# SPATIO-TEMPORAL TRAJECTORY ANALYSIS OF MOBILE OBJECTS FOLLOWING THE SAME ITINERARY

**Laurent ETIENNE[a], Thomas DEVOGELE[a] and Alain BOUJU[b]**

[a]French naval academy research institute (IRENAV), CC 600, 29240 BREST, (laurent.etienne,thomas.devogele)@ecole-navale.fr
[b]Data-processing and Industrial Imagery Laboratory (L3I), Avenue Crépeau, 17042 La Rochelle Cedex 1, abouju@univ-lr.fr

**KEY WORDS:** spatio-temporal patterns, data mining, mobile object, outlier detection, maritime GIS.

**RÉSUMÉ:**

More and more mobile objects are now equipped with sensors allowing real time monitoring of their movements. Nowadays, the data produced by these sensors can be stored in spatio-temporal databases. The main goal of this article is to perform a data mining on a huge quantity of mobile object's positions moving in an open space in order to deduce its behaviour. New tools must be defined to ease the detection of outliers. First of all, a zone graph is set up in order to define itineraries. Then, trajectories of mobile objects following the same itinerary are extracted from the spatio-temporal database and clustered. A statistical analysis on this set of trajectories lead to spatio-temporal patterns such as the main route and spatio-temporal channel followed by most of trajectories of the set. Using these patterns, unusual situations can be detected. Furthermore, a mobile object's behaviour can be defined by comparing its positions with these spatio-temporal patterns. In this article, this technique is applied to ships' movements in an open maritime area. Unusual behaviours such as being ahead of schedule or delayed or veering to the left or to the right of the main route are detected. A case study illustrates these processes based on ships' positions recorded during two years around the Brest area. This method can be extended to almost all kinds of mobile objects (pedestrians, aircrafts, hurricanes, ...) moving in an open area.

## 1 INTRODUCTION

More and more mobile devices are equipped with tracking systems broadcasting accurate information about their movements. Those sensors generate a large amount of data which can be stored in spatio-temporal databases ($STDB$) in order to be analysed. Mobile objects monitoring is commonly used in various fields such as the study of meteorological phenomena, animal migration (Lee et al., 2008), crowd or pedestrian displacement (Knorr et al., 2000), vehicle trips (cars, planes, ships) (Baud et al., 2007, Giannotti et al., 2007). This mobile object monitoring can be linked with intelligent system analysis to improve the system's performance (to ease freight transport planning, for example). Using spatio-temporal databases led to new capabilities. Indeed, the displacement of these mobile objects can be analyzed over a long period of time in order to deduce the general behaviour of mobile objects following the same route. Detecting outliers that behave in an unusual way in such large amounts of data is a very active research field linked to data mining and statistical analysis.

Assuming that mobile objects following a same itinerary behaves in a similar optimized way, these behaviours can be deduced by data mining on $STDB$. It pave the way to analysis of mobile objects' trajectories and detection of unusual behaviour. Different ways to detect outliers in a large dataset could be applied to our issue. These outlier detections are classified according to the method used which can be based on distribution (Barnett and Lewis, 1994), distance (Knorr et al., 2000, Ramaswamy et al., 2000, Lee et al., 2008) or density (Aggarwal and Yu, 2001, Papadimitriou et al., 2003, D'Auria et al., 2006, Kharrat et al., 2008, Lee et al., 2008). The distance and density methods are merged in Lee's works (Lee et al., 2008) based on a "partition and detect framework" that identifies subsets of trajectories which have fewer neighbours. These parts of trajectories are considered as locally unusual regarding density and distance criteria. Unfortunately, time criteria is not taken into account in these methods. In this paper, we propose a process to qualify the position of a mobile object both on spatial and temporal criteria.

The main goal of this study is to define spatio-temporal analysis tools to describe mobile objects' behaviour. Assuming that similar mobile objects following the same itinerary behave in a similar way and move along an optimized main route, it could be useful to analyse the trajectories of these objects in order to deduce spatio-temporal patterns and then, to qualify their behaviour by comparing their trajectories to these patterns. Such trajectory analysis tools coupled with a visualization process could be useful for traffic monitoring operators to focus on outliers (mobile objects behaving in an unusual way) for safety purpose. In some areas, mobile objects' traffic is very dense and the amount of data to be processed in real time can be important. In order to create these tools, notion of trajectories of mobile objects following a same itinerary have to be defined. Then homogeneous subset of trajectories of mobile objects following a same itinerary have to be extracted from the $STDB$. The main goal of this study is to analyze this subset of trajectories in order to infer the behaviour of mobile objects following a similar path.

