

Studies to support
the European Green
Deal - Lot 2
“Vessel density”
Final Technical Report



MarineTraffic
Global Ship Tracking Intelligence



UBITECH
ubiquitous solutions

This study has been conducted by MarineTraffic Operations SA (A. Troupiotis, K. Bereta, G. Spiliopoulos, M. Vodas, G. Karantaidis and D. Zissis) with the contribution of Ubitech (S. Karagiorgou and P. Petrou).

Disclaimer: *This study has been funded by the European Maritime and Fisheries Fund (EMFF) through service contract No. CINEA/EMFF/2020/3.1.16/Lot2/SI2.850940 in response to the invitation to the call EASME/2020/OP/0010 Lot 2 "Studies to support the European Green Deal-Lot 2 Vessel density". 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 EASME, nor the European Commission, guarantee the accuracy of the data included in this study. This study as well as any other results and rights obtained in performance of the ensuing contract, including copyright and other intellectual or industrial property rights, shall be owned solely by the Agency, which may use, publish, assign or transfer them as it sees fit, without geographical or other limitation, except where industrial or intellectual property rights exist prior to the contract being entered into force.*

Please cite this document as: *A. Troupiotis, K. Bereta, G. Spiliopoulos, M. Vodas, G. Karantaidis, S. Karagiorgou, P. Petrou and D. Zissis, "Studies to support the European Green Deal -Vessel density", Final Technical Report, 2022*

Scope of the document

This report summarises the technical achievements of the 9-month study, which has been funded by the European Maritime and Fisheries Fund (EMFF) through service contract No. CINEA/EMFF/2020/3.1.16/Lot2/SI2.850940

This document describes the approach used to improve the accuracy of vessel density maps and provides the corresponding map results and findings.

Executive Summary

This report documents in detail the work carried out in the context of the service contract No. CINEA/EMFF/2020/3.1.16/Lot2/SI2.850940-Studies to support the European Green Deal - Lot 2 "Vessel density" from the 20th August 2021 up until the 20th May 2022.

This project aims to address the general objective of producing accurate digital raster maps of shipping activities in European waters, that will help plan and serve for the analysis of the marine aspects of the Green Deal. The overall goal of this project was to investigate the capabilities and capacities of Satellite based Automatic Identification Systems, while exploring its complementarity with other sources (such as Terrestrial -based AIS and Earth Observation images), so as to improve our overall knowledge of vessel density in European waters. As we show within, there are clear benefits of fusing sensory data from heterogenous sensors, in terms of extended spatial and temporal coverage and increased confidence in the results.

Towards this, we explore image processing techniques based on Artificial Intelligence and Machine Learning approaches to detect vessels from EO images, while we develop algorithmic techniques capable of interpolating the missing data with high accuracy.

Terrestrial and Satellite AIS data from MarineTraffic, have been used in this study to compare reception accuracy and coverage, while EO imagery was acquired from the Copernicus Programme (Sentinel 1 and Sentinel 2).

Although the results are unique in scope and detail, the estimates from this study are in good agreement with the related literature[1]–[5]. The main achievements and findings can be summarised as follows

- We designed and developed a novel algorithmic approach for ship detection in satellite EO images.
- An empirical performance and accuracy analysis has been performed between Satellite AIS, Terrestrial AIS and EO datasets with the aim of revealing and highlighting areas

and conditions that could affect density map generation. Exemplification of complementarities of terrestrial and satellite-based AIS, as well as EO datasets are presented.

- We generate a series of detailed vessel density maps (at different spatial resolutions), based on all previous datasets and “detectability maps” indicating the probability that a vessel can cross an area without potentially being detected (from a single sensory platform).
- We present a novel trajectory reconstruction method, capable of highly accurate “gap filling” based on historical vessel information

As part of this work, we design and implement a prototype toolbox, providing methods covering all phases of the data process; from cleaning and smoothing, to trajectory reconstruction and finally density map generation. We make this software publicly available¹ and release it under a Creative Commons license (Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0)).

The results of this study have appeared in several scientific publications. The overall work reported within has been conducted by MarineTraffic with the support of Ubitech.

