

Predicting gear used in a multi-gear coastal fleet

P. Leitão^{a,b,*}, A. Campos^{a,b}, M. Castro^b

^a IPMA, Instituto Português do Mar e da Atmosfera, Rua Alfredo Magalhães Ramalho 6, Algés 1495-165, Portugal

^b Centre of Marine Sciences (CCMAR/CIMAR LA), Campus de Gambelas, Universidade do Algarve, Faro 8005-139, Portugal

ARTICLE INFO

Handled by A.E. Punt

Keywords:

Multi-gear fleet
Fishing effort
Machine learning
Random forest
Gear prediction

ABSTRACT

Knowledge of the gear used in multi-gear fisheries is crucial for supporting fisheries management. Still, the high complexity and lack of data in the Portuguese multi-gear coastal fleet compromise this task. The present study developed a method to predict main fishing gear used in each fishing trip for the Portuguese multi-gear coastal fleet based on landing records (species caught, port, and month of landing). Landing records were used to predict gear (available for part of the fleet with electronic logbooks) using a machine learning model (random forest). This model was then applied to the remaining trips of the fleet, without electronic logbooks, to predict the gear used. A total of six gear types were considered: bivalve dredges, traps, gillnets, trammel nets, drifting longlines, and bottom longlines. The overall model prediction error was 14 %; bivalve dredges and longlines had the lowest errors, and trammel nets and gillnets were the highest. The study sheds new light on important aspects of the dynamics of this fleet, namely a decreasing trend in the use of longlines, poor electronic logbook coverage for some gear types, and greater diversity in the catches obtained with nets compared to other gear types.

1. Introduction

Multi-gear fisheries are important to local communities worldwide due to their contribution to economic value and food supply, nevertheless, they present many challenges to fisheries management (FAO, 2022). Vessels employ a variety of gears, with most of them not being selective, resulting in the capture of numerous species (Nagasaki and Chikuni, 1989). Moreover, these fleets frequently lack automated data collection through digital systems and on-board monitoring, which impedes the characterization of their activity and impacts (Guyader et al., 2013). The mixed-species nature of these fisheries, coupled with their diverse environmental impacts, requires a shift in fisheries management from single-species approaches, like total allowable catches, to fleet-based strategies (EU, 2022; Ulrich et al., 2012).

Vessel monitoring systems and electronic logbooks provide crucial data for management including the location of fishing activities and the type of fishing gear employed at a haul level. In recent decades, the European Union (EU) has intensified its data collection efforts. Vessel monitoring systems (VMS) and electronic logbooks were introduced in 2004 and 2012 respectively, contributing to the implementation of fleet-based approaches to fisheries management (EU, 2009). However, only vessels above 15 m in length must report their activity through VMS and electronic logbooks. Vessels with a length of less than 12 m, or with a

length between 12 and 15 m if engaged in fishing activities lasting less than 24 hours, are exempt from this obligation (EU, 2023, 2009). This exception leaves most of the fleet without gear usage information, seriously compromising the estimation of gear-specific fishing effort and consequently, the management of these fisheries (EU, 2008).

The Portuguese multi-gear coastal fleet comprises around 500 vessels between nine- and 35-meters length, with a minimum engine power of 26 kilowatts. Of these, two-thirds are larger than 12 m. In 2016, this fleet employed 25 % of licensed fishers, landing over 200 species, accounting for 15 % of the mainland landings in weight and 30 % in value (DGRM, 2016). The most landed species (in weight) were the common octopus (*Octopus vulgaris*), followed by the black scabbardfish (*Aphanopus carbo*), the surf clam (*Spisula solida*), the European hake (*Merluccius merluccius*), and the swordfish (*Xiphias gladius*) (DGRM, 2016).

This fleet operates off the Portuguese continental coast (ICES Division IXa), covering an area of approximately 1.7 million square kilometers and comprising many ecosystems (Leitão et al., 2022). The fishing activity is concentrated over the continental shelf, until 200 m depth (Leitão et al., 2022), mostly between one and six nautical miles from the coast. The area between one nm and the coast is restricted to vessels under nine meters (local fleet). The area outside the six nautical miles limit is characterized by intense activity of the trawling fleet, which seriously compromises the setting of static gears (Campos et al.,

* Corresponding author at: IPMA, Instituto Português do Mar e da Atmosfera, Rua Alfredo Magalhães Ramalho 6, Algés 1495-165, Portugal.

E-mail address: pedro.leitao@ipma.pt (P. Leitão).

<https://doi.org/10.1016/j.fishres.2024.107199>

Received 14 November 2023; Received in revised form 2 October 2024; Accepted 2 October 2024

Available online 23 October 2024

0165-7836/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2021b). However, the multi-gear fleet uses fishing grounds inaccessible to trawling, such as deep canyons, seamounts, and rocky bottoms. Additionally, some vessels engage in deep-sea fisheries over the continental slope and in the northeast Atlantic (Campos et al., 2023, 2019; Farias et al., 2013; Santos et al., 2013).

The fleet adapts to regional and seasonal variations in resource availability by employing a diverse range of fishing gear. Primarily, these vessels operate static gear such as nets (gillnets and trammel nets), traps, and longlines (bottom and drifting longlines), but towed gears, such as bivalve dredges, are also used. The diversity of gears used allows the capture of a wide variety of benthic, demersal, and pelagic species. A vessel may use different gears over time, sometimes on the same trip (Cardoso et al., 2015; Szynaka et al., 2021, 2022). Technical interactions occur when multiple gears target a large group of species, posing a challenge to fisheries management (Cardoso et al., 2015; Duarte et al., 2009). Segmentation studies of the multi-gear fleet have been conducted based on daily vessel sales. However, the identification of fishing operations where a specific gear targets one or more species - referred to as a *métier* - has been limited at the national level. Exceptions of this are the drifting longline fishery targeting swordfish and the deep-water longline fishery for black scabbardfish, which involve distinct fleet segments using specific gear types (Campos et al., 2021a; Duarte et al., 2009). Another example is the bivalve fisheries using dredges (Campos et al., 2021a).

In contrast, studies conducted on the south coast of Portugal, based on landing profiles involving dredges, trammel nets, gillnets, and longlines, defined and validated 11 *métiers* through the use of fisher questionnaires and onboard observations (Szynaka et al., 2021, 2022). These studies showed seasonality for most of the *métiers* identified, as well as annual shifts, such as the change in target species from hake to anglerfish when using gillnets (Szynaka et al., 2021). An example of gear alternation can be seen in the swordfish fleet, which uses drifting longlines to target swordfish in the autumn and winter, and bottom longlines to target wreckfish and European conger in the spring and summer (Campos et al., 2019). Understanding the dynamics of the multi-gear coastal fleet, such as regional and seasonal changes, is crucial for sustainable management.

This study aimed to predict the main fishing gear used on each trip. For part of the multi-gear fleet using electronic logbooks, the gear used is often recorded, but this information is not available for the remaining vessels. Landing records are available for the whole fleet, including all species landed and their weight. The landing information (species, port of landing, and month of the year) was used to predict the main gear used on each trip using a machine learning model, the random forest (Breiman, 2001). After verifying the predictive accuracy of the random forest model, the classification model obtained from the fleet segment with known gear information was applied to fishing trips without gear information.

2. Material and methods

The main purpose of this study is to characterize the fishing operations of the multi-gear coastal fleet operating off the Portuguese mainland, namely, to obtain information on the gear used to assist in gear-based management. Data from 2012 to 2016 were used and the fishing trip associated with a landing event was considered as the unit of analysis. The fleet of interest includes vessels over 9 m, but the reporting of fishing operations in electronic logbooks (time and coordinates of the start and end of each haul, gear used and most important species caught) is only compulsory for vessels over 15 m or between 12 and 15 m when fishing for more than one day, which means that only 33 % of vessels have VMS and electronic logbooks on board. In addition, many vessels in the 12–15-meter segment, even if equipped with electronic devices, do not record daily trips.

