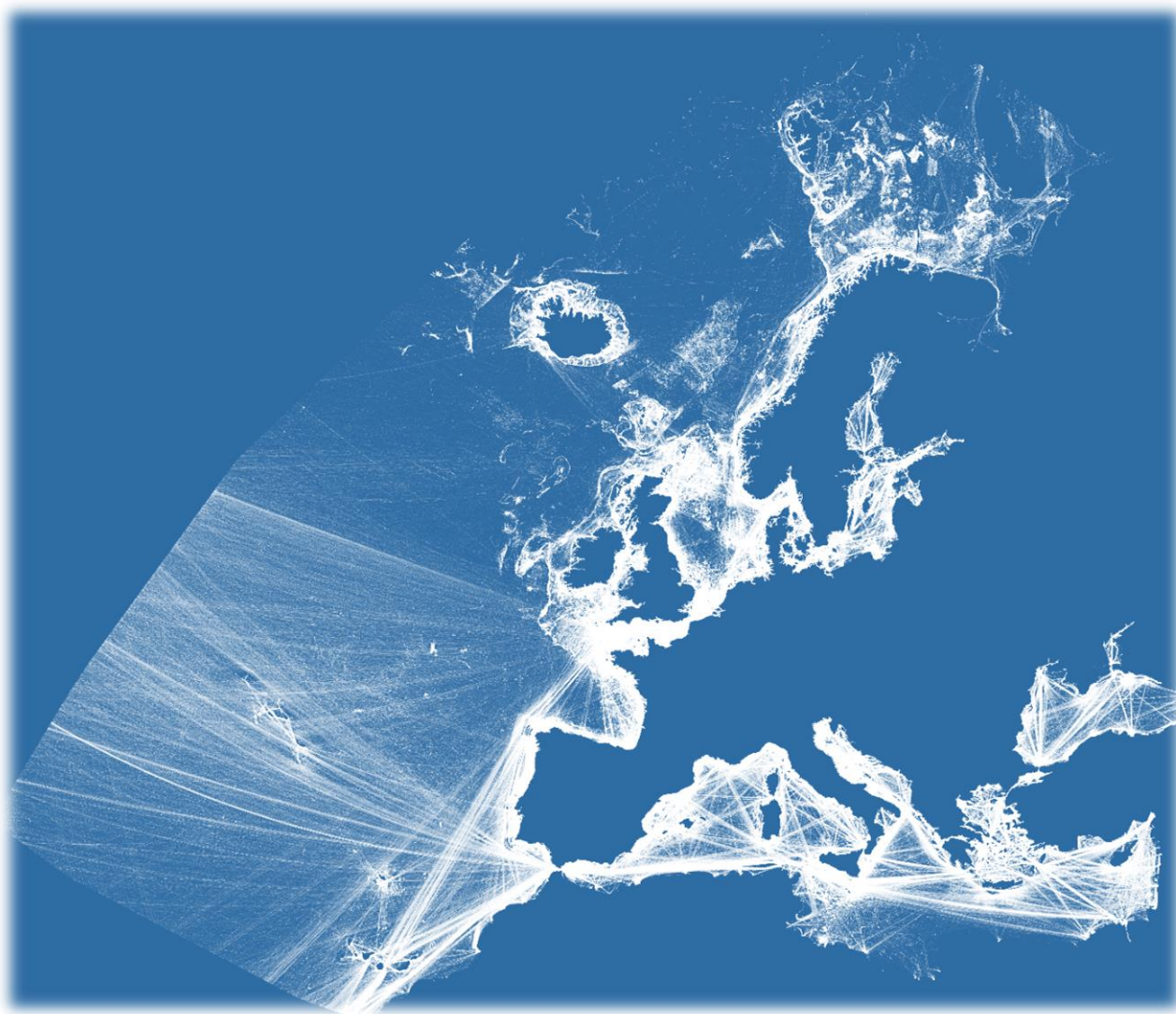


EU Vessel density map Detailed method



v 1.5

March 2019

Contents

Abstract	i
Authors	ii
Acknowledgements.....	ii
Disclaimer.....	ii
List of acronyms	iii
1 Introduction.....	1
2 What is a vessel density map?.....	2
3 What is density?	3
4 How to calculate density.....	5
5 The method	7
5.1 Defining the Area of Interest.....	7
5.2 Creating a grid.....	9
5.3 Setting up a machine	12
5.4 Acquiring and cleaning the AIS data	12
5.5 Importing the AIS data into a database.....	15
5.6 Creating points.....	17
5.7 Creating lines	19
5.8 Calculating density	21
5.9 Creating maps.....	22
6 Publishing the maps.....	23
6.1 Converting to different density units.....	23
7 What's next	23
8 References	25



HUMAN ACTIVITIES
Making use of our oceans

Abstract

The EU-funded “European Marine Observation and Data Network” (EMODnet) aims to make hitherto dispersed European marine data available via a single portal. Under its Human Activities portal, a newly introduced product is vessel density maps of European waters. These maps have been made from ship reporting data of the Automatic Identification System (AIS), collected by coastal stations and satellites. The report describes the method how the vessel density maps were made from the AIS data. The resulting maps provide the total ship presence time on a 1 km grid that follows EEA / Inspire standards, for every month of 2017, separately for 14 different ship categories. The total ship presence times are easily converted to average instantaneous number of ships per square km.

Authors

Luigi FALCO (Cogea)

Alessandro PITITTO (Cogea)

William ADNAMS (Lovell Johns)

Nick EARWAKER (Lovell Johns)

Harm GREIDANUS (Joint Research Centre, European Commission)

Acknowledgements

The EMODnet Human Activities team acknowledge with grateful thanks the input, feedback and expertise provided by the experts who were consulted during the development of the method. Special thanks go to **Maurizio Gibin** from the Joint Research Centre of the European Commission, who spent his time and effort to brainstorm with us and provide invaluable scientific advice; **Manuel Frias** and **Florent Nicolas** from HELCOM, who did a similar exercise in the past and were always ready to share precious insights and comfort us when things were getting difficult; **Johan Pascal** who unveiled to us the mysteries of Linux Shell and helped clean the data.

Without their advice, our work would have been overwhelmingly more difficult and most certainly less accurate.

Disclaimer

The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund.

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the EASME or of the European Commission. Neither the EASME, nor the European Commission, guarantee the accuracy of the data included in this study. Neither the EASME, the European Commission nor any person acting on the EASME's or on the European Commission's behalf may be held responsible for the use which may be made of the information.

List of acronyms

AIS	Automatic Identification System
Aol	Area of Interest
CLS	Collecte Localisation Satellites (private company)
COG	Course Over Ground
CSV	Comma Separated Variable
CPU	Central Processing Unit
EASME	European Agency for Small and Medium Enterprises
EC	European Commission
EEA	European Environment Agency
EMODnet	European Marine Observation and Data Network
EPSG	European Petroleum Survey Group
ETA	Estimated Time of Arrival
ETRS89	European Terrestrial Reference System 1989
EU	European Union
GB	GigaByte
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information System
GISCO	Geographical Information System of the Commission
HELCOM	Helsinki Commission
IMO	International Maritime Organisation
JRC	Joint Research Centre
LAEA	Lambert Azimuthal Equal Area (projection)
MB	MegaByte
MMSI	Maritime Mobile Service Identity
MSFD	Marine Strategy Framework Directive
NMEA	National Marine Electronics Association
RAM	Random Access Memory
ROT	Rate Of Turn
SOG	Speed Over Ground
SOLAS	Safety of Life at Sea (International Convention)
SQL	Structured Query Language
SSD	Solid State Drive
TIFF	Tagged Image File Format
VTs	Vessel Traffic Service
WMS	Web Map Service

1 Introduction

EMODnet Human Activities aims to facilitate access to existing data on activities carried out in EU waters, by building a single entry point for geographic information on human uses of the ocean. The portal makes available information such as geographical position and spatial extent of a series of activities related to the ocean, their temporal variation, time when data was provided, and attributes to indicate the intensity of each activity. The data are aggregated and presented so as to protect personal privacy and commercially-sensitive information. The data also include a time interval, so that historic as well as current activities can be represented.

In March 2017, EMODnet Human Activities was mandated to create vessel density maps of EU waters showing the average number of vessels of certain type (cargo, passenger, fishing etc.) in a given period within a grid cell. Vessel density maps were by far the most requested GIS data product by EMODnet Human Activities users, according to a survey carried out in 2016. The maps went live on EMODnet Human activities (www.emodnet-humanactivities.eu) for visualisation and download on 11 March 2019.

The purpose of this document is to outline the full method that led to the creation of the maps now available online.

For further questions or clarifications, please contact the EMODnet Human Activities team through the online contact form (<http://www.emodnet-humanactivities.eu/contact-us.php>) or by live chat.

2 What is a vessel density map?

A vessel density map is a data product that shows the distribution of ships (i.e., of maritime traffic), based on the instantaneous number of vessels per unit area, such as a square kilometre, a square degree, etc.

Vessel density maps are often generated starting from ship positions retrieved from the Automatic Identification System (AIS). AIS is an automatic ship transponder system used onboard certain vessels and is used by vessel traffic services (VTS). The system was conceived to assist vessels' watchstanding officers and allow maritime authorities to track and monitor vessel movements for purposes such as collision avoidance, maritime security, aid to navigation, search and rescue etc. After its introduction it has found further uses, e.g. in fishing fleet monitoring and control and maritime traffic studies – including production of vessel density maps. Vessels fitted with AIS transponders can be tracked by AIS base stations located along coast lines or, when out of range of terrestrial networks, through a growing number of satellites that are fitted with special AIS receivers which are capable of deconflicting a large number of signatures.

Regulation 19 of the International Convention for the Safety of Life at Sea (SOLAS) sets out navigational equipment to be carried on board ships, according to ship type. In 2000, the International Maritime Organisation adopted a new requirement (as part of a revised new chapter V) for all ships to carry AISs capable of providing information about the ship to other ships and to coastal authorities automatically.

The regulation requires AIS to be fitted aboard all ships of 300 gross tonnage and upwards engaged on international voyages, cargo ships of 500 gross tonnage and upwards not engaged on international voyages and all passenger ships irrespective of size. The requirement became effective for all these ships by 31 December 2004. The regulation requires that AIS shall (i) provide information – including the ship's identity, type, position, course, speed, navigational status and other safety-related information – automatically to appropriately equipped shore stations, other ships and aircraft; (ii) receive automatically such information from similarly fitted ships; (iii) monitor and track ships; (iv) exchange data with shore-based facilities.

In addition to the IMO carriage requirements for AIS that apply globally, individual countries may require further usage. For example, the EU requires that all fishing ships of 15 m and longer use AIS as of 31 May 2014¹.

However, AIS is also increasingly used on a voluntary basis by many other vessels not subject to Regulation 19 of the SOLAS Convention, including smaller leisure and fishing vessels. AIS equipment is divided into two classes – class A for mandatory use and B mostly for voluntary use – depending on the AIS transponder transmitting the AIS information. AIS information from class A transponders will always be prioritised and, thus, be shown to other ships in the area, whereas AIS information from class B transponders will not be shown until or if there is room on the AIS channel.

Vessel density maps have been around for quite a while. As of today, multiple sources make them available either for free or upon payment of a fee. Free vessel density maps tend to be developed in the framework of regional initiatives², whereas maps from commercial sources may be rather expensive. On the other hand, EMODnet's maps are entirely different. They cover an area corresponding to all EU waters (and some adjacent regions), they are 100% free and come with no restrictions to use. This means that the maps can be viewed, downloaded, processed, used and re-used for commercial and non-commercial purposes alike, as long as EMODnet is duly acknowledged as the data source.

