



Ca' Foscari University of Venice
Department of Environmental Sciences, Informatics
and Statistics

Master's Degree Programme in Computer Science

Master's Thesis

Analyses of the fishing effort and of the underwater noise for a sustainable exploitation of the northern Adriatic Sea

Candidate:
Giulia Rovinelli - 867381

Supervisors:
Dr. Alessandra Raffaetà
Dr. Marta Simeoni

Academic year 2021-2022

Acknowledgements

Alone we can do so little; together we can do so much.

– Helen Keller

I would like to spend a few words to thank all those people that have helped me and supported me during this journey.

First of all I would like to thank my supervisors, *Alessandra Raffaetà* and *Marta Simeoni*, for their guidance, continuous support, patience, motivation, enthusiasm and knowledge. I am grateful for the trust placed in me and in this work.

I am particularly grateful to Professor *Claudio Silvestri*, Professor *Fabio Pranovi* and Doctor *Elisabetta Russo* for the help and valuable advice received during the design and implementation of the system.

Thanks also to the *CNR ISMAR* group, in particular Doctors *Fantina Madricardo* and *Marta Picciulin*, for their support in the development of the underwater sound propagation model and their kindness in sharing their knowledge and studies.

Finally, I would like to express my deepest appreciation to my family for being there for me in every moment and because with their love, continuous encouragement and support they allowed me to undertake this journey.

Contents

1	Introduction	1
2	Related works and Background	5
2.1	Related works	5
2.1.1	Fishing activities forecast	5
2.1.2	Underwater noise	8
2.2	Background	9
2.2.1	Multiple aspect trajectories	9
2.2.2	MobilityDB	10
2.2.3	Previous work on fishing vessels in the northern Adriatic Sea	13
3	Fishing Effort Forecast	15
3.1	Computation of the fishing effort over the grid	16
3.2	Dataset	18
3.3	Machine Learning Methods	20
3.3.1	Random Forest	20
3.3.2	Extremely Randomized Trees	21
3.3.3	eXtreme Gradient Boosting	22
3.3.4	Multi-layer Perceptron	22
3.4	Model Hyperparameters	22
3.5	Evaluation metrics	26
3.5.1	Mean Absolute Error	26
3.5.2	Root Mean Squared Error	26
3.5.3	Coefficient of Determination	26
3.6	Results	27
4	Underwater noise	35
4.1	Engine noise from fishing vessels	36
4.2	Regular grid with environmental features	39
4.3	Propagation modelling	40

4.4	Underwater sound propagation model	44
4.5	Model implementation	46
4.5.1	Underwater sound propagation maps	53
5	Conclusions and future works	69
A	Forecast results	73
A.1	Feature Permutation Importance	73
A.2	Evaluation metrics results	74
	Bibliography	79

List of Tables

2.1	Fishing vessels activities and associated ID.	13
3.1	Features used for fishing effort forecast.	18
3.2	Training and testing datasets size.	20
3.3	Random Forest hyperparameters.	23
3.4	Extremely Randomized Trees hyperparameters.	24
3.5	eXtreme Gradient Boosting hyperparameters.	25
3.6	Multi-layer Perceptron hyperparameters.	25
3.7	Training execution time (expressed in seconds) for each machine learning method considering the three training datasets.	27
3.8	Results for each season using three years (2016-2017-2018) as training dataset.	29
3.9	Average seasons results considering each training dataset.	31
4.1	Characteristics of the trawler [65].	36
4.2	Trawler source level with 63 <i>Hz</i> and 125 <i>Hz</i> at cruising speed and high speed [65].	37
4.3	Sound level for different horsepower ranges.	38
4.4	Sound level for different gear types.	38
4.5	Monthly average evolution of the mean of the 1/3 octave bands at 63 and 125 <i>Hz</i> of ambient noise measurements expressed in <i>dB</i> [70].	45
4.6	Number of vessels, trips, AIS data and points relative to the position of the boat every 30 seconds for the years 2019 and 2020.	47
4.7	Underwater noise range for maps and analysis of peaks and averages.	53
4.8	Number of cells per category of the underwater noise peak maps 2019.	59
4.9	Percentage of cells per category of the underwater noise peak maps 2019.	59
4.10	Number of cells per category of the underwater noise average maps 2019.	60

4.11 Percentage of cells per category of the underwater noise average maps 2019.	61
4.12 Number of cells per category of the underwater noise peak maps 2020.	63
4.13 Percentage of cells per category of the underwater noise peak maps 2020.	63
4.14 Number of cells per category of the underwater noise average maps 2020.	64
4.15 Percentage of cells per category of the underwater noise average maps 2020.	64
A.1 Results for each season using one year (2018) as training dataset.	75
A.2 Results for each season using two years (2017-2018) as training dataset.	76

List of Figures

2.1	Representation of the trajectories after their reconstruction.	12
3.1	Overview of the framework for predicting the fishing effort.	16
3.2	Annual fishing effort.	17
3.3	Representation of the temperature grid (black) and the centroids of the squared grid (yellow) where each cell has a size 3×3 kilometers.	19
3.4	Permutation feature importance of the four machine learning models considering three years of training and one of testing.	28
3.5	MAE, RMSE and R^2 scores for each season considering three years of training for the four machine learning models: RF (blue), ET (orange), XGBoost (green) and MLP (red).	30
3.6	MAE, RMSE and R^2 scores resulting from the models trained only with 2018 (blue), 2017-2018 (orange), 2016-2017-2018 (green) datasets.	32
3.7	Fishing effort maps of the actual fishing effort values for 2019 and the predicted ones using the four machine learning models considering three years as training dataset.	33
4.1	Trawler used as reference fishing vessel [65].	37
4.2	Regular grid in which each cell has a size of 1×1 kilometer.	39
4.3	Part of the trajectory 389 carried out by the CARLO ALBERTO II boat with the AIS transmissions and the reconstruction of the points every 30 seconds of the same trajectory.	46
4.4	Points belonging to segments of a trip of the CARLO ALBERTO II boat.	48
4.5	Regions in which the sound generated by a vessel propagates underwater, considering the circle with the maximum propagation radius and the cells of the regular grid in which the sound will then be distributed.	49
4.6	Underwater sound propagation in an instant of time of the boat CARLO ALBERTO II which is carrying out a fishing activity.	50

4.7	Propagation of the underwater noise of the CARLO ALBERTO II fishing vessel while it is navigating (red dot) or fishing (light blue dot) every 30 seconds.	51
4.8	Propagation of underwater noise of fishing vessels in the northern Adriatic Sea on the 30th of January 2019 at 08:40. Vessels are shown with a light blue dot if they are fishing, red otherwise.	52
4.9	Underwater noise peak maps for each month of 2019.	54
4.10	Underwater noise average maps for each month of 2019.	55
4.11	Monthly fishing effort for 2019.	56
4.12	Underwater noise peak maps for each month of 2020.	57
4.13	Underwater noise average maps for each month of 2020.	58
4.14	Percentage of cells divided by category of the underwater noise maps for each month of 2019.	62
4.15	Percentage difference on the number of cell.	65
4.16	Percent difference on peak maps.	66
4.17	Percentage difference on the maps of the averages.	67
A.1	Permutation importance of the four machine learning models considering one year of training and one of testing.	73
A.2	Permutation importance of the four machine learning models considering two years of training and one of testing.	74
A.3	MAE, RMSE and R^2 scores for each season considering one year of training for the four machine learning models: RF (blue), ET (orange), XGBoost (green) and MLP (red).	75
A.4	MAE, RMSE and R^2 scores for each season considering two years of training for the four machine learning models: RF (blue), ET (orange), XGBoost (green) and MLP (red).	76

Chapter 1

Introduction

The secret of getting ahead is getting started.

– Mark Twain

The area of the northern Adriatic Sea has a high productivity rate regarding the fishing activities and it is recognised as one of the most exploited areas of the Mediterranean Sea. In order to make fishing activities sustainable and to guarantee a productive and healthy ecosystem, there is a strong need to develop effective fishery management plans for constant monitoring and for prediction. For this reason, it is of fundamental importance to analyse the data relating to the movements of fishing vessels and their catch.

The thesis builds on previous studies about the fishing activity in the northern Adriatic Sea [73, 74, 14] and aims at exploring two different aspects: the forecast of the fishing effort, that is an essential indicator for monitoring the fishing pressure on an area of interest over time, and the underwater noise modelling. Both the fishing effort and the underwater noise propagation model are of fundamental importance to improve the knowledge of the spatial-temporal aspects of fishing activities in order to reduce unsustainable exploitation. These two goals have been proposed by the ecologists and they allow us to experiment different analysis techniques in the same conceptual framework, based on a particular kind of semantic trajectories, namely *Multiple Aspects Trajectories* [57]. Hence, this work reveals the potentialities of the framework which has been proved to be flexible and adequate to cope with the representation and analyses of various phenomena.

Starting from the AIS data, we reconstruct and enrich the trajectories by

assigning to each segment the activity carried out by the boat (*in port, entering to/exiting from the port, navigation and fishing*). In this way, considering only the portions of the trajectory in which the vessel is fishing, we compute the fishing effort. By partitioning the northern Adriatic Sea with a regular grid of 3×3 kilometers, we distribute the fishing effort over the entire grid. Subsequently, we enrich the initial dataset with daily environmental factors, such as sea surface temperature, spectral significant wave height, wind speed and sea salinity, downloaded from *Copernicus*¹, an European Union’s Earth observation programme which offers information services that draw from satellite Earth Observation and in-situ (non-space) data. After enriching our initial dataset with the environmental data, a variety of prediction methods are used in order to assess their prediction ability related to the fishing effort. Since the fishing effort is real-valued and continuous, the prediction task can be formulated as a regression problem where the output of the regression model is the fishing effort value for a time period and a grid cell.

In order to make the fishing effort forecast for 2019, we consider three different training datasets, increasing their size by adding a year of data each time. In particular, we initially consider only the year 2018, then 2017 and 2018 and finally 2016, 2017 and 2018. This method for splitting the data is chosen due to the sequential nature of the data, for which random sampling is not suitable. The machine learning methods employed for regression modelling are: Random Forest, Extra Trees, XGBoost and Multi-layer Perceptron. Model hyperparameters for each method were chosen by performing random searches on the hyperparameter space and taking the best scoring settings. Finally, to compare the performance of the regressor models, we use the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).

The issue of designing predictive methods to forecast fishing effort is challenging because the dataset contains multivariate and spatio-temporal aspects to cope with. In fact, spatio-temporal data has two unique attributes: spatial properties, which consist of a geographical hierarchy and distance, and temporal properties, that consist of closeness, period and trend [96].

Concerning the northern Adriatic Sea exploitation, another fundamental aspect is the underwater noise generated by vessels, which has a significant short and long term impact on animal species. Using the trajectories and the characteristics of the boats, we build a model for the propagation of underwater noise based not only on the technical characteristics of the fishing boats’ engine but also on environmental

¹<https://marine.copernicus.eu/>

factors that vary in each season, such as sea surface temperature, sea salinity, depth of the sea and sea water potential of Hydrogen (pH). In particular, we used the reconstructed trajectories to overcome the problem of the non-homogeneity of the AIS data transmission. In fact, we get the positions of all the fishing vessels at the same time instants, i.e. every 30 seconds. Then we use these points to calculate the sound propagation radius, (i.e. the distance at which the noise generated by the fishing vessel is no longer perceptible) with respect to the horsepower of the boat, if it is fishing, and the environmental conditions of the cell to which the point under analysis belongs to. Again we partition the northern Adriatic Sea into a regular grid, each cell having a size of 1×1 kilometer, smaller than before because this dimension allows us to better model the effect of the propagation. To measure the noise level we consider the centroid of the cells as listening points. By taking a boat in a precise instant of time and the corresponding propagation radius, we distribute the sound in the cells that are within the calculated radius.

Having this model we can make several analyses. First, using the *QGIS Time-Manager* it is possible to generate animations in which, by selecting the vessels and the number of information per frame, we can visualize these vessels moving in the northern Adriatic Sea and, at the same time, the underwater noise generated by these vessels. Secondly, we can aggregate the sound propagation of individual boats to produce underwater noise maps that are useful for identifying areas where underwater noise can damage the marine environment, even permanently.

The novelty of this model is related to the fact that, instead of taking underwater noise measurements with physical listening points (hydrophones in the water) as it is done in all the works relating to the same topic, see for instance [64, 5, 1], we take “virtual” listening points (the cells centroids), at a distance of 1 kilometer from each other, and using the semantic trajectories of the fishing boats and their characteristics, we propagate the sound produced by vessels.

To accomplish this project we use *MobilityDB* [98], a moving object database that extends the type system of PostgreSQL² and PostGIS³ with abstract data types for representing moving object data, such as *temporal types*. Such data types represent the evolution on time of values of some element type, called the base type of the temporal type. We also use Python and in particular Scikit-learn library to perform the experiments using machine learning models for regression. Furthermore, using the QGIS open source software, we create maps to display, in a simple and clear way, the fishing effort forecast, its prediction, and the areas in the northern

²<https://www.postgresql.org/>

³<https://postgis.net/>

Adriatic Sea characterised by the most intense noise pollution. In particular, we use the *QGIS TimeManager*⁴ which adds time controls to QGIS: using these time controls, we can animate vector features based on a time attribute.

The rest of this work is organized as follows. In Chapter 2 we present some relevant papers related to fishing activities forecast and underwater sound propagation. Moreover, we set up the preliminaries of our work: first we introduce the concept of trajectory and its extension to multiple aspect trajectories, and then we describe how they are represented in MobilityDB. Chapter 3 presents the fishing effort forecast: the creation of the dataset for the prediction, the machine learning models, the choice of hyperparameters, the evaluation metrics and the results. In Chapter 4 we describe the creation of a model for the underwater noise propagation starting from the characteristics of the vessels, their semantic trajectories, the physical models for underwater sound propagation and the daily environmental data. Finally, Chapter 5 concludes this thesis and gives some directions for future work.

⁴<https://plugins.qgis.org/plugins/timanager/>

Chapter 2

Related works and Background

Education is the most powerful weapon you can use to change the world.

– Nelson Mandela

2.1 Related works

The northern Adriatic Sea is one of the most exploited areas of the Mediterranean Sea, causing an over-exploitation of the fish resources. Having a clear representation and understanding of the main factors driving such phenomenon is of paramount importance both for ecologists and for local policymakers [14]. In this way, they can generate plans to make fishing activities sustainable and identify specific areas where it is necessary to reduce the navigation of vessels that can damage the marine environment. In this setting, we can identify two main tasks with the goal of improving the knowledge on the exploitation of the northern Adriatic sea. The first one is the computation of the fishing effort, an essential indicator for monitoring the fishing pressure on an area of interest over time, and in particular its forecast, which allows to study the most fishing exploited areas and protect our seas from overfishing. The second one is the underwater noise generated by vessels which has a significant short and long term impact on animal species [89].

2.1.1 Fishing activities forecast

In the literature there are many works on fishing activities forecast. From a fishing management view, works like [27] propose a seasonal forecasting decision support

tool that is updated automatically on a daily basis and accessible from a tailored website to predict tuna distribution. The work in [42] shows the performance of Convolutional Neural Networks (CNNs) models to forecast 1-month ahead monthly anchovy catches considering only anchovy catches in previous months as inputs to the models. CNNs models are being applied in different fields and usually provide very promising results. Recently, CNNs have been largely applied to time series forecasting, although there are not many studies developed in fisheries sciences [50, 16]. In particular, authors in [16] developed time series based neural network and fuzzy logic models for fish recruitment analysis using fish stock–recruitment data incorporating environmental information. In [20] the authors examined the relationship between sea surface temperature and chlorophyll-a concentration thanks to the integration of satellite data and statistical models output. In [51, 52] the authors focus on mapping dredge gear fishing grounds using complementary data to assess the spatio-temporal distribution and intensity of fishing activity. Similarly to [27] and [20], the work proposed in this thesis uses environmental data for the forecast. Also, similarly to [42] and [14], in this work we exploit the information relating to the fishing activities. In particular, we use the fishing effort to predict the areas that will be the most exploited by fishing activities, instead in [42, 14] the authors use catch information to predict future catches.

From a viewpoint that considers the geolocation technology used to track ships, some works use Vessel Monitoring System (VMS). For example, [54] combines this information with catch data (that are used for modelling species presence) for the estimation of fishing effort. Other works use satellite images [20] or AIS [25, 33, 22, 87] that could be a good alternative option, especially working at a regional scale. Initially, the AIS was designed primarily as a support and aid for navigation to avoid collisions between boats, as it allows their detection. The *International Maritime Organization*¹ (IMO) requires AIS transmission for ships with gross tonnage equal to or greater than 300 tons, for all passenger ships and for all boats over 15 meters in length. Nowadays AIS is also used as a primary resource to obtain useful data for the study of fishing-related activities, probably due to their free display. Furthermore, AIS provides the possibility for ships to exchange, in near real time, state vectors (position, speed, course, rate of turn, etc.), static (vessel identifiers, dimensions, ship-type etc.), and voyage related information (destination, ETA, draught, etc.) [77]. In [33] the suitability of AIS data was tested to assess bottom trawl fishing activity in the Mediterranean Sea over three years. Most of

¹<https://www.imo.org/>

the works that consider the geolocation technology focus on training models to forecast when a vessel performs a fishing activity. Finally, in [14] the authors use AIS data to reconstruct vessels multiple aspect trajectories and exploit the domain knowledge from experts to determine the activity of vessels (e.g., fishing or not) on their trajectory segments. Based on the knowledge of ranges of fishing speed for different types of fishing gears (e.g., trawlers, long-liners, etc.), they encode the specific rules to detect vessel activities. By exploiting this information, they compute the area swept by vessels while fishing, in order to estimate the Catch Per Unit Effort (CPUE).

From the viewpoint of the analysis of models for time-series forecasting, many studies used the Autoregressive Integrated Moving Average [13] (ARIMA), e.g., [6, 47, 56, 67, 94] that has been popular and widely chosen for modelling fisheries science's time series data [42, 82], and the Seasonal version (SARIMA), e.g., [4, 66]. In [95] the authors use fishing locations data of two years and derive chlorophyll-a and sea surface temperature from satellite data to determine potential fishing grounds of the commercial Indian mackerel using a presence-absence data model, Generalized Additive Model (GAM) and presence-only data model, maximum entropy (Maxent). In [84] correlative models were performed for the determination of potential fishing grounds of small-scale fishery combining VMS and environmental variables. In particular, they used Generalized Linear/Additive/Boosting Models, Classification Tree Analysis, Artificial Neural Network, Surface Range Envelope, Flexible Discriminant Analysis, Multiple Adaptive Regression Splines, and Random Forest. Also, the authors in [97], based on the knowledge of fishing chronology among trawlers, design a CNN to predict the short-term fishing effort distribution.

The works discussed in this section use different methods for various purposes. Our goal is to study and use machine learning models to predict the fishing effort as described in Chapter 3. Our approach uses the same datasets employed in [14] for the forecasting (although in our work we forecast the fishing effort rather than the CPUE), but unlike that work, in which the dataset is split used a standard 5-fold cross-validation, we consider the year 2019 for testing and all other previous available years for training. This data splitting was chosen because when data is time-stamped, cross-validation should be avoided as it will not be sensitive to latent concept drift almost always present in real data.

2.1.2 Underwater noise

Underwater noise from human activities is known to have a number of adverse effects on aquatic life [61, 78, 90]. These can range from acute effects such as permanent or temporary hearing impairment [55, 79], to chronic effects such as developmental deficiencies [21, 60] and physiological stress [93, 92, 72]. For these reasons it is of fundamental importance to model the underwater noise produced by boats and provide ecologists with sound maps. In [19] the authors try to summarize the status regarding continuous underwater radiated noise from shipping in European waters to provide recommendations on possible future activities. In particular, the work is focused on four main topics: characteristics and quantification of noise sources from various ship types, impacts on marine fauna, existing policies, including guidelines, decisions, resolutions and regulations and mitigation measures for the abatement of ship noise and noise-related impact. The work in [64] shows two acoustic surveys conducted at 40 listening points distributed along the three inlets that connect the Venice lagoon to the sea, in order to characterize the local noise levels and evaluate the fish spatial distribution by means of its sounds. The study in [5] aimed to determine noise intensity and frequency of traditional fishing boats. In particular, they found out that the variation of traditional fishing boats equipped with various engine types influenced the characteristics of the noise emitted, and that noise intensity based on frequency can be detected at a certain distance with a decreasing pattern: the higher the frequency, the faster the disappearance of the intensity along with increasing distance.

Unlike previous works, authors in [23] developed a real-time low-cost Passive Acoustics Monitoring (PAM) system tailored to assess anthropogenic noise in marine environments. The so called “*COntrollo Rumore MARino*” (*CORMA*) project² is not limited at the already important real-time detection and transmission of the underwater noise coefficient, but it combines a complex and sophisticated mathematical model that intercepts the position of large ships (equipped with AIS, Automatic Identification System), acquires their noise footprint and creates propagation in the water, taking into account the depth and morphology of the seabed. Another approach of real-time monitoring is based on cabled seafloor technologies [59, 1, 32]. Cabled observatories can take advantage of several benefits such as the availability of direct power supply and the possibility to transfer data via a wired link that allows high bandwidth and throughput. Like the previous work, also [24] is based on AIS data and a simple sound transmission model. The

²<http://www.corma.info/>

difference is that in this case the authors try to derive a large-scale noise map which provides an accurate picture as a starting point to identify areas where noise is likely to be and likely not to be a problem for the underwater world. Modelled large-scale maps of shipping noise have been produced for several regions [7, 76]. Differently the aim of [30] is to carry out a multi-site validation of a large-scale shipping noise map constructed using a generic shipping noise model. More recently, the project *SOUNDSCAPE*³ produces an estimate of the spatio-temporal distribution of noise levels generated by human activities at sea, aggregating multiple sources, and assessing short, mid and long term source contributions to the global noise field.

Summing up, many of the works presented try to reconstruct underwater sound using hydrophones and listening points for its measurement. To the best of our knowledge, no work in the literature, starting from AIS data and the reconstruction of multiple aspect trajectories, aims to create an underwater sound propagation model by distinguishing even when a vessel is sailing from when it is fishing. This is exactly the goal of our work, which is described in detail in Chapter 4.

2.2 Background

Moving objects are objects that change their position and can modify some associated characteristics or properties over time. Such objects can be, for example, people, animals, vehicles, boats, etc. The dissemination of location devices and IoT (Internet of Things) technologies has led to the collection of huge amounts of data describing the temporal evolution of these objects. This creates an opportunity to build applications on this data, which require the creation of a Moving Object Database System (MOD) [44].

2.2.1 Multiple aspect trajectories

A trajectory represents the evolution of the position of an object moving in space during a certain time interval in order to reach a certain goal [81]. Moving objects are characterised by continuous movement over time, while their position can only be updated at discrete moments due to the limitation of acquisition, storage and processing technologies [58]. Therefore, starting from a discrete set of spatio-temporal points, it is fundamental to reconstruct the *continuous* movement of the object, thus obtaining a trajectory, called *raw* trajectory, since it is limited to space and time [63]. These trajectories are suitable for applications that aim

³<https://www.italy-croatia.eu/web/soundscape>

only at locating a moving object or at calculating some statistics on the space-time characteristics of the trajectories. However, most applications require to integrate these trajectories with additional information related to the application context. A trajectory that has been enriched with annotations and/or one or more complementary segmentations is called a *semantic trajectory* [80]. The enrichment of trajectories with other data is called *semantic enrichment* process. In this way the trajectories are integrated with additional data called annotations. An *annotation* is any additional data that is related to the entire trajectory or with some of its sub-parts. For example, recording the length of a boat’s journey is an annotation at the trajectory level; instead, recording the different activities of the boat during its trip is an annotation that changes within the same trajectory.

A semantic trajectory becomes a complex object with numerous complex data dimensions that are contextual to the movement and heterogeneous in the form. In particular, we consider the MASTER (Multiple Aspects Trajectory) [57] model because it is more flexible and expressive than other model in the literature. This model is based on the concept of *aspect*. An *aspect* consists of “a real-world fact that is relevant for the trajectory data analysis” [57]. There are different kinds of aspects:

- *volatile* aspects are associated with the points of the trajectory since they vary during the movement of the object;
- *long-term* aspects do not change during an entire trajectory and therefore are related to it;
- *permanent* aspects characterize the entire life of an object, thus they are connected to the moving object.

Based on this notion, a *multiple aspect trajectory* is defined as a sequence of spatio-temporal points of a moving object with a (possibly empty) set of long-term aspects. Each point can have a set of volatile aspects and the moving object can be related to a set of permanent aspects.

These types of data (moving objects and semantic trajectories) can be easily modeled with the MobilityDB database.

