Multiple aspect trajectories: a case study on fishing vessels in the Northern Adriatic sea

Giulia Rovinelli
867381@stud.unive.it
Università Ca' Foscari Venezia
Venezia, Italy

Stan Matwin
stan@dal.ca
Institute for Big Data Analytics, Dalhousie University
Halifax, Canada
Institute of Computer Science
Polish Academy of Sciences
Warsaw, Poland

Fabio Pranovi
Elisabetta Russo
Claudio Silvestri
Marta Simeoni
Alessandra Raffaetà
fpranovi@unive.it
elisabetta.russo@unive.it
silvestri@unive.it
simeoni@unive.it
raffaeta@unive.it
Università Ca' Foscari Venezia
Venezia, Italy

ABSTRACT

In this paper we build, implement and analyze a spatio-temporal database describing the fishing activities in the Northern Adriatic Sea over four years. The database results from the fusion of two complementary data sources: trajectories from fishing vessels (obtained from terrestrial Automatic Identification System, or AIS, data feed) and the corresponding fish catch reports (i.e., the quantity and type of fish caught). We present all the phases of the dataset creation, starting from the raw data and proceeding through data exploration, data cleaning, trajectory reconstruction and semantic enrichment. Moreover, we formalise and compare different techniques to distribute the fish caught by the fishing vessels along their trajectories. We implement the database with MobilityDB, an open source geospatial trajectory data management and analysis platform. Subsequently, guided by our ecological experts, we perform some analyses on the resulting spatio-temporal database, with the goal of mapping the fishing activities on some key species, highlighting all the interesting information and inferring new knowledge that will be useful for fishery management.

1 INTRODUCTION

The Northern Adriatic Sea area 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 policy makers that, together, could use such information for the development of effective fishery management plans, able to make fishing activities sustainable and ensure a productive and healthy ecosystem.

Some interesting objectives relative to the sea monitoring in the Adriatic Sea are:

- improving the knowledge of the fishing activities in the Northern and Central Adriatic Sea,
- evaluating the effectiveness of the current fishery managements, and
- detecting the spatial distribution of commercial fishery catches.

Our aim in this paper is to build, implement and analyze a spatio-temporal database for gaining a sound knowledge about the fishing activities in the Northern Adriatic basin and trying to address the above matters. To accomplish this task, we start from two complementary data sources covering a time period of four years, from January 2015 to December 2018. The first data source is the set of terrestrial Automatic Identification System (AIS) data, i.e., the AIS data sent by ships and received by ground stations on the Italian coast of Northern Adriatic sea. In particular, we focus on the AIS data of the fishing vessels. The second data source is the fish catch reports of the Chioggia fish market, which is the primary market of the Northern Adriatic basin. Such reports contain the quantity and type of fish caught by all vessels selling their landings at the Chioggia fish market.

Similar data have been used in [1] to develop early results on the use of machine learning techniques to predict the future Catch Per Unit Effort (CPUE), an indicator of fishing resources exploitation, from the past data of the Northern Adriatic sea. The mentioned work had some limitations, mainly related to the short temporal horizon – only two years, 2015 and 2016 – of the landing and AIS data. This, in fact, turned out to be a serious problem for the application of prediction methods: using the first year for training and the second one for testing, was not sufficient. The novel database that originates from the present work, thanks to the availability of the data sources for two further years, puts the basis to overcome such limitations and paves the way for a subsequent favorable application of prediction methods.

We present all the phases of the database creation. Trajectories are reconstructed by linear interpolation of the raw AIS data. We first clean the data, then we detect the trips performed by the fishing vessels and we enrich the resulting trajectories with additional information concerning the activities and anomalies occurring during their trips. Moreover, relying on the landing reports of the Chioggia fish market, we add a further, valuable semantic aspect to the trajectories which denotes the quantity of fish caught in each trajectory segment of the fishing vessel. In order to distribute the total catches along the trajectories we define two different approaches, which are subsequently put into action and compared through specific analyses. First, we adopt the *uniform distribution* where the catch of a given species is uniformly distributed along the fishing segments of the corresponding trajectory: each fishing segment is associated with a portion of the total amount of fish, proportional to its length. The uniform distribution is clearly a simplification of the reality. We refined it by considering a *weighted distribution*, whose underlying idea is that the areas where more vessels are fishing during a given time period are more likely to have higher catch rates.

In [4], the authors review current research challenges and trends tied to the integration, management, analysis, and visualization of objects moving at sea. Several strategies have been proposed to deal with the fusion of heterogeneous ocean data properly. For example, the paper [13] shows a platform in the maritime vessel traffic domain for discovering real-time traffic alerts by querying and reasoning across numerous streams (e.g., AIS, weather, ice). The authors use semantic web technologies to integrate heterogeneous data sources. In [3], the authors propose a model for integration and analysis of data for vessel movement in a real-time maritime situation awareness system, also using semantic web techniques and tools. Unlike the previous methods, we represent our trajectory data with a semantic model. By considering data sources such as AIS and landing reports, the trajectory of every fishing vessel becomes a complex object with several data dimensions that are contextual to the movement.

Several semantic models for trajectory data have been proposed, such as the *stops* and *moves* [7], CONSTANT [2], and recently MASTER [6]. In this paper we follow the MASTER semantic model which, among the three proposals, is the more flexible and expressive since it allows for enriching trajectories with complex objects. We represent the trajectory of fishing vessels as a *multiple aspect trajectory*. The AIS data constitute the sequence of spatio-temporal points. Moreover, the MASTER model introduces the concept of *aspect* which consists of "a real-world fact that is relevant for the trajectory data analysis" [6]. It distinguishes between *volatile* aspects — usually associated with the trajectory points, since they vary during the object movement — and *long term* aspects — which do not change during an entire trajectory, and hence they are associated with the whole trajectory. For instance, for vessel trajectories, the speed is a volatile aspect, whereas the fishing gear type is a long term aspect.

