Visual Analytics of Vessel Movement

Natalia Andrienko and Gennady Andrienko

Abstract Visual analytics techniques support the process of data analysis, reasoning, and knowledge building performed by a human analyst. The techniques combine interactive, human-controllable visual displays with interactive operations for data querying and filtering, data transformations, calculation of derived data, and application of computational techniques for analysis and modelling. We demonstrate the use of visual analytics techniques and procedures for analyzing Automatic Identification System (AIS) data. We begin with showing how visual analytics approaches can help in exploring properties of the data, detecting problems, and finding ways to clean and improve the data. Then we describe two analysis scenarios focusing on the events of vessel stopping and on the vessel traffic through the strait between the bay of Brest, France, and the outer sea. Thereby we show how different techniques are applied and combined.

1 Introduction

Human reasoning plays a crucial role in data analysis and problem solving. By means of reasoning, humans build and/or update knowledge in their mind. The knowledge includes understanding of data and understanding of the phenomena reflected in the data. Reasoning requires conveying information to the human's mind, and visual representations are best suited for this. Visual analytics, which is defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [18, p.4], devel-

Natalia Andrienko
Fraunhofer Institute IAIS, Sankt Augustin, Germany, and City, University of London, UK
e-mail: natalia.andrienko@iais.fraunhofer.de

Gennady Andrienko
Fraunhofer Institute IAIS, Sankt Augustin, Germany, and City, University of London, UK
e-mail: gennady.andrienko@iais.fraunhofer.de

ops approaches combining visualizations, interactive operations, and computational processing to support human analytical reasoning and knowledge building.

In this chapter, we demonstrate examples of using visual analytics approaches for exploration and analysis of Automatic Identification System (AIS) data [8]. Our example dataset consists of trajectories of vessels that moved between the bay of Brest, France, and the outer sea [14]. We first investigate the properties of the data and then focus on revealing and understanding patterns of vessel movements. The example analysis task is to study when, where, and for how long the vessels were stopping and to understand whether the events of stopping may indicate waiting for an opportunity to enter or exit the bay (through a narrow strait) or the port of Brest.

2 Exploration of the data properties

Knowing properties of the data that need to be analyzed is essential for performing valid analysis and drawing valid conclusions. Hence, before focusing on the primary analysis goal, it is necessary to explore the data for gaining understanding of their properties, identifying quality problems, and finding ways to solve or mitigate them. Possible quality problems in movement data [3, 17] include, apart from errors in spatial positions of objects and missing records for long time intervals, gaps in spatio-temporal coverage, low temporal and/or spatial resolution, use of the same identifiers for multiple objects, and others.

To explore the properties of the data we are going to analyze, we use aggregated representations of original and derived attributes. Thus, a frequency histogram of the lengths of the temporal spaces between consecutive records [17] shows us that the most frequent spacing is around 10 seconds, and smaller intervals also occurred quite frequently. Smaller peaks around 20, 30, 40, 60, 180, and 360 seconds corresponds to the required frequencies of position reporting depending on the vessel status (moving or stationary), movement speed, type of the vessel, and positioning equipment. Longer gaps between recorded positions may correspond to equipment malfunctioning or being off, or to periods when the vessels were out of the area covered by the data.

When trajectories are represented by lines on a map (Fig. 1, top), long straight line segments can be noticed. Most of these segments correspond to *spatio-temporal gaps* in the trajectories, i.e., absence of recorded vessel positions during long time intervals. Hence, such segments must be excluded when it is necessary to analyze the paths of the vessels or to aggregate the trajectories into overall traffic flows; otherwise, the results will be wrong and misleading. A suitable way to exclude spatio-temporal gaps is to *divide the trajectories* by these gaps: the point preceding a gap is treated as the end of the previous trajectory, and the following point is treated as the beginning of the next trajectory. A gap is defined by choosing appropriate thresholds for the spatial and temporal distances between consecutive trajectory points. Suitable thresholds are chosen based on the statistics of the distances in the data. Thus, for the data presented in Fig. 1, 78.6% of the spatial distances are below

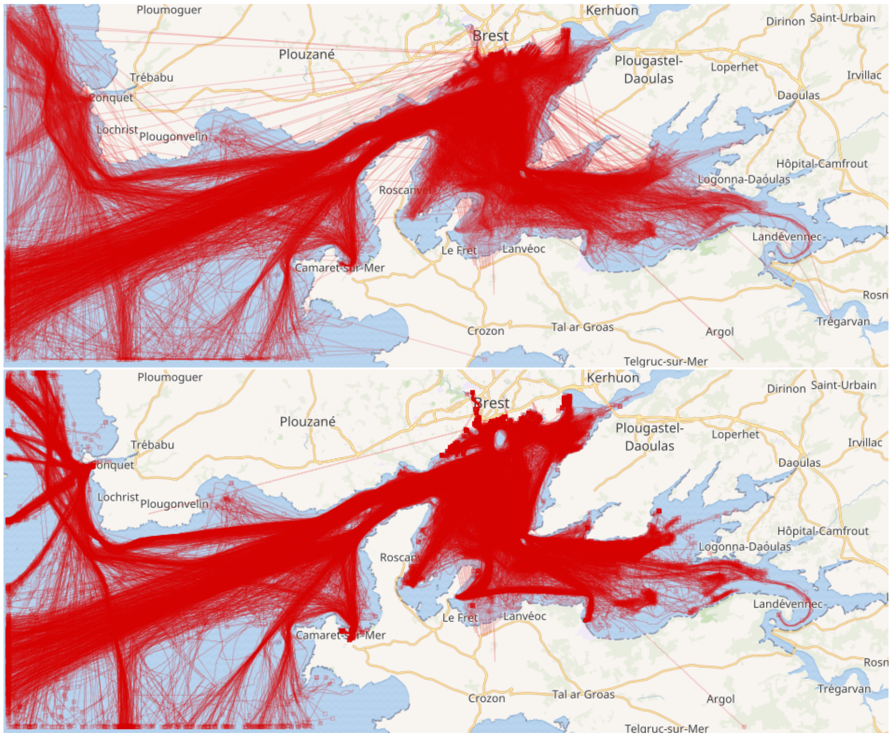

Fig. 1 Top: Long straight line segments in trajectories correspond to *spatio-temporal gaps*, i.e., long time intervals in which position records for the vessels are missing. Bottom: The result of dividing the trajectories by the spatio-temporal gaps in which the spatial distance exceeded 2 km and the time interval length exceeded 30 minutes.

10 meters, 12.1% are from 10 to 50 m, and 6.4% from 50 to 100 m. Only 0.5% of the distances exceed 250 m, 0.2% exceed 500 m, 0.12% are over 1 km, and 0.06% are over 2 km. Hence, a suitable spatial threshold may be from 0.25 to 2 km, depending on the intended spatial scale of the analysis. Only 0.4% of the temporal differences exceed 10 minutes, 0.22% exceed 20 minutes, and 0.19% exceed 30 minutes. Taking the spatial threshold of 2 km and temporal threshold of 30 minutes defines 1,852 spatio-temporal gaps in the data, which is 0.018% of the total number 10,446,156 of the available position records for the territory shown in Fig. 1. The image at the bottom of Fig. 1 shows the result of dividing the trajectories by these gaps. From the original 392 trajectories, the division produced 8,227 trajectories.