¹ <https://www.marinetraffic.com/research/the-marinetraffic-ais-toolbox/>

Contents

- Scope of the document - 3 -**
- Executive Summary - 3 -**
- Acronyms..... - 7 -**
- Partners - 8 -**
- 1. Introduction - 10 -**
- 2. Methodology and overall Approach - 12 -**
- 3. Background..... - 14 -**
 - 3.1 Automatic Identification System..... - 14 -
 - 3.2 Data Fitness - 16 -
 - 3.3 Density maps - 18 -
 - 3.4 Data sources and raster-based spatial analysis..... - 20 -
- 4. Data sources used in this study - 21 -**
 - 4.1 Area of interest - 21 -
 - 4.2 Creating a common grid - 22 -
 - 4.3 Transforming the datasets - 23 -
 - 4.3.1 AIS Transformation..... - 24 -
 - 4.3.2 Transformation of EO images - 28 -
- 5. Spatial Analysis and Comparisons..... - 35 -**
 - 5.1 Comparison between different Ter AIS and Sat AIS - 35 -
 - 5.2 Comparison between Sat AIS, Ter AIS and SAT EO - 41 -
- 6. Creating new sets of digital density maps - 47 -**
 - 6.1 Density maps using AIS - 47 -
 - 6.2 Density maps using EO data..... - 55 -
 - 6.3 Density maps of dark areas - 59 -
 - 6.4 Density maps combining Sat AIS & EO data..... - 63 -
 - 6.5 Corrected trajectories density digital maps..... - 67 -
 - 6.5.1 Trajectory Reconstruction Methods..... - 67 -
 - 6.5.2 Reconstructing incomplete trajectories based on historical information..... - 71 -

6.6 Corrected Density maps..... - 74 -

7. Conclusions and future work - 76 -

7.1 Conclusion - 76 -

7.2 Nest steps..... - 79 -

References..... - 80 -

Appendix..... - 84 -

Acronyms

AI	Artificial Intelligence
AIS	Automatic Identification System
EMFF	European Maritime and Fisheries Fund
EMODnet	European Marine Observation and Data Network
EO	Earth Observation
ML	Machine Learning
MT	MarineTraffic
SaaS	Software as a Service
Sat AIS	Satellite AIS
Ter AIS	Terrestrial AIS

Partners



MarineTraffic (hereinafter MT) is a research-intensive SME founded in 2007 and today a key player in the maritime analytics market, offering vessel tracking services and actionable maritime intelligence. It is the leading maritime Software as a Service (SaaS) in terms of global market penetration & popularity.

MT started as an experiment in late 2007, based on the newly introduced Automatic Identification System (AIS) [4], being the first ever website to publicly display vessel positions in real time. The system leveraged open data and state of the art internet technologies-the novel back then-map “mashups” and the power of crowdsourcing. The MT domain was registered and the website was launched, with some basic functionality: an open system where anyone who had access to AIS data could send the data over and have it automatically displayed on the world map. Due to its popularity, MT quickly became synonymous to vessel tracking at a global scale. MT went on to expand its coastal network to more than 5K AIS coastal receiving stations, having today the most extensive proprietary AIS receivers’ network and establishing alliances with the major Satellite AIS providers (e.g. ExactEarth, Orbcomm and others). Today coverage is more or less global, while MT holds a historical database extending over a period of more than 10 years, providing a valuable source for descriptive, predictive and even prescriptive analytics.

Nowadays, multiple sources make density maps available either for free or for a fee. MT has been one of the pioneers in releasing free to the public open density maps as early as 2013-2014. At the time, these were based both on terrestrial and satellite AIS datasets.

