Article

Reconstruction of Marine Traffic from Sémaphore Data: A Python-GIS Procedure to Build Synthetic Navigation Routes and Analyze Their Temporal Variation

Annalisa Minelli ¹,* Iwan Le Berre ², Ingrid Peuziat ² and Mathias Rouan ²

1 National Research Council—Institute of Marine Biological Resources and Biotechnologies—CNR-IRBIM, Largo Fiera della Pesca 1, 60125 Ancona, Italy
2 LETG-Brest GEOMER, UMR6554 CNRS, IUEM-Université de Bretagne Occidentale, Rue Dumont D’Urville, 29280 Plouzané, France; iwan.bleberre@univ-brest.fr (I.L.B.); ingrid.peuziat@univ-brest.fr (I.P.); mathias.rouan@univ-brest.fr (M.R.)
* Correspondence: annalisa.minelli@irbim.cnr.it

Abstract: Originally designed as a mode of telecommunication, the network of French sémaphore is now dedicated to the continuous monitoring and recording of marine traffic along the entire French coast. Although the observation data collected by sémaphores cover 7/7 days and 24/24 h and could provide precious information regarding marine traffic, they remain underexploited. Indeed, these data concern all types of traffic, including leisure boating and smaller craft that are not usually recorded by the most common means of observation, such as AIS, radar and satellite. Based on sémaphore data, traffic pressure and its spatiotemporal distribution can be fully measured to better analyze its interactions with human activities and the environment. One drawback of these data is their initially semantic nature, which requires the development of an original processing method. The protocol developed to analyze the marine traffic of the Iroise Sea and its first results are presented in this article. It is based on a semi-automatic method aimed to clean the original data and quantify the marine traffic along synthetic routes. It includes a procedure that takes into account the temporal evolution of the traffic based on the Allen’s time framework. The results proved interesting as they provide an overview of marine traffic, including all types of vessels, and may be defined for different time periods and granularity. A description of the numerical and geographic instruments created is given; all the written code is released as Open Source software and freely available for download and testing.

Keywords: marine traffic; sémaphores; route extraction; vessel position data

1. Introduction

The knowledge of maritime traffic has always been a key issue, particularly for the safety of navigation and for coastal states wanting to control their sovereignty within their exclusive economic zones (e.g., the efforts historically made from the SOLAS conventions for Safety of Life At Sea regarding the safety of merchant ships - https://www.imo.org/en/KnowledgeCentre/ConferencesMeetings/Pages/SOLAS.aspx, accessed on 3 March 2021). In addition to these historical concerns, the tremendous development of international shipping, the difficult management of fishery resources and, more recently, the management of marine renewable energies are among the issues that motivate the gradual establishment of a Maritime Spatial Planning (MSP), both to promote economic development (blue growth) and protect the marine and coastal environment [1].

To properly address these concerns, various means have been developed for the observation of maritime traffic. Observation may be based on port, radar or satellite data [2–5], but data are now mostly collected by AIS devices [6–9]. This type of data can be analyzed using different methods, including statistical approaches such as Markov logic...
or the Bayesian networks approach [10–12], those based on GIS [13,14] or Multi-Agent models [15–18]. Applications can vary widely, but much of the literature is dedicated to the monitoring of traffic for greater security, such as monitoring in order to avoid collisions [19–21], or the individuation of atypical boat behavior by different methods [22–24]. Other work has focused on environmental protection [25,26] or on MSP [27–29]. However, although most ships in the world merchant fleet are now equipped with AIS, its use is not widespread on small fishing vessels (boats smaller than 15 m), and it is even rarer on leisure boats. In addition, especially for small boats, other observation methods also have limitations related to the quality of spatial, temporal or typological description of the marine traffic [30]. Remote sensing has limitations related to both the spatial and time resolution of the observations, making it difficult to distinguish the smallest ships (such as small leisure craft or fishing boats) and, consequently, to classify them properly [5,31,32]. Passive acoustic technologies can continuously record underwater noise, but relate only to motorized navigation [33,34]. In the current state of civil technology, they only cover small coastal areas. Moreover, tracking by GPS or mobile phone still appears largely inoperative, or experimental, in the marine environment [30]. Finally, there are participatory methods, whose interest has been effectively demonstrated for monitoring boating frequentation [35] or professional fishing [36–38], but they are still based on samples of the total traffic.

Yet, our focus is precisely on these small boats not equipped with any geographical tracking tools that frequent the coastal sea, where most of the issues related to the exploitation of resources, water activities, and the protection of the environment are concentrated. This is particularly true in marine protected areas and in the busiest coastal zones, which are therefore most at risk of competitions and conflicts.

To improve the knowledge and the monitoring of these coastal activities, the use of alternatives to geolocation data therefore appears justified. In the French case, observations of sémaphore, which are collected continuously along the entire French coastline, are among the data that can be used for this purpose [30]. Sémaphores are tower-shaped structures (Figure 1), often coupled with signalization tools like lighthouses and equipped with instrumentation such as radar and AIS antenna, managed by the military (and specifically by the Formation Opérationnelle de Surveillance et d’Information Territoriale—FOSIT), where officers alternate themselves in shifts covering 24 h per day and 7 days per week. The officers’ monitoring operations consist of manually recording in a spreadsheet each observed passage of boats. Additional information about the route of the boat (departure/destination) could be asked directly by the officers via radio communication with the vessels. This paper presents a GIS Python procedure to exploit sémaphore observation data to reconstruct the spatiotemporal evolution of coastal maritime traffic. After justifying the relevance of these data and describing their original format, the article outlines the procedure developed in three stages: data cleaning and integration into a consistent and exploitable database, the extraction procedure of maritime routes, and the reconstruction of their temporal evolution. We then present our first results, coming from elaborations of traffic for five different datasets collected between 2011 and 2013, before concluding with a discussion on challenges and prospects for future work. The code written and the input data (raw and cleaned) are available as Open Source software (https://doi.org/10.5281/zenodo.4420051, accessed on 3 March 2021) and Open Data (https://doi.org/10.35110/59385ec4-a2ea-4e3c-89e0-e24a770b7358, accessed on 3 March 2021), respectively.
2. Why Is It Relevant to Use Sémaphore Data on Marine Traffic?