As any real data, the data we need to analyze have quite many errors, particularly, wrong positions. Some of these errors, such as positions on the land far from the sea, are very easy to spot visually (Fig. 2). In the trajectories presented in Fig. 1, such obvious errors have been already cleaned. Other positioning errors may be more difficult to detect, especially in a large dataset. A good indication of a recorded position being out of the actual path of a vessel is an unrealistically high value

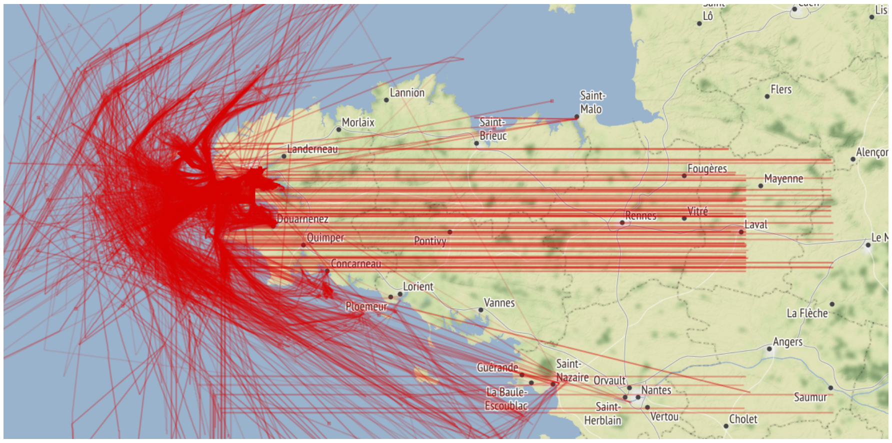

Fig. 2 Some trajectories include positions located on land far from the sea.

of the *computed speed* in the previous position. The speed is computed as the ratio between the distance to the next position and the length of the time interval between the positions. The computed speed may differ from the measured speed, which is recorded in the data. In the data shown in Fig. 1, the maximal recorded value of the measured speed is 102.2 knots (189.2744 km/h) and most of the values (88.5%) are below 36 knots (66.67 km/h), whereas the computed speed values reach as high as 31626 km/h. There are about 4000 points (0.1% of all) with the computed speed values exceeding 200 km/h (about 108 knots). Among these 4000 points, the median being 5161 and the lower and upper quartiles being 2432 and 9863, respectively. These values undoubtedly indicate positioning errors in the data records. Occasionally occurring singular outlying positions are easy to identify and exclude from the trajectories; however, there may be more difficult cases. It is useful to have a close look at trajectories containing many points with extremely high speed values.

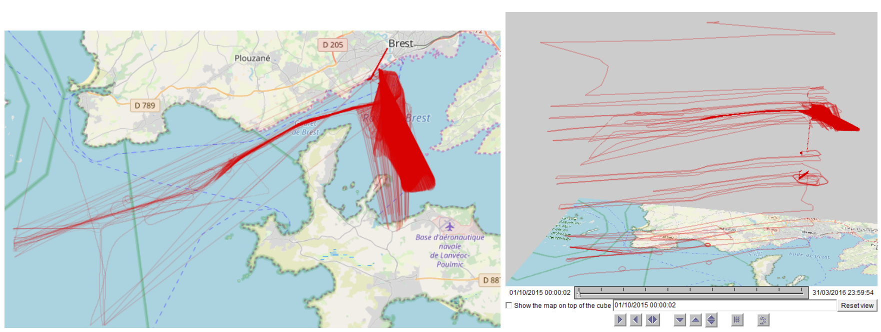

Fig. 3 A trajectory with an extremely high number of outlying points.

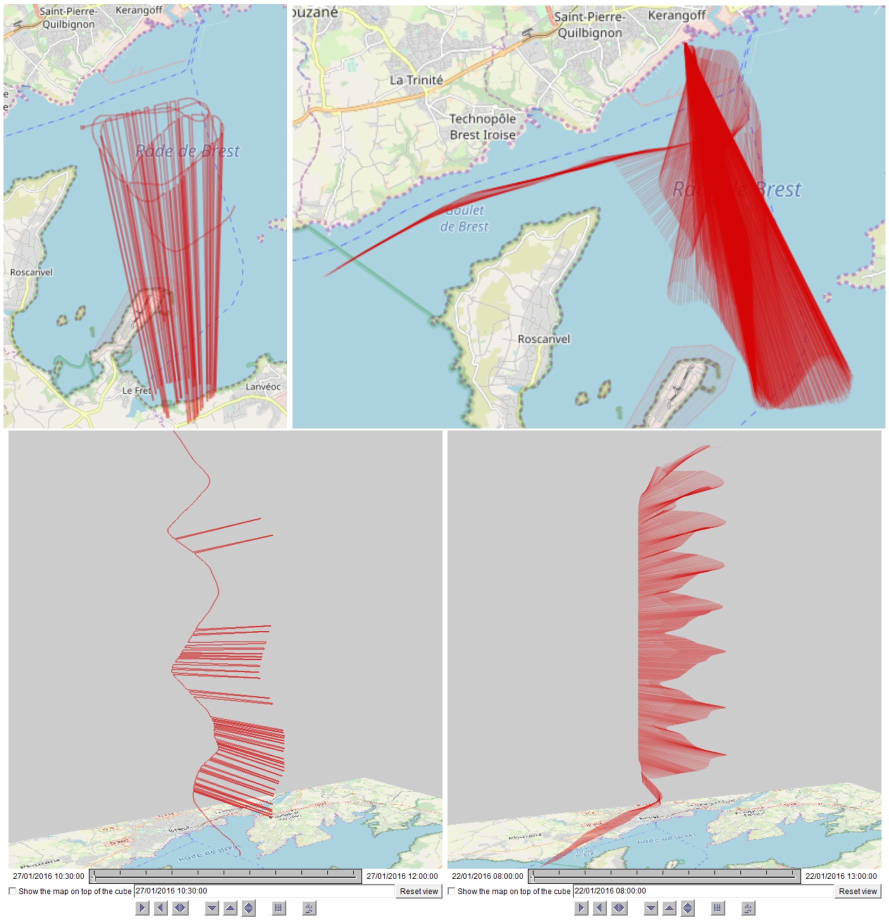

Fig. 4 Detailed investigation of the periods with high numbers of positioning errors.

An example of a trajectory containing extremely many points with very high values of computed speed is shown in Fig. 3. The trajectory has a very long duration – 6 months. By looking at the map only, it is hard to understand what is shown. It is useful to look at this trajectory in a space-time cube (STC), which is a three-dimensional displays with two dimensions in the cube base representing space and the third, vertical dimension representing time. The STC on the right of Fig. 3 shows that there were a few short periods during this time in which the trajectory looks strange. In Fig. 4, the appearance of the trajectory in two such periods is shown in more detail on maps (upper images) and in STC views (lower images). The images indicate that correctly recorded positions alternate with erroneous recording. Moreover, the wrong positions seem to be displaced with respect to the correct positions in a systematic manner rather than randomly. In the part of the trajectory shown on the upper and lower left, the displacements occurred with varying temporal frequency, while the directions and distances of the displacements were similar. The part of the trajectory shown on the upper and lower right has a much more complicated shape. The displacement distances increase and decrease in a periodic manner, while the angles change gradually. Such a pattern may mean that two or more simultaneously moving vessels might refer to the same vessel identifier in reporting their positions. Arranging positions of different vessels in a single trajectory results in an unrealistic path shape and extremely high values of the computed speed in the points due to high distances between consecutive points. Still, it is not unlikely that the strange shape in our case emerged due to malfunctioning of the positioning device.

If such errors may have an impact on the subsequent analysis (e.g., in analyzing paths or flows), it is reasonable to try to exclude the trajectory fragments with high numbers of errors, or even the whole trajectories. However, it is quite difficult to exclude in an automatic way frequently occurring shifts, as in Fig. 4, left, or to separate movements of different vessels, as in Fig. 4, right.