The paper is organized into 6 main sections. The first section of this article introduces the main goal of this paper and related research in data mining mobile objects movement. The second section proposes a method to extract and filter trajectories of mobile objects following a similar itinerary. The third section deals with statistical computation of spatial and spatio-temporal route and channel. This section also describes how to qualify the position of a mobile object following an itinerary using these tools. The fourth section present the results of our process applied to a case study focused on passenger ships in the Brest area followed by some discussions. Finaly, the last section concludes pointing to further future work.

## 2 SPATIO-TEMPORAL TRAJECTORIES EXTRACTION AND FILTERING

The general process proposed to identify unusual mobile object behaviour is presented in figure 1. First of all, information about mobile objects' positions are stored into a spatio-temporal database (figure 1 step 2). The zone graph of the area of interest is set up in the knowledge database (figure 1 step 5). A cluster of trajectories of similar mobile objects following the same itinerary is extracted from the $STDB$ (figure 1 step 3). Then, a statistical analysis is performed to compute spatio-temporal patterns (figure 1 step 4) which are then stored in a knowledge database (figure 1 step 5). Each new position of a mobile object can be compared with these spatio-temporal patterns in order to qualify the mobile object behaviour (figure 1 step 6). The next sections of this article describe this general process.

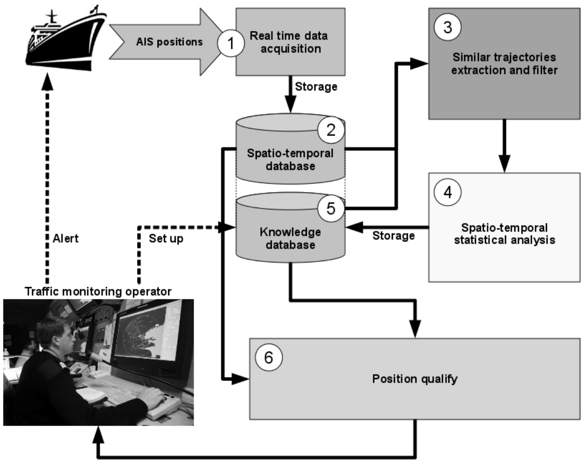


FIG. 1: General process of spatio-temporal trajectory analysis

To analyze a large amount of moving objects' trajectories, both spatial and temporal information about their positions must be stored. $STDB$ are employed to store sets of discrete data having spatial and temporal properties (Güting, 1994). These $STDB$ offer tools to perform queries on these sets of data on spatial and temporal criteria. Formally, the position of a moving object ($O$) is composed of spatial coordinates with a timestamp corresponding to the date on which the moving object was at that position (absolute time). So, the trajectory $T_o$ of a mobile object $O$ can be defined as a sequence of temporally ordered positions $P_{oj}$ so that $T_o = (P_{od}, ..., P_{oj}, ..., P_{oa})$ where $P_{od}$ stands for the departure position of the trajectory and $P_{oa}$ for the arrival one.

### 2.1 Definition of a zone graph

In order to deduce main routes, the trajectories of same-type mobile objects following the same itinerary are extracted from the $STDB$ and then grouped together. The concept of itinerary can be defined as an ordered sequence of spatial zones. In our study, the space is a wide open area which allows mobile objects to navigate from one place to another using the most effective path.

The concept of zone graph used in this section will now be formalized. Zones of this graph represent important areas. These important areas can be manually defined by an operator according to various criteria such as regulations (waiting areas, traffic channels, restricted areas), geography (obstacles, isthmuses, straits, inlets), economy (shops, loading sites, ports, fishing areas). This directed zone graph can be used to describe an itinerary. Using the previously defined vertices of this zone graph ($G_Z$), an itinerary ($I$) is defined as a sequence of ordered zones linked by arcs (a path of the zone graph). An itinerary is made up of at least one arc, therefore it has a departure zone ($Z_D$) and an arrival zone ($Z_A$). A trajectory $T_o$ follows an itinerary $I$ through the vertices of the zone graph $G_Z$ if it satisfies the following conditions :
Let an itinerary be defined as $I = \{Z_D, ..., Z_i, ..., Z_A\}$
Let a trajectory be defined as $T_o = (P_{od}, ..., P_{oj}, ..., P_{oa})$
Trajectory $T_o$ follows the itinerary $I$ if :

$$\forall Z_i \in I, \exists P_{oj} \in T_o, P_{oj} \subset Z_i \tag{1}$$

$$\begin{aligned} \forall P_{oj} \in T_o \wedge P_{oj} \subset Z_l, \forall P_{ok} \in T_o \wedge P_{ok} \subset Z_m, \\ Z_l <_I Z_m \rightarrow P_{oj} < P_{ok} \end{aligned} \tag{2}$$

$$\forall P_oj \in T_o \wedge P_{oj} \subset Z_i \rightarrow Z_i \in I \tag{3}$$

$$P_{oj} \subset Z_D \rightarrow P_{oj} = P_{od} \tag{4}$$

$$P_{oj} \subset Z_A \rightarrow P_{oj} = P_{oa} \tag{5}$$

In other words, for each zone of the itinerary $I$, there is at least one position $P_o$ of the trajectory $T_o$ inside this zone (Eq. 1) which respects the time order relation previously defined (Eq. 2). Taking into account the frequency of trajectory samples and the speed of the mobile object, trajectories that cross a zone of the graph should have at least one position within this zone. No other position $P_o$ of the trajectory $T_o$ is within a zone that does not belong to the itinerary (Eq. 3). Only the first position $P_{od}$ of the trajectory belongs to the departure area of the itinerary $Z_D$ (Eq. 4). In the same way, only the last position $P_{oa}$ of the route belongs to the last area of the route $Z_A$ (Eq. 5).