We also provide a prototype implementation of our spatio-temporal database based on MobilityDB [15], an open source geospatial trajectory data management and analysis platform, specifically developed to support the representation and the analysis of moving objects. On the one hand, the implementation in MobilityDB allows us to perform various analyses on the dataset and assess the appropriateness of the conceptual framework. On the other hand, it reveals the potentialities of MobilityDB for the reconstruction and management of semantic trajectories. In fact, the system offers temporal types that are suited to model points changing their position along a time period. It also provides a lot of spatio-temporal operators to handle trajectories, e.g for getting the position and the associated annotations of a trajectory at a certain time instant, or checking topological relations or computing the distance between trajectories. It supports also the GiST (Generalized Search Tree) and SP-GiST (Space-Partition GiST) indexes, which can be used for accelerating spatial, temporal and spatio-temporal queries. Finally, trajectories can be visualized by traditional tools such as QGIS [11], an Open Source GIS that supports viewing, editing, and analysis of geospatial data.

The spatio-temporal database is used for analysing some phenomena of interest. First, we check the AIS coverage and we detect areas where there are transmission problems. Then, guided by our ecological experts, we map the fishing activities on some key species, highlighting all the interesting information and inferring new knowledge that will be useful for fishery management. The analyses show that spatialising the distribution of catches allows one to single out the fishing grounds and their seasonal and annual variation. This can be useful for the explanation of the fishermen behaviour, as well as to better understand the seasonal migration of the target species.

In summary, the main contributions of the paper are: (i) the use of MobilityDB, a database platform which is proved to be particularly suited for creating and analysing semantic trajectories; (ii) the development of a case study concerning fishing activities based on real and heterogeneous datasets, with the proposal of different approaches for distributing catches along the trajectories; (iii) the execution of various qualitative analyses, proposed and assessed by our ecological experts, for detecting spatial and temporal patterns of the fishing activities.

The paper is organised as follows: Section 2 describes the trajectory reconstruction and enrichment and the creation of a spatio-temporal database by means of MobilityDB. Section 3 reports and discusses the results of some specific analyses performed with MobilityDB on the obtained database. In particular, we show the usefulness of recording and visualizing possible anomalies of the trajectories as well as how to take advantage of the catches distribution for gaining new knowledge on key species in the area. Finally, we draw some concluding remarks in Section 4.

2 FROM RAW DATA TO MULTIPLE ASPECT TRAJECTORIES

In this section we illustrate the various ingredients and steps we followed to produce a spatio-temporal database of fishing vessels' trajectories in the Northern Adriatic sea, enriched with landing data from the Chioggia market. We start by describing the data sources of our case study, that is, the terrestrial AIS data of the Northern Adriatic sea, and the landing reports of the Chioggia fish market. Next, we explain how trajectories can be reconstructed by linear interpolation of the raw AIS data. In this step, we clean the data, we detect the trips performed by the fishing vessels and we enrich the resulting trajectories with additional information concerning the activities and anomalies occurring during the trips. Then we illustrate how to assign landing reports to trajectories and we formalise the two different techniques to distribute the fish catches along the trajectories. Finally, we give some details of the implementation by showing the advantages of using MobilityBD as database to store and analyse trajectories.

The overall view of the process is depicted in Fig. 1: Starting from the raw terrestrial AIS data of the fishing vessels and from the landing reports of the Chioggia's market, we build up on top of MobilityDB a spatio-temporal database of multiple aspect trajectories that enables us to perform analyses on the spatio-temporal and semantic features of the trajectories.

2.1 Data sources

Automatic Identification System (AIS). AIS raw data, provided by the Italian Coast Guard, were obtained for the trawl fishing vessels operating in the Northern Adriatic Sea from January 2015 until December 2018. A total of 70 (2015), 77 (2016), 82 (2017) and 81 (2018) trawlers, with a length overall above 15m, were taken into consideration in this study: in particular, small and large bottom otter trawl (SOTB and LOTB), Rapido, one specific kind of beam trawl (RAP), and midwater pair trawl (PTM). The identification of the vessels was performed by matching the data

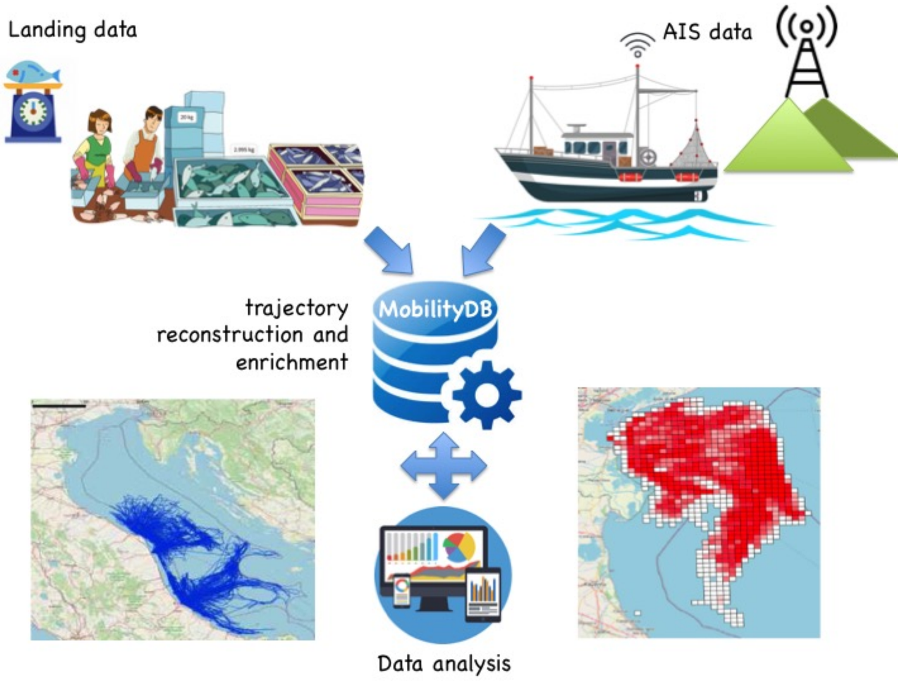


Figure 1: Bird's eye view of the process: data sources, reconstruction and enrichment of trajectories and data analysis