Created as visual communication installations, ancestors of the sémaphore have been well known since the Roman Empire [39], although the methods adopted have evolved over time, from flags and wooden signs to more recent radio transmission. Over the centuries, the functions have also changed: from the mission to speed up communication from one point to another to strategic surveillance, the monitoring of marine traffic and public service [40].

The modern network of marine traffic monitoring consists of 59 sémaphores distributed along the French coast (Figure 2) and grouped into three regions: Mediterranean Sea (red dots), Atlantic Ocean (yellow dots) and English Channel-North Sea (orange dots). This network is supervised and administered by the FOSIT service of the French Navy (Marine Nationale).

The monitoring of marine traffic is continuous (24/24 h, 7/7 days) and based on three types of data: AIS, radar and visual/radio identification, which are grouped into an integrated system called SPATIONAV [41]. This system can identify and track a ship across the French Exclusive Economic Zone (EEZ). It is used for navigation safety purposes in support of Regional Operational Centers for Monitoring and Rescue (CROSS), and for customs surveillance and homeland security missions. Using this system, the navy officers continuously monitor the marine traffic and record all their observations in spreadsheet forms that are archived daily. These forms contain the name, registration number (International Maritime Organization ID for merchant ships; registration ID by maritime quarter for fishing and leisure vessels), type and route of the vessels. Even if these data were not originally spatial but only “semantic” (routes or destinations are described with their toponyms in text form), they provide an interesting source of information, which is fully complementary with geolocalized data such as AIS and the Vessel Monitoring System.
(VMS). Actually, smaller boats, which are more likely to frequent the coastal sea, are poorly equipped with AIS and VMS devices. It is estimated that only 20% of the fishing fleet is equipped with a geolocation system making it possible to track their navigation [8]. For recreational boating, the equipment rate is even lower (less than 10%) and concerns only the largest boats, designed to sail offshore [42]. Yet, leisure boats can account for 50 to 60% of the total maritime traffic of some coastal areas [32,43].

In these conditions, sémaphore data are of particular interest for several reasons: (a) they potentially include all kinds of vessels; (b) their monitoring is based on continuous observation; (c) observed vessels are systematically identified and (d) despite their semantic nature, they provide relevant information for building synthetic navigation flux and their temporal variations. Therefore, they can provide a realistic reconstruction of the total flux of marine vessels and their pressure on the coastal sea.

Using the data from sémaphores, our objective was to develop a GIS model capable of significantly representing and analyzing traffic fluxes in space and time in order to develop more refined research methodologies and specific professional tools. These data thus provide a relevant basis for producing a realistic reconstruction of the entirety of marine traffic in the coastal sea and, therefore, for evaluating its pressure on the natural and human environment. However, their structure and their format require a significant amount of preprocessing before spatio-temporal analysis can be performed.

3. Study Site and Data Structure
3.1. Study Site Description

Data from three sémaphores of the Iroise Sea (Saint-Mathieu, Le Toulinguet and Cap de la Chèvre) were obtained thanks to a convention between the FOSIT and our laboratory. The data contain observations for the years 2011 (Saint-Mathieu, Toulinguet and Cap de la Chèvre), 2012 (Saint-Matthieu) and 2013 (Saint-Mathieu).
The study site chosen to develop a method for processing these data was the Iroise Sea (Figure 3), which is a busy coastal area with a wide range of marine activities. This is due to the presence of both a commercial port and a military port of national importance, dynamic coastal fisheries, mining areas for marine aggregates, plus a variety of recreational and water sports activities that are related to both the touristic attractiveness of the region and to an intensive practice of these activities by a permanent population of more than 400,000 inhabitants. Moreover, its biological, landscape and human characteristics make the Iroise Sea a remarkable site, illustrated by the nature and extent of its protected areas that include a national park (Iroise Marine Natural Park) and a UNESCO Biosphere Reserve. The high level of maritime traffic in these protected areas justified the choice of the study area.

Figure 3. The Iroise Sea. In red, the sémaphores considered in the present work; in grey, the other sémaphores of the study area.

3.2. Dataset Structure

Sémaphore data are recorded daily in two spreadsheets reporting all the military, commercial, fisheries and leisure activities. Using an empty spreadsheet for each recording day, two officers take shifts of 6 h with the task of recording all the marine traffic and noting down the time, name, registration number, type and route of the boats. Although the recordings related to the military traffic are filtered out by the French Navy, all the remaining traffic (commercial, fishing and leisure) is collected for “public service” purposes and has been kindly provided by the FOSIT for use in this research. A total amount of 44,183 boat passages is examined in this work.

The remaining obstacles to use these data are the lack of a completely standardized procedure and the use of a fixed linguistic classification for data collection: the observations are recorded according to different syntaxes, which could generate significant confusion.
in their interpretation (an example of raw data is shown in Table 1). For example, fishing boats are sometimes noted using their specialization name: “pêcheur palangrier”, “pêcheur chalutier” or “pêcheur fileyeur” (longliner, trawler and gillnetter, respectively). For speed-boats, the maker’s name is sometimes specified, but in other cases only the simple name “vedette” is noted. The same problem arises for the route notation, which may include typing and spelling errors, and may be different from one another in formatting and syntax (e.g., “OUESSANT → BREST”, “Ouessan–>Rade de Brest”). For example, in the 2011 Saint-Mathieu recordings alone, composed of about 12,000 observations, there are 961 different routes mentioned when only 100 real routes may be observed, i.e., a 1 to 10 ratio of registration error, which is also completely variable.