Although there is an overlap in the size of vessels that undertake fishing trips with and without records in electronic logbooks, it is

reasonable to question whether classification rules based on larger vessels can be applied to smaller ones. In practice, since the most important variables used in gear prediction are the species landed, it is important to know whether species diversity is the same in both groups of vessels. Data to answer this question were specifically collected during this study, based on questionnaires in fishing ports. Differences between the species composition of vessels with and without electronic logbooks, by gear, were evaluated. Since the null hypothesis of no differences in species composition could not be rejected, the gear classification rules obtained with the random forest algorithm were applied to the trips without electronic logbooks to predict the gear used.

The data were analyzed at trip level, but the link between gear used and landings was not straightforward because more than one type of gear can be used on a fishing trip. In practice, a vessel may use two distinct types of static gear on the same trip, such as traps and gillnets, and the catch is landed together. The frequency of gear combinations for trips with electronic logbooks is shown in Appendix 1 and in 28 % of trips more than one gear were used. The option of basing the model only on trips with a single gear was considered undesirable as it would eliminate 28 % of the trips. In cases where more than one gear was used, the most important gear was selected based on the analysis of the catch of each haul and its contribution to the catch of the trip (by weight). The consequences of considering only one gear per trip were assessed by comparing the predictive ability of the random forest model on two sets of data: (1) trips using only one gear or (2) all trips, assigning the main gear used to each trip.

All vessels in this fleet use the same types of gear, but the length of the gear, such as the number of panels and pots, varies according to the size of the vessel. Portuguese regulations allow larger vessels to have up to three times more gear than smaller vessels (Diário da República, 2023). Before data analysis, catches were standardized by dividing the catch by the length of the vessel, thus removing the effect of vessel size (Le Pape, 2001; Leitão et al., 2022).

2.1. Datasets

This study used different datasets covering vessels of the coastal fleet, over nine meters in length, operating along the Portuguese continental coast between 2012 and 2016. All data on fishing activity and technical characteristics of the fleet were provided by the Directorate-General for Natural Resources, Safety and Maritime Services (DGRM) in an anonymized format, using vessel identifiers to link information across datasets. The following datasets were used in this study:

- Fleet records for all vessels, consisting of vessel identifiers, fishing licenses, and length. The total number of vessels for the period 2012–2016 was 497, ranging from 460 in 2012–444 in 2016. During this period, the number of vessels in each length category was 222, 123, and 152 for vessels between [9–12[meters, [12–15[meters, and ≥ 15 m, respectively.
- Electronic logbook records were available for a subset of vessels including vessel identifier, fishing trip identifier, port of return and timestamp of return, fishing gear used, and approximate weight of each species caught. The total number of vessels with electronic logbooks and corresponding landing records was 155, increasing from 62 in 2012–154 in 2016. The dataset contained a total of 279,712 observations (species by haul) corresponding to 45,470 fishing trips. From this database, only information on fishing gear was used in this study. A total of six gear types were considered (in parenthesis, the FAO gear classification, FAO, 2021): towed dredge (DRB), trap (FPO), gillnet (GNS), trammel net (GTR), drifting longline (LLD), and bottom longline (LLS).
- Landing records – consisting of landing port and species including the vessel identifier, landing date, trip unique identifier, landing port, common name, and weight of all species landed, available for 478 of the 497 vessels in the fleet. Each landing operation was

considered to correspond to one trip. The dataset contained a total of 1402,059 observations (species by trip). In this study species composition (in weight), month (12 months), and port of landing (22 ports) were used. Due to the considerable number of species involved (over 200), only the 45 most important species landed were considered. Campos et al., 2021a selected the 50 most important species in the landings of the coastal fleet, based on a combination of landed weight and landed value. This study included 45 of these species for which species identification was unambiguous.

2.2. Data merging

The correspondence between the two fisheries-dependent datasets (landing records and electronic logbooks) was obtained by matching the vessel identifier and the date and port of landing. We build the random forest algorithm with gear information from electronic logbooks and species catches from landing records. From the fleet records, vessel length was merged with the previous data and used to standardize the catches to kilograms per vessel length (kg m^{-1}).

The resulting dataset, called the 'landings dataset', brings together information on the main gear used (where available), the quantities of each species landed in kilograms per meter of a vessel, the date, and the port of landing. This dataset was split into two subsets: 45,470 trips from 155 vessels with gear information and 175,890 trips from 469 vessels without gear information. Of the 478 vessels whose trips were analyzed, nine always used the electronic logbook, 146 used the electronic logbook for part of the trips (and are repeated in both subsets of data) and 323 never used the electronic logbook.

2.3. Questionnaires at fishing ports

One of the methodological issues discussed earlier is the need to base the gear allocation of trips without electronic logbook information on trips with electronic logbooks using random forest. These two subgroups of trips are associated with vessels of different sizes, the smaller ones not being equipped with electronic devices. As the random forest model is essentially based on the species present in the landings, it was important to understand whether, for the same gear, the species composition of the landings was similar for vessels with and without electronic logbooks (the same as saying for vessels of 12–15 m and vessels of 9–12 m).

This information was supplemented by landings records (weight and value by species) provided by Docapesca, the state company responsible for the first sale of fishery products. The questionnaires were conducted between April and August 2023 in the ports of Aveiro (North), Peniche and Sesimbra (Centre), and Portimão (South), which are the main ports for the multi-gear fleet. Each questionnaire corresponded to a fishing trip and was completed after the vessel arrived in port, during the landing operations. The questionnaire was divided into two parts. The first collected interview identification, vessel size, and the availability of electronic logbooks. The second section collected information on fishing operations, including gear used and species associated with each gear. Fishing operations using bivalve dredges and drifting longlines were not included in the port samples, as they are associated with unique species landed, namely clams and swordfish, and represent specific métiers well identified in previous studies (Campos et al., 2019, 2021a; Oliveira et al., 2013; Santos et al., 2014).

2.4. Statistical analysis

2.4.1. Comparison of species composition for vessels with and without electronic logbooks based on port questionnaires

The species composition for each gear type was compared between vessels with and without electronic logbooks. For each trip, the gear information provided by the skipper was linked to the species caught as recorded in the corresponding landing records and standardized by vessel length (kg m^{-1}). If the skipper reported more than one gear type,

an expert-based rule was applied to assign the gear type to the trip based on the most important species landed in weight: gillnet was assigned if hake, horse mackerel or pouting were the main species; trap was assigned if octopus was the main species; and trammel net was assigned if cuttlefish, soles or anglerfish were the main species landed.

A Permutational Multivariate Analysis of Variance, the Adonis function (Anderson, 2001), was used to test for differences in species composition between trips with and without electronic logbooks within the same gear. This non-parametric statistical test does not assume normality and is used to compare differences in multivariate datasets (Anderson, 2001). It uses distance metrics and is particularly effective in analyzing complex data structures, making it ideal for ecological studies (Anderson, 2001; Anderson et al., 2008). Its permutation method and resistance to outliers compared to parametric tests make it a robust tool for testing (Anderson and Walsh, 2013). The input data consist of Bray–Curtis distance matrices and 999 permutations were used to obtain p-values (Anderson, 2001).

2.4.2. Random forest algorithm

The random forest algorithm is a supervised machine learning technique that classifies observations based on multiple variables (predictors) by randomly generating decision trees (Breiman, 2001). Trees have a pre-defined number of divisions (nodes), each based on the value of a single variable. The model is first applied to a set of observations of a known class (training phase). Hundreds of trees are generated, and bootstrapping is used on a subset of the variables to find the variable and value at each node. Once the "forest" is obtained, each observation is run through each tree and the assigned classification corresponds to the most frequent class obtained in all trees. The performance of the model is evaluated using various parameters related to the proportion of correctly classified observations. If the performance of the model is deemed satisfactory, the classification trees are then used to classify new observations with unknown class. In our case, the observations are the fishing trips, the predictor variables are the species caught (kg m^{-1}), the month and the port of landing, and the classification variable is the main gear used (a random tree of the model is presented in Appendix 3). The model was built and tested using trips with known gear from electronic logbooks. The random forest algorithm was then applied to the remaining trips.

Before building the random forest algorithm, several parameters must be determined, including the number of trees generated and the number of nodes (variables) in each tree. To choose the number of trees needed to build the model, we evaluated the accuracy from 100 to 500 trees (step of 50 trees). Although stability was achieved with 200 trees, the number 500 was chosen because it originated a slight increase in accuracy without a significant increase in computation time. The number of classification variables used in each simulation was nine, which is the square root of the total number of classification variables available. This is suggested in the literature and was found to be the most accurate for prediction (Hastie et al., 2009).