### 2.2 Extraction of an homogeneous group of trajectories

Now that the concepts of trajectory and itinerary have been formalized, the criteria used to extract trajectories following the same arc $A$ of an itinerary $I$ can be detailed. The goal of this part is to extract the $STDB$ trajectories of same type $T$ objects moving along the same arc $A$ of an itinerary $I$. This set is called homogeneous group of trajectories of same type mobile objects following the same arc of an itinerary ($HGT_{AIT}$). Thus, the first selection criterion is the type of the mobile object. The second selection criterion is a geographical one. The first position of the trajectory must be the only one within the departure zone ($Z_D$) of the arc (Eq. 4), and the last position of the trajectory must be the only one within the arrival zone ($Z_A$) of the arc (Eq. 5). Finally, the last selection criterion used is time. Some moving objects can follow this itinerary periodically, these different trajectories can be distinguished using a time interval. These selection criteria applied to the $STDB$ are used to extract all the spatio-temporal positions of a same mobile object in the meantime between positions $P_{od}$ and $P_{oa}$ forming the trajectory of the mobile object ($Tr_o$) ordered by timestamp. Finally, the trajectory should not intersect zones of the graph $G_Z$ that do not belong to the itinerary $I$ (Eq. 3). All valid trajectories previously extracted from the $STDB$ compose the $HGT_{AIT}$ to be analyzed.

### 2.3 Erroneous trajectory filtering

Once the $HGT_{AIT}$ has been extracted from the database, trajectories with an important gap between two consecutive positions or erroneous positions are filtered from the $HGT_{AIT}$ in order to improve statistical analysis. First of all, trajectories containing important communication loss compared to normal transmission rate of the studied group of trajectories are discarded. Then, some tracks may contain erroneous positions due to a malfunction of the geolocation system or transmission errors. These erroneous positions can be detected using the calculated speed of the position compared to the maximum speed of a moving object of this type. Trajectories having either erroneous positions or transmission gaps are removed from the $HGT_{AIT}$.

### 2.4 Spatial shifting

In order to compute trajectories for which departure and arrival positions are independent from time of transmission, starting and ending positions of the trajectory within the departure and arrival zones must also be filtered. Without this filtering, a bias can be measured in the spatio-temporal patterns defined in the next section of this article. The cloud of initial starting positions of the $HGT_{AIT}$ is represented in figure 2.a. The new starting positions are computed by interpolation between a virtual starting line (border of $Z_D$) and each trajectory of the $HGT_{AIT}$. The same process is applied to the arrival zone $Z_A$. The result of space shifting applied to our example is illustrated in figure 2.b.

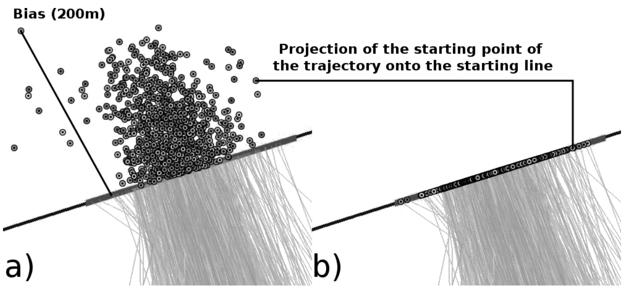


FIG. 2: Spatial shifting of trajectories

### 2.5 Spatio-temporal Douglas & Peucker filter

Once the spatial shifting is done, in order to optimize the computation time, trajectories can be both indexed according to a spatio-temporal method (Rasetic et al., 2005) and simplified using a filter initially proposed by Douglas & Peucker (Douglas and Peucker, 1973). Several different algorithms are based on this work. Some of them have been compared by Wu (Wu and Pelot, 2007). In this study, a spatio-temporal Douglas & Peucker filter (Bertrand et al., 2007, Cao et al., 2006, Meratnia and de By, 2004) is used. The goal of this filter is to retain only significant positions of a trajectory while keeping information about speed or heading changes. To do this, the greatest distance $d_{max}$ between each positions $P_i$ of the trajectory and their spatio-temporal projections $P'_i$ on the line between the starting positions $P_d$ and arrival $P_a$ is calculated. If this distance $d$ between $P_i$ and $P'_i$ exceeds a threshold, the farthest position $P_{max}$ is retained. The trajectory is then split at that position ($P_{max}$) and the algorithm is recursively applied to both trajectory subparts. If the distance $d$ is smaller than the threshold, only positions $P_d$ and $P_a$ are kept. This algorithm also filters inaccuracies of measuring devices (Bertrand et al., 2007).