Even more importantly, EMODnet's vessel density maps are based on a fully-documented bespoke method, which is different from the methods normally used to calculate shipping density.

¹ Council regulation (EC) No 1224/2009 of 20.11.2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy

² For instance, <http://maps.helcom.fi/website/AISexplorer/>

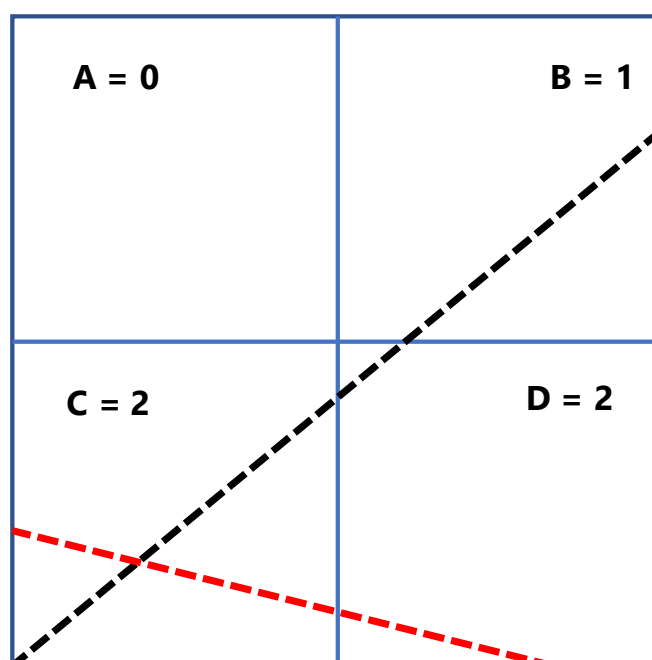
3 What is density?

Trivial as it may seem, the first question anybody needs to ask when resolving to make a vessel density map is: what is density?

One of the most recurring methods to make a vessel density map consists of generating ship track lines from AIS positions, intersect them with a grid and count the number of lines that cross each cell of the grid in a given time interval, usually a month. As no method is in principle better than the other, it all boils down to what one wants to represent and for which purposes.

Back in 2017, the EMODnet Human Activities team started wondering how to calculate vessel density. After a few brainstorming sessions, it emerged that, while the above-mentioned method was valid and straightforward, it did not produce the results we expected. See for instance the following picture:

Figure 1 – Calculating density based on number of ship tracks



The picture shows a grid made up of four cells. The black and the red lines represent two different ship tracks. If we counted the number of lines crossing each cell, we would have the following density values in a given time period:

A = 0 neither the black nor the red ship cross cell A, so density is 0.

B = 1 only the black ship crosses the cell, so density is 1.

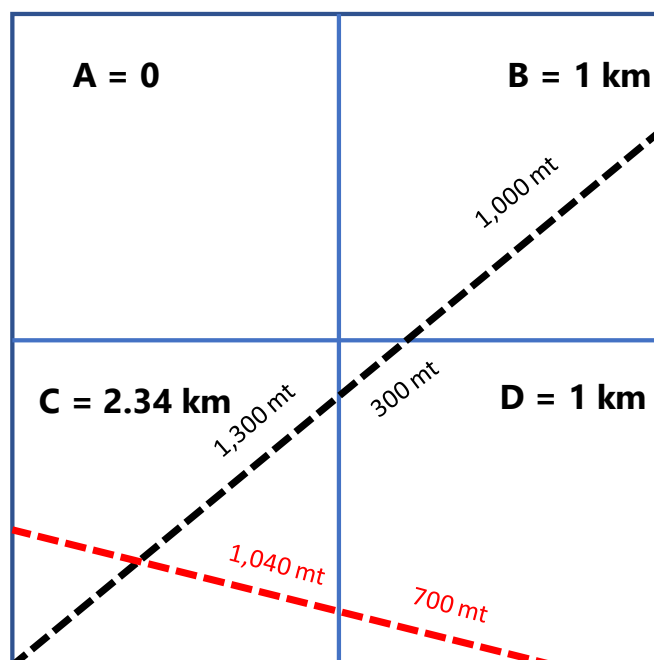
C = 2 both the black and the red ship cross the cell, so density is 2.

D = 2 both the black and the red ship cross the cell, so density is 2.

Cell C and D have the same value of density, even though it is visible to the naked eye that the length of both lines in cell D is considerably shorter. Thus, this method simply counts the number of tracks that cross each cell. Rather than measuring “ship density”, counting ship tracks that cross grid cells instead measures something that can be referred to as “ship crossing density”.

What immediately came to mind, then, was how to factor in line length in our measure of density. For instance, following the example above:

Figure 2 – Calculating density based on ship track length



Supposing our cell measures 1 km², if we took into account line length, the results would be completely different:

A = 0 no ship crosses the cell, so the density value is still 0.

B = the black ship crosses the cell and sail 1 km in it, so density is 1 km.

C = both the black and the red ship cross the cell, for a total length of 2.34 km.

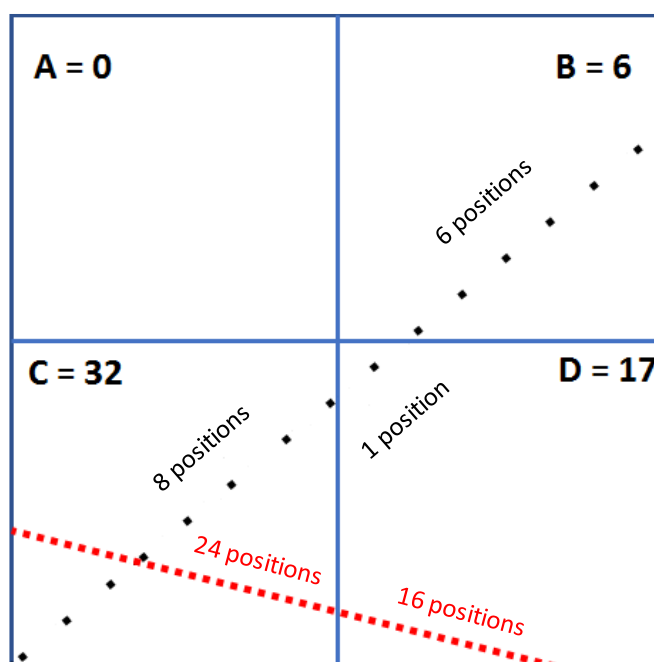
D = both the black and the red ship cross the cell, for a total length of 1 km.

Compared with the previous method, we can observe that the density value of cell C is now more than two times the value of cell D, whereas they were exactly the same before. Furthermore, cell B and cell D now have the same density value, even though two ships cross cell D, something which would have been impossible under the previous method.

However, this method does not produce the expected results either. This is because density should be a quantity used to describe the degree of concentration of countable objects (so, in this case ships) in a physical space (so, in this case a grid cell). Neither of the above-mentioned methods take this definition of density into account.

Vessel density should thus be interpreted as “how many vessels one expects to find within the area of reference during a given time period”. One of the main problems is that the actual number of vessels within that area is never constant over time, because ships always move around at different speeds. So, vessel density should be the expected number of ships in that area, i.e. the average number of ships in that area, taken from a large number of samples in time.

Figure 3 - Calculating density from number of AIS positions



Assuming that ship positions are taken at regular time samples, the figure above shows that the black ship is sailing faster than the red ship. Therefore, even though both ships sail the same distance as in the previous figures, over time the expected probability of finding the black ship in any given cell is comparatively lower. This example shows that speed should also be taken into account when calculating density over a time period. So, once again, using a different method produces entirely different results:

A = 0 no ship crosses the cell, so the density value is still 0.

B = the black ship is sampled 6 times, so the density value is 6.

C = the black ship is sampled 8 times and the red ship is sampled 24 times over the same time period, so the density value is 32.

D = the black ship is sampled 1 time and the red ship is sampled 16 times over the same time period, so the density value is 17.

Therefore, over a time period, the average number of ships in the reference area is different. Now, the density value of cell C is about 2 times higher than cell D, whereas it was 2.34 times higher in the former example. The difference is even more marked if one looks at cell B and cell D. While they had the same density value in the former example, now cell D has a density value nearly 3 times higher than cell B.

The conclusion is that the first two methods may be representations of vessel density, but they do not give results that measure the average number of vessels per unit area, so they are not true densities. The third method indeed leads to a vessel density map with numbers that are the average instantaneous number of vessels per unit area. The average is unavoidably taken over time, because the vessels are changing positions all the time.

4 How to calculate density

The third method presented in the previous paragraph is believed to produce a true measure of density, according to the definition proposed. However, it also comes with a strong limitation: counting ship positions in a cell over a time period can be done only if ships are sampled at regular intervals. In the real world, this seldom occurs. First of all, the default transmit rate for class A and class B is different. This problem could be solved easily, by making separate calculations for class A and class B messages. But there also is another challenge: in theory, the default transmit rate for class A messages is once every 2-10 seconds; in practice, when

one acquires an AIS dataset, they will soon realise that the messages are almost never received at the expected transmit rate. This may be due to several reasons: from simple transmission errors, to not being in range of any ground- or space-based receiver, the latter being the main reason for data gaps.

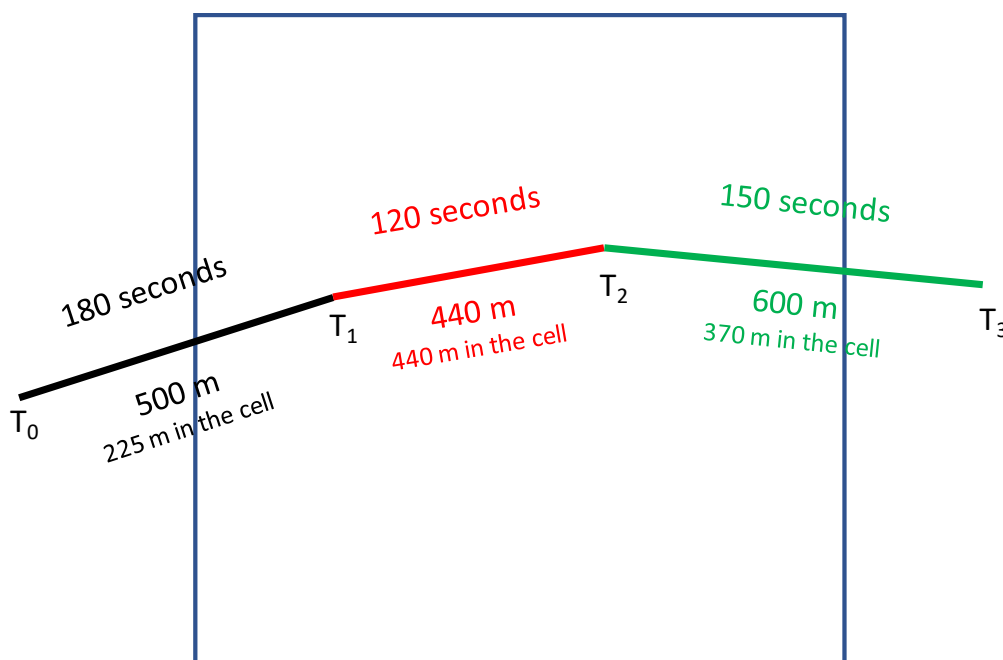
The set of AIS data acquired by EMODnet Human Activities confirmed that the actual time samples are often completely random, making it impossible to simply count ship positions in a grid cell to calculate density.

To overcome this issue, one solution is to interpolate the data by reconstructing ship routes from received position points and then sampling the tracks at regular time intervals. Depending on how it is done exactly, interpolation may be quite intensive in terms of computational power, because there might be billions of positions to process.