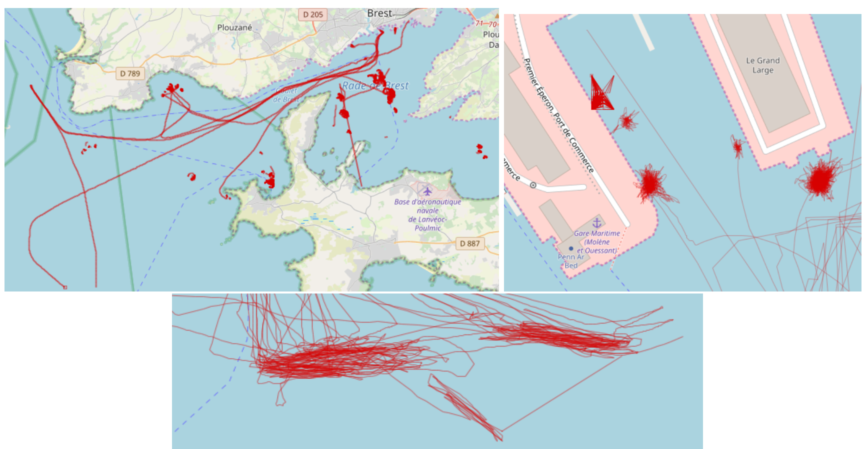

Fig. 5 Fragments of trajectories with wrongly reported navigation status. The reported status in the upper left and bottom images is “at anchor”, whereas the vessels were actually moving. On the upper right, the reported status is “under way using engine”, while the vessels remained at the same places and should have reported “at anchor”.

Errors may occur not only in positions but also in attribute values associated with the positions. For our intended study, the values of the attribute ‘navigational status’ are relevant. Particularly, the value 1 means “at anchor”, which may help us to find the anchoring events, and the value 7 means “engaged in fishing”, which may help us to exclude the trajectories of fishing boats from our analysis. We want to exclude the fishing boats because we expect their movement behaviors to be quite different from those of the vessels purposefully travelling from a certain origin to a certain destination and not performing any activities on the way. Thus, we can expect that anchoring of fishing boats may be related to their fishing activities rather than with a busy traffic situation or crowded port area.

While the attribute reporting the navigational status is of high interest to us, it turns out that its values may be unreliable. Figure 5 demonstrates a few examples of wrongly reported navigational statuses. The upper left image shows a selection of points from multiple trajectories in which the navigational status equals 1, i.e., “at anchor”. It is well visible that the selection includes not only stationary points but also point sequences arranged in long traces, which means that the vessels were moving rather than anchoring. On the opposite, the upper right image demonstrates several trajectory fragments that look like hairballs. Such shapes are typical for stops, when the position of a moving object does not change but the tracking device reports each time a slightly differing position due to unavoidable errors in the measurements. The locations of the hairballs also signify that the vessels were at anchor or moored at the shore. However, the recorded navigational status is 0, which means “under way using engine”. The lower image shows trajectory fragments of several vessels. The character of their movements (back and forth repeated multiple times) indicates that they were fishing, i.e., the navigational status should be 7, but the value attached to the positions is 0, i.e., “under way using engine”.

Hence, in our study, we should not fully rely on the attribute values, and should also find ways to mitigate the possible impacts of the other errors we have detected.

3 Transformations of movement data

In the following analysis, we apply multiple transformations of the data. Movement data can be considered from several complementary perspectives [1]: trajectories, spatial events, dynamically changing situations over a territory, variation of presence of moving objects in selected places, or aggregated movements (flows) between places. For example, a sequence of positions of a vessel can be treated as its trajectory. An anchorage with its spatial position and temporal interval is an event. A situation can represent a spatial distribution of vessels at a given time or movement flows of vessels over a time interval. These diverse perspectives are supported by techniques for transformations between different possible representations of movement data, as illustrated Fig. 6.

The transformation scheme can be explained as follows. Typically, movement data are originally available as collections of records specifying spatial positions of moving objects (e.g. vessels) at different times. Such records describe *events* of presence (or appearance) of the moving objects at certain locations and specify the times when these events occurred. When all records referring to the same moving object are put in a chronological sequence, they together describe a trajectory of this object. Hence, trajectories are obtained by integrating spatial events of object appearance at specific locations. The trajectories can be again disintegrated to the component events. Particular events of interest, such as stops or zigzagged movement, can be detected in trajectories and extracted from them. A trajectory describing movements of an object during a long time period can be divided into shorter trajectories, for example, representing different trips of the object.

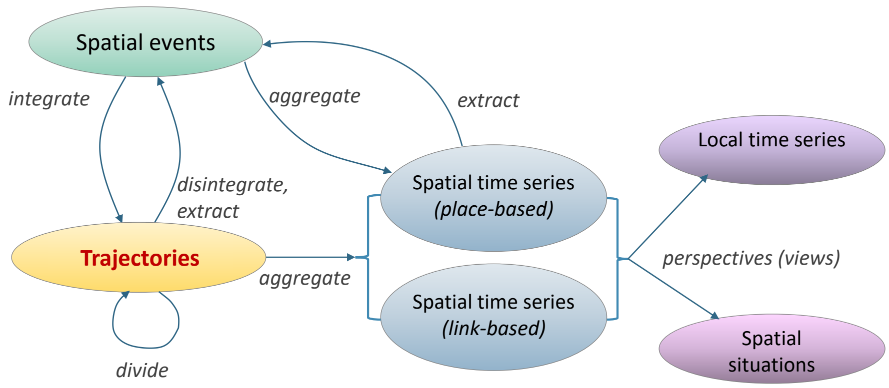

Fig. 6 Different representations of movement data and transformations between them (source: [1]).

Having divided the space into *compartments* (shortly, places) and time into *intervals*, it is possible to aggregate either spatial events or trajectories by places and time intervals. Place-based aggregation involves counting for each pair of place and time interval (1) the events that occurred in this place during this interval, or (2) the number of visits of this place by moving objects and the number of distinct objects that visited this place or stayed in it during the interval. Additionally, various summary statistics of the events or visits can be calculated, for example, the average or total duration of the events or visits. The result of this operation is time series of the aggregated counts (e.g. counts of stops or counts of distinct visitors) and statistical summaries associated with the places.

Link-based aggregation summarizes movements (transitions) between places and, thus, can be applied to trajectories. For each combination of two places and a time interval, the number of times when any object moved from the first to the second place during this interval and the number of the objects that moved are counted. Additionally, summary statistics of the transitions can be computed, such as the average speed or the duration of the transitions. The result of this operation is time series of the counts and statistical summaries associated with the pairs of places. The time series characterize links between the places; therefore, they can be called link-based. The term “link between place A and place B” refers to the existence of at least one transition from A to B.

Both place-based and link-based time series can be viewed in two complementary ways: as spatially distributed local time series (i.e., each time series refers to one place) and as temporal sequence of spatial situations, where each situation is a particular distribution of the counts and summaries over the set of places or the set of links. These perspectives require different methods of visualisation and analysis. Thus, the first perspective focuses on the places or links, and the analyst compares the respective temporal variations of the attribute values such as counts of distinct vessels in ports over days. The second perspective focuses on the time intervals, and the analyst compares the respective spatial distributions of the values associated with the places or links.

4 Detection and analysis of the anchoring events

The goal of our analysis is to investigate when, where, and for how long the vessels were stopping and to understand whether the stops may indicate waiting for an opportunity to enter or exit the bay of Brest through a narrow strait or the port of Brest. If many stops might have happened due to waiting, it may mean that the management of the traffic through the straight or the port services is sub-optimal and requires improvement.

4.1 Data selection and preparation

For our study, we selected the trajectories of the vessels that passed the strait connecting the bay of Brest to the outer sea at least once. We excluded the vessels that had, at some point, the navigational status 7, i.e., "engaged in fishing", for the reasons explained at the end of Section 2. We also excluded the vessels that never (during the time period covered by the data) had the navigational status 0, i.e., "under way using engine". The resulting selection consists of trajectories of 346 vessels.