Table 1. Extract from the raw data (anonymized) for the Saint-Mathieu sémaphore. The English version of this table is available in the supplementary file.

| Journée Du Samedi 18 Août 2012 | NOM | IMMAT | BD Plaisance TYPES | ROUTE |
|-------------------------------|-----|-------|---------------------|------|
| HEURE                         |     |       |                     |      |
| 12:39:00 a.m.                 | NoName | NoName | Chalutier          | IROISE–>CONCARNEAU |
| 2:40:00 a.m.                  | NoName | NoName | Fileyeur           | LE CONQUET–>ST MATHIEU |
| 3:39:00 a.m.                  | NoName | NoName | Fileyeur           | LE CONQUET–>IROISE |
| 4:16:00 a.m.                  | NoName | NoName | Fileyeur           | LE CONQUET–>IROISE |
| 4:21:00 a.m.                  | NoName | NoName | Fileyeur           | LE CONQUET–>IROISE |
| 4:35:00 a.m.                  | NoName | NoName | Péche              | LE CONQUET–>IROISE |
| 5:59:00 a.m.                  | NoName | NoName | Fileyeur           | Le Conquet–>Sud Iroise |
| 6:40:00 a.m.                  | NoName | NoName | Sloop 17 m        | St Malo–>Sud Bretagne |
| 7:42:00 a.m.                  | NoName | NoName | Vedette SNSM      | Iroise |
| 8:10:00 a.m.                  | NoName | / | S/V                | CAMARET–>Le Four |
| 8:10:00 a.m.                  | NoName | / | M/Y                | CAMARET–>Sud Iroise |
| 8:40:00 a.m.                  | NoName | NoName | Course/Régate     | AV Goulet–> Le Four |
| 8:45:00 a.m.                  | NoName | NoName | Semi-rigide       | Vieux Moines |
| 8:48:00 a.m.                  | NoName | NoName | Bow Man 42 coque bleu | AV Goulet–> Le Four |
| 8:55:00 a.m.                  | NoName | NoName | Coque Open jaune  | AV Goulet–> Vieux Moines |
| 8:58:00 a.m.                  | NoName | NoName | Fileyeur           | Le Conquet–>St Mathieu |

4. Data Processing

4.1. Data Cleaning and Database Integration

The first step of the data processing consists of the concatenation of the daily recorded files (two files per day, thus giving 730 spreadsheets per year) in a single annual file, properly formatted for date and time information. This work was automated by writing a macro in VisualBasic in Excel.

The next step aims at cleaning the annual file in order to make the data readable and exploitable for various purposes. Since each spreadsheet, for each year and for each sémaphore, can contain several thousand lines, a semi-automatic procedure was used in order to make the first language standardization for the type of boats, usage and routes. For each element, the basic assortment (i.e., for the type of boats: “speedboat”, “lifeboat”, “scow”, etc.) was selected according to the occurrences encountered in the Iroise sémaphores [44]. Later, users could expand this list to suit other regional contexts. The cleaning procedure, reported in the following flowchart (Figure 4), was implemented as Python code using Qt library (https://riverbankcomputing.com/software/pyqt/, accessed on 3 March 2021), and includes a graphical user interface (GUI) in order to be as user friendly as possible. As input, the code takes an annual spreadsheet file and three dictionaries containing the fundamental information: types of usage, types of vessel and routes related to the raw data occurrences for each element. These dictionaries are created automatically, verifying the occurrence of routes in a first core of data (represented by Saint-Mathieu sémaphore traffic for the year 2011), and during the routing it is possible to enlarge these dictionaries to take into account new situations. In case of totally new
elements, new values of the original keys or new keys can be specifically added. The output is made up of a clean annual data file and the updated dictionaries.

**Figure 4.** Flowchart illustrating raw data cleaning and dictionary updating procedures.

### 4.2. Automatic Route Extraction Procedure

Starting from cleaned sémaphore data, it is then possible to quantify traffic fluxes on synthetic routes. In all cases, the data recorded by officers only describe the origin and destination of the boats: these data alone do not allow us to reconstruct their actual routes. A geometric square-cell grid was created in order to connect the “origin” and “destination” points for each boat observed and reconstruct a synthetic route that could potentially have been followed by the boat. This geometric grid, representing all the synthetic routes belonging to each sémaphore, was created using GRASS GIS [45]. The tool we developed allows the user to interactively set the resolution of the grid (the base segment length) and to represent the traffic fluxes by recording the passage of each boat in each network segment located between the origin and destination of the recorded route. This allows the quantification of the traffic at different time scales. However, in this procedure, the routes are not explicitly described. In fact, the navigation of each boat can only be reconstructed from the indications of origin and destination. The synthetic route is therefore calculated from two gates: an entry point (origin) and a way out (destination). The direction of the route is assigned automatically by the code and the route is calculated according to the principle of the shortest route (shortest path route logic from [46]). A GRASS Python tool was developed to calculate the shortest route (called v.createRoutes.py).

The calculation performed in v.createRoutes.py can be divided into two main parts (Figure 5), both of which are written in Python. The first part is dedicated to data preparation; the second concerns data spatialization and is based on the use of the routine “grass_script”.

In the first part of the flowchart, the original data are read and formatted: spaces are removed; the names of the gates are edited according to the routes, etc. Moreover, an analysis of the occurrences of all gates is performed in order to read the geographic coordinates of the identified gates from a text file. Basically, from sémaphore records, a gate can be represented in three ways: a specific and exact point (e.g., a port), a zone (e.g., a group of islands), or a direction.
Figure 5. v.createRoutes.py flowchart: the first part (in Python, bordered in green) concerns the data preparation; the second part (in GRASS GIS, bordered in red) relates to data spatialization.

In fact, the gates are not often easy to represent “spatially”, although the location of ports is not a problem. For the representation of areas, the simplest solution (adopted in this work) is to examine the coordinates of the centroid of the zone itself in order to position the associated gate. Figure 6a shows the placement, at the center of the area surrounded in blue, of the point generically representing “The Islands”. In other cases, the route indicates only a direction: “North”, “South”, “East” or “West”. Therefore, the mean capacity of visibility of the sémaphore (fixed experimentally at 15 km) is considered in order to trace a circle and locate the gates associated with each direction (see Figure 6b for the case of the Saint-Mathieu sémaphore). These first approximations are necessary to process the semantic uncertainties of the cleaned annual data file. As the semantic uncertainties are quite difficult to evaluate [47,48] and are not the main aim of this paper, they are not considered further at this stage, allowing the present paper to focus on the spatio-temporal modeling. The quantification of these semantic uncertainties will be the subject of future studies.