The landings dataset was divided into two groups of trips based on whether the gear used was known: L1 with the gear used reported and L2 without. To build the random forest algorithm, L1 was further divided into two subsets: L1-train (80 %, 36,376 trips) for model building and L1-test (20 %, 9094 trips) for cross-validation and evaluation. Using the L1-train dataset, 500 decision trees were generated. The random forest obtained by applying the random forest algorithm to the L1-train was then used in the L1-test to evaluate the predictive ability of the model. The outputs were: (1) global and individual gear classification error, known as out-of-bag error, the proportion of incorrectly assigned gears; (2) sensitivity, the proportion of true records correctly predicted; and (3) identification of the most important variables in the classification process (Genuer et al., 2010; Hong Han et al., 2016).

To determine whether the agreement between predicted and true gear could be justified by chance alone, we applied Cohen's kappa coefficient to the L1-test confusion matrix (Ben-David, 2008). This

coefficient provides an empirical indication of deviations from random classification, with high values associated with the better predictive ability of the random forest model (McHugh, 2012).

Information on the main gear used, obtained by the random forest algorithm, was analyzed using descriptive statistics to characterize the activity of the fleet.

2.4.3. Software used

All analyses were performed in R and implemented through RStudio (R Core Team, 2021). The libraries {dplyr} (Wickham et al., 2022) and {reshape2} (Wickham, 2017) were used to reshape and clean the data, {superml} (Saraswat, 2022) to find the best random forest parameters, {randomForest} (Liaw and Wiener, 2002) to build the predictive model, {parallelMap} (Bischl and Lang, 2016) to improve computational performance, {vcd} (Meyer D. et al., 2023) to compute Cohen's Kapp, and {ggplot2} (Wickham, 2009) for graphical representation. Finally, the packages {MASS} and {vegan} were used for nonparametric multidimensional scaling and analysis of variance (Oksanen et al., 2022; Venables and Ripley, 2002).

3. Results

3.1. Comparison of species composition for vessels with and without electronic logbooks based on port questionnaires

A total of 68 questionnaires were conducted to collect information on the fishing gear used and the corresponding catches per trip. These surveys were carried out in four major ports known for their significant quantities of fish landed: Aveiro (North), Peniche-Sesimbra (Centre) and Olhão (South), with 21, 40 and 7 questionnaires, respectively. The vessels surveyed included those with and without electronic logbooks. Of the four gear types considered, only FPO (traps) showed significant differences in species composition between the two groups of vessels (Table 1).

The significant differences for traps are due to confounding between vessel size and gear used. The gear type FPO covers a large group of gears that can be further subdivided into shelter traps and cage traps. Shelter traps catch only octopuses and are used on larger vessels, whereas cage traps, which target octopuses but also catch various fish and crustaceans, are used on smaller vessels (based on our interview data). In the questionnaires, shelter traps were most reported by vessels with electronic logbooks, while cages were more commonly reported by vessels without electronic logbooks. As this study does not distinguish between these two trap types, the significant difference in catch composition does not affect the prediction of trap type, which is based on the quantity of octopus landed, the dominant species for both trap types. Given the results obtained, it was considered possible to base the prediction of the gear on larger vessels.

Table 1
Permutational multivariate analysis of variance using distance matrices on standardized landed weight (kilograms per meter of vessel length). Each p-value is based on more than 999 permutations (the lowest p-value reported is 0.001).

Gear	p-value
FPO	0.001
GNS	0.067
GTR	0.130
LLS	0.933

3.2. The random forest model

3.2.1. Model evaluation

Table 2 shows the results of applying the random forest algorithm to the L1 test dataset in the form of a confusion matrix. The confusion matrix is a key tool in classification, providing a detailed breakdown of the actual versus predicted classifications made by the model. Correct membership (row labels) and random forest membership (column labels) can be used to assess model performance. Classification error (out-of-bag error) can be obtained by quantifying the proportion of trips that are incorrectly classified for each gear. The sensitivity represents the proportion of correctly classified trips, also known as the true positive rate.

The overall classification error was 14 %; the most accurate predictions were associated with dredges (DRB), drifting lines (LLD), and bottom longlines (LLS) with errors of 0 %, 1.1 %, and 2.8 % respectively, while the least accurate predictions were associated with gillnets (GNS) and trammel nets (GTR) with errors of 27.1 % and 22.7 % respectively. Traps (FPO) had a classification error of 7.1 %. Most of the misclassification errors concerned gillnets and trammel nets; 19 % of gillnet trips were misclassified as trammel nets and 14 % of trammel net trips were misclassified as gillnets. When both nets were considered together, 8 % were misclassified as traps. Cohen's kappa was 0.81, indicating good agreement between predicted and true values (McHugh, 2012).

The importance of variable classification is determined by its absolute and relative mean decrease in accuracy (aMDA and rMDA, respectively), which quantify the impact on predictive performance when the variable is removed (Table 3). When surf clam (*Spisula solida*) was used as a classification variable, trips were consistently classified as bivalve dredges. The next variable, swordfish (*Xiphias gladius*), is associated with drifting longlines. Another important variable is black scabbardfish (*Aphanopus carbo*), associated with a bottom longline fishery. The common octopus (*Octopus vulgaris*) is important when attributing trap as the main gear.

3.2.2. Effect of single gear assignment in trips using multiple gears

We compared the classification errors of single-gear trips with the results presented previously, based on all vessels with electronic logbooks and assigning the most important gear to trips that used multiple gears (Appendix 4). The overall out of bag error was higher for single gear trips (31 % compared to 14 %). For dredges (DRB), the error was 0 % as before. For traps (FPO), gillnets (GNS), and bottom longlines (LLS), the error was much higher at 25.7 % (compared to 6.6 %), 65 % (compared to 28.6 %) and 7.6 % (compared to 2.4 %), respectively. For drifting lines (LLD) the error was slightly higher, 1.7 % (compared to 0.4 %), and for trammel nets (GTR) the error was slightly lower, 18.8 % (compared to 24.3 %). Cohen's Kappa was 0.56 for the model using only single gear trips and 0.81 when considering all trips. Given that the errors were higher for the single-gear trips only, we believe that assigning the most important gear in multigear trips did not affect the predictive ability of the random forest model.

3.2.3. Gear prediction using random forest

After considering a 14 % error as acceptable, the classification rules obtained with L1 were applied to the L2 dataset to obtain information for the main gear in all trips. The results were used to quantify the use of fishing licenses (Table 4). Almost all vessels licensed for trammel nets used them, as well as, but to a lesser extent, bivalve dredges, gillnets, and traps. Longlines were the least used main type of gear, used by only half of the vessels with this license.

The number and percentage of trips (observed and predicted) were estimated for each gear. Traps represent almost half of the total number of trips, followed by trammel nets. Gillnets and bottom longlines were found to be much less used. Finally, drifting longlines are the least used gear type.

Fig. 1 shows the number of trips for each gear type for the L1 and L2

Table 2

Confusion matrix from the prediction of the L1-test subset with 9094 fishing trips together with the percentage of classification error (Class. Error) and sensitivity (proportion correctly classified). The overall classification error was 14.0 %. The blue cells represent the number of correctly classified trips. Cohen’s kappa was 0.81.

	Number of trips with observed gears						Total	Class. Error (%)	Sensitivity
	DRB	FPO	GNS	GTR	LLD	LLS			
DRB – bivalve towed dredges	120	0	0	0	0	0	120	0	1
FPO – traps	0	2534	63	126	0	6	2729	7.1	0.93
GNS – gillnets	0	125	1240	327	0	10	1702	27.1	0.73
GTR – trammel nets	0	223	332	1898	0	3	2456	22.7	0.77
LLD – drifting longlines	0	0	0	1	94	0	95	1.1	0.99
LLS – bottom longlines	0	16	31	8	1	1936	1992	2.8	0.97

Table 3

The mean decrease in accuracy of the cross-validation of the top ten variables with at least 100 trips shows the absolute (aMDA) and relative (rMDA) importance of each variable in predicting the gear using (L1-train). The aMDA value represents the number of trips misidentified when the variable is removed, and rMDA is its relative percentage. Higher values indicate greater importance in the classification process. Variables are in descending order of rMDA.