Year	Number of vessels	Number of transactions
2015	71	64180
2016	79	70017
2017	80	71716
2018	76	72165

Table 1: Dimension of the Landing dataset

Year	AIS data	Number of trajectories
2015	29757601	11280
2016	38519864	11130
2017	21247207	35335
2018	25098120	9549

Table 2: Raw AIS data vs trajectories

present in the AIS (MMSI code, vessel name and the call sign) with those of the European Fleet Register, which supplies specific information on the vessels (i.e., primary and secondary gear, length overall, gross tonnage, etc.). All the data given by the AIS (i.e., data position, speed, time, MMSI) were used to identify the fishing tracks and analyze the fishing activities (fishing, not fishing).

Daily landing reports. Landing dataset was obtained from the Chioggia's Fish Market, whose harbor hosts one of the main fishery fleets of the Adriatic Sea. This dataset consists of daily landings (catch amounts in kilogram) for 104 commercial species caught during four years, from January 2015 to December 2018 in the Northern Adriatic Sea. The records pertain around 80 fishing vessels, and contains a total of 278078 transactions over the four years, as detailed in Table 1.

2.2 Trajectory reconstruction and enrichment

Trajectories are reconstructed by linear interpolation of the raw AIS data. While performing the reconstruction raw data are cleaned: all the points implying movements that are not physically feasible due to a maximum possible boat speed are removed. Next, in order to organize the data into distinct trajectories followed by the fishing boats, we apply two criteria: a new trip begins *a*) when the vessel is inside a port area and there is no transmission for longer than a fixed time, or *b*) there is an AIS datum outside a port area and the immediate previous AIS datum is inside a port area and the time period between the two AIS data is greater than 20 minutes. The first condition corresponds to the fact that the vessel ends a trip, it switches off the AIS, it is docked at the port and after a while it starts a new trip. The second one corresponds to a situation in which a vessel leaves out of the port and then it starts transmitting when it is outside the port (20 minutes is the minimum time a vessel takes to leave the port). A detailed analysis reveals that some fishing vessels, after entering the port area at the end of a trip, continue to transmit their position. In this way, none of the above criteria is met. This causes a wrong trip reconstruction in which two or more trips are considered as a unique trip with a duration of several days. Hence, to avoid this phenomena we remove the AIS data transmitted inside the port when the vessel returns to a port. In Table 2 we report the dimension of the original AIS datasets and the resulting number of trajectories.

A trajectory, resulting from the reconstruction, is a sequence of segments, obtained by connecting consecutive AIS points. It is enriched with the following information:

- MMSI, boat identifier;
- trip duration (in hours);
- trip length (in meters);
- total time of fishing activity (in hours);
- total length of the fishing activity (in meters);
- date and time of the trip departure and conclusion;

ID	Description
0	normal trip
1	no transmission for more than 30 minutes outside a port area
2	trip always inside a port area
3	trip duration exceeds the 24 hours.

Table 3: Values of the *anomaly* attribute

ID	Activity description
0	in port
1	exiting from port
2	entering to port
3	fishing
4	navigation.

Table 4: Values of the *activity* attribute

- total number of segments with more than 30 minutes between two consecutive AIS transmissions;
- *anomaly*, a code specifying whether the trip presents an anomaly or not and the kind of anomaly.

The *anomaly* attribute highlights some strange behaviour of the fishing vessel. Possible anomalies are:

- the time interval between two consecutive AIS data is longer than 30 minutes outside the port, suggesting some points could be missing (*anomaly* is set to 1);
- a boat remains inside a port area for the whole trip (*anomaly* is set to 2);
- the duration of the trip exceeds the 24 hours (*anomaly* is set to 3);

If none of the above cases occurs, the trip is considered as normal and *anomaly* is set to 0. Table 3 summarizes the possible values of the *anomaly* attribute.

It is worth noting that through the MMSI, we can obtain further information on the vessel, such as its name and the fishing gear. Each segment in the trajectory is in turn annotated with:

- speed;
- position of the segment with respect to the port areas;
- activity of the boat within the segment;
- length of the segment;
- time spent in the segment;
- transmission.

The *activity* attribute describes what the vessel is doing according to Table 4. The *in port*, *exiting from* port and *entering to* port situations can be deduced from the position of the extremes of the segment w.r.t. 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. The considered gears and their minimum and maximum speed during the fishing activity are reported in Table 5.

The attribute *transmission* records whether the end points of the segment have a time distance greater than 30 minutes. If this happens the attribute is set to 1, otherwise to 0. As explained above, the presence of segments with transmission set to 1 allows for the detection of an anomalous behaviour of the trajectory.

These trajectories are modeled as a *multiple aspect trajectory*, following MASTER model [6]. Actually, as minimum granularity to attach semantic information, we do not consider a single

Gear description	ID	min. speed	max. speed
Small bottom otter trawl	SOTB	3.704	8.334
Large bottom otter trawl	LOTB	3.704	8.334
Pelagic pair trawl	PTM	3.704	10.186
Rapido	RAP	7.408	12.964

Table 5: Gears and their minimum and maximum fishing speed (in km/h)

spatio-temporal point as in the original MASTER model, but segments. This is motivated by the fact that we want to highlight the presence of homogeneous trajectory portions, which are the appropriate granularity level for our analyses. According to the MASTER model classification, the information listed above can be classified as *long-term aspects*, (those associated with the full trajectory), and *volatile aspects* (those associated with the segments).

