Chapter 1
Visual Analytics in the Aviation and Maritime Domains

Gennady Andrienko, Natalia Andrienko, Georg Fuchs, Stefan Rüping,
Jose Manuel Cordero Garcia, David Scarlatti, George A Vouros, Ricardo Herranz
and Rodrigo Marcos

Abstract Visual analytics is a research discipline that is based on acknowledging the power and the necessity of the human vision, understanding, and reasoning in data analysis and problem solving. It develops a methodology of analysis that facilitates human activities by means of interactive visual representations of information. By examples from the domains of aviation and maritime transportation, we demonstrate the essence of the visual analytics methods and their utility for investigating properties of available data and analysing data for understanding real-world phenomena and deriving valuable knowledge. We describe four case studies in which distinct kinds of knowledge have been derived from trajectories of vessels and airplanes

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

Natalia Andrienko
Fraunhofer Institute IAIS, Sankt Augustin, Germany, and City, University of London, UK

Georg Fuchs
Fraunhofer Institute IAIS, Sankt Augustin, Germany

Stefan Rüping
Fraunhofer Institute IAIS, Sankt Augustin, Germany

Jose Manuel Cordero Garcia
CRIDA (Reference Center for Research, Development and Innovation in ATM), Madrid, Spain

David Scarlatti
Boeing Research & Development Europe, Madrid, Spain

George A Vouros
University of Piraeus, Piraeus, Greece

Ricardo Herranz
Nommon Solutions and Technologies, Madrid, Spain

Rodrigo Marcos
Nommon Solutions and Technologies, Madrid, Spain

and related spatial and temporal data by human analytical reasoning empowered by interactive visual interfaces combined with computational operations.

1.1 Introduction

Visual Analytics (VA) has been defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [25, p. 4]. Visual analytics is a research discipline that is based on acknowledging the power and the necessity of the human vision, understanding, and reasoning in data analysis and problem solving. An essential idea of visual analytics is to combine the power of human reasoning with the power of computational processing. It thus aims at developing methods, analytical workflows, and software systems that can support the unique capabilities of humans by providing appropriate visual displays of data and involving as much as possible the capabilities of computers to store, process, analyse, and visualise data.

As facilitators of human understanding and reasoning, VA techniques and tools can greatly support analysts in all stages of a typical analytical process. They can be used for the following purposes:

- gain awareness of properties and problems of available data and understand how the data need to be corrected, transformed, enriched, and/or complemented to become suitable for the intended analysis;
- comprehend the phenomena reflected in the data, grasp essential features, relationships, patterns, trends, and understand how to represent these in models;
- create valid and useful models of the phenomena by involving human critical thinking in model design, preparation, configuration, evaluation, comparison, and iterative improvement.

A substantial body of research in VA has been focusing on data and problems related to mobility and transportation [3]. This chapter includes several examples of applying visual analytics approaches to data and tasks in the domains of air and maritime transportation. The aim is to demonstrate how interactive visual displays in combination with relatively simple computational techniques support the involvement of human understanding and reasoning in the analytical process.

1.2 Related work

A particularly active sub-field of research in visual analytics deals with the analysis of movement data [6, 3], with approaches including trajectory-centred techniques [2, 4, 5], representation and analysis of overall mobility patterns [13, 27], discovery of interactions between moving objects [15], and support of domain-specific decision making processes [5, 9, 20] in complex transportation systems.

A comprehensive survey of the visual analytics research dedicated to mobility and transportation has been published recently [3]. Particularly, there have been research works focusing on the aviation or maritime transportation domains.

VA approaches have been proposed for various specific problems in air traffic analysis. Methods for detection of holding loops, missed approaches, and other aviation-specific events and patterns were implemented in a system integrating a moving object database with a visual analytics environment [22]. Albrecht et al. [1] calculate air traffic density and, considering aircraft separation constraints, assess the conflict probability and potentially underutilized air space. The traffic density and conflict probability are aggregated over different time scales to extract fluctuations and periodic air traffic patterns. Hurter et al. [14] propose a procedure for wind parameter extraction from the statistics of the speeds of planes that pass the same area at similar flight levels in different directions. Buchmüller et al. [10] describe techniques for studying the dynamics of landings at Zurich airport with the goal to detect cases of violating the rules imposed for decreasing the noise in populated areas. The detected violations can be examined in relation to weather conditions and air traffic intensity. Sophisticated domain-specific analyses can be done by applying clustering to interactively selected relevant parts of aircraft trajectories [5]. Andrienko et al. presented an approach to detection of deviations of the routes of actual flights from the planned routes and exploration of the distributions of the deviations over space, time, set of flights, trajectory structures, and spatio-temporal contexts [9].

Related to the maritime domain, a state of the art survey [7] 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. Andrienko et al. [8] 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. Scheepens et al. [23] have designed special glyphs for visualizing maritime data. Tominski et al. [26] 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 vessel routes. Lundblad et al. [18] employ visual and interactive techniques for analyzing vessel trajectories together with weather data. Variants of dynamic density maps combined with specialized computations and techniques for interaction [17, 24, 28] were proposed to support exploration of the density and other characteristics of maritime traffic. Kernel density estimation can be used to compute a volume of the traffic density in space and time [12], which can be represented visually in a space-time cube [16] with two dimensions representing the geographical space and one dimension the time.

Apart from these researches aiming at understanding of the phenomena, events, and processes pertinent to the domains of aviation and maritime traffic, there has been research focusing on exploration of properties of movement data and the use of VA techniques for detection of various quality problems that may occur in such data [4].

In the next section, we shall demonstrate several examples of visual detection of some quality problems in aviation and maritime data.

1.3 Visual exploration of data quality

Possible quality problems in movement data include errors in spatial positions of objects, gaps in spatio-temporal coverage, low temporal and/or spatial resolution, use of the same identifiers for multiple objects, and others [4]. It is essential to reveal such problems before starting to use the data in the planned analysis. The use of inappropriate data or reliance on unchecked assumptions concerning data properties can lead to invalid analysis results and wrong conclusions.

Let us present several examples of quality problems that may occur in data concerning vessel or aircraft movement. Figure 1.1 demonstrates obvious errors in recorded spatial positions: here, many points from trajectories of vessels are located on land. Such wrong records need to be removed from the data, e.g., by filtering.

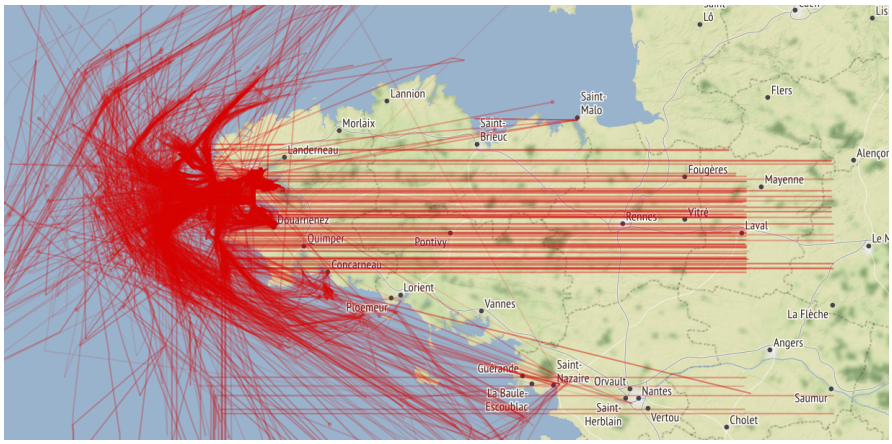

Fig. 1.1: Some trajectories of vessels include positions located on land far from the sea.

Positioning errors in trajectories are not always so obvious. A good indication of a recorded position being out of the actual path of a vessel is an unrealistically high value of the computed speed in the previous position. The computed speed is the ratio between the distance to the next position and the length of the time interval between the positions. If wrong position records detected in this way occur occasionally in the data, they are not difficult to filter out. However, trajectories containing many positions supposedly reached at unrealistic speeds require special investigation. Thus, it may happen that the same identifier is assigned to two or more

simultaneously moving objects. Connecting consecutive points of such trajectories result in zigzagged or more complex shapes, as demonstrated in Fig. 1.2. Such problems may occur due to errors in manually entered data fields, such as flight call signs.