2.2.2 MobilityDB

MobilityDB [98], which uses the MOD abstract datatype approach [43], is an open source extension of PostgreSQL and PostGIS that supports temporal types and

space-time operators to manage moving objects. MobilityDB defines abstract data types (ADTs) for representing data of moving objects [8]. In particular, it provides the following temporal data types: `tbool`, `tint`, `tfloat`, `ttext`, `tgeompoint` and `tgeogpoint`. These temporal types are based, respectively, on the basic types `bool`, `int`, `float` and `text` provided by PostgreSQL and on `geometry` and `geography` provided by PostGIS (limited to 2D or 3D points). The time type on the other hand can be `timestamp`, `timestampset`, `period`, or `periodset`. Every combination of a base type and a time type defines one temporal type. The types are supported with spatiotemporal index access methods by extending GiST (Generalized Search Tree) and SP-GiST (Space Partitioning GiST) [99].

MobilityDB was created on the concept of trajectory to allow the reconstruction and processing of moving objects starting from coordinates or points in space obtained from modern GPS systems or other tracking systems. Consequently, in MobilityDB it is possible to reconstruct the trajectories starting from a set of points (lon, lat, t) , where lon represents the longitude, lat the latitude and t the instant in time in which that particular object is in that specific position (lon, lat) . By grouping the spatial coordinates and associated time instants, we can build a `tgeompoint`, that is a sequence of points (lon, lat, t) that belong to a single trajectory using the following functions:

```
tgeompointseq(array_agg(tgeompointinst(ST_Transform(
  ST_SetSRID(ST_MakePoint(lon, lat), 4326), 5676), t)
  ORDER BY t))
```

In particular:

- `ST_MakePoint` constructs a point starting from the specified longitude (lon) and latitude (lat);
- `ST_SetSRID` sets the SRID of the point to 4326, which is the standard WGS 84 GPS coordinates;
- `ST_Transform` transforms GPS spherical coordinates into planar coordinates;
- `tgeompointinst` creates a temporal point taking as arguments the associated temporal points and instants;
- `array_agg` stores in an array all the temporal points belonging to a particular trip and sorts them based on the time instant (as specified by the `ORDER BY` clause);

- `tgeompointseq` takes as argument the array of temporal points and returns a temporal point whose duration is expressed by the `sequence` time value.

A `sequence` represents the evolution of a value during a sequence of time instants in which the values between these instants are obtained with piecewise or linear interpolation.

Finally, the trajectories can be reconstructed using the function `trajectory` [99] which requires the sequence of temporal points just created as argument (a `tgeompoint`):

```
trajectory(tgeompoint): geometry(line)
```

This function returns a `geometry` which allows the visualization of the trajectory in its spatial reference system (Figure 2.1).



Figure 2.1: Representation of the trajectories after their reconstruction.

The ADT model implemented in MobilityDB encapsulates the complete trajectory within an object, which can be stored in a single attribute. To limit the number of instants within the trajectories in the MobilityDB, we can use various functions, for example:

```
startInstant(ttype)      : ttypeinst
endInstant(ttype)       : ttypeinst
instantN(ttype, int)    : ttypeinst
instants(ttype)         : ttypeinst[]
```

In these functions, `ttype` is any temporal type, and the result given by the function must have the same base type as the temporal type given as argument. These functions allow us to access individual instants within the trajectory, choose a specific instant or return them in an SQL array that allows random access, joining and sorting.

Another important function that allows us to have the point in the trajectory at the time instant given in input is `valueAtTimestamp`. It is defined as follows:

```
valueAtTimestamp(ttype, timestamptz) : base
```

where `ttype` represents any temporal type and `timestamptz` represents an instant in time. For example, the input of the function is the trajectory (of type `geometry`) and a time instant (of type `timestamp`) and the output is the point of the trajectory in that instant of time (of type `geometry`).

2.2.3 Previous work on fishing vessels in the northern Adriatic Sea

This thesis builds on the work we have done in [73, 74, 14]. The first objective of that work was the reconstruction of the trajectories of fishing vessels in the northern Adriatic Sea starting from AIS (Automatic Identification System) data for the years 2015-2018. The AIS data contains the position of the boat, the instant in time of detection and other information such as the speed of the boat. After the reconstruction of these trajectories, we have performed the enrichment of the trajectories: the trajectories have been divided into segments and we have assigned semantic annotations to each segment. In particular, we annotated each segment with the activity carried out by the boat distinguishing between *in port*, *entering to* and *exiting from* the port, *navigation* and *fishing*. Table 2.1 presents the activities of the fishing vessels and their associated ID.

ID	Fishing vessels activities
0	in port
1	exiting from port
2	entering to port
3	fishing
4	navigation

Table 2.1: Fishing vessels activities and associated ID.

The *in port*, *exiting from port* and *entering to port* situations can be deduced from the position of the extremes of the segment with respect to the port area. If none of the previous cases applies, the *fishing* or *navigation* activities are established on the basis of the average speed of the boat. More precisely, if the average speed is in the range of the fishing speed of the gear the boat is equipped with, the boat is assumed to be in a fishing phase, otherwise, it is assumed to be in a navigation phase.

In this way it was possible to calculate, for each trajectory, the duration, the length of the trip and the actual fishing activity. We had also a dataset containing the daily landings (catch amounts in kilogram) for the fishing vessels obtained from the Chioggia's Fish Market. These daily landings were used to distribute the catch using two different approaches. In particular, we used a uniform distribution and a weighted one with respect to the number of boats in the same area. As expected, the weighted distribution allowed us to identify fishing areas more precisely.

The reconstruction of the trajectories starting from the AIS data, and in particular their enrichment, is of fundamental importance for the work developed in this thesis. In particular, the segmentation of the trajectories with their activity is useful for the calculation of the fishing effort and its prediction, which will be described in Chapter 3. Similarly, also for the creation of the underwater sound propagation model, which will be described in Chapter 4, it is of paramount importance to know when the boat is fishing since, in this case, the sound is much more intense. Consequently, before the construction of the model, we performed the segmentation of the trajectories and their enrichment.

Chapter 3

Fishing Effort Forecast

The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present.

– Paul Saffo

Oceans and seas are essential components of the Earth’s ecosystem. Due to their continuous exploitation by human activities, there is the need to develop strategies to monitor the entire ecosystem ensuring its health and the availability of the marine resources for the future generations. In this context, the analysis of fishing effort data has long occupied fisheries researchers in deriving appropriate indices of abundance for identifying changes in the size of fish populations [12] in order to ensure a healthy and productive ecosystem. In fact, fishing effort is a measure that can be used by experts to understand whether the exploitation of the marine environment and fishing activities are sustainable and do not damage the marine world. For the aforesaid reasons, in [74, 14] we have used AIS data to reconstruct, in time and space, the trajectories of the fishing vessels, and also to enrich the resulting trajectories with additional information. Moreover, by considering the daily fish catch reports of the Chioggia fish market, we have distributed the quantity of fish caught along the trajectories. Finally, using the spatio-temporal database and some relevant environmental factors, we have tried to predict the Catch Per Unit Effort (CPUE) using a variety of machine learning methods.

In this chapter, starting from the works described above, I will present the fishing effort forecast, by first describing the creation of the dataset used for the prediction, then the machine learning methods employed and finally the results

obtained with their limitations and future works. Figure 3.1 shows the main steps we have carried out to be able to predict the fishing effort (which corresponds to the bottom section of the Figure).

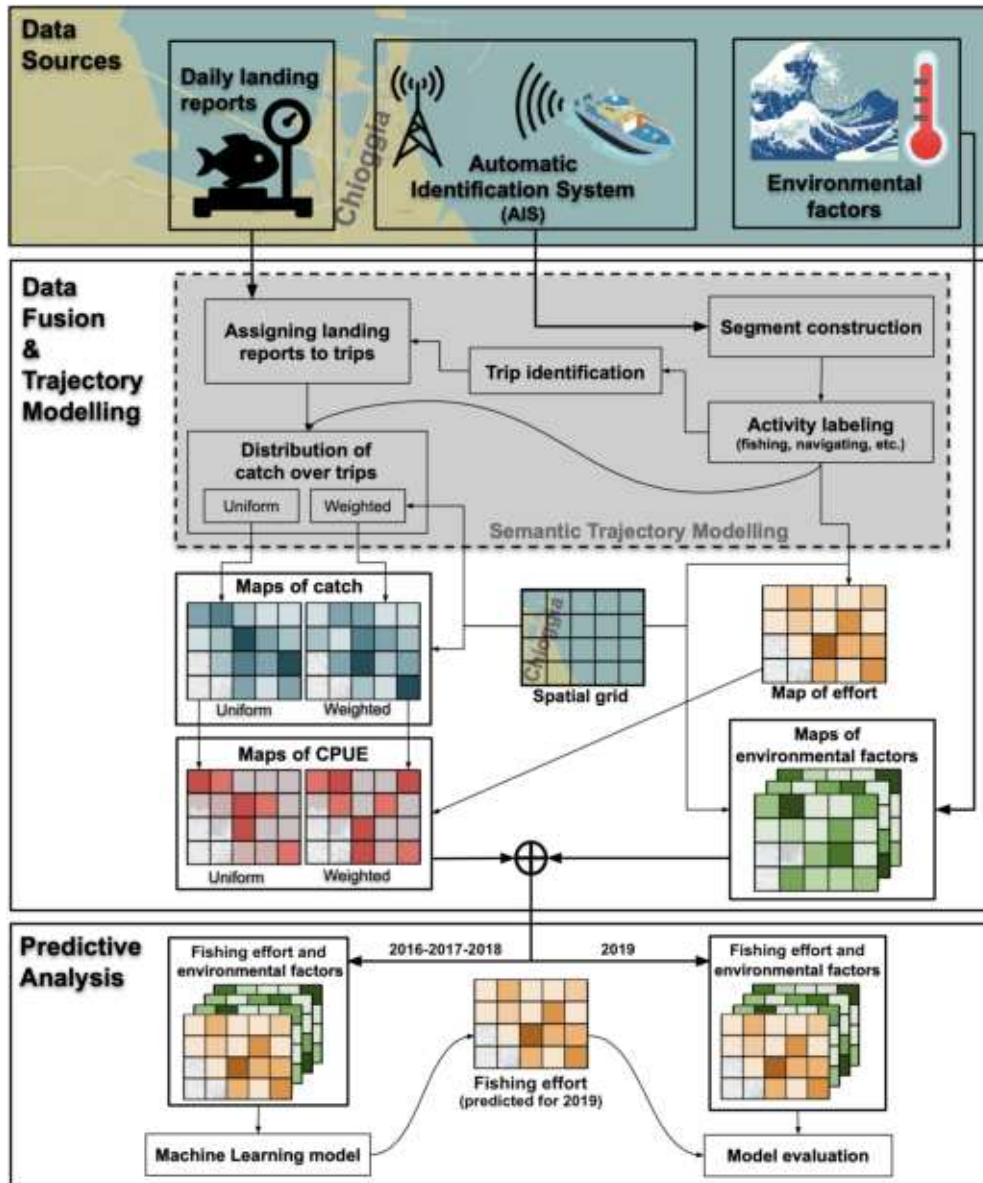


Figure 3.1: Overview of the framework for predicting the fishing effort.

3.1 Computation of the fishing effort over the grid

After the creation of the multiple aspect trajectories, we proceed with the computation of the fishing effort, an essential indicator for monitoring fishing pressure on an area of interest over time. First of all, we partition the Northern Adriatic Sea into a regular grid where each cell has a size of 3×3 kilometers. The

fishing effort for a spatio-temporal cell is defined as the ratio between the area of the cell “swept” by vessels while fishing during the associated time period and the total area of the cell itself [14]. The swept area depends on the employed gear which can be obtained from a specific dataset where each vessel, identified by its MMSI, is associated with its gear (small and large bottom otter trawl, mid-water pair trawl and Rapido).

Let c be a spatio-temporal cell and g a gear. The *fishing effort* with respect to the gear g in the cell c is defined as follows:

$$fe(c, g) = \frac{(\sum_{tr \in T, gear(tr)=g} len(tr \cap c)) \times gear_width(g)}{area(c)} \quad (3.1)$$

where T is the set of multiple aspect trajectories, $len(tr \cap c)$ returns the sum of the lengths of the fishing segments of trajectory tr falling in the spatio-temporal cell c , $gear_width(g)$ is the width of the net of gear g and $area(c)$ is the total area of the spatial component of the cell c .

At this point, we have computed the total fishing effort in each spatio-temporal cell c by summing up the fishing effort for each gear. Moreover, for the evaluation of fishing exploitation by environmental experts, we have also built fishing effort annual maps. In Figure 3.2 there are the fishing effort of 2018 (Figure 3.2(a)) and 2019 (Figure 3.2(b)). From the two figures we can easily see how the most exploited areas are the same for both years.

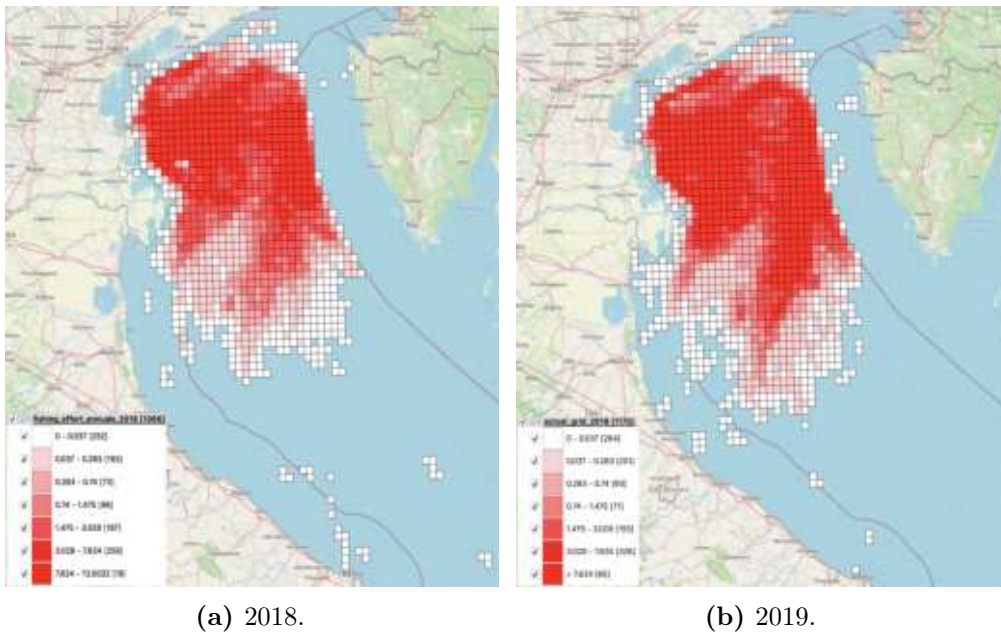


Figure 3.2: Annual fishing effort.

3.2 Dataset

In order to predict the fishing effort, it was first necessary to create a dataset with environmental, spatial and temporal information presented in Table 3.1. In particular, the spatial information are the longitude and latitude grid cell centre, considering a grid size of 3×3 kilometers. As temporal variables, we consider the day of the year, the month of year, the week of the year and the season, converted to four binary features using one-hot encoding. Hot-encoding essentially transforms categorical into numeric attributes. Finally, environmental information was obtained thanks to the *Copernicus Programme of the European Union*¹ which provides regular and systematic reference information on the physical and biogeochemical ocean and sea-ice state for the global ocean and the European regional seas [53]. As suggested by environmental experts, for the environmental features we took into consideration the daily sea surface temperature (expressed in degrees celsius), the daily spectral significant wave height (in meters), the monthly average wind speed (in m/s) and the daily sea salinity (expressed in psu, Practical Salinity Units). These environmental factors can affect fishing activities on a daily basis.

	Attribute description	Type	Unit
Environmental	daily sea surface temperature	float	celsius
	daily spectral significant wave height	float	meter
	monthly average wind speed	float	m/s
	daily sea salinity	float	psu
Spatial	longitude of the grid cell centre	float	degree
	latitude of the grid cell centre	float	degree
Temporal	day of the year (1-365)	int	
	month of the year (1-12)	int	
	week of the year (1-53)	int	
	season (1-4)	int	

Table 3.1: Features used for fishing effort forecast.

Copernicus Marine Service environmental data are organized as follows:

- longitude of the grid cell centre;
- latitude of the grid cell centre;

¹<https://marine.copernicus.eu/>

- value of the environmental feature considered.

Clearly the size of the regular grid of the environmental data downloaded from *Copernicus* is not the same as ours. For example, the information relating to the spectral significant wave height is characterized by a squared grid size of one kilometer, or the sea surface temperature by cells with a larger size (about four kilometers). For this reason, we first made a correction of the data in order to have the environmental information associated with each cell of the grid characterized by a given dimension (in our case $3\text{ km} \times 3\text{ km}$).

In particular, if the grid with the environmental data, in Figure 3.3 the temperature in black, is larger than the $3\text{ km} \times 3\text{ km}$ grid used, we took the centroids of the 3×3 kilometers cells (the yellow points in the Figure) and we assigned to each centroid the temperature value equal to the one of the cell in which it is contained. If a point was exactly on the intersection of two temperature cells (as shown in the enlargement of Figure 3.3), to assign the correct value to the centroid, we computed the average of the temperature cells that intersect the centroid analyzed. Eventually, each $3\text{ km} \times 3\text{ km}$ cell has the same environmental features as its centroid.

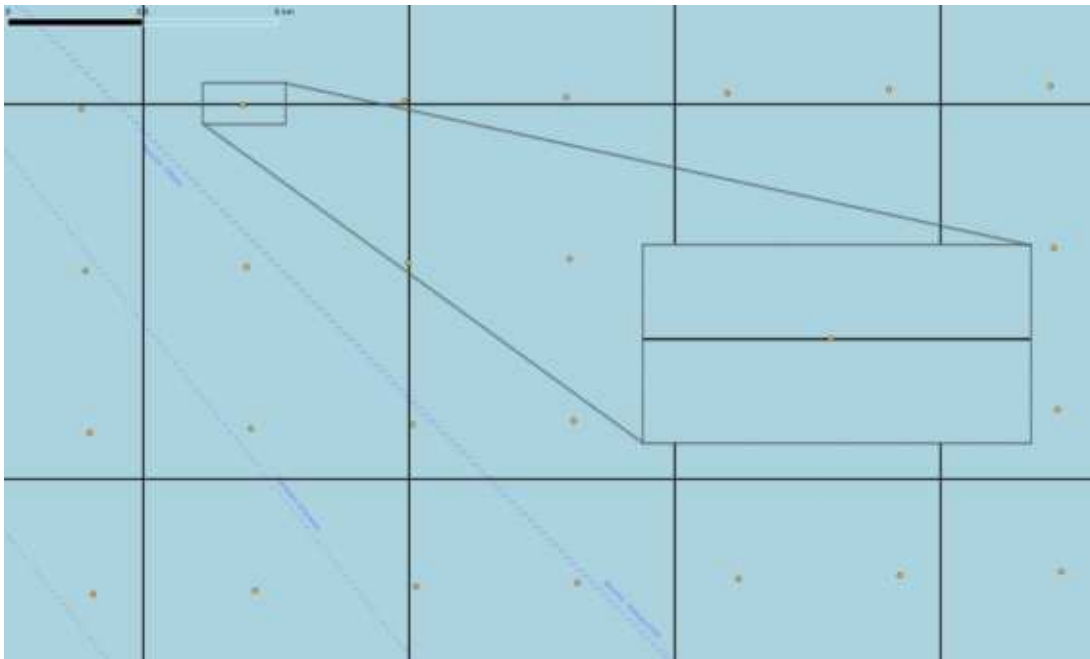


Figure 3.3: Representation of the temperature grid (black) and the centroids of the squared grid (yellow) where each cell has a size 3×3 kilometers.

Since the fishing effort is real-valued and continuous, the prediction task can be formulated as a regression problem where the output of the regression model is the value of the fishing effort for a time period p and a grid cell c . In order to make the fishing effort forecast for 2019, three different training datasets were considered,

increasing their size by adding a year of data each time. In particular, the training and testing datasets and their dimensions are illustrated in Table 3.2. This method for splitting the data is chosen due to the sequential nature of the data, for which random sampling is not suitable.

	Year	Size
Training	2018	77577
	2017 - 2018	152766
	2016 - 2017 - 2018	243080
Testing	2019	83539

Table 3.2: Training and testing datasets size.

Some machine learning methods (explained in Section 3.3) were applied to these three different training datasets to try to understand whether, and to what extent, more information from previous years could have a positive impact on the forecast and therefore an increase in model reliability.

3.3 Machine Learning Methods

Starting from the work done in [14], four Machine Learning methods were used to predict the fishing effort rather than the CPUE. These methods, each of which was part of *Scikit-learn* software package [62] (an open-source machine learning library for Python²), are:

- Random Forest [15];
- Extremely randomized Trees (Extra Trees) [38];
- eXtreme Gradient Boosting (XGBoost) [17];
- Multi-layer Perceptron [75].

All the machine learning algorithms used to forecast the fishing effort are described in detail below.

3.3.1 Random Forest

Random Forest (RF) is an example of *ensemble method*. An ensemble for classification is a composite model, made up of a combination of classifiers: the

²<https://www.python.org/>

individual classifiers vote, and a class label prediction is returned by the ensemble based on the collection of votes [45]. Therefore, RF is an ensemble tree-based machine learning algorithm [15]. When used for classification, a Random Forest obtains a class vote from each tree, and then classifies using majority vote. When used for regression, the predictions from each tree are simply averaged [46]. The trees are created by drawing a subset of training samples through replacement (a bagging approach). This means that the same sample can be selected several times, while others may not be selected at all [9].

Random Forest is a modification of the standard tree growing algorithm that works as follows:

- Bagging is exploited to improve accuracy of single decision trees.
- Trees are fully grown until pure leaves. Bagging does not reduce bias, therefore we need accurate (low bias) trees.
- Random input selection is used: each node is built on a small random subset of the feature set.

Aside from being simple to use, RF is generally recognized for its accuracy and its ability to deal with small sample sizes and high-dimensional feature spaces. At the same time, it is easily parallelizable and has therefore the potential to deal with large real-life systems [10].

3.3.2 Extremely Randomized Trees

Extra Trees (ET) is a tree-based ensemble method for supervised classification and regression problems [38]. ET and RF have a lot in common, but they have also some differences. Both of them are composed of a large number of decision trees, and the output is obtained by considering each tree. In particular, for regression problems, it is calculated as the average of outputs of individual regression trees, while for classification problems it is obtained by majority vote. The main difference between the two is how the randomness is introduced during the training. To train an ET, multiple trees are trained: each tree is trained on all training data. Similar to the Random Forest the best split at a node is found by analyzing a subset of all available features, but instead of searching for the best threshold for each feature, a single threshold for each feature is selected at random [41].

Experiments in [38] show that this method is most of the time competitive with Random Forest in terms of accuracy, and sometimes it achieves better results. Since it removes the need for the optimization of the discretization thresholds, it has also a clear advantage in terms of computing times and ease of implementation [39].

3.3.3 eXtreme Gradient Boosting

XGBoost is a tree-based ensemble method, proposed by Chen and Guestrin, which is an improved version of gradient boosting with higher computation efficiency and better capability to deal with over-fitting problems [17]. XGBoost aims to prevent over-fitting but also optimize the computation resources [28]. The processes of additive learning in XGBoost is the following: the first learner is fitted to the entire space of the data given in input, and a second model is then fitted to these residuals to improve the performance of a weak learner. This fitting process is repeated a few times until the stopping criterion is met. Then an aggregation of outputs of all the models is considered to be the final output.

The success of XGBoost is largely due to its scalability in all scenarios. The system runs more than ten times faster than existing popular solutions on a single machine and scales to billions of examples in distributed or memory-limited settings [17]. Moreover, it performs parallel computing and has the merit of optimally using memory resources and effectively taking care of missing values during the training process [48].

3.3.4 Multi-layer Perceptron

The Multi-layer Perceptron (MLP) consists of a system of simple interconnected neurons, or nodes, representing a nonlinear mapping between an input vector and an output vector [37]. The neurons are organized in layers which are connected by weights. The MLP has one or more hidden layers between its input and output layers, and the information flows unidirectionally from the input layer to the output layer, passing through the hidden layers [11]. Perceptrons for the same layer have the same activation function. In general, it is a sigmoid for the hidden layers. Depending on the application, the output layer can also be a sigmoid or a linear function [83]. Learning for the MLP is the process to adapt the connections weights in order to obtain a minimal difference between the network output and the desired output. The algorithm used for this purpose is the so called *Back-propagation* algorithm which is based on gradient descent techniques [68]. MLP can be applied to complex non-linear problems and it works well with large input data.

3.4 Model Hyperparameters

Model hyperparameters for each method were chosen by performing random searches on the hyperparameter space and taking the best scoring settings. Specif-

ically, a *Randomized Grid Search* was performed on the combination of different values for hyperparameters, and the values that resulted with the highest coefficient of determination score were selected. Models hyperparameters and their selected values will be discussed below.