From these trajectories, we selected only the points located inside the bay of Brest, in the strait, and in the area extending to about 20 km west of the strait. The shapes of the resulting trajectories are shown in Fig. 1. As we explained previously (Section 2), the long straight line segments correspond to periods of position absence. To exclude these segments, we divided the trajectories into sub-trajectories by the spatio-temporal gaps with distance thresholds 2 km in space and 30 minutes in time. This means the following: if there is a trajectory point such that its distance to the next point exceeds 2 km and the time difference exceeds 30 minutes, the trajectory is divided into two smaller trajectories. The first point is taken as the end of the first trajectory, and the next point is taken as the beginning of the second trajectory. The division gave us 6,346 smaller trajectories. To clean these trajectories from outliers, we removed the points whose distances from the previous and next points were more than 2 km. We further divided the trajectories by stops (segments with low speed) within the Brest port area and selected from the resulting trajectories only those that passed through the strait and had duration at least 15 minutes. As a result of these selections and transformations, we obtained 1,718 trajectories for further analysis. Of these trajectories, 945 came into the bay from the outer area, 914 moved from the bay out, and 141 trajectories include incoming and outgoing parts.

4.2 Detection and extraction of anchoring events

Our study aims at the detection of anchoring events in the trajectories. As explained in Section 2, we cannot rely on the values of the attribute 'navigational status', because they may be wrong. Assuming that, despite the errors we detected (Fig. 5), many vessels reported their anchoring correctly, we take the following approach. We extract from the database all trajectory points where the navigation status equals 1 (i.e., "at anchor") and the speed over ground is less than 2 knots. The extracted points are represented in Fig. 7 by purple circles drawn with a high level of transparency (90%). The circle symbols appear dark and thick where the points are clustered in space. We define 11 anchorage areas relevant to our study by creating buffer zones around the point clusters, excluding the clusters on the main traffic lanes (these clusters consist of points of singular vessels and may thus result from errors in data or indicate abnormal situations), within the port or attached to the shore. The zones, labelled by numbers from 1 to 11, are shown in Fig. 7.

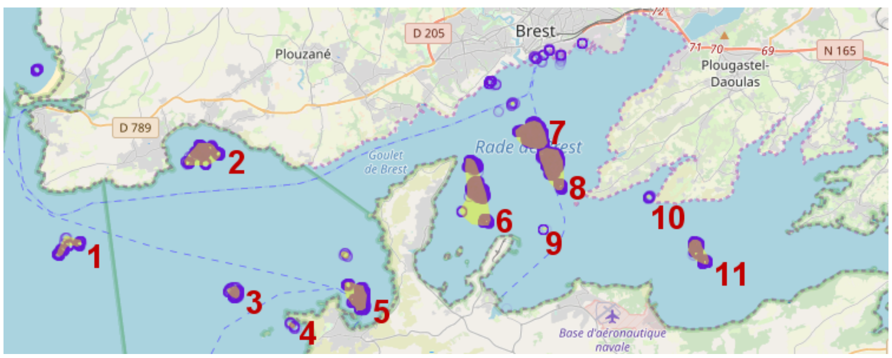

Fig. 7 Delineation of anchoring zones.

Having defined the anchorage zones, we marked the points of the trajectories as *probable anchoring* if, first, they belong to one of the zones and, second, have the speed over ground below 2 knots. There are 158 trajectories that contain points satisfying these conditions. Pitsikalis et al. [13] describe another approach to detecting anchoring events, which can be used for online detection of such events, while the analysis described in this chapter is performed offline.

Figure 8, top, shows the previously selected set of 1,718 selected trajectories. The segments corresponding to probable anchoring are colored in red. The lower left image shows the shapes of the trajectory segments corresponding to probable anchoring in more detail. In the lower right, only those 158 trajectories that contain probable anchoring events are shown in an STC view. The trajectories have been aligned in time by putting together their starting times. Such transformation of the absolute time references to relative makes the segments corresponding to probable

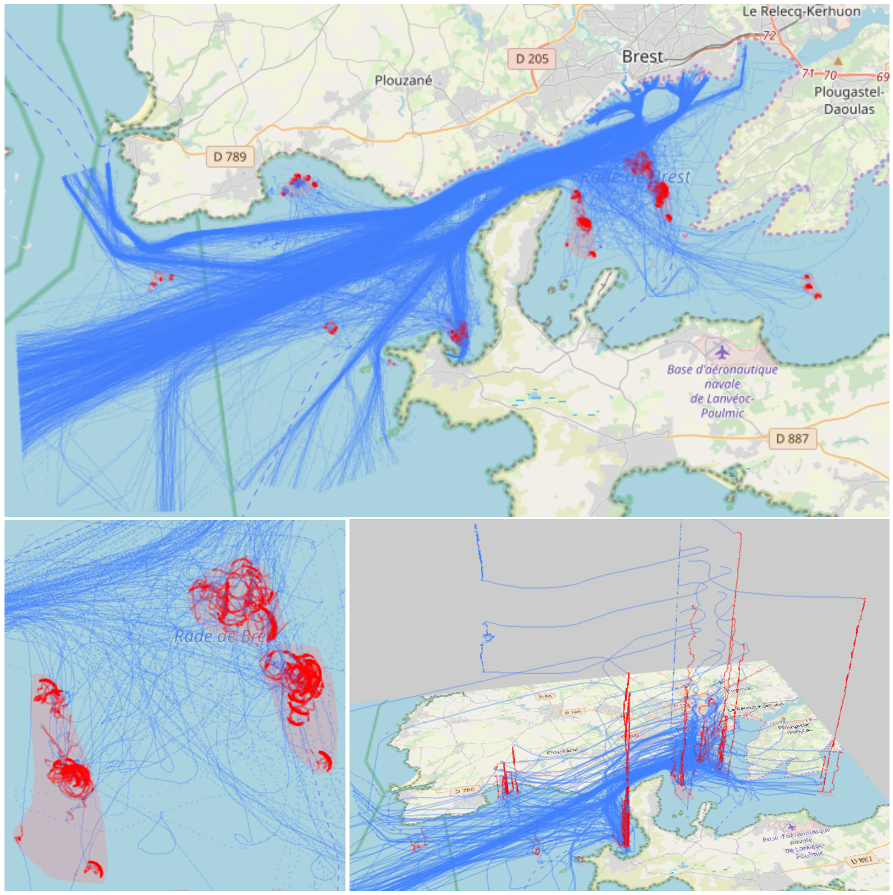

Fig. 8 The trajectories selected for analysis with the segments corresponding to probable anchoring marked in red. The lower left image shows an enlarged map fragment in more detail, and the STC on the right shows only the trajectories that contain probable anchoring; the trajectories have been temporally aligned.

anchoring better visible (they appear as vertical lines in the cube) and their duration (which is represented by the lengths of the vertical lines) easier to compare.

We extract the probable anchoring events (i.e., the segments identified as probable anchoring) from the trajectories to a separate dataset. We obtained 327 events in total with the duration ranging from 1 minute to 130.5 hours. We deem it unlikely that a vessel would actually stay at anchor for a very short time, since the anchoring and unanchoring procedures are not instantaneous. So, we want to exclude unrealistically short events. To determine the minimal realistic duration, we find the shortest event that included points with the reported navigational status 1, i.e., “at anchor”. Its duration was slightly less than 6 minutes. This gives us a ground to assume that the events shorter than 5 minutes may not correspond to real staying at anchor.

After filtering these events out, we obtain 212 events with the median duration 134 minutes and the lower and upper quartiles being 18 minutes and 896 minutes (14.9 hours), respectively. These events happened in 126 trajectories (7.33% of the 1,718 trajectories under study).

4.3 Exploration of the anchoring events in relation to strait passing

Since we want to know how the anchoring events are related to passing the strait between the bay and the outer sea, we find the part corresponding to strait passing in each trajectory. Figure 9, left, demonstrates how they have been identified. We outlined interactively the area of the strait. We applied a spatial computation operation that determined for each trajectory point whether it lies inside the outlined area. The parts of the trajectories lying within the strait area are marked in a darker grey shade in Fig. 9, top left. These parts of the trajectories will be further referred to as *strait passing events*. The duration of these events ranges between 7.8 and 88 minutes, the median being 16.7 minutes.