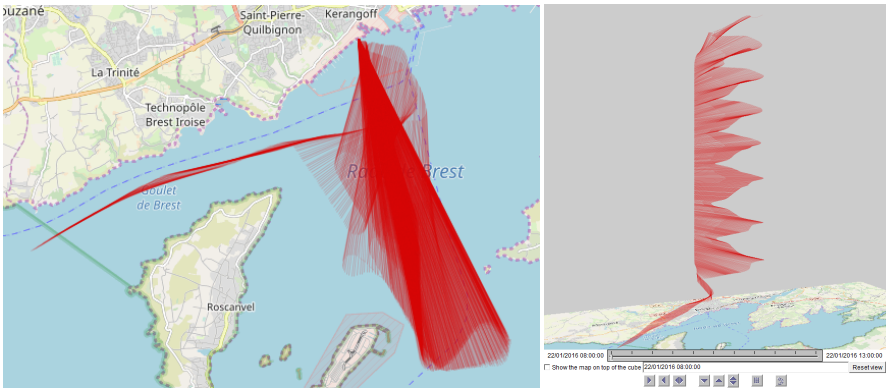

Fig. 1.2: Visual investigation of a trajectory with a high number of erroneous positions. On the left, the trajectory is shown on a map, and on the right in a space-time cube, where the base represents the geographic space and the vertical dimension represents the time.

Zigzagged shapes may also result from incorrect integration of data from multiple sources, such as different radars. Thus, the shape in Fig. 1.3 resembles a mixture of two flight trajectories, but it is unlikely that there were two simultaneous flights following parallel routes and keeping a constant distance between them, as could be deduced from the shape. It is more likely that the same flight is represented twice in the dataset, and at least one set of records contains systematically shifted positions with respect to the real flight trajectory.

When such positioning errors are identified, it is necessary either to devise special, case-specific algorithms for data correction or to discard the problematic trajectories from the analysis.

The example in Fig. 1.4 prominently demonstrates the problem of gaps in the spatial coverage of a dataset consisting of flight trajectories. The trajectories are drawn in a semi-transparent mode. Respectively, darker colours reflect higher density of the flights. Some regions where we expect flights to be frequent, appear as completely empty on the map. In other regions, the density of flights is lower than in neighbouring areas. Obviously, large pieces of data describing the flights are missing, which makes the dataset unsuitable for any meaningful analysis.

Another example in Fig. 1.5 comes from the maritime domain. Here, there are spatio-temporal gaps in some trajectories, that is, absence of position records for long time intervals of vessel movement. These gaps appear as long straight lines when the trajectories are drawn on a map (Fig. 1.5, top). Such segments must be

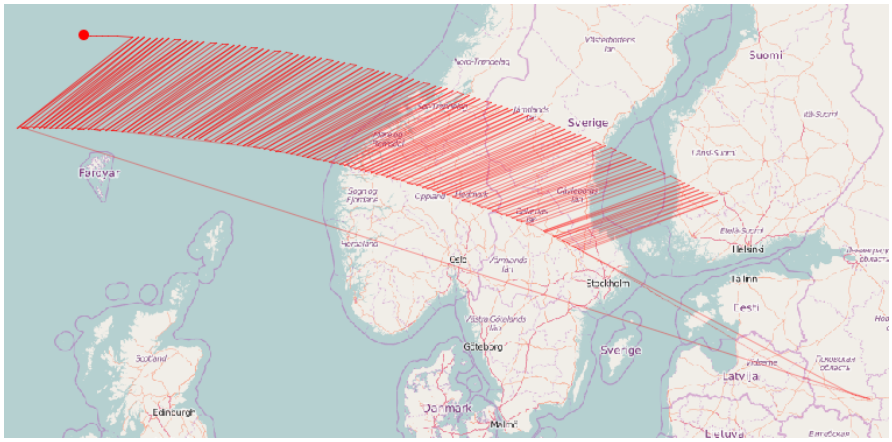

Fig. 1.3: An unrealistic shape of a trajectory indicates either systematically occurring positioning errors (displaced positions) or a mixture of movements of two airplanes.

Fig. 1.4: An example of problems with spatial coverage in a dataset with aircraft trajectories.

excluded from the trajectories 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 with the gaps into several smaller trajectories, so that 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 can be chosen based on domain-specific knowledge, such as usual speeds of vessel movement and normally expected frequencies of position reporting, and taking into account the statistics of the distances in the data.

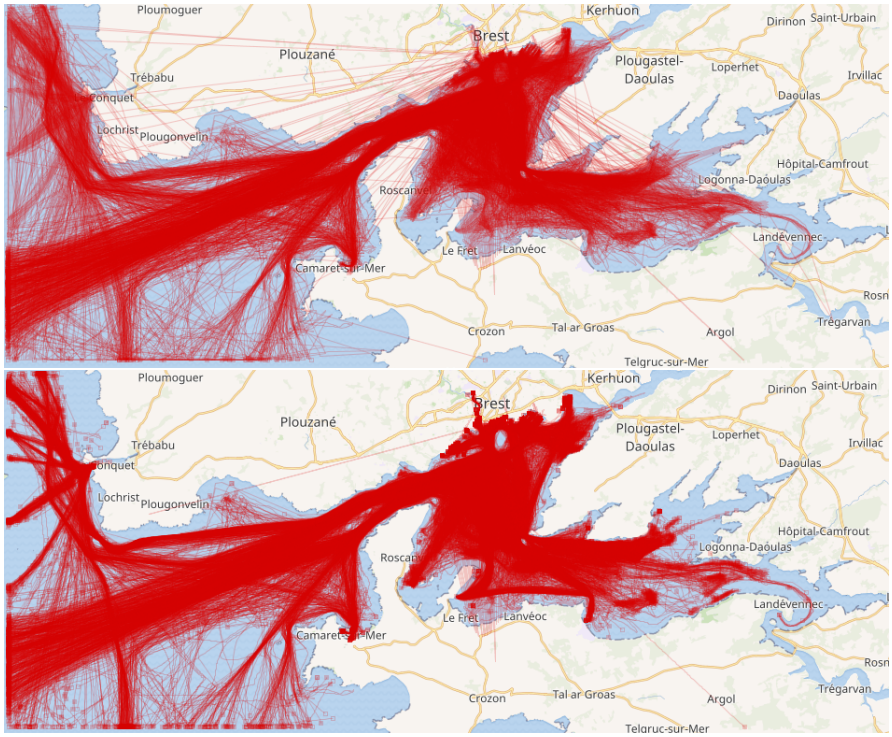

Fig. 1.5: Top: Long straight line segments in vessel trajectories correspond to *spatio-temporal gaps*, i.e., long time intervals in which position records for the vessels are missing. Bottom: The result of dividing the trajectories by the spatio-temporal gaps in which the spatial distance exceeded 2 km and the time interval length exceeded 30 minutes.

Many trajectory analysis methods assume that temporal resolution of position records is constant. Very often data sets do not comply to this requirement. Figure 1.6 presents an example of a data set where the sampling rates of 15, 30 and 60 seconds occur the most frequently, but other values occur as well. Hence, before applying an analysis method that assumes equal-length time steps between positions or attribute

values, it is necessary to re-sample the data to make the time steps equal; otherwise, the method results may be invalid.

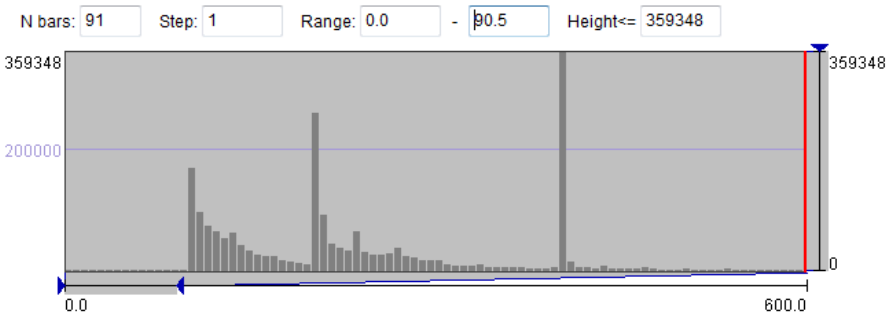

Fig. 1.6: Variability of sampling rates in trajectories.

Figure 1.7 demonstrates that errors may occur not only in positions or identifiers but also in attribute values associated with the positions. In this example, the values of the attribute reporting the navigation status of vessels are unreliable. Hence, if the analysis requires the navigation status to be taken into account, it is necessary to determine the actual status based on movement characteristics rather than attribute values. For example, when analysis requires extraction of stops, they can be identified by finding parts of trajectories where the positions are nearly the same during a chosen minimal stop time.