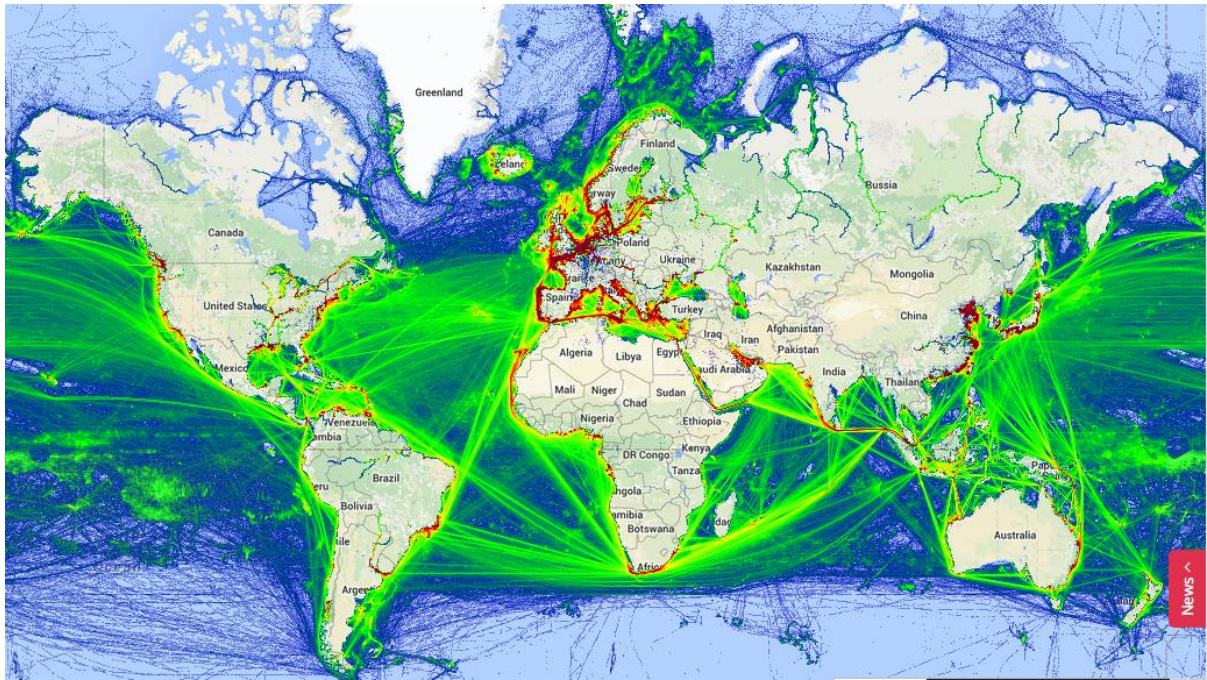


Figure 1. MarineTraffic 2015 Global Density maps

This report outlines the challenges of generating accurate density maps based on sparse AIS datasets and provides a number of solutions that address these challenges, based on technologies and approaches built and tested within MT at an industrial level. With the support of **UBITECH (UBI)**, a leading, highly innovative software house, systems integrator and technology provider. UBITECH provides leading edge intelligent technical solutions and consulting services to businesses, organizations and government in order to allow the efficient and effective secure access and communication with various heterogeneous information resources and services, anytime and anywhere. UBITECH enables real-time valid information processing and decision making, the realization of intelligent business environments, and B2B and B2C transactions by providing high added-value business –oriented and –based solutions.

1. Introduction

The European Commission has set a goal of achieving zero emissions of greenhouse gases by 2050[6], in line with the Paris Agreement[7], where countries globally have agreed to pursue efforts to limit global warming to below 2 degrees Celsius, compared to pre-industrial levels. Scenarios explored in this context, suggested that achieving this will require doubling electricity production with about a quarter of it being produced offshore[6]. resulting in enormous changes to European waters. Up to a quarter of certain countries' waters could be devoted to wind farms. This will inevitably impact other marine activities. Current sea related uses and activities involve maritime shipping, sea-fishing, extended aquaculture, oil and gas exploration and drilling, leisure and boating activities, cultural heritage conservation initiatives and many more. Improving our understanding of activities that take place at sea, including their spatial and temporal aspects, is vital in the light of an intensified use of maritime space[8].

The aim of this study is to add to the evidence base that will underpin planning for this transition.