Figure 6. The semantic uncertainties in gate placement. (a): the placement of gates related to a zone (left) and a port (right); (b): the placement of gates related to a direction with respect to the Saint-Mathieu sémaphore (the mean visibility considered is 15 km).

In the second part of the process (framed in red in Figure 5), these coordinates are used by GRASS GIS to create a vector file describing the gates where traffic will be recorded. The resulting file also defines the geographic extent of the work area based on the following...
principles: the center of the grid is that of the vector file gates, and its spatial extent is calculated to include all gates. In addition, the mainland and islands are removed from the network, which is then cleaned of any dangling lines.

The gates are then connected to the network to obtain a single repository. Finally, the column “npass”, in which the traffic will be recorded, is added to both the gates file and the network vector file. In summary, the loop procedure is applied to each line of the cleaned observations file, and it records the traffic in each segment and each gate according to the following steps:

- Extraction of the gates from the cleaned file to select the origin, median (or medians, if present) and destination points;
- Selection of these points in the gates vector file, and searching for the shortest route between these points in the geometric network using the routine “v.net.path” of GRASS GIS;
- For each segment of the grid and each affected gate, the incrementation of the “npass” parameter to record the number of ship passages.

When all records have been examined, the procedure provides a gate file and a grid where each feature (gate points and network segment) contains the number of vessel passages. The size of the visualization symbol (line or point) is proportional to the value of the “npass” parameter. This allows us to instantly visualize the busiest routes, origins and destinations. In Figure 7, the traffic value is explicitly displayed on each gate, while it appears in a query window for each network segment.

![Figure 7. Screenshot of a sample result in output from v.createRoutes.py for the Toulinguet sémaphore.](image)

In order to define the paths, a graph approach could also be used. In this specific case, at a technical level, using a graph is equivalent to using a regular grid. In fact, the vector grid, with respect to the graph, originates as a vector object in a georeferenced world. Conversely, the graph is originally an image (raster file) without the specificity of georeferencing. In fact, most graphs are used to represent non georeferenced objects (an interesting comparison between the two formats can be found in [49]). In this specific case, when speaking about the “grid”, the meaning is a “lines vector file”, and all the algorithms used or created ad hoc for this analysis are applicable to both a grid and a vector (georeferenced) graph.
4.3. The Temporal Framework

The last part of the work concerns the introduction of the time variable into the analysis. Indeed, the v.createRoutes.py model makes it possible to spatialize observational data and to quantify the overall traffic on each synthetic route, but it does not describe their temporal variations, although the time and date of each observation are duly recorded.

The v.createRoutes.py model was therefore changed to exploit the capabilities of the “TGRASS” branch of GRASS GIS [50]. TGRASS allows us to integrate raster and vector maps in a temporal framework where each map is considered as an “event”, each of them having a temporal topology. Based on this topology, the events interact together following the Allen theory [51]. In particular, this theory individuates seven different levels of interactions (mutual event behavior) which make the temporal topology “correct”. The events can be:

- Equivalent: same duration, same starting and ending points in time;
- Adjacent: the events follow or precede one another, such that they share a starting or ending point in time;
- Overlapped: the events partially occur in the same time period, and no sharing of starting/ending point occurs;
- Before/after: the events occur one before/after the other, without sharing any time period or starting/ending point;
- Contained: one event contains the other;
- Started together: the events share the starting point;
- Finished together: the events share the ending point.

From these principles, TGRASS supports a simplified Allen’s topology [52], taking into account only the situations where the events have at least one shared point (Figure 8).

![Figure 8](image_url)

Figure 8. The Allen temporal topology schema, simplified by [52] and implemented in TGRASS. Reproduced with permission from Gebbert, S. and E. Pebesma, A temporal GIS for field based environmental modeling. *Env. Modelling & Software*; published by Elsevier, 2014.

In accordance with this theory, each boat passage recorded by the officers (each line of the input spreadsheet) is seen as an “event”. For each one, a vector map containing the route is produced. These vector lines are cumulated on the geometric vector grid over a user-specified period of time, $P$. The granularity, $g$, of the analysis must also be specified in order to compute the number of boats in each sub-period and to produce a number of maps equal to $P/g$. Each map has a timestamp with the starting and ending time in order to reference its temporality.

Figure 9 summarizes the process. In (A) the “boat passages” events are placed over a timeline and some examples of date and time information, extracted from the original data, are reported; (B) defines the period $P$ of interest and a granularity $g$ value (in this case, the period is from 1 to 25 January 2015 and the granularity specified is 5 days); (C) lists the
This model was implemented in TGRASS from the module t.vect.createRoutes.py, and allows us to perform both an instantaneous analysis (create a timestamped vector map reporting the instantaneous passage of boats on a geometric grid) and a period analysis, with a specified granularity. In the first case, the output will be a single timestamped map reporting the instantaneous traffic situation for a specific sémaphore. In the second case, the output will be a set of timestamped maps, following each other over a user-specified period and granularity. All the vector maps produced are saved into a GRASS database in order to speed up the geographical processing, and can be exported in different formats for a wide variety of purposes. For instance, it is possible to produce an animation of the traffic evolution and save it as a .gif file. The module was written in the Python language and involves most of the v.createRoutes.py routines; the graphical user interface is reported in Figure 10. The temporal information required for the period or moment parameter must be formatted as “dd:mm:yyyy hh:mm”, and the granularity expressed in minutes. The presence (or absence) of the flag “a” determines the production of the animation.
All the code written for this work (including both cleaning data and geoprocessing routines) has been released as Open Source software (license GNU GPL v.3). It can be freely downloaded from https://github.com/annalisapg/maritimeTraffic/tree/v.1.0 (accessed on 3 March 2021) and it is published in Zenodo (https://doi.org/10.5281/zenodo.4420051, accessed on 3 March 2021).