For each event of strait passing, we identified the direction of the vessel movement by means of spatial queries involving the pair of interactively defined areas corresponding to the outer and inner sides of the strait; see Fig. 9, right (the areas are painted in light green). The trajectory fragments that passed first the outer side and then the inner side received the label 'inward' (colored in red in Fig. 9, top right), and the fragments that passed the areas in the opposite order were labelled as 'outward' (colored in blue on the bottom left). After performing this operation, we detected a fragment that was not labelled. We inspected it separately and found that it belongs to a vessel that entered the strait at the inner side but then returned back into the bay. This fragment is colored in yellow in Fig. 9, bottom right. We labelled this fragment 'in2in', which means "from inner area back to inner area".

Next, we determined the temporal relationships of the stops to the strait passing events of the same vessels. We used temporal queries to find for each stop the nearest straight passing event of the same vessel that happened in the past and in the future with respect to the time of the stop. Then we categorized the stops according to the directions of the past and future straight passing; see the legend on the right of Fig. 10. The most common category (105 stops) is 'inward;none', which means that the anchoring took place after passing the strait in the inward direction and there was no other strait passing after the anchoring, i.e., the vessels finally came in the port of Brest. There were 36 'outward;inward' stops, i.e., the vessels exited the bay through the strait, anchored in the outer area, and then moved back into the bay. 34 stops took place before entering the bay ('none;inward'), 18 happened after exiting the bay ('outward;none') and 11 before exiting the bay ('none;outward'). In 7 cases, vessels entered the bay from the outside, anchored, and then returned back without visiting the port ('inward;outward'), and there was one stop that happened after entering the strait at the inner side and returning back ('in2in;none').

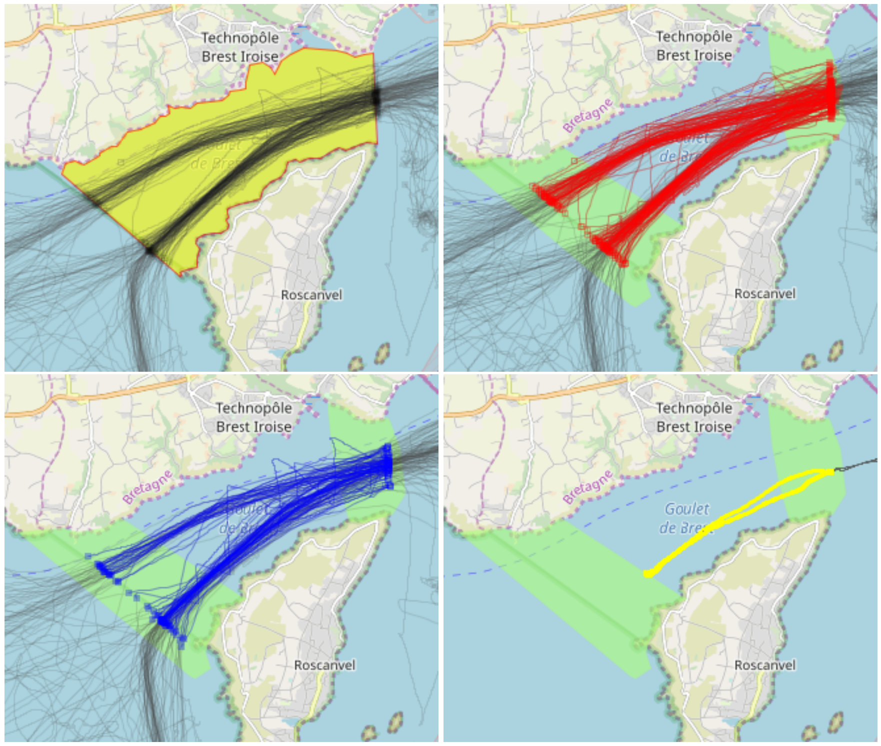

Fig. 9 Top left: Finding the trajectory parts corresponding to strait passing. Remaining images: Identifying the directions of the strait passing. Top right: inward, bottom left: outward, bottom right: a vessel entered the strait from the bay and returned back before reaching the outer side of the strait. This trajectory is labelled “in2in”.

The sizes of the pie charts in Fig. 10 are proportional to the total counts of the stops in the respective zones; the largest chart corresponding to 70 events is located inside the bay. The pie segments represent the proportions of the stops according to the directions of the past and future strait passing. The stops that happened before entering the bay and/or after exiting are located in the outer area, and the events that happened after entering the bay and/or before exiting are located inside the bay.

We see that the majority of the events (yellow pie segments) happened after entering the bay and, moreover, a large part of the stops that took place in the outer area happened after exiting the bay and before re-entering it (orange pie segments). It appears probable that the vessels stopped because they had to wait for being served in the port. Most of them were waiting inside the bay but some had or preferred to wait outside. Hence, the majority of the anchoring events can be related to waiting for port services rather than to a difficult traffic situation in the strait.

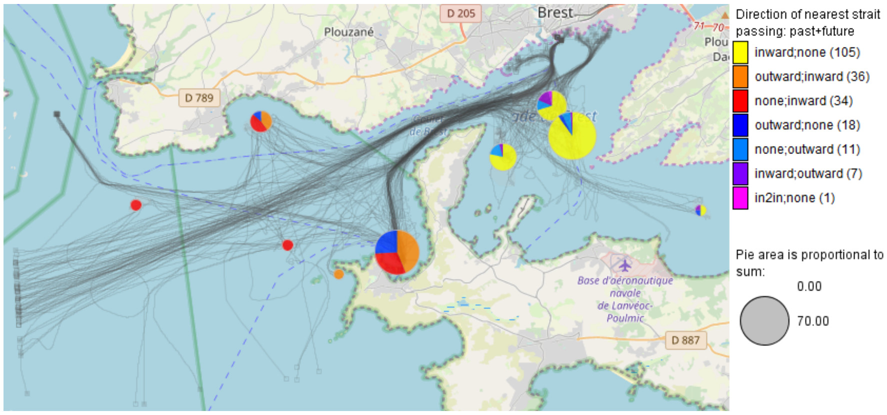

Fig. 10 The pie charts represent the counts of the stops in the anchoring zones categorized according to the temporal relationships to the strait passing by the vessels. The largest pie area represents 70 stops (see the legend in the lower right corner).

4.4 Temporal distribution of the anchoring events

Apart from the spatial distribution, we look at the distribution of the stops over time using two-dimensional histograms (Figs. 11 and 12) the dimensions of which correspond to time components: hours of the day represented by the horizontal dimension versus dates (Fig. 11), and days of the week (Fig. 12) represented by the vertical dimension. The lengths of the dark gray bars are proportional to the numbers of distinct anchoring vessels. The maximal bar length in Fig. 11 corresponds to 4 simultaneously anchoring vessels, which is quite few. The pattern of the distribution by the dates and times of the day tells us that there were days when some vessels were anchoring during the whole days but also many days when there were no anchoring vessels at all or a few vessels anchoring for short times. We do not see any pattern regarding busy and less busy hours. However, the histogram in Fig. 12 shows us that the number of anchoring vessels tends to decrease starting from the morning of Wednesday (the third row from the bottom of the histogram) till the morning of Thursday (the fourth row), and then it starts increasing again. The weekend (two top rows) and Monday (the bottom row) are the busiest days in terms of vessel anchoring. We looked separately at the temporal distribution of the stops that happened after entering the bay and before going to the port or before entering the bay and found that the patterns observed in Figs. 11 and 12 are primarily made by these events, which is not surprising since they are the most numerous. The accumulation of the anchoring vessels by the weekend and gradual decrease of their number during the weekdays supports our hypothesis that the stops may be related to the vessels waiting for being served in the port.