In light of the above, the method used for EMODnet's vessel density maps consists of reconstructed ship routes (lines) from the ship positions (points), by using a unique identifier of a ship. A line is created for every two consecutive received positions of a ship. For each line, length and duration (note that it is possible to calculate duration because each AIS message comes with a timestamp) are calculated and added as attributes. The lines obtained can then be intersected with a cell representing a unit area. Because each line has length and duration as attributes, it is possible to calculate how much time each ship spends in a given cell over a time period by intersecting the line records with the cell.

For the sake of simplicity, the figure below shows an example with a single ship sailing at different speeds in the same cell:

Figure 4 - Calculating density with EMODnet's method



Black line: $(225 \text{ m in the cell} / 500 \text{ m total length}) * 180 \text{ seconds} = 81 \text{ seconds in the cell}$

Red line: $(440 \text{ m in the cell} / 440 \text{ m total length}) * 120 \text{ seconds} = 120 \text{ seconds in the cell}$

Green line: $(370 \text{ m in the cell} / 600 \text{ m total length}) * 150 \text{ seconds} = 92.5 \text{ seconds}$

Total time in the cell: 293.5 seconds.

From this time value, one may calculate the number of expected positions (i.e. messages received) if the ships had been sampled at regular time intervals by dividing the total time spent in the cell by the hypothetical time sample. Assuming an interpolation sampling period of 30 seconds, one may divide the seconds spent by the ship in the cell by 30, that is $293.5 / 30 = 9.78$ which is the number of expected ship positions over a time period.

To get a proper density value (i.e. the average instantaneous number of ships per cell over a time period), one should divide the time spent by all ships in a cell by the total time in the period of reference. For instance, since EMODnet's vessel density maps are calculated over a month, supposing that the time spent by all ships in a given cell is 6,000 hours, the average instantaneous density would be obtained by:

$$6,000 \text{ hours} / (30 \text{ days} * 24 \text{ hours}) = 8.34 \text{ ships per cell}$$

Of course, one should correct for the effect of 28- and 31-day months. A density value per square km is obtained by dividing through the cell area.

To be noted that the vessel density maps that are available online on EMODnet Human Activities only have a time value (in hours), i.e. no normalisation was carried out. This is because, it is believed that in this way users are allowed to normalise at their preferred time rate, depending on their needs. Whatever the time factor used for normalisation, the relative values between the cells will remain the same.

The following sections provide a detailed description of the various steps of the method used to calculate the vessel density maps, with density expressed in hours per square kilometre per month and broken down by ship type.

5 The method

The method developed to create vessel density maps based on the considerations outlined in the previous sections is made up of the following steps:

1. Defining the Area of Interest (Aoi)
2. Creating a grid
3. Setting up a machine
4. Acquiring and cleaning the AIS data
5. Importing the AIS data in a database
6. Preparing the data for the calculations
7. Creating points
8. Creating lines
9. Calculating density
10. Creating maps

5.1 Defining the Area of Interest

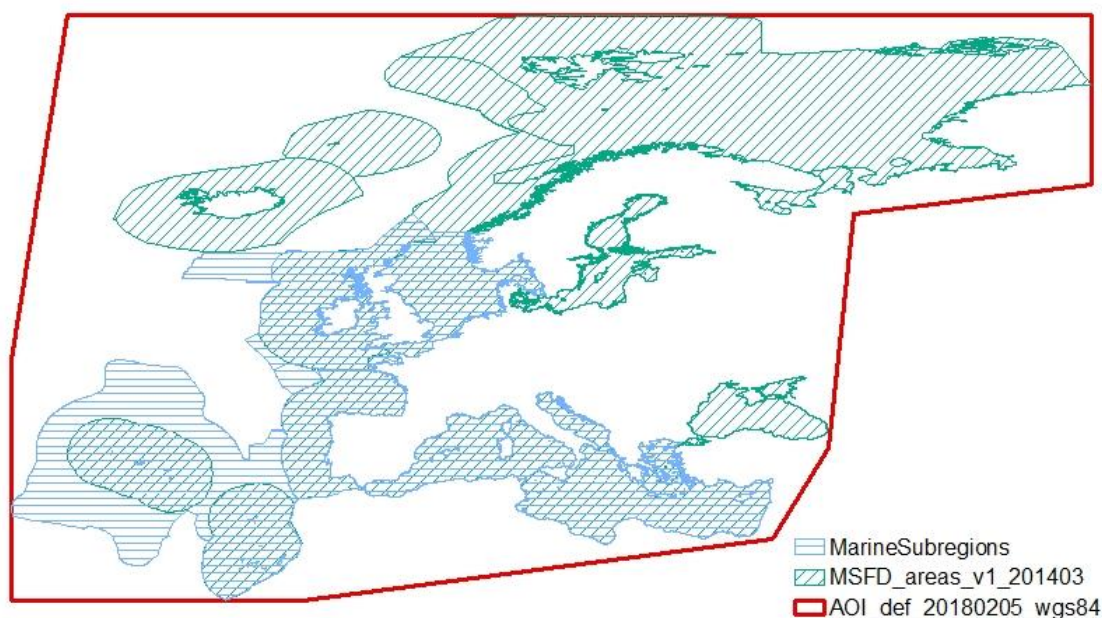
Before any other operation, one should define the desired Area of Interest (Aoi), i.e. the area for which vessel density will be calculated. EMODnet Human Activities was mandated to create vessel density map for all EU waters, so we started from the marine regions and subregions map layers³, which provide information about the geographic boundaries of the areas listed under Article 4 of the Marine Strategy Framework Directive (MSFD)⁴.

We defined the Aoi (the red polygon in the image below) according to the spatial extent of the MSFD marine subregions shapefile and according to the spatial extent of the MSFD_areas_v1_201403 layer.

³ <https://www.eea.europa.eu/data-and-maps/data/msfd-regions-and-subregions-1#tab-gis-data>

⁴ Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy.

Figure 5 – Area of Interest



The polygon in the picture above, however, also includes land areas. We obtained the actual Aol by intersecting the Aol polygon layer with the polygons of the European coastline⁵ and the polygons of the world countries shapefile⁶ at the lowest scale (1:1 million) available on GISCO. It was necessary to complete the EEA coastline polygon by adding the missing land of Greenland and remote Russian regions from the GISCO dataset.

Figure 6 – Area of Interest and coastline

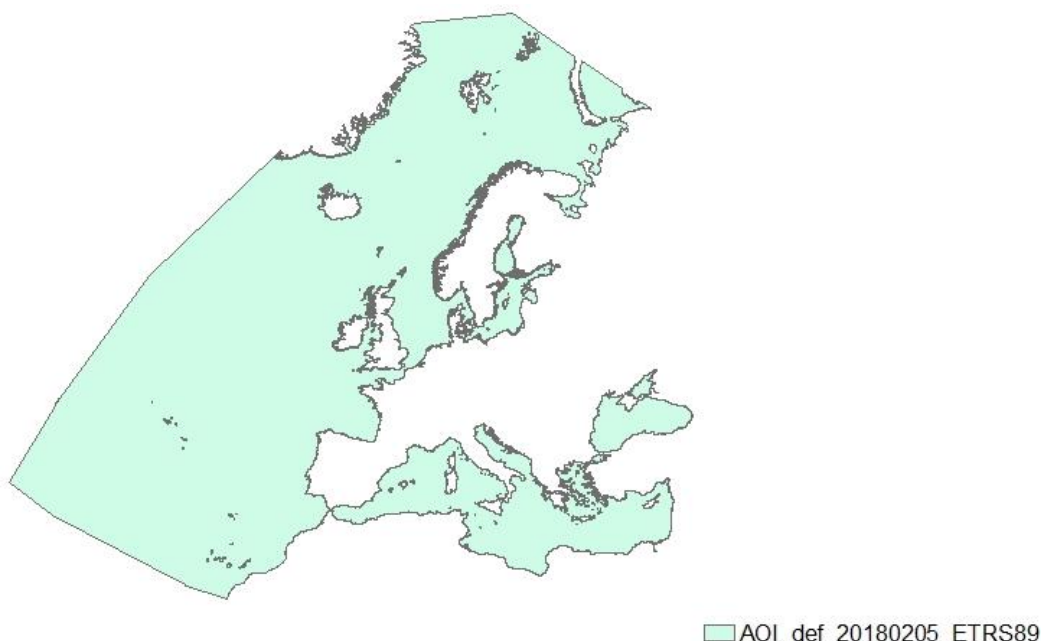


We re-projected all layers into the ETRS89-LAEA coordinate system and we erased land from the whole Aol. The final result is shown in the picture below:

⁵ <https://www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis-1>

⁶ <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries>

Figure 7 – Area of Interest without land



The final Aol covers all EU waters and does not contain land areas. This is necessary to define the grid in the following step, as well as to speed up data processing.

5.2 Creating a grid

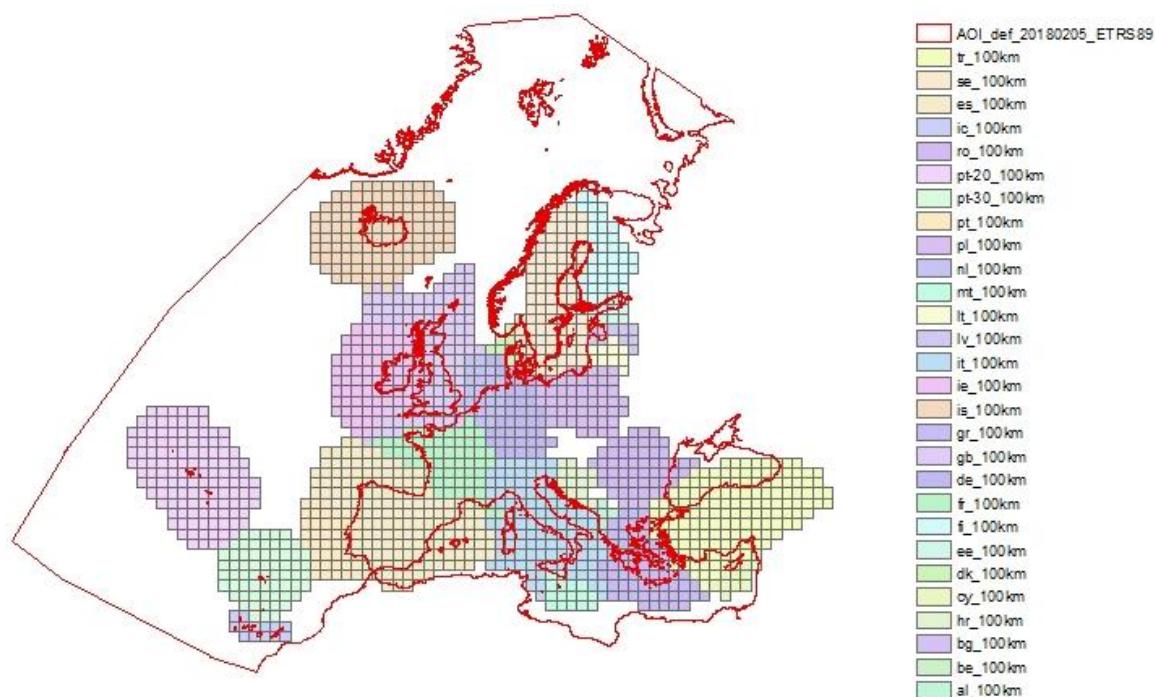
After defining the Aol, it is necessary to create a grid that will be used in the final steps of the method to calculate vessel density. We started from the EEA national grids⁷ for two reasons:

- the EEA grids are based on the recommendations from the 1st European Workshop on Reference Grids in 2003 and on INSPIRE geographical grid systems;
- the EEA grids reference coordinate system is ETRS89-LAEA, i.e. an equal area projection, which makes the grid suitable for generalising data, statistical mapping and analytical work whenever a true area representation is required. Recommended grid resolutions are 100 m, 1 km, 10 km and 100 km.

