Mediterranean Marine Science

Vol 23, No 4 (2022)

VOL 23, No 4 (2022)

Fishing the waves: comparing GAMs and random forest to evaluate the effect of changing marine conditions on the energy performance of vessels

ALESSANDRO COLOMBELLI, JACOPO PULCINELLA, SARA BONANOMI, EMILIO NOTTI, FABRIZIO MORO, ANTONELLO SALA

doi: 10.12681/mms.29721

Research Article

Mediterranean Marine Science
Indexed in WoS (Web of Science, ISI Thomson) and SCOPUS
The journal is available on line at http://www.medit-mar-sc.net
www.hcmr.gr
DOI: http://doi.org/10.12681/mms.29721

Fishing the waves: comparing GAMs and random forest to evaluate the effect of changing marine conditions on the energy performance of vessels

Alessandro COLOMBELLI, Jacopo PULCINELLA, Sara BONANOMI, Emilio NOTTI, Fabrizio MORO and Antonello SALA

CNR-IRBIM, National Research Council - Institute of Marine Biological Resources and Biotechnologies, Ancona, Italy

Corresponding author: Alessandro COLOMBELLI; alessandro.colombelli@irbim.cnr.it

Contributing Editor: Stelios SOMARAKIS

Received: 23 February 2022; 22 August 2022; Published online: 18 November 2022

Abstract

The optimization of consumption and the reduction of gas emissions in fisheries rely on a thorough examination of all factors affecting the energy balance of fishing vessels. Engines, propellers, or the hydrodynamic characteristics of vessels and gears are unquestionably the primary factors affecting this balance, and an improvement in energy efficiency based on these factors is typically attained through technical modifications to existing technologies. Behavioral modifications, such as a reduction in operational speeds or the selection of closer fishing grounds, are another option. There may still be room for improvement in behavioral responses, for instance by adapting fishing strategies in response to changing weather and sea conditions. As far as the authors are aware, the influence of varying sea state and wind conditions on the energy expenditure of fishing vessels has not yet been investigated and is the focus of this research. In this study, wind and wave actions were associated with the observed activity of three fishing vessels operating in the northern Adriatic Sea: an OTB, a PTM, and a TBB trawler. The analysis made use of a comparison between two different approaches, generalized additive models (GAMs) and random forest, in order to quantify the significance of each variable on the response and generate consumption forecasts. In our analysis, the observed influence of predictors was significant albeit occasionally ambiguous. Wave height had the most obvious impact on energy expenditure, with the towing and gear handling phases being the most affected by wave action. Conversely, navigation seemed to be mostly unaffected by significant wave heights up to 1.5 meters, with unclear effects on consumption estimated above this threshold. The relationship between winds and fuel consumption was found to be nonlinear and ambiguous; hence, its significance should be investigated further.

Keywords: Fisheries; bottom trawling; fuel consumption; Adriatic Sea; GAM; random forest.

Introduction

Typically, fishing is an energy-intensive activity with high rates of greenhouse gas emissions, the bulk of which are produced directly from the consumption of non-renewable energy sources. The amount of fuel required in fisheries varies significantly between and within fleets (Parker & Tyedmers, 2015; Parker *et al.*, 2018; Pelletier *et al.*, 2011). Both the economic costs to fishermen and the environmental effects of excessive greenhouse gas emissions from energy-intensive operations are a cause for concern. Outdated technology is often the main cause of high energy costs (Buglioni *et al.*, 2011), and its replacement through the purchase of new vessels is hindered by European Commission restrictions for fishing capacity control (EU, 2013) and prohibitive vessel pricing. The re-adaptation of existing technologies, namely the adoption of minor technical adjustments to reduce energy costs, is currently the most feasible solution.

Over the past few decades, different methods have been explored, such as the optimization of hull geometry and propulsion systems (Gabiña *et al.*, 2016; Lin *et al.*, 2018; Notti & Sala, 2012; Schau *et al.*, 2009) and fouling control on vessels' hulls to diminish ship resistance (Notti *et al.*, 2019). Trawl gear performance and modifications have also been investigated (Thierry *et al.*, 2020a; Wileman, 1984; Wileman & Hansen, 1988). Different trawl designs, larger meshes, or alternative netting materials (Thierry *et al.*, 2020b; Parente *et al.*, 2008; Sala *et al.*, 2008; Verhulst & Jochems, 1993) have been shown to be effective in reducing fuel consumption while maintaining a gear's ability to catch target species. In addition, behavioral modifications, such as the modulation of towing speed in response to rising fuel costs (Poos *et al.*, 2013), lower steaming speeds (Sala *et al.*, 2011), or the selection of closer fishing grounds (Bastardie *et al.*, 2010; Sampson, 1991) have also been considered. Although it is now evident that consumption increases exponentially with speed (Corbett *et al*., 2009; Ronen, 1982) and that it primarily depends on the hull's hydrodynamic characteristics, engine power, and gear drag, secondary factors such as bad weather and rough seas could still contribute, albeit to a lesser extent, to the fluctuation of operating costs (Schau *et al*., 2009; Thrane, 2004). Moreover, power requirements can be influenced by weather conditions in certain vessel operational modes (Swider *et al*., 2019; Swider & Pedersen, 2019). The interaction between fishing vessels and changing environmental conditions, such as the seakeeping performance of vessels in waves (Tello *et al*., 2011), vessels' stability and hull design (Neves *et al*., 2003, 1999), or stability variations owing to the combined action of waves and fishing gear pull (Mantari *et al*., 2009) have all been studied in the past.

Here, we investigate whether changes in sea state and wind conditions affect the fishing activity and energy expenditure of individual vessels, and we attempt to quantify the effects of these environmental variables. To do so, we analyzed the activities of three different fishing vessels that are representative of three very common *metiérs* in Mediterranean fisheries (Santiago *et al*., 2015; Tsagarakis *et al*., 2017): i) a pelagic midwater pair trawler, ii) a beam trawler, and iii) a bottom otter trawler. We used an accurate description of their consumption profiles associated with the different phases of their fishing activities, and we created predictive models to examine the importance of predictors on the response variable and to forecast consumption for each vessel. We acknowledge that there are some limitations to this approach, such as the regional scope of this study. The geographic configuration of the northern Adriatic Sea makes it unlikely that fishing vessels will frequently operate in extreme conditions, as is the case for fisheries in the oceans or the North Sea. Nonetheless, local fleets are geared to a specific range of working conditions and a nebulous limit determines the suspension of fishing activities, even for larger vessels. When sea state conditions are on the verge of this limit, it is the skippers who decide whether it is advisable to set a course or not. Above and beyond adverse working conditions, potential differences in catches, and the probable variation in market prices as a result, it would be useful to consider potential changes in energy requirements due to unfavorable weather. Our modeling approach is based on a comparison of two different statistical methods: generalized additive models (GAMs) (Hastie & Tibshirani, 1986) and random forest (Breiman, 2001; Ho, 1995). GAMs allow for the definition of a nonlinear function between each predictive variable and the response. Automatic modeling of nonlinear relationships has the potential to give accurate estimations of the variable of interest. In addition, their additive nature enables them to examine each variable's effect on the response while holding the other variables constant (James *et al*., 2013). These aspects make GAMs an optimal choice for inference, as they offer greater flexibility than standard linear models while retaining a reasonable degree of interpretability. The downside of this approach is GAM's additive structure, which may constitute a limitation in terms of interactions between predictive variables (Faraway, 2016; James *et al*., 2013). Even if these can be manually incorporated into the model, important interactions may still be overlooked. To overcome these drawbacks, we conducted a comparison with a random forest model, a machine learning approach that ensures greater flexibility at the expense of some model interpretability. Comparisons between these two methods already exist in the literature, and they have been applied successfully in different fields (González-Irusta *et al*., 2015; Liu *et al*., 2018; Marmion *et al*., 2008). This comparison between modeling methodologies was devised to highlight the importance ascribed to predictors by each model and to examine the contribution of different variables to the observed variability in consumption. We also looked into the models' forecasting capabilities by evaluating their predictive performance. Both models were tested on a test dataset, and their estimates were compared to the actual consumption values in our records.

Materials and Methods

Data Collection

In this study, we collected operational data characterizing the activities of three fishing vessels operating in the Ancona harbour. The documented fishing activity took place entirely in the central-northern area of the Adriatic Sea, within FAO GSA 17 (Fig. 1). Except for the OTB vessel, which also fished in the 100-200m depth layer, all vessels carried out their activity within the 0-50 m and the 50-100 m depth strata. The variables recorded during the data collection process include vessel ID, date, time, latitude, longitude, vessel speed, course, and fuel consumption.

Due to disparities between vessels in terms of dataset size and varied time scales that do not entirely overlap, we randomly picked one hundred fishing trips for each vessel, sampling the observations from 2011 to 2016, excluding fishing ban periods. Bans in the region usually take effect between August and September, with start dates, end dates, and duration that vary annually. Between 2011 and 2016, the average length of the ban period was 46 days, with a maximum of 60 days in 2011 and a minimum of 42 days in 2014 and 2016.

The selected vessels were chosen from a pool of monitored ships participating in a larger data collection activity conducted by the National Research Council (CNR), which was designed to gather data on the energy efficiency of fishing vessels and led to the production of other scientific contributions over the years (Buglioni *et al*., 2011; Sala *et al*., 2022; 2019; 2011). Since these specific vessels had the longest and most continuous time series, they were ideal for matching with several years of environmental data from the Copernicus platform. The length of the time series was one of the decisive factors in our selection process in order to ensure that a broad range of operational conditions was matched with the widest possible range of weather and sea conditions.