4.4. Storing and Displaying Data in INDIGEO SDI

In order to visualize these data, we used the scientific Spatial Data Infrastructure (SDI) INDIGEO [52] at the LETG (Littoral, Environnement, Télédétection, Géomatique) research unit. The functionalities of this SDI include cataloging, storage and distribution of geographic information produced by LETG and its partners. The SDI consists of a metadata catalog and a geo-referenced data server backed by a web portal with a viewer. The deployed solution is based on geOrchestra free tools (geonetwork, geoserver, openlayers...), an initiative of GéoBretagne (a French regional SDI, http://cms.geobretagne.fr/, accessed on 3 March 2021). INDIGEO is interoperable with other regional, national and scientific SDIs through the use of Open Geospatial Consortium standards (WMS, WFS, WCS). It also benefits from the development of an ergonomic and scalable viewer (GeoCMS) to meet the specific needs of science (series of temporal data, spatiotemporal data, etc.) and thus make them public in accordance with the principles of the European Directive INSPIRE (http://inspire.ec.europa.eu/, accessed on 3 March 2021).

Several INDIGEO functions are useful to our project: support for spatiotemporal data, an interactive and customizable way to query data, support for graphic time-series representation, a folder structure to save and share maps, and a user-friendly administration interface. The ability to navigate through the temporal dimension of a spatiotemporal dataset stored on a PostGIS database (Figure 11) and the visualization of statistical data through a chart module are other specificities.

Figure 11. Visualization of the output data stored in INDIGEO (https://portail.indigeo.fr/geocms/projects/cartahu, accessed the 3 March 2021).
All the data related to this project (raw data of sémaphores and data with cleaned routes) are deposited and published with a CC-BY license through the INDIGEO SDI, and this is available for download at the following address: https://doi.org/10.35110/59385ec4-a2ea-4e3c-89e0-e24a770b7358 (accessed on 3 March 2021).

5. Results and Discussion

5.1. Marine Traffic Mapping

Figure 12 shows an example of the output of the automatic route extraction procedure by the modality “moment” for the grid, and gates associated with the Saint-Mathieu sémaphore at 12 PM on 3 January 2011.

![Figure 12. Screenshot of a sample result in output from t vect.createRoutes.py for the “moment” mode.](image)

The routes are represented as proportional to the “npass” parameter. The gate vector file only contains information regarding the frequentation if the boat is stationed at the gate itself. Thus, if the observed vessel is stationed at the Le Conquet gate, the routine will count +1 on the parameter “npass” of this gate (for example, at the point in time depicted in Figure 12, no boats are stationary); otherwise, if the vessel is on its way from Le Conquet to “The Islands”, “npass” will be incremented in the vector grid. It is thus possible to differentiate vessels in transit from those that are stationary. For the “period” mode, the output is represented by several timestamped pairs of vector maps (a line map of the routes and a point map of the gates), cumulating traffic over a specified period of time, following a user-defined granularity. These data can be exported from GRASS in various vector formats (e.g., shapefile or kml) in order to be analyzed by alternative means. For example, they can be archived into a PostGIS database or imported into specific applications to be visualized in a 3D environment.

Figure 13 shows the traffic evolution over four months of the year 2011, each one representative of a season, for the Saint-Mathieu sémaphore. This output shows the most frequented routes during the year and, consequently, it allows us to individualize the relative pressure on the coastal environment. Principal routes differ between different times of the year, such as Le Conquet-Westward, Brest Bay to Le Conquet, Le Conquet to “The Islands” or to the North.
Figure 13. Screenshots of a sample result in output from t.vect.createRoutes.py for the “period” mode at a monthly rate. Samples for each season of 2011 and the Saint-Mathieu sémaphore are reported. A complete version of this image (with all the 12 months for the year 2011) is available in the supplementary file.

In order to correctly interpret these maps, it should be remembered that the original data are semantic; so, taking into account the potential visibility of the sémaphore observers’ annotations, it would not be surprising if the Le Conquet-Westward route meant “high seas” when visibility is perfect, or “The Islands” when visibility is poor.
5.2. Quantitative Analysis

The integration of these outputs into INDIGEO makes it possible to build simple graphical representations, which provide relevant support for their quantitative analysis. For example, two-thirds of the total traffic for the Saint-Mathieu sémaphore and the year 2011 (total $n = 12,231$) is related to fishing, shipping and passenger vessel movements (Figure 14a). Monthly analyses of the traffic could also be performed (Figure 14b), which confirmed the high temporal variability of maritime flux observed.

![Figure 14. Characterization of the boat traffic recorded by Saint-Mathieu sémaphore for 2011 from the clean input data of t.vect.createRoutes.py. (a): type of boats observed in 2011 (% annual traffic); (b): monthly traffic by type of maritime activities (2011); (c): weekly traffic observed by Saint-Mathieu sémaphore (2011). Legend for (a) and (b): in blue = fishing boats, in orange = recreational motor boats, in yellow = shipping/passenger boats, in green = sailing boats, in brown = other type.](image)

By comparing Figure 14b with the maps of Figure 13 it is possible to qualify the boat traffic more accurately. The monthly variability of the traffic can be directly related to recreational boating flux, which shows important seasonal variations (average 304 boats/month with a standard deviation $s = 87$) and is concentrated between April and August (81% of traffic). The shipping and fishing traffic is more stable over the whole year as a whole (average $= 533$ boats/month with a standard deviation $s = 120$), except in December, January and February when the navigation conditions are rougher. Finally, the weekly distribution of boat traffic (Figure 14c) shows a large variability, linked to both the socio-economic context and weather conditions. For instance, in spring (weeks 12 to 25) and autumn (weeks 37 to 51) 2011, Brittany had particularly unstable weather, which could explain such variability.

5.3. Discussion

In summary, the model we developed firstly reads the raw data and performs the cleaning procedures through cleanDataByDicts.py; in a second step, the temporal information is added to the spreadsheet by the addTime.py tool; and, finally, the data are processed through t.vect.createRoutes.py. The output of these processes is either a couple of time stamped vector maps, if t.vect.createRoutes.py is used in “moment” mode, or a set of sequential timestamped vector maps, if the “period” mode is set. In this case, a temporal granularity must be specified in order to assign a temporal resolution. The vector maps
produced in this way show the aggregation of traffic fluxes in a specific moment or over a
period of time. Each map contains quantitative information about the traffic through the
parameter “npass” calculated both for the line vector file of routes, and the point vector file
of gates, to distinguish boats in transit from stationary ones.