### 2.6 Position normalized relative timestamps computation

In order to ease distance and time comparison between trajectories, a relative timestamp is computed for each position of a trajectory. Timestamps of positions are very useful to compute speed and order each position within a trajectory. Initial positions of trajectories are all set up with an absolute timestamp ($t_A$). In order to compare these trajectories, a new relative timestamp ($t_R$) is computed for each position. This relative timestamp stands for the interval of time since the starting position of the trajectory. Thereby, every starting position of the trajectories of the $HGT_{AIT}$ have a null relative timestamp ($t_0 = 0$). Finally, to avoid spatial distortions introduced by slightly different speeds of mobile objects of the $HGT_{AIT}$, timestamps of all the trajectories of the $HGT_{AIT}$ must be normalized. This relative normalized timestamp $t_{NR}$ stands for the normalized time elapsed since the starting position of a trajectory. To compute this relative normalized timestamp, first of all, the median duration $D_{med}$ of the $HGT_{AIT}$ is calculated. The choice of the median duration is less disturbed by outliers. Using this duration, a normalization process is applied to all trajectory positions so that each trajectory begins at a time $t_0 = 0$ and ends at the same relative normalized time $t_m = t_0 + D_{med}$.

## 3 STATISTICAL ANALYSIS OF TRAJECTORIES

Once the $HGT_{AIT}$ has been extracted and filtered, it is worthwhile to perform a statistical analysis of this group of trajectories. This statistical analysis aims at qualifying positions and trajectories of moving objects following an itinerary using spatial and temporal criteria. To do this, spatio-temporal patterns are defined to compare positions and trajectories of a moving object with patterns which stand for normal behaviour of mobile objects of the same type following the same itinerary.

### 3.1 Main route computation

First of all, a main route followed by most of the trajectories of the $HGT_{AIT}$ is computed by statistical analysis. The first stage of this process consists in setting up a new relative normalized timestamp for each position of each trajectory of the $HGT_{AIT}$ as explained in section 2.6. Then, for each position of each trajectory of the $HGT_{AIT}$, positions of other trajectories of the $HGT_{AIT}$ are interpolated using their normalized time. This second step of the main route computation generates a subset of positions at each normalized time (note that only meaningful positions kept by the spatio-temporal Douglas & Peucker algorithm are used, so that the computation process is only applied on subparts of trajectories where mobile object behaviour changes). Median positions are computed at each normalized time using median values of coordinates (latitudes and longitudes) of each position subset. Then, these computed median positions are ordered according to their normalized time to create the main route of the itinerary. Finally, this main route is also filtered using the Spatio-temporal Douglas & Peucker algorithm (section 2.5). Algorithm 1 summarizes the main route computation steps.

**Algorithm 1** Main route computation

**Require:**
**for** each trajectory $Tr$ of the $HGT_{AIT}$ **do**
    Delete erroneous trajectories
    Spatial shifting of starting and ending positions
    Douglas_Peucker_ST(Trajectory $Tr$)
    Temporal normalization using median duration $t_m$
**end for**
**Algorithm** Main_Route_Computation($HGT_{AIT}$)
**for** each trajectory $Tr_i$ of the $HGT_{AIT}$ **do**
    **for** each position $P_i$ of $Tr_i$ **do**
        Let $tn_i$ be the normalized time of $P_i$
        **for** each other trajectoiries $Tr_j$ of the $HGT_{AIT}$ **do**
            Interpolate the positions $P_j$ at normalized time $tn_i$
            Add $P_j$ to the subset of positions $EP_i$
        **end for**
        Compute median position $P_{med}$ of $EP_i$
        Add $P_{med}$ to the main route $R_{IT}$ at normalized time $t_{ni}$
    **end for**
**end for**
**return** Douglas_Peucker_ST(Trajectory $R_{IT}$)

### 3.2 Spatial channel computation

As the studied mobile objects move in an open area, some of them can move away from the main route. These slight deviations must be distinguished from outliers. The goal of the spatial channel computation is to detect outlier positions of trajectories that spread out of this spatial channel. These unusual deviations affect a small subset of positions within some trajectories of the $HGT_{AIT}$. In order to distinguish normal and unusual trajectories, a spatial channel is calculated using a statistical analysis of all the trajectories of the $HGT_{AIT}$ compared to the main route.