By using the MASTER model we are able to represent different aspects of our trajectories in a uniform and simple way. Moreover, this representation allows us to perform complex queries merging spatial, temporal and semantic features. In the rest of the paper, we denote by T the resulting set of multiple aspect trajectories.

2.3 Catch distribution

We next describe how to merge the trajectories of the fishing vessels with the daily landing reports provided by the Chioggia fish market. The latter dataset contains information about each trading transaction, including the landing date, MMSI of the seller, the species, and the quantity of fish. Note that we work on a subset of the set of reconstructed multiple aspect trajectories. In fact, we exclude from our analysis, fishing vessels that do not sell their fish in Chioggia, trajectories with anomaly 2, i.e., the ones that do not leave the port area, and trajectories that do not have any fishing activity.

In order to perform the merge we need to associate each fish market transaction with a trajectory of the vessel having the specified MMSI. To accomplish this task, for each transaction, we select the vessel trip with the most recent arrival in the port (before 4 PM of the landing date). Arrivals after 4 PM are associated with transactions occurring the next day. The quantity (weight) of fish assigned to a trajectory is called a *catch*.

In order to distribute the fish associated with a trajectory over its fishing segments we follow two different approaches:

- uniform distribution, and
- weighted distribution.

In the first case, the catch is uniformly distributed along the fishing segments of the corresponding trajectory. Each fishing segment of the trajectory is associated with a fraction of the total amount of fish, proportional to its length. We consider separately each species that the fishing vessel caught.

Definition 2.1 (Uniform distribution). Let tr be a trajectory and let *catch* the record containing the quantities of the different species associated with the trajectory tr. Given a segment s belonging to tr with activity set as *fishing* and a species sp, the *uniform catch* for segment s and species sp is defined as

$$d_U(s, sp) = \frac{s.len}{tr.len\_fishing} * catch.sp \quad (1)$$

where

- $tr.len\_fishing$ is the attribute storing the total length of the fishing activity for the trajectory tr;

- *s.len* is the length of the segment;
- *catch.sp* selects the quantity of a certain species *sp*.

Clearly the assumption of uniform catch distribution is a simplification of reality. We consider also a refinement based on a so called *weighted distribution*. The idea is that the areas where more vessels are fishing, during a given time period, are more likely to have higher catch rates.

In order to implement this technique, we need to suitably partition the fishing area because it becomes crucial to evaluate the number of fishing vessels present in a certain zone. We decided to divide the Northern Adriatic sea into a square grid with 3×3 km cell size. The size has been chosen in agreement with the environmental scientists based on the dimension of the fishing vessels and their behaviour during the fishing activity.

The introduction of the grid leads to a further segmentation of the trajectories. In fact, each segment that spatially crosses one or more cells of the grid needs to be split into smaller segments in such a way that each portion is completely inside a single cell. Moreover, since we deal with a spatio-temporal grid, all segments spanning over two days are split into two smaller segments by taking as extra point the interpolated position at midnight.

In order to compute the weighted distribution, we associate a coefficient to each spatio-temporal cell of the grid.

Definition 2.2 (fishing coefficient). Let *c* be a spatio-temporal cell and *sp* a species. The *fishing coefficient* of cell *c* for the species *sp* is defined as follows:

$$\alpha(c, sp) = |\{tr \in T \downarrow sp \mid tr \cap c \neq \emptyset\}| * \Sigma_{tr \in T \downarrow sp} \Sigma_{s \in tr \cap c \wedge s.activiy=fishing} s.len \quad (2)$$

where

- $T \downarrow sp$ is the set of trajectories having a landing report with the species *sp*;
- $tr \cap c$ returns the intersection between the trajectory *tr* and the cell *c*;
- *s.activity* and *s.len* are respectively the attributes of segment *s* storing the activity and the length of the segment.

The coefficient $\alpha(c, sp)$ combines the number of fishing vessels and the amount of fishing activity they perform in the cell, hence it provides a measure of the fishing activity in the cell. Note that the coefficient depends on the species. Hence, for each species *sp*, we select only the trajectories having a landing report for the given species *sp*.

Since it is natural to expect that vessels will mostly concentrate in fishy areas, the intuition is that cells where the fishing coefficient is higher will have higher catch rates. This leads to the idea, formalised below, of using such coefficient as a weight when distributing catches over a trajectory.

Definition 2.3 (Weighted distribution). Let tr be a trajectory and let *catch* the record containing the quantities of the different species associated with the trajectory tr. Given a segment *s* belonging to *tr* with activity set as *fishing* and a species *sp*, the *weighted catch* for segment s and species *sp* is defined as

$$d_W(s, sp) = \frac{\alpha(s.cell, sp) * s.len}{\Sigma_{s' \in tr \wedge s'.activity=fishing}(\alpha(s'.cell, sp) * s'.len)} * catch.sp \quad (3)$$

where *s.cell* is the unique cell the segment *s* belongs to.

When distributing the catch over the segments of the trajectory tr, again only segments which are classified as fishing are considered. The difference is that in this case each segment *s* receives a weight which is proportional not only to the length *s.len* of the segment but also to the fishing coefficient $\alpha(s.cell, sp)$ of the cell the segment belongs to.

2.4 Implementation

To construct and store the set of multiple aspect trajectories, we used MobilityDB [15], an open source extension to the PostgreSQL database system [10] and its spatial extension PostGIS [9]. It provides temporal types and spatio-temporal operators that ease the management of moving objects.