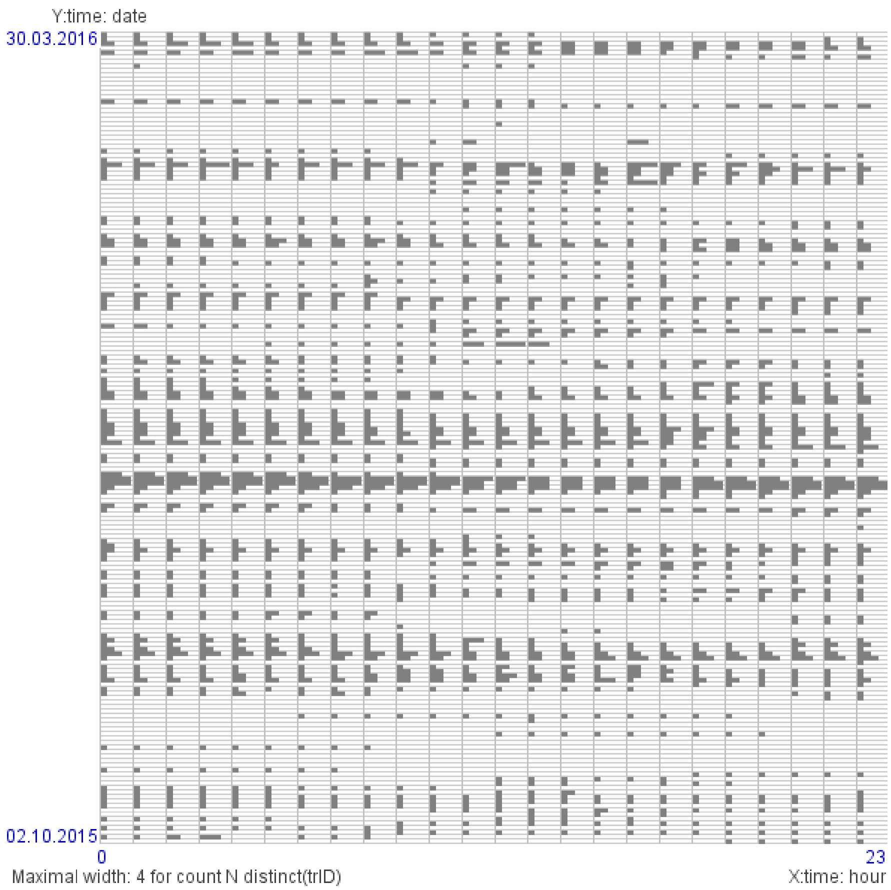

Fig. 11 The temporal distribution of the stops by the dates (rows of the matrix) and hours of the day (columns of the matrix). The bar lengths in the matrix cells are proportional to the counts of the anchored vessels; the maximal length represents 4 vessels.

4.5 Exploration of the anchoring events in the context of the trajectories

Now we want to look at the movements of the vessels that made stops on their way. We aggregate the vessel trajectories by a set of interactively defined areas, which include the anchoring zones, the port area, the areas at the outer and inner ends of the strait as shown in Fig. 9, right, and a few additional regions in the outer sea. The aggregation connects the areas by vectors and computes for each vector the number of moves that happened between its origin and destination areas. The result is shown on a flow map, where the vectors are represented by flow symbols with the widths proportional to the move counts. In our example (Fig. 13), the flow symbols are

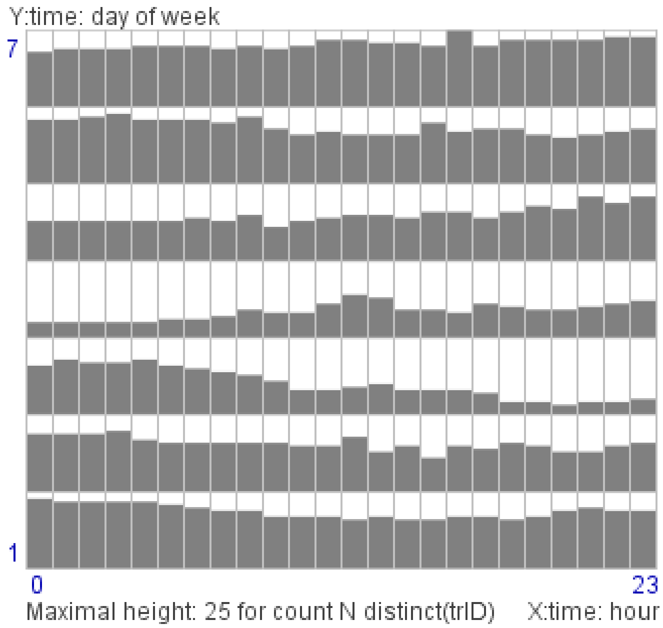

Fig. 12 The temporal distribution of the stops by the days of the week (rows of the matrix; 1 corresponds to Monday and 7 to Sunday) and hours of the day (columns of the matrix). The bar heights in the matrix cells are proportional to the counts of the anchored vessels; the maximal height represents 25 vessels.

curved lines with the curvature being lower at the vector origins and higher at the destinations.

The aggregation we have applied to the trajectories is dynamic in the sense that it reacts to changes of the filters that are applied to the trajectories. As soon as the subset of trajectories selected by one or more filters changes, the counts of the moves between the areas are automatically re-calculated, and the flow map representing them is immediately updated. The six images in Fig. 13 represent different states of the same map display corresponding to different query conditions. The upper left image represents the 1592 trajectories (92.67% of all initially selected trajectories) that did not include stops. This flow map can be considered as showing uninterrupted traffic to and from the port of Brest. The flows between any two areas look symmetric, i.e., the lines have equal widths, which means approximately the same numbers of moves in the two opposite directions.

The remaining images show the flows obtained from the trajectories that included stops. The sizes of the red circles located in the anchoring zones are proportional to the numbers of the stops in these zones. The upper right image represents all 126 trajectories that contained stops. Here, the flows are asymmetric, showing more movements from the outside into the bay than from the bay. Please note that the scales of the widths of the flow symbols differ among the images. The maximal

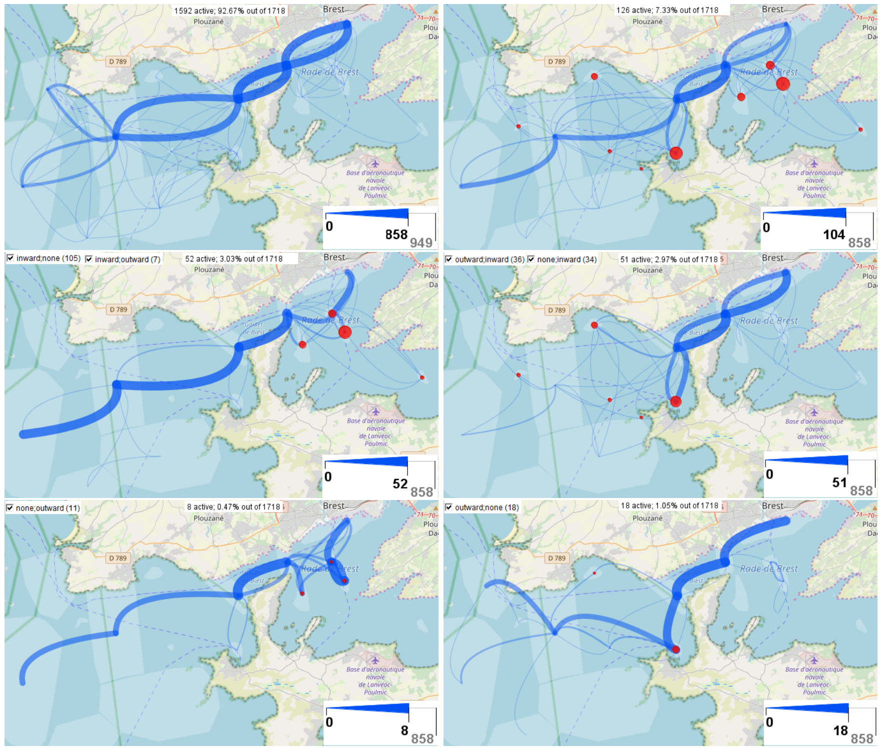

Fig. 13 The trajectories under study are represented in an aggregated form on flow map. Top left: trajectories without stops; top right: all 126 trajectories containing stops. The remaining images represent subsets of the trajectories having stops after entering the bay (middle left), before entering the bay (middle right), before exiting the bay (bottom left), and after exiting the bay (bottom right).

width, which is the same in all images, is proportional to the individual maximal value attained in each image; see the legend in the lower right corner of each image. Thus, the thickest line represents 858 moves on the top left, 104 moves on the top right, 52 and 51 moves in the two images below, and 8 and 18 moves in the lower two images.