Positions of all trajectories of the $HGT_{AIT}$ are ordered by distance and side to the main route using crossing positions between trajectories of the $HGT_{AIT}$ and the line perpendicular to the heading $(LPH)$ of each previously calculated position of the main route. On each side of the main route, the positions of trajectories which intersect with the $LPH$ are ordered by distance from the main route's position. The sorted position corresponding to the ninth decile of each side of the main route is used to create the border of the channel. Positions outside of this spatial limit are considered as outliers. The choice of this statistical decile provides a channel within which most of the mobile objects following this itinerary move along. Algorithm 2 summarizes the different steps used to calculate the spatial channel $(SC)$.

**Algorithm 2** Spatial channel computation

**Algorithm** Spatial_Channel_Computation($HGT_{AIT}$)
**for** each position $P_i$ of the main route $R_{IT}$ **do**
    Compute line $[LPH]$ perpendicular to the heading of $P_i$
    **for** each trajectory $Tr_j$ of the $HGT_{AIT}$ **do**
        Compute crossing position $P_i'$ between $Tr_j$ and $[LPH]$
        **if** $P_i'$ is right $P_i$ **then**
            Store $P_i'$ in array $A_{right}$
        **else**
            Store $P_i'$ in array $A_{left}$
        **end if**
    **end for**
    Sort $A_{right}$ and $A_{left}$ by distance to $P_i$
    $Pfi_{right}$ = ninth decile of $A_{right}$
    $Pfi_{left}$ = ninth decile of $A_{left}$
    Set $Pfi_{right}$ and $Pfi_{left}$ timestamp to $P_i$ one
    Add $Pfi_{right}$ to the right border $Tr_{right}$ of spatial channel
    Add $Pfi_{left}$ to the left border $Tr_{left}$ of spatial channel
**end for**

### 3.3 Spatio-temporal zones calculation

Given that a moving object is travelling in the spatial channel of a main route, one other interesting element is to know wheter this object is on time compared to the main route. As for positions, temporal zones can be computed in order to temporally qualify the mobile object's position (ahead of schedule, on time, late).

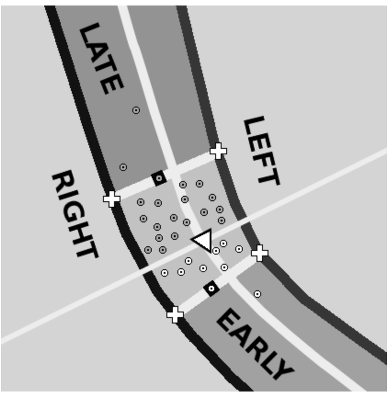


FIG. 3: Spatio-temporal zone at a relative time

To generate these temporal zones, once the spatial channel is computed, the trajectories outside the spatial channel are first removed from the $HGT_{AIT}$. Then, for each position of the main route $P_{Ri}$ (represented by a white triangle on figure 3) using its relative time $tP_{Ri}$, all other positions of the $HGT_{AIT}$ are interpolated. These positions are converted into a new polar system using $P_{Ri}$ as pole and $P_{Ri}$'s heading as polar axis. This conversion defines a total order for each position subset according to distance. Distances $r_{ij}$ and azimuth $\theta_{ij}$ of each interpolated position from the $HGT_{AIT}$ are then divided into two subsets according to the azimuth $((\theta_{ij} > 90^\circ \wedge \theta_{ij} <= 270^\circ) \rightarrow P_j delayed)$ and then sorted by distance (white dots for early positions and grey dots for late ones as shown in figure 3). Finally the positions whose distances $r_{ij}$ match the ninth decile of each subset ($P_{Li}$ for late positions and $P_{Ei}$ for early positions) are selected (shown as black squares in figure 3).Then, the projected positions of $P_{Ei}$ and $P_{Li}$ on the main route are computed ($P'_{Ei}$ and $P'_{Li}$). The crossing positions (white crosses in figure 3) between the spatial channel and the lines perpendicular to $P'_{Ei}$ and $P'_{Li}$ are used to create the temporal normality zone $Z_N$ for each $tP_{Ri}$. Spatial channel and temporal zones at each relative time can be combined to create the spatio-temporal channel which is then stored in the knowldege database. As new positions are frequently acquired by the system, this spatio-temporal channel can be improved by updating it periodically.