All vessels belong to the medium-large category of the

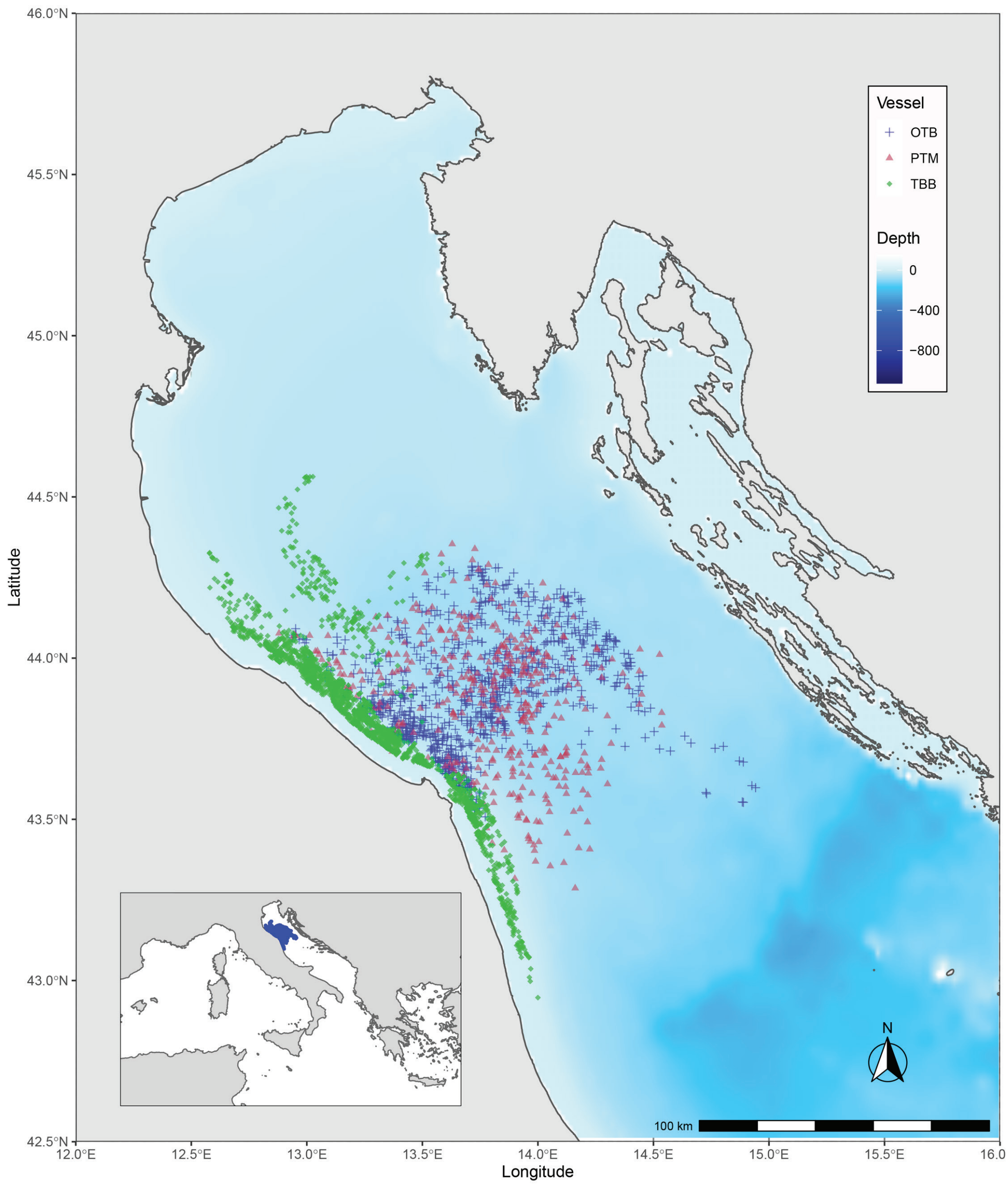

Fig. 1: Spatial distribution of haul midpoints for PTM, OTB, and TBB vessels.

Italian fleet and represent three very common *metiérs* in Mediterranean fisheries: a bottom otter trawler (OTB), a pelagic midwater pair trawler (PTM), and a beam trawler (TBB). The OTB has a LOA of 21.5 m, a breadth extreme of 5.7 m, a gross registered tonnage of 82 GT, and an engine power of 478 kW. This vessel hauls a single cone-shaped net that remains in constant contact with the seabed. The horizontal opening of the trawl is maintained by means of otter boards, while floats placed on the headline and a weighted footrope are used for the vertical opening. This type of trawl is mainly used to capture a wide variety of benthic and demersal species.

The PTM is the largest of the vessels under observation, with a hull of 29.0 m LOA, a breadth extreme of 6.9 m, a gross registered tonnage of 138 GT, and an engine power of 940 kW. It functions in tandem with another vessel with similar technical features. During fishing operations, the pair navigate in tandem, while jointly dragging a large semi-pelagic trawl. This fishing technique primarily targets small pelagic species such as sardines (*Sardina pilchardus*) and anchovies (*Engraulis encrasicolus*).

The TBB uses a "*Rapido"* type of net, a distinctive Italian trawl with a fixed opening that remains in constant contact with the seabed. The opening of the trawl is secured by a metal frame with teeth, which is designed to dig into the seabed in order to retrieve and capture the target species, which are mainly bivalves (*Pecten jacobaeus*) or flatfish species (*Solea spp., Psetta maxima*, and *Scophthalmus rhombus*). Up to four *Rapido* trawls are simultaneously hauled during towing operations. With a hull of 23.3 m LOA, a breadth extreme of 7.0 m, a gross registered tonnage of 86 GT, and an engine power of 780 kW, the TBB vessel is the middle ground between the previous ships in terms of size and engine power.

The PTM and the OTB vessels' fuel consumption were measured with a Race Technology Ltd GPS datalogger coupled with two flowmeters (Hendress & Hauser) and a multichannel recorder. For the TBB vessel, a GPS and flowmeters were used for data acquisition, and data were collected by a core control unit based on Raspberry technology that is capable of handling signal inputs from different data sources, storing information in a local micro-SD, and creating an automatic data transfer procedure. Data acquisition software consisted of: a GPS module, which managed spatial data and synchronized data from different sources (system time, coordinates, course, consumption data, etc.); a fuel consumption module (KRAL module) for the acquisition of fuel flow data from the sensors; a communication module for the internet connection for data transfer; and a MySQL module, which managed the connection between the software and a local MySQL database. The data collection systems used were designed and developed by CNR-IRBIM Ancona (Italy). The information collected for each vessel was organized in a matrix whose rows represent a sequential series of "pings," or observations concerning time, latitude, longitude, route, and navigation speed, completed by each vessel's consumption. Vessels in the dataset are identified by means of a unique vessel identifier. The temporal resolution of the collected observations is of one ping/min.

Environmental variables

The selected environmental variables included wind and sea state conditions. We also considered seasonality as a categorical variable added to the dataset and extrapolated from each record's date. This variable was introduced in order to account for any potential seasonal pattern that may eventually exist within the dataset. Sea bottom depth was also taken into account, as shorter cables and shallower waters reduce drag, often allowing vessels to navigate slightly faster while fishing. Wind conditions were reported using reprocessed six-hourly wind observations from 2012 to 2016 that were downloaded from the Copernicus Marine Service (product ID WIND_GLO_WIND_L4_REP_OBSERVATIONS_012_006). The analyzed variables included Wind Speed (ms^{-1}) and Wind U and V components (ms^{-1}), transformed into a unique wind direction (degrees) using the rWind R package (Fernández-López & Schliep, 2018). The dataset has a horizontal grid size of 0.25°.

Sea state conditions were described using reprocessed one-hourly wave observations, characterized by a horizontal grid size of 0.42°. Observations were downloaded from the Copernicus Marine Service, product ID MEDSEA_HINDCAST_WAV_006_012, (Zacharioudaki *et al*., 2019). The variables considered in this case were Sea Surface Significant Wave Height (m) and Sea Surface Wave from Direction (VMDR), expressed in degrees. Using the recorded vessel's heading, two additional variables were taken into account: the difference (in degrees) between the direction of travel and the respective wind and wave directions. This information was summarized by two categorical variables added to the analysis. Winds were classified as "headwinds," "tailwinds," or "crosswinds" depending on whether they were coming from within a ±15° interval from the vessels' heading, a ±15° interval from the stern, or from the side. The same classification was applied to waves, categorizing their direction as "head wave," "tail wave," or "cross wave". Bathymetric data was included using the ETOPO1 dataset (Amante & Eakins, 2009) hosted on the NOAA servers, and handled with the R-package marmap. The spatial analysis of downloaded datasets was performed in R, while NetCDF-4 and shapefiles were processed using the R-packages raster (Hijmans & van Etten, 2014) and sf (Pebesma, 2018), respectively.

Data analysis and fishing activity classification

R version 4.1.1 (R Core Team, 2017) was used for all data analyses. Pings containing improbable values, potential erroneous entries, and/or outliers were removed from the dataset as part of an initial data cleansing process. Fishing trips were identified using the pings corresponding to the beginning and end of a trip, that is, the observations falling within a buffer of half a nautical mile from the harbor. These values correspond to the entry and exit times of a vessel from the harbor.

Within the context of this study, a typical fishing trip can be divided into a succession of phases that have varying effects on the observed consumption. These phases can be summarized briefly as follows: high-speed movements towards and away from the fishing grounds; maneuvering phases of fishing gear set, gear handling, and recovery; the actual trawling phase (haul); and shorter, moderate-speed displacements often observed within the fishing grounds between one haul and another (searching). To determine this activity classification, we analyzed vessel speed as a time series of consecutive observations with a frequency of one ping per minute. The raw data were subjected to a noise reduction process to eliminate any potential noise and facilitate the activity's classification. Raw time series data were smoothed using a moving average based on a centered two-sided rolling window function, namely a rolling window with a width of n=3, which groups each observation in the series with its preceding and leading observations. A simple arithmetic average was then applied to the observations within the rolling window. This process produces a smoothed version of the original time series. Time series clustering, that is, the partition of time series data into phases based on distance or similarity, was performed using Euclidean distance metric to calculate a distance matrix and by then applying a hierarchical average-linkage clustering technique to define clustering structures. The optimal number of clusters was determined using the *pamk()* function on the calculated distance matrix. The function is part of the R-package fpc (Henning, 2015) and operates a partitioning around medoids clustering estimating the number of clusters by optimum average silhouette width. Finally, a small number of misclassified pings were corrected by applying filters to the clustering result. These erroneous entries mostly resulted from observations with ambiguous speed values that placed a ping halfway between two contiguous activity phases.