Today the most valuable resource for understanding shipping and marine related activities is vessel traffic data, as collected through the Automatic Identification System. To support improved understanding of this data, commonly density maps are used as an information visualisation method. The added value of density maps is that they support a bird's eye view of vessel traffic, through providing an overview of vessel behaviour either at a regional or global scale. Analysing density maps as they evolve over time also supports understanding traffic changes and pattern distribution.

From density maps one can determine the patterns of life in a given study area. Patterns of Life are understood as observable human activities that can be described as patterns in the maritime domain related to a specific activity (e.g., fishing) [9]. Essentially, vessel-based maritime activity can be described in space and time, while classified to a number of known activities at sea (fishing etc). The spatial element describes recognised areas where maritime activity takes place; thus, including ports, fishing grounds, offshore energy infrastructure, dredging areas, etc. The transit paths to and from these areas also describe the spatial element (e.g., commercial shipping, ferry routes, etc.), while the temporal element often holds additional information for categorising these activities (e.g., fishing period, time of year, etc.)[10].

Capturing the accurate and complete trajectory of a moving object though, is almost impossible in real conditions, due to the inherent limitations of data acquisition and storage mechanisms. As a result, the continuous movement of an object is usually obtained as an approximate form of discrete samples of spatiotemporal locations [4]. Sometimes the error is

acceptable for a given use case while in other occasions it may lead to erroneous interpretations.

The overall goal of this study to investigate the capabilities and capacities of AIS datasets for accurate density maps generation, and especially Satellite AIS capabilities and limitations, while exploring its complementarity with other sources (such as Terrestrial AIS and EO images), and designing methods for improving its “data fitness” for the specific use. The rest of this study explores the above theme.

2. Methodology and overall Approach

Creating accurate vessel density requires overcoming issues pertinent in the underlying data. Our approach is organised around three main tasks, while following an overall raster-based spatial analysis approach: Step 1 involves all the steps towards homogenising all different spatial datasets across a common grid; Step 2 involves the tasks for performing the quantitative and qualitative comparison of all available data sources; while Step 3 involves the tasks of generating improved density maps.

Our overall implementation methodology is shown graphically in Figure 2.