**Algorithm 3** Temporal zones computation

**Require:**
Let $R_{IT}$ be the main route of the $HGT_{AIT}$
Let $SC$ be the spatial channel of the $HGT_{AIT}$
**Algorithm** Temporal_Zone_Computation($HGT_{AIT}$)
Remove every trajectory of the $HGT_{AIT}$ which lies out of $SC$
**for** each position $P_{Ri}$ of the main route $R_{IT}$ **do**
    Let $tP_{Ri}$ be the relative time of $P_i$
    Let $H_{Ri}$ be the heading of $P_i$
    Change the polar system using $P_{Ri}$ as pole and $H_{Ri}$ as polar axis
    **for** each trajectory $Tr_j$ of the $HGT_{AIT}$ **do**
        Interpolate position $P_j$ of $Tr_j$ at relative time $tP_{Ri}$
        Compute $r_{ij}$, the distance between the pole and $P_j$
        Compute $\theta_{ij}$, the angle between the polar axis and $P_j$
        **if** $(\theta_{ij} > 90^\circ \wedge \theta_{ij} <= 270^\circ)$ **then**
            Store $r_{ij}$ in array $A_{late}$
        **else**
            Store $r_{ij}$ in array $A_{early}$
        **end if**
        Sort $A_{late}$ and $A_{early}$ by distance $r_{ij}$ to $P_i$
        $r_{late}$ = ninth decile of $A_{late}$
        $r_{early}$ = ninth decile of $A_{early}$
        Store $r_{late}$ and $r_{early}$ for relative time $tP_{Ri}$
        Using $P_{Ri}$ speed, $r_{late}$ and $r_{early}$, interpolate positions on $SC$ and $R_{IT}$
        Create normality zone $Z_N$ using interpolated positions
    **end for**
**end for**

## 4 EXPERIMENT

This section presents the results of the process exposed in previous sections applied to a maritime context. The shipping freight traffic is constantly increasing and traffic surveillance operators can have to visualy monitor up to 250 ships displayed simultaneously on theirs displays. For safety purposes, ships are fited out with *Automatic Identification System* (AIS) to track ships' positions in real time using GPS receivers and VHF transmission systems (IMO, 2007). The spatio-temporal database studied in this example contains 1005 ships and 4 821 447 positions stored since May 2007 in the Brest area (Iroise sea). This spatio-temporal database works using a PostgreSQL/PostGIS server. Each position is associated to a ship whose features are also stored in this database.

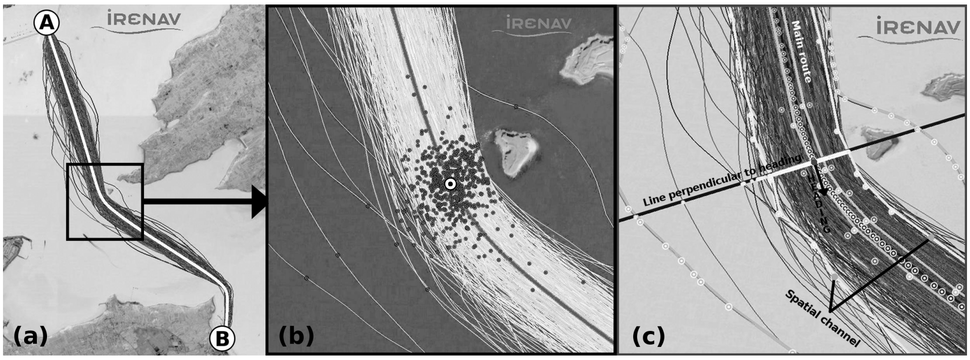


FIG. 4: Main route and spatial channel computation, position cloud at same normalized time

Using the $STDB$ spatial extraction tools, ship trajectories can be distinguished and extracted from this database. As explained in section 2.1, a spatial zone ($Z$) can be defined and represented by geometric areas ($Z.g$) of points of interest. In a place where mobile objects usually stop or interact, where traffic is limited by the geography or by regulations, a zone is defined. As the mobile object move in an open space, there is no forced network between these zones (except for limited traffic due to regulation or geography), the space is a wide open area which allows ships to navigate from one place to another using the most effective path. A position $P_o$ is included into an area $Z_i$ if its coordinates are included into the geometrical surface $Z_i.g$ of the zone. The geometry of the zone must also be large enough to include at least one position of each trajectory that cross this zone (otherwise interpolated positions may have to be calculated). The zone graph of our example is depicted in figure 5 where labeled white circles stands for zones of interest.

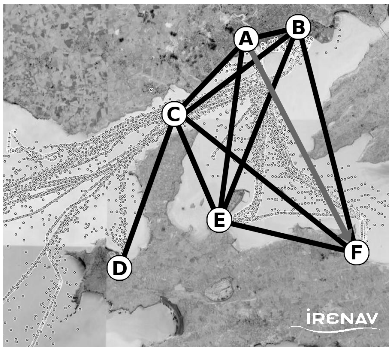


FIG. 5: Zone graph of the $STDB$

Thus, the itinerary shown in Figure 5 by the arc (A, F) (Brest Arsenal $\rightarrow$ Lanvéoc Naval Academy) of the zone graph $G_Z$ is different from the one represented by the string ( A, E, F) (Brest Arsenal $\rightarrow$ Ile Longue $\rightarrow$ Lanvéoc Naval Academy). The zone graph is incomplete and directed, all its vertices are not connected directly with each other by an edge and the way back of the itinerary may be different as navigation rules can set distinct channels in order to avoid collisions. Once set up, this zone graph is saved in the knowledge database (figure 1 step 5). The numerous dots shown in figure 5 represent positions of ships. The main routes used by most of the ships are visually noticeable as dense areas.