The examples shown in this section do not cover all possible kinds of errors and problems that may occur in movement data, in particular, in trajectories of aircraft and vessels. Our intention was to demonstrate the utility of visual displays for detecting existing problems, understanding their likely reasons and possible impacts on the analysis, and finding suitable remedies.

1.4 Examples of visual analytics processes

This section briefly presents several case studies intended to demonstrate the use of visual analytics approaches for gaining understanding of different phenomena in the air and maritime traffic. As mentioned in the introduction, visual analytics combines interactive visual displays with computational techniques for data selection, transformation, and automated derivation of various analytical artefacts that can supply relevant information for human reasoning.

A case study performed in the maritime traffic domain mostly aims at demonstrating the use of data transformations for supporting visual analysis and reasoning. The transformations include extraction of relevant parts of trajectories, detection

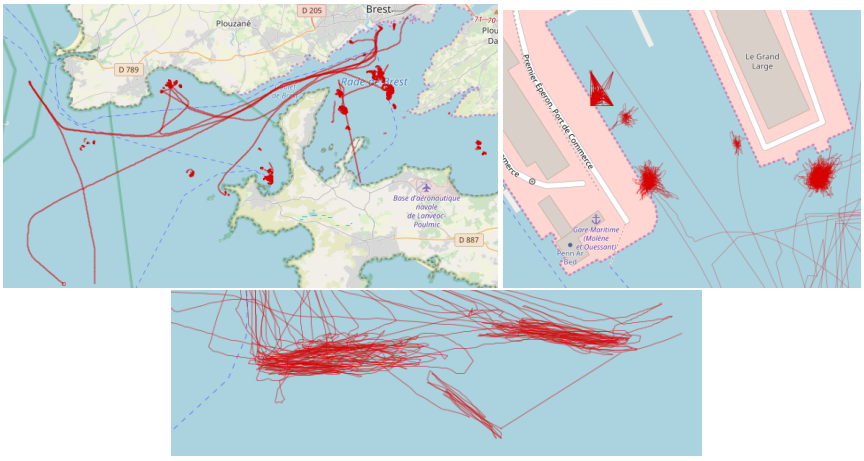

Fig. 1.7: Fragments of vessel trajectories with wrongly reported navigation status. The reported status in the upper left and bottom images is "at anchor", whereas the vessels were actually moving. On the upper right, the reported status is "under way using engine", while the vessels remained at the same places and should have reported "at anchor".

and extraction of particular events, and spatio-temporal aggregation of events and movements.

As an example of using a computational analysis method, two case studies include clustering of flight trajectories based on geometric similarity of the routes. The general approach is to use a density-based clustering algorithm with a special distance function that matches corresponding points and segments of trajectories according to their spatial proximity. The specifics of the case studies we undertook was that not all parts of trajectories might be relevant to the analysis goals. Thus, in studying route choices, the initial and final parts of trajectories were irrelevant because these parts depend on the wind direction and not subject to choice by airlines. In studying the separation scheme of the approach routes to multiple airports of London, we needed to disregard the holding loops as inessential parts of the routes. To be able to apply clustering only to task-relevant parts of trajectories, we adapted the distance function so that it could account for results of interactive filtering of trajectory segments [5]. The main idea is that the distance function receives two trajectories to compare together with two binary masks specifying which points of the trajectories to take into account and which to ignore.

One case study in the aviation domain demonstrates exploratory analysis made with three different kinds of data, planned flight trajectories, geometries of airspace configurations, and temporal succession of the configurations. The aim of the analysis was to understand the relationships between characteristics of the air traffic and the choices of the airspace configurations for controlling the traffic. The analytical

process involved seeking evidence to support or refute hypotheses generated by the analyst based on patterns observed.

All examples presented in this section do not merely demonstrate application of visual analytics techniques to data. They also highlight the importance of human perception, interpretation, understanding, and analytical reasoning and show how visual analytics techniques provide inputs to these cognitive processes.

1.4.1 Detection and analysis of anchoring events in maritime traffic

In this example, visual analytics approaches are used for exploration and analysis of trajectories of vessels that moved between the bay of Brest and the outer sea [21]. The specific analysis task is to study when, where, and for how long the cargo vessels were anchoring and understand whether the events of anchoring may indicate waiting for an opportunity to enter or exit the bay (through a narrow strait) or the port of Brest. The data set consists of about 18M positions of 5,055 vessels during 6 months from the 1st of October, 2015 till March 31, 2016. The exploration of the data properties revealed many problems, some of which have been demonstrated in the previous section.

After cleaning the data, we selected the task-relevant data subset consisting of trajectories of 346 cargo vessels that passed the strait connecting the bay of Brest to the outer sea at least once. From these trajectories, we took 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. To exclude the long straight line segments corresponding to periods of position absence (Fig. 1.5, top), we divided the trajectories into sub-trajectories by the spatio-temporal gaps with distance thresholds 2 km in space and 30 minutes in time (Fig. 1.5, bottom). Next, we further divided the trajectories by stops (segments with low speed) within the Brest port area and then selected from the resulting trajectories only those that passed through the strait and had duration of at least 15 minutes. An outcome of these selections and transformations is a set 1718 trajectories suitable for the 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.

The analysis goal requires us to identify the *anchoring events*. As we cannot rely completely on the navigation status in position records (see Fig. 1.7), we apply the following heuristics. First, we identify areas where many trajectories had records marked as anchoring. We consider these areas as only anchoring zones (Fig. 1.8), thus ignoring occasional records marked as anchoring but located in unusual places. Next, we assume that any sufficiently (at least 5 minutes) long stop in an anchoring zone corresponds to anchoring. So, we get a set of 212 anchoring events (we shall further call them shortly "stops") that happened in 126 trajectories. Fig. 1.9 shows these trajectories in light blue and the positions of the stops in red.

Since we want to understand how the stops are related to passing the strait between the bay and the outer sea, we find the part corresponding to strait passing in each

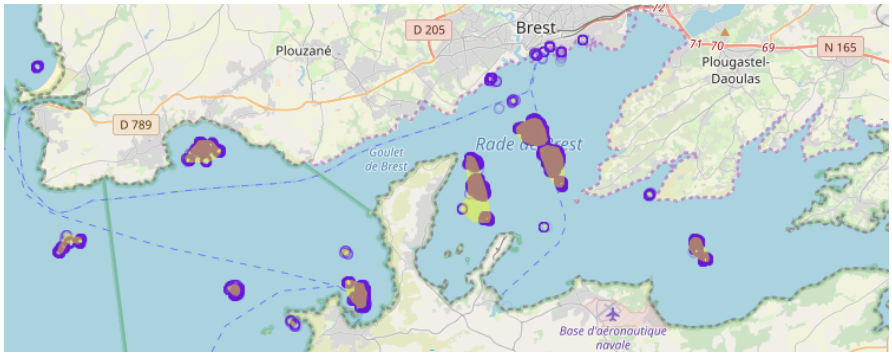

Fig. 1.8: Delineation of anchoring zones: violet points show positions from all trajectories marked as anchoring in the data, tan polygons outline anchoring zones containing dense enough concentrations of the anchoring points outside the port and major traffic lanes.