Generalized Additive Models (GAMs)

Generalized additive models (GAMs) (Hastie & Tibshirani, 1986) provide a useful way to extend a standard multiple linear regression model by incorporating nonlinear functions of the predicting variables while maintaining additivity (James *et al*., 2013). Nonlinear relationships between each predictor and the response are modeled by substituting the linear components $\beta_i x_i$ of a standard linear model with a smooth nonlinear function $f_i(x_i)$ (James *et al*., 2013). Additivity is then granted by calculating the singular contribution of each variable and adding them together to predict the response. The relationship between the combination of predictors and the response is established by a link function, and the basic form of the model can be found in (Hastie & Tibshirani, 1990):

$$g(\mu) = \beta_0 + \sum_{i=1}^{p} f_i(x_i)$$

where $g(\mu)$ represents the link function, β_0 is an unknown constant, and f_i is a non-parametric smoothing function of the continuous non-linear variable x_i. GAM models provide greater flexibility than standard linear models while retaining interpretability. The smoothing functions f_i in the model can be plotted to observe the marginal relationship between each predictor and the response (Faraway, 2016). In this research, the model was trained using the R package *mgcv* (Wood, 2018). Following a random 80/20 dataset split, only 80% of the dataset was used to train our models. The remaining 20% of the dataset was used to test the accuracy of predictions. Singular variables and interactions among predictions were both evaluated. Model selection was performed using a backward stepwise selection based on Akaike's Information Criterion (AIC) (Akaike, 1998). An Anova test of significance was used to evaluate interactions between predictors. Furthermore, the maximum degrees of freedom of the smoothing functions (number of knots *k*) for single variable smoothers as well as for interactions was limited to $k = 5$ to avoid overfitting. A summary of the variables used in the model can be found in Table 1.

Random Forest

Random forest is a widely-used ensemble learning method for classification and regression that is applied in several real-world applications and research projects in different domains (Oshiro *et al*., 2012). The method relies on a combination of tree predictors, where each tree is a classification or regression model. Individual trees are weak learners characterized by low bias and high variance. Bagging, which involves bootstrap sampling many random subsets from the training dataset, building a separate model for each subset, and averaging the resulting predictions, can reduce prediction variance and increase model accuracy (James *et al*., 2013). Two-thirds of the observations are used by bagged trees for training, while the remaining third, known as out-of-bag (OOB) observations can be used to observe prediction accuracy. The response variable predicted by the model for a specific observation equals the average of all estimates returned by the trees trained without using that observa-

Table 1. Summary of the variables used in the GAM and RF models.

	Variable	Variable type	Unit of measure - variable description
Vessel variables	Vessel_ID	numeric	-
	lat	numeric	Decimal degrees
	lon	numeric	Decimal degrees
	speed	numeric	knots
	Acceleration	numeric	knots
	Activity	categorical	levels: Towing, Steaming, Searching, Maneuvering
	FC	numeric	l/h
Environmental variables	Season	categorical	levels: Fall, Winter, Spring, Summer
	depth	numeric	meters
	Significant wave height	numeric	meters
	Wave direction	numeric	degrees
	Wind speed	numeric	ms-1
	Wind direction	numeric	degrees
Calculated/added variables	Crosswind	categorical	levels: Headwind,Tailwind, Crosswind
	Crosswave	categorical	levels: Headwave, Tailwave, Crosswave

tion. Random forest algorithms also use a small tweak to decorrelate the trees, that is, at each split in the trees, they select a random sample of n predictors from among all predictors. This prevents strong predictors from always being used at the top split of each tree, generating a forest of very similar trees (James *et al*., 2013). Our RF model was trained with the R-package randomForest (Liaw & Wiener, 2002), which implements Breiman's original random forest algorithm (Breiman, 2001; Breiman *et al.*, 1984). Although the RF function's default values deliver reliable estimates, the RF algorithm was fine-tuned in an effort to improve performance. The evaluation of the optimal hyper-parameter configuration could be computationally expensive and very long to execute; consequently, the fine-tuning of the algorithm was attempted using a random sample of the original dataset (approximately 10% of the training dataset). Although the adopted procedure was not optimum, it may nonetheless generate rough estimates that can be used as general guidelines. We attempted to optimize three parameters: the number of randomly sampled variables at each split; the maximum number of terminal nodes; and the final number of trees generated. Standard 10-fold cross validation was applied to optimize parameters. While understanding a single tree is straightforward, an ensemble of trees greatly reduces model interpretability, making it harder to determine the importance of individual variables. Nevertheless, the importance of different predictors can be observed either as: 1) a mean drop in the accuracy of predictions for out-of-bag samples; or, 2) a measure of the total decrease in node impurity. Here, we employed the first approach, addressing the importance of individual variables as the RSS decrease caused by splits over a given predictor. A variable importance score was computed as the variable's mean importance across all trees. Typically, the different RF implementations display the normalized importance of the variables, that is, each score is divided by its standard error. Here, we considered the unscaled version of variable importance, as recent studies (Díaz-Uriarte & Alvarez de Andrés, 2006; Strobl *et al*., 2008) have suggested that raw importance has better statistical properties.

Prediction

Using the remaining 20% of the dataset, which included 66,678 observations, we examined the capacity of GAM and RF to predict the amount of fuel consumed by vessels. The models' predictions were compared with observed real consumption values, and model accuracy was determined by computing a simulated R^2 coefficient, which is useful for evaluating the models' performance and roughly describing how well predicted values fit observed values. As an additional measure of model fit, we compared the Root Mean Squared Error (RMSE) computed by both models on training data (train RMSE) with the RMSE computed on previously unseen data (test RMSE).

Results

The dataset used contained over 333,386 entries (one-minute observations) documenting the activity of three fishing vessels between 2011 and 2016. The clustering process divided the activity of the PTM and OTB vessels into four phases: into gear-handling moments (maneuvering); short within-fishing ground vessel displacements (searching); high-speed navigation from and to fishing grounds (steaming); and the actual trawling (towing). Pings were categorized, that is, they were assigned to one of these four phases. The activities of the TBB vessel, on the other hand, were categorized into only three phases, excluding vessel displacements between hauls. During its fishing activity, the moment of gear retrieval and the moment of gear shooting for the next haul occur in rapid succession, with no discernible intervening phases. The average values of consumption and operating speeds representative of the different phases have been highlighted for each vessel. Table 2 provides a summary of these observed values, sorted by vessel.

Fishing activity in the northern-central Adriatic Sea took place at an average latitude of 43°.85 (range 42.93°, 44.63°), an average longitude of 13.57° (range 12.56°, 15.25°), and a depth range of 0 to -107 m with a mean value of -42 ±18 meters. The OTB vessel fished in the deepest waters (up to -107m, mean -54±21m), while the PTM and TBB vessels reached more modest depths of -87m (mean -53m, ±23m) and -76m (mean -18m ±12m), respectively. Within this spatial and temporal range, the Significant Wave Heights observed ranged from 0 to about 2.3 m, with an average value of 0.42m ± 0.34. About 90% of the activity fell within the 0-1 m range, with only 10% of the data (27,282 observations) exceeding this limit. No activity was recorded for Significant Wave Heights exceeding 2.34 m. The wind speed and direction variables mainly reported south-southeast winds with speeds ranging between 0.03 to 13.34 ms^{-1} (0.05 – 25.93 kn), with an average of 3.31 ± 1.96 ms^{-1} (6.43 ± 3.8 kn). Figure 2 provides a brief summary of the data obtained for each vessel and the environmental variables explored in our analysis.

GAM

The final GAM model was fit on the 80% of the observations randomly sampled from the dataset, and the accuracy of its predictions was tested using the remaining 20% of the dataset. The final model contained 14 variables with two interaction terms as described in the formula:

FC~ Vessel + te(speed, acceleration, by=Activity) + te(speed, depth, by=Activity) + s(acceleration, by=Activity) + s(speed, by=Activity) + s(depth, by=Activity) + s(W.Height, by=Activity) + s(W.Direction, by=Activity) + s(ws, by=Activity) + s(wd, by=Activity) + s(lat, by=Activity) + s(lon, by Activity) + Crosswind + Crosswave + Season + family=gaussian(link=log), data=train).

Table 2. Summary of values summarizing the each vessel's fishing activities.

Vessels	PTM	OTB	TBB
Maneuvering%	16.52%	11.46%	20.45%
Searching%	6.31%	6.59%	-
Steaming %	47.31%	9.2%	8.81%
Towing %	29.86%	72.74%	70.74%
Mean Maneuvering speed (kn)	1.47 (± 0.89)	2.68 (± 1.1)	5.05 (± 2.01)
Mean Searching speed (kn)	6.85 (± 1.32)	5.89 (± 1.31)	-
Mean Steaming speed (kn)	9.3 (± 0.75)	9.99 (± 0.47)	9.79 (± 1.38)
Mean Towing speed (kn)	4.28 (± 0.39)	3.89 (± 0.29)	6.9 (± 0.20)
Mean Maneuvering FC/hour (liters)	18 (± 9.09)	15.2 (± 8.51)	57.8 (± 32.2)
Mean Searching FC/hour (liters)	45.7 (± 30.6)	29.7 (± 20.5)	-
Mean Steaming FC/hour (liters)	70.8 (± 20.5)	57.8 (± 4.87)	100.0 (± 33.3)
Mean Towing FC/Hour (liters)	119.0 (± 6.42)	58.5 (± 4.34)	127.0 (± 15.9)