Random Forest

As specified in Section 3.3, for the implementation of the fishing effort prediction we used the *Scikit-learn*³ library. In particular, for Random Forest machine learning model the *RandomForestRegressor* was used. Hyperparameters that are selected via random search are as follows.

- `n_estimators` specifies the number of trees in the RF.
- `min_samples_split` is the minimum number of samples needed to split a node.
- `min_samples_leaf` specifies the minimum number of samples required to be at a leaf node.
- `max_features` corresponds to the number of features that the algorithm considers when looking for the best split.
- `max_depth` specifies the maximum depth of each tree.

Table 3.3 specifies Random Forest hyperparameters, their searched values and the best value for each training dataset.

Parameter	Best value			Searched values
	2018	2017-18	2016-17-18	
<code>n_estimators</code>	1000	1000	300	100, 200, 300, 500, 1000
<code>min_samples_split</code>	8	10	8	8, 10, 12
<code>min_samples_leaf</code>	3	4	3	3, 4, 5
<code>max_features</code>	auto	auto	auto	auto, sqrt, log2
<code>max_depth</code>	150	150	200	80, 90, 100, 150, 200

Table 3.3: Random Forest hyperparameters.

³<https://scikit-learn.org/stable/>

Extremely Randomized Trees

In order to forecast the fishing effort we used the *Scikit-learn* library. In particular, for Extra Trees machine learning model the *ExtraTreesRegressor* was used. Hyperparameters, that are selected via random search, are the same used for the Random Forest (described in the previous paragraph). Table 3.4 specifies Extra Trees hyperparameters, their searched values and the best value for each training dataset.

Parameter	Best value			Searched values
	2018	2017-18	2016-17-18	
<code>n_estimators</code>	200	300	1000	100, 200, 300, 500, 1000
<code>min_samples_split</code>	8	8	8	8, 10, 12
<code>min_samples_leaf</code>	3	4	4	3, 4, 5
<code>max_features</code>	auto	auto	auto	auto, sqrt, log2
<code>max_depth</code>	150	100	100	80, 90, 100, 150, 200

Table 3.4: Extremely Randomized Trees hyperparameters.

eXtreme Gradient Boosting

As specified in Section 3.3, for the implementation of the fishing effort prediction we used the *Scikit-learn* library. In particular, for eXtreme Gradient Boosting machine learning model the *GradientBoostingRegressor* was used. Hyperparameters that are selected via random search are as follows.

- `n_estimators` is the number of boosting stages that will be performed.
- `min_samples_split` corresponds to the minimum number of samples required to split an internal node.
- `max_depth` is the maximum number of nodes in the tree.
- `learning_rate` corresponds to how much the contribution of each tree will get smaller.

Table 3.5 specifies eXtreme Gradient Boosting hyperparameters, their searched values and the best value for each training dataset.

Parameter	Best value			Searched values
	2018	2017-18	2016-17-18	
n_estimators	500	600	600	500, 600, 700
min_samples_split	5	5	6	5, 6
max_depth	6	6	4	4, 5, 6
learning_rate	0.1	0.1	0.1	0.001, 0.01, 0.1, 0

Table 3.5: eXtreme Gradient Boosting hyperparameters.

Multi-layer Perceptron

In order to forecast the fishing effort we used the *Scikit-learn* library. In particular, for Multi-layer Perceptron machine learning model the *MLPRegressor* was used with Adam optimizer (a stochastic gradient-based optimizer proposed by Kingma, Diederik, and Jimmy Ba [49]). Hyperparameters that are selected via random search are as follows.

- `learning_rate` corresponds to how much we move in the direction opposite to the gradient.
- `hidden_layer_sizes` corresponds to the number of hidden layers.
- `activation` corresponds to the activation function that we want to use for hidden layers.

Table 3.6 specifies Multi-layer Perceptron hyperparameters, their searched values and the best value for each training dataset.

Parameter	Best value			Searched values
	2018	2017-18	2016-17-18	
learning_rate	invscaling	invscaling	constant	constant, invscaling, adaptive
hidden_layer_sizes	40	40	20	10, 20, 30, 40, 50, 60, 70, 80, 100, 200, 300
activation	tanh	tanh	tanh	identity, logistic, relu, softmax, tanh

Table 3.6: Multi-layer Perceptron hyperparameters.

3.5 Evaluation metrics

In order to compare the performance of the regressor models, we use the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). They are described in detail below.

3.5.1 Mean Absolute Error

The Mean Absolute Error is formally define as follows.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3.2)$$

where \hat{y} are the predicting values and y corresponds to the ground truth. In our case the MAE is computed for each period (i.e., season) as the mean of the absolute errors of the predicted average fishing effort for all cells in that period. Therefore, n denotes the number of cells with a fishing effort value in the specified period p and \hat{y}_i and y_i correspond respectively to the predicted and the actual fishing effort values for a given cell and a time period p . The MAE ranges between 0 and $+\infty$, where 0 is the best case.

3.5.2 Root Mean Squared Error

The RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3.3)$$

where \hat{y} and y are defined as above. Also the RMSE is computed for each period (i.e., season) as the mean of the errors of the predicted average fishing effort for all cells in that period. The RMSE ranges between 0 and $+\infty$, where 0 is the best case.

3.5.3 Coefficient of Determination

The Coefficient of Determination (R^2) is formally define as follows [29].

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.4)$$

where \hat{y} and y are defined as above and \bar{y} is the mean of the observed data, i.e. $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. The coefficient of determination can be interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variables [18]. In our case, it is computed as the MAE and RMSE.

The Coefficient of Determination is defined between 0 and 1 [29], where 1 is the best-case scenario and any value lower than zero points out arbitrarily worse results. The fact that positive values of the R^2 range between 0 and 1 is one of the main advantage of the Coefficient of Determination. For example, values like $R^2 = 0.8$ clearly indicate a very good regression model performance, regardless of the ranges of the ground truth values and their distributions [18]. A value of MAE or RMSE equal to 0.9 is less informative about the quality of the regression model since these two measures are scale-dependent.

3.6 Results

After selecting machine learning methods hyperparameters and choosing the evaluation metrics, we can proceed with the fishing effort forecast.

Training dataset	ML model	Execution Time
2018	Random Forest	283.75
	Extra Trees	15.49
	XGBoost	17.51
	Multi-layer Perceptron	2.79
2017-2018	Random Forest	604.64
	Extra Trees	53.86
	XGBoost	203.01
	Multi-layer Perceptron	4.57
2016-2017-2018	Random Forest	361.84
	Extra Trees	387.69
	XGBoost	249.22
	Multi-layer Perceptron	4.93

Table 3.7: Training execution time (expressed in seconds) for each machine learning method considering the three training datasets.

First, we need to train machine learning models with the selected hyperparameters. Table 3.7 shows the training execution time, expressed in seconds, for

each machine learning method considering three training datasets. It is possible to observe that the training execution time for the Multi-layer Perceptron is much lower than the other algorithms and, by increasing the size of the dataset the execution time remains almost the same (the increase is at most 1-2 seconds). This is not the case for the other methods, where an increase in the dataset size also results in an increase in the time required. In particular, Random Forest and Extra Trees, considering three years of training, have an execution time of about 6 minutes. The fact that the execution time for Random Forest with two years as training dataset is almost double that of three years as training dataset is due to the number of trees in the RF. In fact, Table 3.3 specifies the value of the `n_estimator` hyperparameter which in the case of 2018 and 2017-18 is 1000, while for the dataset that contains the years (2016-17-18) it is 300.

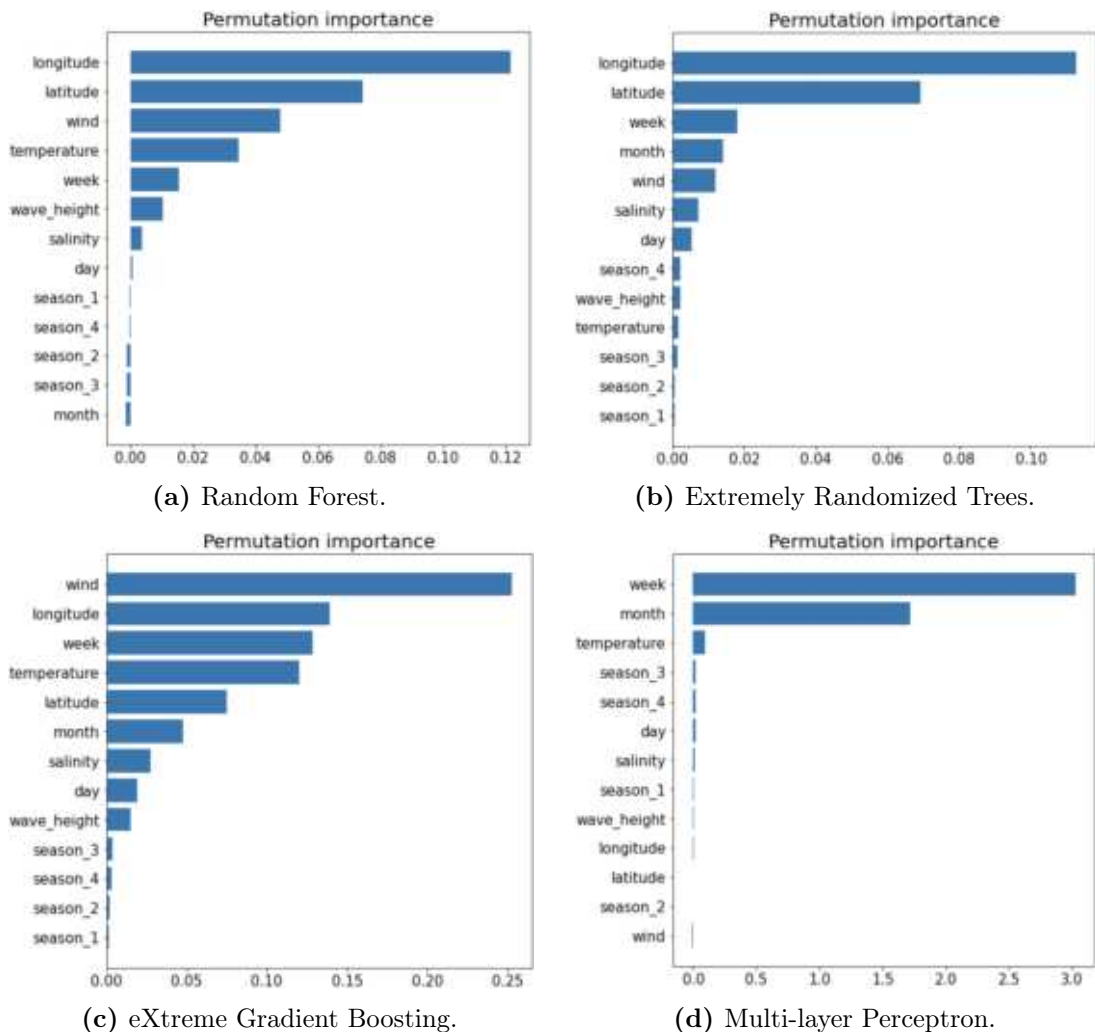


Figure 3.4: Permutation feature importance of the four machine learning models considering three years of training and one of testing.

In order to understand the importance that features have in forecasting, we considered the *Permutation Feature importance* method. This method measures the importance of a feature as follows. It computes the baseline coefficient of determination (R^2) score (if we have a regression problem) by passing a validation set through the machine learning method. Then it permutes the column values of a single predictor feature and it passes all test samples back through the machine learning method. Then it computes again the coefficient of determination, and finally the importance of that feature is defined as the difference between the baseline R^2 score and the one obtained after permuting the column. The larger this difference, the greater the importance of the feature.

In Figure 3.4 we can see the Permutation feature importance of the four machine learning methods considering three years of training and one of testing. We can observe how longitude and latitude are important features for the fishing effort forecast in all algorithms except for the Multi-layer Perceptron. Other features with high relevance for the forecast are temperature, week and wind. Features importance considering one year or two years as training dataset can be found in Appendix A.

ML model	Season	MAE	RMSE	R^2
Random Forest	1	0.34288	0.59689	0.60993
	2	0.23278	0.44007	0.71202
	3	0.18499	0.33447	0.70299
	4	0.29397	0.49519	0.67454
Extra Trees	1	0.34065	0.59862	0.60767
	2	0.24192	0.45078	0.69782
	3	0.18134	0.33225	0.70692
	4	0.30244	0.51496	0.64804
XGBoost	1	0.34496	0.60383	0.6008
	2	0.25722	0.48682	0.64757
	3	0.18512	0.35155	0.67187
	4	0.31094	0.52982	0.62743
Multi-layer Perceptron	1	0.51125	0.87958	0.15297
	2	0.35011	0.62568	0.41785
	3	0.22812	0.42971	0.50974
	4	0.38251	0.63475	0.46524

Table 3.8: Results for each season using three years (2016-2017-2018) as training dataset.

In Table 3.8 there are the results for the four machine learning models for each season considering three years of training and one of testing with respect to all the evaluation metrics examined (MAE, RMSE and R^2). It can be easily seen that Multi-layer Perceptron has a very low coefficient of determination score for all seasons compared to the other three models which have a value around 0.65. Furthermore, Random Forest has the best values for all seasons compared to the other algorithms. Regarding MAE and RMSE evaluation metrics (for which the best value is 0) we can make similar considerations. Multi-layer Perceptron achieves the worst results while Random Forest remains the most accurate model also in this case with a small difference compared to the other two models. Also in this case, in Appendix A there are the results for the other two training datasets considering all the evaluation metrics and each season.

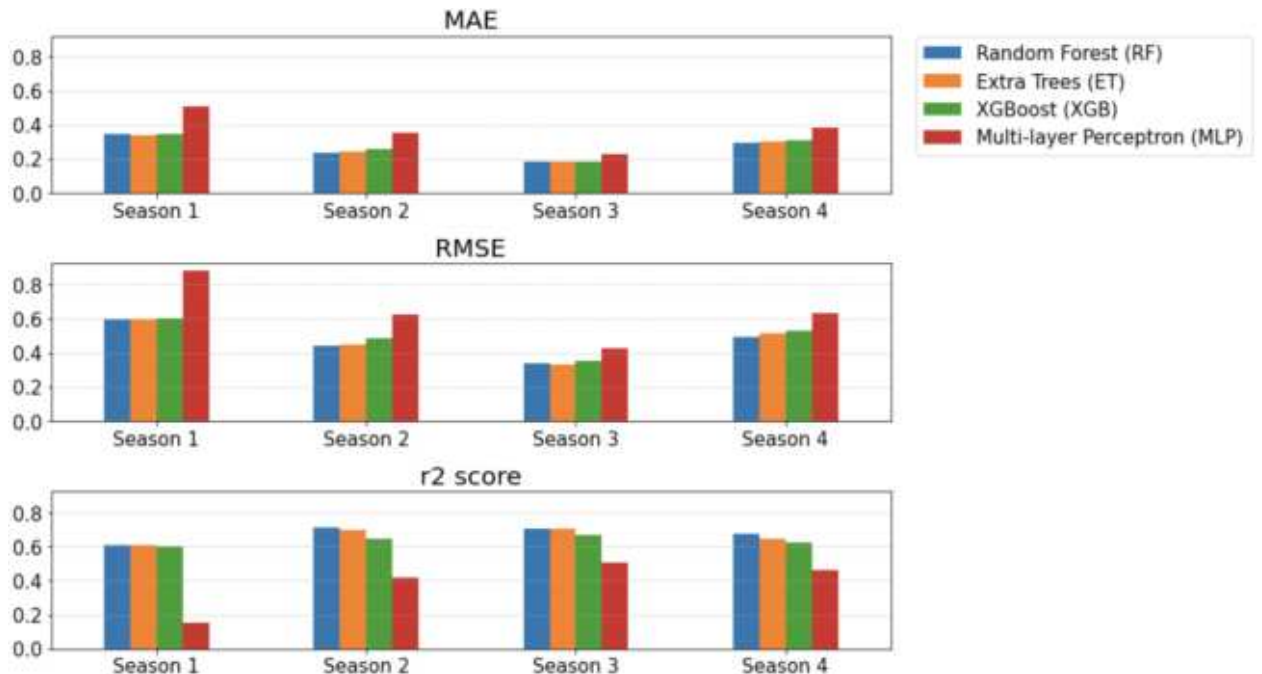


Figure 3.5: MAE, RMSE and R^2 scores for each season considering three years of training for the four machine learning models: RF (blue), ET (orange), XGBoost (green) and MLP (red).

In Figure 3.5 it is possible to understand clearly the results presented in Table 3.8. In particular, we can analyse in a better way the relationship between each machine learning method for each season and a given evaluation metric. For example, if we consider the Coefficient of determination we can see that Random Forest gets the best value, i.e. the closest to one, in all seasons except the third one in which, however, there is not a big difference compared to the resulting value obtained using Extra Trees. The same reasoning can be done for RMSE (where

the best value is 0) and the coefficient of determination (where the best value is 1). The bar charts considering one year and two years as training dataset, respectively, can be found in Appendix A.

In order to compare the three different training datasets and understand whether, and to what extent, more information from previous years could have a positive impact on the forecast, Table 3.9 shows the average season results considering all the machine learning methods and the three different training datasets used.

Training dataset	ML model	MAE	RMSE	R^2
2018	Random Forest	0.277695	0.484947	0.648415
	Extra Trees	0.292423	0.518442	0.599190
	XGBoost	0.356368	0.588965	0.461297
	Multi-layer Perceptron	0.467982	0.707218	0.199368
2017-2018	Random Forest	0.269230	0.468997	0.670283
	Extra Trees	0.274450	0.479762	0.658693
	XGBoost	0.272552	0.477483	0.659347
	Multi-layer Perceptron	0.337900	0.596575	0.453630
2016-2017-2018	Random Forest	0.263655	0.466655	0.674870
	Extra Trees	0.266588	0.474152	0.665112
	XGBoost	0.274560	0.493005	0.636918
	Multi-layer Perceptron	0.367997	0.642430	0.386450

Table 3.9: Average seasons results considering each training dataset.

We can observe that Random Forest is the best machine learning algorithm for our problem since all evaluation metrics get closer to the best case (0 for MAE and RMSE and 1 for R^2). Again for all three training datasets there is not much difference between Random Forest, Extra Trees and XGBoost while the difference is significant compared to Multi-layer Perceptron. As we can also see from Figure 3.6, in which there are the MAE, RMSE and R^2 scores resulting from the models trained only with 2018 (blue), 2017-2018 (orange), 2016-2017-2018 (green) datasets, it seems that considering more years of data, Random Forest and Extra Trees increase their accuracy in forecasting. Concerning XGBoost and Multi-layer Perceptron, if we consider two years instead of one as our training dataset, there is a significant improvement with respect to all evaluation parameters. However, this does not happen if we consider one more year (2016).

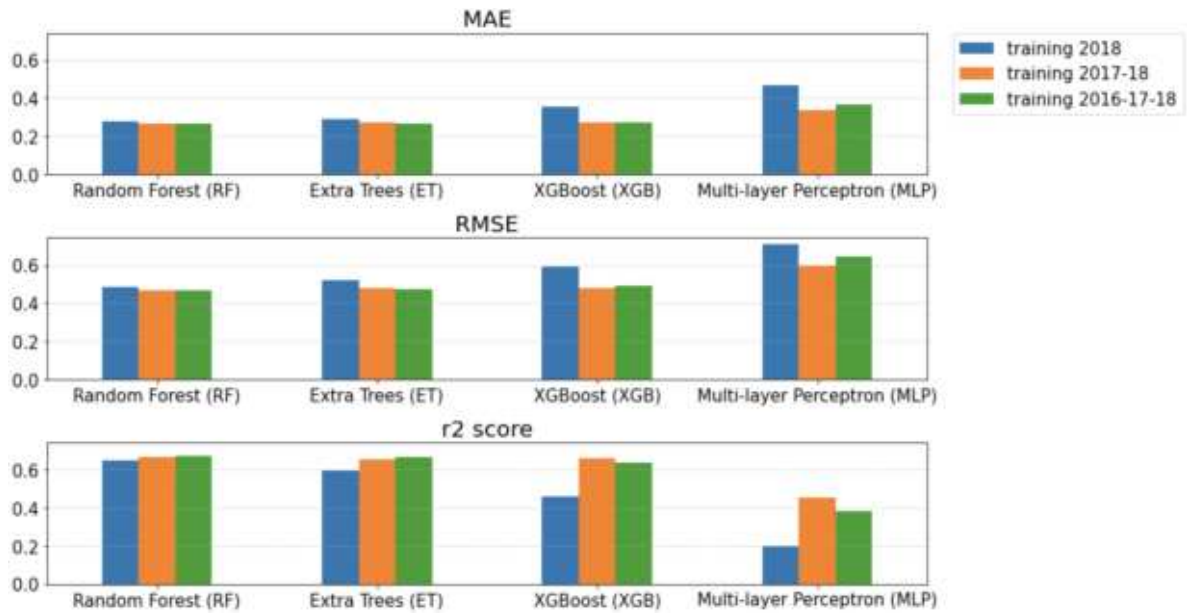


Figure 3.6: MAE, RMSE and R^2 scores resulting from the models trained only with 2018 (blue), 2017-2018 (orange), 2016-2017-2018 (green) datasets.

Figure 3.7 shows the fishing effort maps of the actual fishing effort values for 2019 and the predicted ones using the four machine learning models considering three years as training dataset. In Figure 3.7(a) we can see the actual fishing values for 2019 which corresponds to the ground truth. It is possible to observe that using Random Forest and also Extra Trees the forecast is very close to the actual fishing effort (Figure 3.7(b) and 3.7(c) respectively). In the rectangle delimited by dashed lines (Figures 3.7(a), 3.7(b), 3.7(c)) we can see how both Random Forest and Extra Trees are not so accurate in predicting values in areas where there is no over-exploitation. XGBoost fishing effort map (Figure 3.7(d)) does not correctly define the most exploited areas (or cells). Finally, using Multi-layer Perceptron (Figure 3.7(e)) this phenomenon is even more accentuated since there is not a marked difference between the various cells but the fishing effort seems to be spread uniform over the entire grid. In fact, using Multi-layer perceptron as machine learning model, in the predicted cells the value of the fishing effort never exceeds 7.624 (dark red cells), while in the actual fishing effort these cells (in which the fishing effort value is greater than 7.624) are 65. Moreover, the number of cells in the range between 3.038 and 7.624 are much fewer than the actual ones (the actual ones are 326, while in the Multi-layer perceptron we have only 10 cells in this range).

These results, especially those of Random Forest and Extra Trees, are considered a good achievement because the problem is challenging, given the fact that the

dataset contains multivariate and spatio-temporal aspects to cope with. In this work, autocorrelation has not been exploited much. For example, we could include among the dataset variables, the fishing effort in the surrounding areas, or in some previous periods of time. In Chapter 5 we will discuss possible future works.

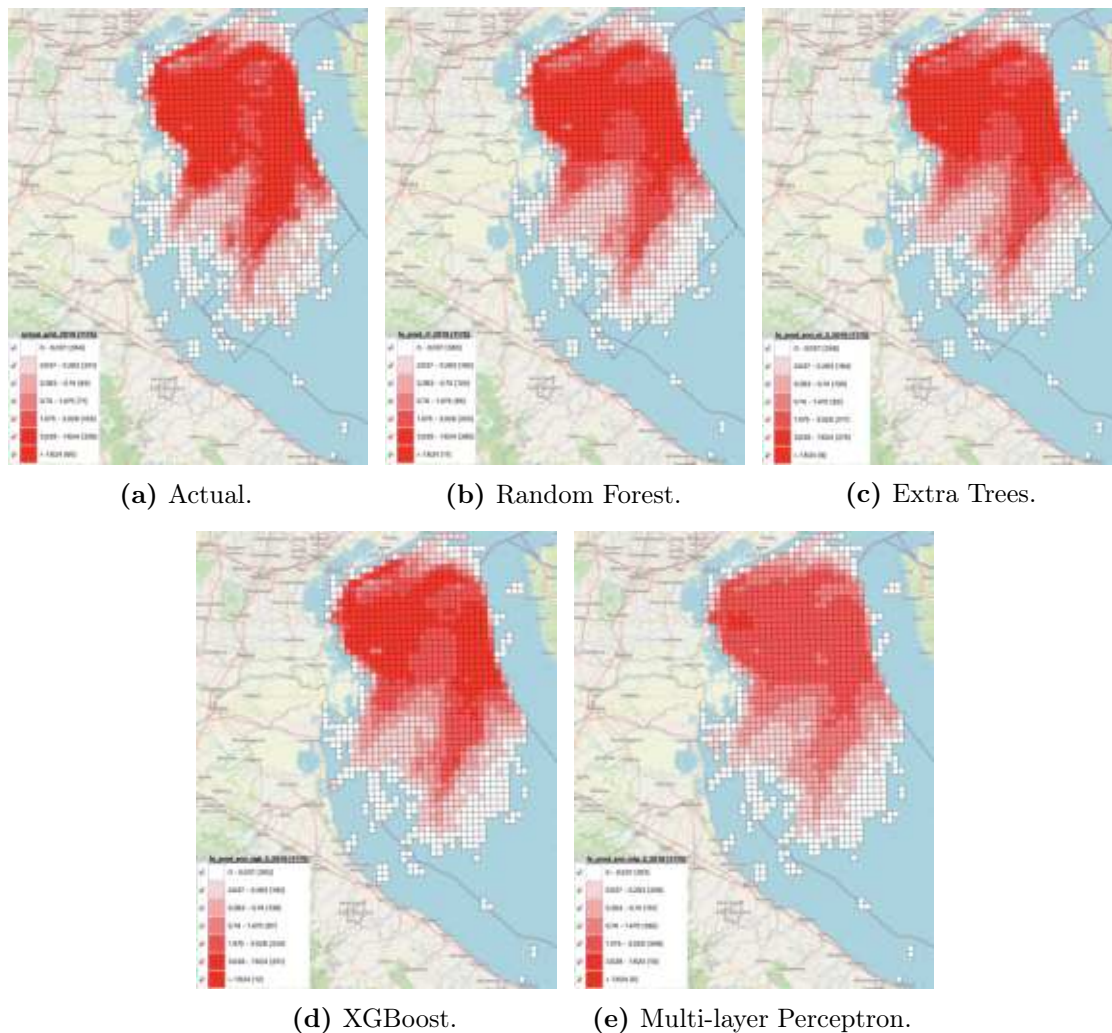


Figure 3.7: Fishing effort maps of the actual fishing effort values for 2019 and the predicted ones using the four machine learning models considering three years as training dataset.