Common name	Number of trips	aMDA (num)	rMDA (%)
Surf clam	417	389	93
Swordfish	579	288	50
Black scabbardfish	6641	2312	35
Shortfin mako shark	630	186	29
Common octopus	17,568	4738	27
Olhão port	703	190	27
Sesimbra port	9243	1547	17
Lowfin gulper shark	2109	309	15
Blackbellied angler	3494	446	13
Silver scabbardfish	342	45	13

datasets combined. The number of trips using bivalve dredges increased by over 80 % and the use of bottom longlines decreased by 30 %. The use of drifting longlines (LLD), and traps (FPO) present a similar trend, decreasing in 2015 and increasing in 2016. The number of trips using gillnets (GNS), and trammel nets (GTR) varied over time, but the high classification error associated with these gears makes it difficult to suggest a real trend.

Fig. 2 shows the species caught by different types of fishing gear. Trips with electronic logbooks and without are plotted separately.

The diversity of species landed varies according to the type of fishing gear employed. In bivalve dredge fishing trips (DRB), only one species was reported, namely the surf clam (ULO – *Spisula solida*). In drifting longline (LLD) six species were reported. In contrast, in fishing trips

Table 4

Number of licensed vessels for each gear type in the coastal multi-gear fleet, together with the number of vessels and fishing trips using each gear type (observed: L1 dataset and predicted: L2 dataset). DRB: bivalve dredge; FPO: trap; GNS: gillnet; GTR: trammel net; LLD: drifting longline; LLS: bottom longline. Vessels may hold licenses for more than one type of gear.

	DRB	FPO	GNS	GTR	LLD	LLS	
Number of vessels with license	52	389	320	341	60	434	Number of vessels
Number of vessels with observed gear (with e-logs, L1 dataset)	2	85	97	101	24	48	155
Number of vessels with predicted gear (without e-logs, L2 dataset)	47	344	286	333	29	164	469
Total number of vessels (observed + predicted gear)*	47	348	289	334	41	174	478
Percentage of vessels using licensed gear	90.4	89.5	90.3	97.9	68.3	40.1	Number of trips
Number of trips with observed gear (e-logs, L1 dataset)	537	13,480	8489	12,270	586	10,108	45,470
Number of trips with predicted gear (without e-logs, L2 dataset)	7767	81,054	21,700	54,920	221	10,228	175,890
Total number of trips (with observed + predicted gear)	8304	94,534	30,189	67,190	807	20,336	221,360
Percentage of trips	3.7	42.7	13.6	30.4	0.4	9.2	

* The total number of vessels does not correspond to the sum of the values in the two previous rows because many vessels are repeated in L1 and L2 because they do not use the electronic logbook on every trip.

employing gillnets, trammel nets, and traps over fifteen species were reported. The proportion of fishing trips that employ electronic logbooks varies among fishing gears and among species within each gear. The level of availability of electronic logbooks for bivalve dredges (DRB) and traps (FPO) is low. Conversely, most drifting longlines (LLD) are associated with trips with electronic logbook data. Finally, bottom longline (LLS) has an overall low coverage with electronic logbooks but a high coverage of trips reporting black scabbardfish (BSF – *Aphanopus carbo*), lowfin gulper shark (CPL – *Centrophorus lusitanicus*), and smooth-hound (SMD – *Mustelus mustelus*). This can be attributed to the larger size of the vessels involved in this fishery, which was identified previously as a specific fleet component.

4. Discussion

This study looks at the fishing gear used by the multi-gear fleet on the Portuguese coast. Knowing which gears are used is the first step towards managing this fleet more efficiently. Gear information, combined with georeferenced data (location of the fishing grounds), knowledge of the species caught, and the impacts caused by each gear (Fennell et al., 2021; Grabowski et al., 2014) will allow gear-based management, which is fundamental for an ecosystem-based management approach (Chuenpagdee et al., 2003). In Europe, such management policies are important to comply with by-catch reduction and the landing obligation (Prellezo and Villasante, 2023). Gear information, combined with economic data on the value of the landings and operation costs of each gear, can be used to predict the consequences of forbidding or reducing the use of a given gear in space or time or introducing gear substitutions (Chuenpagdee et al., 2003). At present, considering this approach is not possible because most of the vessels in this fleet (about 2/3), do not have electronic logbooks; for these vessels, only landing records are available and thus it is not possible to know which fishing gear is being used.

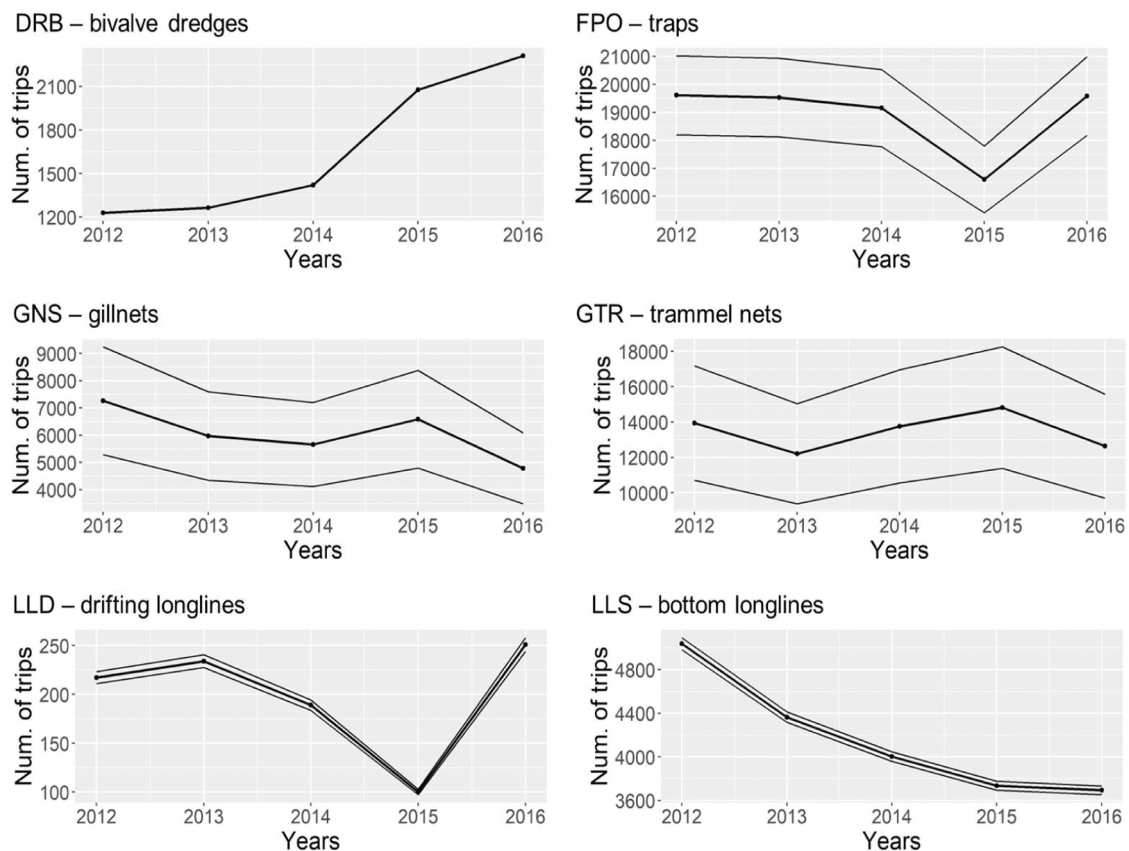


Fig. 1. Number of trips from 2012 to 2016 for each gear type for datasets L1 and L2 combined. The upper and lower lines show the standard deviation. Bivalve dredge (DRB) has no error lines since the prediction error was 0 %.

