CETRA^{2022} 11-13 May 2022, Pula, Croatia
7th International Conferenceon Road and Rail Infrastructure

STATISTICAL ANALYSIS OF FLUVIAL TRAJECTORIES BASED ON AIS DATABASE FOR THE CONSTRUCTION OF A BRIDGE

Pierre-Olivier Vandanjon[1], Alex Coiret[2], Tristan Lorino[3]
[1] *AME-SPLOTT, Univ Gustave Eiffel, Bouguenais, France*
[2] *COSYS-SII, Univ Gustave Eiffel, Bouguenais, France*
[3] *AME-UMRAE, Univ Gustave Eiffel, Bouguenais, France*

Abstract

The French metropolis, Rouen Normandie, has a project of a new bridge requiring temporary pier in the river Seine during the building phase. The location of this pier in the river is constrained by mechanical reasons related to the construction of the bridge. The purpose of this study is to find a position of this pier in the mechanically constrained area which minimizes vessel traffic obstruction and therefore collision risk. This type of study is classically carried out by using complex and multiple vessel dynamics simulation software in order to assess the risk for a given vessel to crash into the pier. The novelty of this study is to propose a statistical study based on the database of Automatic Identification System (AIS) of the Vessel Traffic Service (VTS) that is an automatic tracking system relying on ships transceivers. The technical objective is to identify low risk areas according to vessel speeds and sizes. Both factors should have an impact on the maneuverability of ships to avoid pier collision risk. Among all trajectories, straight line trajectories are selected based on statistical methods. The computation of prediction intervals of these trajectories delimits navigation zone. If all the straight trajectories are taken into account, the navigation zone extends to the whole river surface, which is not an helping result. However, by focusing on large and high speed vessels trajectories the navigation zone of these weak maneuverable ships is more centered in the middle of the river and represents only 25 % of the river area. The complement of this area delineates possible locations of the temporary pier of the bridge in the river that do not disrupt vessel traffic.

Keywords: automatic identification system, linear regression, prediction interval, R software, vessel trajectories, collision risk

1 Introduction

The French Metropolis “Rouen Normandie” has a project of bridge over the Seine for active mean of transportation, such as walking, bicycling and in-line skating. The future implantation of this bridge is displayed on Figure 1. Although It is not considered to install a permanent pier in the Seine the recourse to intermediate supports during the construction phase of the work is unavoidable. Similarly, the definition of a navigable channel gauge influences the definition of the lower limit of the structure’s intrados and therefore the design of the bridge’s accesses. Only a detailed knowledge of the river’s vessel trajectories can make it possible to gain in precision on the definition of the possible establishments of provisional supports, then on the definition of the definitive river height gauge.

DOI: https://doi.org/10.5592/CO/CETRA.2022.1433

This knowledge of trajectories is now possible. Indeed, the vessels concerned by the Solas (Safety of Life at Sea) convention of the International Maritime Organization must be equipped with an identification system called AIS (Automatic Identification System) [4] which transmits the position of the vessel. In addition, vessels that do not fall within the scope of Solas are spontaneously equipped with this type of system for safety reasons. Thus, the fluvial authority (Grand Port Maritime de Rouen, GPMR) has access to the trajectories of ships and records them in a database since 2006.
From this database, a statistical analysis of ship trajectories in the study area defined by the GPMR was conducted. First, this article presents the context of the data acquisition, as well as the data nature. Then, it provides the different steps of the statistical analysis procedure, notably, the identification of individual trajectories by direction of navigation, gauge and speed, and their adjustment by a linear regression model. Finally, it provides illustrations of the various maritime flows and concludes with the determination of suitable areas for the installation of the temporary supports of the bridge.