Chapter 4

Underwater noise

Even in the most beautiful music, there are some silences, which are there so we can witness the importance of silence. Silence is more important than ever, as life today is full of noise. We speak a lot about environmental pollution but not enough about noise pollution.

– Andrea Bocelli

As we have extensively discussed in Section [2.1.2](#) the fish resources of seas and oceans are often recognized to be over-exploited. In order to make fishing activities sustainable and to guarantee a productive and healthy ecosystem, there is a strong need to develop adequate monitoring systems, including the development of a model for the propagation of underwater noise. In fact, the reconstruction of underwater sound propagation of vessels is an important challenge for the preservation of the marine environment, since it is known to be a primary contributor to anthropogenic noise in the seas. For this reason, in this chapter, an underwater noise modelling is presented. First of all, the underwater noise generated by a boat depending on the characteristics of its engine will be analyzed, then we will describe in detail the implementation of the model and the use of environmental features for the propagation of underwater sound, and finally we will discuss the results and we will analyze the underwater noise maps.

4.1 Engine noise from fishing vessels

Before implementing a model for the propagation of underwater sound, it is necessary to understand the underwater noise generated by a boat depending on the characteristics of its engine. There are many factors that influence this noise, such as, for example, the type of boat (trawler, gillnetter, tour boat, sailing boat, motor boat, etc.), the power of the engine, the position of the engine under the boat, the frequency considered, etc. Our AIS dataset includes only fishing boats but with different characteristics such as the length overall (LOA) of the boat, the horsepower of the engine and also the fishing gear used. Since we do not have the information on the underwater noise generated by all fishing vessels, we started from a fishing boat of which we know the horsepower, the LOA and the underwater noise it generates. Table 4.1 shows the characteristics of the trawler taken as reference boat (in Figure 4.1), a vessel that was used in the SOUNDSCAPE project [65].

Characteristics of the hull	
Length	13.70m
Breadth	4.2m
Draught	1.8m
Building material	wood
Type of hull	monohull
Characteristics of the engine	
Type of engine	IVECO
Number of engines	1
HP and type of mounting	220 <i>HP</i> inboard diesel
Power	109 <i>kW</i>
Maximum speed/ <i>rpm</i>	9.2kn/1900rpm (1:1.48 gearbox ratio)
Characteristics of the propulsion	
Number of propellers	1
Number of blades	4
Type of propeller	fixed
Speed/RPM during test runs	
Cruising speed	7kn/1100rpm
High speed	7.1kn/1200rpm

Table 4.1: Characteristics of the trawler [65].



Figure 4.1: Trawler used as reference fishing vessel [65].

In particular, we are interested in the horsepower of the boat which, in this case, is 220 *HP*. As already mentioned, the noise depends on the frequency in which it is measured, so the same object can generate a different noise depending on the frequency. Since the *European Marine Strategy Framework Directive* (MSFD) uses 63 *Hz* and 125 *Hz* frequencies as standard, we consider these two frequencies for the vessel noise measure. Table 4.2 shows the noise generated by the boat (source level) for the two frequencies. In our model we used the source level (SL) total correct (expressed in *dB*) with frequency 63 *Hz*, therefore we consider that this fishing vessel generates 153.6 *dB* while sailing.

Frequency (Hz)	Cruising speed: 7kn		High speed: 7.1kn	
	63	125	63	125
SL-Port - dCPA1 (dB)	150	146.6	149.6	146.7
SL-Starboard - dCPA1 (dB)	153	148	153.3	151.8
SL-Port - dCPA2 (dB)	155.4	147.9	157.7	147.8
SL-Starboard - dCPA2 (dB)	156.2	151.9	159.1	153.4
SL total correct (dB)	153.6	148.6	154.9	149.9

Table 4.2: Trawler source level with 63 *Hz* and 125 *Hz* at cruising speed and high speed [65].

Since we have no measurements of the noise produced by the boats in our dataset, we have made some assumptions. Clearly, as soon as we have precise measurements of the noise produced, it will only be necessary to change the input parameters in our model. Since the 69 different vessels in our dataset have a horsepower (*HP*) ranging from 146.9 to 1084.1 *HP*, we have decided to divide the

vessels into four groups according to their horsepower and multiply the underwater noise they generate by a multiplication factor that increases with the *HP* of the vessels. The four groups correspond to different ranges for horsepower and they are listed in Table 4.3 with the associated multiplicative factors and the sound level expressed in *dB*.

HP range	Multiplicative factor	dB
0 – 250	1	153.6
250 – 500	1.04	159.744
500 – 750	1.06	162.816
> 750	1.08	165.888

Table 4.3: Sound level for different horsepower ranges.

Furthermore, when the vessel is performing a fishing activity, the noise generated is more intense since there is a greater effort of the engine due to the use of the fishing gear, and also the fishing gear itself produces some noise. The fishing gear used by fishing vessels are the following: small bottom otter trawl (SOTB), large bottom otter trawl (LOTB), pelagic pair trawl (PTM) and rapido (RAP). The ecological experts suggested some factors to weigh the increasing noise during the fishing activity. In Table 4.4 the different types of fishing gear with the associated multiplication factors are shown.

Gear types	Multiplicative factor
SOTB	1.02
LOTB	1.04
PTM	1.02
RAP	1.06

Table 4.4: Sound level for different gear types.

For example, the boat named ADELINDA has horsepower equal to 276 *HP* and it uses the RAP as fishing gear. Consequently, when it is sailing its source noise level is assumed to be 159.74 *dB*, while when it is fishing, the source level is assumed to be 169.33 *dB*.

4.2 Regular grid with environmental features

In order to distribute the underwater sound produced by boats, it is necessary to partition the northern Adriatic Sea into a regular grid. For this work, the cells of the regular grid have a size of $1\text{km} \times 1\text{km}$, as we can see in Figure 4.2. The regular grid is also necessary for the introduction of environmental data. In fact, the grid is not only spatial but spatio-temporal: each cell for each day is characterized by different environmental data that change on a daily basis.

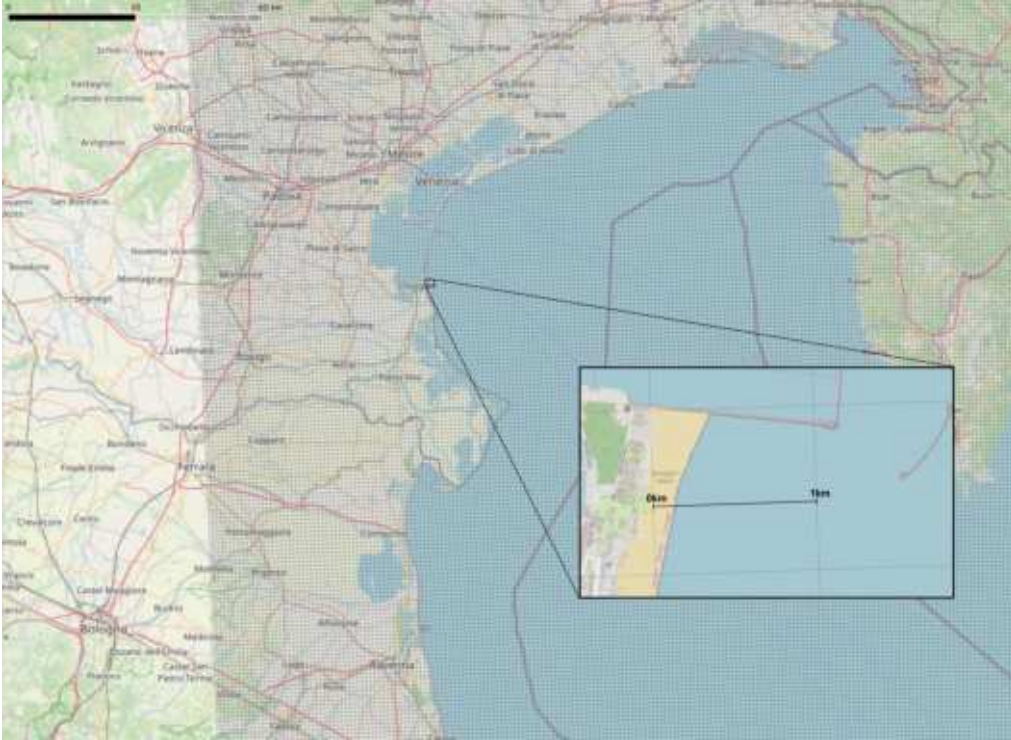


Figure 4.2: Regular grid in which each cell has a size of 1×1 kilometer.

Environmental information was obtained thanks to the *Copernicus Programme of the European Union*, as we have already described in Chapter 3. The environmental variables necessary to calculate the propagation of underwater sound in our model (which will be presented in detail in Section 4.4) are the following:

- daily sea surface temperature (expressed in celsius);
- daily sea salinity (in psu, Practical Salinity Units);
- depth of the sea, which we consider as a constant (35 meters);
- sea water potential of Hydrogen (pH).

The data of these environmental features downloaded from Copernicus are spatio-temporal and, therefore, for each cell, we have different values for our features. Clearly the size of the cells does not correspond to the one chosen for our work ($1km \times 1km$): for example, the information relating to the sea surface temperature is characterized by a grid size of four kilometers. For this reason, we first made a correction of the data in order to have all the daily environmental information on the regular grid of the chosen size (as described in Section 3.2).

4.3 Propagation modelling

The basic objective of noise modelling is to assess how much noise a particular activity will generate in the surrounding area [31]: the aim is to model the received noise level (RL) at a given point (or points), based on the sound source level (SL) of the noise source, and the amount of sound energy which is lost as the sound wave propagates from the source to the receiver (transmission loss or propagation loss, TL). The relationship between these quantities is encapsulated in the classic sonar equation [86]:

$$RL = SL - TL \quad (4.1)$$

This straightforward expression is fundamental to modelling underwater noise, and its simplicity belies considerable complexity in the task of computing the transmission loss in order to estimate the received noise.

Sound propagation is profoundly affected by some factors such as the conditions of the surface and bottom boundaries of the sea as well as by the variation of sound speed within the ocean volume [26]. The most important acoustical variable in the sea is sound speed. Air has a density 800 times lower than the density of water, therefore a sound that propagates inside the water has a higher propagation speed, equal to about $1500 m/s$, against about $340 m/s$ of air [40]. Sound propagation speed is influenced by various chemical-physical factors such as temperature, salinity and pressure [71], varying both during the day and with the seasons in the superficial part [69], and with depth. In particular:

- *temperature* is greatly influenced by the solar radiation; an increase in temperature of $1^\circ C$ determines an increase in the speed of sound of 3-4 m/s ;
- *salinity* is influenced by the supply of fresh water from rivers, lakes and the melting of ice; the increase of one unit of salinity compared to the average value of 35 psu (Practical Salinity Units) increases the speed of sound by

1.1 m/s ;

- *pressure* is linearly related to depth, as the increase of 100 m in depth (about 100 kPa) increases the speed of sound by 1.7 m/s .

The computation of the transmission loss considering all these parameters is not a simple task and for this reason various models have been introduced.

Transmission loss

In the simplest scenario, transmission loss is often estimated using straightforward spreading laws of the form:

$$TL = N \times \log_{10}(R) \quad (4.2)$$

where R is the distance from the noise source in meters, and N is a scaling factor [31]. Clearly, this simplistic approach does not take into account all the environmental characteristics that are fundamental for a more accurate estimation of transmission loss, therefore it can only be used in uncomplicated propagation scenarios.

In more realistic models, what we want to consider are all the environmental aspects that influence the sound propagation underwater. In particular, instead of considering only the distance R (in meters) and a scaling factor (N), [85] presents a model for the calculation of the transmission loss that considers the absorption of sound coefficient (α expressed in dB/km) to include environmental features in its estimation:

$$TL = N \times \log_{10}(R) + \alpha \times R \quad (4.3)$$

In the literature there are three main models for predicting the absorption of sound in sea water which retain the essential dependence on temperature, pressure, salinity, acidity and other environmental features.

In the Francois and Garrison model [35] the general equation for the absorption of sound in sea water, at a given frequency f (in kilohertz), is given as the sum of contributions from boric acid, magnesium sulfate, and pure water (contributions from other reactions are small and are not included):

$$\alpha = \frac{A_1 P_1 f_1 f^2}{f_1^2 + f^2} + \frac{A_2 P_2 f_2 f^2}{f_2^2 + f^2} + A_3 P_3 f^2 \quad dB \ km^{-1} \quad (4.4)$$

The first two terms represent chemical relaxation processes, the first for boric acid and the second for magnesium sulfate. The third term represents the absorption

from pure water. The pressure dependences are given by P_1 , P_2 , and P_3 and the relaxation frequencies are f_1 and f_2 (expressed in kHz). A_1 , A_2 (both expressed in $dB km^{-1} kHz^{-1}$) and A_3 (expressed in $dB km^{-1} kHz^{-2}$) were originally intended as constants but there is experimental evidence that they vary with water properties [36]. The three contributions from boric acid, magnesium sulfate, and pure water are computed as follows:

- Boric acid Contribution

$$A_1 = \frac{8.86}{c} \times 10^{(0.78pH-5)}$$

$$P_1 = 1$$

$$f_1 = 2.8(S/35)^{0.5} \times 10^{(4-1245/\theta)}$$

where c is the sound speed (m/s), given approximately by:

$$c = 1412 + 3.21T + 1.19S + 0.0167D$$

T is the temperature ($^{\circ}C$), $\theta = 273 + T$, S is the salinity (‰) and D is the depth (m).

- $MgSO_4$ Contribution

$$A_2 = 21.44 \times \frac{S}{c} \times (1 + 0.025T)$$

$$P_2 = 1 - 1.37 \times 10^{-4}D + 6.2 \times 10^{-9}D^2$$

$$f_2 = \frac{8.17 \times 10^{(8-1990/\theta)}}{1 + 0.0018(S - 35)}$$

- Pure Water Contribution

– for $T \leq 20^{\circ}C$

$$A_3 = 4.937 \times 10^{-4} - 2.59 \times 10^{-5}T + 9.11 \times 10^{-7}T^2 - 1.50 \times 10^{-8}T^3$$

– for $T > 20^{\circ}C$

$$A_3 = 3.964 \times 10^{-4} - 1.146 \times 10^{-5}T + 1.45 \times 10^{-7}T^2 - 6.5 \times 10^{-10}T^3$$

$$P_3 = 1 - 3.83 \times 10^{-5}D + 4.9 \times 10^{-10}D^2$$

Fisher and Simmons [34] presented another equation for the computation of sound absorption in sea water as a function of frequency, temperature, and pressure based on laboratory data. Simmons showed that his results for absorption in Lyman and Fleming sea water (i.e. $S = 35$ psu, $pH = 8$) due to $MgSO_4$ agreed well with those he calculated from data in [91] for absorption in natural sea water. As in [35, 36] sound absorption in sea water may be considered as the sum of the absorption due to boric acid, magnesium sulfate and water. When using water other than the Lyman and Fleming standard, the Fisher and Simmons algorithm is invalid. Assuming the three contributions are linearly additive, the total absorption coefficient can be written as:

$$\alpha = \frac{A_1 f_1 f^2}{f_1^2 + f^2} + \frac{A_2 P_2 f_2 f^2}{f_2^2 + f^2} + A_3 P_3 f^2 \quad m^{-1} \quad (4.5)$$

where the A_i are atmospheric pressure values, the P_i describe the pressure dependence of the A_i , f is the acoustic frequency and f_i the relaxation frequency. In particular, the values are calculated as follows:

$$\begin{aligned} A_1 &= 1.03 \times 10^{-8} + 2.36 \times 10^{-10}T - 5.22 \times 10^{-12}t^2 \\ f_1 &= 1.32 \times 10^{-3} \times (T + 273.1) \times e^{-1700/(T+273.1)} \\ A_2 &= 5.62 \times 10^{-8} + 7.52 \times 10^{-10}T \\ f_2 &= 1.55 \times 10^7 \times (T + 273.1) \times e^{-3052/(T+273.1)} \\ P_2 &= 1 - 10.3 \times 10^{-4}P + 3.7 \times 10^{-7}P^2 \\ A_3 &= (55.9 - 2.37T + 4.77 \times 10^{-2}T^2 - 3.48 \times 10^{-4}T^3) \times 10^{-15} \\ P_3 &= 1 - 3.84 \times 10^{-4}P + 7.57 \times 10^{-8}P^2 \end{aligned}$$

where f and f_i are in Hz , T in degree centigrade, P is in atm and A_i in $sec m^{-1}$. In order to convert α to dB/km , it is necessary to multiply it by 8.686.

Finally, Ainslie and McColm [2] presented a simplified expression for viscous and chemical absorption in sea water. Ocean sound is attenuated by two main mechanisms: chemical relaxation effects (due to boric acid at low frequency and magnesium sulphate at intermediate frequencies up to a few 100 kHz) and viscous absorption, which is significant at high frequency. In order to simplify the Francois and Garrison model [35, 36], Ainslie and McColm set temperature, salinity and acidity to reference values of $T = 0^\circ C$, $S = 35 ppt$ (parts per thousand), and $pH = 8$ respectively. This immediately simplifies the boric acid and magnesium

sulphate relaxation frequencies (in kHz):

$$f_1 = 0.78(S/35)^{1/2}e^{T/26} \quad (4.6)$$

$$f_2 = 42e^{T/17} \quad (4.7)$$

for boron and magnesium respectively. Applying a similar analysis to the coefficient terms, the authors obtained the following value for the absorption coefficient in sea water in dB/km (at a given frequency f in kHz , a depth z in km and a temperature T expressed in $^{\circ}C$):

$$\begin{aligned} \alpha = & 0.106 \frac{f_1 f^2}{f^2 + f_1^2} e^{(pH-8)/0.56} \\ & + 0.52 \left(1 + \frac{T}{43}\right) \left(\frac{S}{35}\right) \frac{f_2 f^2}{f^2 + f_2^2} e^{-z/6} \\ & + 0.00049 f^2 e^{-(T/27+z/17)} \end{aligned} \quad (4.8)$$

This simplified formula retains reasonable accuracy (to within 10% of Francois and Garrison model between 100 Hz and 1 MHz) for the following oceanographic conditions:

- $-6 < T < 35^{\circ}C$ ($S = 35$ ppt, $pH = 8$, $D = 0$ km)
- $7.7 < pH < 8.3$ ($T = 10^{\circ}C$, $S = 35$ ppt, $D = 0$ km)
- $5 < S < 50$ ppt ($T = 10^{\circ}C$, $pH = 8$, $D = 0$ km)
- $0 < D < 7$ km ($T = 10^{\circ}C$, $S = 35$ ppt, $pH = 8$)

In order to calculate how the sound propagates underwater, we use the Francois and Garrison model since it does not make any assumptions on the parameters and it provides a good approximation of the underwater sound even in different scenarios.

4.4 Underwater sound propagation model

Before describing in detail the implementation of the underwater sound propagation model, in this section we want to discuss the formulas for calculating the noise received at a given distance and the fundamental role that background noise plays in water. The received noise level (RL) at a given point is computed starting

from Equation (4.1). The problem is that the background noise, which is present in the marine environment, is not taken into account in this formula. In fact, when the noise generated by a source is the same as that of the background, this noise is no longer distinguishable and, consequently, it is no longer perceived. The formula for calculating the received noise level (RL) then becomes the following:

$$RL = SL - TL - AN \quad (4.9)$$

where SL is the sound source level, TL the transmission loss and AN the ambient noise. In [70] the authors measure the ambient noise in the Mediterranean Sea, and in Table 4.5 their measurements (expressed in dB) for each month with respect to two frequencies (63 and 125 Hz) are reported.

Month	63 Hz	125 Hz
January	97.9	100.3
February	101	101.8
March	103.3	103.7
April	101.1	102.6
May	98.3	101.7
June	99.17	102.2
July	103.3	103.6
September	103.1	103.9
October	101.8	103.3
November	100.4	101.7
December	98.4	100.4
Average	100.7	102.3

Table 4.5: Monthly average evolution of the mean of the 1/3 octave bands at 63 and 125 Hz of ambient noise measurements expressed in dB [70].

The month of August is not present in Table 4.5, because of the fishing ban that was in force during that time.

At this point we know all the parameters needed to compute the sound received at a given distance following Equation (4.9). The source level SL changes according to the vessel considered and whether it is sailing or fishing. The ambient noise changes depending on the month, and finally the transmission loss is computed according to Equation (4.3), considering the Francois and Garrison model (discussed

in Section 4.3).

4.5 Model implementation

The development of a model for underwater sound propagation requires the construction of multiple aspect trajectories starting from the AIS data. Hence, the previous reconstruction and enrichment of trajectories performed in [74, 14] turns out to be fundamental also in this context. In particular, in [74, 14] we first reconstructed all the trajectories of the fishing boats starting from their AIS data. Subsequently, we segmented the trajectories and then moved on to their enrichment. In particular, we are interested in assigning vessel's activity: in port, entering to/exiting from the port, fishing or navigation (as described in Section 2.2.3). This is of fundamental importance for the underwater sound propagation model we want to develop, because when a boat is fishing it produces a much more intense sound.

For the underwater noise propagation model we take the AIS data of 2019 and 2020, so we used the model developed in [74] to construct the multiple aspect trajectories using these new years of data. AIS data is not transmitted at regular intervals: there may be an interval of a few seconds to a few minutes between two AIS transmissions. Furthermore, analyzing the trajectories, we have noticed how in certain cases, due either to a malfunction of the transmission or to a voluntary interruption, between one transmission and another there was a time interval of more than half an hour (in some cases even a few hours).

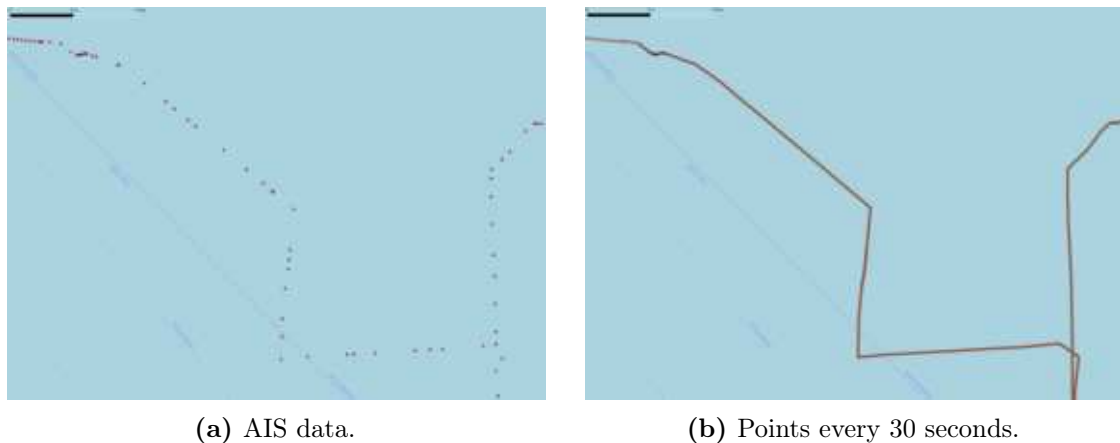


Figure 4.3: Part of the trajectory 389 carried out by the CARLO ALBERTO II boat with the AIS transmissions and the reconstruction of the points every 30 seconds of the same trajectory.