For each EEA member country, and for Europe as a whole (except Norway), three polygon shapefiles are made available, according to grid resolutions of 1, 10 and 100 km. The coordinate reference system is ETRS89-LAEA Europe, also known in the EPSG Geodetic Parameter Dataset under the identifier EPSG:3035. The Geodetic Datum is the European Terrestrial Reference System 1989 (EPSG:6258).

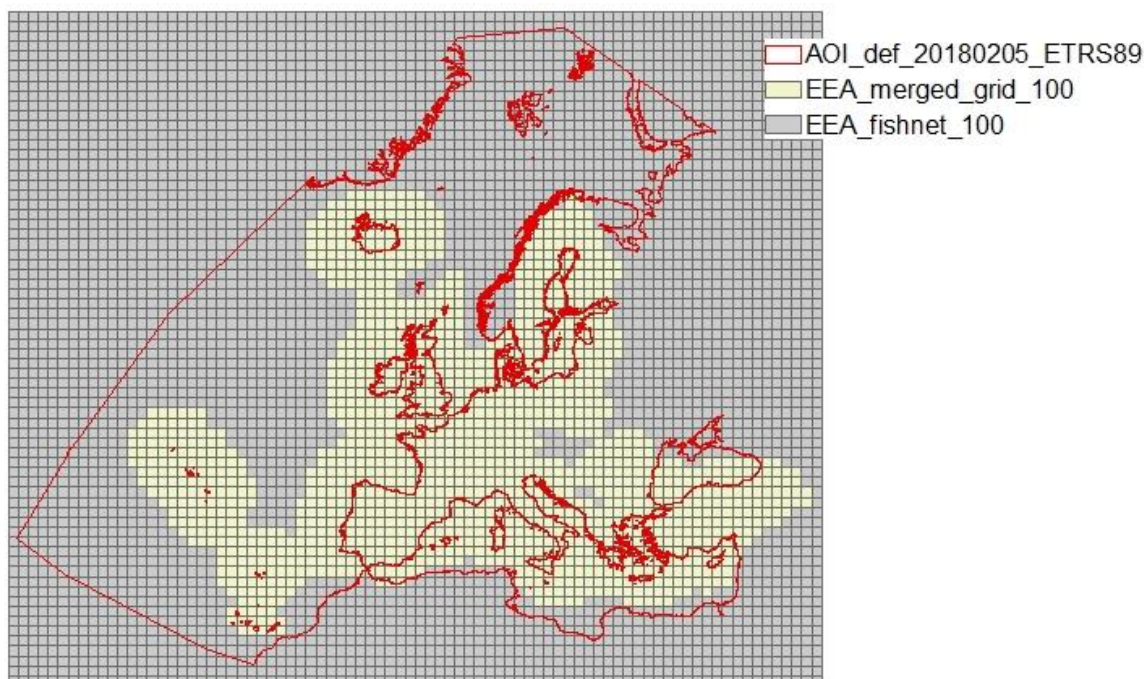
⁷ <https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2>

Figure 8 – EEA national grids (100 km cells)



In some cases, country grids overlap each other and, as can be seen in the map above, the available data do not fully cover the Aoi (for an easier visualisation only the 100km grids are shown). So, it was necessary to merge all the national grids and extend the results to the Aoi bounding box. The obtained grid fully corresponds to and perfectly overlaps the cells of the merged EEA national grids.

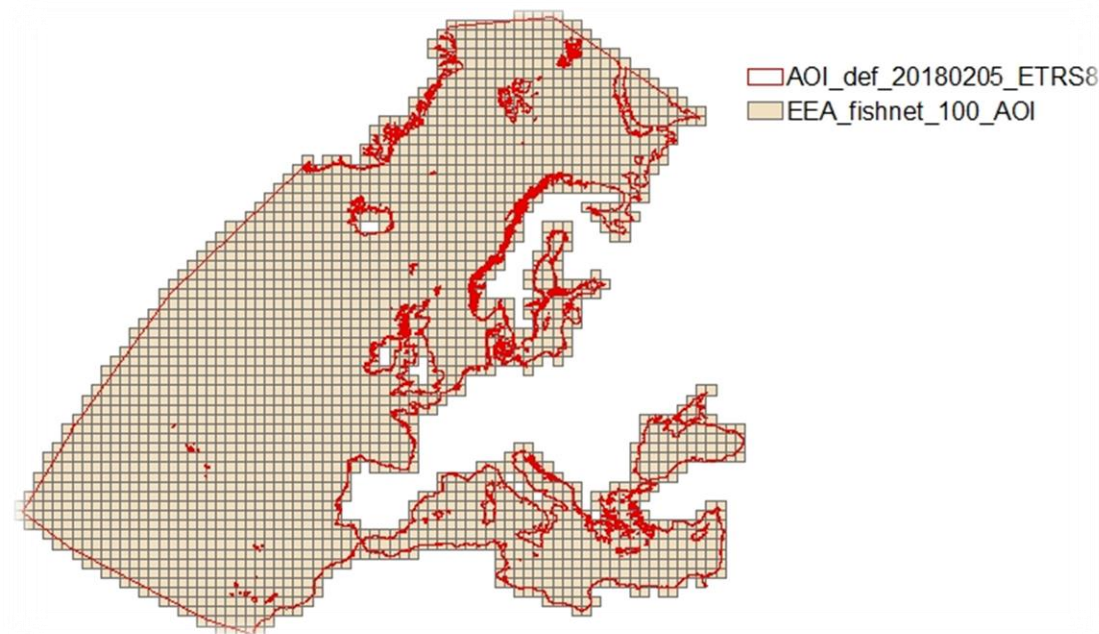
Figure 9 – Extended EEA grid (100 km cells)



The same process was repeated for the other four resolutions, 1km, 10 km, 200 km and 500 km. Only the cells that had their centroid in the Aoi or are within an inner buffer of 1,000 metres from the Aoi borders (N.B. the coastline is a border of the Aoi) were selected, so as to obtain a fully INSPIRE-compliant 1x1km

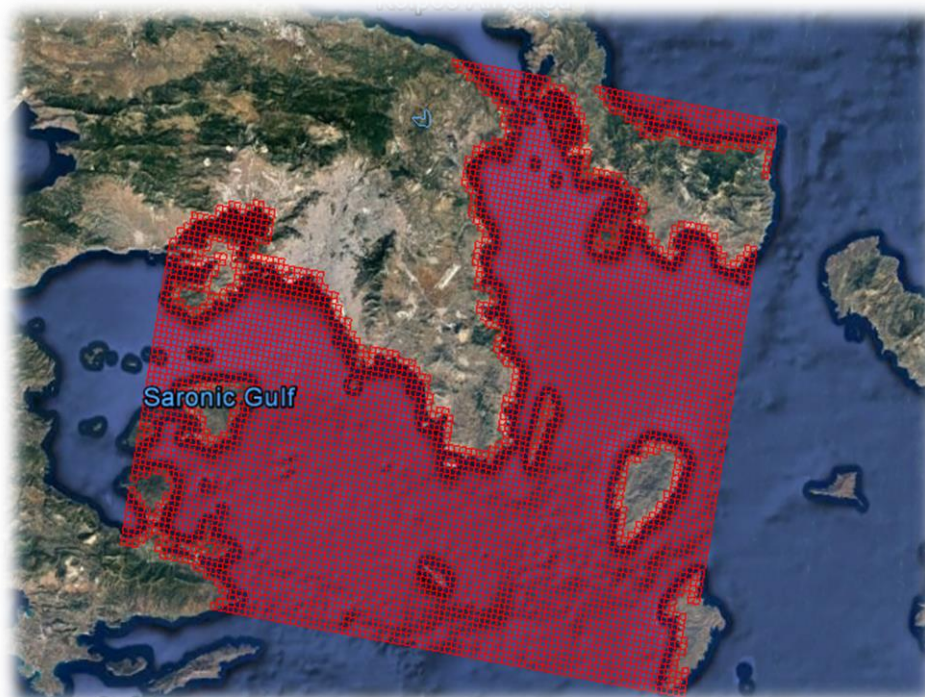
grid for calculating density. The 1,000-metre inner buffer was set to avoid loss of data along the coastline due to the resolution of the coastline layers used, and to limit the calculation only to cells that correspond to ocean areas. The result was a grid with 20,969,020 cells, stored as a feature class in a File Geodatabase.

Figure 10 – Final grid (100 km cells)



Also the 1km grid exactly overlaps the EEA national grids. By way of an example, see the zoom on the 1x1 km grid layer of the Attic peninsula (plotted on Google Earth).

Figure 11 – Grid zoomed on the Attic peninsula (1 km cells)



5.3 Setting up a machine

When dealing with a large Aol and a long time period, processing AIS data to create vessel density map might require considerable computing power. It is advised to set up a dedicated machine or to buy cloud computing power. In our case, we decided setting up a dedicated machine, with the following characteristics:

CPU: Intel Xeon W-2155 @3.30GHz 10 cores

RAM: 32 GB DDR4-2666

It is recommended to have several TB of disk space, because the files generated by the calculations tend to be quite large. SSD is recommended, as the transfer rate to disk can be a limiting factor in the processing. We used a 1TB SSD to process data on the fly, and a dedicated back-up server to store all the files generated throughout the various steps of the method.

The following software was installed on the machine:

- PostgreSQL 9.5 with PostGIS extension
- ArcGIS Desktop Basic 10.6.1
- QGIS 3.4.5 'Madeira'

It is recommended to install or virtualise a Linux distro, because certain data cleaning operations can be more efficiently executed through Shell commands.

5.4 Acquiring and cleaning the AIS data

Generally speaking, AIS data are not easily available. Some countries make them available publicly⁸, some others grant access to selected users. In most cases, only terrestrial AIS data are available, as satellite AIS data are property of private companies which have their own satellite constellation.