The four images in the second and third rows in Fig. 13 represent different subsets of the trajectories containing stops. The images in the second row represent the trajectories that had stops after entering the bay (left, 52 trajectories) and before entering the bay (right, 51 trajectories). Among the latter 51 trajectories, 30 began in the port area, moved through the strait to the outside region, stopped mostly either south of the strait entrance (21 trajectories) or northwest of it (6 trajectories), and then re-entered the bay and moved again into the port. This behavior may mean that the vessels were unloaded in the port and then moved to the outer area for waiting until the loads for their next trips are prepared in the port.

The images in the third row represent the trajectories having stops before (left) and after (right) exiting the bay, 8 and 18 trajectories, respectively. On the left, all 8 trajectories began in the port, 6 of them stopped south of the port and two southwest of the port before going out through the strait. However, three out of these 8 trajectories moved back to the port soon after exiting the bay (they had no stops in the outer area). This behavior may mean that the vessels were waiting for some port services inside the bay and then were going to relocate but changed their intention, perhaps, after being notified that the port is ready to serve them. The image on the right shows that some vessels had stops outside of the bay before going to the outer sea. These vessels, evidently, did not have to wait for port services and stopped for some other reasons.

Concluding the exploration of the stops, we can summarize that most of them are likely to have happened because the vessels had to wait for being served in the port of Brest. The vessels coming from the outer sea were waiting mostly inside the bay (Fig. 13, middle left), and the vessels that had been unloaded in the port and had to wait for the next load or another service were waiting mostly outside of the bay (Fig. 13, middle right). There were at most 4 simultaneously anchoring vessels (Fig. 11). The vessels that had to wait for port services tended to accumulate over the weekend, and their number reduced during the weekend (Fig. 12).

4.6 Exploration of the traffic through the strait

Although we found out that the vessel anchoring events are unlikely to be related to the traffic in the strait, we are nevertheless interested in exploring the intensity and the temporal patterns of the traffic. From all available data, we select the trajectory fragments contained inside the strait area shown on the left of Fig. 9. There are 2,891 trajectory fragments satisfying the spatial query. To guarantee taking into account not only the times of the vessels being inside the strait but also the entering and exiting times, we extend each fragment by adding the preceding and following 45-minutes parts of the trajectories. We obtain 1,366 trajectories, i.e., some trajectories include two opposite movements through the strait. The trajectories are represented by dashed dark gray lines in Fig. 14, top.

We use the two areas shown on the right of Fig. 9 for obtaining aggregate flows through the strait in two directions. The total flow magnitudes are represented by the widths of the curved lines in Fig. 14, top. Please note that flow symbols, such as the curved lines in our map, only connect the origins and destinations of flows but do not represent the movement paths. Therefore, it should not be thought that the red flow symbol in our map represent movements of vessels through the land.

There were 1,010 inward and 995 outward movements (red and blue, respectively). However, we have calculated not only the totals but have applied spatio-temporal aggregation based on these areas and hourly intervals within the weekly time cycle. Hence, for each direction, we obtained a time series consisting of 168 values of the traffic volume in different hours along the weekly cycle. The aggregation over the weekly cycle puts together the movements that happened in the same hours of different weeks.

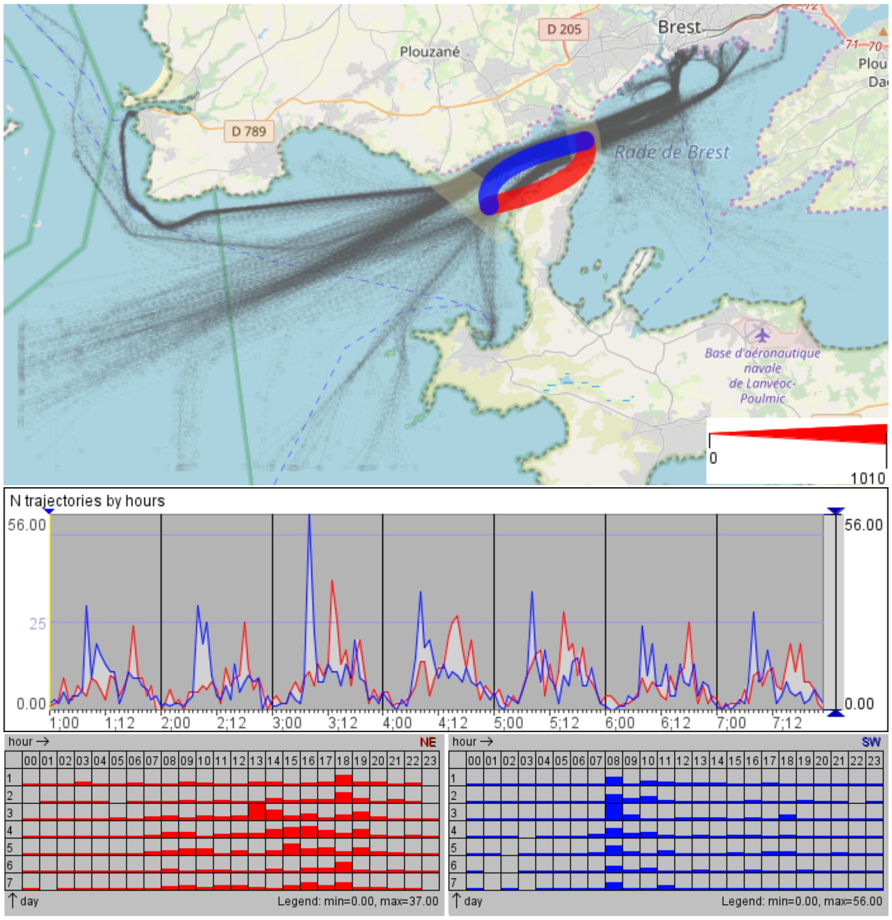

Fig. 14 Top: the trajectories going through the strait and the aggregate flows derived from them. Middle and bottom: the temporal distribution of the flow magnitudes over the hours of the week is shown in a line graph and in two 2D histograms where the rows correspond to the days of the week (from 1 for Monday to 7 for Sunday) and the columns to the hours of the day. In all images red is used for the inward flow and blue - for outward.

The time series of the traffic volumes in the inward (red) and outward (blue) directions are shown in a line graph and two two-dimensional histograms below the map in Fig. 14. We see that the highest intensity of the outward movements was reached in the hour 07:00-8:00 of all days, especially on Wednesday (day 3). The highest peaks of the inward movements happened from 12:00 to 13:00 on Wednesday and early afternoons on Thursday and Friday. The inward traffic was also higher in the hour 18:00-19:00 of the days from Wednesday to Friday and in the hour 17:00-18:00 of the remaining days.

Interestingly, the increase of the inward traffic in the middle of the day and early afternoon from Wednesday to Friday (Fig. 14, bottom left) correlates with the decrease of the number of anchoring vessels in these days, which we observed in Fig. 12. It is probable that both patterns are related to the port operation schedule. Please note that the histogram rows corresponding to the days of the week are arranged from top to bottom in Fig. 14 and from bottom to top in Figs. 11 and 12.

5 Discussion and conclusion

Visual analytic techniques are meant to support data exploration and analytical reasoning by a human. The human plays the key role in analysis. It is the task of the human to understand what data are available and how to make them appropriate to the analysis goal, what methods can be applied and what their results mean, and how the results differ depending on the methods or the parameter settings applied. It is also the task of the human to gain knowledge of the phenomena represented by the data. Visualization is important for supporting the analysis because it can convey information to the analyst in the most efficient way that enables perceiving and interpreting patterns, abstraction, and generalization.