Since the sound propagates continuously in the water, to solve the problem of

the non-homogeneity of the AIS data transmission, we considered the reconstructed trajectories of the fishing boats and we got their position every 30 seconds. In Figure 4.3 we can see how the AIS data of part of the trajectory 389 carried out by the CARLO ALBERTO II boat are not homogeneous (Figure 4.3(a)), while in Figure 4.3(b) the reconstruction of the same trajectory with the points every 30 seconds allows for greater continuity in the movement of the boat during its trip. In Table 4.6 there are the number of vessels, the number of trips, the number of AIS data and the number of points resulting after splitting the trajectory every 30 seconds. As expected, the points we are going to use for the sound propagation model are plenty more than the simple AIS data (6, 588, 990 points more for 2019 and 8, 339, 289 more for 2020).

Year	Vessels	Trips	AIS data	Points every 30 seconds
2019	69	6, 570	16, 627, 201	23, 216, 191
2020	66	5, 285	10, 385, 742	18, 725, 031

Table 4.6: Number of vessels, trips, AIS data and points relative to the position of the boat every 30 seconds for the years 2019 and 2020.

The partition of the trajectory into points every 30 seconds is possible thanks to the function `valueAtTimestamp` defined in the *MobilityDB*¹ (described in Section 2.2.2). Once we have these new points, we need to assign to each point the segment it belongs to. This is of fundamental importance in order to be able to detect the activity carried out by each vessel at that point. In fact, once a point is associated with the segment, it is possible to trace the activity of the fishing vessel. For the noise propagation, however, we are only interested in distinguishing when the vessel is fishing or navigating. Consequently, in addition to attributing the correct segment to the point under consideration, if in that segment the boat is fishing (light blue segments in Figure 4.4(b)), then also all points on that segment are labeled with the fishing activity, otherwise the points are assigned as navigation (segments in red in Figure 4.4(a)).

Each point belongs not only to a segment but also to a cell ($1km \times 1km$) of the regular spatio-temporal grid, therefore it was necessary to assign each point to the corresponding cell. This is of paramount importance because each cell has different environmental characteristics every day. As described in Section 4.2, each cell has the environmental features that are used to compute the coefficient of absorption α

¹<https://www.mobilitydb.com/>

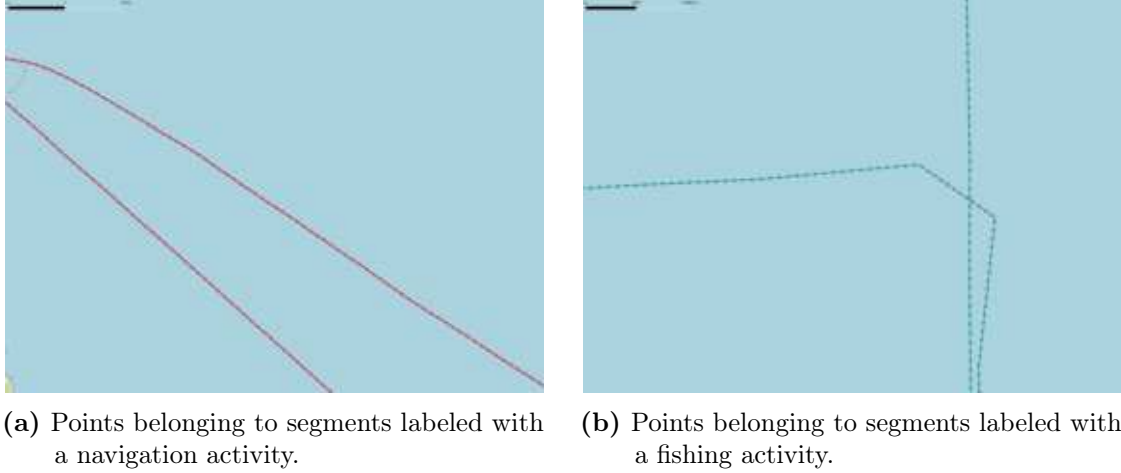


Figure 4.4: Points belonging to segments of a trip of the CARLO ALBERTO II boat.

of Equation (4.3). This means that, before being able to compute the underwater sound propagation for each point, we have to calculate this coefficient for each cell and for each day, because it will be used for the transmission loss computation.

At this point we can calculate the propagation of underwater sound given a noise source (the fishing boat). In particular, given a point of a trajectory, we want to understand what is the sound propagation radius, i.e. what is the distance at which the noise generated by the fishing vessel is no longer perceptible (or rather, it is no longer possible to distinguish it from the background noise). To calculate this distance (the radius R expressed in meters) we used Equation (4.9), where the received noise (RL) is equal to 0:

$$0 = SL - TL - AN \quad (4.10)$$

The source level (SL) and the ambient noise (AN) are two constants, then the transmission loss (TL) is given by Equation (4.3), so we will have:

$$\begin{aligned} SL - AN &= TL \\ &= N \times \log_{10}(R) + \alpha \times R \end{aligned} \quad (4.11)$$

The greatest contribution to the calculation of the transmission loss is given by $N \times \log_{10}(R)$, since the coefficient of absorption α is a very small number (on the order of 10^{-7}). For this reason, for the calculation of the radius R we will have:

$$N \times \log_{10}(R) = SL - AN \quad (4.12)$$

and finally the radius is given as follows:

$$R = 10^{(SL-AN)/N} \quad (4.13)$$

where $N = 20$ [85]. For example, the NUOVA TIRRENIA boat, while it is sailing in the month of June, in which we have a background noise equal to 99.17 dB and the noise of the source is equal to 159.744 dB , the maximum distance in which the sound generated by the boat is still perceptible is 1068.3173 meters.

Since we want to distribute the noise on a regular grid, the centroids of the cells of the grid have been considered as listening points (we have 4,195 of these points), and consequently the noise “perceived” by a centroid is then the same that is considered in the cell at that instant of time. Therefore, for each point we compute the sound propagation radius R (as previously described) and we construct the circle of radius R . All the centroids within the considered circle correspond to points where the noise generated by the boat is still perceptible. In Figure 4.5(a) we can see an example of the maximum spread of sound propagation and the centroids of the grid where the sound generated by the vessel is still distinguishable and therefore are points within the circle. At this point we consider only the cells corresponding to those centroids inside the circle to compute the sound inside the cells (as we can see in Figure 4.5(b)), since in the other cells of the grid the perceived noise is zero (or indistinguishable from the background noise).



(a) Circular region in which the underwater sound generated by a vessel propagates.

(b) Regular grid in which the underwater sound generated by a vessel propagates.

Figure 4.5: Regions in which the sound generated by a vessel propagates underwater, considering the circle with the maximum propagation radius and the cells of the regular grid in which the sound will then be distributed.

After finding the cells in which the vessel’s underwater sound propagates at that

instant of time, we can compute the noise received in each cell with Equation (4.9). The transmission loss is calculated according to Equation (4.3). Since we have to compute the logarithm of the radius R , without loss of generality, we assume that if the radius is less than 1, the transmission loss will be 0 (otherwise we would have a negative value, and this is unfeasible).



Figure 4.6: Underwater sound propagation in an instant of time of the boat CARLO ALBERTO II which is carrying out a fishing activity.

In Figure 4.6 we can see an example of the underwater sound propagation. The fishing boat CARLO ALBERTO II at a time instant ‘2019-09-09 00:00:00’ is carrying out a fishing activity with the gear RAP. It is possible to observe how the sound is clearly more intense in the cells closest to where the fishing boat is and the sound becomes less and less intense as the distance from the boat increases.

Using the *QGIS TimeManager* it is possible to generate animations in which, by selecting the vessels and the number of information per frame, we can visualize these vessels moving in the northern Adriatic Sea. Figure 4.7 shows the sequence of this animation, in which the movements of the fishing vessel CARLO ALBERTO II can be seen every 30 seconds (starting from ‘2019-01-30 01:02:30’ to ‘2019-01-30 01:06:30’). The fishing vessel is represented with a light blue dot if it is fishing (from Figure 4.7(f) to Figure 4.7(i)), red otherwise (from Figure 4.7(a) to Figure 4.7(e)). Moreover, from Figure 4.7 we can easily observe how the noise generated by the fishing boat when it is fishing is much more intense than when it is sailing, and consequently the sound propagates much more.

In order to understand how the sound generated by all fishing vessels propagates during the day, a more adequate visualisation is the one in Figure 4.8. This figure depicts a time instant of an animation produced with *QGIS TimeManager*, in which we can see all the vessels labeled with a light blue dot when they are fishing, red

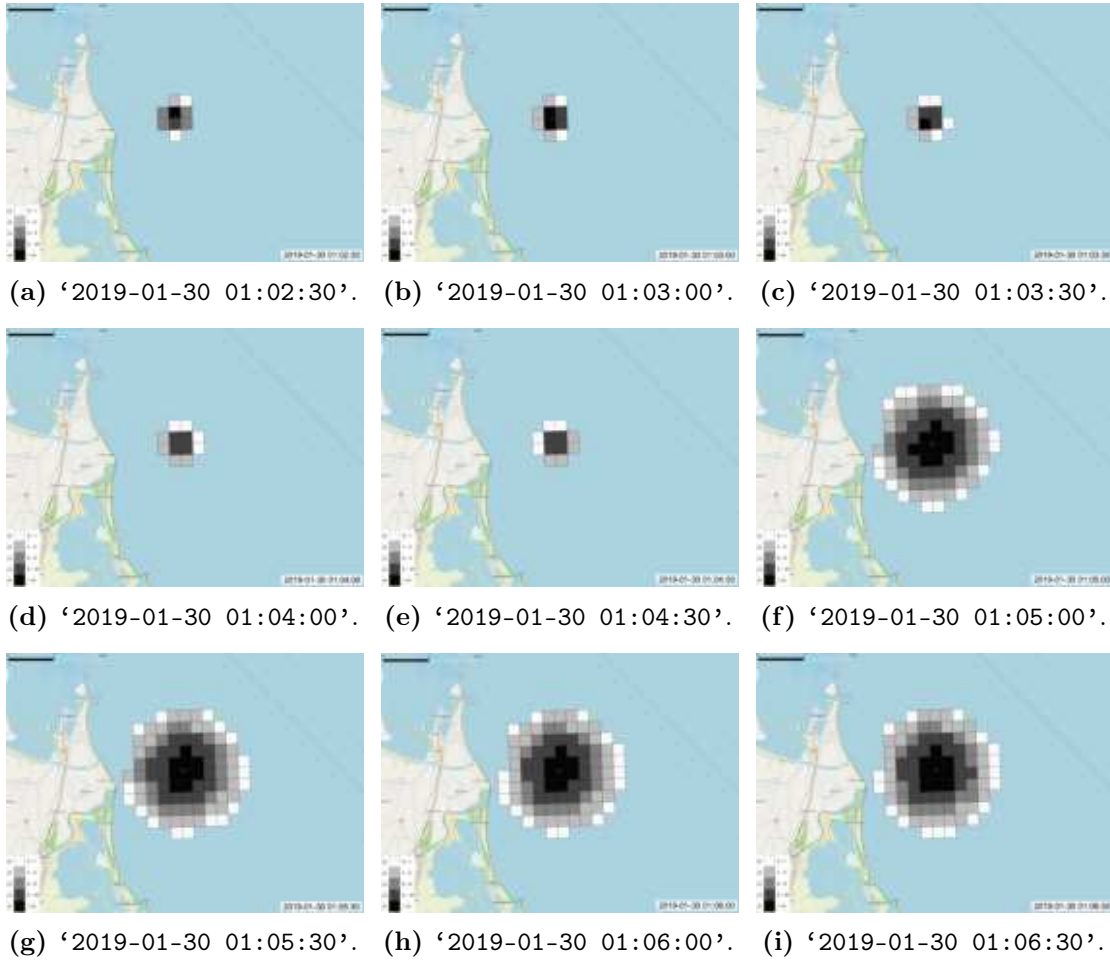


Figure 4.7: Propagation of the underwater noise of the *CARLO ALBERTO II* fishing vessel while it is navigating (red dot) or fishing (light blue dot) every 30 seconds.

otherwise. Furthermore, the noise generated by a boat is considered in isolation, i.e. the simultaneous presence of two boats in the same cell and in the same time instant does not affect on the underwater sound propagation of the single boat. Only when we build the sound propagation map for each time instant and cell, if two boats are in the same cell in the same time instant, the sound perceived by the listening point of that cell is the sum of the noise produced by each boat.

In Figure 4.8 we can observe the different underwater noise propagation between the various fishing vessels. In particular, the two boats highlighted by circles *A* and *B* in orange, are both fishing, have the same horsepower (590.9 *HP*) but they use different gears: the boat in *A* uses the LOTB gear, while the one in *B* the RAP. The boats using the RAP gear generate a much more intense noise (Table 4.4), in fact the cells in which the noise is propagated for the boat using the RAP gear (*B*) are 94, while for the one using LOTB are 44 (*A*). Conversely, there may be boats that are fishing, use the same gear but their engines have different horsepower. This

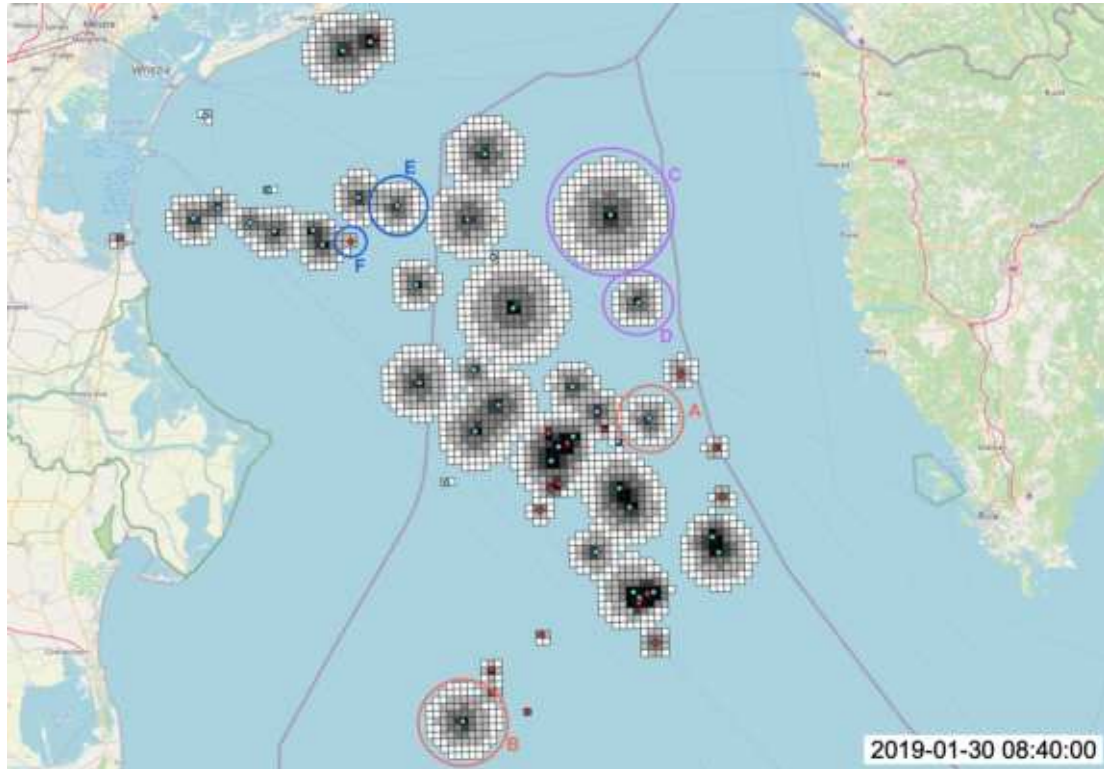


Figure 4.8: Propagation of underwater noise of fishing vessels in the northern Adriatic Sea on the 30th of January 2019 at 08:40. Vessels are shown with a light blue dot if they are fishing, red otherwise.

is precisely the case for boats in circles *C* and *D* (in light purple), and therefore the only difference between the two boats is the power of the engine: boat in *C* has an engine with 1084.1 *HP*, while the one in *D* with 296.1 *HP*. We can see how the sound propagates in a different way according to the different horsepower of the engine (the cells in which the noise is propagated for the boat in *C* are 196, while for the one in *D* are 43), and therefore the more horsepower, the more the sound will propagate underwater (Table 4.3). Finally, in circles *E* and *F* (in blue) there are respectively two boats that have the same number of horsepower but one is fishing (the one in *E*) while the other is not (in *F*). Also in this case the difference is considerable: if the boat is fishing the sound propagates more (in this case the cells in which the sound is perceptible are 44), otherwise the sound is limited in the region in which the boat is navigating (the cells in which the sound propagates are 4). Using this visualization we can observe how the underwater sound propagation is different when the boats are fishing or sailing, and also the dissimilarity of underwater sound propagation between vessels that are fishing but use a different fishing gear or, while using the same gear, they have more/less horsepower and, consequently, the propagation is more/less widespread.

4.5.1 Underwater sound propagation maps

After the creation of a simple underwater sound propagation model, it is of fundamental importance for environmental experts to be able to study and analyze underwater sound propagation maps in order to understand how noise can impact marine fauna. Consequently, we have built two different types of maps. In particular, Figure 4.9 reports the peak maps for each month: each cell is characterized by the maximum noise detected in that cell (the peak of the month in the cell). Figure 4.10 reports the average underwater noise value measured in each cell for all the months of 2019. The average value for a cell is calculated by summing up the underwater noise measured every 30 seconds in that cell and dividing it by the number of measurements.

In order to perform a more quantitative analysis based on the maps in Figure 4.9 and 4.10, Tables 4.8 and 4.10 show, for each month, the number of cells for each category. In particular, as regards the peak and average maps, the categories are divided by range of underwater noise measured, expressed in decibel (*dB*), as depicted in Table 4.7. Table 4.9 and 4.11 show, for each month, the percentages of cells characterized by a value in the various noise ranges for the peaks and the averages respectively.






Symbol	Range (dB)	
	Peak	Average
	0 – 10	0 – 2.5
	10 – 25	2.5 – 3.5
	25 – 35	3.5 – 4.5
	35 – 45	4.5 – 5.5
	> 45	> 5.5

Table 4.7: Underwater noise range for maps and analysis of peaks and averages.

First of all, from the underwater noise maps it is possible to observe the area of the northern Adriatic Sea where the fishing boats carry out their fishing activities. The maps in Figure 4.9 show that there are evident temporal differences with the areas characterized by very high noise peaks: in the months of January, May, June and December there are many areas with a high concentration of noise, while in the months of March, July and September these areas decrease, and therefore in these months there are fewer areas characterized by intense noise peaks. Table 4.9 underlines the latter observation.

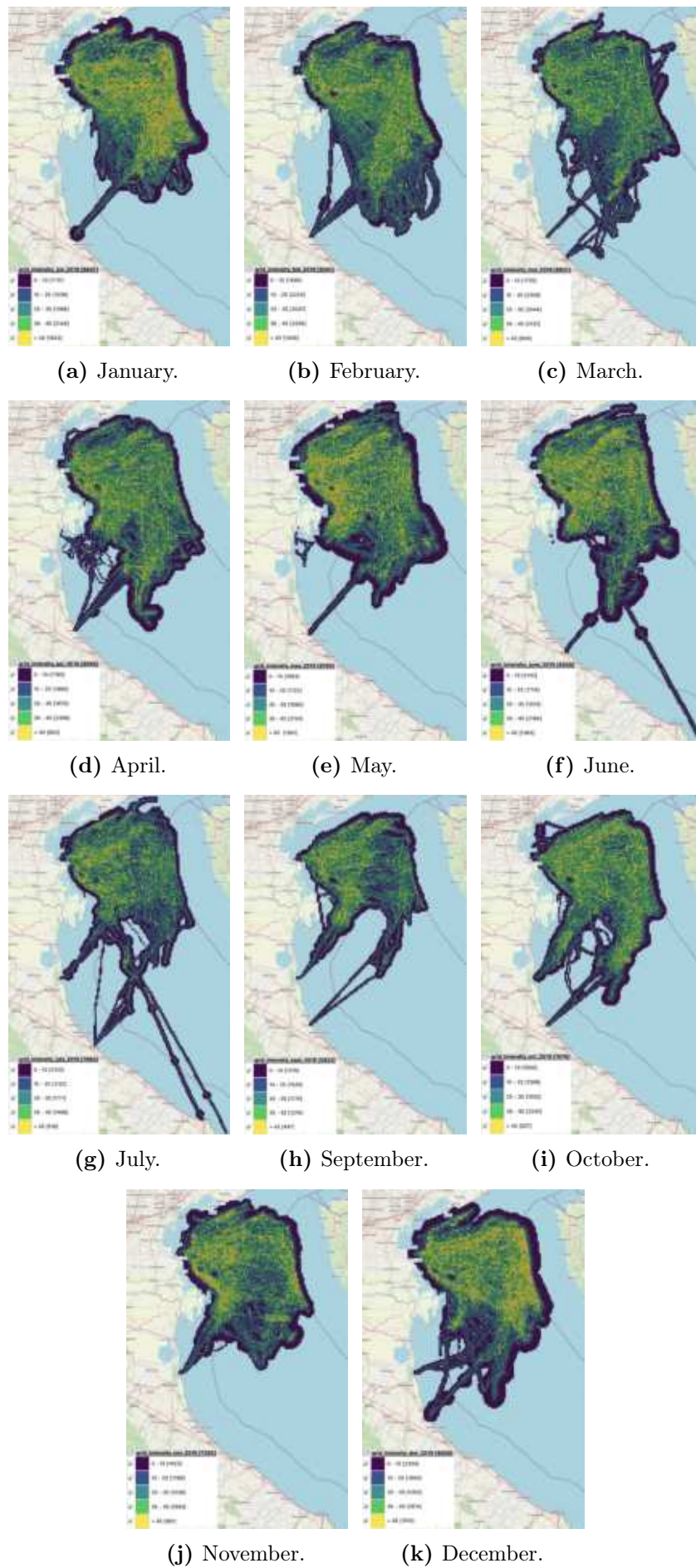


Figure 4.9: Underwater noise peak maps for each month of 2019.

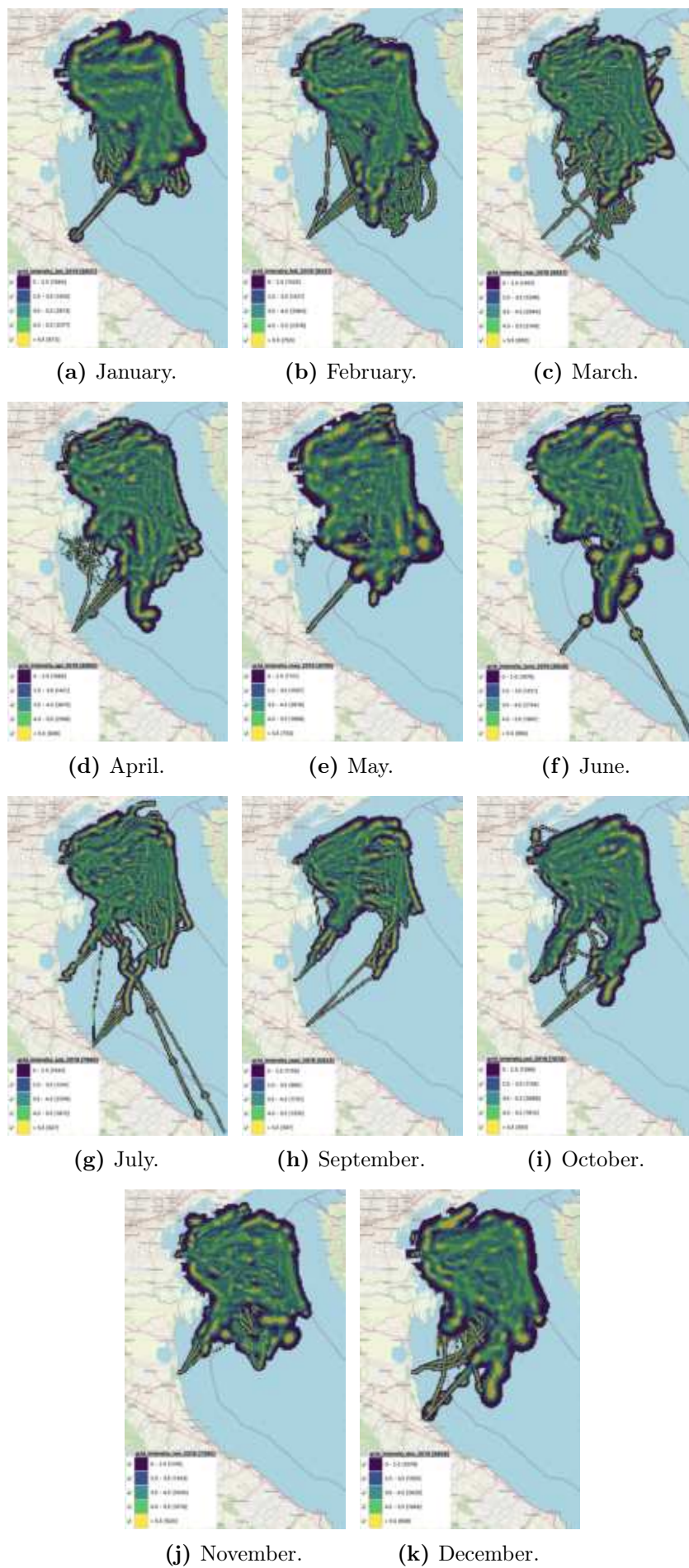


Figure 4.10: Underwater noise average maps for each month of 2019.