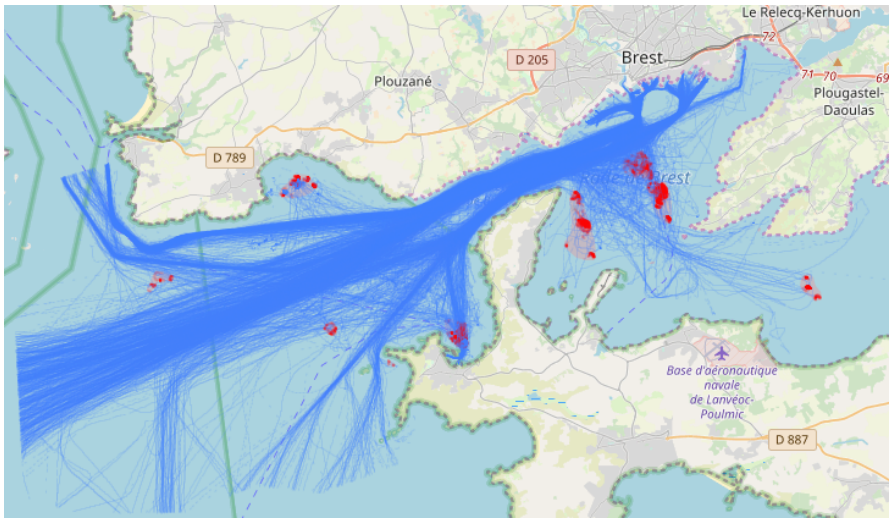

Fig. 1.9: The trajectories selected for analysis with the anchoring events (stops) marked in red.

trajectory. For this purpose, we interactively outline the area of the strait as a whole and, separately, two areas stretching across the strait at the inner and outer ends of it. The segments of the trajectories located inside the whole strait area are treated as *strait passing events*. For these events, we determine the times of vessel appearances in the areas at the inner and outer ends of the strait. Based on the chronological order of the appearances, we determine the direction of the strait passing events: inward or

outward with respect to the bay. Then we categorise the stops based on the directions of the preceding and following events of strait passing.

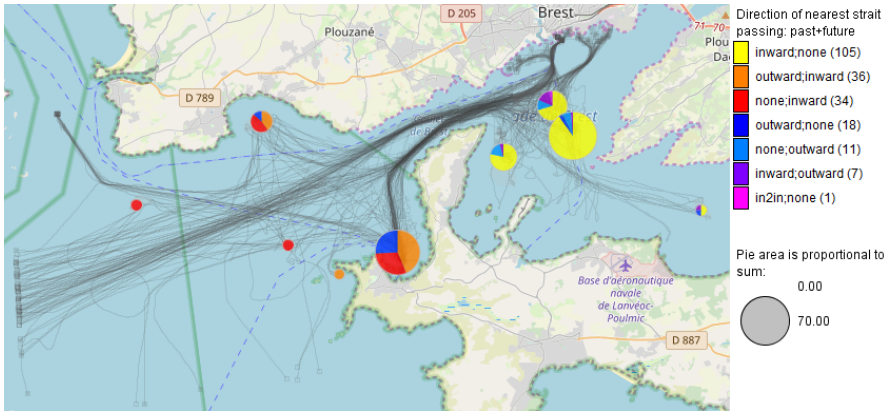

Fig. 1.10: The pie charts represent the counts of the stops in the anchoring zones categorized with regard to the directions of the strait passing by the vessels.

The pie charts on the map in Fig. 1.10 represent the counts of the different categories of the stops that occurred in the anchoring areas. The most numerous category 'inward;none' (105 stops) includes the stops of the vessels that entered the bay, anchored inside the bay, and, afterwards, entered the port. The category 'outward;inward' (36 stops) contains the stops of the vessels that exited the bay, anchored in the outer area, then returned to the bay and came in the port. 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').

We see that the majority of the stop 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.

Additional evidence can be gained from the 2D time histogram in Fig. 1.11. It shows us that the number of anchoring vessels reaches the highest levels on the weekend (two top rows) and on Monday (the bottom row). It 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 accumulation of the anchoring vessels by the weekend and gradual decrease of their

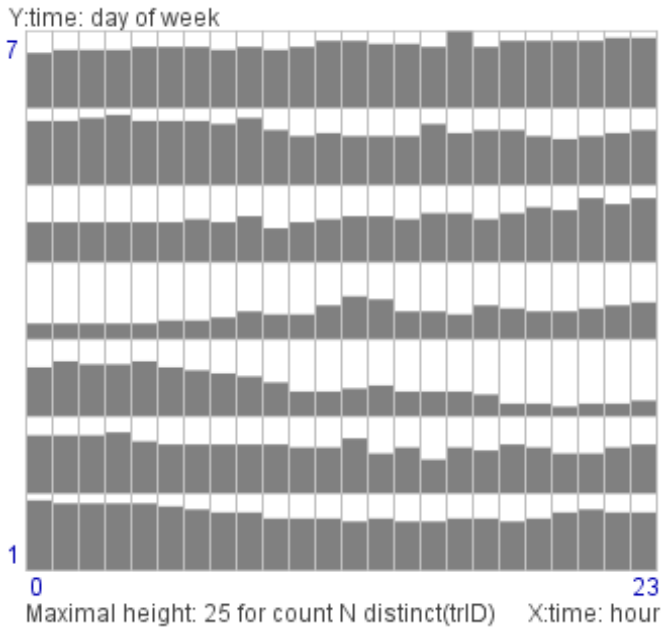

Fig. 1.11: A 2D time histogram represents the counts of the anchoring events by the hours of the day (horizontal axis) and days of the week (vertical axis) by the heights of the corresponding bars.

number during the weekdays supports our hypothesis that the stops may be related to the port operation.

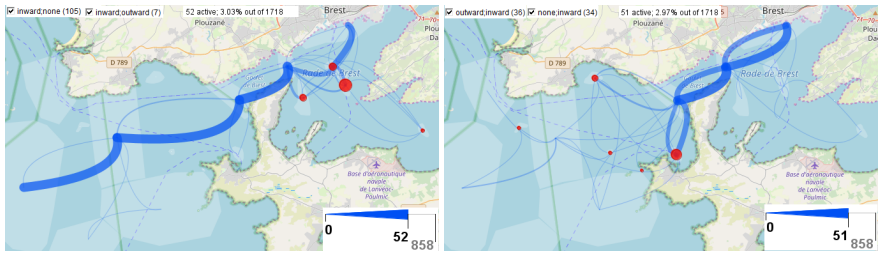

Fig. 1.12: Subsets of the trajectories under study are represented in an aggregated form on flow map. Left: The trajectories having stops after entering the bay of Brest. Right: The trajectories having stops before entering the bay.

To refine our conclusions, we also look at the routes of the vessels that made stops after entering the bay of Brest and before that. In Fig. 1.12, the trajectories of the vessels having stops after (left) and before (right) entering the bay have been

aggregated into flows between interactively defined areas. The flows are represented by curved lines with the widths proportional to the move counts and the curvature increasing in the direction of the flow. The image on the left shows us that most of the vessels that stopped after entering the bay came from the outer sea. After stopping, they eventually moved into the port of Brest. Evidently, the reason for the stops was waiting for the port services. The image on the right shows that a large fraction of the vessels that anchored in the outer area before entering the bay previously came from the port. After stopping, they returned in the bay and moved back into the port. A likely explanation could be that these vessels were unloaded in the port and had to move to the outside area for waiting until the next cargo to be transported is ready for loading.

1.4.2 Exploring separation of airport approach routes

This case study was conducted using 5,045 trajectories of actual flights that arrived at 5 different airports of London during 4 days from December 1 to December 4, 2016. The goals were, first, to reconstruct the major approach routes, second, to determine which of them may be used simultaneously and, third, to study how the routes that can be used simultaneously are separated in the three-dimensional airspace, i.e., horizontally and vertically.

A suitable approach to identifying the major approach routes is clustering of the trajectories by route similarity. A problem we had to deal with was the presence of holding loops in many trajectories (Fig. 1.13). It was necessary to identify the loops in the trajectories and filter them out so that they could not affect the clustering. We have found a combination of query conditions involving derived attributes of trajectory segments, such as sum of turns during 5 minutes, which allowed us to separate the loops from the main paths and filter them out [5]. The clustering was then applied to the remaining parts of the trajectories.

By means of clustering, we have identified 34 distinct routes, 16 of which were used only on the first day out of four. A major change in the use of the routes happened at about 10AM on the second day, when the east-west component of the wind direction changed from the western to the eastern. This refers to all airports except Stansted, where the approach routes changed on the first day at about 18:25 in response to a change of the north-south component of the wind. This was due to the northeast-southwest orientation of the runway in Stansted, which is different from the east-west orientation of the runways in the other airports.

Knowing when each route was used, we could investigate the groups of the routes that were used simultaneously. Figure 1.14 shows the routes that were used on the first day till 18:25 (top) and the routes that were used after 10:00 on the second day, i.e., after the wind change. Using the 3D representation of the trajectories, we observe that the routes coming to the same airport from different sides join in their final parts.