As of 2019, there is no source that makes available free terrestrial and satellite AIS data for the whole Europe. Therefore, it was decided to purchase a set of data from a commercial provider, Collecte Localisation Satellites (CLS)⁹, a French company specialised in space-based added value products and services for maritime applications. The dataset covered the year 2017 for an area covering the Aol described above and included both terrestrial and satellite messages. CLS were given the coordinates of the vertices of the polygon corresponding to the Aol:

```
35.5793000002947,84.1457766111018
68.6079635620688,84.1457766111018
68.6079635620688,66.7866189035703
44.4963834818124,63.8033254953754
41.8082307313632,39.9969181476544
36.2158415629379,30.8193040544386
-11.7531667987087,24.6036292163154
-35.5793000002947,24.6036292163154
-41.3099169996698,24.5935451406957
-41.3099169996698,48.9976000010394
-35.5793000002947,84.1457766111018
```

A partial pre-processing of the data was carried out by CLS, based on a method agreed with the EMODnet Human Activities team¹⁰:

⁸ See for instance Denmark: <https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/default.aspx> and the US: <https://marinecadastre.gov/ais/>

⁹ <https://maritime-intelligence.cls.fr/>

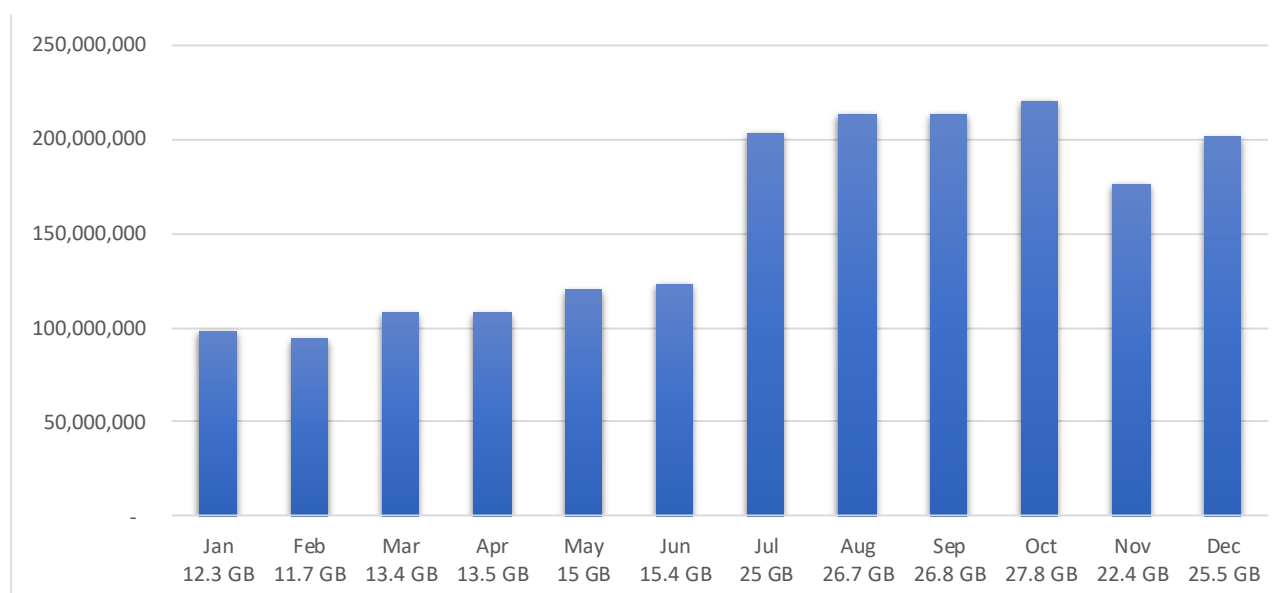
¹⁰ HELCOM's experience was very helpful to define the pre-processing steps.

- The only AIS messages delivered by CLS were the ones relevant for assessing shipping activities (AIS messages 1, 2, 3, 5, 18, 19, 24 and 27). This helped reduce the number of messages to be processed in the subsequent steps.
- Static and voyage data from messages 5 and 24 were associated to the position reports of messages 1, 2, 3, 18, 19 and 27 on the basis of the MMSI number.
- Points on land (clearly errors) were removed.
- The AIS data were down-sampled to 3 minutes, meaning that two consecutive messages from the same vessel could not have a time interval shorter than 3 minutes; this also helped reduce the number of messages to be processed considerably.
- Duplicate signals were removed.
- Messages with wrong MMSI numbers were removed (i.e. more or less than 9 digits or equal to 000000000, 111111111, 222222222, 333333333, 444444444, 555555555, 666666666, 777777777, 888888888, 999999999, 123456789, 0.12345, 1193046).
- Special characters and diacritics were removed from names and call signs.
- Signals with erroneous speed over ground (SOG) were removed (negative values or more than 80 knots)
- Signals with erroneous course over ground (COG) were removed (negative values or more than 360°)
- A Kalman filter (developed by CLS) was applied to remove satellite noise. The Kalman filter was based on a correlated random walk fine-tuned for ship behaviour. The consistency of a new observation with the modelised position is checked compared to key performance indicators such as innovation, likelihood and speed.
- A footprint filter (developed by CLS) was applied to check for satellite AIS data consistency. All positions which were not compliant with the ship-satellite co-visibility were flagged as invalid.

The AIS data were converted from their original format (NMEA) to CSV, and split into 12 files, each corresponding to a month of 2017..

Overall, the dataset included 1,879,626,250 records, from 859,156 unique vessels for a total size of 235 GB.

Figure 12 – Number of records by month



The following fields were available in the acquired dataset:

Table 1 – Available fields in the AIS dataset¹¹

Field	Explanation
mmsi	The Maritime Mobile Service Identity is a 9-digit number used to identify vessels in VHF radio communications. The first three digits denote the country of registry.
source	Satellite or terrestrial
locDate	Date when the movement was recorded
locTime	Time when the movement was recorded
lon	Longitude
lat	Latitude
sog	Speed over ground in 1/10 knot steps
cog	Course over ground. The direction of motion with respect to the ground that a vessel has moved relative to true north, in steps of 0.1°.
trueHeading	The direction that a vessel is pointed at any time relative to true north. Degrees (0-359)
repeatIndicator	Used by the repeater to indicate how many times a message has been repeated.
maneuverIndicator	Indicates whether a vessel is on a special manoeuvre (e.g. regional passing arrangements or inland waterway)
navigationStatus	0 = under way using engine, 1 = at anchor, 2 = not under command, 3 = restricted manoeuvrability, 4 = constrained by her draught, 5 = moored, 6 = aground, 7 = engaged in fishing, 8 = under way sailing, 9 = reserved for future amendment of navigational status for ships carrying DG, HS, or MP, or IMO hazard or pollutant category C, high speed craft (HSC), 10 = reserved for future amendment of navigational status for ships carrying dangerous goods (DG), harmful substances (HS) or marine pollutants (MP), or IMO hazard or pollutant category A, wing in ground (WIG); 11 = power-driven vessel towing astern (regional use); 12 = power-driven vessel pushing ahead or towing alongside (regional use)
rot	Rate of Turn. Right or left (degrees per minute)
positionAccuracy	1 = high (<= 10 m). 0 = low (> 10 m). 0 is the default value.
imo	IMO number (7 digits)
callSign	An alphanumeric identifier used for radio communications. Each flag authority has a designated call sign range, from which the national radio authority issues a call sign for each vessel in that fleet.
shipName	Vessel Name in English format, as recorded by the vessel's registration authority
aisShipType	Ship type and cargo type reported by the ship at the time of the movement. In total 256 values are possible, but only some 50 are in use.
typeOfElectronicPositionFixingDevice	Flag of electronic position fixing device; 0 = RAIM not in use = default; 1 = RAIM in use
eta	Estimated time of arrival as entered by personnel on board the ship
destination	Destination as entered by personnel on board the ship.
shipDraught	Ship draught
shipLength	Ship length
shipWidth	Ship width
flagISOcode	ISO code of the flag state

¹¹ <https://www.navcen.uscg.gov/?pageName=AISMessagesA> for more details.

5.5 Importing the AIS data into a database

Considering the size of the dataset, it was decided to import the csv files into a relational database to store, analyse and query the data. The database used was PostgreSQL.

To this end, an initial set of empty tables, one per month, was created with unspecialised data types for all of the fields listed in the table above.

Upon importing the first csv into its table, it emerged that some messages still contained invalid characters, which caused the import process to fail. It was thus necessary to remove the invalid characters from the csv files. This was done through the ‘sed’ and ‘awk’ utilities launched from a Linux Shell.

The csv files were then imported into their corresponding tables. To speed up data analysis and query it was decided to remove unnecessary fields and keep only those that would be used in the subsequent steps of the method. The remaining fields were assigned specialised data types:

- mmsi: numeric
- locDate: date
- locTime: text
- lon: double precision
- lat: double precision
- imo: character varying
- aisShipType: character varying

Indexes were created on mmsi and aisShipType fields to speed up querying. We also added an ID column (integer) as primary key.

By querying the data (counting records by MMSI and AIS ship type), it turned out that some MMSI numbers were associated to more than one ship type during the year. In principle, this is not an anomaly, as the MMSI number is associated to a device and not to a ship, and the device can be moved from one ship to another. However, in some cases, MMSI numbers changed repeatedly throughout the year, which is more unlikely. In these instances, we attributed to an MMSI the most recurring ship type, or when one of the ship types was an error or not available, we attributed to an MMSI the known ship type.

The aisShipType field admits a wide number of ship types according to the AIS specifications. Therefore, it was decided to create an MMSI register (reporting for each MMSI the unique AIS ship type code and the EMODnet code) with macro-categories of ship types, as per the following table:

Table 2 – EMODnet ship types

EMODnet Ship Type	AIS Ship Type Code	AIS Ship Type Description	Vessel count
0 Other	0	Not available	7,875
	1	Reserved for future use	45
	2	Reserved for future use	25
	3	Reserved for future use	18
	4	Reserved for future use	34
	5	Reserved for future use	13
	6	Reserved for future use	45
	7	Reserved for future use	39
	8	Reserved for future use	12
	9	Reserved for future use	82
	10	Reserved for future use	56
	11	Reserved for future use	4
	12	Reserved for future use	20
	13	Reserved for future use	7
	14	Reserved for future use	14
	15	Reserved for future use	10
	16	Reserved for future use	6
	17	Reserved for future use	9
	18	Reserved for future use	6
	19	Reserved for future use	23

EMODnet Ship Type	AIS Ship Type Code	AIS Ship Type Description	Vessel count
	20	Wing in ground (WIG), all ships of this type	259
	21	Wing in ground (WIG), Hazardous category A	4
	22	Wing in ground (WIG), Hazardous category B	12
	23	Wing in ground (WIG), Hazardous category C	4
	24	Wing in ground (WIG), Hazardous category D	10
	25	Wing in ground (WIG), Reserved for future use	13
	26	Wing in ground (WIG), Reserved for future use	8
	27	Wing in ground (WIG), Reserved for future use	5
	28	Wing in ground (WIG), Reserved for future use	7
	29	Wing in ground (WIG), Reserved for future use	13
	34	Diving ops	385
	38	Reserved	173
	39	Reserved	79
	56	Spare - Local Vessel	111
	57	Spare - Local Vessel	56
	59	Non combatant ship according to RR Resolution No. 18	38
	90	Other Type, all ships of this type	2,418
	91	Other Type, Hazardous category A	28
	92	Other Type, Hazardous category B	14
	93	Other Type, Hazardous category C	12
	94	Other Type, Hazardous category D	7
	95	Other Type, Reserved for future use	36
	96	Other Type, Reserved for future use	37
	97	Other Type, Reserved for future use	151
	98	Other Type, Reserved for future use	27
	99	Other Type, no additional information	2,796
	100-255	Not available (local coding system)	448
Total			15,494
1 Fishing	30	Fishing	15,012
2 Service	50	Pilot Vessel	981
	51	Search and Rescue vessel	2,195
	53	Port Tender	866
	54	Anti-pollution equipment	200
	58	Medical Transport	43
Total			4,285
3 Dredging or underwater operations	33	Dredging or underwater ops	1,447
4 Sailing	36	Sailing	29,155
5 Pleasure craft	37	Pleasure Craft	23,845
6 High-speed craft	40	High speed craft (HSC), all ships of this type	684
	41	High speed craft (HSC), Hazardous category A	12
	42	High speed craft (HSC), Hazardous category B	6
	43	High speed craft (HSC), Hazardous category C	5
	44	High speed craft (HSC), Hazardous category D	8
	45	High speed craft (HSC), Reserved for future use	7
	46	High speed craft (HSC), Reserved for future use	9
	47	High speed craft (HSC), Reserved for future use	18
	48	High speed craft (HSC), Reserved for future use	7
	49	High speed craft (HSC), No additional information	146
Total			902
7 Tug and towing	31	Towing	533
	32	Towing: length exceeds 200m or breadth exceeds 25m	80
	52	Tug	3,712
Total			4,325
8 Passenger	60	Passenger, all ships of this type	4,449
	61	Passenger, Hazardous category A	35
	62	Passenger, Hazardous category B	22
	63	Passenger, Hazardous category C	13
	64	Passenger, Hazardous category D	25

EMODnet Ship Type	AIS Ship Type Code	AIS Ship Type Description	Vessel count
	65	Passenger, Reserved for future use	35
	66	Passenger, Reserved for future use	31
	67	Passenger, Reserved for future use	85
	68	Passenger, Reserved for future use	22
	69	Passenger, No additional information	2,502
Total			7,219
9 Cargo	70	Cargo, all ships of this type	12,469
	71	Cargo, Hazardous category A	1,325
	72	Cargo, Hazardous category B	138
	73	Cargo, Hazardous category C	79
	74	Cargo, Hazardous category D	200
	75	Cargo, Reserved for future use	55
	76	Cargo, Reserved for future use	30
	77	Cargo, Reserved for future use	27
	78	Cargo, Reserved for future use	12
	79	Cargo, No additional information	6,199
Total			20,534
10 Tanker	80	Tanker, all ships of this type	4,718
	81	Tanker, Hazardous category A	497
	82	Tanker, Hazardous category B	399
	83	Tanker, Hazardous category C	213
	84	Tanker, Hazardous category D	266
	85	Tanker, Reserved for future use	30
	86	Tanker, Reserved for future use	8
	87	Tanker, Reserved for future use	11
	88	Tanker, Reserved for future use	17
	89	Tanker, No additional information	1,934
Total			8,093
11 Military and law enforcement	35	Military ops	732
	55	Law Enforcement	754
Total			1,486
12 Unknown	Error (-1)		238,98
Total			370,779

The number of vessels is different from the one reported above (859,156), because vessels with only one recorded position throughout the year were removed; they were most certainly errors.