The model was able to explain about 88.1% of the observed total deviance, and the Anova $\chi 2$ test indicated that all terms included were significant. Only two interaction terms were retained, represented by the tensor product *te()* in the formula. These terms accounted for the interactions between the variables speed and acceleration and speed and depth. Interactions between the remaining variables were tested and found to be non-significant; therefore, they were omitted from the model. The effect of the *Vessel* and *Season* categorical variables is included in the model's parametric components. Also included among the parametric terms are the three-level categorical variables representing the difference between the heading of vessels and the direction of waves and winds. Smooth functions were used to describe the principal effects of continuous variables. The term "by=Activity" was included in the definition of smoothing terms within the GAM model formula. This element produces a replicate of the smooth for each categorical level to account for the marked differences in consumption between the various phases of fishing activity. A summary of the model outcome is depicted in Table 3, while the estimated influence of interactions and single predictors on consumption is shown in Figure 3 for the PTM vessel and in the supplementary material for the OTB and TBB vessels. The interaction between vessel speed and acceleration, considered by activity, is explored in Figure 3a. As expected, vessel *speed* was the variable that had the greatest influence on the amount of fuel consumed (Figs 3a and c). Acceleration had nonlinear effects on consumption across all activities (Figs 3a and d). In general, the model predicted minor energy expenditure associated with minor variations in speed and slower speeds. The interaction between *speed* and *depth* is explored in Figure 3b. The model suggested minor or no substantial differences in consumption associated with bathymetric shifts during high-speed navigation phases (mostly steaming). During this activity, changes in consumption are almost linearly correlated solely with the vessels' speed. In contrast, for manoeuvring and towing phases, increased depth contributed to an evident increase in PTM and OTB vessel consumption. In general, the model indicated higher energy requirements at greater depths (Figs 3b and 3e; Supplementary material Figs S1b and S1e) and faster trawling speeds, showing that this is the least energy-efficient combination. With regard to the TBB vessel (Supplementary material Figs S2b and S2e), the effect of depth on its consumption resulted in a nonlinear trend. In this case, the model predicted an increase in consumption up to depths of about 40-50 meters, followed by a decrease for trawling operations conducted at greater depths. It should be noted that only ~ 3-4% of the activity took place at depths greater than 50 meters, with almost 80% taking place within 10-30 meters and about 10% occurring within the 40-50 meter range. This is due to the characteristic behaviour of *Rapido* trawlers, which closely follow bathymetric bands during trawling operations.

Changes in Significant Wave Height had a positive

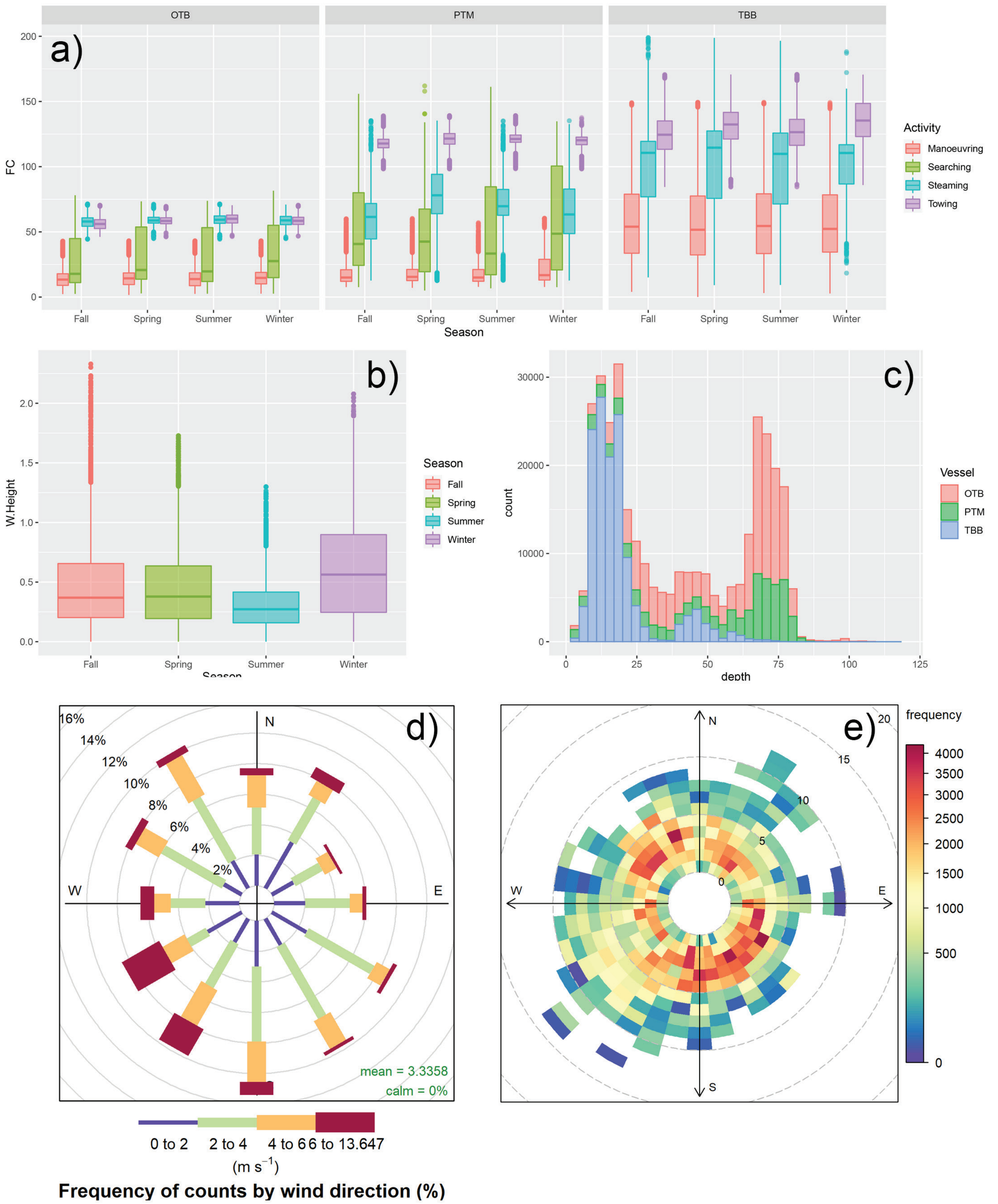

Fig. 2: Summary of the major variables described in the final dataset: a) boxplot with the observed energy expenditure divided by vessel, season, and activity; b) Significant Wave Heights observed, summarized by season; c) activity depth ranges; and d) windRose plot describing wind speed/direction frequencies. Speeds are broken down into the intervals shown by the scale in each panel. The grey circles show % frequencies; e) polarFreq plot to observe wind speed/directions. Each cell displays the total number of hours that the wind was blowing in a certain speed/direction. The number of hours are coded as a nonlinear color scale, while dashed circular grey lines depict the wind speed scale. The date range covered by the data is shown in the strip.

nonlinear relationship with the amount of fuel consumed (Fig. 3f) during the maneuvering and towing phases. Although the influence of this variable was minor in comparison to other variables, the model predicted a fluctuating yet steadily increasing trend in consumption associated with increasing wave heights for the towing and maneuvering phases. Conversely, for the navigation phases, the model suggested none to slight increments

Table 3. GAM model summary.

Parametric coefficients					
	Estimate	**Std.**	**Error**	**Pr(>|t|)**	
(Intercept)	3.2458682	0.0469166	69 184	< 2e-16	***
VesselPTM	0.6257599	0.001732	361 286	< 2e-16	***
VesselTBB	0.810675	0.0034827	232 774	< 2e-16	***
CrosswindHeadwind	0.0203944	0.0007466	27 315	< 2e-16	***
CrosswindTailwind	-0.0087206	0.0012192	-7 153	8.54E-13	***
CrosswaveHeadwave	-0.0137797	0.0007523	-18 316	< 2e-16	***
CrosswaveTailwave	0.0108273	0.001142	9 481	< 2e-16	***
SeasonSpring	0.0601551	0.000892	67 441	< 2e-16	***
SeasonSummer	0.0375851	0.0009366	40 130	< 2e-16	***
SeasonWinter	0.0461042	0.0011919	38 682	< 2e-16	***
Approximate significance of smooth terms:					
	edf	**Ref.df**	**F**	**p-value**	
te(speed,acceleration):ActivityManeuvering	21 556	21 811	604 157	< 2e-16	***
te(speed,acceleration):ActivitySearching	21 018	21 499	27 650	< 2e-16	***
te(speed,acceleration):ActivitySteaming	12 859	18 000	35 865	< 2e-16	***
te(speed,acceleration):ActivityTowing	18 534	19 935	92 777	< 2e-16	***
te(speed,depth):ActivityManeuvering	14 221	19 000	21 657	< 2e-16	***
te(speed,depth):ActivitySearching	6 987	19 000	2 880	< 2e-16	***
te(speed,depth):ActivitySteaming	12 380	16 000	73 907	< 2e-16	***
te(speed,depth):ActivityTowing	17 752	19 000	175 491	< 2e-16	***
s(acceleration):ActivityManeuvering	4 044	4 000	648 469	< 2e-16	***
s(acceleration):ActivitySearching	3 377	3 847	20 456	< 2e-16	***
s(acceleration):ActivitySteaming	3 805	3 890	71 902	< 2e-16	***
s(acceleration):ActivityTowing	1 195	1 160	9 770	0.001191	**
s(speed):ActivityManeuvering	3 912	4 000	224 810	< 2e-16	***
s(speed):ActivitySearching	3 606	3 704	21 186	< 2e-16	***
s(speed):ActivitySteaming	3 953	3 973	385 985	< 2e-16	***
s(speed):ActivityTowing	4 724	4 000	166 898	< 2e-16	***
s(W.Height):ActivityManeuvering	3 946	3 997	66 353	< 2e-16	***
s(W.Height):ActivitySearching	3 950	3 998	40 165	< 2e-16	***
s(W.Height):ActivitySteaming	3 928	3 997	92 619	< 2e-16	***
s(W.Height):ActivityTowing	3 710	3 949	39 137	< 2e-16	***
s(depth):ActivityManeuvering	8 341	8 639	33 067	< 2e-16	***
s(depth):ActivitySearching	8 032	8 626	5 686	3.44E-05	***
s(depth):ActivitySteaming	6 548	7 003	44 106	< 2e-16	***
s(depth):ActivityTowing	8 661	8 857	113 420	< 2e-16	***
s(W.Direction):ActivityManeuvering	1 002	1 004	25 616	4.25E-07	***
s(W.Direction):ActivitySearching	3 952	3 998	70 352	< 2e-16	***
s(W.Direction):ActivitySteaming	3 905	3 994	98 221	< 2e-16	***
s(W.Direction):ActivityTowing	3 853	3 987	23 944	< 2e-16	***
s(ws):ActivityManeuvering	3 633	3 926	5 588	0.000144	***
s(ws):ActivitySearching	3 923	3 996	34 627	< 2e-16	***
s(ws):ActivitySteaming	3 900	3 994	26 737	< 2e-16	***
s(ws):ActivityTowing	3 523	3 875	22 119	< 2e-16	***
s(wd):ActivityManeuvering	1 011	1 023	0.091	0.769912	