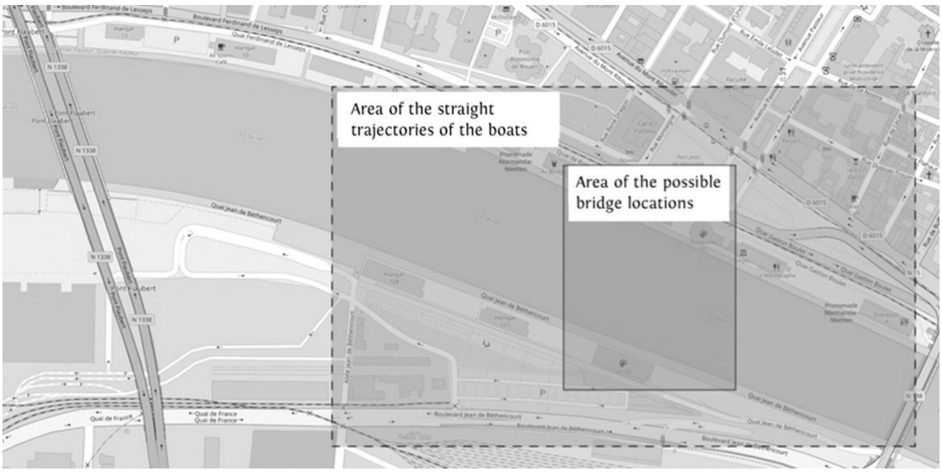

Figure 1 Implantation project across the Seine river at Rouen city.

2 Data processing

2.1 Data description

The data were provided by the fluvial Authority. The data are vessel locations provided in the format of AIS messages. Two types of messages are transmitted by the vessels and stored in text format: dynamic messages (the time interval between each message is a function of the vessel's speed and gyration) and static messages relating to the vessel's identity, geometry and general characteristics. Among the information contained in these messages, the following features, necessary for the determination of river trajectories, are:

- for the dynamic message : identification MMSI (Maritime Mobile Service Identity), the speed in knots (kn), the abscissa and ordinate in Lambert I north, the date and time of the recording of the message;
- for the static message : identification MMSI, the distances between the bow and the position of the GPS device (m) and between the stern and the position of the GPS device (m), the distances port side to GPS and starboard side to GPS.

Only the 2016 data were retained for statistical analysis, due to their large number (Tab.1), and thus their representativeness of all current maritime flows. The choice of 2016 was confirmed by fluvial authority.

Table 1 Data amount for the 2016 year: messages and vessels

	Dynamic data		Static data	
Time period	**Messages**	**Vessels**	**Messages**	**Vessels**
January-March	3, 657, 602	232	3, 651, 976	230
April-June	3, 448, 677	241	3, 389, 631	240
July-September	2, 280, 994	641	2, 271, 854	630
Oct.-December	2, 980, 323	178	2, 945, 159	178

2.2 Data processing

The data processing, carried out with the "R" statistical computing software [2] and libraries specific to the manipulation of large data tables [1, 3], proceeded according to the following six steps.
Step 1: Dynamic Data: Import data; Extraction of the years, months and days of record; Selection of 3 consecutive months of 2016.
Step 2: Static data: Importing the data; Extraction of the years, months and days of recording; Selection of the same 3 consecutive months of 2016; Compute the width of the vessels (distance from port side to GPS antenna + distance from starboard side to GPS antenna), then their size (small size if the width is less than 10 m, large size otherwise).
Step 3: Determination of the banks: Determination of the linear equations of the polylines constituted by the banks on their rectilinear part (Lambert ordinate as a function of the Lambert abscissa).
Step 4: Truncation of trajectories: Restriction of the data to the straight part of the trajectories; Restriction of the data to the trajectories located between the two banks (removal of outliers due to GPS inaccuracy).
Step 5: Determining the direction of travel: Add an IdN identifier for each modality of the MMSI/month/day crossing; For each IdN, compute the difference Δ between the Lambert ordinates of two temporally successive records (taking into account the hours and seconds of each record). If $\Delta = 0$: ship is motionless ; if $\Delta > 0$: displacement in downstream direction, otherwise in upstream direction this choice being valid locally, i.e. for the studied meander.
Step 6: Determination of the linear trajectories: Add an IdT identifier for each modality of the MMSI/month/day/direction crossing; Removal of motionless ships ($\Delta = 0$).

2.3 Data filtering and statistical analysis

Step 6 of the data processing requires data filtering which is presented in the first part of this subsection. In order to calculate a set of trajectories, we use a confidence interval build on a linear regression. This method is presented in the second part of this subsection. This method is applied after segmenting the trajectories into several categories.
Determination of linear trajectories
In order to distinguish vessels moving in the same direction on a permanent basis from other vessels (moored, docking or tacking, etc.) which do not impact the bridge design, the following assumption is made. Given the geometry of the study area (see Fig. 2 and 3), the vessels that evolve durably and have therefore to be taken into account in the bridge design have a straight trajectory. A linear regression model is applied to each of their trajectories. This model is written:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

where Y and X are respectively the Lambert 1 ordinates and abscissas of a ship's positioning point, β_0 and β_1 are coefficients to be estimated, and ε is an error term assumed to follow a centered Gaussian distribution of variance σ^2. The determination coefficient R^2, associated with the equation 1, measures the adequacy of the regression model to the observed data that allowed to establish this model. By definition:

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

with $\hat{y}_i$ the ordinate i from the linear fit, $\overline{y}$ the mean of yi and n the sample size.