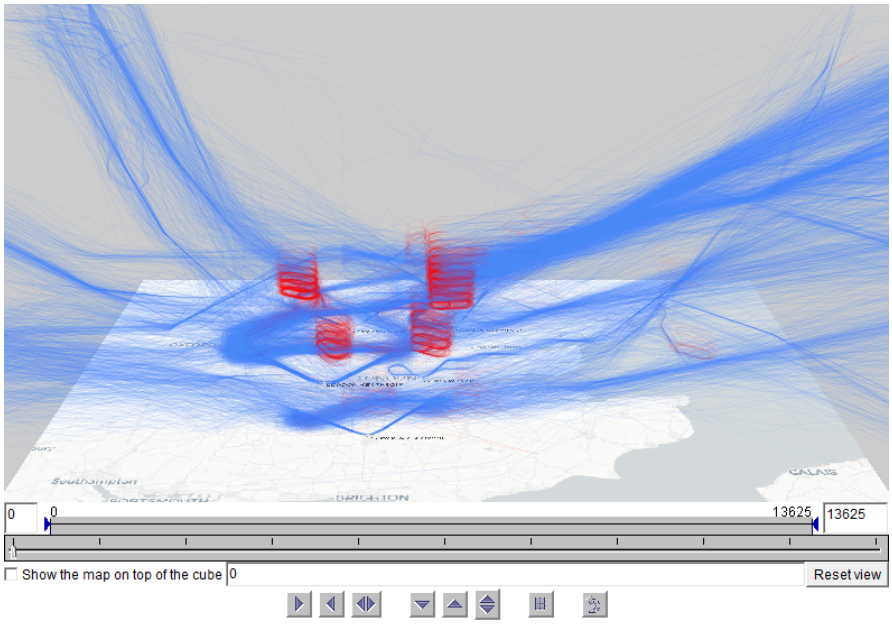

Fig. 1.13: Holding loops in the trajectories of the flights arriving to London are marked in red.

Some routes going to different airports intersect or overlap on the 2D map. To investigate whether they are separated vertically, we repeatedly applied a spatial filter for selecting various groups of intersecting and overlapping trajectories. An example is shown in Fig. 1.15. The filter (Fig. 1.15, top) selects two partly overlapping routes ending at Luton and Stansted (pink and orange, respectively) that apparently intersect two routes ending at Heathrow. In a 3D view (Fig. 1.15, bottom), we see that the former two routes overlap also in the vertical dimension but there is no intersection with the routes to Heathrow due to differences in the flight levels. Our interactive investigation shows that it is a general pattern: where segments of different routes overlap in the horizontal dimension, their altitude ranges overlap as well, and routes intersecting in 2D are separated vertically. Hence, relevance-aware clustering of trajectories and interactive exploration with the use of temporal and spatial filters and a combination of a geographic map and a 3D view helped us to understand how air traffic services organise and manage a huge number of flights following diverse routes within a small densely packed air space.

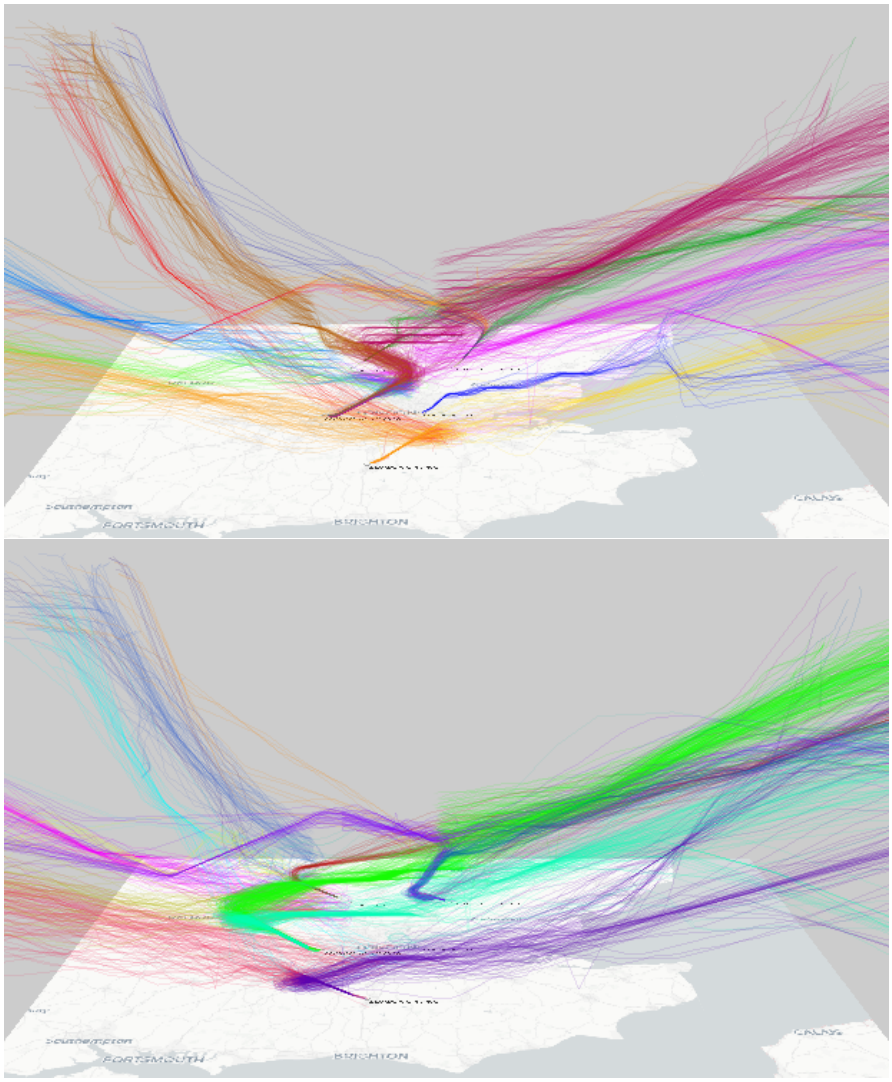

Fig. 1.14: The routes that were used on the first day till 18:25 (top) and on the following days after the wind change (bottom).

1.4.3 Revealing route choice criteria of flight operators

In this study, we wish to reveal the criteria used by airlines in choosing particular flight routes from many possible routes connecting a given origin-destination pair. This translates to a significant improvement in terms of predictability at pre-tactical phase (in particular for routes near local airspace boundaries, for which subtle route

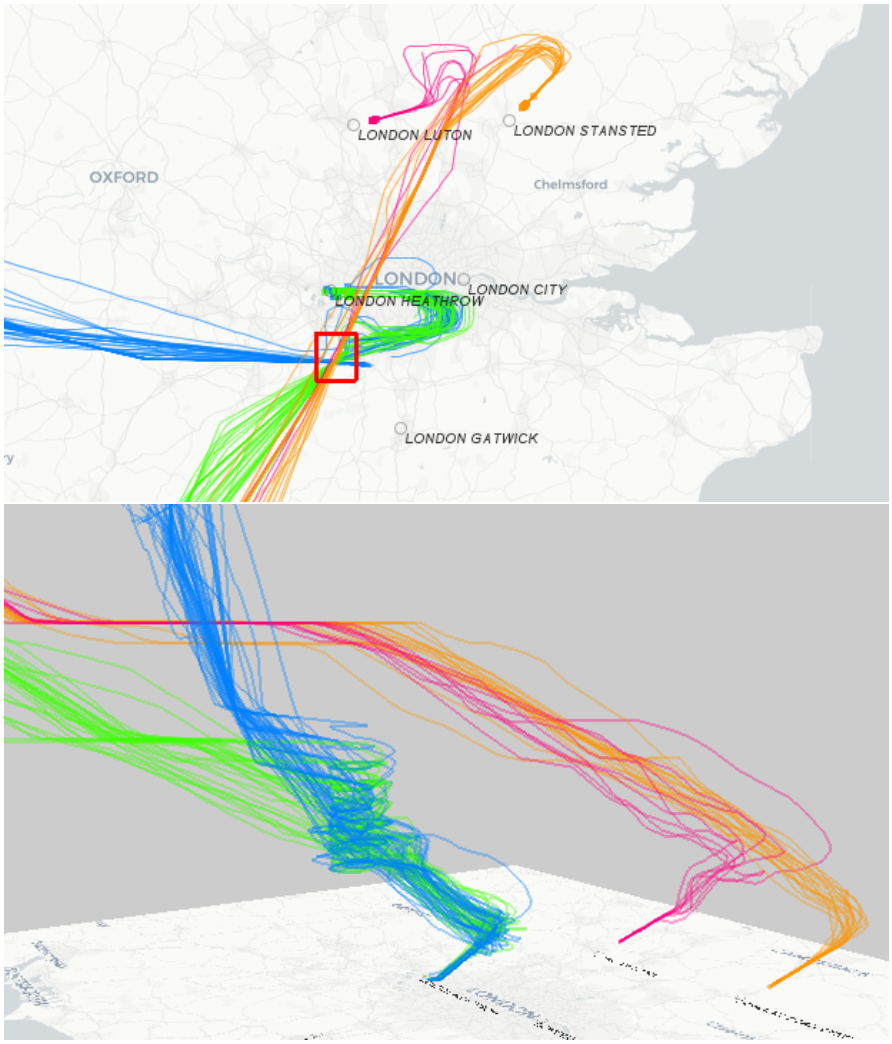

Fig. 1.15: Investigation of the route separation.