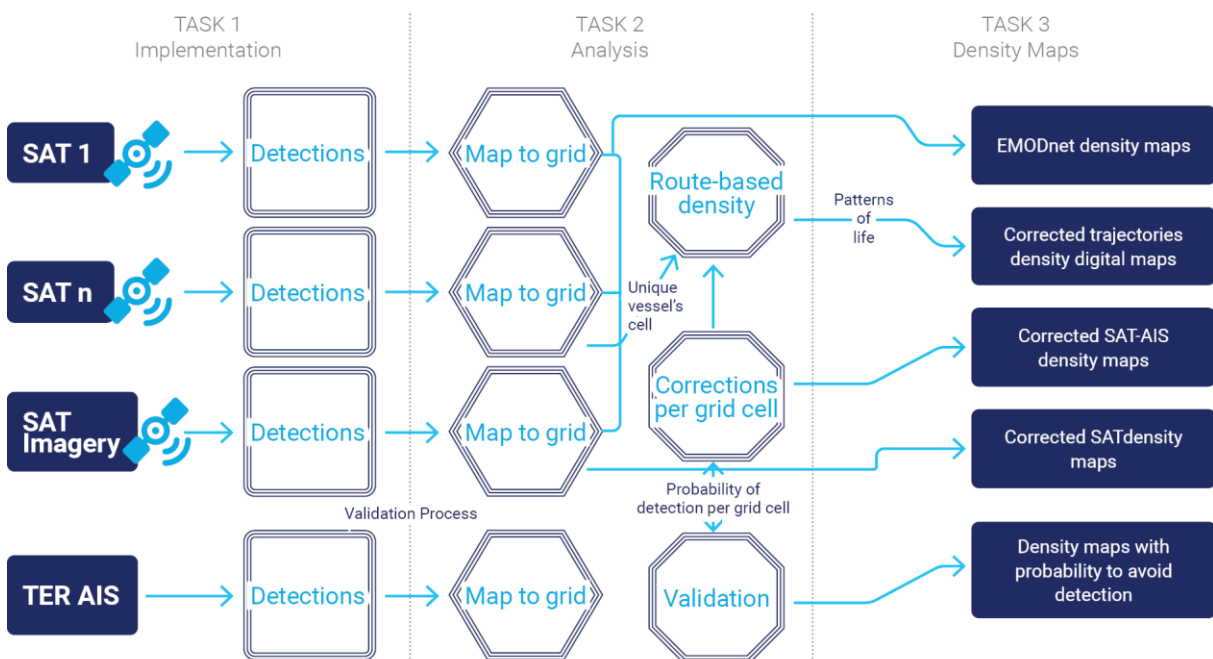


Figure 2. Approach

In the context of Task 1, we implement a Python workflow that processed data from the aforementioned sources, detects vessel positions, and maps them to a common grid, in order to perform calculations per grid cell, so that we are able to compare vessel density data from different sources in Task 2. The output of Task 1 is a set of vessel positions and detections for each data source, stored in related database table, named "density tables".

The output of Task 1 is provided as input to the analysis tasks described in Task 2. To facilitate the analysis of vessel density and compare the different sources, we use a uniform grid and we map the derived positions of Task 1 to this grid. Then, we perform calculations per grid cell. Next, we will compare the different sources per grid cell, which leads to corrections with respect to the real number of vessels that are located in a grid cell. In the analysis step we also calculate the probability of detection per grid cell for Sat AIS, by using data from the other sources.

The analysis output of Task 2 is then used to create a new set of density maps in Task 3. In this task, we enhance the AIS-based approaches used by EMODnet, using the results of Task 1 and Task 2, as we explain in more detail later on, and create a set of density maps based on Sat AIS, as well as other sources such as EO. The corrected density maps will be provided as sets of raster tiles that can be deployed on GeoServer², so that they can be easily accessible and used, complying to the OGC standards.

² <http://geoserver.org>

3. Background

3.1 Automatic Identification System

A trajectory can be captured as a time stamped series of location points denoted as $p_0(x_0, y_0, t_0)$, $p_1(x_1, y_1, t_1)$, ..., $p_n(x_n, y_n, t_n)$, where x_i, y_i represents geographic coordinates of the moving object at time t_i and N is the total number of elements in the series. Simply put, to generate the trajectory, a sensor needs to acquire its coordinates x, y at time t .

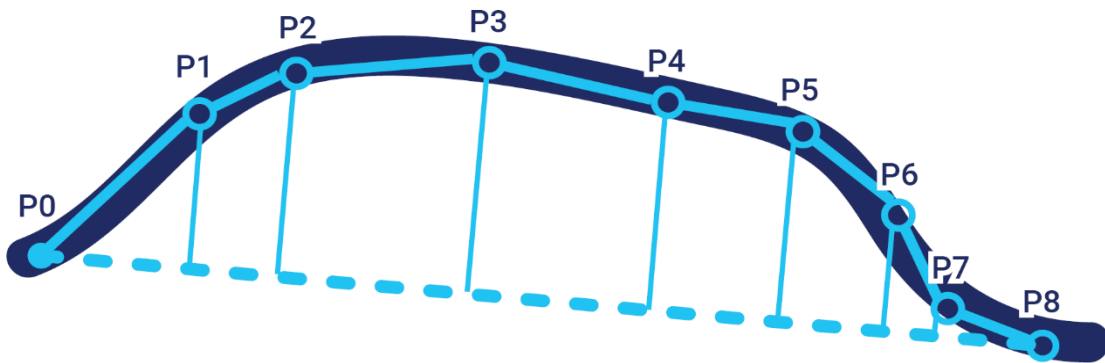


Figure 3. In red the true trajectory, while in black the approximated trajectory captured by a given sensor

Notice that the approximated trajectory can also be represented as a series of line segments between the stamped positions (given that there is a unique identifier grouping these positions into the same trajectory)

$$\sum^8 \text{traj}_1 = \underline{p_0p_1}, \underline{p_1p_2}, \underline{p_2p_3}, \underline{p_3p_4}, \underline{p_4p_5}, \underline{p_5p_6}, \underline{p_6p_7}, \underline{p_7p_8}$$