The calculation of R^2 led to the division of the trajectories into two groups: on the one hand the linear trajectories ($R^2 > 90$ %), on the other hand the non-linear trajectories ($R^2 \leq 90$ %). This value of 90 %, which reflects an excellent fit of the model to the data, led to the identification of 70 % of the totality of the trajectories as being linear trajectories.
Figure 2 and the following include: the William the Conqueror and Flaubert bridges, represented by parallelograms at both ends of the figure; the banks connecting these bridges, a famous tower in Rouen city, called XXL, symbolized by an oval, serving as a geographical reference point.
Figure 2 displays non-linear trajectories, which characterize docking maneuvers: the points symbolizing the trajectories indeed touch the banks. In contrast, Figure 3 illustrates the linearity of trajectories for which the coefficient of determination exceeded 90 %.

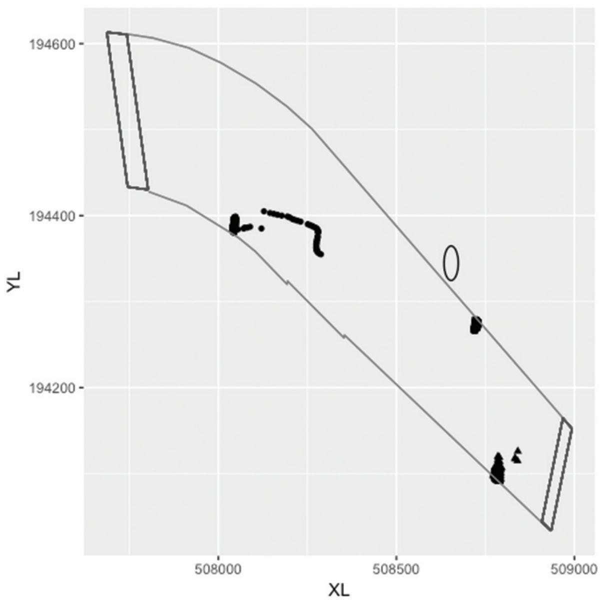

Figure 2 Non-linear trajectories

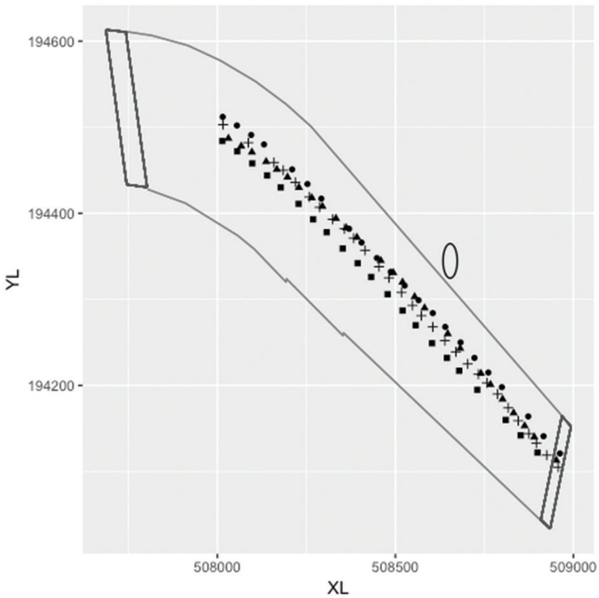

Figure 3 Linear trajectories examples

Vessels classification

Of all the ships, only those with low maneuverability and large size have an influence on the bridge design. The size is directly given by the database. It is assumed that maneuverability is correlated with the speed and size of vessels. The faster a ship goes and the larger it is, the more difficult it is to change its direction. The statistical analysis therefore required splitting the set of linear trajectories into 32 groups, according to the values taken by the following four factors:

1) the speed V of the ship: constitution of two modalities (V › 5 kn and V › 10 kn). Note: the knot (kn) is a unit of speed equal to one nautical mile per hour, 1 kn = 1.852 km/h;
2) the gauge: two modalities, small gauge when the width of the ship is lower than 10 m and large gauge in the contrary case;
3) the direction of navigation: two modalities (upstream and downstream).
4) the quarter of 2016: constitution of four modalities.