changes might imply the appearance or disappearance of hotspots), among other potential applications. As a representative example, we consider the flights from Paris to Istanbul. This example provides rich information for the study: there are many flights conducted by multiple airlines, which take diverse routes crossing the air spaces of different European countries whose navigation charges greatly vary. Some airlines may prefer such flight routes that minimize the navigation costs by avoiding expensive airspaces or travelling shorter distances across such airspaces. One of the questions in the study was to check if indeed some airlines are likely to have such preferences.

We apply our analysis to trajectories constructed from flight plans, because the route choices are made at the stage of planning. We use the plans of 1,717 flights performed during 5 months from January to May, 2016. Additionally, we use a dataset specifying the boundaries of the navigation charging zones in Europe and the unit rate in each. The map background in Fig. 1.16 represents the navigation rates by proportional darkness of shading. The labels show the exact values, in eurocents per mile. On top of this background, coloured lines represent the result of clustering of the trajectories by route similarity excluding the initial and final parts. On the bottom left, the area around Paris is enlarged; the initial parts of the trajectories are shown in dashed lines. The lines are coloured according to their cluster membership. Through clustering, we have revealed 9 major routes. The most frequent was route 1 shown in red; it was used 1,031 times, i.e., in 60% of the flights. Route 2 (green) was used 217 times (12.6% flights), and the others were much less frequent.

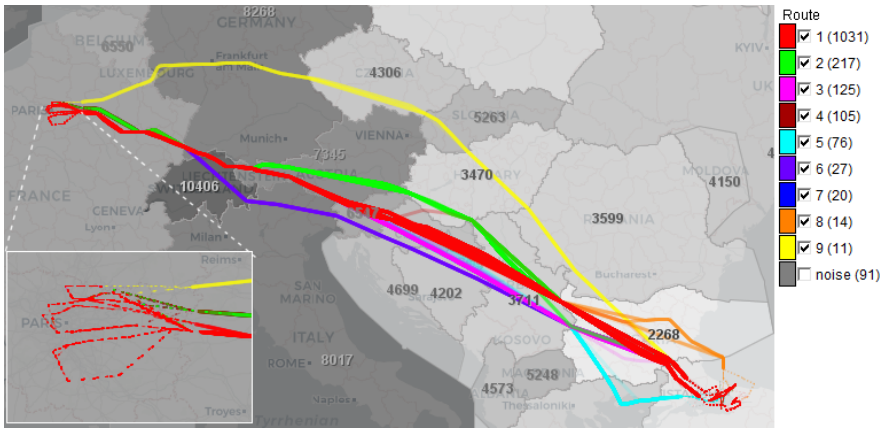

Fig. 1.16: Trajectories according to flight plans have been clustered by route similarity to reveal the major flight routes from Paris to Istanbul. The initial and final parts of the trajectories, which are represented by dashed lines, were disregarded in the clustering.

It can be observed that the green route goes through cheaper airspaces than the other routes. This is the "cheapest" route among all, with the total navigation cost ranging from 434.9 to 492.8 euro, with the median 459.4 euro. The most popular route 1 costs from 472.2 to 547.3 euros, with the median 515.6 euros. Route 2 is the longest among all, except route 9 (yellow) that was taken only 11 times; however, the difference from route 1 is not dramatic, only about 12 km.

The graph in Fig. 1.17 shows how many times each of the 6 major flight operators (airlines) conducting flights from Paris to Istanbul chose each of the routes. The operators are labelled FOP1 to FOP6. It can be seen that FOP4 used only the cheapest route 2. This route was also occasionally used by FOP1, who conducted the largest number of flights (41.9% of all) but not by any other airline. Possibly,

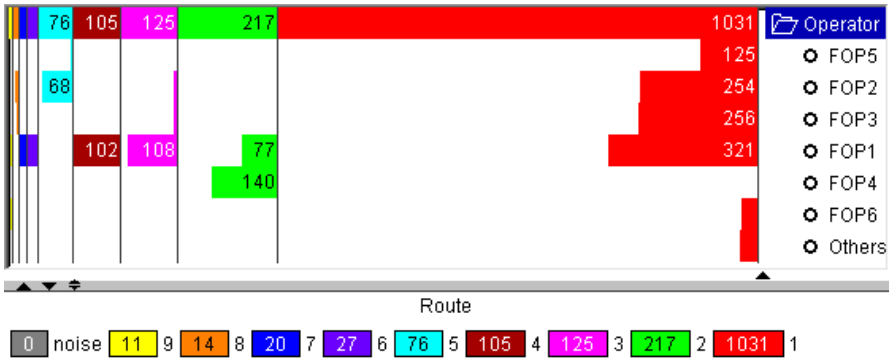

Fig. 1.17: Route choices by 6 major flight operators labelled FOP1 to FOP6. The length of each coloured bar represents the frequency of using the corresponding route by the flight operator specified in the respective row of the graph. The topmost row corresponds to all operators in total.

this route has disadvantages that overweight the navigation cost saving. Apart from the path length difference, which is not very large, it may be lower flight levels or frequent deviations from the flight plans. Indeed, the flight levels on route 2 are lower than on route 1 by about 6 levels on the average and the difference between the third quartiles is 20. We have also calculated the deviations of the actual flights from the planned routes (i.e., the distances between the corresponding points in the planned and actual trajectories) and found that they are higher on route 2 than on route 1 by about 0.8 km on the average while the third quartiles differ by 3.2 km. Route 2 may also have other disadvantages that are not detectable from the available flight data.

Hence, we see that the navigation costs is not the main route choice criterion for most airlines, but it has high importance for some airlines.

Further details on analysis and modelling route choice preferences can be found in paper [19].

1.4.4 Understanding airspace configuration choices

A sector configuration is a particular division of an airspace region into sectors, such that each sector is managed by a specific number of air traffic controllers (typically two, Executive and Planning Controllers). The number of active sectors depends, on the one hand, on the expected traffic features (such as number of flights within a time interval and their associated complexity/workload given the traffic complexity) and, on the other hand, on the available number of controllers for that given shift (which depends on the strategical demand forecast, which diverges from actual flights for a set of reason).

On the other hand, often there are multiple ways to divide a region into a given number of sectors. The choice of a particular division depends on the flight routes within the region. Sector configurations schedule is continuously refined as getting closer to operation, when the available flight plan information is progressively refined. The flight plan information available the day before operation, while is sure to change in tactical phase, already allows to prepare a schedule of sector configurations for the next shifts of the air traffic controllers.

Ideally, configurations should be chosen so that the demand for the use of the airspace in each sector does not exceed the sector capacity, while making efficient and balanced use of resources (controllers). In reality, demand-capacity imbalances happen quite often for a set of reasons (deviations of actual flights from flight plans, weather conditions, etc.), causing flight regulations and delays. In search for predictive models that might support enhanced pre-tactical planning (able to forecast deviations), researchers would like to understand how configuration choices are made by airspace managers. They would also like to find a way to predict which configuration will be used at each time moment during the day of operation, considering uncertainty caused by operational factors in search for a more accurate sector configuration schedule in the day before operation (or earlier), allowing better management of demand-capacity imbalances. However, it is unclear what features should be used for building a predictive model. We utilised visual analytics approaches to gain understanding of the configuration system, patterns of change, and probable reasons for preferring one configuration over another. We performed interactive visual exploration of configurations used in several regions.