Csv files were created for each EMODnet ship type and for each month, by using the MMSI number to join the MMSI register with each of the 12 monthly tables.

The total number of csv files obtained was given by the 13 EMODnet ship types multiplied by 12 months, i.e. 156.

5.6 Creating points

The positions stored in the 156 csv files needed to be converted into points in a GIS environment. The software used to do this was ArcGIS 10.6.1 and its Model Builder tools. The ArcGIS environment was configured for best processing performance:

- 64-bit background processing was installed
- 'Background Processing' was enabled within the geoprocessing options
- Script/model property 'always run in foreground' was set to off (although foreground processing is useful for initial testing)
- The scratch file geodatabase was stored on an SSD

We developed a script that made it possible to calculate and store point feature classes in a file geodatabase. A file geodatabase with 13 feature classes (one per ship type) was created for each month of the year.

The size of each file ranges from 30 GB to 65 GB, depending on the month of the year (some months are busier than others).

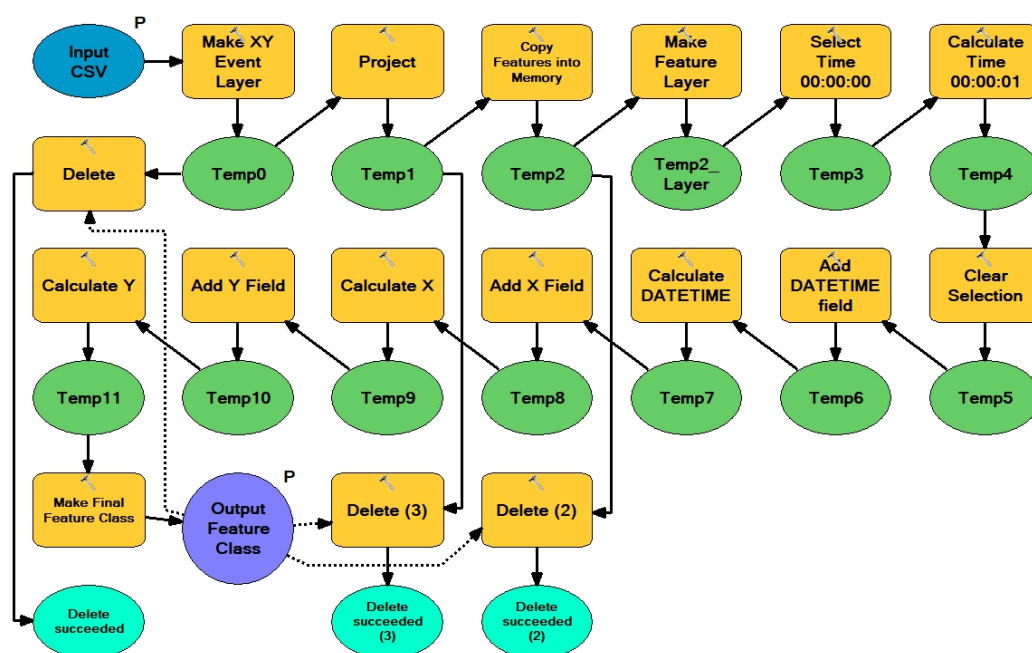
The script developed consists of the following steps:

- Make XY event layer from the input CSV file
- Project from WGS84 to ETRS_LAEA, applying the appropriate geographic transformation
- Copy point features into RAM and make a feature layer
- Select times that are 00:00:00 (ArcGIS treats these values as a blank time rather than exactly midnight)
- Calculate selected times to be 00:00:01 (one second passed midnight)
- Clear Selection
- Add DATETIME field
- Concatenate date and time into DATETIME field
- Add X1 field (type double)
- Calculate X1 values to be the X coordinate in ETRS_LAEA
- Add Y1 field (type double)
- Calculate Y1 values to be the Y coordinate in ETRS_LAEA
- Copy features from RAM into final output layer
- Delete temporary layers from RAM

Overall it took between 12 and 26 hours to create points for a month, depending on the number of records.

The picture below shows the work flow of the script:

Figure 13 – Work flow of the Create Points script



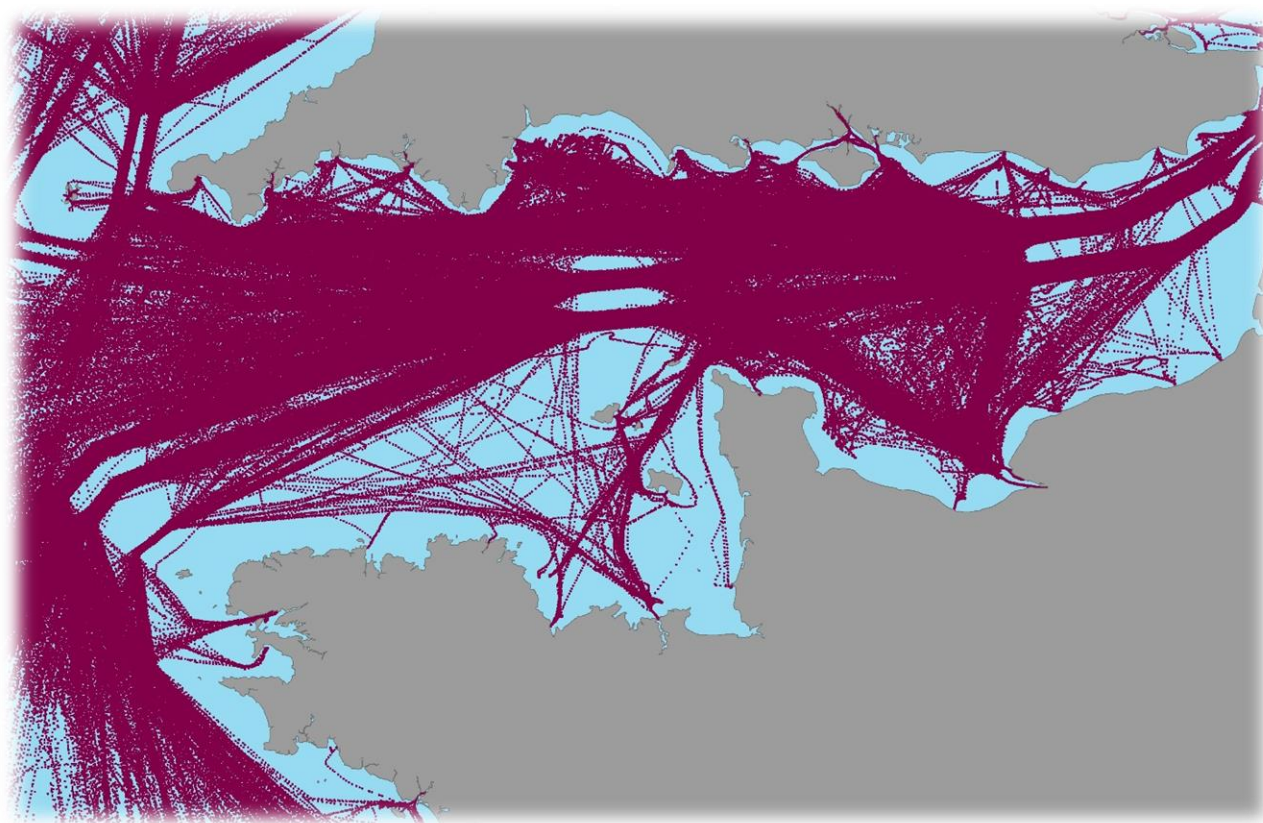
The obtained feature classes had the following fields:

Table 3 – Fields in the point feature class

Field name	Data type
OBJECTID	Object ID
Shape	Geometry
MMSI	Long integer
LocDate	Date
LocTime	Text
Lon	Double
Lat	Double
DateTime	Date
X1	Double
Y1	Double

Essentially, the script was used to project data from WGS84 to ETRS-LAEA, and concatenate date and time into a single field.

Figure 14 – Example of points (positions) in the English Channel (Cargo, December 2017)



5.7 Creating lines

The original input csv files (and therefore our points) were already ordered by vessel ID then by time, saving any effort in re-ordering of records. The points obtained from the previous step were used to reconstruct ship routes (lines), by using the MMSI number as a unique identifier of a ship. Another custom-made script made it possible to calculate and store line feature classes in a file geodatabase. A file geodatabase with 13 feature classes (one per each ship type) was created for each month of the year. The size of each file ranges from 8 GB to 20 GB, depending on the month of the year.

The script firstly added to each record the MMSI number, coordinates and the timestamp of the preceding record with the same MMSI. It then calculated the time interval in hours between one record and the other. A

line was then created for every two consecutive positions of a ship. In addition, for each line the script calculated its length (in km). If the distance between two consecutive positions of a ship was longer than 30 km or if the time interval was longer than 6 hours, no line was created. This was a necessary precaution, because in many cases a ship might stop sending messages while on voyage, and then start again even after days. As can be grasped intuitively, interpolating the data in such cases would have produced inaccurate results and implausible routes. On the other hand, the interpolation method adopted – i.e. creating a line only between two consecutive positions of a ship and not for her whole route – produced better, more accurate results.