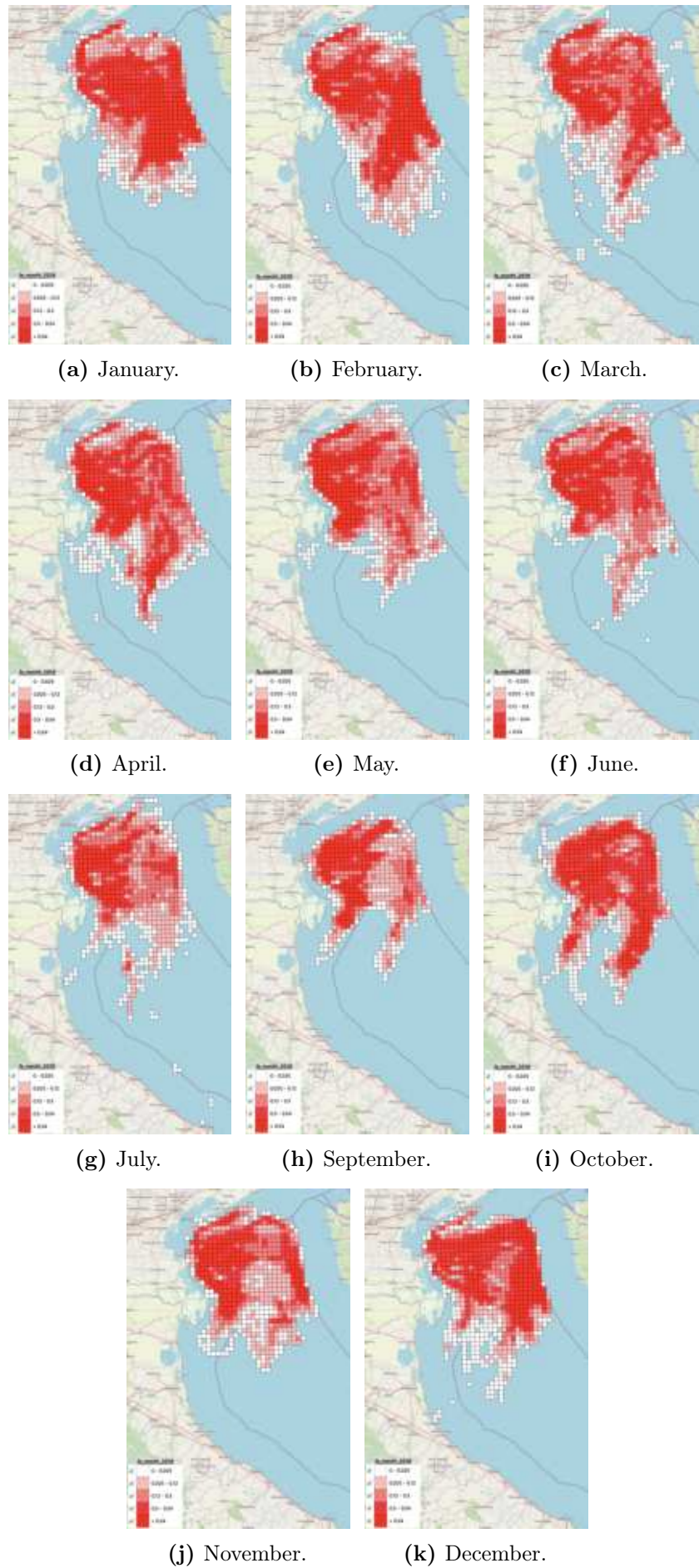


Figure 4.11: Monthly fishing effort for 2019.

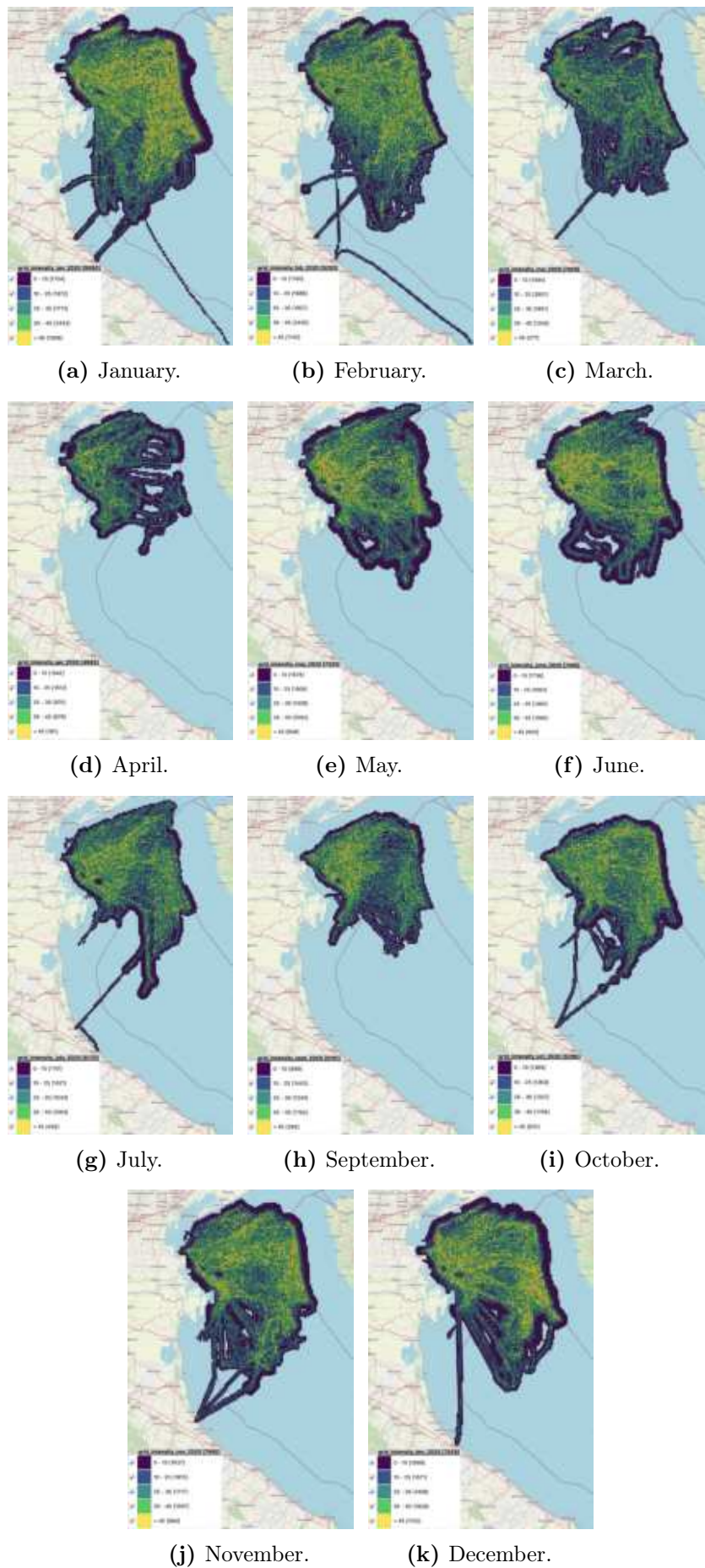


Figure 4.12: Underwater noise peak maps for each month of 2020.

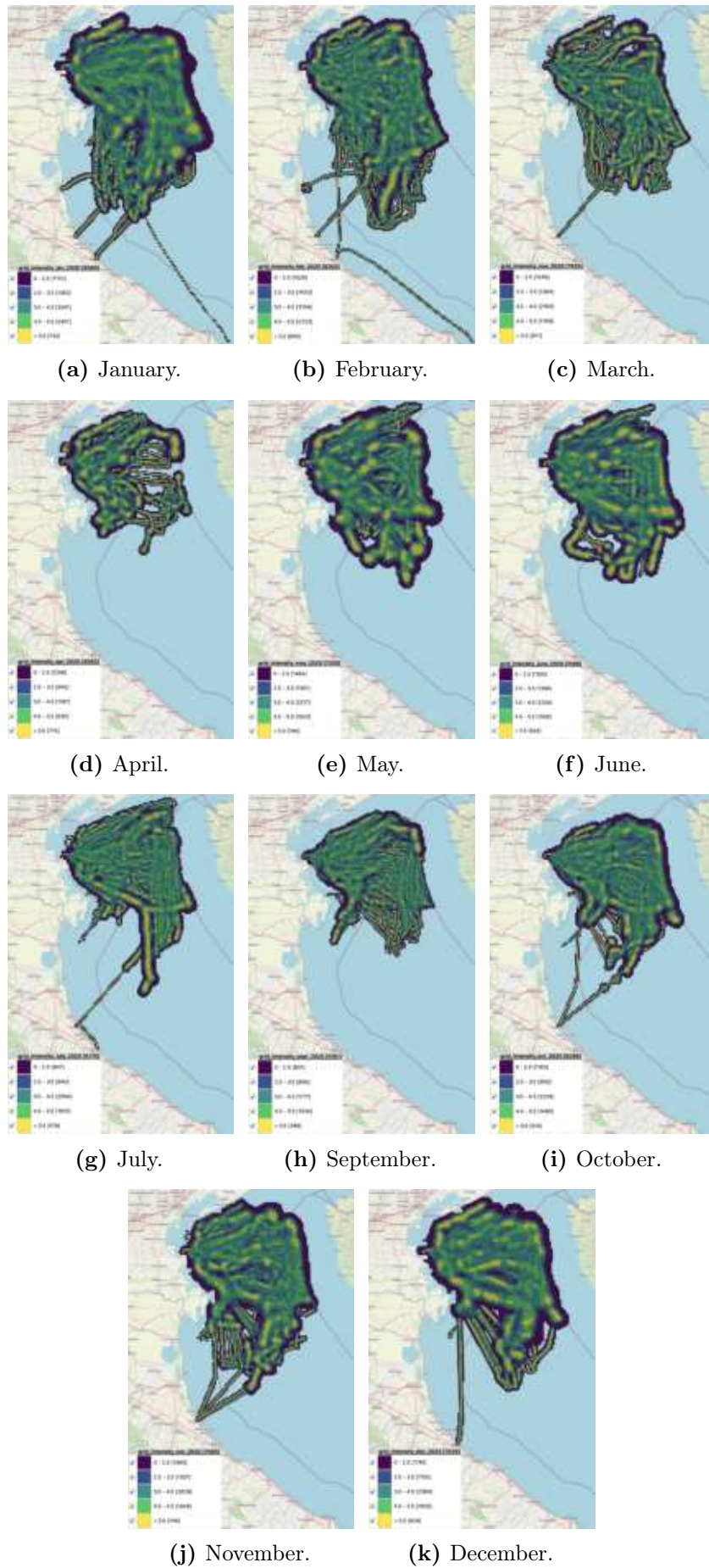


Figure 4.13: Underwater noise average maps for each month of 2020.

In fact, if we focus on the percentage of cells that are characterized by a high peak value in the months of January, May, June and December, they are 21.3%, 15.8%, 15.3% and 17.1%, while in the months of March, July and September they are 6.8%, 6.5%, 7.7% respectively.

Month	Number of cells per category					Number of cells
	0 – 10	10 – 25	25 – 35	35 – 45	> 45	
January	1731	1539	1388	2146	1843	8647
February	1499	2224	2037	2536	1045	9341
March	1735	2306	2044	2132	604	8821
April	1785	1990	1870	2398	952	8995
May	1863	1722	1695	2124	1391	8795
June	2142	1714	1523	2186	1363	8928
July	2103	2142	1711	1486	518	7960
September	1378	1545	1175	1278	447	5823
October	1658	1399	1552	2240	827	7676
November	1423	1788	1530	1693	961	7395
December	2334	1840	1250	1874	1510	8808

Table 4.8: Number of cells per category of the underwater noise peak maps 2019.

Month	Percentage of cells per category				
	0 – 10	10 – 25	25 – 35	35 – 45	> 45
January	20	17.8	16.1	24.8	21.3
February	16	23.8	21.8	27.2	11.2
March	19.7	26.1	23.2	24.2	6.8
April	19.8	22.1	20.8	26.7	10.6
May	21.2	19.6	19.3	24.1	15.8
June	24	19.2	17	24.5	15.3
July	26.4	26.9	21.5	18.7	6.5
September	23.7	26.5	20.2	21.9	7.7
October	21.6	18.2	20.2	29.2	10.8
November	19.2	24.2	20.7	22.9	13
December	26.5	20.9	14.2	21.3	17.1

Table 4.9: Percentage of cells per category of the underwater noise peak maps 2019.

Furthermore, from the same maps, it can be observed that in the months of July, September and November the noise is more intense near the coast rather than offshore. In the months of January and December there are areas offshore (on the Croatian border line) where the peak values are very high (> 45 dB). These well-defined areas, characterized by high sound peaks, correspond to the areas of greatest exploitation of the catch in 2019. Figure 4.11 shows the monthly fishing effort maps for 2019 (with a square grid in which each cell has a size of 3×3 kilometers). By analyzing these maps and those of the peaks, it is evident how they are related. In particular, this correlation is most visible in the months of January, February, July, September, November and December. This correlation is justified by the fact that, while the vessel is fishing, the underwater noise generated is much more intense than that produced when the boat is simply sailing (as previously described in Section 4.1).

Another consequence of the fact that while the boat is fishing it produces a greater underwater noise, is that, in the maps in Figure 4.9, the cells in the areas where the boats are entering or exiting from the ports (the lines that go from the offshore towards the coast) are characterized by very low noise peaks since they are simply in navigation.

Month	Number of cells per category					Number of cells
	0 – 2.5	2.5 – 3.5	3.5 – 4.5	4.5 – 5.5	> 5.5	
January	1684	1400	2813	2077	673	8647
February	1425	1421	3364	2378	753	9341
March	1461	1346	2944	2140	930	8821
April	1585	1421	3015	2166	808	8995
May	1721	1507	2818	1996	753	8795
June	1876	1421	2754	1997	880	8928
July	1542	1241	2378	1872	927	7960
September	1159	995	1737	1335	597	5823
October	1399	1135	2668	1913	561	7676
November	1316	1343	2434	1678	624	7395
December	2079	1355	2620	1946	808	8808

Table 4.10: Number of cells per category of the underwater noise average maps 2019.

Finally, by comparing Figures 4.9 and 4.10, we can observe that the maps with the average values are very different from those of the peaks. First of all, we see

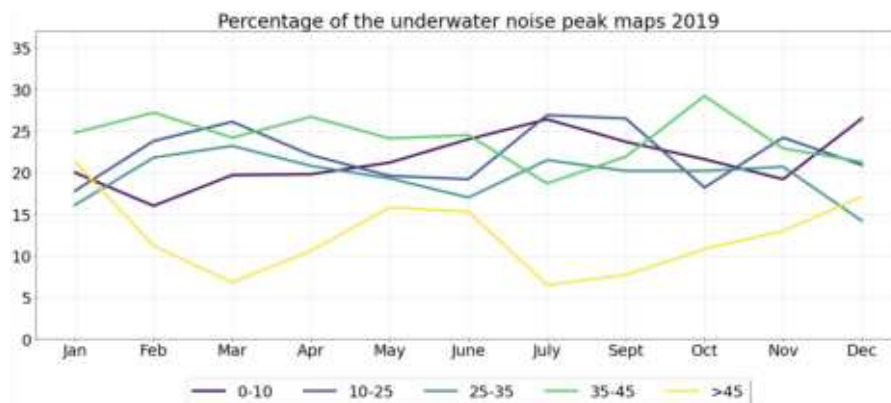
that the noise distribution on the northern Adriatic Sea is much more uniform if we consider the average values. There are no cells with very high average values close to cells with an almost zero average value: we can observe a propagation of underwater sound with areas in which the noise is more intense and gradually it becomes weaker. Furthermore, there are not many areas with a high average value (> 5.5 dB), but on average the monthly noise in the cells is around 3.5-4.5 decibel. In fact, in Table 4.11, the percentage of the number of cells is higher in this range (almost 1/3 of all cells).

Month	Percentage of cells per category				
	0 – 2.5	2.5 – 3.5	3.5 – 4.5	4.5 – 5.5	> 5.5
January	19.5	16.2	32.5	24	7.8
February	15.2	15.2	36	25.5	8.1
March	16.6	15.2	33.4	24.3	10.5
April	17.6	15.8	33.5	24.1	9
May	19.6	17.1	32	22.7	8.6
June	21	15.9	30.8	22.4	9.9
July	19.4	15.6	29.9	23.5	11.6
September	19.9	17.1	29.8	22.9	10.3
October	18.2	14.8	34.8	24.9	7.3
November	17.8	18.2	32.9	22.7	8.4
December	23.6	15.4	29.7	22.1	9.2

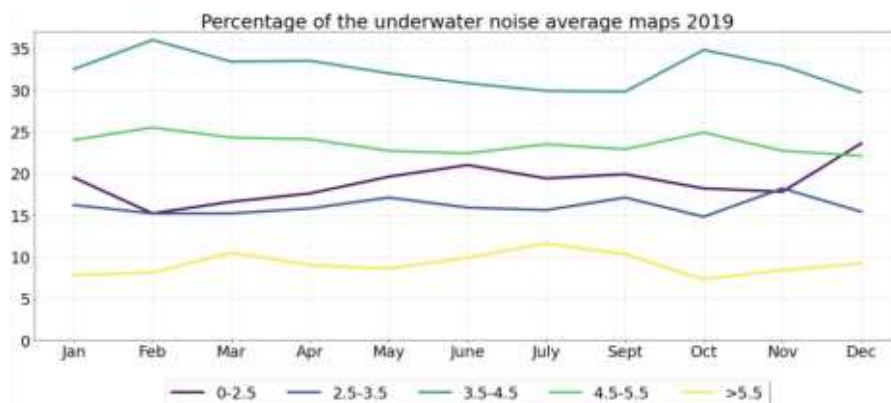
Table 4.11: Percentage of cells per category of the underwater noise average maps 2019.

Figure 4.14 represents the percentage of underwater noise for each month in each category, considering the peaks (Figure 4.14(a)) and the average values (Figure 4.14(b)). Thanks to Figure 4.14 we can easily analyze the distribution of underwater sound. In particular, as regards the average values, the noise distribution is uniform for all months, with about 7% of cells with average values above 5.5 dB and 32% of cells with values around 3.5-4.5 dB. In Figure 4.14(a) we can notice that the peak values vary greatly depending on the months of the year and they are not as uniform as the average values. In particular, the percentage of intense peaks (> 45 dB) are very high in the months of January, May, June and December and much lower (about 10% less) in the months of March, July and September (as previously described in the analysis of the maps related to Figure 4.9).

With the recent availability of AIS data also for the year 2020, we have reconstructed the trajectories, enriched them and built underwater sound propagation maps using the same model described above in this chapter. This allows us to compare the underwater noise of 2019 with that of 2020 characterized by the coronavirus disease 2019 (COVID-19) and, consequently, by several containment measures adopted by governments. In Italy, a lockdown period was imposed from March to May 2020, during which many activities had to close or limit their business. All of this caused a strong reduction in seafood requests and consequently, a decrease in fishing activities.



(a) Percentage of cells divided by category considering the maps of the underwater noise peak values in Figure 4.9.



(b) Percentage of cells divided by category considering the maps of the underwater noise average values in Figure 4.10.

Figure 4.14: Percentage of cells divided by category of the underwater noise maps for each month of 2019.

In Figure 4.12 we can visualize the peak maps for each month: each cell is characterized by the maximum noise that we have detected in that cell (the peak of the month in the cell). In Figure 4.13 we can observe the average underwater noise

value that we have measured in each cell for all the months of 2020. Moreover, for a more detailed quantitative analysis of these maps, Tables 4.12 and 4.14 show, for each month, the number of cells for each category (defined in Table 4.7). Table 4.13 and 4.15 show, for each month, the percentages of cells characterized by a value in the various noise ranges for the peaks and the averages respectively.

Month	Number of cells per category					Number of cells
	0 – 10	10 – 25	25 – 35	35 – 45	> 45	
January	1754	1872	1713	2433	1808	9580
February	1740	1986	1927	2430	1120	9203
March	1494	2651	1801	1206	277	7429
April	1542	1612	970	678	181	4983
May	1678	1809	1438	1592	808	7325
June	1736	1662	1363	1695	950	7406
July	1151	1421	1543	1563	492	6170
September	899	1443	1244	1192	383	5161
October	1369	1263	1327	1706	631	6296
November	1627	1810	1717	1957	884	7995
December	1899	1671	1408	1829	1132	7939

Table 4.12: Number of cells per category of the underwater noise peak maps 2020.

Month	Percentage of cells per category				
	0 – 10	10 – 25	25 – 35	35 – 45	> 45
January	18.31	19.54	17.88	25.40	18.87
February	18.91	21.58	20.94	26.40	12.17
March	20.11	35.69	24.24	16.23	3.73
April	30.94	32.35	19.47	13.61	3.63
May	22.91	24.70	19.63	21.73	11.03
June	23.44	22.44	18.40	22.89	12.83
July	18.66	23.03	25.00	25.33	7.98
September	17.42	27.96	24.10	23.10	7.42
October	21.74	20.06	21.08	27.10	10.02
November	20.35	22.64	21.47	24.48	11.06
December	23.92	21.05	17.73	23.04	14.26

Table 4.13: Percentage of cells per category of the underwater noise peak maps 2020.

Month	Number of cells per category					Number of cells
	0 – 2.5	2.5 – 3.5	3.5 – 4.5	4.5 – 5.5	> 5.5	
January	1722	1383	3341	2401	733	9580
February	1528	1433	3154	2223	865	9203
March	1246	1364	2193	1709	917	7429
April	1206	945	1187	930	715	4983
May	1464	1301	2217	1603	740	7325
June	1350	1398	2326	1508	824	7406
July	947	940	2094	1610	579	6170
September	801	856	1771	1334	399	5161
October	1163	900	2229	1490	514	6296
November	1465	1307	2629	1848	746	7995
December	1781	1155	2369	1805	829	7939

Table 4.14: Number of cells per category of the underwater noise average maps 2020.

Month	Percentage of cells per category				
	0 – 2.5	2.5 – 3.5	3.5 – 4.5	4.5 – 5.5	> 5.5
January	17.97	14.44	34.88	25.06	7.65
February	16.60	15.57	34.27	24.16	9.40
March	16.77	18.36	29.52	23.00	12.35
April	24.20	18.97	23.82	18.66	14.35
May	19.99	17.76	30.27	21.88	10.10
June	18.23	18.88	31.41	20.36	11.12
July	15.35	15.23	33.94	26.10	9.38
September	15.52	16.59	34.31	25.85	7.73
October	18.47	14.30	35.40	23.67	8.16
November	18.32	16.35	32.88	23.12	9.33
December	22.43	14.55	29.84	22.74	10.44

Table 4.15: Percentage of cells per category of the underwater noise average maps 2020.

From the maps of 2020 (Figures 4.9 and 4.12), we can easily observe that in March (when restrictions have been imposed) there is a noteworthy decrease in the areas affected by fishing activities. In fact, this difference can be seen in the months of March, April, July and October, while in the months of January and February

we can observe how the fishing areas of the fishing boats are very extensive even in 2020. In order to better understand the differences over the two years, we have defined a percentage difference d , according to the following equation:

$$d = \frac{v_{2020} - v_{2019}}{v_{2019}} \times 100 \quad (4.14)$$

where v corresponds to the value of which we want to analyze the difference. For example, if we want to calculate the percentage difference with respect to the number of cells between January 2019 and 2020, the value v will be the number of cells in the two months and, therefore, 8647 and 9580 respectively. In Figure 4.15 we can see this difference for each month.