At the present stage of our work, the procedure shows encouraging results: reasonable
initial language standardization, cleaning procedure and representation of data over space
and time. One quite innovative aspect of the methodology used is the spatialization of se-
monic traffic data. In fact, this spatialization technique can be re-used each time a “moving
object” needs to be spatialized, and can be refined by applying more precise spatialization
rules, as mentioned above. Moreover, it is possible not only to create a representation
of the traffic, but also to export timestamped data in order to obtain a complete traffic
dataset ready to be analyzed and processed for any further purpose. Therefore, considering
that there are semaphores positioned along the entire French coastline and that these data
are readily available, an analysis at national level could now be pursued. Additionally,
the results of this study could be compared with the data from other monitoring systems,
such as AIS or GPS data, with the general aim to achieve more accurately than now the
“real” amount of leisure boating in the total marine traffic. In fact, in current conditions,
the contribution of leisure boating to the marine traffic, and the pressure it exerts on the
coastal environment entity, could only be summarily estimated since not all recreational
craft (especially the small ones) carry an AIS or GPS on board.

The methodology has now been widely tested over five different datasets (Saint-
Mathieu, years: 2011, 2012, and 2013; Toulinguet, 2011; Cap La Chèvre, 2011). The total
amount of traffic observed and processed during this period was about 44,000 passages
of vessels in the Iroise Sea. In addition, some simulations were performed using different
configurations, for example, short simulations on one week of traffic, with a temporal
granularity going from 30 min to one day, or long period simulations on one year, with a
temporal granularity of one month. Moreover, various “instantaneous” situation maps and
animations have been produced. Some statistical analyses have been conducted in order to
qualify and quantify the traffic over the observed period. These data will be presented in
detail in a later publication concentrating on this theme.

5.4. Challenges and Possible Improvements

Three major challenges could be addressed from this work. The first is relative to the
time of calculation, which is a little long: half an hour is required to process about 500 boats
on a standard workstation (8 Gb RAM and 8 cores at 2.2 GHz). These long calculation times
are not due to the grid resolution, since the operation of adding a boat passage to the grid
is quite rapid, but are mainly the result of the algorithm that searches for the shortest path
for each boat, as this takes some time to compute the route. To improve this step, new al-
gorithms with better performance could be studied, and a multiprocessing or distributed
computing solution, for example, via the OpenMOLE platform [53], could be explored.

The second limitation does not yet, to our knowledge, have an easy or clear solution.
It concerns the splitting of the same routes into more than one path of equi-probability.
Indeed, since the grid is regular, the same semantic route can be represented by more
than one synthetic path, causing a division, rather than a grouping, of traffic fluxes. In a
temporal fixed environment, this could be easily solved by computing the shortest path
for the same semantic route once and then repeating it at each occurrence without recalculating
it. In a temporally-domained environment, however, particularly because of the tide
effect, paths change over time, so the solution suggested above is not viable. Trying to
provide an answer to this problem, the ongoing work concerns the introduction into the
analysis of various constraints in order to limit equi-probability routes as much as possible.
For example, we have excluded zones where the circulation is forbidden or dangerous,
so the navigable space is reduced and the number of routes from origin to destination
gates is also reduced. In order to circumscribe the navigable areas, a bathymetry map with
5 m resolution has been built for the Iroise Sea and the tide recordings over time have
been collected for its ports. Tide level data can be interpolated to obtain a raster surface of water height, with the aim of introducing further spatial restrictions related to vessel type. Additionally, as small leisure vessels have a relatively high freedom of movement, we are currently focusing on the integration of the GIS methodology described above with Multi-Agent modelling [54]. This kind of method would assign “artificial intelligence” settings to each boat to improve the representation of individual behavior, making possible interactions with its natural or human environment (agents).

The third challenge stems from the fact that the observation of neighboring sémaphore zones is generally designed to overlap. This allows people to follow a boat from sémaphore to sémaphore, which is essential for maritime or homeland security purposes. However, this makes it problematic to reconstruct boat routes that pass from one area to another because the observation is not necessarily recorded at the same time (concurrency problem), nor is the description of the boat necessarily identical between sémaphores (semantic variability). To trace the route of the same boat between two adjacent (or more) sémaphores, a method of automatic recognition could be outlined as follows, by defining some theoretical steps:

- Combine the available data from all sémaphores for the same period;
- Recognize the boats from one sémaphore to another using their name, route, date and the time of observation;
- Rebuild the input file to unify records of vessels identified in this way.

These steps would be a basic requirement to prepare the data by eliminating any double traces and rebuilding the full itinerary of each boat. After this preparatory phase, a spatially extended geometric network should be built in order to compile the traffic observed by all sémaphores concerned and, ultimately, to reconstruct the individual routes of each boat.

6. Conclusions

This work is original in several ways. First, it is based on original, unexploited data collected from sémaphore monitoring of marine traffic. Despite their semantic nature, these data are relevant for describing synthetic marine traffic flux in the spatiotemporal dimension, especially because the observation made by the network of sémaphores is continuous in space (French coast) and in time (24/24 h, 7/7 d). In fact, the results obtained from these data and the approach we propose are able to provide information for a most comprehensive summary on maritime traffic. Indeed, much of this traffic is composed of boats that are not currently observed by any other device than the visual and radio observations made by the sémaphore network.

The production of such information could open at least two interesting prospects for public policy support, including the identification of areas of potential conflict between several kinds of marine activities, and the assessment of the cumulative pressure generated by marine activities.

Although such conflicts and pressures are not widespread in the coastal area, the analysis of sémaphore observation data can provide a useful contribution to identifying the most vulnerable areas in order to anticipate, avoid or control the potential impacts of marine activities, especially in marine protected areas.

From a strictly methodological point of view, the method presented is also quite original. In fact, a natural temporal framework underlies all the analysis, and there is a slight but perceptible difference between artificially considering a specific situation placed in a temporal moment and the situation where the map is time stamped and treated as a (X, Y, Z, T) fact, and is thus geo- and temporally referenced in an univocal manner.
Supplementary Materials: The following are available online at https://www.mdpi.com/2077-1312/9/3/294/s1. The translation in English of Table 1: A complete version of Figure 13 including the representation of the routes and gates frequentation for all the 12 months of the year 2011 at the Saint-Mathieu sémaphore.