One main feature of MobilityDB is that it offers a construct for representing the evolution of a value during a sequence of time instants. The values between successive instants are interpolated using a linear function. Clearly, this construct perfectly suites the representation of trajectories, which are reconstructed from a sequence of spatio-temporal data. In our case, the spatio-temporal points are the AIS data aggregated on the basis of the trajectory id. We created a set of objects of type `tgeompoint`, which is a temporal type modelling a point changing its position along a time period.

Next, the function `trajectory` is applied to these objects, and a `geometry` value is returned. In this way the trajectory can be visualized. In our work, for visualizing trajectories and the result of our analyses, we used QGIS [11], an Open Source GIS that supports viewing, editing, and analysis of geospatial data. For instance, Figure 2(left) shows the sequence of AIS data, i.e., the sequence of spatio-temporal points, related to the trip of a fishing boat, whereas Figure 2(center) illustrates the *continuous* representation of the same trip obtained by using the MobilityDB construct. The interpolation is internally implemented by the system, with the dual advantage of raising the user from this task and simplifying queries and analyses.

MobilityDB provides a lot of spatio-temporal operators to handle trajectories. For instance, `startTimestamp` and `endTimestamp` return respectively the first and last time instant among a set of time instants and this can be useful to extract the beginning and ending points of a trajectory; `getValue` returns a value at a particular time instant. There are operators to check topological relations between trajectories, like `tintersects`, `tdisjoint`, and others to compute distances. Interestingly the results of these operators are values changing in time. In fact, it can happen that at certain time periods trajectories enjoy the relations whereas at other ones they do not, and the distance between the objects can vary depending on the movement of the objects themselves. For instance, the user can check whether a fishing vessel respects the rule that it can fish only at a distance greater than three nautical miles from the coast and eventually detect where and when the ban has not been observed.

MobilityBD allows for an easy representation of semantic trajectories where semantic attributes can be modelled as temporal types. This means that we can model in a single table both the sequence of spatio-temporal points forming a trajectory and information associated with the whole trajectory itself, such as the MMSI of the vessel, the duration and length of the trajectory. Moreover, a trajectory can be segmented and each segment can be stored as a temporal type. Even in this case we can add other attributes modelling features of the segment itself, such as the speed, the activity, the transmission and the quantity of caught fish. In Figure 2(right) the different colours describe the activities of the fishing vessel. They allow the user to immediately detect where the vessel is fishing and also the shape of the movement.

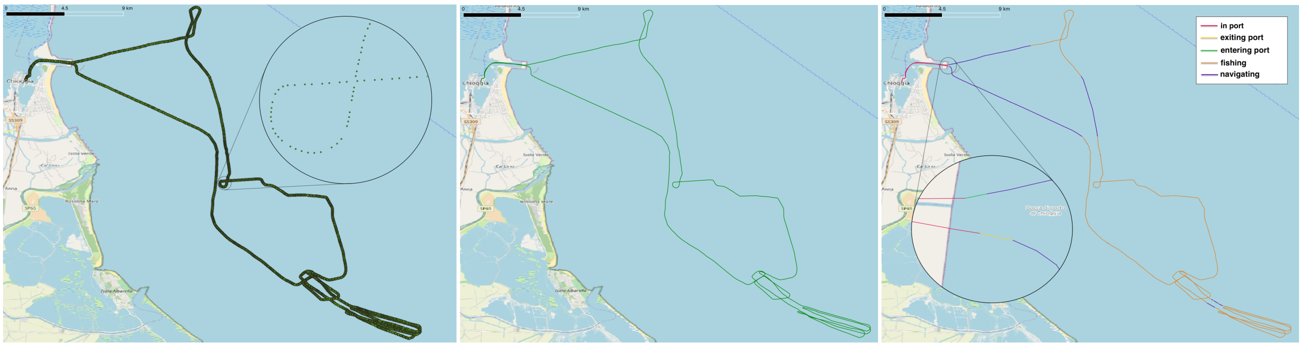


Figure 2: Trajectory visualisation as a sequence of spatio-temporal points (left), as a *continuous* function (center), and as a semantic object where the *activity* attribute is highlighted (right)

For instance, the figure highlights several circular movements and the experts have confirmed that they are typical of this kind of fishing activity.

Finally, MobilityDB provides support for the GiST (Generalized Search Tree) and SP-GiST (Space-Partition GiST) indexes, which can be created for table columns of temporal types. We used such indexes for accelerating spatial, temporal and spatio-temporal queries.

3 DATA ANALYSIS AND DISCUSSION

In this section we present some analyses performed with MobilityDB on the obtained spatio-temporal database of the Northern Adriatic sea.

The first analysis aims at visualizing the regions where there are transmission problems. We exploit the *anomaly* attribute and in particular we investigate trajectories having this attribute set to 1. In Figure 3 we show for each cell, the percentage of trajectories that got disconnected from the AIS for a time period greater than 30 minutes while crossing that cell with respect to the total number of trajectories passing through the cell. Looking at Figure 3, it is evident that the no-transmission anomaly has decreased a lot from 2015 to 2018. In fact, in 2015 the area where this percentage is over 50% is very large and it covers almost the whole fishing zone. Instead, in 2018 this phenomenon is localized in few areas, i.e., close to the coasts and along the territorial waters borders. Moreover, in 2018 there are also some isolated cells in the southern part.

The low spatial coverage of AIS is a well-known issue and the amount of missing data can vary substantially between vessels as discussed in [12]. Our analysis reveals that data from 2018 are more reliable and can be useful for detecting areas where the AIS signal is not received well, like the isolated cells in the southern portion of the sea area under investigation.