Previous studies have addressed fleet segmentation in this fleet based on landed species composition and vessel characteristics (Campos et al., 2021a; Duarte et al., 2009). Duarte et al. (2009) used a non-hierarchical clustering approach, partition around medoids (PAM), and multivariate regression trees (MRT) to identify variables important in separating groups of vessels, by region. Regression trees were better for grouping vessels based on variables like vessel size, gear, or port of landing. The vessel groups could not be associated with specific gears or target species. In the study by Campos et al. (2021a), another non-hierarchical classification technique, CLARA (Clustering LARge Applications), was used (Castro et al., 2010; Castro et al., 2007). Fourteen landing profiles were found, associated with target species, nine of which were classified as strong, based on a quality index. Gear was not considered in this study because the electronic logbooks were unavailable. Recently, Szyńska et al. (2021) used CLARA and MRT to identify eleven métiers from fishing trips in the south of Portugal, and ten of these were later validated with surveys (Szyńska et al., 2022).

The segmentation of fishing trips undertaken by multi-gear fleets, based on landing records and classification techniques, has been employed in other regions in southern Europe. González-Álvarez et al. (2016), in Asturias, Spain, used an integrated method of analysis combining administrative data on fishing licenses, daily sales statistics, and information from surveys, to identify twenty-one métiers with special relevance of the trammel net (targeting *Lophius spp.*), the *Merluccius* longline and the *Merluccius* gillnet métiers. In Italy, Russo et al. (2016), used Self Organizing Maps, a machine learning technique, in conjunction with VMS data and logbook data, to identify specific métiers. This represents a significant advancement in the identification of fishing gears and métiers utilized in the Mediterranean region. It diverges from the previous studies in that it is an unsupervised method that employs unlabeled data to infer métiers (Russo et al., 2016). Nine

landing profiles associated with métiers were identified. In Mallorca, Spain, Palmer et al. (2017) employed a methodology based on the input of experts to identify seven métiers. Subsequently, seventeen binary classification algorithms were applied to the data with the k-nearest neighbor classifier (IBk) being considered as the most optimal, based on agreement with expert classification.

This study proposes the use of random forest, a supervised classification algorithm, to predict the gear used based on landings information such as species composition, month, and port. The availability of complete species records only at the trip level determined the adoption of the fishing trip as the study unit instead of the haul. The model was tested on a set of fishing trips for which the gear was known (the training phase), and the results were applied to all other trips. This model was chosen due to its robustness to noise and its non-parametric approach to classification (Hamza and Larocque, 2005). Machine learning classification techniques using multiple classifiers are considered better than the ones using single classifiers but have an increased risk of overfitting (Kotsiantis et al., 2006). The random forest solves the overfitting problem by selecting a distinct set of classifiers in each tree, chosen with bootstrap sampling from a subset of the available classifiers (Zhu, 2020). Although not commonly used in fisheries, the random forest demonstrated to be an important tool for predicting fishing events (Behivoke et al., 2021), identifying fishing points and quantifying fishing efforts (Rufino et al., 2023), and detecting fishing operations from surveillance data (Meeanan et al., 2023).

The random forest algorithm had an overall classification error of 14 % when classifying trips in six gear types (in parenthesis, the FAO gear classification, FAO, 2021): towed dredge (DRB), trap (FPO), gillnet (GNS), trammel net (GTR), drifting longline (LLD), and bottom longline (LLS). Cohen's kappa was 0.81, indicating good agreement between predicted and true values (McHugh, 2012). Two gear types, the gillnets

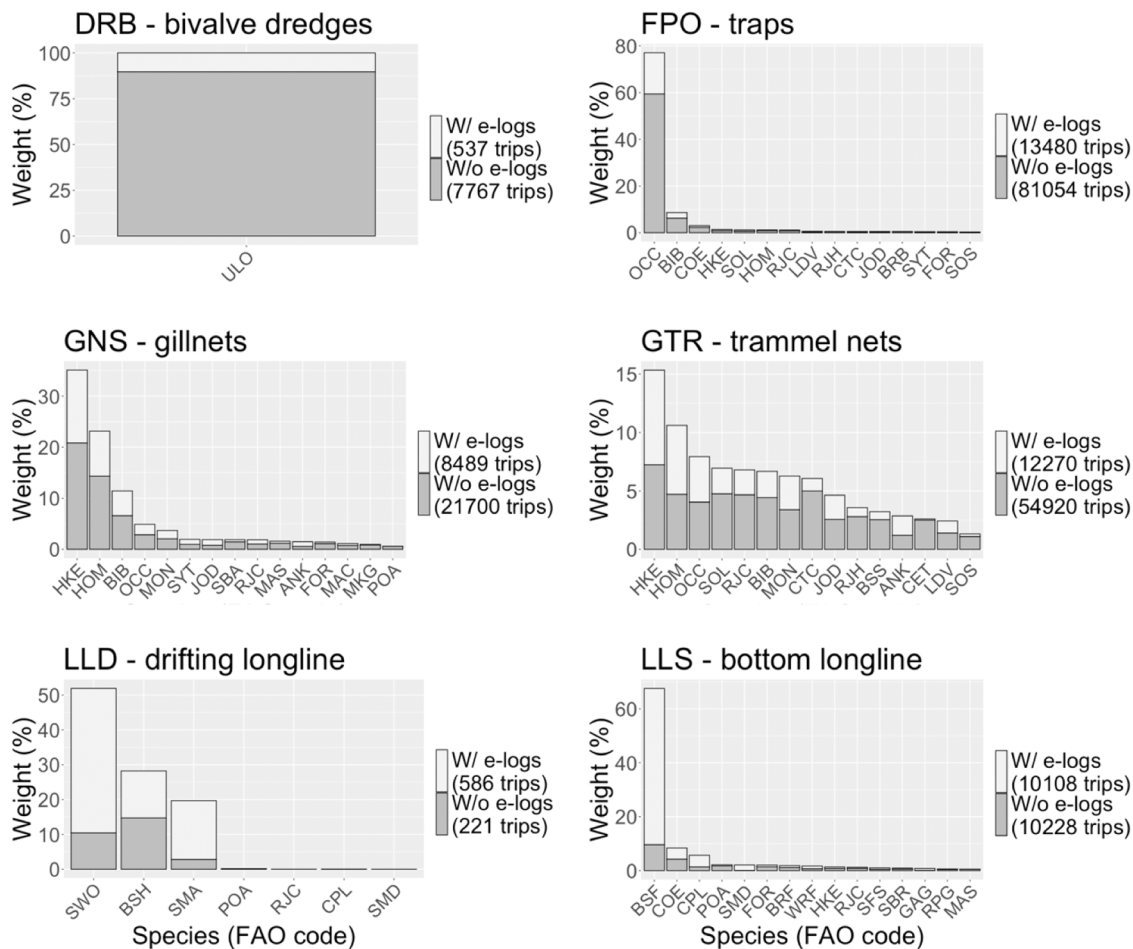


Fig. 2. Species composition in percentage of landed weight by gear type. Only the most important fifteen species for each gear are represented. Landings with gear information (w/ e-logs) are in white and landings without gear information (w/o e-logs; gear predicted) are in grey. Species 3-digit codes in order of appearance: ULO – surf clam; OCC – common octopus; BIB – pouting; COE – European conger; HKE – European hake; SOL – common sole; RJC thornback ray; HOM – Atlantic horse mackerel; LDV – Large-scaled gurnard; RJH – Blonde ray; JOD – John dory; BRB – Black seabream; CTC – Common cuttlefish; SYT – Nursehound; FOR – Forkbeard; SOS – Sand sole; MON – Angler; SBA – Axillary seabream; MAS – Chub mackerel; ANK – Blackbellied angler; MAC – Atlantic mackerel; MKG – Thickback sole; MUR – Surmullet; BSS – European seabass; CET – Wedge sole; SWO – Swordfish; BSH – Blue shark; SMA – Shortfin mako; POA – Atlantic pomfret; CPL – Lowfin gulper shark; SMD – Smooth-hound; BSF – Black scabbardfish; BRF – Blackbelly rosefish; WRF – Atlantic wreckfish; SFS – silver scabbardfish; SBR – Blackspot seabream; GAG – Tope shark; RPG – Red porgy.

and trammel nets, had the highest errors, respectively 27.1 % and 22.7 %. The source of higher errors is associated with three gears, traps, gillnets, and trammel nets. Misclassifications associated with these gears represent 94 % of all errors. The misclassification of gillnets as trammel nets and vice versa is the most common classification error representing 52 % of the errors. This may be related to more than one gear being used in the same fishing trip since in 7 % of the trips trammel nets and gillnets were used together. Misclassification involving traps and nets accounts for 42 % of the errors. The confusion among these gears comes from overlap in the most important species landed. As an example, hake and horse mackerel are the most abundant species for both gillnets and trammel nets and hake is also the fourth most important species for traps. Another important species for classification, the black-bellied angler, is found in both gillnets and trammel nets. The common octopus, which is the most significant species for traps, ranks as the fourth most important species for gillnets. Three fishing gear types were associated with small classification errors: the bivalve dredges, bottom, and drifting longlines.