Continued

Table 3 continued

	edf	Ref.df	F	p-value	
s(wd):ActivitySearching	3 805	3 975	36 841	< 2e-16	***
s(wd):ActivitySteaming	3 971	3 999	100 249	< 2e-16	***
s(wd):ActivityTowing	3 986	4 000	186 514	< 2e-16	***
s(lat):ActivityManeuvering	3 861	3 987	57 908	< 2e-16	***
s(lat):ActivitySearching	2 880	3 407	7 159	3.91E-05	***
s(lat):ActivitySteaming	3 917	3 996	17 290	< 2e-16	***
s(lat):ActivityTowing	3 988	4 000	826 586	< 2e-16	***
s(lon):ActivityManeuvering	3 872	3 987	14 105	< 2e-16	***
s(lon):ActivitySearching	3 836	3 979	37 802	< 2e-16	***
s(lon):ActivitySteaming	3 915	3 995	37 237	< 2e-16	***
s(lon):ActivityTowing	3 814	3 980	57 916	< 2e-16	***

R-sq.(adj)= 0.88   Deviance explained=88.1%
fREML=1.0829e+06 Scale est.=195.29   n=66708

in consumption for Significant Wave Heights up to 0.5-1 meter, and a consumption decrease for all vessels for wave heights of more than one meter.

As for the direction of the waves in relation to the direction of navigation, the model's estimates revealed an ambiguous pattern with negligible differences in consumption, at least within the wave height range considered. There was also no discernible pattern of change in consumption associated with changes in wind speed and direction. Although significant, the relationship between wind speed and consumption in this context could be described as nonlinear and lacking a clear trend (Fig. 3g). Nevertheless, the interaction between wind direction and navigation direction confirms the significance of the wind variable. The model suggested higher (albeit slightly) consumption values for observations concerning navigation in the opposite direction of the wind (Fig. 3j). Latitude had a visible effect on the energy expenditure of vessels, with higher consumption values predicted for maneuvering and towing at the lowest latitudes for all vessels. Longitude, on the other hand, seems to have a minimal effect on consumption only during the towing phase, where a drop in energy expenditure is expected at higher longitude values. The effect of this variable on the remaining phases is nonlinear and less clear. The relationship between consumption and seasonality (Fig. 3i) was significant, but seasonal variations should be minimal. As illustrated in Table 3 by the estimates (logarithmic scale) provided by the model's parametric coefficients, spring operations were anticipated to have the highest consumption values, while consumption values were forecast to be the lowest in the fall.

Random Forest

Random forest training followed the exact 80/20 random split already used for our GAM model, as both methods used the same training dataset. The standard 10-fold cross validation procedure used during the training phase to define 1) the optimal number of variables sampled at each split, 2) the number of terminal nodes, and 3) the optimal number of trees in the forest, indicated a random sampling of 9 out of 14 explanatory variables to be selected at each split and a maximum number of 650 nodes. The optimal number of trees was derived by evaluating 100 to 2,000 trees. The reductions in MAE and RMSE values, which were employed as the model's performance indicators, suggested 800 trees as the best compromise, with only minor improvements in model performance beyond this number (Fig. 4). With this configuration, our final RF model explained 90.06% of the observed variance, with an RMSE of 12.75 and an MAE of about 8.28. Unlike conventional regression methods, a graphical representation of an RF model is hard to achieve, as this method is based on an ensemble of classification or regression trees. Additionally, the graphical representation of individual trees would just be indicative, as each tree is unique and generated by random resampling processes. Feature importance, calculated during the training phase, is better suited for describing the relative importance of each variable on the model's final estimates. The results of these processes are described in Figure 5. As expected, and in agreement with the GAM model, the RF model attributed the major differences in the response to the activity effect (steaming vs. towing), the vessel effect, and the speed and acceleration variables. Significant wave height, longitude, wave direction, depth, and latitude variables influenced the response variable to a lesser extent and had a similar effect on model accuracy. Given the geographical configuration of the Adriatic coast, the effects of longitude, latitude, and depth should be evaluated holistically, as an increase in longitude and a decrease in latitude roughly correspond to an increase in depth. Finally, The RF deemed the influence of the remaining variables to be minimal, as indicated by the minimal increase in %MSE. Among these were the effects of wind speed and direction, seasonality, and the direction of navigation relative to wind and waves.

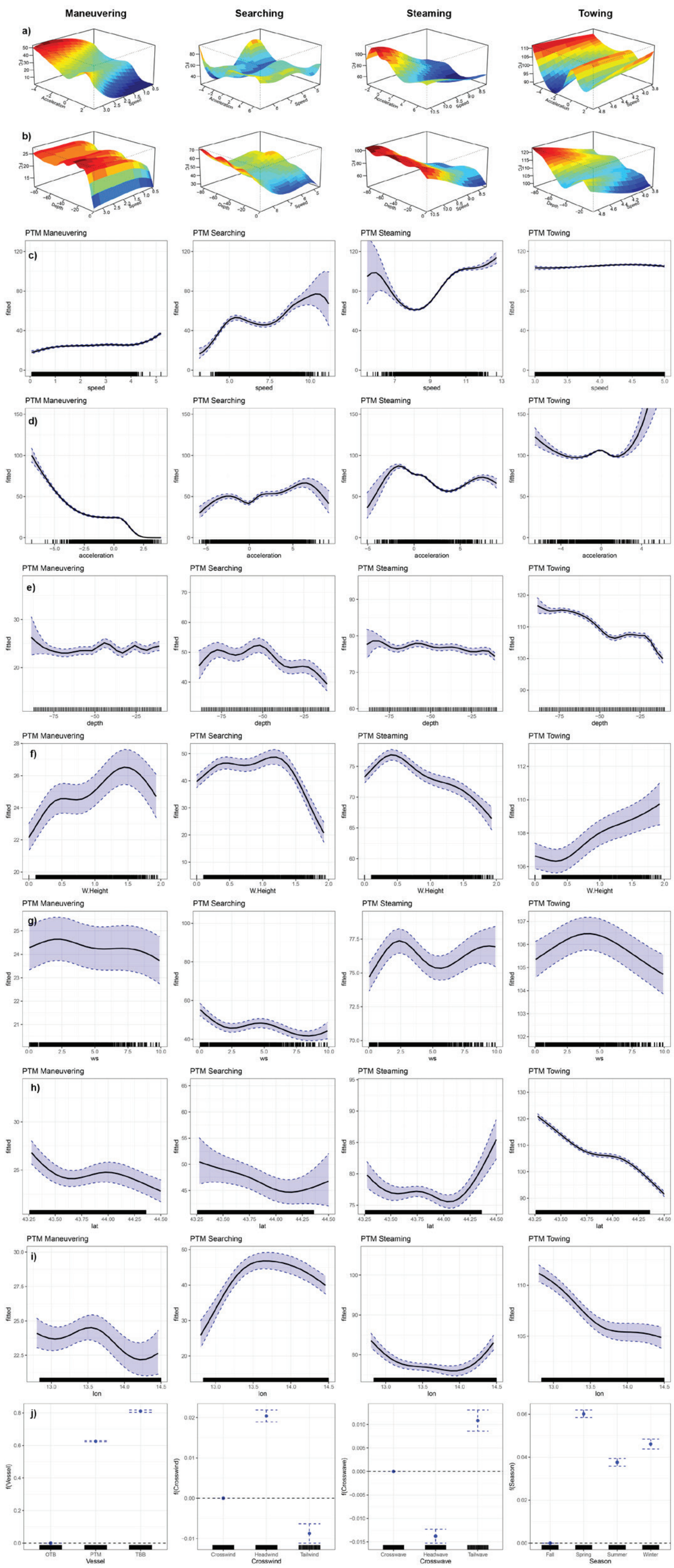

Fig. 3: Example of the estimated influence of both interactions and single predictors on consumption. The shown estimates for consumption (l/h) are the predictions of the GAM model for the pelagic pair trawler. Predictions for the response variable are generated by examining the response shape of the focal predictor of interest while holding the other variables constant at their mean values. A) and b) tensor product splines represent the interaction between acceleration and speed (a) and speed and depth (b) and their combined effect on consumption observed by activity; from c) to j), estimated smooth effects of the individual predictors by activity, independent of the other variables. The rug beneath each smoother's plot represents the actual distribution of the training dataset's observations. The variables shown include speed (c), acceleration (d), depth (e), significant wave height (W.Height, f), wind speed (ws, g), latitude (lat, h), and longitude (lon, i). The dotted lines represent the 95% confidence intervals around the response curve. On the bottom, (j) represents the parametric effect of the vessel, crosswind, crosswave, and season categorical variables.

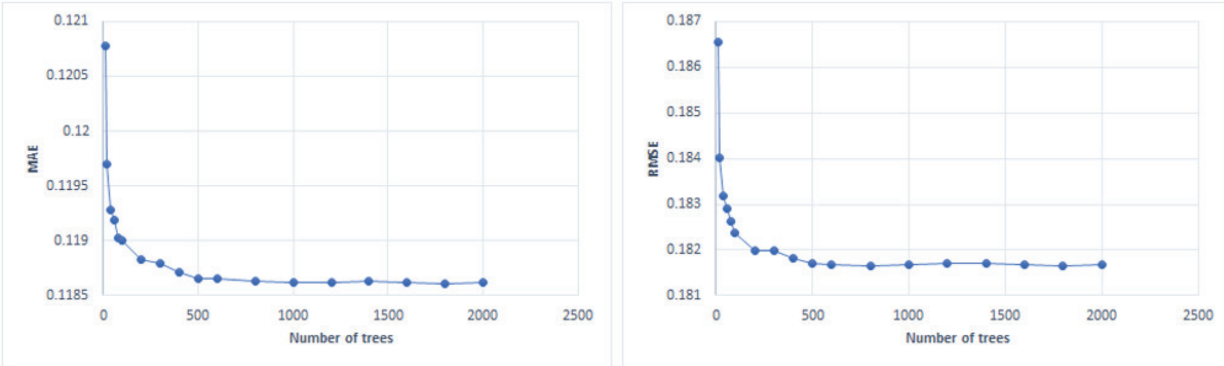

Fig. 4: Relationship between the number of trees in the forest and model performance. Here, model performance is reported in terms of the reduction in the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE).