As an example, the upper image in Fig. 1.18 shows the configurations that were used in one of the regions in Spain (namely, LECMCTAS) during one month. The configurations are denoted by labels starting with a digit showing the number of sectors in which the region is divided. Almost for each number of sectors, there are two or more variants, some of which are used quite rarely. The lower image shows the use of the different configurations over time. The configurations are represented by coloured segments of horizontal bars. The light colours correspond to small numbers of sectors and dark blue to dark purple colours to 7 and 8 sectors, respectively. The positions of the segments correspond to the times when the configurations were used. The rows correspond to time intervals of one week length. The temporal bar graph shows that the changes of the configurations happen quite periodically. The configurations with small numbers of sectors are used in nights, when the air traffic is low. The configurations with 7 and 8 sectors are usually used from 07:30 till 22:30.

While the choices between configurations differing in the number of sectors can be explained by differences in the traffic volume, the reasons for choosing between configuration variants with the same number of sectors are not obvious. To understand how configurations differ from each other, we used a 3D view as shown in Fig. 1.19. The example in Fig. 1.19 shows two configurations in which the region is divided into 8 sectors, CNF8A1 on the left and CNF8A2 on the right. The sectors are represented by distinct colours. The configurations are almost identical, except the vertical division of the sub-region on the west. In CNF8A1, the sub-area

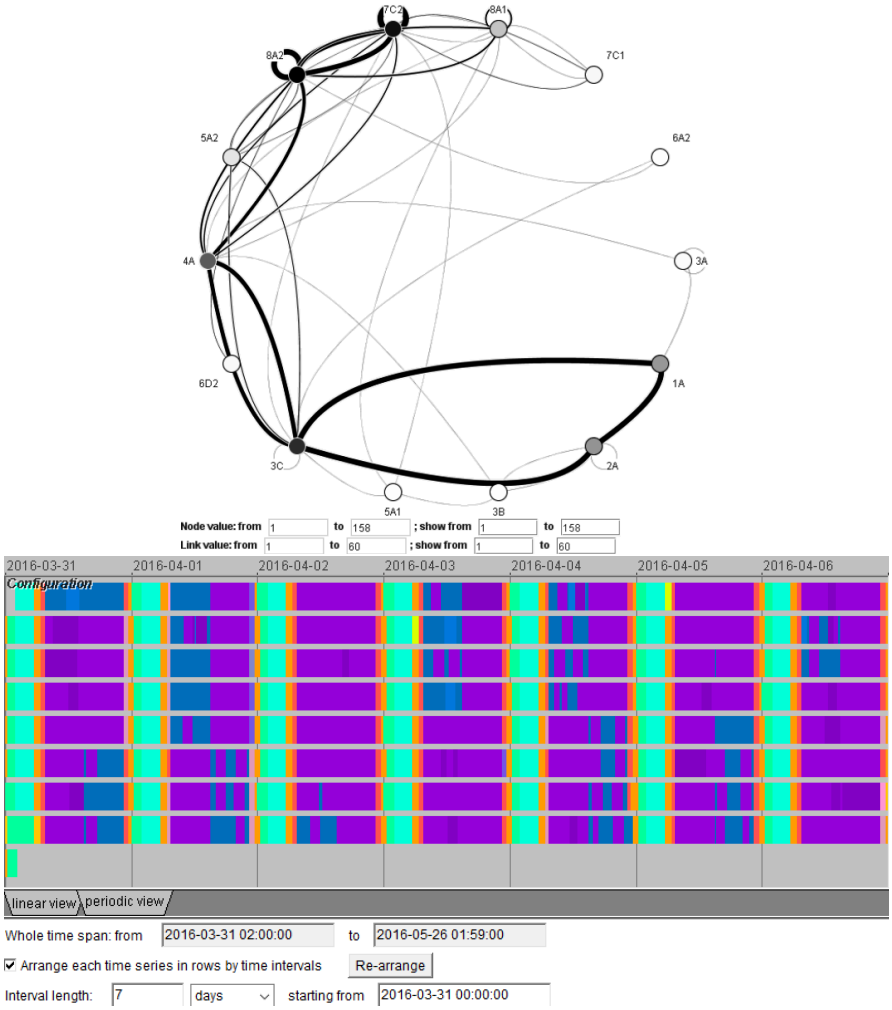

Fig. 1.18: Top: A state transition graph shows changes of airspace configurations in one region during a month. Bottom: The configurations are represented by differently coloured bar segments in a periodic time view. The rows correspond to time intervals of one week length.

is divided into two sectors at the flight level 325, and in CNF8A2 at the flight level 345. These two configurations are often used interchangeably during a day.

The density graph in Fig. 1.20, in which the horizontal dimension represents time and the vertical dimension flight level, shows the traffic intensity in the western sub-region in one day when the configuration CNF8A1 was used in time interval from 12:30 till 14:00 and CNF8A2 in the remaining time from 07:30 till 22:30. These times are marked in the graph by vertical lines. The horizontal lines mark the

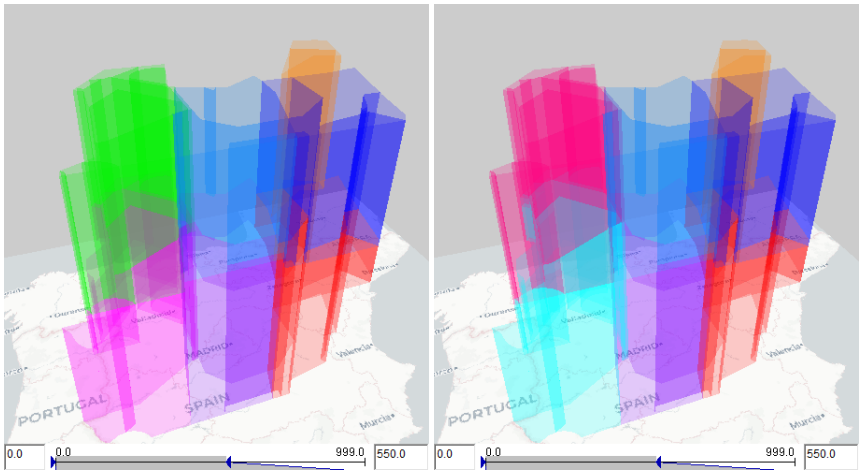

Fig. 1.19: Two configurations with the same number of sectors differ only in the vertical division of the sub-region on the west.

flight levels 325 and 345. The flight intensity is represented by shading from light yellow (low) to dark red (high). The upper image shows the temporal density of all trajectory positions within the western sub-region and the lower image shows the density of the positions where the flight level changed with respect to the previous positions.

A reasonable hypothesis for explaining the choice between different subdivisions would be that the traffic managers strive to balance the workload among the operators controlling different sectors, according to the behaviour of the specific traffic. Indeed, we see that the traffic intensity at the flight levels above 345 decreased after 12:30, and the division level was lowered from 345 to 325. However, after 14:00, when the division level returned to 345, there was no corresponding increase of traffic at the higher levels; so, our hypothesis would not be supported by this exclusive factor. Another possible decision rationale would be to choose such a division level that fewer flights have to cross this level while they are within the area. However, this hypothesis is not supported by the lower image in Fig. 1.20, where we see many intersections of both level 325 and level 345 at the time of using either of the two configurations. Hence, the vertical distribution of the flights does not explain the reasons for preferring one configuration over the other, and further investigation is needed. Domain expert suggest that the sector configuration change was motivated by controller workload, not always precisely represented by traffic counts or intensity. For this model, controller workload was not an input so this factor could only be taken into account indirectly through traffic.