To apply the random forest algorithm, it was necessary to assume that vessels only operate a single gear per fishing trip. In 28 % of the trips analyzed, more than one gear was employed. In such instances, the gear assigned to the trip was that which constituted most of the catch, as

determined by logbook data. The decision to build the model using all trips for which gear information was available was based on two arguments: (1) it was considered undesirable to exclude 28 % of the trips, and (2) it was assumed that multiple gears should be present in the trips to which the model was applied. This approach ensures consistency between the data used in building the model and the data used in its application. However, the accuracy of the classification was evaluated by comparing classification errors based on all trips with classification errors based on trips involving a single gear. The overall out-of-bag error was significantly higher for single-gear trips (31.2 % compared to 14.6 %). Cohen’s Kappa was 0.56 for the model utilizing solely single-gear trips, in comparison to 0.81. For dredges (DRB) the error was 0 % as before. For traps (FPO) and gillnets (GNS) the error was much higher, when using single-gear trips, respectively 25.7 % (compared with 6.6 %), 65 % (compared with 28.6 %). For bottom longline (LLS) and drifting longlines (LLD) the error was slightly higher, respectively 7.6 % (compared with 2.8 %) and 1.7 % (compared with 0.4 %), and regarding trammel nets (GTR) the error was slightly lower, 18.8 % (compared with 22.7 %). The reason for higher errors in traps and gillnets, when training the model with single-gear trips, is not clear. The errors were expected to be similar. One possibility is that electronic logbooks are not completed correctly. Often the data is entered at the

end of the trip and the only information considered important is the declared catch, which can be verified by port authorities on landing and subject to fines if misreported. Little importance is given to the gear used. Based on these results, we concluded that considering the main gear for trips with multiple gear was acceptable.

Bivalve dredges and bottom longline fishing trips correspond to previously identified *métiers* belonging to the bivalve dredge fishery and black scabbard fishery respectively (Campos et al., 2021a). Both the bivalve dredges and the black scabbard longlines are highly selective gears, thus justifying a low classification error. The dredges catch a small group of bivalve species and operate off Aveiro, Matosinhos, Olhão, and Setúbal (Oliveira et al., 2010), while the deep-water longline targets black scabbardfish along the continental slope, mainly off Sesimbra (Bordalo-Machado and Figueiredo, 2009; Farias et al., 2013). The drifting longline fishery, targeting swordfish in open sea areas and seamounts (Santos et al., 2013), was identified in previous studies (Campos et al., 2023, 2019) corresponding to a seasonal fishery practiced by a group of longliners that can also operate bottom longlines.

The errors for the different gears agree with the important variables for classification found in random forest. These variables are strongly associated with gears with low classification error. Considering the important species for each gear from landing records, surf clam is associated with dredges, swordfish, shortfin mako shark, lowfin gulper shark with drifting longlines, black scabbardfish and silver scabbardfish with bottom longlines, common octopus with traps, and to lesser extent gillnets, blackbellied angler is associated with gillnets and trammel nets. The ports of Olhão and Sesimbra are associated, respectively, with the landing of octopus and scabbardfish, contributing to the classification of trips as traps or bottom longline. Similarly, Duarte et al., (2009) also found that landing port was among the most important variables to predict gear.

Fishing gear characteristics of the multi-gear fleet affect species composition of caught species, thus if the gear information were available at a higher level of detail, such as mesh size or type of trap, misclassification could be reduced. The same applies to trammel nets and traps, with the second and third higher classification errors, respectively. In this study, *métiers* were only classified according to the type of gear used (level 4 according to FAO) due to the lack of gear characteristics such as mesh size and hook number, impacting fleet-based management. In gillnets and trammel nets, more than one mesh size is known to be used depending on whether the targeted species are small, such as soles and cuttlefish, or large, such as anglers and skates (Duarte et al., 2009; Moura et al., 2016; Szynaka et al., 2021). Since this information is not available, the definition of *métiers* to a higher level requires the implementation of nationwide questionnaires to fishers. This was the case of the studies carried out by Castro et al. (2021) through interviews to skippers of the Portuguese small-scale multi-gear fleet, and Szynaka et al. (2021) with a similar approach in the multi-gear coastal fleet operating off the south coast.

Almost all vessels in this fleet are licensed for more than one gear type, with more than thirty different license combinations (DGRM, 2016). In the trips analyzed in this work and considering the gear declarations in electronic logbooks, an average of 3.3 gears were used per vessel. However, owning these licenses does not imply their effective use. In this study, information was obtained on the most important gears used by vessels with multiple licenses and how their usage evolved. Since the number of gear licenses attributed to each vessel is mostly constant, changes in the number of gears used indicate gear preferences. It was demonstrated that not all vessels use the gears they are licensed for, like the studies by Duarte et al. (2009) and Szynaka et al. (2022). While most vessels used trammel nets, bivalve dredges, gillnets, or traps, longlines were much less used. This was more evident for bottom longlines, used by only one-third of the vessels holding this license. Gear preferences changed over time; while fishing trips using bivalve dredges increased over the study period, those using bottom longlines decreased. We hypothesize that the low-risk, steady income of the bivalve dredges

fishery made this gear attractive, with multiple vessels changing their activity towards this fishery. The fishing gear shifts between 2014 and 2015 were further explored during this study by questioning fishers operating traps and drifting longlines. Two changes were registered during this period: first, the price of the main target species in traps - octopus - dropped in 2015. This made fishers switch to trammel nets or stop fishing during this year. Secondly, regarding drifting longlines, a new law entered into force in 2015 restricting the use of this fishing gear. A minimum weight of fifteen kilograms for swordfish landed was added and the transaction of individual quotas between vessels was facilitated (Portuguese legislation *Diário da República*, 2014). Vessels that had a quota for swordfish but did not want to use it could sell their quotas to other vessels. We hypothesized that this reduced the size of the fleet operating drifting longlines in 2015 and thus the number of trips.

We found that catch species composition, complemented with information on the port of landing (an indicator of the location of fishing grounds) can be used to predict fishing gear when this information is available for part of the fishing trips. Such a methodology is important because it allows gear-based management. The results obtained in this study also show that information from the fleet component with electronic logbooks can be generalized to the rest of the fleet. This approach can be applied before electronic logbooks are introduced in smaller vessels, which will be mandatory in the EU fleet up to 2030 (European Commission, 2023). Further improvements to minimize misclassification of trips can be achieved by exploring modifications to the random forest base model (Modak et al., 2024; Zhou and Mentch, 2023).

5. Conclusions and recommendations

In this study, a random forest model based on landing records was employed to predict the fishing gear utilized in thousands of fishing trips, with a high level of accuracy. The selectivity of the various fishing gears and, consequently, the species composition of the catches they yield varies considerably. Therefore, information on the gear used is of foremost importance for fisheries advice. Nets have a higher catch rate but are less selective, resulting in a greater number of species caught and higher discards and the catch is often in poor condition. Hooks are more selective and result in fewer discards and much of the catch is of good quality. These nuances demonstrate that the impact of gear varies and highlight the importance of understanding the specific fishing gear utilized in each trip. The deployment of this model has contributed to the understanding of the fishing gears employed and the quantification of the fishing effort within the multi-gear fishing fleet. Furthermore, the utilization of electronic logbooks was pivotal; however, it is recommended that additional gear-related information, such as mesh and hook sizes, traps, and longline types, should be registered in electronic logbooks to facilitate the identification of higher-level *métiers* and enhance the quality of data for fisheries advice and conservation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The authors do not have permission to share data.

Acknowledgments

This study was carried out under the scope of the project TECPESCAS – Project Mar 2020 16–01–04-FMP-0010–IPMA.