This analysis is an example of how the semantic knowledge hidden in a single attribute, such as the *anomaly* attribute, can be useful to greatly improve the general spatio-temporal knowledge of the domain of interest. On one hand the progressive low-coverage reduction of AIS data is *per se* a highly valuable information for ecologists and policy makers, since this ensures the reliability of the collected data. On the other hand, the proposed implementation allows the experts to continuously monitor the degree of coverage and eventually decide to add further terrestrial AIS receivers.

The second and third analyses take advantage of the catches distribution and try to infer some knowledge on key species in the area. In fact, spatializing the distribution of catches has several important applications. For instance, it allows us to obtain knowledge about the seasonal variation of the fishing grounds and this, in turn, is useful for explaining the fisherman behaviour, as well as to better understand the seasonal migration of a target species. Figure 4 reports the seasonal spatial distribution of cuttlefish, *Sepia officinalis*, aggregated by fishing gears (SOTB, LOTB and RAP) in 2018. Cuttlefish is one of the main target species of the Adriatic Sea, hence it is an ideal case study for showing a seasonal migratory behaviour. It is worth noting that the most productive seasons were autumn and winter, with two high density areas, one nearer the coast and the other one more offshore, at the border with the Croatian waters. In spring the catches resulted more scattered, while in summer the catch area was more defined and localized closer to the Italian coast. This is in line with the general ecological knowledge about the behaviour of the species, hence, the catches data correctly reflect cuttlefish seasonal spatial distribution behaviour. Figure 4 reports also the comparison between the uniform (A) and the weighted (B) distribution maps of cuttlefish *Sepia officinalis* in 2018. It is evident that the maps obtained with the weighted distribution (B) result more defined, allowing to better identify the fishing grounds of cuttlefish.

Another important application of the spatial distribution of catches is the detection of different fishing grounds among years. As an example, the catches of anchovy, *Engraulis encrasicolus*, recorded in winter 2015, 2016, 2017 and 2018 and distributed according with the weighted distribution are reported in Figure 5. The maps clearly show how the fishing grounds, and consequently the distribution of anchovies, changed along the years. In particular, a gradual reduction of the fishing grounds is observed from 2016 to 2018. This is clearly a relevant information for both ecologists and policy makers: if the fishing ground reduction is the result of an over exploitation of the species they can adopt appropriate countermeasures.

To end up, we would like to point out that these are only a few examples of the analyses that can be performed by using the dataset of multiple aspect trajectories. For instance, we can focus on vessels equipped with a specific fishing gear (i.e., LOTB, SOTB, RAP and PTM) and determine their fishing grounds and the corresponding degree of exploitation. This fine-grained analysis could help to reveal different efficiency degrees of fisheries that, in turn, could constitute a basis to implement specific management actions for these activities. Moreover, we can vary our analysis according to different time periods and consider only certain sea

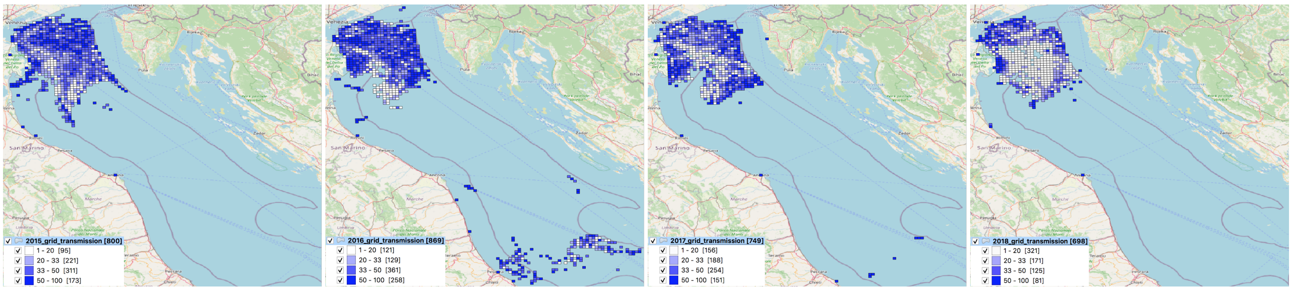
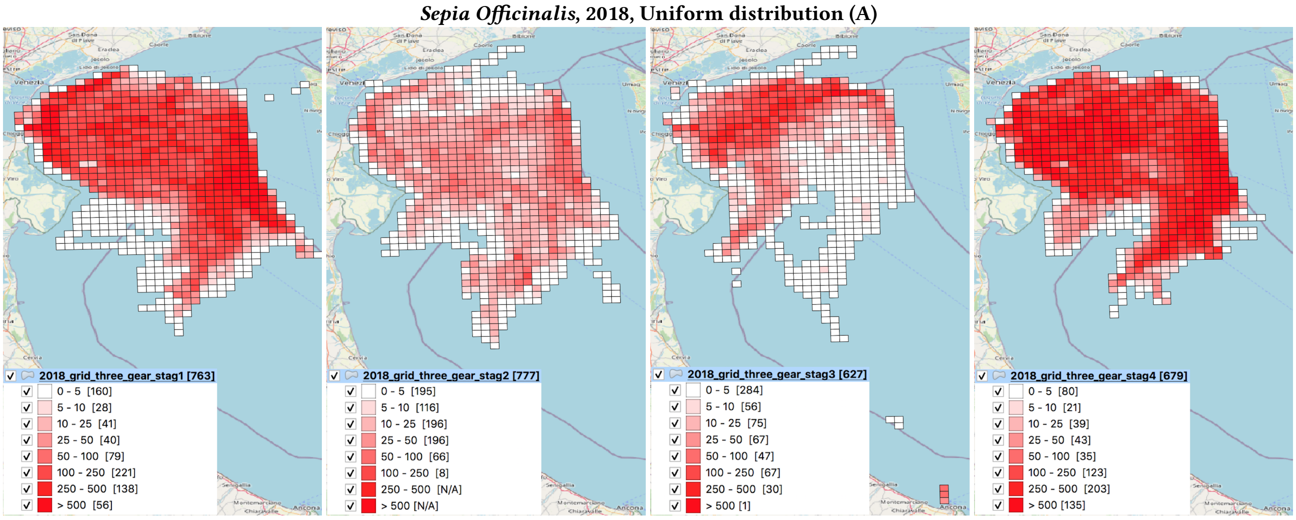


