal, F.J., López, J., Murua, H., Alayon, P.J.P., Ramos, M.L., Ruiz, J., Mérigot, B., 2020. Elasmobranch bycatch distributions and mortality: insights from the European tropical tuna purse-seine fishery. Glob. Ecol. Conserv. 24 e01211, ISSN 2351-9894.

Curnick, D.J., Andrzejaczek, S., Jacoby, D.M., Coffey, D.M., Carlisle, A.B., Chapple, T.K., Ferretti, F., Schallert, R.J., White, T., Block, B.A., Koldewey, H.J., Collen, B., 2020. Behavior and ecology of silky sharks around the Chagos archipelago and evidence of Indian Ocean wide movement. Front. Mar. Sci. 7, 596619 https://doi.org/10.3389/fmars.2020.596619.

Dagorn, L., Holland, K.N., Restrepo, V., Moreno, G., 2012. Is it food or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? Fish Fish. https://doi.org/10.1111/j.1467-2979.2012.00478.x.

Druon, J.N., Chassot, E., Murua, H., López, J., 2017. Skipjack tuna availability for purse seine fisheries is driven by suitable feeding habitat dynamics in the Atlantic and Indian Oceans. Front. Mar. Sci. 4, 315. https://doi.org/10.3389/fmars.2017.00315.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification. Wiley, New York, NY, USA.

Erauskin-Extramiana, M., Arrizabalaga, H., Cabre, A., Coelho, R., Rosa, D., Ibaibarriaga, L., Chust, G., 2020. Are shifts in species distribution triggered by climate change? A swordfish case study. Deep Sea Res. Part II Top. Stud. Oceanogr. 175, 104666 https://doi.org/10.1016/j.dsr2.2019.104666.

Erauskin-Extramiana, M., Chust, G., Arrizabalaga, H., Cheung, W.W., Santiago, J., Merino, G., Fernandes-Salvador, J.A., 2023. Implications for the global tuna fishing industry of climate change-driven alterations in productivity and body sizes. Glob. Planet. Chang. 222, 104055 https://doi.org/10.2139/ssrn.4059543.

FAO, 2022. The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. FAO, Rome. https://doi.org/10.4060/cc0461en.

Fernandes, J.A., Irigoien, X., Goikoetxea, N., Lozano, J.A., Inza, I., Pérez, A., Bode, A., 2010. Fish recruitment prediction, using robust supervised classification methods. Ecol. Model. 221 (2), 338–352. https://doi.org/10.1016/j.ecolmodel.2009.09.020.

Fernandes, J.A., Cheung, W.W., Jennings, S., Butenschön, M., de Mora, L., Frölicher, T.L., Barange, M., Grant, A., 2013a. Modelling the effects of climate change on the distribution and production of marine fishes: accounting for trophic interactions in a dynamic bioclimate envelope model. Glob. Chang. Biol. 19 (8), 2596–2607. https://doi.org/10.1111/gcb.12231.

Fernandes, J.A., Lozano, J.A., Inza, I., Irigoien, X., Pérez, A., Rodríguez, J.D., 2013b. Supervised pre-processing approaches in multiple class variables classification for fish recruitment forecasting. Environ. Model. Softw. 40, 245–254. https://doi.org/10.1016/j.envsoft.2012.10.001.

Fernandes-Salvador, J.A., Oanta, G.A., Olivert-Amado, A., Goienetxea, I., Ibaibarriaga, L., Aranda, M., Cuende, E., Foti, G., Olabarrieta, I., Murua, J., Prellezo, R., Iñarra, B., Quincoces, I., Caballero, A., Sobrino-Heredia, J.M., 2022. Research for PECH Committee – Artificial Intelligence and the fisheries sector, European Parliament, Policy Department for Structural and Cohesion Policies, Brussels (11) (PDF) Research for PECH Committee – Artificial Intelligence and the fisheries sector. Available from: https://www.researchgate.net/publication/361391358_Research_for_PECH_Committee_-_Artificial_Intelligence_and_the_fisheries_sector#fullTextFileContent.

Fiedler, P.C., Bernard, H.J., 1987. Tuna aggregation and feeding near fronts observed in satellite imagery. Cont. Shelf Res. 7 (8), 871–881. https://doi.org/10.1016/0278-4343(87)90003-3.

Fonteneau, A., Hallier, J.P., 2015. Fifty years of dart tag recoveries for tropical tuna: a global comparison of results for the western Pacific, eastern Pacific, Atlantic, and Indian Oceans. Fish. Res. 163, 7–22.