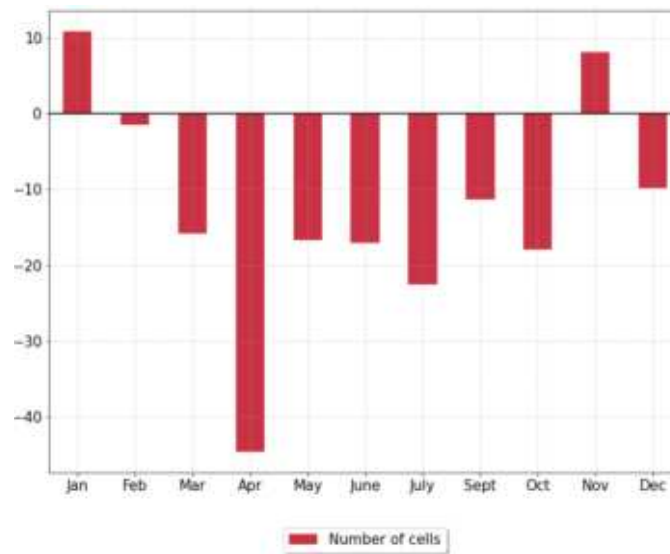


Figure 4.15: Percentage difference on the number of cell.

As we can see from Figure 4.15, except for the months of January and November, the areas affected by fishing activities have decreased significantly. In particular, in the month of April 2020 we have 44.6% fewer cells than in the same month of 2019, while for the months of March, May, June and July we have about 15-20% fewer cells. The decrease in fishing areas reflects the lockdown imposed by the government and the period of uncertainty that has led fishermen to stay closer to the coast or reduce the length of their trips.

Using Equation (4.14) we calculate the percentage difference for underwater noise peak and average maps for each category and each month. In Figure 4.16 there are the percentage differences for underwater noise peak maps.

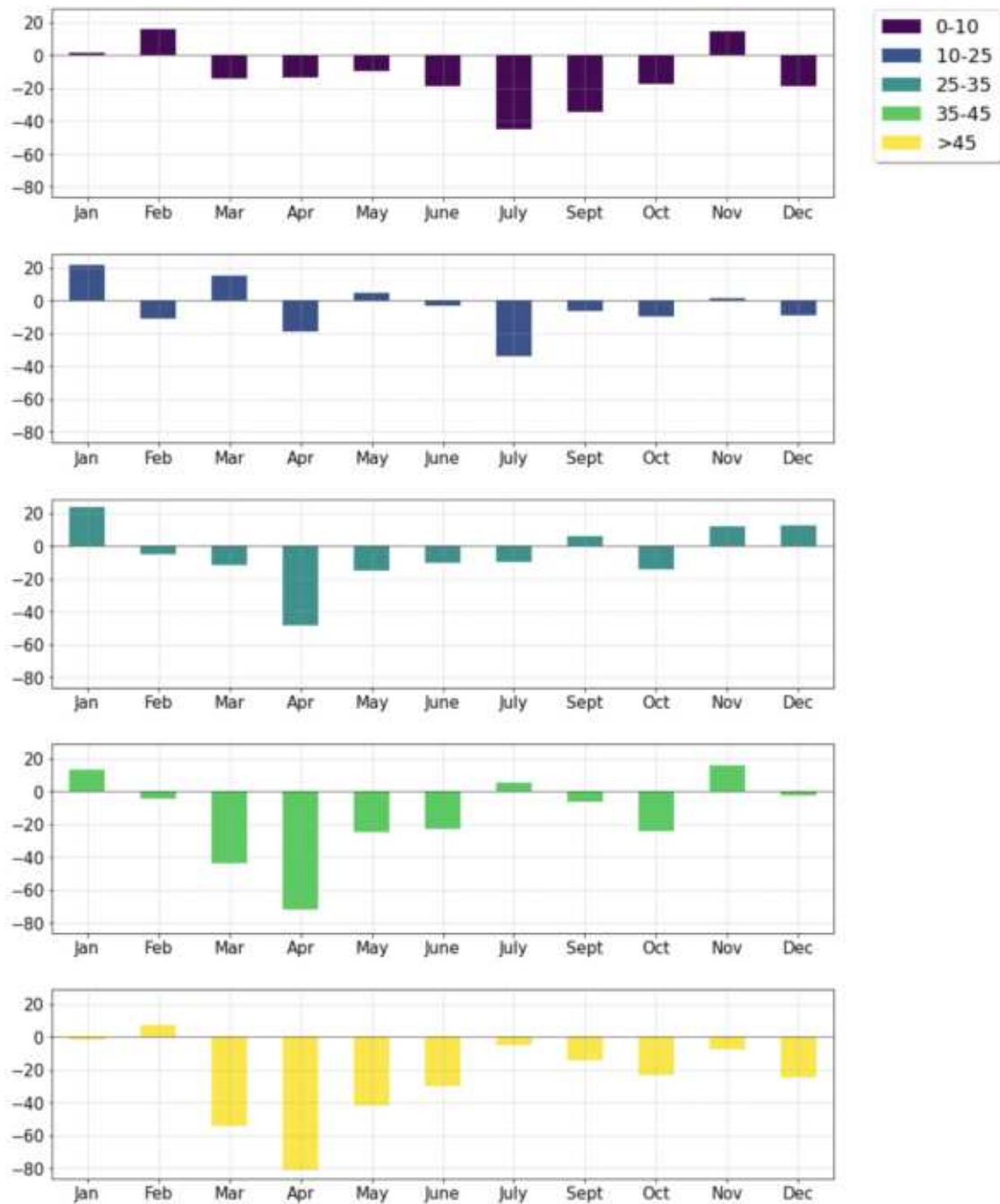


Figure 4.16: Percent difference on peak maps.

In particular, we can observe a decrease in the number of cells for peaks greater than 45 *dB* and also for peaks between 35 and 45 *dB*, even if this difference becomes less and less noticeable in the categories where the peaks have lower values. This phenomenon is due to the fact that the number of boats carrying out fishing activities were fewer than those before COVID-19, and consequently their concentration in the same areas (cells) has decreased. Since the noise perceived by the virtual listening points of the cells is given by the sum of the noise produced by

the individual boats, the lower the concentration of the boats in the same cell is, the lower the resulting peak is, because it is calculated considering the contribution of fewer boats.

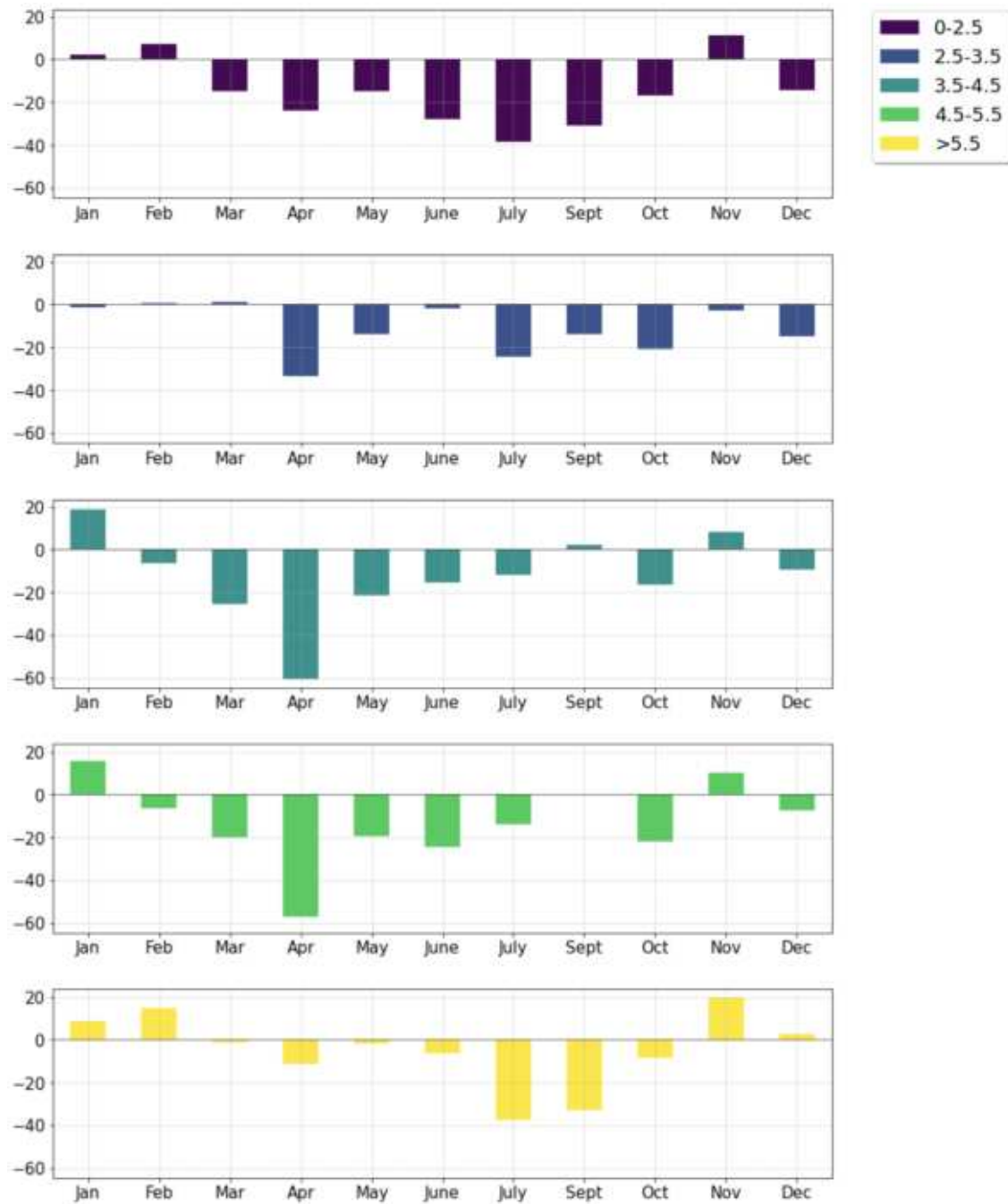


Figure 4.17: Percentage difference on the maps of the averages.

In Figure 4.17 there are the percentage differences for underwater noise on the maps of the noise average values. Also from these graphs we can observe how there is a decrease even with respect to the average noise generated by boats. This difference is clearly visible for values below 2.5 dB and between 3.5 and 5.5 dB. For

example, for the month of April (which is also characterized by a strong decrease in peaks values) we have 60% fewer cells characterized by an average value between 3.5 and 5.5 *dB*.

From Figures 4.16 and 4.17 it is possible to observe how during the COVID-19 pandemic there was a strong decrease in underwater noise in the northern Adriatic Sea. The sound propagation model developed in this work is able to capture these differences and it can help environmental experts to understand how certain circumstances can affect the marine environment and animal species.

Chapter 5

Conclusions and future works

Every trail has its end, and every calamity brings its lesson!

– James Fenimore Cooper, *The Last of the Mohicans*

In this work we wanted to tackle two hot topics related to the monitoring of the exploitation of the seas: the prediction of the fishing effort and the modelling of the underwater noise produced by vessels. We have started with a deep investigation of the literature in these two settings, by pointing out the main approaches and the differences with respect to our proposal. Then we have focused on the two tasks separately. First of all, taking as basis the work already carried out in [74, 14, 73] we have processed the dataset of the years 2019 and 2020: starting from the AIS data, we have reconstructed the trajectories and introduced semantic attributes capable of unveiling interesting information and aspects of the original data themselves. Thus, we have concentrated on fishing effort: we have computed it over a square grid with a size of 3×3 kilometers for each day. Then, we have enriched such a grid with environmental data that change on a daily basis. After the creation of this dataset, we have built predictive models. In particular, to forecast the fishing effort we have used four different Machine Learning models: Random Forest, Extra Trees, XGBoost and Multi-layer Perceptron. The results are based on three metrics, including error-based (MAE and RMSE) and coefficient of determination. The best machine learning method for our problem is Random Forest which achieves a score of about 0.67 out of 1 for R^2 . The result is considered a good achievement because the problem is challenging, given the fact that the dataset contains multivariate and spatio-temporal aspects to cope with, but it can be improved. For example,

we could include among the dataset variables, the fishing effort in the surrounding areas, or in some previous periods of time. So, as future work, we would like to study and use Statistical, Machine Learning and Deep Learning spatio-temporal models to improve the prediction. Environmental spatio-temporal data are usually characterized by spatial, temporal, and spatio-temporal correlations and capturing these dependencies is an extremely important task [3]. Deep learning has achieved remarkable success in this type of problems due to the powerful ability in automatic feature representation learning [88], because of its capability to automatically extract features both in the spatial domain (with Convolutional Neural Networks) and in the temporal domain (with the recurrent structure of Recurrent Neural Networks) [3].

As far as the underwater noise modelling is concerned, to avoid the problem of the non-homogeneity of the AIS data transmission, we have considered the positions of the fishing vessels every 30 seconds. As in the previous analysis we have partitioned the northern Adriatic Sea into a regular grid where each cell has a smaller size, namely 1×1 kilometer, because in this way we can distribute the underwater noise more accurately since, with smaller cells, we consider more listening points. Then we enriched the grid with environmental data that change on a daily basis. At this point we have calculated the propagation of underwater sound given a noise source. The noise source corresponds to a vessel: the higher the horsepower of the engine is, the more intense is the noise produced by the boat, and with the same horsepower, if the boat is fishing it emits more noise. Using this underwater sound propagation model we are able to understand how the single boat produces underwater noise depending on the activity it is carrying out. We have also generated maps relating to the sound peaks perceived by the listening points in each month and the monthly average noise in each cell.

By using this underwater noise propagation model we have analysed the fishing vessels for the years 2019 and 2020. This allows us to compare the underwater noise generated by the vessels in two completely different scenarios: the year 2019 before the COVID-19 pandemic emergency and 2020, characterized by several containment measures adopted by the government and the financial and economic crisis. It turns out that considering the peaks we have, on average, 12.6% less cells for each category, while for the average noise values we have, on average 11.3% less cells considering each category. Furthermore, the areas characterized by fishing activities have decreased by 12.5% (in 2020).

This model for the propagation of underwater noise, based on the technical

characteristics of the fishing boats' engine and on some environmental factors, can be refined in several directions. First of all, the computation of the noise could be improved by considering other environmental factors such as sediments, bathymetry and more accurate measurements of the noise generated by fishing vessels. Moreover, we should consider possible interaction between the noise produced by the fishing vessels since at the moment they are taken in isolation.

Another aspect that can lead to the improvement of both the fishing effort forecast and the underwater sound propagation is the semantic trajectories enhancement. Ecologists are interested in specific movements made by the fishing vessels during their trips. For instance, trawlers make circular movements when fishing. Representing this event as a pattern would be very important to better localize the places where vessels are fishing. Therefore, one objective is to define mining algorithms to detect recurrent patterns and classify them according to the activity the vessels are performing. It is also worthwhile to detect routine traffic of vessels and abnormal vessel behaviour.

Finally, it is important to remark that in this work we have used the AIS data and the environmental variables from the northern Adriatic Sea because the Italian Coast Guard VII Department-ITC and Traffic Monitoring has provided the raw AIS data for that area. Nevertheless, the proposed framework is general because it could be applied in all contexts in which we have the AIS data and the required environmental variables.

Appendix A

Forecast results

This appendix contains some results obtained in the fishing effort forecast described in Section 3.6

A.1 Feature Permutation Importance

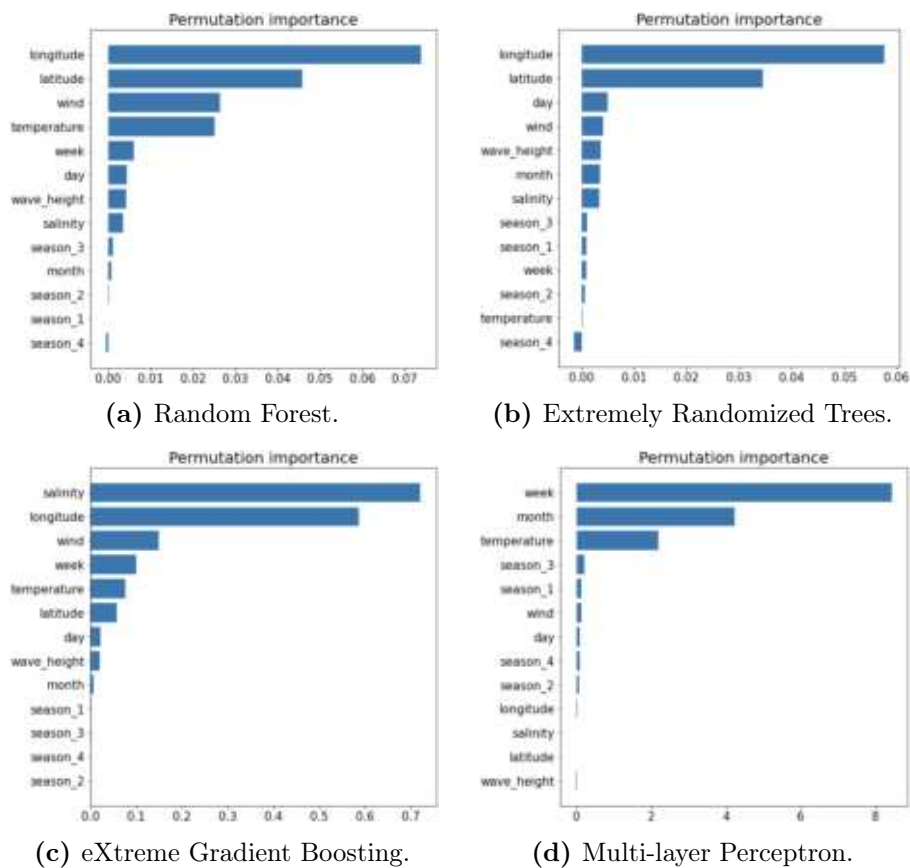


Figure A.1: Permutation importance of the four machine learning models considering one year of training and one of testing.

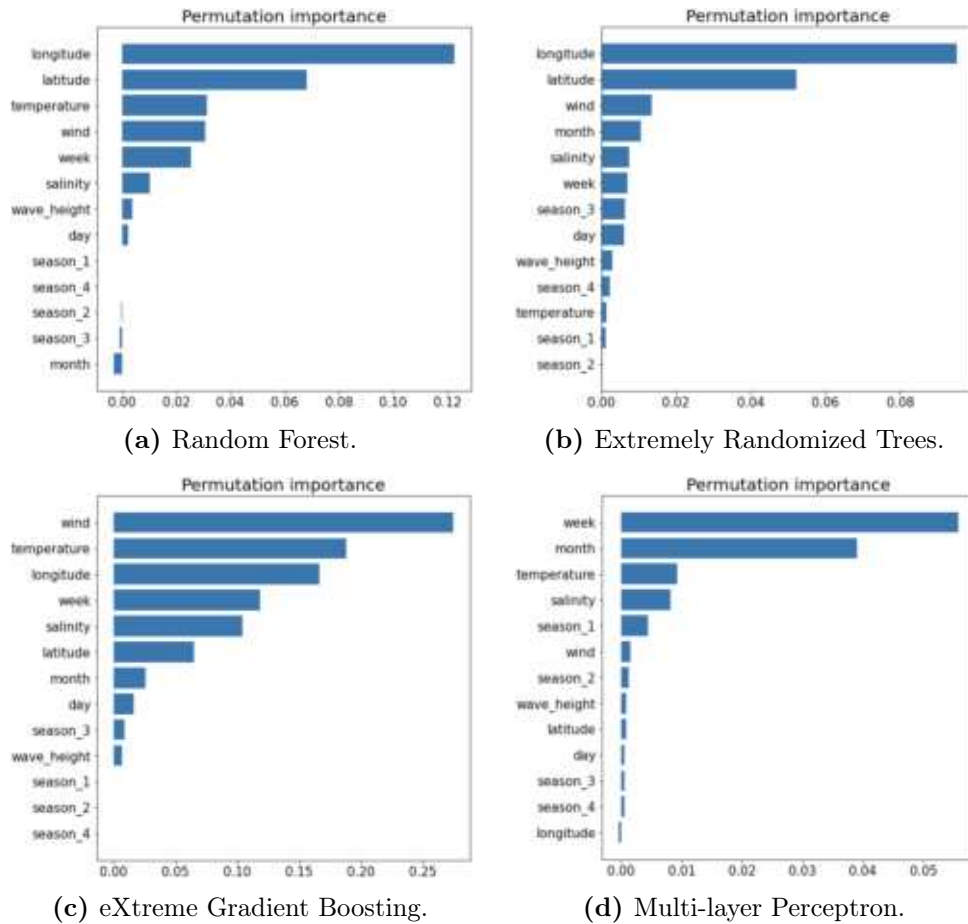


Figure A.2: Permutation importance of the four machine learning models considering two years of training and one of testing.

Figures A.1 and A.2 show the Permutation feature importance for the four machine learning algorithms used for fishing effort forecast for 2019, using one year (2018) and two years (2017-2018) as training dataset respectively. It is possible to notice how, by adding information to the initial dataset, the most relevant features are the longitude, latitude and wind for RF and ET for all training datasets. Regarding XGBoost machine learning model, the most important features are the longitude, salinity, wind and temperature. Finally for the Multi-layer perceptron the three most relevant features, in all training datasets, are the month, week and temperature.

A.2 Evaluation metrics results

Tables A.1 and A.2 present the results for each season using one year (2018) and two years (2017-2018) as training dataset respectively.

ML model	Season	MAE	RMSE	R ²
Random Forest	1	0.34082	0.60134	0.60409
	2	0.26311	0.48206	0.65443
	3	0.19944	0.34858	0.67739
	4	0.30741	0.50781	0.65775
Extra Trees	1	0.35159	0.62057	0.57837
	2	0.28375	0.52513	0.58993
	3	0.19126	0.35784	0.66003
	4	0.34309	0.57023	0.56843
XGBoost	1	0.35884	0.64295	0.54741
	2	0.30364	0.54059	0.56541
	3	0.34454	0.50443	0.32443
	4	0.41845	0.66789	0.40794
Multi-layer Perceptron	1	0.38032	0.56692	0.64812
	2	0.58788	0.91161	-0.23581
	3	0.30231	0.30231	0.5048
	4	0.60142	0.91846	-0.11964

Table A.1: Results for each season using one year (2018) as training dataset.

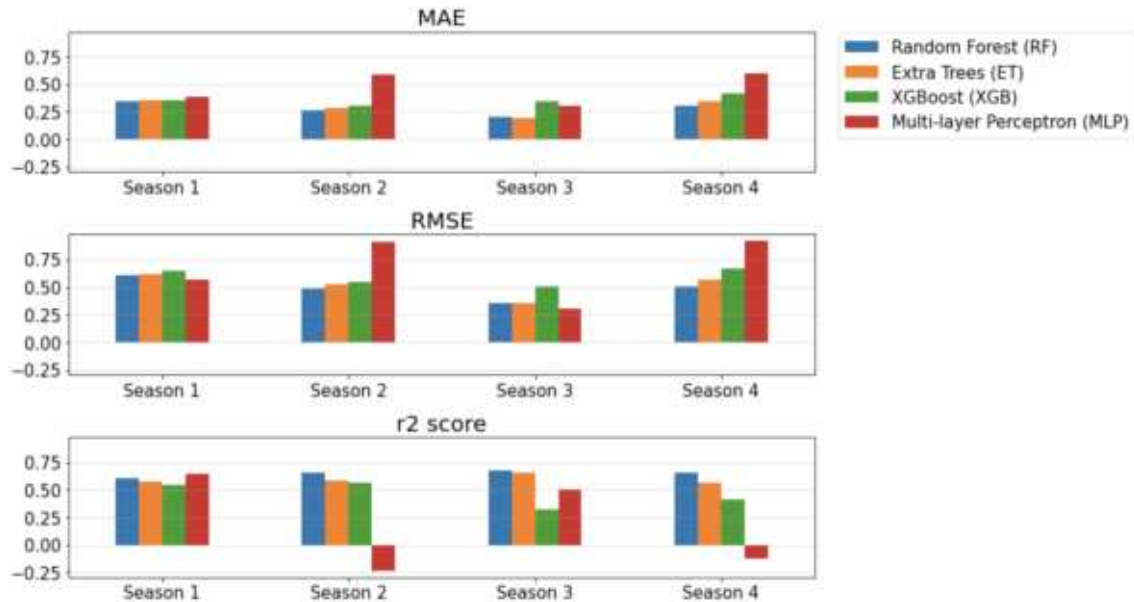


Figure A.3: MAE, RMSE and R^2 scores for each season considering one year of training for the four machine learning models: RF (blue), ET (orange), XGBoost (green) and MLP (red).

ML model	Season	MAE	RMSE	R^2
Random Forest	1	0.33835	0.59006	0.61881
	2	0.234	0.43804	0.71467
	3	0.2	0.34634	0.68153
	4	0.30457	0.50155	0.66612
Extra Trees	1	0.34579	0.60755	0.59588
	2	0.253	0.46356	0.68045
	3	0.18149	0.31919	0.72951
	4	0.31752	0.52875	0.62893
XGBoost	1	0.34371	0.60031	0.60545
	2	0.23739	0.44528	0.70514
	3	0.19504	0.34324	0.68721
	4	0.31407	0.5211	0.63959
Multi-layer Perceptron	1	0.37003	0.64349	0.54665
	2	0.31479	0.56218	0.53002
	3	0.24147	0.47843	0.3923
	4	0.42531	0.7022	0.34555

Table A.2: Results for each season using two years (2017-2018) as training dataset.