Figure 3: Spatial distribution of the no-transmission anomaly, years 2015, 2016, 2017 and 2018 (from left to right)

***Sepia Officinalis*, 2018, Uniform distribution (A)**



***Sepia Officinalis*, 2018, Weighted distribution (B)**

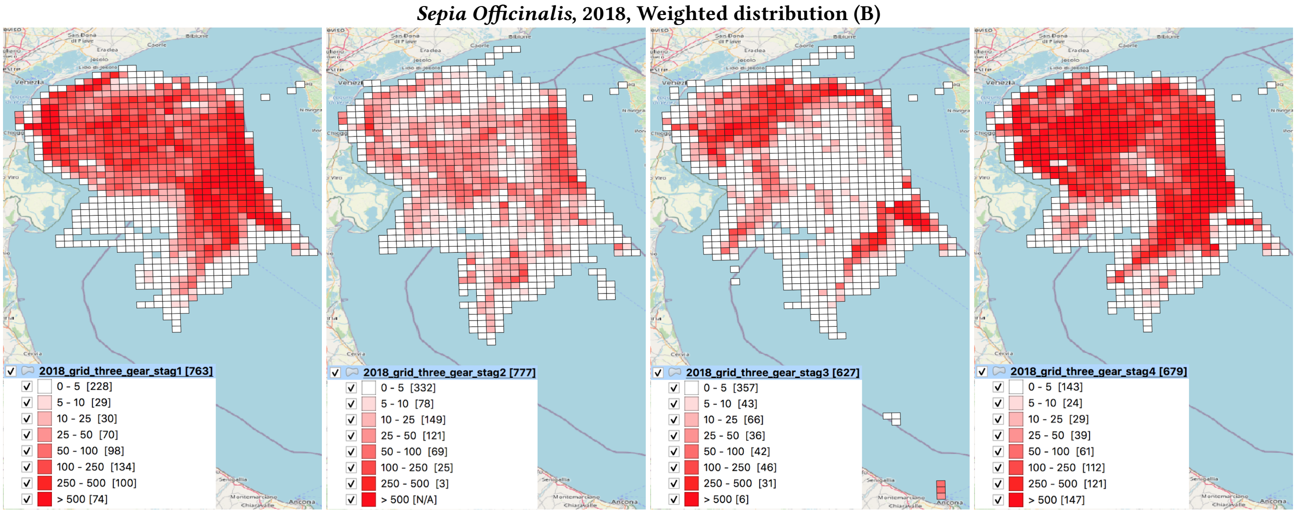


Figure 4: Comparison between uniform (A) and weighted (B) distribution of cuttlefish *Sepia officinalis*, aggregated by seasons (winter, spring, summer and autumn 2018)

areas. For instance one could focus on protected areas, like the Pomo Pit or the Sole Sanctuary. We can also select the behaviour of single trajectories satisfying complex conditions concerning both their movements and their semantic annotations by using the operators available in MobilityDB.

4 CONCLUSIONS

In this paper we built, implemented and analysed a spatio-temporal database of the vessels trajectories in the Northern Adriatic sea. We started from the terrestrial AIS data of the area of interest and the fish reports of the main fish market, Chioggia, for the years 2015, 2016, 2017, 2018. We determined the trajectories and introduced semantic attributes able to unveil interesting information and aspects of the original data themselves. Moreover,

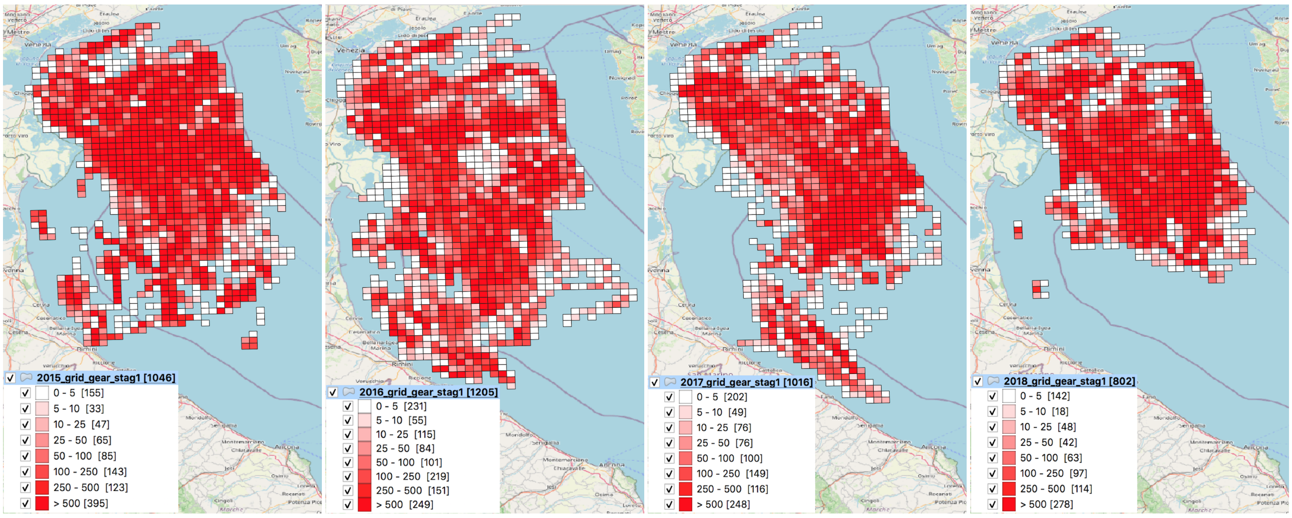


Figure 5: Spatial distribution of anchovy *Engraulis encrasicolus* in winter, years 2015, 2016, 2017 and 2018 (from left to right)

we gave a formal definition of two different catch distribution techniques, the *uniform* and *weighted*, with the aim of putting them at work and comparing their behavior.