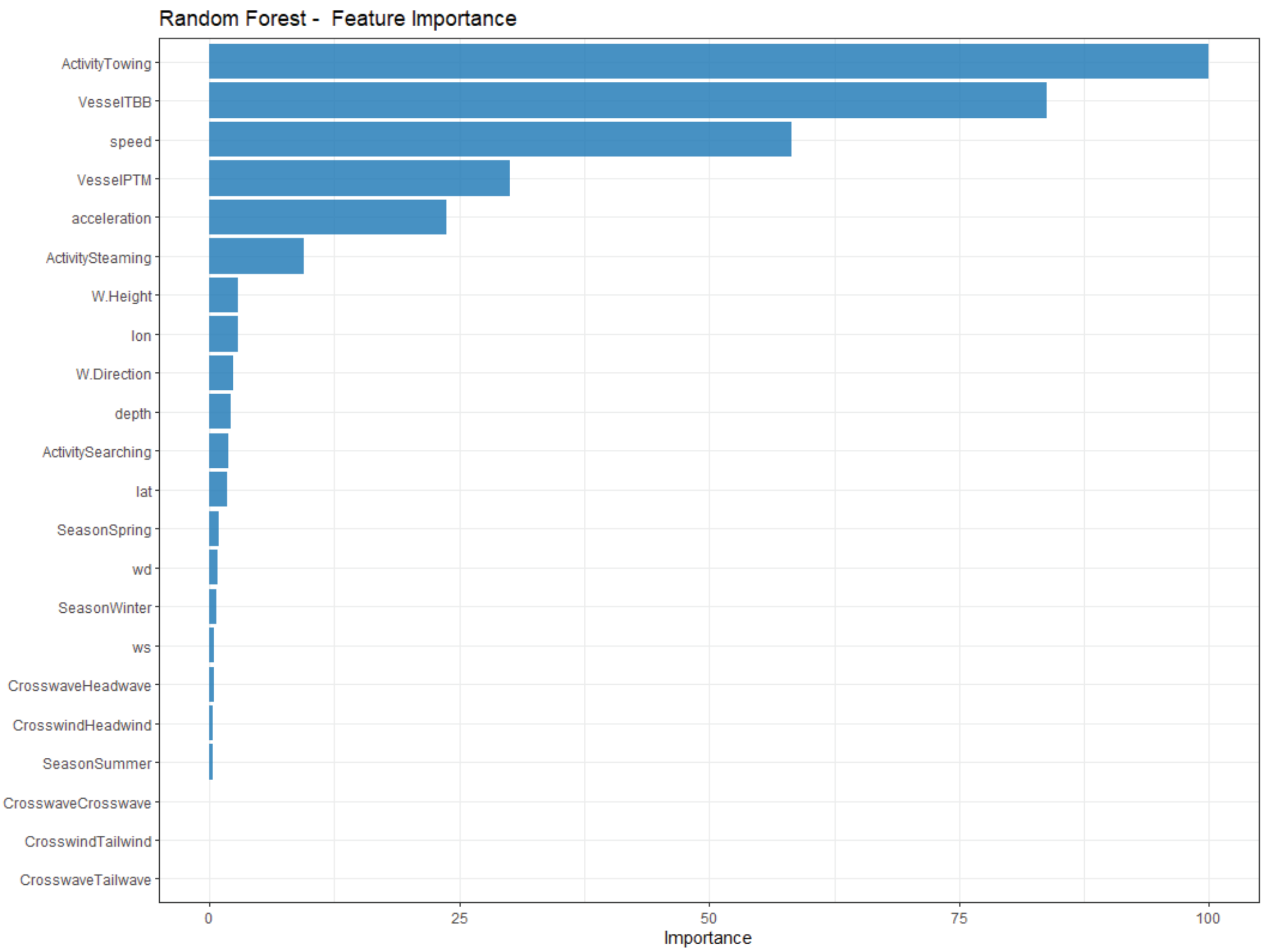

Fig. 5: Feature importance for random forest variables. The importance of predictors in the model is given as the mean % increase in Mean Squared Error, i.e., the mean decrease in model accuracy.

GAM and Random Forest predictive performance evaluation

The test dataset was used to evaluate the capacity of the models to predict the energy expenditure of observed vessels. Both models achieved a good fit, with GAM performing slightly better and with better R-squared (RF=0.86, GAM=0.88) and test RMSE (RF=16.47, GAM=13.98) values. Each model's training RMSE and test RMSE were compared to observe prediction accuracy, where the former is obtained by making the models predict consumption values using training data and comparing the predictions with actual consumption values, and the latter is obtained by repeating the process using a new test dataset. A significant discrepancy between training and test RMSE would imply that the models don't perform equally well on training and previously unseen data. Also in this case, GAM performed slightly better with a training RMSE=13.97 vs. a test RMSE=13.98, compared to RF's training RMSE=12.75 and test RMSE=16.47. A summary of the predictive performance of both models is available in Figure 6. A comparison between predicted vs. observed values highlighted a small bias in the predictions. Occasionally, both GAM and RF either underestimated or overestimated the energy expenditure of vessels with visible patterns for observations defined by low speed and low consumption values.

Discussion

It is important to consider every aspect weighing on the energy balance of fishing vessels in order to optimize consumption and reduce the carbon footprint. The key technical features influencing this balance have already been addressed, and the optimizations for engines, propellers, or the hydrodynamic properties of vessels and gears are accomplished through several technical adaptations to existing technologies. Behavioral modifications such as reducing operational speeds (Poos *et al.*, 2013) or selecting closer fishing grounds (Poos *et al.*, 2013) have also been previously considered. Possibilities for additional behavioral changes include altering fishing activity in response to changing weather and sea conditions. In this study, we evaluated the influence of varying sea states and wind conditions on fishing activities and the energy expenditure of vessels. As far as the authors are aware, this subject is still largely unexplored and has, therefore, served as the focus of this research. The impact of winds and waves was investigated using two different approaches, GAM and random forest. The comparison was specifically designed to capture most of

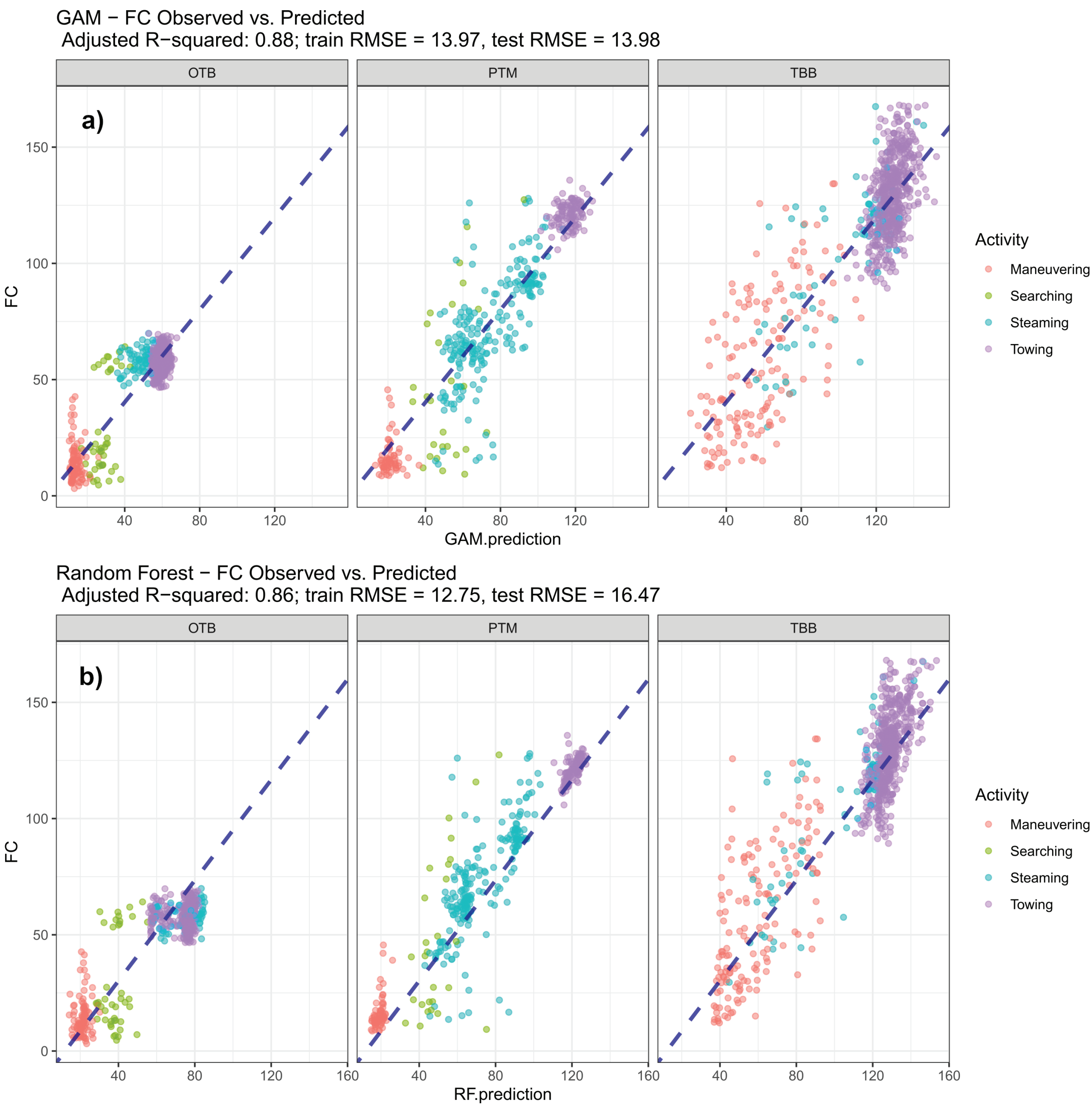

Fig. 6: Predictive performance of the (a) GAM model and (b) Random Forest. The image depicts predicted vs. observed values of consumption FC (l/h) for a random sample of 2,000 observations extracted from the test dataset.