We have demonstrated the use of visual analytics techniques in examples of analysis scenarios that involved discovery of patterns, interpretation of these patterns, and further analytical reasoning. We applied a number of different visualization techniques in combination with interactive querying and filtering of the data, data transformations, and computational derivation of new data, such as events and temporal relationships between them. Besides data transformations, visual analytics workflows often include methods for computational analysis and modelling, which stem from statistics, data mining, geographic information science, and other areas concerned with data analysis. We want to emphasize that such methods need to be supported by interactive visualizations enabling interpretation and comparison of the results, and that they need to be applied as a part of an iterative procedure in which the human analyst examines the effects of the parameter settings and chooses the most suitable ones.

Due to the interactive character of the visual analytics techniques, they typically require the data under analysis to be loaded in the main memory of the computer, which limits the scalability of these techniques to very large data volumes. Recent research in visual analytics develops approaches to overcoming this problem by combining data processing in a database with interactive operations on data samples, aggregates, or other kinds of derivatives. As an example, paper [4] proposes a scalable approach to clustering of trajectories according to the travelled routes.

Another challenge is application of visual analytics techniques to streaming data such as AIS streams. A possible approach is the use of dynamically updated visual displays in combination with computational analysis techniques specially developed for streaming data. An example is interactive real-time detection and tracing of spatial event clusters [7].

Visual analytics processes need to be performed prior to building computer models. Obtaining a good computer model requires that a human analyst understands what patterns it needs to capture and what modelling methods are suitable for this. The analyst should also understand how the data need to be prepared to the modelling. When a computer model is built, visual analytics techniques can greatly help in investigating its performance and finding ways to improve it. A model developed and tested in this way can be trusted and applied for justifiable decision making.

6 Bibliographical Notes

The book by Andrienko et al. [1] gives a detailed presentation of a broad spectrum of visual analytics techniques supporting analysis of movement data. A state of the art survey [5] uses a set of vessel trajectories as a running example to show how different visual analytics techniques can support understanding of various aspects of movement. A more recent survey [2] focuses on the application of visual analytics in transportation studies. The number of visual analytics papers proposing various approaches for analyzing movement data is very large and continues growing. Some of them deal specifically with data describing movements of vessels. Variants of dynamic density maps combined with specialized computations and techniques for interaction [11, 16, 20] support exploration of not only the density but also other characteristics of maritime traffic. Kernel density estimation can be used to compute a volume of the traffic density in space and time [9], which can be represented visually in a space-time cube [10] with two dimensions representing the geographical space and one dimension the time. Tominski et al. [19] apply a 3D view to show similar trajectories as bands stacked on top of a map background. The bands consist of colored segments representing variation of dynamic attributes along the routes. Scheepens et al. [15] propose special glyphs for visualizing maritime data. Lundblad et al. [12] employ visual and interactive techniques for analyzing vessel trajectories together with weather data. Andrienko et al. [6] use vessel movement data to demonstrate the work of an interactive query tool called TimeMask that selects subsets of time intervals in which specified conditions are fulfilled. This technique is especially suited for analyzing movements depending on temporally varying contexts.

Acknowledgements This research was supported by Fraunhofer Cluster of Excellence on "Cognitive Internet Technologies" and by EU in project Track&Know (grant agreement 780754).

References

1. Andrienko, G., Andrienko, N., Bak, P., Keim, D., Wrobel, S.: Visual Analytics of Movement. Springer (2013). DOI 10.1007/978-3-642-37583-5
2. Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R., Zhao, Y.: Visual analytics of mobility and transportation: State of the art and further research directions. IEEE Transactions on Intelligent Transportation Systems **18**(8), 2232–2249 (2017). DOI 10.1109/TITS.2017.2683539

3. Andrienko, G., Andrienko, N., Fuchs, G.: Understanding movement data quality. Journal of Location Based Services **10**(1), 31–46 (2016). DOI 10.1080/17489725.2016.1169322
4. Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., Giannotti, F.: Interactive visual clustering of large collections of trajectories. In: 2009 IEEE Symposium on Visual Analytics Science and Technology, pp. 3–10 (2009). DOI 10.1109/VAST.2009.5332584
5. Andrienko, N., Andrienko, G.: Visual analytics of movement: An overview of methods, tools and procedures. Information Visualization **12**(1), 3–24 (2013). DOI 10.1177/1473871612457601
6. Andrienko, N., Andrienko, G., Camossi, E., Claramunt, C., Garcia, J.M.C., Fuchs, G., Hadzagic, M., Jousselme, A.L., Ray, C., Scarlatti, D., Vouros, G.: Visual exploration of movement and event data with interactive time masks. Visual Informatics **1**(1), 25 – 39 (2017). DOI https://doi.org/10.1016/j.visinf.2017.01.004
7. Andrienko, N., Andrienko, G., Fuchs, G., Rinzivillo, S., Betz, H.: Detection, tracking, and visualization of spatial event clusters for real time monitoring. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 1–10 (2015). DOI 10.1109/DSAA.2015.7344880
8. Bereta, K., Chatzikokolakis, K., Zissis, D.: Maritime Reporting Systems, chap. 1. Springer (2020)
9. Demšar, U., Virrantaus, K.: Space–time density of trajectories: exploring spatio-temporal patterns in movement data. International Journal of Geographical Information Science **24**(10), 1527–1542 (2010). DOI 10.1080/13658816.2010.511223
10. Kraak, M.J.: The space-time cube revisited from a geovisualization perspective. Proceedings of the 21st International Cartographic Conference pp. 1988–1996 (2003)
11. Lampe, O.D., Hauser, H.: Interactive visualization of streaming data with kernel density estimation. In: IEEE Pacific Visualization Symposium, PacificVis 2011, Hong Kong, China, 1-4 March, 2011, pp. 171–178 (2011). DOI 10.1109/PacificVis.2011.5742387
12. Lundblad, P., Eurenius, O., Heldring, T.: Interactive visualization of weather and ship data. In: Proceedings of the 13th International Conference on Information Visualization IV2009, pp. 379–386. IEEE Computer Society (2009)
13. Pitsikalis, M., Artikis, A.: Composite Event Recognition for Maritime Monitoring, chap. 9. Springer (2020)
14. Ray, C., Dreo, R., Camossi, E., Jousselme, A.L.: Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance (2018). DOI 10.5281/zenodo.1167595
15. Scheepens, R., van de Wetering, H., van Wijk, J.J.: Non-overlapping aggregated multivariate glyphs for moving objects. In: IEEE Pacific Visualization Symposium, PacificVis 2014, Yokohama, Japan, March 4-7, 2014, pp. 17–24 (2014). DOI 10.1109/PacificVis.2014.13
16. Scheepens, R., Willems, N., van de Wetering, H., Andrienko, G.L., Andrienko, N.V., van Wijk, J.J.: Composite density maps for multivariate trajectories. IEEE Trans. Vis. Comput. Graph. **17**(12), 2518–2527 (2011). DOI 10.1109/TVCG.2011.181
17. Tampakis, P., Sideridis, S., Nikitopoulos, P., Pelekis, N., Theodoridis, Y.: Maritime Data Analytics, chap. 4. Springer (2020)
18. Thomas, J., Cook, K.: Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE (2005)
19. Tominski, C., Schumann, H., Andrienko, G., Andrienko, N.: Stacking-based visualization of trajectory attribute data. IEEE Transactions on Visualization and Computer Graphics **18**(12), 2565–2574 (2012). DOI 10.1109/TVCG.2012.265
20. Willems, N., van de Wetering, H., van Wijk, J.J.: Visualization of vessel movements. Computer Graphics Forum **28**(3), 959–966 (2009). DOI 10.1111/j.1467-8659.2009.01440.x