An estimate of a set of trajectories for each category For each of these 32 groups, a linear regression model is identified on the Lambert coordinates of all trajectories belonging to the same modality. Then the 95 % prediction interval associated with this model is determined. This interval defines the limits between which would fall, in 95 % of cases, $\hat{y}_k$ the value of Lambert ordinate predicted by the model for a new xk value of Lambert abscissa. This model predicts the trajectories of the ships for the considered modality. This confidence interval is the estimate of the set of trajectories. This procedure is based on the following formulas.

$$\left[\hat{\beta}_0 + \hat{\beta}_{1x0} \pm t_{1-\alpha/2}^{n-2} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}\right)}\right] \quad (3)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \; \hat{\beta}_1 = \overline{y} - \hat{\beta}_1 \overline{x} \tag{4}$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{5}$$

$t_{1-\alpha/2}^{n-2}$ Follows the Student law.

3 Results

Figure 4 illustrates the totality of the trajectories for the first quarter of 2016. The set of straight trajectories covers the entire width of the Seine. This result does not answer the questions concerning the positioning of the temporary supports during the construction phase of the bridge, nor the height of the intrados.

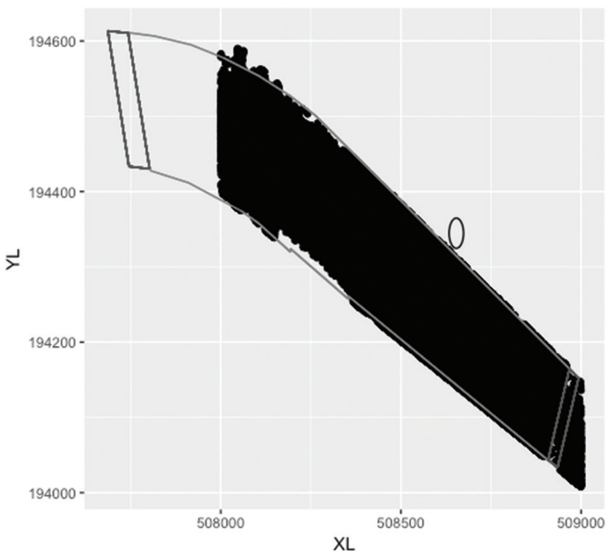

Figure 4 Whole set of vessel trajectories over the first trimester of 2016.

The results are then presented according to to the speed and size of the vessels. On these figures, the regression line has been added (light bold line), as well as the 95 % prediction interval of the trajectories (gray parallelogram). This parallelogram is the set of trajectories determined by the method. Its surface estimates the percentage of the river area used by traffic for the different modalities. Only the results for downstream-oriented trajectories recorded during the first quarter of 2016 are presented because these results are similar to those obtained for the upstream direction and for the other quarters.

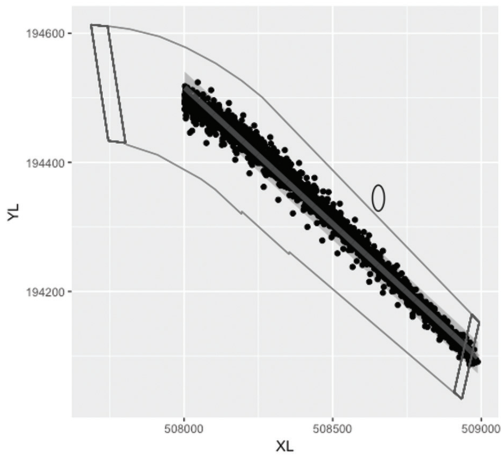

Figure 5 Small vessels, speed › 10 kn (1 kn = 1.852 km/h)

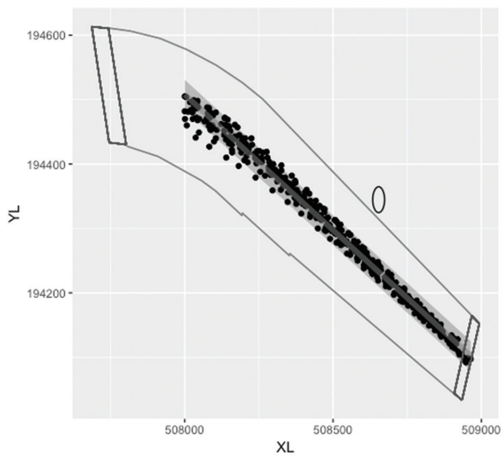

Figure 6 Large vessels, speed › 10 kn (1 kn = 1.852 km/h)