Author Contributions: Conceptualization and methodology, A.M., I.L.B., I.P. and M.R.; software, A.M. and M.R.; data curation, A.M., I.L.B., I.P. and M.R.; all the authors contributed to writing (original draft preparation, review and editing); visualization, A.M., M.R. and I.P.; supervision, I.L.B. and I.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fondation de France, project CARTAHU (number 1116) in the frame of the financement line: “Quels littoraux pour demain” (2011–2016). This work also benefited from a State subsidy managed by the Agence Nationale de la Recherche (ANR, National Agency for Research) within the framework of the “Investments for the Future” program, reference ANR-10-LABX-19-01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable. Our study does not concerns human beings but only boats anonymized and not identifiable.

Data Availability Statement: The data (raw and elaborated) presented in this study are openly available in the INDIGEO portal (http://indigeo.fr/, accessed on 3 March 2021) at https://doi.org/10.35110/59385ec4-a2ea-4e3c-89e0-e24a770b7358 (accessed the on 3 March 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Douvere, F. The importance of marine spatial planning in advancing ecosystem-based sea use management. Mar. Policy 2008, 32, 762–771. [CrossRef]
2. Wahl, T.; Høye, G.K.; Lyngvi, A.; Narheim, B.T. New possible roles of small satellites in maritime surveillance. Acta Astronaut. 2005, 56, 273–277. [CrossRef]
3. Greidanus, H.; Kourti, N. Findings of the DECLIMS project—Detection and classification of marine traffic from space. In Proceedings of the Advances in SAR Oceanography from Envisat and ERS Missions, Frascati, Italy, 23 January 2006; ESA Publications Division: Noordwijk, The Netherlands, 2006. Available online: https://earth.esa.int/workshops/seasar2006/proceedings/papers/s5_12_gre.pdf (accessed on 8 January 2021).
4. Reggiannini, M.; Bedini, L. Multi-Sensor Satellite Data Processing for Marine Traffic Understanding. Electronics 2019, 8, 152. [CrossRef]
5. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. IEEE Access 2020, 8, 120234–120254. [CrossRef]
6. Ray, C.; Devogèle, T.; Noyon, V.; Petit, M.; Fournier, S.; Claramunt, C. GIS technology for Maritime Traffic Systems. In European Research Consortium for Informatics and Mathematics News: Special Theme on Traffic Planning and Logistics; Kunts, P., Ed.; ERCIM: Sophia-Antipolis, France, 2007; Volume 68, pp. 41–42. ISBN 0926-4981.
7. Høye, G.K.; Eriksen, T.; Meland, B.J.; Narheim, B.T. Space-based AIS for global maritime traffic monitoring. Acta Astronaut. 2008, 62, 240–245. [CrossRef]
8. Le Guyader, D. Modélisation des Activités Humaines en mer Côtière. Ph.D. Thesis, EDM, IUEM—Université de Bretagne Occidentale, Brest, France, 2012. Available online: https://tel.archives-ouvertes.fr/tel-00717420v2/document (accessed on 11 January 2021).
9. Ferrà, C.; Tassetti, A.N.; Grati, F.; Pellini, G.; Polidori, P.; Scarcella, G.; Fabi, G. Mapping change in bottom trawling activity in the Mediterranean Sea through AIS data. Mar. Policy 2018, 94, 275–281. [CrossRef]
10. Snidarô, L.; Visentini, I.; Bryan, K. Fusing uncertain knowledge and evidence for maritime situational awareness via Markov Logic Networks. Inf. Fusion 2015, 21, 159–172. [CrossRef]
11. Chen, Z.; Xue, J.; Wu, C.; Qin, L.Q.; Liu, L.; Cheng, X. Classification of vessel motion pattern in inland waterways based on Automatic Identification System. Ocean Eng. 2018, 161, 69–76. [CrossRef]
12. Furlan, E.; Slanzi, D.; Torresan, S.; Critto, A.; Marcomini, A. Multi-scenario analysis in the Adriatic Sea: A GIS-based Bayesian network to support maritime spatial planning. Sci. Total Environ. 2020, 703, 134972. [CrossRef]
13. Claramunt, C.; Devogèle, T.; Fournier, S.; Noyon, V.; Petit, M.; Ray, C. Maritime GIS: From monitoring to simulation systems. In Information Fusion and Geographic Information Systems, Lecture Notes in Geoinformation and Cartography; Popovich, V.V., Schrenk, M., Korolenko, K.V., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 34–44. [CrossRef]
14. Siljander, M.; Venäläinen, E.; Goerlandt, F.; Pellikka, P. GIS-based cost distance modelling to support strategic maritime search and rescue planning: A feasibility study. Appl. Geogr. 2015, 57, 54–70. [CrossRef]
15. Huet, T.; Osman, T.; Ray, C. Modelling traffic navigation network with a multi-agent platform. In Proceedings of the 17th European Simulation Multiconference (ESM2003), Nottingham, UK, 9–11 June 2003; SCS-European Publishing House: Delft, The Netherlands, 2003; pp. 111–117, ISBN 3936150257.

16. Xiao, F.; Ligteringen, H.; Van Gulijk, C.; Ale, B.J.M. Artificial force fields for multi-agent simulations of maritime traffic and risk estimation. In Proceedings of the 11th International Probabilistic Safety Assessment and Management Conference (PSAM 11) & The Annual European Safety and Reliability Conference Scandic Marina Congress Center (PSAM 11), Helsinki, Finland, 25–29 June 2012; Available online: http://resolver.tudelft.nl/uuid:0e68246e-f259-4b70-81a6-38d327bab35 (accessed on 8 January 2021).