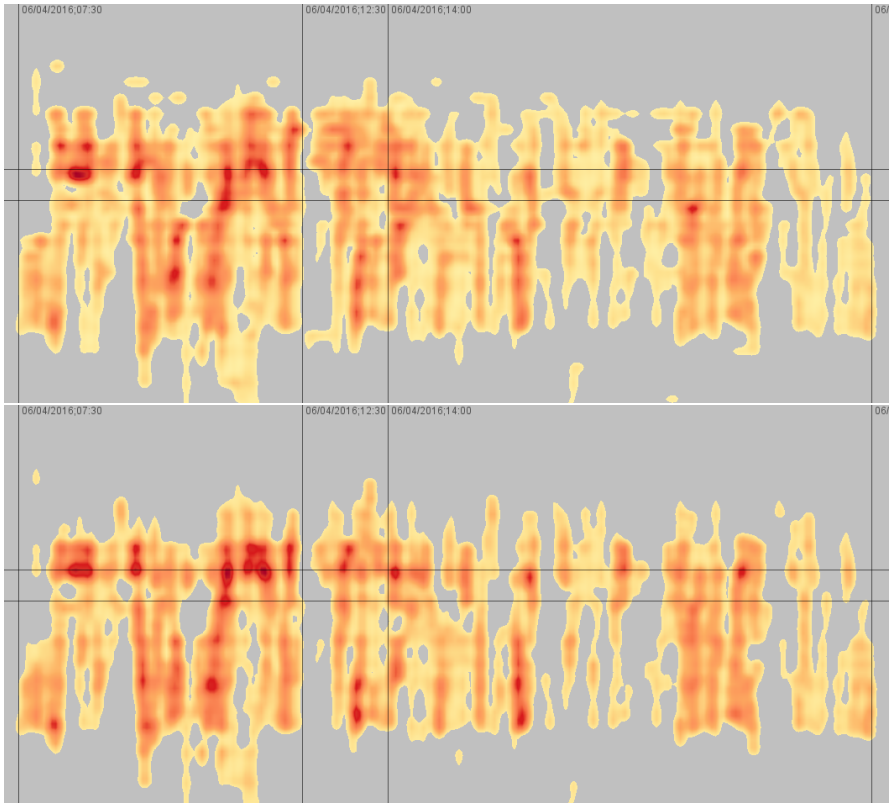

Fig. 1.20: The horizontal and vertical dimensions of the graph represent the time and flight level, respectively. The vertical lines mark the times 07:30, 12:30, 14:00, and 22:30. The horizontal lines mark the flight levels 325 and 345. The shading shows the variation of the traffic intensity in the western sub-region; top: all trajectory segments; bottom: segments where the flight level changed with respect to the previous position.

1.5 Discussion and conclusion

This chapter fulfils several purposes. First, it introduces the concept of visual analytics as a methodology of data analysis where the key role belongs to the human reasoning. The methodology involves the use of interactive visual representations of information for facilitating the cognitive activities of human analysts. Second, the chapter provides multiple examples that can help readers grasp the essence of the VA methodology and see its utility in investigating properties of data (Section 1.3) and in analysing data for understanding real-world phenomena (Section 1.4). Third, it describes analysis scenarios in the domains of maritime and air traffic that resulted in gaining valuable knowledge concerning the design and planning of the business ac-

tivities in these domains. This kind of knowledge can be potentially used for building predictive models and/or for improvement of the businesses.

We would like to emphasise that it is humans, not machines, who can generate new knowledge. Although the term "knowledge discovery" is commonly applied to computational techniques for data analysis, their outcomes are not yet knowledge. They require human apprehension and reasoning for being transformed to knowledge. Therefore, it is absolutely necessary that human reasoning is involved in analyses aimed at gaining new knowledge and finding possible or new, better ways to solve problems. Visual analytics techniques, which support human reasoning, have therefore high importance and high potential.

This potential has been illustrated by four different cases corresponding to diverse operating environments and different data sources. The results have been discussed and validated with domain experts to ensure applicability to operational needs. Particularly, in the domain of air traffic management (ATM), the analysis scenarios demonstrated the value of the VA methods to identify decision criteria as key aspects of the ATM system, able to feed predictive or analytic models applicable in planning phase. The scenarios especially highlighted the power of these techniques to derive knowledge from spatio-temporal patterns. The VA techniques also proved their utility for assessment of data quality. The domain experts admitted that in some cases, as well as in data quality assessment, similar results can be achieved by means of non-visual techniques, but at a significantly higher cost of data preparation and analysis. Visual analytics techniques have proven as time-efficient for these purposes.

In the aviation domain, several Single European Sky ATM Research (SESAR) projects concluded that visual analytics is an important instrument for data analysis and modelling. The white paper [11] supports the use of visual analytics for performance modelling. The improvement in data quality and reliability at planning stages that SESAR new concepts will deliver (i.e., by means of shared business trajectory (SBT), reference business trajectory (RBT), and Trajectory-Based Operations) will only enhance the benefits demonstrated by reducing data uncertainty. However, current day data are already usable by this kind of techniques, delivering applicable results.

Acknowledgements This research was supported by Fraunhofer Cluster of Excellence on "Cognitive Internet Technologies"

References

1. Albrecht, G., Lee, H.T., Pang, A.: Visual Analysis of Air Traffic Data Using Aircraft Density and Conflict Probability. DOI 10.2514/6.2012-2540
2. Andrienko, G., Andrienko, N., Bak, P., Keim, D., Wrobel, S.: Visual Analytics of Movement. Springer (2013). DOI 10.1007/978-3-642-37583-5
3. 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
4. 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
5. Andrienko, G., Andrienko, N., Fuchs, G., Garcia, J.M.C.: Clustering trajectories by relevant parts for air traffic analysis. IEEE Transactions on Visualization and Computer Graphics **24**(1), 34–44 (2018). DOI 10.1109/TVCG.2017.2744322
6. Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M., MacEachren, A., Wrobel, S.: Geovisual analytics for spatial decision support: Setting the research agenda. International Journal of Geographical Information Science **21**(8), 839–857 (2007). DOI 10.1080/13658810701349011
7. 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
8. Andrienko, N., Andrienko, G., Camossi, E., Claramunt, C., Cordero Garcia, J.M., 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
9. Andrienko, N., Andrienko, G., Cordero Garcia, J.M., Scarlatti, D.: Analysis of flight variability: a systematic approach. IEEE Transactions on Visualization and Computer Graphics **25**(1), 54–64 (2019). DOI 10.1109/TVCG.2018.2864811
10. Buchmüller, J., Janetzko, H., Andrienko, G., Andrienko, N., Fuchs, G., Keim, D.A.: Visual analytics for exploring local impact of air traffic. Computer Graphics Forum **34**(3), 181–190 (2015). DOI 10.1111/cgf.12630
11. Cordero Garcia, J., Herranz, R., Marcos, R., Prats, X., Ranieri, A., Sanchez-Escalonilla, P.: Vision of the future performance research in sesar. White paper. SESAR **2020** (2018)
12. 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
13. Huang, X., Zhao, Y., Ma, C., Yang, J., Ye, X., Zhang, C.: Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. IEEE Transactions on Visualization and Computer Graphics **22**(1), 160–169 (2016). DOI 10.1109/TVCG.2015.2467771
14. Hurter, C., Alligier, R., Gianazza, D., Puechmorel, S., Andrienko, G., Andrienko, N.: Wind parameters extraction from aircraft trajectories. Computers, Environment and Urban Systems **47**, 28 – 43 (2014). DOI https://doi.org/10.1016/j.compenvurbsys.2014.01.005. Progress in Movement Analysis - Experiences with Real Data
15. Konzack, M., McKetterick, T., Ophelders, T., Buchin, M., Giuggioli, L., Long, J., Nelson, T., Westenberg, M.A., Buchin, K.: Visual analytics of delays and interaction in movement data. International Journal of Geographical Information Science **31**(2), 320–345 (2017). DOI 10.1080/13658816.2016.1199806
16. Kraak, M.J.: The space-time cube revisited from a geovisualization perspective. Proceedings of the 21st International Cartographic Conference pp. 1988–1996 (2003)
17. 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
18. 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)
19. Marcos, R., Cantu Ros, O., Herranz, R.: Combining visual analytics and machine learning for route choice prediction. application to pre-tactical traffic forecast. Proceedings of the 7th SESAR innovation days. Belgrade (Serbia) (2017)
20. Min Lu, Chufan Lai, Tangzhi Ye, Jie Liang, Xiaoru Yuan: Visual analysis of route choice behaviour based on gps trajectories. In: 2015 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 203–204 (2015). DOI 10.1109/VAST.2015.7347679

21. Ray, C., Dreo, R., Camossi, E., Jousselme, A.L.: Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance (2018). DOI 10.5281/zenodo.1167595
22. Sakr, M., Andrienko, G., Behr, T., Andrienko, N., Güting, R.H., Hurter, C.: Exploring spatiotemporal patterns by integrating visual analytics with a moving objects database system. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '11, pp. 505–508. ACM, New York, NY, USA (2011). DOI 10.1145/2093973.2094060
23. 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
24. 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
25. Thomas, J., Cook, K.: Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE (2005)
26. 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
27. von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G., Kerren, A.: Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. IEEE Transactions on Visualization and Computer Graphics **22**(1), 11–20 (2016). DOI 10.1109/TVCG.2015.2468111
28. 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