Once the graph zone established, an homogeneous group of trajectories is extracted from the $STDB$ as explained in section 2.2. The first selection criterion used to extract this set of trajectories is the type of the mobile object. Applied to our maritime example, only "passenger ships" are selected (30 vessels out of 1005) then the data mining extraction method identified 554 trajectories of passenger ships following the itinerary "Brest Arsenal $\Rightarrow$ Lanvéoc Naval Academy" represented by the arc (A, F) on figure 5. Next, trajectories containing important communication loss compared to normal transmission rate (No position for 1 minute for the AIS system), erroneous positions or transmission gaps are discarded from the $HGT_{AIT}$ as explained in section 2.3. Among 554 trajectories, 506 trajectories were kept after filtering out erroneous trajectories, which is enough to apply statistical analysis to this set of trajectories. The starting and ending position of the remaining trajectories are spatialy shifted as exposed in section 2.4. This spatial shifting avoid a maximum 200-meter distance between farthest starting positions and the projected one on the starting line as shown on figure 2.a. The spatio-temporal Douglas & Peucker filter exposed in section 2.5 applied to the $HGT_{AIT}$ reduced the number of positions from 104 201 to 16 110 (compression rate of 84.54 %) for a threshold of 10 m (precision of a GPS device).

The extracted and filtered $HGT_{AIT}$ composed of 506 trajectories plotted in black in figure 4.a is then used to compute spatio-temporal patterns presented in sections 3.1 and 3.2. Looking at figure 4.a visually shows that same-type mobile objects with the same itinerary globally follow a main route. The cloud of dark dots shown in figure 4.b represents the subset of positions at a same normalized time, the large white dot indicates the median position of the whole subset. All these median positions ordered by theirs normalized time compose the main route plotted in white in figure 4.

Once the main route calculated, the spatio temporal channel can be statisticaly computed using algorithms 2 and 3 presented in sections 3.2 and 3.3. Figure 4.c shows the calculated borders of the spatial channel applied to our example. Thus, unusual positions outside the spatial channel can be highlighted for each $HGT_{AIT}$. The distances between the main route and the spatial channel borders (right and left) are different. Indeed, it is easier for a moving object to deviate outward than to get closer to an obstacle in an open space area. Similarly, the width of the channel provides information about the trajectories spreading from the main route. In our example, this spreading is narrower at the start, the end and in the curves of the itinerary. However, in straight parts of trajectories, spatial channel width increases. The choice of the statistical decile used to compute the spatio-temporal channel gives a more or less wide spread of this channel within which most of the mobile objects following this itinerary move along.

Finaly, as shown in Figure 3, positions of the trajectory of a passenger ship going from Brest to Lanvéoc can be qualified using the five spatio-temporal zones previously-defined in section 3.3. Only 30 positions are displayed in order to keep Figure 3 readable. Positions of the ship are spatially and temporally qualified in order to alert the traffic operator about the unusual behaviour of a ship. The operator can then focus on a few ships within a huge set of vessels cruising in a wide area. Note that the distances between the main route's position and the early and late zone borders are quite different as it is more frequent for a moving object to be delayed than to be ahead of schedule. Moreover, a position outside the spatial channel cannot be temporally qualified as early or late, indeed the moving object moving away from the route can either take a shortcut or make a detour.

## 5 DISCUSSIONS

The novelty of the method is the use of meta-knowledge ($HGT_{AIT}$, main route, spatio-temporal channels) to describe the behaviour of mobile objects following an itinerary on both spatial and temporal criteria. Moreover, these meta-data could be used to qualify new mobile object's positions in real time. The graph zone used to define arc of itineraries can bridge this study to the network based approach of trajectory analysis. However, matching a position to an itinerary in real time remains a complex problem to solve as some arcs of an itinerary can be shared. Previous position of the mobile object coupled with its destination can facilitate the matching to an itinerary but every time a new position is obtained, this matching may change. Tracks for future research include extending our analysis to sections of trajectories. By analogy to the zone graph, it would be interesting to split trajectories into subsections to enhance analysis of the behaviour of a ship on a subpart of the trajectory sharing common properties (speed, heading, rate of turn...). Sections of trajectories could be compared to the main routes. Furthermore, computation time could be decreased by filtering the whole $BDST$ using the Douglas and Peucker spatio-temporal algorithm and adding a trajectory index. Indeed, 50,04% of CPU was used to extract and filter the $HGT_{AIT}$. Finally, the main selection criteria used in this analysis is the type of the ship which does not take into account the environment of the ship (such as the tide, wind or season).