The most commonly used dataset for tracking vessel activities at sea is the Automatic Identification System (AIS), a collaborative, self-reporting system that allows vessels to broadcast their identification information, characteristics and destination, along with other information originating from on-board devices and sensors, such as location, speed and heading [2]. AIS messages are broadcast periodically and can be received by other vessels equipped with AIS transceivers, as well as by on the ground or satellite-based sensors. This information is transmitted at regular intervals ranging anywhere from 2 seconds to 3 minutes, according to the vessel's behaviour. Since the 31st of December 2004, AIS is compulsory for all vessels of 300 gross tonnage and upwards engaged on international voyages, cargo vessels of 500 gross tonnage and upwards not engaged on international voyages and all passenger vessels irrespective of size.

Class A systems	
Ships Dynamic Conditions	Reporting Rate
At anchor or moored	3 minutes
0–14 knots	10 Seconds
0–14 knots and changing course	3½ seconds
14–23 knots	6 seconds
14–23 knots and changing course	2 seconds
Faster than 23 knots	2 seconds
Faster than 23 knots and changing course	2 seconds
Class B systems	
Ships Dynamic Conditions	Reporting Rate
0–2 knots	3 minutes
Above 2 knots	30 seconds
Other AIS sources	
Source	Reporting Rate
Search and Rescue (SAR) aircraft	10 seconds
Aids to navigation	3 minutes
AIS base station	10 seconds or 3.33 seconds, depending on operating parameters

Table 1. The AIS reporting rate depends on the vessels behaviour

The vessels which are required to carry AIS are equipped with Class A AIS transponders, whereas other vessels can carry either Class A or Class B AIS transponders. The Class A transponder has a more powerful signal and transmits messages more frequently than the Class B transponders; therefore, Class A transponders typically have a finer spatial and temporal resolution. Vessels which may carry Class B transponders include recreational vessels, fishing vessels, or small passenger vessels [11]

Ground-based and satellite-based AIS dataset have some considerable differences,

- Terrestrial-based AIS (Ter AIS): Terrestrial receivers are land-based stations which receive messages from vessels within their line of sight. Once the message is received, it is relayed via network connection to a computer for storage, processing,

and visualization. Typically, with optimal terrestrial receiver setup, messages from up to 40 – 60 nautical miles away can be received. Terrestrial receivers are land-based stations which receive messages from vessels within their line of sight.

- Space-based AIS (Sat AIS): Satellite receivers function similarly to terrestrial receivers by transmitting the received AIS message to a computer for data storage, processing, and visualization. Satellite receivers do not require line of sight; therefore, they have a large field of view (up to 5,000 km) [12].

Apart from AIS receivers installed on coastal facilities or satellites, AIS receivers may also be mounted on oil rigs, wind turbines and other offshore structures, in order to monitor the vessel traffic in areas away from the coast. Towards this direction, roaming stations are also used. A roaming station is essentially a vessel that is equipped with an AIS receiver and satellite internet connection and, thus, it is able to transmit not only its own AIS messages but also the ones that it receives from nearby vessels.

To date, several challenges remain to accurately producing AIS-based density maps, mostly due to the inherent limitations, inaccuracies and complexities of managing AIS data, as will be discussed later in this study.

3.2 Data Fitness

The concept of data quality is somewhat vague, but an effective and widely used definition for data quality is “fitness for use”, which is the ability of the data collected to meet user requirements[13]. In this context, specific applications may allow some imprecision based on their requirements. AIS datasets have long been used for maritime density maps and researchers have identified some of the underlying difficulties affecting the data fitness [14]–[16]

Although AIS datasets are one of the most important sources of information for maritime traffic, the resulting spatial trajectories may have several missing data points, due to several reasons including design features.