Fonteneau, A., Bargain, R.-M., Nordstrom, V., Pallares, P., 2002. Atlas of Indian ocean purse seine fisheries 1982-2001 with a special emphasis for yellowfin tuna taken on FAD and free schools. IOTC Proceedings 5, 282–309.

Frank, E., Hall, M., Witten, I., 2016. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Fourth edition. Morgan Kaufmann.

Fromentin, J.M., Reygondeau, G., Bonhommeau, S., Beaugrand, G., 2014. Oceanographic changes and exploitation drive the spatio-temporal dynamics of Atlantic bluefin tuna (Thunnus thynnus). Fish. Oceanogr. 23 (2), 147–156. https://doi.org/10.1111/fog.12050.

Galappaththi, E.K., Susarla, V.B., Loutet, S.J., Ichien, S.T., Hyman, A.A., Ford, J.D., 2022. Climate change adaptation in fisheries. Fish Fish. 23 (1), 4–21. https://doi.org/10.1111/faf.12595.

Georganos, S., Kalogirou, S.A., 2022. Forest of forests: a spatially weighted and computationally efficient formulation of geographical random forests. ISPRS Int. J. Geo Inf. 11, 471. https://doi.org/10.3390/ijgi11090471.

Granado, I., Hernando, L., Galparsoro, I., Gabiña, G., Groba, C., Prellezo, R., Fernandes, J.A., 2021. Towards a framework for fishing route optimization decision support systems: review of the state-of-the-art and challenges. J. Clean. Prod. 320, 128661 https://doi.org/10.1016/j.jclepro.2021.128661.

Granado, I., Hernando, L., Uriondo, Z., Fernandes-Salvador, J.A., 2024. A fishing route optimization decision support system: the case of the tuna purse seiner. Eur. J. Oper. Res. 312 (2), 718–732. https://doi.org/10.1016/j.ejor.2023.07.009.

Groba, C., Sartal, A., Vázquez, X.H., 2015. Solving the dynamic traveling salesman problem using a genetic algorithm with trajectory prediction: an application to fish aggregating devices. Comput. Oper. Res. 56, 22–32. https://doi.org/10.1016/j.cor.2014.10.012.

Groba, C., Sartal, A., Vázquez, X.H., 2018. Integrating forecasting in metaheuristic methods to solve dynamic routing problems: evidence from the logistic processes of tuna vessels. Eng. Appl. Artif. Intell. 76, 55–66. https://doi.org/10.1016/j.engappai.2018.08.015.

Hall, M.A., 2000. Correlation-based feature selection of discrete and numeric class machine learning. In: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 359–366.

Hall, M., Roman, M., 2013. Bycatch and non-tuna catch in the tropical tuna purse seine fisheries of the world. FAO Fisheries and Aquaculture Technical Paper No. 568. Rome, FAO, 249 pp.

Hallier, J.P., Fonteneau, A., 2015. Tuna aggregation and movement from tagging data: a tuna "hub" in the Indian Ocean. Fish. Res. 163, 34–43.

Hampton, J., Bigelow, K., Labelle, M., 1998. A summary of current information on the biology, fisheries and stock assessment of bigeye tuna (*Thunnus obesus*) in the Pacific Ocean, with recommendations for data requirement and future research. http://oceanfish.spc.int/en/publications/doc_download/540-1998-biology-fisheries-and-stock-assessment-of-bigeye-tuna-thunnus-obesus-in-the-pacific-ocean [accessed Nov 8 2023].

Haykin, S., 1994. Neural Networks: A Comprehensive Foundation. Prentice Hall PTR.

Heneghan, R.F., Galbraith, E., Blanchard, J.L., Harrison, C., Barrier, N., Bulman, C., Cheung, W., Coll, M., Eddy, T.D., Erauskin-Extramiana, M., Everett, J.D., Fernandes-Salvador, J.A., Gascuel, D., Guiet, J., Maury, O., Palacios-Abrantes, J., Petrik, C.M., du Pontavice, H., Richardson, A.J., Steenbeek, J., Tai, T.C., Volkholz, J., Woodworth-Jefcoats, P.A., Tittensor, D.P., 2021. Disentangling diverse responses to climate change among global marine ecosystem models. Prog. Oceanogr. 198, 102659 https://doi.org/10.1016/j.pocean.2021.102659.