## 6 CONCLUSION

This article focused on the specific problem of outlier detection in mobile object displacements in an open area. It was applied to a maritime context as shown in our case study based on an important dataset. Once the notion of itinerary and trajectory following an arc of an itinerary formaly defined, a general process to qualify mobile object behaviour based on spatio-temporal data mining was defined as previously exposed in figure 1. First of all, position data are acquired and a knowledge database is set up with the zone graph. Then, trajectories of same-kind mobile objects are clustered according to arcs of itineraries. A statistical analysis of each cluster allows to define the main route and spatio-temporal channel of this cluster. These meta-data are stored in the knowledge database. Each new position can be spatially and temporally qualifyed. These processes have been tested on an important dataset applied to the maritime context in different area. Thus, statistical analysis of a $GHT_{AIT}$ gives us information about mobile object's behaviour. Thanks to the spatio-temporal channels, positions of a trajectory can be qualified on both spatial and temporal criteria. It could be worthwhile to validate this study by providing these tools to traffic surveillance operators who can monitor up to 250 ships displayed simultaneously in order to decrease the operator's cognitive load. However, real time analysis tools have not yet been implemented to this prototype.

## RÉFÉRENCES


Aggarwal, C. C. and Yu, P. S., 2001. Outlier detection for high dimensional data. SIGMOD 30(2), pp. 37–46.

Barnett, V. and Lewis, T., 1994. Outliers in Statistical Data. John Wiley & Sons New York.

Baud, O., El-Bied, Y., Honore, N. and Taupin, O., 2007. Trajectory comparison for civil aircraft. In : Aerospace Conference, 2007 IEEE, pp. 1–9.

Bertrand, F., Bouju, A., Claramunt, C., Devogele, T. and Ray, C., 2007. Web and Wireless Geographical Information Systems. Lecture Notes in Computer Science, Vol. 4857, Springer Berlin / Heidelberg, chapter Web Architecture for Monitoring and Visualizing Mobile Objects in Maritime Contexts, pp. 94–105.

Cao, H., Wolfson, O. and Trajcevski, G., 2006. Spatio temporal data reduction with deterministic error bounds. VLDB Journal 15, pp. 221–228.

D'Auria, M., Nanni, M. and Pedreschi, D., 2006. Time-focused density-based clustering of trajectories of moving objects. Journal of Intelligent Information Systems 27(3), pp. 267–289.

Douglas, D. H. and Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica : The International Journal for Geographic Information and Geovisualization 10, pp. 112–122.

Giannotti, F., Nanni, M., Pinelli, F. and Pedreschi, D., 2007. Trajectory pattern mining. In : KDD '07 : Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, NY, USA, pp. 330–339.

Güting, R. H., 1994. An introduction to spatial database systems. VLDB Journal 3, pp. 357–399.

IMO, 2007. Development of an e-navigation strategy. Technical report, International Maritime Organization.

Kharrat, A., Popa, I. S., Zeitouni, K. and Faiz, S., 2008. Clustering Algorithm for Network Constraint Trajectories. Springer Berlin Heidelberg, chapter Clustering Algorithm for Network Constraint Trajectories, pp. 631–647.

Knorr, E. M., Ng, R. T. and Tucakov, V., 2000. Distance-based outliers : algorithms and applications. The VLDB Journal 8(3-4), pp. 237–253.

Lee, J., Han, J. and Li, X., 2008. Trajectory outlier detection : A partition-and-detect framework. In : Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on Data Engineering, pp. 140–149.

Meratnia, N. and de By, R. A., 2004. Spatiotemporal compression techniques for moving point objects. In : Advances in Database Technology - EDBT 2004, Lecture Notes in Computer Science, Vol. 2992, Springer Berlin / Heidelberg, pp. 561–562.

Papadimitriou, S., Kitagawa, H., Gibbons, P. and Faloutsos, C., 2003. Loci : fast outlier detection using the local correlation integral. In : Data Engineering, 2003. Proceedings. 19th International Conference on Data Engineering, pp. 315–326.

Ramaswamy, S., Rastogi, R. and Shim, K., 2000. Efficient algorithms for mining outliers from large data sets. In : SIGMOD '00 : Proceedings of the 2000 ACM SIGMOD international conference on Management of data, ACM, New York, NY, USA, pp. 427–438.

Rasetic, S., Sander, J., Elding, J. and Nascimento, M. A., 2005. A trajectory splitting model for efficient spatio-temporal indexing. In : VLDB '05 : Proceedings of the 31st international conference on Very large data bases, VLDB Endowment, pp. 934–945.

Wu, Y. and Pelot, R., 2007. Geomatics Solutions for Disaster Management. Jonathan Li and Sisi Zlatanova and Andrea G. Fabbri, chapter Comparison of Simplifying Line Algorithms for Recreational Boating Trajectory Dedensification, pp. 321–334.