This study received Portuguese national funds from FCT - Foundation for Science and Technology through projects UIDB/04326/2020 (DOI: 10.54499/UIDB/04326/2020) and LA/P/0101/2020 (DOI: 10.54499/

LA/P/0101/2020) and the scholarship 2022.11214.BDANA.

The fisheries data were supplied to IPMA by the Portuguese Directorate-General for Natural Resources, Safety and Maritime Services (DGRM) and DOCAPESCA. The authors would like to thank the contributions of three anonymous reviewers who significantly helped to improve this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.fishres.2024.107199](https://doi.org/10.1016/j.fishres.2024.107199).

References

- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>.
- Anderson, M.J., Walsh, D.C.I., 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol. Monogr.* 83, 557–574. <https://doi.org/10.1890/12-2010.1>.
- Anderson, M.J., Gorley, R.N., Robert Clarke, K., 2008. PERMANOVA+ for primer: Guide to software and statistical methods. PRIMER-E, Plymouth.
- Behivoke, F., Etienne, M.-P., Guitton, J., Randriatsara, R.M., Ranaivoson, E., Léopold, M., 2021. Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests. *Ecol. Indic.* 123, 107321. <https://doi.org/10.1016/j.ecolind.2020.107321>.
- Ben-David, A., 2008. Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Syst. Appl.* 34, 825–832. <https://doi.org/10.1016/j.eswa.2006.10.022>.
- Bischi, B., Lang, M., 2016. parallelMap: Unified Interface to Parallelization Back-Ends. Bordalo-Machado, P., Figueiredo, I., 2009. The fishery for black scabbardfish (*Aphanopus carbo* Lowe, 1839) in the Portuguese continental slope. *Rev. Fish. Biol. Fish.* 19, 49–67. <https://doi.org/10.1007/s11160-008-9089-7>.
- Breiman, L., 2001. Random Forest. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Campos, A., Lopes, P., Fonseca, P., Figueiredo, I., Henriques, V., Gouveia, N., Delgado, J., Gouveia, L., Amorim, A., Araújo, G., Drago, T., dos Santos, A., 2019. Portuguese fisheries in seamounts of Madeira-Tore (NE Atlantic). *Mar. Policy* 99, 50–57. <https://doi.org/10.1016/j.marpol.2018.10.005>.
- Campos, A., Fonseca, P., Araújo, G., Parente, J., 2021a. Definition of landing profiles in the Portuguese coastal multi-gear fleet. In: *Developments in Maritime Technology and Engineering*. CRC Press, London, pp. 693–697. <https://doi.org/10.1201/9781003216599-73>.
- Campos, A., Henriques, V., Erzini, K., Castro, M., 2021b. Deep-sea trawling off the Portuguese continental coast—spatial patterns, target species and impact of a prospective EU-level ban. *Mar. Policy* 128, 104466. <https://doi.org/10.1016/j.marpol.2021.104466>.
- Campos, A., Leitão, P., Sousa, L., Gaspar, P., Henriques, V., 2023. Spatial patterns of fishing activity inside the Gorringe bank MPA based on VMS, AIS and e-logbooks data. *Mar. Policy* 147, 105356. <https://doi.org/10.1016/j.marpol.2022.105356>.
- Cardoso, I., Moura, T., Mendes, H., Silva, C., Azevedo, M., 2015. An ecosystem approach to mixed fisheries: technical and biological interactions in the Portuguese multi-gear fleet. *ICES J. Mar. Sci.* 72, 2618–2626. <https://doi.org/10.1093/icesjms/fsv138>.
- Castro, J., Punzón, A., Pierce, G.J., Marín, M., Abad, E., 2010. Identification of métiers of the Northern Spanish coastal bottom pair trawl fleet by using the partitioning method CLARA. *Fish. Res.* 102, 184–190. <https://doi.org/10.1016/j.fishres.2009.11.011>.
- Castro, N., Romão, F., Castro, J.J., Pereira, T.J., Pedro, S., Viegas, V., Costa, J.L., 2021. Catches, sales and discards: Small-scale fisheries in a Portuguese Marine park. *Reg. Stud. Mar. Sci.* 42, 101643. <https://doi.org/10.1016/j.rsma.2021.101643>.
- Chuenpagdee, R., Morgan, L.E., Maxwell, S.M., Norse, E.A., Pauly, D., 2003. Assessing collateral impacts of fishing methods in US waters. *Front. Ecol. Environ.* 1, 517–524. [https://doi.org/10.1890/1540-9295\(2003\)001\[0517:SGACIO\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2003)001[0517:SGACIO]2.0.CO;2).
- Castro, J., Abad, E., Artetxe, I., Cardador, F., Duarte, R., García, D., Hernández, C., Marín, M., Murta, A., Punzón, A., Quincoces, I., Santurtún, M., Silva, C. and Silva, L. Castro, J., et al., 2007. Identification and Segmentation of Mixed-Species Fisheries Operating in the Atlantic Iberian Peninsula Waters. IBERMIX Project, final report (Contract Ref.: FISH/2004/03-33). European Commission - Directorate General for Fisheries and maritime Affairs 1–260.
- DGRM, 2016. Multi-gear fisheries data from 2012 to 2016. Daily sales, electronic logbooks, vessels licenses and characteristics delivered on request [dataset]. Diário da República, 2014. Portaria n.º 119/2014 de 3 de junho. Diário da República 3050–3052.
- Diário da República, 2023. Portaria n.º 227/2023 do Ministério da Agricultura e Alimentação, Diário da República: I série, n.º 141.
- Duarte, R., Azevedo, M., Afonso-Dias, M., 2009. Segmentation and fishery characteristics of the mixed-species multi-gear Portuguese fleet. *ICES J. Mar. Sci.* 66, 594–606. <https://doi.org/10.1093/icesjms/bsp019>.
- EU, 2008. Commission Decision of 6 November 2008.
- EU, 2009. Council Regulation (EC) No 1224/2009. Establishing a Community control system for ensuring compliance with the rules of the common fisheries policy. Official Journal of the European Union L343, 1–50.
- EU, 2022. The implementation of ecosystem-based approaches applied to fisheries management under the CFP. European Commission.
- EU, 2023. Regulation (EU) 2023/2842 of the European Parliament and of the Council of 22 November 2023 amending Council Regulation (EC) No 1224/2009, and amending Council Regulations (EC) No 1967/2006 and (EC) No 1005/2008 and Regulations (EU) 2016/1139, (EU) 2017/2 2842, 1–105.
- European Commission, 2023. Regulation (EU) 2023/2842 of the European Parliament and of the Council of 22 November 2023 amending Council Regulation (EC) No 1224/2009, and amending Council Regulations (EC) No 1967/2006 and (EC) No 1005/2008 and Regulations (EU) 2016/1139, (EU) 2017/2403 and (EU) 2019/473 of the European Parliament and of the Council as regards fisheries control.
- FAO, 2021. Classification and illustrated definition of fishing gears, Classification and illustrated definition of fishing gears. <https://doi.org/10.4060/cb4966en>.
- FAO, 2022. The State of World Fisheries and Aquaculture 2022. The State of World Fisheries and Aquaculture 2022. FAO.
- Farias, I., Morales-Nin, B., Lorance, P., Figueiredo, I., 2013. Black scabbardfish, *Aphanopus carbo*, in the northeast Atlantic: distribution and hypothetical migratory cycle. *Aquat. Living Resour.* 26, 333–342. <https://doi.org/10.1051/alr/2013061>.
- Fennell, H., Sciberras, M., Hiddink, J.G., Kaiser, M.J., Gilman, E., Donnan, D., Crawford, R., 2021. Exploring the relationship between static fishing gear, fishing effort, and benthic biodiversity: a systematic review protocol. *Environ. Evid.* 10, 27. <https://doi.org/10.1186/s13750-021-00242-y>.
- Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2010. Variable selection using random forests. *Pattern Recognit. Lett.* 31, 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>.
- González-Álvarez, J., García-de-la-Fuente, L., García-Flórez, L., Fernández-Rueda, M. del P., Luis Alcázar-Álvarez, J., 2016. Identification and Characterization of Métiers in Multi-Species Artisanal Fisheries. A Case Study in Northwest Spain. *Nat. Resour.* 07, 295–314. <https://doi.org/10.4236/nr.2016.76026>.
- Grabowski, J.H., Bachman, M., Demarest, C., Eyras, S., Harris, B.P., Malkoski, V., Packer, D., Stevenson, D., 2014. Assessing the Vulnerability of Marine Benthos to Fishing Gear Impacts. *Rev. Fish. Sci. Aquac.* 22, 142–155. <https://doi.org/10.1080/10641262.2013.846292>.
- Guyader, O., Berthou, P., Koutsikopoulos, C., Alban, F., Demanèche, S., Gaspar, M.B., Eschbaum, R., Fahy, E., Tully, O., Reynal, L., Curtil, O., Frangoudes, K., Maynou, F., 2013. Small scale fisheries in Europe: A comparative analysis based on a selection of case studies. *Fish. Res.* 140, 1–13. <https://doi.org/10.1016/j.fishres.2012.11.008>.
- Hamza, M., Larocque, D., 2005. An empirical comparison of ensemble methods based on classification trees. *J. Stat. Comput. Simul.* 75, 629–643. <https://doi.org/10.1080/00949650410001729472>.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *Springer Series in Statistics The Elements of Statistical Learning - Data Mining, Inference, and Prediction*. Springer.
- Hong Han, Xiaoliang Guo, Hua Yu, 2016. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest, in: 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE, pp. 219–224. <https://doi.org/10.1109/ICSESS.2016.7883053>.
- Kotsiantis, S.B., Zaharakis, I.B., Pintelas, P.E., 2006. Machine learning: a review of classification and combining techniques. *Artif. Intell. Rev.* 26, 159–190. <https://doi.org/10.1007/s10462-007-9052-3>.
- Le Pape, O., 2001. The influence of vessel size and fishing strategy on the fishing effort for multispecies fisheries in northwestern France. *ICES J. Mar. Sci.* 58, 1232–1242. <https://doi.org/10.1006/jmsc.2001.1121>.
- Leitão, P., Sousa, L., Castro, M., Campos, A., 2022. Time and spatial trends in landing per unit of effort as support to fisheries management in a multi-gear coastal fishery. *PLoS One* 17, e0258630. <https://doi.org/10.1371/journal.pone.0258630>.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest.
- McHugh, M.L., 2012. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 22, 276–282.
- Meenan, C., Noranartragoon, P., Sinanun, P., Takahashi, Y., Kaewnern, M., Matsushita, T.F., 2023. Estimation of the spatiotemporal distribution of fish and fishing grounds from surveillance information using machine learning: The case of short mackerel (*Rastrelliger brachysoma*) in the Andaman Sea, Thailand. *Reg. Stud. Mar. Sci.* 62, 102914. <https://doi.org/10.1016/j.rsma.2023.102914>.
- Meyer D., Zeileis A., Hornik K., Friendly M., 2023. vcd: Visualizing Categorical Data.
- Modak, M., Jadhav, M., Kadam, R., Arwade, M., Mane, S., 2024. Improving Diagnostic Accuracy: A Deep Dive into Random Forest Optimization for Clinical Data, in: 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT). IEEE, pp. 1–6. <https://doi.org/10.1109/AIIoT58432.2024.10574720>.
- Moura, T., Fernandes, A., Alpoim, R., Azevedo, M., 2016. Unravelling the dynamics of a multi-gear fleet – Inputs for fisheries assessment and management under the Common Fisheries Policy. *Mar. Policy* 72, 219–230. <https://doi.org/10.1016/j.marpol.2016.07.015>.
- Nagasaki, F., Chikuni, S., 1989. Management of multispecies resources and multi-gear fisheries. FAO Fisheries Technical Paper. FAO Fisheries and Aquaculture Department.
- Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solyomos, P., Stevens, M.H.H., Szocs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlenn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F., Weedon, J., 2022. *vegan: Community Ecology Package*.
- Oliveira, M.M., Camanho, A.S., Gaspar, M.B., 2010. Technical and economic efficiency analysis of the Portuguese artisanal dredge fleet. *ICES J. Mar. Sci.* 67, 1811–1821. <https://doi.org/10.1093/icesjms/fdq110>.