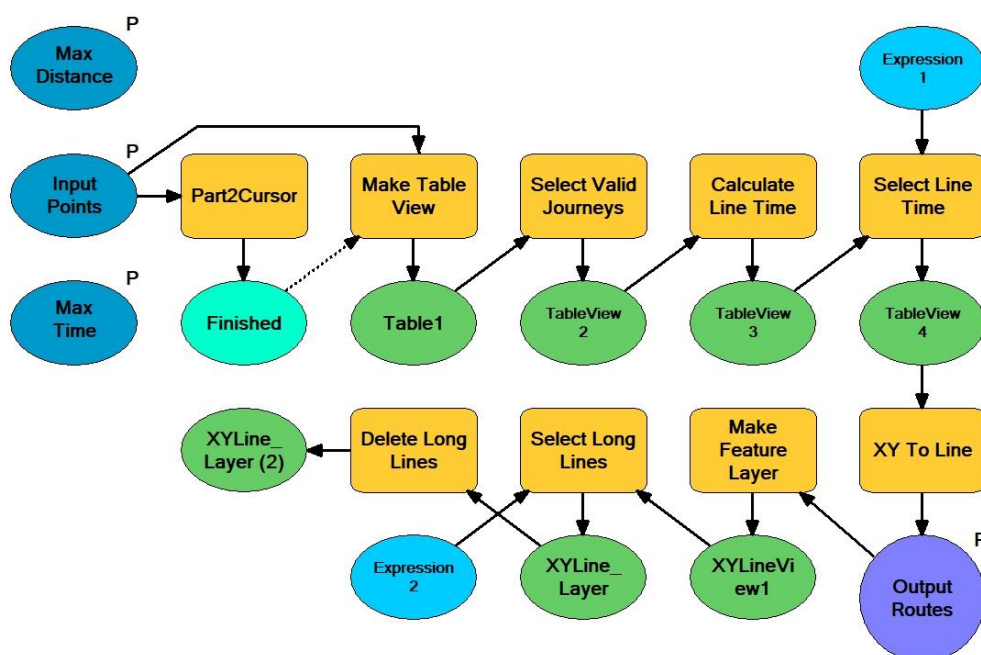
The script developed consists of the following steps:

- Add blank fields, X2, Y2, ID2, datetime2.
- Create variables X1, Y1, ID and datetime1 and set to zero.
- Then for each row
 - o Calc $X2 = X1$, $Y2 = Y1$, $ID2 = ID$, $datetime2 = datetime1$.
 - o Read/store new values from current row for X1 (from X), Y1 (from Y), ID (from MMSI), datetime1 (from datetime) and repeat the step.
- Select Valid Journeys: Copy all records to new feature class where $ID = ID2$. Note: if $ID \neq ID2$, these would be lines that join routes from different vessels together – including the first row.
- Create new LINETIME field (long)
- Calculate LINETIME (difference between datetime and datetime1 in hours)
- Select LINETIME rows where time is within the parameter set in the tool options
- Create straight line routes for each selected record using X1, Y1, X2, Y2. Carry through LINETIME as the ID, so we don't have to re-join the previous table.
- Select lines greater in length than specified in the parameter set in the tool options and remove.

Overall it took between 8 and 18 hours to create lines for a month, depending on the number of records.

The picture below shows the work flow of the script. The section 'Part2Cursor' contained the python code which added to each record the MMSI number, coordinates and the timestamp of the preceding record.

Figure 15 – Work flow of the Create Lines script

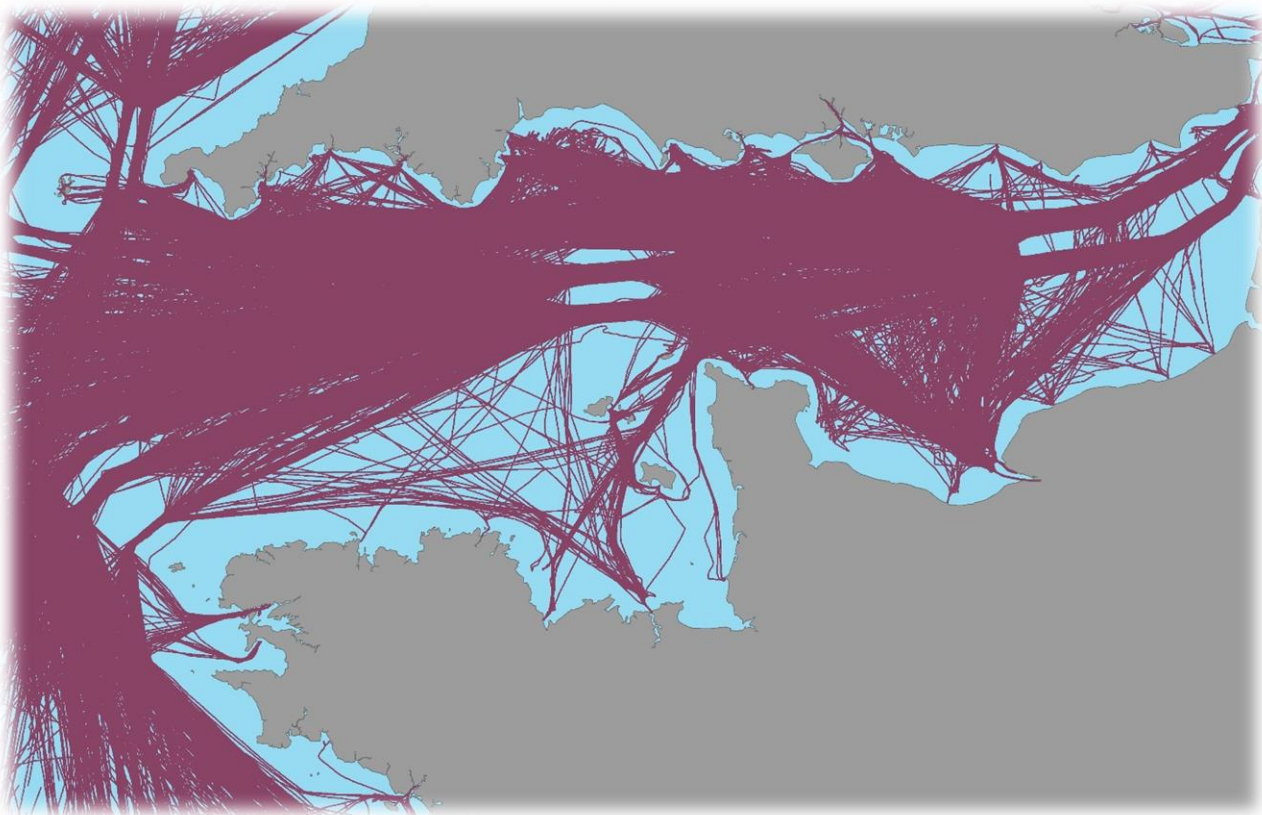


The obtained feature classes had the following fields:

Table 4 – Fields in the line feature class

Field name	Data type
OID	Object ID
Shape	Geometry
X1	Double
Y1	Double
X2	Double
Y2	Double
LineTime	Double
Shape_length	Double

Figure 16 - Example of lines (routes) in the English Channel (Cargo, December 2017)



5.8 Calculating density

The lines obtained in the previous step were intersected with the grid. Because each line had length and duration as attributes, it was possible to calculate how much time each ship spent in a given cell over a month. This is a distinctive feature of our method; when a line crossed two or more cells of the grid, we were able to establish how much time the ship spent in each cell, by calculating the length of the line segment that crossed a cell, dividing it by the total line length, and multiplying it by the total line time. Density is thus expressed in hours per square kilometre per month. For each cell of the grid, we then summed the time value of each “segment” in it, thus obtaining the density value associated to that cell.

To do so, we imported the 1x1km grid polygon feature as well as the lines obtained in the previous step in a new PostgreSQL (with the PostGIS extension enabled) database, using the ogr2ogr GDAL command. This tool also created a spatial index for the imported features.

A SQL script was run in PostgreSQL with the following steps:

- Intersect lines with the grid using the PostGIS function `st_intersection`
- Add a field `SEGLENGTH` and calculate the length of each segment using the PostGIS function `st_length`
- Add a field `SEGTIME` and calculate the segment time to be $(SEGLENGTH / Shape_length) * LineTime$
- Sum the segment time within each grid square
- Join the values to the 1km grid and make a new grid table

The result was a new grid polygon object with a field containing the number of hours spent in all cells by each ship type per each month. Overall, there were 156 files, 3.5 GB each.

Overall it took around 24 hours to calculate density for a month.

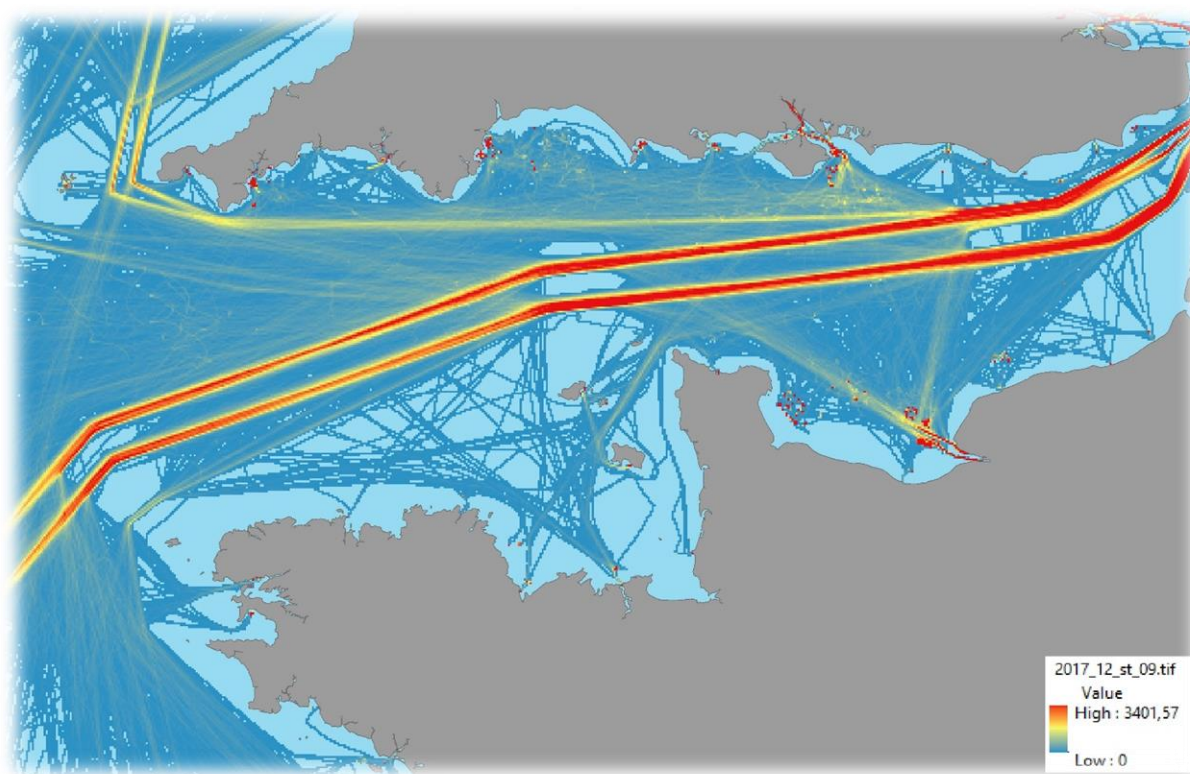
5.9 Creating maps

The final step consisted of creating raster files (TIFF file format) from the PostgreSQL vessel density tables, using the 'rasterize' GDAL command (setting the grid polygon feature class as output extent and assigning -9999 as 'nodata' value). With 14 ship types (including the sum 'all ship types') multiplied by 12 months, we had 168 raster files. Then we added 14 more raster files (one per each ship type) with average density values over the whole year. What is called 'vessel density map' is in fact a set of 182 different maps.

The monthly raster datasets with 'all ships types' and the yearly averages were calculated with QGIS raster calculator.

Each raster (uncompressed) is made up of 21 million cells, for a size of 184 MB.