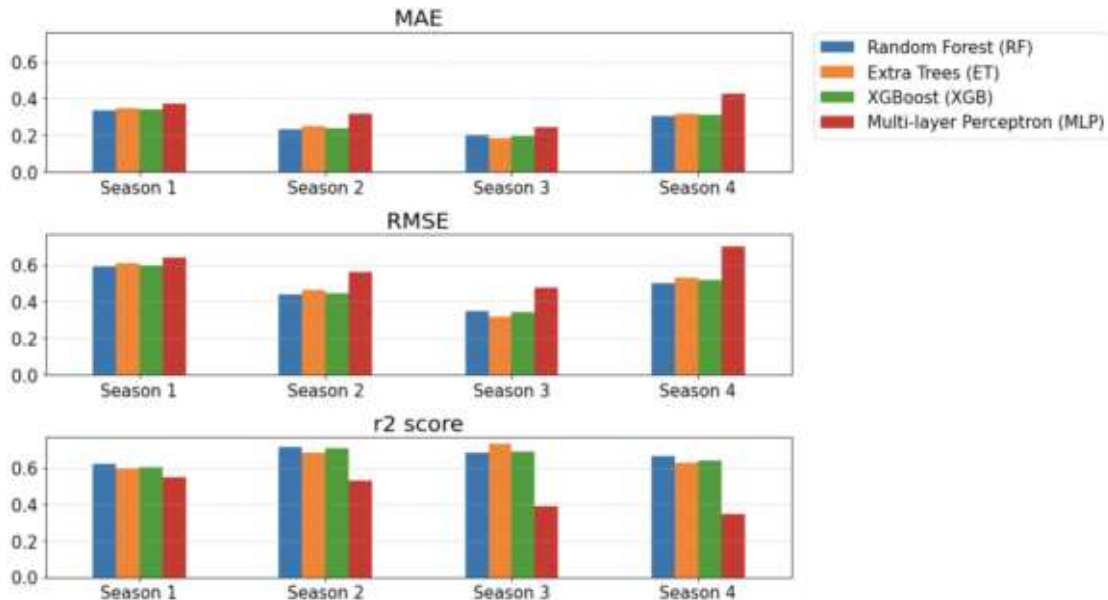


Figure A.4: MAE, RMSE and R^2 scores for each season considering two years of training for the four machine learning models: RF (blue), ET (orange), XGBoost (green) and MLP (red).

The first thing that we can observe is that using Multi-layer Perceptron machine learning algorithm considering one year as training dataset, the coefficient of determination assumes even a negative value for season 2 and 4. This means that using this method we do not achieve good results. We can easily observe this phenomenon also from Figure A.3. Concerning the other algorithms, using Random Forest we get the best values with respect to all the evaluation metrics. Also Extra Trees works well, while XGBoost does not achieve good results. From Figure A.3 we can very easily see the behaviour of each machine learning algorithm in each season and understand that using the Random Forest the results that we get are good.

From Table A.2 and Figure A.4 it is possible to understand the behaviour of the four machine learning methods in each season considering two years as training dataset. Also in this case Random Forest performs better than the other models except for season 3 where Extra Trees achieves better results (even if there is not much difference between the two). XGBoost evaluation metrics are closer to the ones that are considered as the best values, and this means that, by adding information to the initial dataset (and therefore adding also the year 2017), XGBoost increases its reliability. Finally, Multi-layer Perceptron still does not provide good results.

Bibliography

- [1] O. Adam and H. Glotin. Passive acoustic storey of the Antares neutrino detector for real-time cetaceans detection, localization and behavior studies. In *2008 New Trends for Environmental Monitoring Using Passive Systems*, pages 1–6. IEEE, 2008.
- [2] Michael A Ainslie and James G McColm. A simplified formula for viscous and chemical absorption in sea water. *The Journal of the Acoustical Society of America*, 103(3):1671–1672, 1998.
- [3] Federico Amato, Fabian Guignard, Sylvain Robert, and Mikhail Kanevski. A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific reports*, 10(1):1–11, 2020.
- [4] K Amiri, N Shabanipour, and S Eagderi. Forecasting the catch of kilka species (*Clupeonella* spp.) using Time Series SARIMA models in the Southern Caspian Sea. *Caspian Journal of Environmental Sciences*, 16(4):349–358, 2018.
- [5] Amron Amron, Rizqi Rizaldi Hidayat, Maria Dyah Nur Meinita, and Mukti Trenggono. Underwater noise of traditional fishing boats in Cilacap waters, Indonesia. *Heliyon*, 7(11):e08364, 2021.
- [6] A Anuja, VK Yadav, VS Bharti, and NR Kumar. Trends in marine fish production in Tamil Nadu using regression and autoregressive integrated moving average (ARIMA) model. *Journal of Applied and Natural Science*, 9(2):653–657, 2017.
- [7] Florian Aulanier, Yvan Simard, Nathalie Roy, Cédric Gervaise, and Marion Bandet. Effects of shipping on marine acoustic habitats in Canadian Arctic estimated via probabilistic modeling and mapping. *Marine pollution bulletin*, 125(1-2):115–131, 2017.
- [8] Mohamed Bakli, Mahmoud Sakr, and Esteban Zimanyi. Distributed moving object data management in MobilityDB. In *Proceedings of the 8th ACM*

- SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, pages 1–10, 2019.
- [9] Mariana Belgiu and Lucian Drăguț. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31, 2016.
- [10] Gérard Biau and Erwan Scornet. A random forest guided tour. *Test*, 25(2):197–227, 2016.
- [11] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.
- [12] Pedro Bordalo-Machado. Fishing effort analysis and its potential to evaluate stock size. *Reviews in Fisheries Science*, 14(4):369–393, 2006.
- [13] George EP Box and Gwilym M Jenkins. Time series analysis: Forecasting and control San Francisco. *Calif: Holden-Day*, 1976.
- [14] Bruno Brandoli, Alessandra Raffaetà, Marta Simeoni, Pedram Adibi, Fateha Khanam Bappee, Fabio Pranovi, Giulia Rovinelli, Elisabetta Russo, Claudio Silvestri, Amilcar Soares, et al. From multiple aspect trajectories to predictive analysis: a case study on fishing vessels in the Northern Adriatic sea. *GeoInformatica*, pages 1–29, 2022.
- [15] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [16] Ding-Geng Chen and Steven R Hare. Neural network and fuzzy logic models for pacific halibut recruitment analysis. *ecological modelling*, 195(1-2):11–19, 2006.
- [17] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [18] Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.
- [19] E. Cruz, T. Lloyd, J. Bosschers, F.H. Lafeber, P. Vinagre, and G. Vaz. Study on inventory of existing policy, research and impacts of continuous underwater noise in Europe. *EMSA report EMSA/NEG/21/2020*. *WavEC Offshore Renewables and Maritime Research Institute Netherlands*, 2021.

- [20] Menggunakan Satelit Penderiaan Jauh dan Teknik, Suhartono Nurdin, Muzneena Ahmad Mustapha, Tukimat Lihan, and Mazlan Abd Ghaffar. Determination of potential fishing grounds of Rastrelliger kanagurta using satellite remote sensing and GIS technique. *Sains Malaysiana*, 44(2):225–232, 2015.
- [21] Natacha Aguilar De Soto, Natali Delorme, John Atkins, Sunkita Howard, James Williams, and Mark Johnson. Anthropogenic noise causes body malformations and delays development in marine larvae. *Scientific reports*, 3(1):1–5, 2013.
- [22] Erico N de Souza, Kristina Boerder, Stan Matwin, and Boris Worm. Improving fishing pattern detection from satellite AIS using data mining and machine learning. *PloS one*, 11(7):e0158248, 2016.
- [23] Paolo Diviacco, Antonio Nadali, Massimiliano Iurcev, Mihai Burca, Rodrigo Carbajales, Matteo Gangale, Alessandro Busato, Fabio Brunetti, Lorenzo Grio, Alberto Viola, et al. Underwater Noise Monitoring with Real-Time and Low-Cost Systems, (The CORMA Experience). *Journal of Marine Science and Engineering*, 9(4):390, 2021.
- [24] Christine Erbe, Alexander MacGillivray, and Rob Williams. Mapping cumulative noise from shipping to inform marine spatial planning. *The Journal of the Acoustical Society of America*, 132(5):EL423–EL428, 2012.
- [25] Mohammad Etemad, Zahra Etemad, Amílcar Soares, Vania Bogorny, Stan Matwin, and Luis Torgo. Wise sliding window segmentation: a classification-aided approach for trajectory segmentation. In *Canadian Conference on Artificial Intelligence*, pages 208–219. Springer, 2020.
- [26] Paul C Etter. *Underwater acoustic modeling and simulation*. CRC press, 2018.
- [27] J Paige Eveson, Alistair J Hobday, Jason R Hartog, Claire M Spillman, and Kirsten M Rough. Seasonal forecasting of tuna habitat in the Great Australian Bight. *Fisheries Research*, 170:39–49, 2015.
- [28] Junliang Fan, Xiukang Wang, Lifeng Wu, Hanmi Zhou, Fucang Zhang, Xiang Yu, Xianghui Lu, and Youzhen Xiang. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy conversion and management*, 164:102–111, 2018.
- [29] Julian J Faraway. *Linear models with R*. Chapman and Hall/CRC, 2004.

- [30] Adrian Farcas, Claire F Powell, Kate L Brookes, and Nathan D Merchant. Validated shipping noise maps of the Northeast Atlantic. *Science of the Total Environment*, 735:139509, 2020.
- [31] Adrian Farcas, Paul M Thompson, and Nathan D Merchant. Underwater noise modelling for environmental impact assessment. *Environmental Impact Assessment Review*, 57:114–122, 2016.
- [32] Paolo Favali, Francesco Chierici, Giuditta Marinaro, Gabriele Giovanetti, Adriano Azzarone, Laura Beranzoli, Angelo De Santis, Davide Embriaco, Stephen Monna, Nadia Lo Bue, et al. NEMO-SN1 abyssal cabled observatory in the Western Ionian Sea. *IEEE Journal of Oceanic Engineering*, 38(2):358–374, 2013.
- [33] Carmen Ferrà, Anna Nora Tassetti, Fabio Grati, Giulio Pellini, Piero Polidori, Giuseppe Scarcella, and Gianna Fabi. Mapping change in bottom trawling activity in the mediterranean sea through ais data. *Marine Policy*, 94:275–281, 2018.
- [34] FH Fisher and VP Simmons. Sound absorption in sea water. *The Journal of the Acoustical Society of America*, 62(3):558–564, 1977.
- [35] RE Francois and GR Garrison. Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulfate contributions. *The Journal of the Acoustical Society of America*, 72(3):896–907, 1982.
- [36] RE Francois and GR Garrison. Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption. *The Journal of the Acoustical Society of America*, 72(6):1879–1890, 1982.
- [37] Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15):2627–2636, 1998.
- [38] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.
- [39] Pierre Geurts and Gilles Louppe. Learning to rank with extremely randomized trees. In *Proceedings of the Learning to Rank Challenge*, pages 49–61. PMLR, 2011.

- [40] Robert C Gisiner. Workshop on the Effects of Anthropogenic Noise in the Marine Environment, 10-12 February 1998. Proceedings. Technical report, Office of Naval Research Arlington Va Marine Mammal and Biological, 1998.
- [41] Michael Goetz, Christian Weber, Josiah Bloecher, Bram Stieltjes, Hans-Peter Meinzer, and Klaus Maier-Hein. Extremely randomized trees based brain tumor segmentation. *Proceeding of BRATS challenge-MICCAI*, pages 006–011, 2014.
- [42] Juan Carlos Gutiérrez-Estrada, Claudio Silva, Eleuterio Yáñez, Nivaldo Rodríguez, and Inmaculada Pulido-Calvo. Monthly catch forecasting of anchovy *Engraulis ringens* in the north area of Chile: non-linear univariate approach. *Fisheries Research*, 86(2-3):188–200, 2007.
- [43] Ralf Hartmut Güting, Michael H Böhlen, Martin Erwig, Christian S Jensen, Nikos A Lorentzos, Markus Schneider, and Michalis Vazirgiannis. A foundation for representing and querying moving objects. *ACM Transactions on Database Systems (TODS)*, 25(1):1–42, 2000.
- [44] Ralf Hartmut Güting and Markus Schneider. *Moving objects databases*. Elsevier, 2005.
- [45] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- [46] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Random forests. In *The elements of statistical learning*, pages 587–604. Springer, 2009.
- [47] OH Kehinde and GE Joseph. Time series modelling for forecasting artisanal fish production in Nigeria. *Int. J. Adv. Multidiscip. Res*, 5(7):10–17, 2018.
- [48] Sonia Kahiomba Kiangala and Zenghui Wang. An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. *Machine Learning with Applications*, 4:100024, 2021.
- [49] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [50] Teruhisa Komatsu, Ichiro Aoki, Isamu Mitani, and Takeo Ishii. Prediction of the catch of Japanese sardine larvae in Sagami Bay using a neural network. *Fisheries science*, 60(4):385–391, 1994.

- [51] Damien Le Guyader, Cyril Ray, and David Brosset. Identifying 15 small-scale fishing zones in France using AIS data. *Advances in shipping data analysis and modeling: Tracking and mapping maritime flows in the age of big data*, page 251, 2017.
- [52] Damien Le Guyader, Cyril Ray, Françoise Gourmelon, and David Brosset. Defining high-resolution dredge fishing grounds with Automatic Identification System (AIS) data. *Aquatic Living Resources*, 30:39, 2017.
- [53] Pierre Yves Le Traon, Antonio Reppucci, Enrique Alvarez Fanjul, Lotfi Aouf, Arno Behrens, Maria Belmonte, Abderrahim Bentamy, Laurent Bertino, Vittorio Ernesto Brando, Matilde Brandt Kreiner, et al. From observation to information and users: The Copernicus Marine Service perspective. *Frontiers in Marine Science*, 6:234, 2019.
- [54] I Maina, S Kavadas, S Katsanevakis, S Somarakis, G Tserpes, and S Georgakarakos. A methodological approach to identify fishing grounds: A case study on Greek trawlers. *Fisheries Research*, 183:326–339, 2016.
- [55] Robert D McCauley, Jane Fewtrell, and Arthur N Popper. High intensity anthropogenic sound damages fish ears. *The Journal of the Acoustical Society of America*, 113(1):638–642, 2003.
- [56] Qaisar Mehmood, Maqbool Hussain Sial, Saira Sharif, Abid Hussain, Muhammad Riaz, and Nargis Shaheen. Forecasting the Fisheries Production in Pakistan for the Year 2017-2026, using Box-Jenkin’s Methodology. *Pakistan Journal of Agricultural Research*, 33(1):140, 2020.
- [57] Ronaldo dos Santos Mello, Vania Bogorny, Luis Otavio Alvares, Luiz Henrique Zambom Santana, Carlos Andres Ferrero, Angelo Augusto Frozza, Geomar Andre Schreiner, and Chiara Renso. Master: A multiple aspect view on trajectories. *Transactions in GIS*, 23(4):805–822, 2019.
- [58] Nirvana Meratnia et al. Spatiotemporal compression techniques for moving point objects. In *International Conference on Extending Database Technology*, pages 765–782. Springer, 2004.
- [59] Kate Moran, Brigitte Boutin, S Kim Juniper, Benoît Pirenne, and Adrian Round. A multi-use and multi-stakeholder ocean observing platform system. In *OCEANS 2019 MTS/IEEE SEATTLE*, pages 1–5. IEEE, 2019.

- [60] Sophie L Nedelec, Andrew N Radford, Stephen D Simpson, Brendan Nedelec, David Lecchini, and Suzanne C Mills. Anthropogenic noise playback impairs embryonic development and increases mortality in a marine invertebrate. *Scientific Reports*, 4(1):1–4, 2014.
- [61] Douglas P Nowacek, Lesley H Thorne, David W Johnston, and Peter L Tyack. Responses of cetaceans to anthropogenic noise. *Mammal Review*, 37(2):81–115, 2007.
- [62] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [63] Dieter Pfoser and Yannis Theodoridis. Generating semantics-based trajectories of moving objects. *Computers, Environment and Urban Systems*, 27(3):243–263, 2003.
- [64] Marta Picciulin, Chiara Facca, Riccardo Fiorin, Federico Riccato, Matteo Zucchetta, and Stefano Malavasi. It Is Not Just a Matter of Noise: Sciaena umbra Vocalizes More in the Busiest Areas of the Venice Tidal Inlets. *Journal of Marine Science and Engineering*, 9(2):237, 2021.
- [65] Falkner Raffaella, Picciulin Marta, Armelloni Enrico, Gaggero Tomaso, Rako Gospić Nikolina, Pleslić Grgur, Radulovic Marko, and Vučur Blazinić Tihana. Report on the source level assessment for recreational boats. <https://www.italy-croatia.eu/documents/>, 2021.
- [66] Rohan Kumar Raman and Basanta Kumar Das. Forecasting shrimp and fish catch in Chilika lake over time series analysis. In *Time Series Analysis-Data, Methods, and Applications*. IntechOpen, 2019.
- [67] Rohan Kumar Raman, TV Sathianandan, AP Sharma, and BP Mohanty. Modelling and forecasting marine fish production in Odisha using seasonal ARIMA model. *National Academy Science Letters*, 40(6):393–397, 2017.
- [68] Hassan Ramchoun, Youssef Ghanou, Mohamed Ettaouil, and Mohammed Amine Janati Idrissi. Multilayer perceptron: Architecture optimization and training. 2016.
- [69] W John Richardson, Charles R Greene Jr, Charles I Malme, and Denis H Thomson. *Marine mammals and noise*. Academic press, 2013.

- [70] FJ Rodrigo-Saura, P Poveda, J Carbajo, and J Ramis. Monitoring Long-Term underwater acoustic pollution in Mediterranean Sea waters. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 259, pages 5877–5888. Institute of Noise Control Engineering, 2019.
- [71] PH Rogers and M Cox. Underwater sound as a biological stimulus. In 'Sensory Biology of Aquatic Animals'. (Eds J. Atema, RR Fay, AN Popper and WN Tavolga.) pp. 131–149, 1988.
- [72] Rosalind M Rolland, Susan E Parks, Kathleen E Hunt, Manuel Castellote, Peter J Corkeron, Douglas P Nowacek, Samuel K Wasser, and Scott D Kraus. Evidence that ship noise increases stress in right whales. *Proceedings of the Royal Society B: Biological Sciences*, 279(1737):2363–2368, 2012.
- [73] Giulia Rovinelli. Ricostruzione, arricchimento semantico e analisi delle traiettorie dei pescherecci nel mar Adriatico settentrionale. Bachelor's thesis, Ca' Foscari University of Venice, 2020.
- [74] Giulia Rovinelli, Stan Matwin, Fabio Pranovi, Elisabetta Russo, Claudio Silvestri, Marta Simeoni, and Alessandra Raffaetà. Multiple aspect trajectories: a case study on fishing vessels in the Northern Adriatic sea. In *EDBT/ICDT Workshops*, 2021.
- [75] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [76] Hüseyin Özkan Sertlek, Hans Slabbekoorn, Carel Ten Cate, and Michael A Ainslie. Source specific sound mapping: Spatial, temporal and spectral distribution of sound in the Dutch North Sea. *Environmental pollution*, 247:1143–1157, 2019.
- [77] Richard L Shelmerdine. Teasing out the detail: How our understanding of marine AIS data can better inform industries, developments, and planning. *Marine Policy*, 54:17–25, 2015.
- [78] Hans Slabbekoorn, Niels Bouton, Ilse van Opzeeland, Aukje Coers, Carel ten Cate, and Arthur N Popper. A noisy spring: the impact of globally rising underwater sound levels on fish. *Trends in ecology & evolution*, 25(7):419–427, 2010.
- [79] Brandon L Southall, Ann E Bowles, William T Ellison, James J Finneran, Roger L Gentry, Charles R Greene Jr, David Kastak, Darlene R Ketten,

- James H Miller, Paul E Nachtigall, et al. Marine mammal noise-exposure criteria: initial scientific recommendations. *Bioacoustics*, 17(1-3):273–275, 2008.
- [80] Stefano Spaccapietra and Christine Parent. Adding meaning to your steps (keynote paper). In *International Conference on Conceptual Modeling*, pages 13–31. Springer, 2011.
- [81] Stefano Spaccapietra, Christine Parent, Maria Luisa Damiani, Jose Antonio de Macedo, Fabio Porto, and Christelle Vangenot. A conceptual view on trajectories. *Data & knowledge engineering*, 65(1):126–146, 2008.
- [82] KI Stergiou, ED Christou, and G Petrakis. Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. *Fisheries Research*, 29(1):55–95, 1997.
- [83] Hind Taud and JF Mas. Multilayer perceptron (MLP). In *Geomatic approaches for modeling land change scenarios*, pages 451–455. Springer, 2018.
- [84] Edgar Torres-Irineo, Silvia Salas, Jorge Iván Euán-Ávila, Leopoldo E Palomo, Daniel R Quijano Quinones, Eva Coronado, and Rocío Joo. Spatio-Temporal Determination of Small-Scale Vessels’ Fishing Grounds Using a Vessel Monitoring System in the Southeastern Gulf of Mexico. *Frontiers in Marine Science*, 8:542, 2021.
- [85] Lesa Triwahyanti, Ajeng Sekar Cyndana, Yessy Hurly Sefnianti, Ratna Juita Sari, and Amron Amron. Transmission loss estimation of underwater sound based on the noise intensity emitted by MV. Pengayoman IV in Tanjung Intan cruise line, Cilacap. In *E3S Web of Conferences*, volume 47, page 04011. EDP Sciences, 2018.
- [86] Robert J Urick. Principles of underwater sound 3rd edition. *Peninsula Publishing Los Atlos, California*, 22:23–24, 1983.
- [87] Michele Vespe, Maurizio Gibin, Alfredo Alessandrini, Fabrizio Natale, Fabio Mazzearella, and Giacomo C Osio. Mapping EU fishing activities using ship tracking data. *Journal of Maps*, 12(sup1):520–525, 2016.
- [88] Senzhang Wang, Jiannong Cao, and Philip Yu. Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*, 2020.

- [89] Linda S Weilgart. A brief review of known effects of noise on marine mammals. *International Journal of Comparative Psychology*, 20(2), 2007.
- [90] Robert Williams, Andrew J Wright, Erin Ashe, LK Blight, R Bruintjes, R Canessa, CW Clark, S Cullis-Suzuki, DT Dakin, Christine Erbe, et al. Impacts of anthropogenic noise on marine life: Publication patterns, new discoveries, and future directions in research and management. *Ocean & Coastal Management*, 115:17–24, 2015.
- [91] OB Wilson Jr and Robert W Leonard. Measurements of sound absorption in aqueous salt solutions by a resonator method. *The Journal of the Acoustical Society of America*, 26(2):223–226, 1954.
- [92] Andrew J Wright, Natacha Aguilar Soto, Ann Linda Baldwin, Melissa Bateson, Colin M Beale, Charlotte Clark, Terrence Deak, Elizabeth F Edwards, Antonio Fernández, Ana Godinho, et al. Do marine mammals experience stress related to anthropogenic noise? *International Journal of Comparative Psychology*, 20(2), 2007.
- [93] Lidia Eva Wysocki, John P Dittami, and Friedrich Ladich. Ship noise and cortisol secretion in European freshwater fishes. *Biological conservation*, 128(4):501–508, 2006.
- [94] Anil Kumar Yadav, Kishore Kumar Das, Pronob Das, Rohan Kumar Raman, Jeetendra Kumar, and Basanta Kumar Das. Growth trends and forecasting of fish production in Assam, India using ARIMA model. *Journal of Applied and Natural Science*, 12(3):415–421, 2020.
- [95] Syazwani Mohd Yusop, Muzzneena Ahmad Mustapha, and Tukimat Lihan. Determination of spatio-temporal distribution of *Rastrelliger kanagurta* using modelling techniques for optimal fishing. *Journal of Coastal Conservation*, 25(1):1–17, 2021.
- [96] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, and Xiuwen Yi. DNN-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 1–4, 2016.
- [97] Zhongning Zhao, Feng Hong, Haiguang Huang, Chao Liu, Yuan Feng, and Zhongwen Guo. Short-term prediction of fishing effort distributions by discovering fishing chronology among trawlers based on VMS dataset. *Expert Systems with Applications*, 184:115512, 2021.

- [98] Esteban Zimányi, Mahmoud Sakr, and Arthur Lesuisse. MobilityDB: A mobility database based on PostgreSQL and PostGIS. *ACM Transactions on Database Systems (TODS)*, 45(4):1–42, 2020.
- [99] Esteban Zimányi, Mahmoud Sakr, Arthur Lesuisse, and Mohamed Bakli. Mobilitydb: A mainstream moving object database system. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, pages 206–209, 2019.