Hilborn, R., Agostini, V.N., Chaloupka, M., Garcia, S.M., Gerber, L.R., Gilman, E., Hanich, Q., Himes-Cornell, A., Hobday, A.J., Itano, D., Kaiser, M.J., Murua, H., Ovando, D., Pilling, G.M., Rice, J.C., Sharma, R., Schaefer, K.M., Severance, C.J., Taylor, N.G., Fitchett, M., 2022. Area-based management of blue water fisheries: current knowledge and research needs. Fish Fish. 23 (2), 492–518. https://doi.org/10.1111/faf.12629.

Honarmand Ebrahimi, S., Ossewaarde, M., Need, A., 2021. Smart fishery: a systematic review and research agenda for sustainable fisheries in the age of AI. Sustainability 13 (11), 6037. https://doi.org/10.3390/su13116037.

ISSF, 2020. Status of the World Fisheries for Tuna ISSF Technical Report 2020–16. International Seafood Sustainability Foundation, Washington, D.C., USA.

IUCN, 2021. The IUCN Red List of Threatened Species. Version 2021-3. Available at: www.iucnredlist.org [Accessed: 10 March 2022].

John, G.E., Langley, P., 1995. Estimating continuous distributions in Bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, pp. 338–345.

Kaplan, D.M., Chassot, E., Amandé, J.M., Dueri, S., Demarq, H., Dagorn, L., Fonteneau, A., 2014. Spatial management of Indian Ocean tropical tuna fisheries: potential and perspectives. ICES J. Mar. Sci. 71 (7), 1728–1749. https://doi.org/10.1093/icesjms/fst233.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28 (5), 1–26. https://doi.org/10.18637/jss.v028.i05. https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

Lau-Medrano, W., 2020. grec: Gradient-Based Recognition of Spatial Patterns in Environmental Data. R package version 1.4.1. https://CRAN.Rproject.org/package=grec.

Lezama-Ochoa, N., Murua, H., Chust, G., Van Loon, E., Ruiz, J., Hall, M., Chavance, P., Delgado de Molina, A., Villarino, E., 2016. Present and future potential habitat distribution of Carcharhinus falciformis and Canthidermis maculata by-catch species in the tropical tuna purse-seine fishery under climate change. Front. Mar. Sci. 3, 34. https://doi.org/10.3389/fmars.2016.00034.

Li, Z., 2022. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost, Computers, Environment and Urban Systems,Volume 96, 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845. ISSN 0198–9715.

Lopetegui-Eguren, J., Poos, J.J., Arrizabalaga, H., Guirhem, G.L., Murua, H., Lezama-Ochoa, N., Griffiths, S.P., Ruiz Gondra, J., Sabarros, P.S., Báez, J.C., Juan-Jordá, M. J., 2022. Spatio-temporal distribution of juvenile oceanic whitetip shark incidental catch in the Western Indian Ocean. Front. Mar. Sci. 9, 863602 https://doi.org/10.3389/fmars.2022.863602.

López, J., Alvarez-Berastegui, D., Soto, M., Murua, H., 2020. Using fisheries data to model the oceanic habitats of juvenile silky shark (Carcharhinus falciformis) in the tropical eastern Atlantic Ocean. Biodivers. Conserv. 29 (7), 2377–2397. https://doi.org/10.1007/s10531-020-01979-7.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Proces. Syst. 30. ISBN: 9781510860964.

Maury, O., Gascuel, D., Marsac, F., Fonteneau, A., Rosa, A.L.D., 2001. Hierarchical interpretation of nonlinear relationships linking yellowfin tuna (Thunnus albacares) distribution to the environment in the Atlantic Ocean. Can. J. Fish. Aquat. Sci. 58 (3), 458–469. https://doi.org/10.1139/f00-261.

McCauley, D.J., Woods, P., Sullivan, B., Bergman, B., Jablonicky, C., Roan, A., Hirshfield, M., Boerder, K., Worm, B., 2016. Ending hide and seek at sea. Sci 351 (6278), 1148–1150.

Nataniel, A., Lopez, J., Soto, M., 2021. Modelling seasonal environmental preferences of tropical tuna purse seine fisheries in the Mozambique Channel. Fish. Res. 243, 106073 https://doi.org/10.1016/j.fishres.2021.106073.

Nielsen, J.R., Thunberg, E., Holland, D.S., Schmidt, J.O., Fulton, E.A., Bastardie, F., Waldo, S., 2018. Integrated ecological–economic fisheries models—evaluation, review and challenges for implementation. Fish Fish. 19 (1), 1–29.