17. Xiao, Z.; Fu, X.; Zhang, L.; Ponnambalam, L.; Goh, R.S.M. Data-driven multi-agent system for maritime traffic safety management. In Proceedings of the IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6. [CrossRef]

18. Singh, A.J.; Kumar, A.; Lau, H.C. Hierarchical Multiagent Reinforcement Learning for Maritime Traffic Management. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, Auckland, New Zealand, 5 May 2020; International Foundation for Autonomous Agents and MultiAgent Systems: Richland, SC, USA, 2020; pp. 1278–1286. [CrossRef]

19. Goerlandt, F.; Kujala, P. Traffic simulation based on ship collision probability modeling. Reliab. Eng. Syst. Saf. 2011, 96, 91–107. [CrossRef]

20. Hänninen, M. Bayesian networks for maritime traffic accident prevention: Benefits and challenges. Accid. Anal. Prev. 2014, 73, 305–312. [CrossRef][PubMed]

21. Xuan, S.; Xi, Y.; Huang, C.; Hu, S.; Zhang, L. Spatial Markov Chain simulation model of accident risk for marine traffic. In Proceedings of the 4th International Conference on Transportation Information and Safety (ICTIS), Banff, AB, Canada, 8–10 August 2017; pp. 664–669. [CrossRef]

22. Kazemi, S.; Abghari, S.; Lavesson, N.; Johnson, H.; Ryman, P. Open data for anomaly detection in maritime surveillance. Expert Syst. Appl. 2013, 40, 5719–5729. [CrossRef]

23. Osekowska, E.; Johnson, H.; Carlsson, B. Grid size optimization for potential field based maritime anomaly detection. Transp. Res. Procedia 2014, 3, 720–729. [CrossRef]

24. Ferrà, C.; Tassetti, A.N.; Armelloni, E.N.; Galdelli, A.; Scarcella, G.; Fabi, G. Using AIS to Attempt a Quantitative Evaluation of Unobserved Trawling Activity in the Mediterranean Sea. Front. Mar. Sci. 2020, 7, 1036. [CrossRef]

25. Hauser, D.D.; Laidre, K.L.; Stern, H.L. Vulnerability of Arctic marine mammals to vessel traffic in the increasingly ice-free Northwest Passage and Northern Sea Route. Proc. Natl. Acad. Sci. USA 2018, 115, 7617–7622. [CrossRef]

26. Eronat, A.H.; Bengil, F.; Nezer, G. Shipping and ship recycling related oil pollution detection in Çandarlı Bay (Turkey) using satellite monitoring. Ocean Eng. 2019, 187, 106157. [CrossRef]

27. De Vivero, J.L.S.; Mateos, J.C.R. The Spanish approach to marine spatial planning. Marine strategy framework directive vs. EU integrated maritime policy. Mar. Policy 2012, 36, 18–27. [CrossRef]
40. Lecouturier, Y.; Lucas, S. Les sémaphores. Avènement d’un réseau moderne/Veiller la mer aujourd’hui. *Chass. Marée* **1988**, *35*, 2–17.

41. Michoux, L. Spationav, pierre angulaire de la protection et de la sauvegarde maritime. *REE Rev. L’électricité L’électronique* **2007**, *6–7*, 81–84.

42. Lancaster, D.; Dearden, P.; Haggarty, D.R.; Volpe, J.P.; Ban, N.C. Effectiveness of shore-based remote camera monitoring for quantifying recreational fisher compliance in marine conservation areas. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2017**, *27*, 804–813. [CrossRef]

43. Smallwood, C.B.; Beckley, L.E.; Moore, S.A.; Kobryn, H.T. Assessing patterns of recreational use in large marine parks: A case study from Ningaloo Marine Park, Australia. *Ocean Coast. Manag.* **2011**, *54*, 330–340. [CrossRef]

44. Gohn, C. _Mobiliser les Savoir-Faire Pour L’analyse Spatiale et Dynamique des Activités et des Flux en Mer Côtière; Application à la Mer d’Iroise, Rapport de Stage, Master 1 SIGAT, Université de Rennes: Rennes, France, 2013; Available online: https://www-iuem.univ-brest.fr/pops/attachments/192 (accessed on 11 January 2021).

45. Neteler, M.; Mitasova, H. *Open Source GIS: A GRASS GIS Approach*, 3rd ed.; Springer: New York, NY, USA, 2008. [CrossRef]

46. Biggs, N.L.; Wilson, R.J.; Lloy, E.K. *Graph Theory 1736–1936*; Clarendon Press: Oxford, UK, 1976; ISBN 978-0198539162.

47. Brosset, D.; Claramunt, C. An experimental ant colony approach for the geolocation of verbal route descriptions. *Knowl. Based Syst.* **2011**, *24*, 484–491. [CrossRef]

48. Belouaer, L.; Brosset, D.; Claramunt, C. Modeling spatial knowledge from verbal descriptions. In *Spatial Information Theory*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 338–357. [CrossRef]

49. Chakrabarty, K.; Iyengar, S.S.; Qi, H.; Cho, E. Grid coverage for surveillance and target location in distributed sensor networks. *Comput. IEEE Trans.* **2002**, *51*, 1448–1453. [CrossRef]

50. Gebbert, S.; Pebesma, E. A temporal GIS for field based environmental modeling. *Environ. Model. Softw.* **2014**, *53*, 1–12. [CrossRef]

51. Allen, J.F. Maintaining knowledge about temporal intervals. *Commun. ACM* **1983**, *11*, 832–843. [CrossRef]

52. Rouan, M. A Spatial Data Infrastructure dedicated to scientific research and observation of the coastal environment. In Proceedings of the IODE International Coastal Atlas Network (ICAN) Workshop 7: Coastal Web Atlases—Supporting Ecosystem Based Management, Cape Town, South Africa, 19–20 April 2015; Available online: https://hal.archives-ouvertes.fr/hal-01159108/file/INDIGEO_ICAN2015.pdf (accessed on 11 January 2021).

53. Reuillon, R.; Leclaire, M.; Rey-Coyrehourcq, S. OpenMOLE, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Gener. Comput. Syst.* **2013**, *29*, 1981–1990. [CrossRef]

54. Schüle, M.; Herrler, R.; Klügl, F. Coupling gis and multi-agent simulation—towards infrastructure for realistic simulation. In *Multiagent System Technologies*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 228–242. [CrossRef]