Additionally, we implemented the spatio-temporal database using MobilityDB, thus ensuring a suitable environment for storing, querying and visualizing trajectories of moving objects.

The ecological experts proposed some analyses on the obtained database. We started with the analysis of the transmission anomalies – stored as a new semantic feature – that allowed us to acknowledge a concrete and progressive improvement of the data completeness in the years 2015-2018, thanks to the growing use of the AIS transmission systems in the fishing vessels and to the increasing AIS data receiving coverage.

We proceeded then with the analysis of the two proposed distribution techniques. It turned out that the weighted distribution is actually a refinement of the uniform one, able to better define the fishing ground of the species of interest. Besides, we showed how the use of semantic trajectories can provide an assessment of the fishing activities, capturing spatial and temporal patterns.

All these results put the best possible basis for a favorable application of prediction methods, which is the next step to be done on the obtained database. In particular, first we would like to test whether the Random Forest prediction results reported in [1] improve thanks to the availability of 2017 and 2018 data. Then, we would like to experiment other prediction techniques, such as lag variables [14], or modern time series prediction.

Finally, another interesting line of research is to extract fishing patterns, like the circular one illustrated in Figure 2, or anomaly behaviour, as investigated in [5, 8].

ACKNOWLEDGEMENTS

This paper is supported by the MASTER project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Sklodowska Curie grant agreement N. 777695.

REFERENCES

[1] Pedram Adibi, Fabio Pranovi, Alessandra Raffaetà, Elisabetta Russo, Claudio Silvestri, Marta Simeoni, Amílcar Soares, and Stan Matwin. 2019. Predicting Fishing Effort and Catch Using Semantic Trajectories and Machine Learning. In *Multiple-Aspect Analysis of Semantic Trajectories - First International Workshop, MASTER 2019 (Lecture Notes in Computer Science)*, Vol. 11889. Springer, 83–99.

[2] Vania Bogorny, Chiara Renso, Artur Ribeiro de Aquino, Fernando de Lucca Siqueira, and Luis Otavio Alvares. 2014. Constant–a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS* 18, 1 (2014), 66–88.

[3] Stefan Brüggemann, Konstantina Bereta, Guohui Xiao, and Manolis Koubarakis. 2016. Ontology-based data access for maritime security. In *European Semantic Web Conference*. Springer, 741–757.

[4] Christophe Claramunt, Cyril Ray, Elena Camossi, Anne-Laure Jousselme, Melita Hadzagic, Gennady L. Andrienko, Natalia V. Andrienko, Yannis Theodoridis, George A. Vouros, and Loïc Salmon. 2017. Maritime data integration and analysis: recent progressand research challenges. In *Proceedings of the 20th International Conference on Extending Database Technology*. 192–197.

[5] Ioannis Kontopoulos, Konstantinos Chatzikokolakis, Dimitris Zissis, Konstantinos Tserpes, and Giannis Spiliopoulos. 2020. Real-time maritime anomaly detection: detecting intentional AIS switch-off. *Int. J. Big Data Intell.* 7, 2 (2020), 85–96.

[6] 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. 2019. MASTER: A multiple aspect view on trajectories. *Transactions in GIS* 23, 4 (2019), 805–822.

[7] Christine Parent, Stefano Spaccapietra, Chiara Renso, Gennady Andrienko, Natalia Andrienko, Vania Bogorny, Maria Luisa Damiani, Aris Gkoulalas-Divanis, Jose Macedo, Nikos Pelekis, et al. 2013. Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)* 45, 4 (2013), 42.

[8] Lucas May Petry, Amílcar Soares, Vania Bogorny, Bruno Brandoli, and Stan Matwin. 2020. Challenges in Vessel Behavior and Anomaly Detection: From Classical Machine Learning to Deep Learning. In *Advances in Artificial Intelligence - 33rd Canadian Conference on Artificial Intelligence (Lecture Notes in Computer Science)*, Vol. 12109. Springer, 401–407.

[9] PostGIS [n.d.]. *PostGIS: Spatial and Geographic objects for PostgreSQL.* https://postgis.net/

[10] PostgreSQL [n.d.]. *PostgreSQL Open Source Relational Database.* https://www.postgresql.org/

[11] QGIS [n.d.]. *QGIS: a free and open source Geographic Information System.* https://qgis.org/en/site/

[12] Jennifer L Shepperson, Niels T Hintzen, Claire L Szostek, Ewen Bell, Lee G Murray, and Michel J Kaiser. 2018. A comparison of VMS and AIS data: the effect of data coverage and vessel position recording frequency on estimates of fishing footprints. *ICES Journal of Marine Science* 75, 3 (2018), 988–998.

[13] Amílcar Soares, Renata Dividino, Fernando Abreu, Matthew Brousseau, Anthony W Isenor, Sean Webb, and Stan Matwin. 2019. CRISIS: Integrating AIS and Ocean Data Streams Using Semantic Web Standards for Event Detection. In *International Conference on Military Communications and Information Systems*.

[14] Hristos Tyralis and Georgia Papacharalampous. 2017. Variable Selection in Time Series Forecasting Using Random Forests. *Algorithms* 10 (2017), 114.

[15] Esteban Zimányi, Mahmoud Sakr, Arthur Lesuisse, and Mohamed Bakli. 2019. MobilityDB: A Mainstream Moving Object Database System. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*. 206–209.