- Oliveira, M.M., Camanho, A.S., Gaspar, M.B., 2013. The influence of catch quotas on the productivity of the Portuguese bivalve dredge fleet. *ICES J. Mar. Sci.* 70, 1378–1388. <https://doi.org/10.1093/icesjms/fst098>.
- Palmer, M., Tolosa, B., Grau, A.M., Gil, M. del M., Obregón, C., Morales-Nin, B., 2017. Combining sale records of landings and fishers knowledge for predicting métiers in a small-scale, multi-gear, multispecies fishery. *Fish. Res.* 195, 59–70. <https://doi.org/10.1016/j.fishres.2017.07.001>.
- Prellezo, R., Villasante, S., 2023. Economic and social impacts of the landing obligation of the European Common Fisheries Policy: a review. *Mar. Policy* 148, 105437. <https://doi.org/10.1016/j.marpol.2022.105437>.
- R Core Team, 2021. R: A language and environment for statistical computing.
- Rufino, M.M., Mendo, T., Samarão, J., Gaspar, M.B., 2023. Estimating fishing effort in small-scale fisheries using high-resolution spatio-temporal tracking data (an implementation framework illustrated with case studies from Portugal). *Ecol. Indic.* 154, 110628. <https://doi.org/10.1016/j.ecolind.2023.110628>.
- Russo, T., Carpentieri, P., Fiorentino, F., Arneri, E., Scardi, M., Cioffi, A., Cataudella, S., 2016. Modeling landings profiles of fishing vessels: an application of Self-Organizing Maps to VMS and logbook data. *Fish. Res.* 181, 34–47. <https://doi.org/10.1016/j.fishres.2016.04.005>.
- Santos, M.N., Coelho, R., Lino, P.G., 2013. Standardized CPUE for swordfish (*Xiphias gladius*) caught by the portuguese pelagic longline fishery in the north. *Atlantic* 70, 1783–1791.
- Santos, M.N., Coelho, R., Lino, P.G., 2014. STANDARDIZED CPUE FOR SWORDFISH (*Xiphias gladius*) CAUGHT BY THE PORTUGUESE PELAGIC LONGLINE FISHERY IN THE NORTH ATLANTIC. *Collect. Vol. Sci. Pap. ICCAT*.
- Saraswat, M., 2022. superml: Build Machine Learning Models Like Using Python's Scikit-Learn Library in R.
- Szynaka, M.J., Erzini, K., Gonçalves, J.M.S., Campos, A., 2021. Identifying métiers using landings profiles: an octopus-driven multi-gear coastal fleet. *J. Mar. Sci. Eng.* 9, 1022. <https://doi.org/10.3390/jmse9091022>.
- Szynaka, M.J., Fernandes, M., Anjos, M., Erzini, K., Gonçalves, J.M.S., Campos, A., 2022. Fishers, Let Us Talk: Validating Métiers in a Multi-Gear Coastal Fishing Fleet. *Fishes* 7. <https://doi.org/10.3390/fishes7040174>.
- Ulrich, C., Wilson, D.C.K., Nielsen, J.R., Bastardie, F., Reeves, S.A., Andersen, B.S., Eigaard, O.R., 2012. Challenges and opportunities for fleet- and métier-based approaches for fisheries management under the European Common Fishery Policy. *Ocean Coast Manag* 70, 38–47. <https://doi.org/10.1016/j.ocecoaman.2012.06.002>.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*, Fourth. ed. Springer, New York.
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*. Media. Springer.
- Wickham, H., 2017. tidy: Easily Tidy Data with “spread(.)” and “gather(.)” Functions. R package version 0.6.1. <https://CRAN.R-project.org/package=tidy>.
- Wickham, H., François, R., Henry, L., Müller, K., 2022. dplyr: A Grammar of Data Manipulation.
- Zhou, S., Mentch, L., 2023. Trees, forests, chickens, and eggs: when and why to prune trees in a random forest. *Stat. Anal. Data Min.: ASA Data Sci. J.* 16, 45–64. <https://doi.org/10.1002/sam.11594>.
- Zhu, T., 2020. Analysis on the Applicability of the Random Forest. *J. Phys. Conf. Ser.* 1607, 012123. <https://doi.org/10.1088/1742-6596/1607/1/012123>.