the observed data variances while maintaining reasonable model interpretability. Depth and seasonality were also examined as additional sources of variability, namely the potential differences in consumption owing to a variable bathymetry and potential variations in fishing behavior due to seasonal changes. Both GAM and RF achieved a reasonable fit, with RF performing marginally better in terms of variance explained and GAM achieving better results in terms of prediction accuracy. The minor difference between the two approaches suggested that the GAM model fit was capable of capturing all the main effects and no important interactions were overlooked. A pattern of slightly biased and homoscedastic GAM residuals suggests that some explanatory variable may still be missing from the model. We observed a small bias visible in the predictions within the range of low speed and low consumption values, where sometimes both GAM and RF either underestimated or overestimated the energy expenditure of vessels with visible patterns. We attribute this bias to the presence of secondary on-board operations of a short temporal duration, such as gear shooting, gear hauling, or gear cables recovery, here summarized as the maneuvering phase. As previously mentioned by Sala *et al*. (2011), these operations often generate visible patterns during the activity's low consumption phases. These patterns frequently deviate from the expected energy demands of towing or steaming. Occurring at the beginning and end of each haul, these phases are characterized by variable temporal cadence; their duration depends on multiple factors and varies from one vessel to another. This complicates their accurate identification and inclusion within the already identified phases. Accurately pinpointing these moments would certainly improve the accuracy of prediction models, thereby rectifying the minor discrepancies observed in the predictions.

Both the GAM and RF approaches concurred on the importance of variables, identifying the vessel and the activity phase as those responsible for the most pronounced differences observed in consumption values. In all cases, operational speed undoubtedly remains the most important variable, capable of explaining most of the variability observed in the dataset on its own. As previously discussed (Corbett *et al*., 2009; Ronen, 1982), an increase in navigation and towing speed exponentially increases energy requirements. As already reported (Notti *et al*., 2012), a judicious choice of towing and navigation speed remains a reliable behavioral response, capable of yielding significant reductions in fuel consumption with a considerate containment of costs and emissions.

Although the influence of environmental conditions on consumption was significant, we found their interaction with fishing activity to be ambiguous, and their overall impact was generally minor when compared to the variables considered previously.

Significant wave height had the most recognizable effect on consumption from among all the environmental predictors. However, its effect was variable and dependent on the phase of activity being observed. Operations, such as maneuvering and towing, involving the presence of fishing equipment in water (trawl gears, cables, otter boards, etc.), generally appeared to be more affected by this variable. For both maneuvering and trawling, our GAM model estimated a fluctuating but steadily increasing consumption pattern associated with increments in significant wave height. In contrast, the energy requirements of vessels during navigation appear to be subject to nonlinear fluctuations attributable to the action of waves. For significant wave heights up to 0.5-1 meter, which are comparable to background noise, the model predicted zero or slight increases in observed consumption. Beyond this threshold, a rapid fall in consumption was observed and predicted by the model, perhaps as a result of the skippers' ability to take advantage of more sustained wave motions by adjusting their course. Consequently, within the range of observed values, the effect of wave motion on consumption appears to be greater for techniques where maneuvering and trawling are prevalent, such as beam trawling, with minor effects on fishing techniques where navigation is a substantial part of daily activities.

In this study, the sum of maneuvering and trawling phases accounted for an average of 46.38%, 84.2%, and 91.19% of fishing trips for PTM, OTB, and TBB, respectively. As previously reported by Parente *et al*. (2008), the duration of navigation phases varies and largely depends on the skipper's strategy, while trawling emerges as the most important component of a fishing trip and the most important phase for reduction efforts. Even if minimal, the effect of any variable that increases or decreases energy efficiency during trawling and maneuvering operations should be regarded as important and studied further.

It should be noted that the observed significant wave height range is rather narrow (0-2.3 m) when compared to more prohibitive conditions encountered by other fisheries in different geographic areas. This range roughly corresponds to the regional tolerance interval typical of fishing vessels. The activity of the fleet segment in GSA 17, as the gathered dataset seems to confirm, seldom takes place near the upper limit of this range, and it is difficult to observe any trawling activity in the presence of significantly higher wave height values.

In future, it may be beneficial to further analyze the effects of waves under a broader range of sea-state conditions, since we anticipate that the significance of this variable will be greater in worse sea-state conditions.

The influence of wind speed on consumption was nonlinear and unclear. Although the results described this variable as significant, both GAM and RF indicate that the contribution of this variable is minor. Nevertheless, the GAM model predicted an increase in consumption for any ship displacement against the wind, confirming the importance of this variable, which should be examined further. As with waves, wind speeds also spanned a very narrow speed range (0–13.34 ms^{-1}), and most of the activity took place in the lower speed portion of this range. Moreover, the number of observations reporting wind speed values that are close to the upper limit of this range was relatively low. Finally, the temporal resolution of wind speed and direction variables was rather coarse, as winds were assumed to be constant over 6-hour intervals. The results observed represent an attempt to explore the wind-consumption relationship utilizing the available finer time resolution. This temporal resolution was potentially inadequate for the scope and a finer temporal scale may be required.

As expected, bathymetric conditions do not appear to affect consumption during navigation. During the trawling and maneuvering phases, however, greater depths were projected to result in an increase in consumption, at least for the PTM and OTB vessels. For these vessels, the GAM interaction that accounts for the combined action of speed and depth has shown how, at greater depths, the additional effect of an increased gear drag leads to an increase in energy requirements if not compensated by a drop in towing speed. On the other hand, shallower waters, reduced gear drag, and moderate trawling speeds improve, albeit slightly, energy efficiency while fishing.

The analysis of consumption, with respect to seasonality, revealed minor but significant differences between seasons, with fall and spring representing the minimum and maximum extremes in consumption estimates, respectively. This difference is to be considered attributable to slight local seasonal variations in the fishing strategies employed by skippers. Nonetheless, neither the observed spatial-temporal activity patterns nor the other descriptive variables adequately explained the disparities highlighted in the estimates. The contribution of seasonality should be considered on a strictly regional basis, and different outcomes should be expected for different fisheries in other geographic areas.

Future steps

In future, it would be worth applying the same analytic approach to a larger number of vessels from different fleet segments and to observe the activity in other geographic regions. The effect of shifting marine conditions should be explored, including a broader range of wave and wind conditions, possibly exceeding those already observed in the northern Adriatic Sea. Finer temporal scales, especially for winds, should also be considered. Finally, it would be worthwhile to observe the influence of these variables on catches, investigate potential trends, and determine if and how the weather-driven modulation of consumption is accompanied by specific trends in catches and revenues.

Declarations of interest: none

Funding: this work was supported by the Italian CNR IRBIM, National Research Council – Institute of Marine Biological Resources and Biotechnologies, Ancona, Italy. The funding source was not involved in the study design or in the collection, analysis, and interpretation of data.

References

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. p. 199-213. In: Parzen, E., Tanabe, K., Kitagawa, G. (Eds). *Selected Papers of Hirotugu Akaike*. Springer, New York.

Amante, C., Eakins, B.W., 2009. *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis*. NOAA Technical Memorandum NESDIS, NGDC-24, 19 pp.

Bastardie, F., Nielsen, J.R., Andersen, B.S., Eigaard, O.R., 2010. Effects of fishing effort allocation scenarios on energy efficiency and profitability: an individual-based model applied to Danish fisheries. *Fisheries Research*, 106, 501-516.

Breiman, L., 2001. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Scences*, 16, 199-231.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and regression trees*. The Wadsworth statistics/probability series. Chapman & Hall/CRC press, Boca Raton, Florida, 366 pp.

Buglioni, G., Notti, E., Sala, A., 2011. E-Audit: Energy use in Italian fishing vessels. p. 1043-1047. In: *Sustainable Maritime Transportation and Exploitation of Sea Resources*. Rizzuto & Guedes Soares (Eds).Taylor & Francis Group, London.

Corbett, J.J., Wang, H., Winebrake, J.J., 2009. The effectiveness and costs of speed reductions on emissions from international shipping. *Transportation. Research. Part D: Transport and Environment.* 14, 593-598.

Díaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7, 3.

EU, 2013. Regulation No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC, n.d. 40.

Faraway, J.J., 2016. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC press, 413pp.

Fernández-López, J., Schliep, K., 2018. rWind: Download, edit and include wind data in ecological and evolutionary analysis. *Ecography* 42, 804-810.

Gabiña, G., Basurko, O.C., Notti, E., Sala, A., Aldekoa, S. *et al.*, 2016. Energy efficiency in fishing: Are magnetic devices useful for use in fishing vessels? *Applied Thermal Engineering,* 94, 670-678.

González-Irusta, J.M., González-Porto, M., Sarralde, R., Arrese, B., Almón, B. *et al.* P., 2015. Comparing species distribution models: a case study of four deep sea urchin species. *Hydrobiologia*, 745, 43-57.

Hastie, T.J., Tibshirani, R.J., 1986. Generalized additive models (with discussion). *Statistical Science*, 1 (3), 297-318.

Hastie, T.J., Tibshirani, R.J., 1990. *Generalized additive models, vol. 43. Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC press, Boca Raton, Florida, 352 pp.

Henning, C., 2015. fpc: Flexible Procedures for Clustering. R package version 2.1-10.