Orue, B., Lopez, J., Pennino, M.-G., Moreno, G., Santiago, J., Murua, H., 2020. Comparing the distribution of tropical tuna associated with dirfting fish aggregating devices (DFADs) resulting from catch dependent and independent data. Deep-Sea Res. Part II: Top. Stud. Oceanogr. 175 https://doi.org/10.1016/j.dsr2.2020.104747.

Palomares, I., Martínez-Cámara, E., Montes, R., García-Moral, P., Chiachio, M., Chiachio, J., Alonso, S., Melero, F.J., Molina, D., Fernández, B., Moral, C., Marchena, R., Pérez de Vargas, J., Herrera, F., 2021. A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: progress and prospects. Appl. Intell. 51 (9), 6497–6527.

Parker, R.W., Tyedmers, P.H., 2014. Fuel consumption of global fishing fleets: current understanding and knowledge gaps. Fish Fish. 16 (4), 684–696.

Parker, R.W., Blanchard, J.L., Gardner, C., Green, B.S., Hartmann, K., Tyedmers, P.H., Watson, R.A., 2018. Fuel use and greenhouse gas emissions of world fisheries. Nat. Clim. Chang. 8 (4), 333–337.

Platt, J., 1998. Fast training of support vector machines using sequential minimal optimization. In: Schoelkopf, B., Burges, C., Smola, A. (Eds.), Advances in Kernel Methods - Support Vector Learning.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.

Reilly, S.B., Fiedler, P.C., 1994. Interannual variability of dolphin habitats in the eastern tropical Pacific. I: Research vessel surveys, 1986-1990. Fish. B.-NOAA 92 (2), 434–450.

Reygondeau, G., Maury, O., Beaugrand, J.M., Fromentin, A., Fonteneau, A., Cury, P., 2012. Biogeography of tunas and billfish communities. J. Biogeogr. 39, 114–129. https://doi.org/10.1111/j.1365-2699.2011.02582.x.

Rivas, A.L., Pisoni, J.P., 2010. Identification, characteristics and seasonal evolution of surface thermal fronts in the Argentinean Continental Shelf. J. Mar. Syst. 79, 134–143. https://doi.org/10.1016/j.jmarsys.2009.07.008.

Rodriguez, J.D., Perez, A., Lozano, J.A., 2009. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32 (3), 569–575. https://doi.org/10.1109/TPAMI.2009.187.

Rojon, I., Smith, T., 2014. On the attitudes and opportunities of fuel consumption 512 monitoring and measurement within the shipping industry and the identification and 513 validation of energy efficiency and performance interventions, 18 pp. Sebastiani, P., Abad, M. M., Ramoni, M. F., 2005. Bayesian networks. In: Maimom