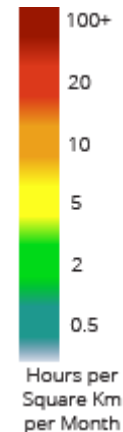
Figure 17 - Example of cargo density map of the Mediterranean (December 2017)



6 Publishing the maps

The maps were made available on EMODnet Human Activities. The rasters in WGS84 projection with raw time values were stored on the server and published using GeoServer. The rasters are symbolised dynamically assigning colours (as hexadecimal values) to time values using a logarithmic scale.

```
<RasterSymbolizer>
  <ColorMap>
    <ColorMapEntry color="#0b2c7a" quantity="0" opacity="0" />
    <ColorMapEntry color="#20998f" quantity="0.5" />
    <ColorMapEntry color="#00db00" quantity="2" />
    <ColorMapEntry color="#ffff00" quantity="5" />
    <ColorMapEntry color="#eda113" quantity="10" />
    <ColorMapEntry color="#dd3a1b" quantity="20" />
    <ColorMapEntry color="#9b1800" quantity="100" />
  </ColorMap>
  <Opacity>1.0</Opacity>
</RasterSymbolizer>
```



The rasters are also projected dynamically to the WGS84/Pseudo-Mercator map projection so they display correctly in the 'View Data' page.

The rasters were made available as WMS and are provided monthly by vessel type along with an average for the whole year.

6.1 Converting to different density units

As explained, the units of density in the published maps are total ship time in hours spent in a cell over the whole month. As the cells are 1 km², this is immediately per square km.

In order to obtain the density as the average number of instantaneous ships per square km, the values of the published monthly maps must be divided by the total number of hours in the month, which is 28, 30 or 31 times 24 hours depending on the month. For the published yearly average maps, the values must be multiplied by 12 and divided by 365 times 24 hours (for 2017).

To convert to ship density per square degree instead of per square km, the published numbers can be multiplied by the factor $(60 * 1.852)^2 \cos(\text{latitude})$.

7 What's next

Creating EMODnet's vessel density maps was a daunting task, but also an amazing journey. Working with big data is not necessarily rocket science, but it certainly poses many a challenge: the amount of data to process is ginormous, so figuring out what to do and how to do it is more often than not a trial and error process. The problem is that trial and error might be quite time-consuming when processing a large amount of data, even more so if there is a looming deadline. Nevertheless, overall it was a terrific experience; we learned a lot and we got to know a bunch of wonderful people, who helped us focus on our goal. The European marine data community is a close-knit group of experts, many of whom took a friendly interest in our work and never hesitated to help out and share their experience.

EMODnet's vessel density maps went live on 11 March 2019, and so it is time for us to figure out how we can capitalise on this experience and improve the services offered.

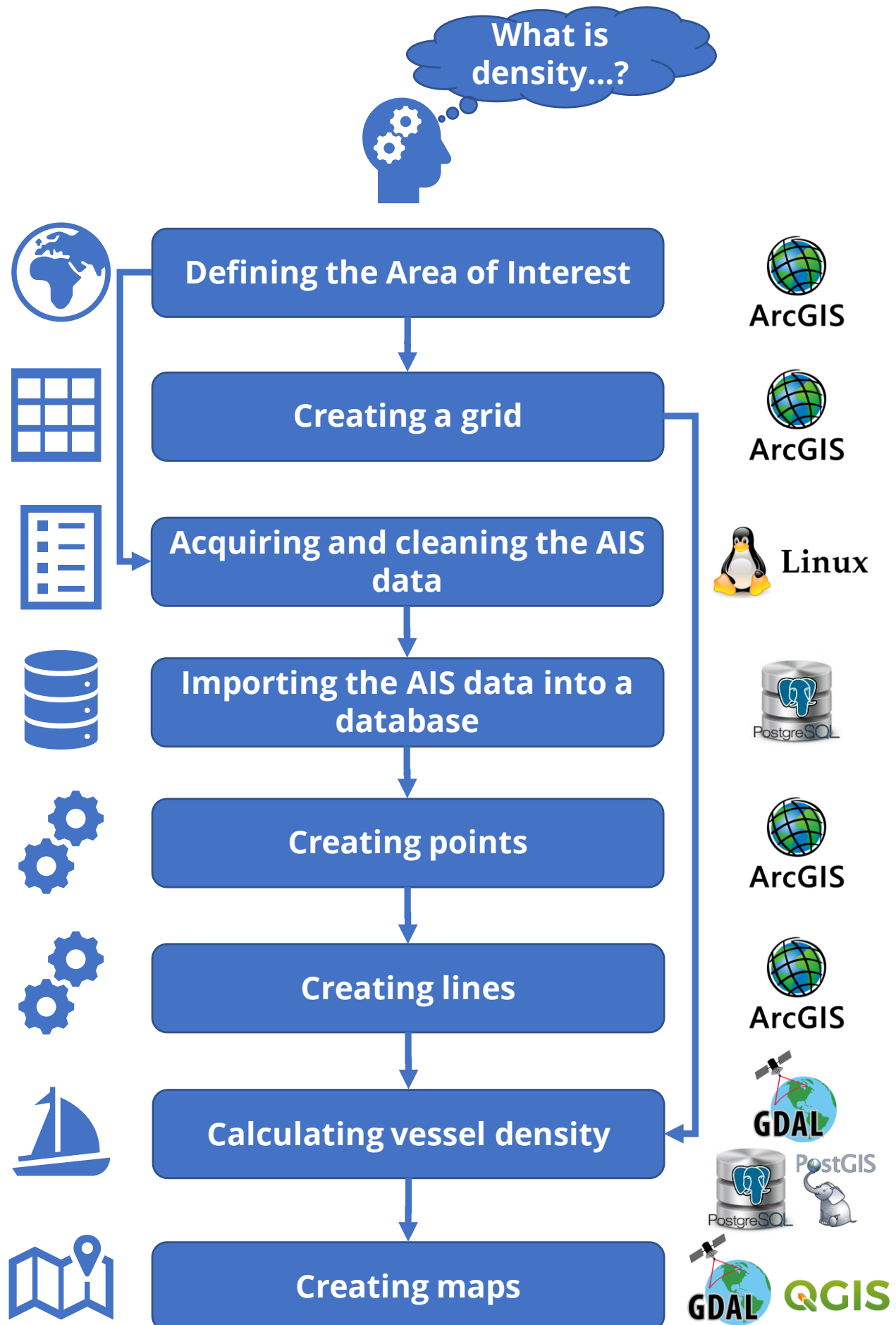
Since the current phase of EMODnet runs until 2021, our first goal is to provide regular updates to the density maps. An update with 2018 data is due later in 2019. Other possible developments, depending on time and budget, might include:

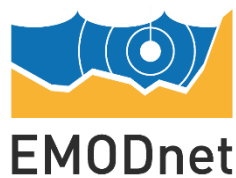
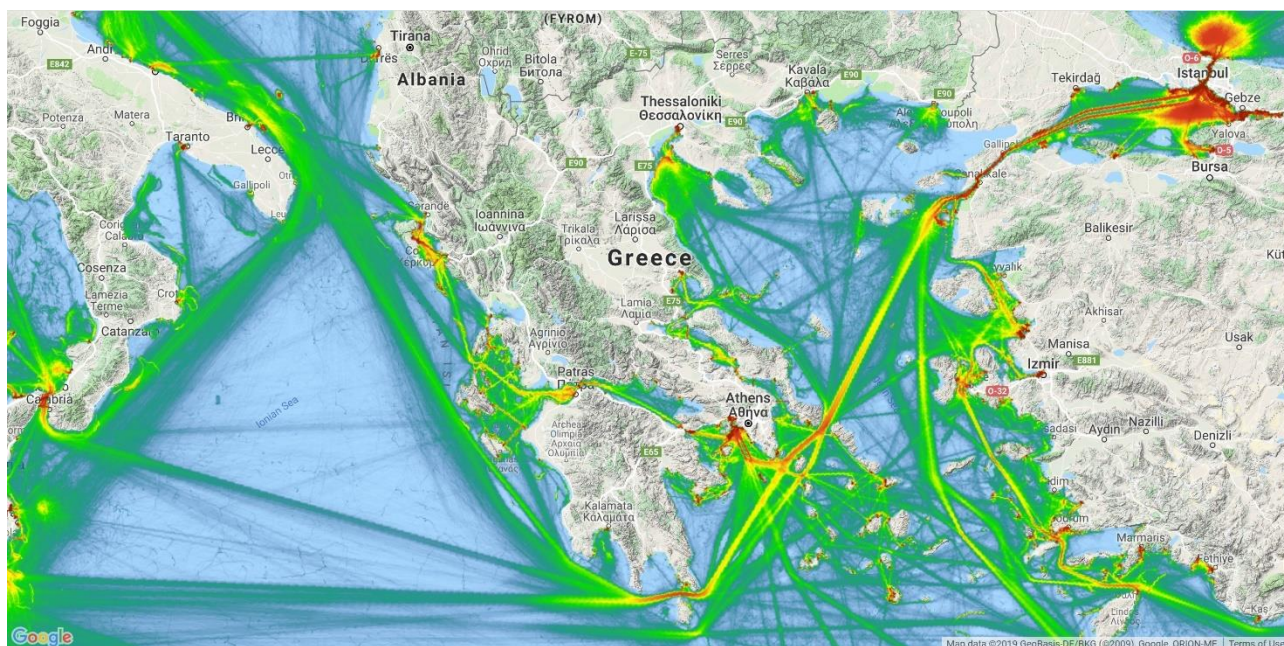
- Including inland waterway traffic.
- Purchasing historical datasets, so to enrich our time series.
- Complementing AIS data with data from other databases, so as to develop other maps such maps on shipping emissions and noise maps (in collaboration with EMODnet physics).

As usual, feedback from users will be crucial. Rather than deciding how to move forward ourselves, we would like to hear what researchers, businesses, maritime spatial planners, policy makers and other users have to say, so as to understand how we can better serve the marine community with our products.

8 References

- Bonham C., Noyvirt A., Tsalamanis I., Williams S., 'Analysing port and shipping operations using big data', Data Science Campus, ONS, 2018.
- Bureau of Ocean Energy Management, National Oceanic and Atmospheric Administration, 'Tutorial:How to Build Vessel Density Maps with AIS', 2015.
- Fiorini M., Capata A., Bloisi D., 'AIS Data Visualization for Maritime Spatial Planning (MSP)', International Journal of e-Navigation and Maritime Economy 5 (2016), pp. 45-60, 2016.
- Hilliard C., Pelot R., 'Analysis of Marine Traffic along Canada's coasts', Phase 1 Final Report, 2012.
- Marine Management Organisation, 'Mapping UK shipping density and routes technical annex', 2014.
- Mazzarella F., Vespe M., Damalas D., Osio G., 'Discovering vessel activities at sea using AIS data: Mapping of fishing footprints', 17th International Conference on Information Fusion (FUSION), 2014.
- Natale F., Gibin M., Alessandrini A. Vespe M., Paulrud A., 'Mapping Fishing Effort through AIS Data', PLoS ONE 10(6), 2015.
- Nicolas F., Frias M., Backer H., 'Mapping maritime activities within the Baltic Sea', Baltic Scope, 2016.
- Pallotta G., Vespe M., Bryan K., 'Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction', Entropy 15(6), pp. 2218-2245, 2013.
- Posada M., Greidanus H., Alvarez M., Vespe M., Cokacar T., Falchetti S., 'Maritime awareness for counter-piracy in the Gulf of Aden', IEEE International Geoscience and Remote Sensing Symposium, 2011.
- Vespe M., Greidanus H., Santamaria C., Barbas T., 'Knowledge Discovery of Human Activities at Sea in the Arctic using Remote Sensing and Vessel Tracking Systems', ShipArc Proceedings, 2015.
- Vespe M., Gibin M., Alessandrini A., Natale F. Mazzarella F., Osio G., 'Mapping EU fishing activities using ship tracking data', Journal of Maps, 12:sup1, pp. 520-525. 2016





HUMAN ACTIVITIES

Making use of our oceans

www.emodnet-humanactivities.eu