Hijmans, R., van Etten, J., 2014. Raster: Geographic data anal-

ysis and modeling. R Package Version 517, 2-12.
Ho, T.K., 1995. Random decision forests. p. 278-282. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, 14-16 August 1995.* Vol 1, IEEE.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An introduction to statistical learning*. Springer, New York, 426 pp.
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R news*, 2/3, 18-22.
Lin, Y., He, J., Li, K., 2018. Hull form design optimization of twin-skeg fishing vessel for minimum resistance based on surrogate model. *Advances in Engineering Software*, 123, 38-50.
Liu, X., Song, Y., Yi, W., Wang, X., Zhu, J., 2018. Comparing the Random Forest with the Generalized Additive Model to Evaluate the Impacts of Outdoor Ambient Environmental Factors on Scaffolding Construction Productivity. *Journal of Construction Engineering and Management*, 144, 6.
Mantari, J.L., Ribeiro e Silva, S., Guedes Soares, C., 2009. Variations on transverse stability of fishing vessels due to fishing gear pull and waves. In: *Proceedings of the XXI Pan American Congress of Naval Engineering (COPINAVAL), Montevideo*. COPINAVAL, Uruguay.
Marmion, M., Hjort, J., Thuiller, W., Luoto, M., 2008. A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. *Earth Surface Processes and Landforms*, 33 (14), 2241-2254.
Neves, M.A.S., Pérez, N.A., Valerio, L., 1999. Stability of small fishing vessels in longitudinal waves. *Ocean Engineering*, 26, 1389-1419.
Neves, M.A.S., Perez, N., Lorca, O., Rodrıguez, C., 2003. Hull design considerations for improved stability of fishing vessels in waves. p. 291-304. In: *Proceedings of Eighth International Conference on the Stability of Ships and Ocean Vehicles (STAB'2003), Madrid, 15-19 September 2003*. STAB'2003, Madrid.
Notti, E., Sala, A., 2012. On the opportunity of improving propulsion system efficiency for italian fishing vessels. In: *Second International Symposium on Fishing Vessel Energy Efficiency, E-Fishing*, *Vigo*, *22-24 May 2012*. Spain.
Notti, E., Buglioni, G., Sala, A., 2012. Energy performance evaluation for fishing vessels. p. 85-94. In: *Proceedings of the 17th International Conference on Ships and Shipping Research*, *Naples*, 17-19 *October 2012*. NAV, Italy.
Notti, E., Figari, M., Sala, A., Martelli, M., 2019. Experimental assessment of the fouling control coating effect on the fuel consumption rate. *Ocean Engineering*, 188, 106233.
Oshiro, T.M., Perez, P.S., Baranauskas, J.A., 2012. How Many Trees in a Random Forest? p. 154-168. In: *Machine Learning and Data Mining in Pattern Recognition*, *Lecture Notes in Computer Science*. Perner, P. (Ed.). Springer, Berlin, Heidelberg.
Parente, J., Fonseca, P., Henriques, V., Campos, A., 2008. Strategies for improving fuel efficiency in the Portuguese trawl fishery. *Fisheries Research*, 93, 117-124.
Parker, R.W., Tyedmers, P.H., 2015. Fuel consumption of global fishing fleets: current understanding and knowledge gaps. *Fish and Fisheries*, 16, 684-696.
Parker, R.W.R., Blanchard, J.L., Gardner, C., Green, B.S., Hartmann, K., *et al.*, 2018. Fuel use and greenhouse gas emissions of world fisheries. *Nature Climate Change*, 8, 333-337.
Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10 (1), 439-446.
Pelletier, N., Audsley, E., Brodt, S., Garnett, T., Henriksson, P., *et al.*, 2011. Energy intensity of agriculture and food systems. *Annual Review of Environment and Resources*, 36, 223-246.
Poos, J.J., Turenhout, M.N.J., van Oostenbrugge, H.A.E., Rijnsdorp, A.D., 2013. Adaptive response of beam trawl fishers to rising fuel cost. *ICES Journal of Marine Science.*, 70, 675-684.
R Core Team, 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Ronen, D., 1982. The effect of oil price on the optimal speed of ships. *Journal of the Operational Research Society*, 33, 1035-1040.
Sala, A., Hansen, K., Lucchetti, A., Palumbo, V., 2008. Energy saving trawl in Mediterranean demersal fisheries. p. 961-964. In: *Ocean Engineering and Coastal Resources.* Guedes Soares & Kolev (Eds). Taylor Francis Group, London.
Sala, A., De Carlo, F., Buglioni, G., Lucchetti, A., 2011. Energy performance evaluation of fishing vessels by fuel mass flow measuring system. *Ocean Engineering*, 38, 804-809.
Sala, A., Notti, E., Bonanomi, S., Pulcinella, J., Colombelli, A., 2019. Trawling in the Mediterranean: An Exploration of Empirical Relations Connecting Fishing Gears, Otterboards and Propulsive Characteristics of Fishing Vessels. *Frontiers in Marine Science*, 6, 534.
Sala, A., Damalas, D., Labanchi, L., Martinsohn, J., Moro, F. *et al.*, 2022. Energy audit and carbon footprint in trawl fisheries. *Scientific Data*, 9, 428.
Sampson, D.B., 1991. Fishing tactics and fish abundance, and their influence on catch rates. *ICES Journal of Marine Science,* 48, 291-301.
Santiago, J.L., Ballesteros, M.A., Chapela, R., Silva, C., Nielsen, K.N. *et al.*, 2015. Is Europe ready for a results-based approach to fisheries management? The voice of stakeholders. *Marine Policy*, 56, 86-97.
Schau, E.M., Ellingsen, H., Endal, A., Aanondsen, S.A., 2009. Energy consumption in the Norwegian fisheries. *Journal of Cleaner Production*, 17, 325-334.
Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9, 307.
Swider, A., Pedersen, E., 2019. Data-Driven Methodology for the Analysis of Operational Profile and the Quantification of Electrical Power Variability on Marine Vessels. *IEEE Transactions on Power Systems*, 34, 1598-1609.
Swider, A., Langseth, H., Pedersen, E., 2019. Application of data-driven models in the analysis of marine power systems. *Applied Ocean Research*, 92, 101934.
Tello, M., Ribeiro e Silva, S., Guedes Soares, C., 2011. Seakeeping performance of fishing vessels in irregular waves. *Ocean Engineering*, 38, 763-773.
Thierry, N.N.B., Tang, H., Achile, N.P., Xu, L., Hu, F. *et al.*, 2020a. Comparative study on the full-scale prediction performance of four trawl nets used in the coastal bottom trawl

fishery by flume tank experimental investigation. *Applied Ocean Research*, 95, 102022.
Thierry, N.N.B., Tang, H., Liuxiong, X., You, X., Hu, F. *et al.*, 2020b. Hydrodynamic performance of bottom trawls with different materials, mesh sizes, and twine thicknesses. *Fisheries Research*, 221, 105403.
Thrane, M., 2004. Energy Consumption in the Danish Fishery: Identification of Key Factors. *Journal of Industrial Ecology*, 8, 223-239.
Tsagarakis, K., Carbonell, A., Brčić, J., Bellido, J.M., Carbonara, P. *et al.*, 2017. Old Info for a New Fisheries Policy: Discard Ratios and Lengths at Discarding in EU Mediterranean Bottom Trawl Fisheries. *Frontiers in Marine Science*, 4, 99.
Verhulst, N., Jochems, J., 1993. *Final Confidential report for the project TE-1.102 hp NET'92*. Research Project Financed by the Commission of the European Communities within the Frame of the EEC research Programme in the fisheries sector ("FAR").
Wileman, D.A., 1984. *Project" Oilfish": Investigation of the Resistance of Trawl Gear. Danish Institute of Fisheries Technology*. Technical Report n. NP-6750170, 123 pp.
Wileman, D.A., Hansen, K., 1988. *Estimation of the drag of trawls of known geometry*. Hirtshals Denmark: The Danish Fisheries Technology Institute, 51 pp.
Wood, S., 2018. Package 'mgcv'. 2016. URL http://cran. r-project. org/web/packages/mgcv/mgcv. pdf. R package version (2018): 1-0.
Zacharioudaki, A., Ravdas, M., Korres, G., 2019. *Mediterranean Production Centre MEDSEA_HINDCAST_WAV_006_012*, 51 pp. https://www.cmcc.it/wp-content/uploads/2021/06/CMEMS-MED-QUID-006-012.pdf

Supplementary Data

The following supplementary information is available online for the article:

Fig. S1: Estimated influence of both interactions and single predictors on consumption. The estimates for consumption (l/h) shown represent the predictions of the GAM model observed for the bottom otter trawler. Predictions for the response variable are generated exploring the response shape of the focal predictor of interest while retaining the other variables constant at their mean values. a) and b) tensor product splines representing the interaction between the variables acceleration and speed(a), and speed and depth (b) and their combined effect on consumption observed by activity; From c) to j) Estimated smooth effects, observed by activity, of the individual predictors independent of the other variables. The rug under each smoother's plot represent the actual distribution of the observations within the training dataset. The variables shown include speed(c), acceleration (d), depth (e), significant wave height (W.Height, f), wind speed (ws, g), latitude (lat, h) and longitude (lon, i). The dotted lines represent the 95% confidence intervals around the response curve. On the bottom (j) is represented the parametric effect of the categorical variables Vessel, Crosswind, Crosswave and Season.

Fig. S2: Estimated influence of both interactions and single predictors on consumption. The estimates for consumption (l/h) shown represent the predictions of the GAM model observed for the beam trawler. Predictions for the response variable are generated exploring the response shape of the focal predictor of interest while retaining the other variables constant at their mean values. a) and b) tensor product splines representing the interaction between the variables acceleration and speed(a), and speed and depth (b) and their combined effect on consumption observed by activity.; From c) to j) Estimated smooth effects, observed by activity, of the individual predictors independent of the other variables. The rug under each smoother's plot represent the actual distribution of the observations within the training dataset. The variables shown include speed(c), acceleration (d), depth (e), significant wave height (W.Height, f), wind speed (ws, g), latitude (lat, h) and longitude (lon, i). The dotted lines represent the 95% confidence intervals around the response curve. On the bottom (j) is represented the parametric effect of the categorical variables Vessel, Crosswind, Crosswave and Season.