O., Rokach O. (ed.). The Data Mining and Knowledge Discovery Handbook. Springer, New York, NY, USA, pp.193–230.
Roll, K.H., Asche, F., Bjørndal, T., 2022. The effect of introducing fuel tax to the Norwegian fishery industry. Mar. Policy 135, 104829. https://doi.org/10.1016/j.marpol.2021.104829.
Rubbens, P., Brodie, S., Cordier, T., Destro Barcellos, D., Devos, P., Fernandes-Salvador, J.A., Irisson, J.O., 2023. Machine learning in marine ecology: an overview of techniques and applications. ICES J. Mar. Sci. 80 (7), 1829–1853. https://doi.org/10.1093/icesjms/fsad100.
Rubio, I., Hobday, A.J., Ojea, E., 2022. Skippers' preferred adaptation and transformation responses to catch declines in a large-scale tuna fishery. ICES J. of Mar. Sci. 79, 532–539. https://doi.org/10.1093/icesjms/fsab065.
Sabarros, P.S., Romanov, E.V., Bach, P., 2015. Vertical behavior and habitat preferences of yellowfin and bigeye tuna in the South West Indian Ocean inferred from PSAT tagging data. IOTC Working Party on Tropical Tuna (WPTT), Montpellier, France, pp. 23–28.
Schaefer, K.M., Fuller, D.W., Block, B.A., 2009. Vertical movements and habitat utilization of skipjack (Katsuwonus pelamis), yellowfin (Thunnus albacares), and bigeye (Thunnus obesus) tunas in the equatorial eastern Pacific Ocean, ascertained through archival tag data. In: Nielsen, J.L., Arrizabalaga, H., Fragoso, N., Hobday, A., Lutcavage, M., Sibert, J. (Eds.), Tagging and Tracking of Marine Animals with Electronic Devices. Springer, New York, pp. 121–144. https://doi.org/10.1007/978-1-4020-9640-2.
Scientific, Technical and Economic Committee for Fisheries, 2021. In: Guillen, J., Virtanen, J., Prellezo, R., Carvalho, N. (Eds.), The 2021 annual economic report on the EU fishing fleet (STECF 21–08). Publications Office, 2021. https://data.europa.eu/doi/10.2760/60996.
Scott, G.P., Lopez, J., 2014. The Use of Fads in Tuna Fisheries. European parliament, Brussels. Available from. https://www.europarl.europa.eu/RegData/etudes/note/join/2014/514002/IPOL-PECH_NT(2014)514002_EN.pdf [Accessed 7 November 2023].
Sigrist, F., 2022. Gaussian process boosting. J. Mach. Learn. Res. 23 (1). Article 232 (January 2022), 46 pages.
Song, L.M., Zhang, Y.U., Xu, L.X., Jiang, W.X., Wang, J.Q., 2008. Environmental preferences of longlining for yellowfin tuna (Thunnus albacares) in the tropical high seas of the Indian Ocean. Fish. Oceanogr. 17 (4), 239–253. https://doi.org/10.1111/j.1365-2419.2008.00476.x.
Stramma, L., Prince, E.D., Schmidtko, S., Luo, J., Hoolihan, J.P., Visbeck, M., Wallace, D.W.R., Brandt, P., Körtzinger, A., 2012. Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes. Nat. Clim. Chang. 2 (1), 33–37.
Sun, C., Hobday, A.J., Condie, S.A., Baird, M.E., Eveson, J.P., Hartog, J.R., Richardson, A.J., Stevem, A.D.L., Wild-Allen, K., Babcock, R.C., Yang, D., Yu, R., Mongin, M., 2022. Ecological forecasting and operational information systems support sustainable ocean management. Forecasting 4, 1051–1079. https://doi.org/10.3390/forecast4040057.
Sund, P.N., Blackburn, M., Williams, F., 1981. Tunas and their environment in the Pacific Ocean: a review. Oceanogr. Mar. Biol. Annu. Rev. 19, 443–512.
Suuronen, P., Chopin, F., Glass, C., Løkkeborg, S., Matsushita, Y., Queirolo, D., Rihan, D., 2012. Low impact and fuel efficient fishing-looking beyond the horizon. Fish. Res. 119-120, 135–146.
Tolotti, M.T., Filmalter, J.D., Bach, P., Travassos, P., Seret, B., Dagorn, L., 2015. Banning is not enough: the complexities of oceanic shark management by tuna regional fisheries management organizations. Glob. Ecol. Conserv. 4, 1–7. https://doi.org/10.1016/j.gecco.2015.05.003.
Vahabnezhad, A., Taghavimothlagh, S.A., Salarpouri, A., Mohammadreza, M., 2023. Identifying the ecologically significant habitats of yellowfin tuna (Thunnus albacares, Bonnaterre, 1788) of Iranian purse seine fishery in the Gulf of Oman and Indian Ocean: an approach using satellite imagery and fishery data. Reg. Stud. Mar. Sci. 68, 103257 https://doi.org/10.1016/j.rsma.2023.103257.
Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M., Fuso Nerini, F., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat. Commun. 11 (1), 1–10.
Watson, J.T., Ames, R., Holycross, B., Suter, J., Somers, K., Kohler, C., Corrigan, B., 2023. Fishery catch records support machine learning-based prediction of illegal fishing off US west coast. PeerJ 11, e16215. https://doi.org/10.7717/peerj.16215.
Young, J.W., Hunt, B.P.V., Cook, T.R., Llopiz, J.K., Hazen, E.L., Pethybridge, H.R., Ceccarelli, D., Lorrain, A., Olson, R.J., Allain, V., Menkes, C., Patterson, T., Nicol, S., Lehodey, P., Kloser, R.J., Arrizabalaga, H., Choy, C.A., 2015. The trophodynamics of marine top predators: current knowledge, recent advances and challenges. Deep-Sea Res. II 113, 170–187. https://doi.org/10.1016/j.dsr2.2014.05.015.
Zainuddin, M., Farhum, A., Safruddin, S., Selamat, M.B., Sudirman, S., Nurdin, N., Syamsuddin, M., Ridwan, M., Saitoh, S.I., 2017. Detection of pelagic habitat hotspots for skipjack tuna in the Gulf of Bone-Flores Sea, southwestern Coral Triangle tuna, Indonesia. PLoS One 12 (10), e0185601. https://doi.org/10.1371/journal.pone.0185601.