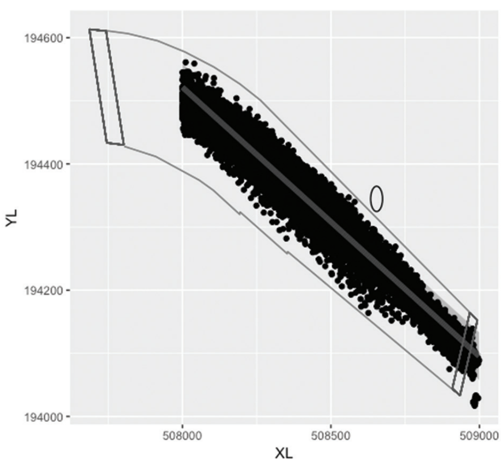

Figure 7 Small vessels, speed ‹ 5 kn (1 kn = 1.852 km/h)

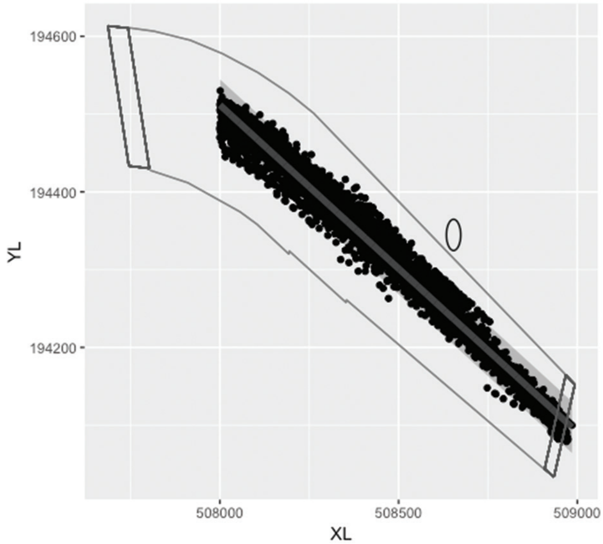

Figure 8 Large vessels speed < 5 kn (1 kn = 1.852 km/h)

At first glance, for high speeds, it appears that the set of trajectories is restricted to a central portion of the river: this portion is of the order of 30 % for small gauges (Fig. 5) and of 25 % for large gauges (Fig. 6). The situation is less favourable to the constraints of bridge location for low speeds: indeed, small-gauge vessels are occupying about 80 % of the area of the Seine (Fig. 7) and large-gauge vessels are occupying about 60 % of the river surface (Fig. 8). A secondary result is that the distinction between large and small vessels does not determine the location of vessel tracks as significantly as speed.

4 Conclusion

River crossings are expensive engineering structures: it is fundamental to design them as well as possible. Since a few years, ships are geo-located and their trajectories are recorded. In this paper, a method using this large amount of available data is proposed in order to help the design of a bridge. First of all, the ships that sailed in a straight line, i.e. that were not maneuvring, were selected thanks to a linear regression model and the calculation of the associated regression coefficient.

This first selection, which is only statistical, appears to be not sufficient to determine a set of trajectories that could be used by the project owner. Then the trajectories are segmented according to the speed and size of the vessels. The assumption underlying this segmentation is that the larger and faster the vessels, the less they maneuver and the closer they are to a nominal path through the center of the river. For each of the speed and gauge modalities, a linear model for predicting vessel trajectories is identified. The 95 % prediction interval associated with this model determines the set of trajectories.

The results confirm the underlying assumption that large vessels, which travel above 10kn, have a straight path through the center of the river. They use only a quarter of the river's area, which is a relevant information for the project owner. With this information, the bridge designer knows where to place the temporary pier during the construction phase so as not to disrupt river traffic. This estimation of a navigable channel gauge influences also the definition of the lower limit of the structure's intrados which is a sizing parameter for engineers.

Acknowledgement

This study was carried out with the financial and technical support of the Rouen Normandie Metropolis.

References

[1] Hadley Wickham, H., Francois, R., Henry, L., et al: Package dplyr: A Grammar of Data Manipulation, R Project, 2022.

[2] The R Project for Statistical Computing: https://www.r-project.org, 7.3.2022.

[3] Dowle, M., Srinivasan, A., Gorecki, J. et al.: Package data.table: Extension of data.frame, R Project, 2022.

[4] Harati-Mokhtari, A., Wall, A., Brooks, P., et al.: Automatic Identification System (AIS): Data Reliability and Human Error Implications, The Journal of Navigation, 60 (2007), pp. 373-389