- I. Firstly, datasets that have been collected by Satellites and those by Terrestrial stations will have different granularities and resolutions. Earth orbiting Satellites collecting AIS messages are easily congested when there is a large number of vessels within their given field of view. AIS is based on the Time Division Multiple Access (TDMA) radio access scheme which ensures that no two ships within radio range of each other are transmitting at the same time. The TDMA defined in the AIS standard creates 4,500 available time-slots in each minute but this can be easily overwhelmed by the large satellite reception footprints and the increasing numbers of AIS transceivers, resulting in message collisions, which the satellite receiver cannot process. Schemes such as the TDMA were designed for successful ship-to-ship or ship-to-shore communication, not for shipto-satellite communication which heavily degrades their efficiency[17]. However, in the case of the satellite segment of the AIS, the efficiency of implemented access schemes is heavily degraded due to the high ratio of the AIS packets collisions.

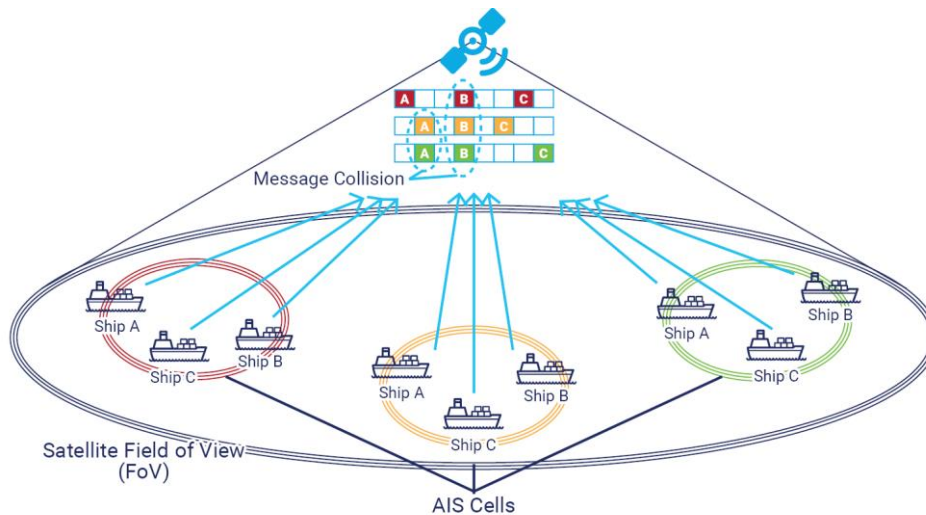


Figure 4. SAT field of vies and AIS self organising cells

- II. Additionally, according to the AIS specifications, Class A transceivers reserve their time slots for transmission via Self Organized Time Division Multiple Access (SOTDMA). After performing a scan to ascertain which slots have already been reserved by other vessels, they reserve an empty slot. The device lets nearby AIS devices know that it intends to use this slot for future broadcasts. On the other hand, Class B transceivers are permitted to transmit via Carrier Sense Time Division Multiple Access (CSTDMA). Unlike SOTDMA, slots are not reserved. They instead simply scan for available space and transmit when a free one is determined to be available. Transmission priority is given to Class A transceivers which use SOTDMA since they reserve time slots. The timing of Class B transmissions via CSTDMA must work around the time slots reserved by Class A transceivers. If a Class B transceiver is unable to find an empty space, their transmissions are delayed.
- III. Furthermore, Class B and Class A transmitters fitted aboard vessels have a critical difference which also affects Satellite reception. According to the ITU specifications, provision should be made for two levels of nominal power (high power, low power) as required by some applications. The default operation of the AIS station should be on the high nominal power level. The two power settings should be 1 W and 12.5 W or 1 W and 5 W for Class B “SO”. Evidently, the weaker signal of Class B devices means it is more difficult to receive these signals from space.

Additionally, Sat AIS cannot capture at once the signal of all transmitting vessels; several orbits are required in order to capture a representative density sample. In previous studies [1] the data compilation approach for acquiring information relevant to vessel-population, was based on the generation of vessel position ‘snapshots’ in a specified time window. As the duration of this window increases, the amount of information gained in terms of distinct vessels, newly detected by the available sensors, tends to decrease.



Figure 5. Issues affecting Sat AIS reception

On the other hand, coastal reception from the terrestrial receivers is only possible if a vessel is within the line of sight or approximately 40 NM in ideal conditions, affected by bad weather and other conditions[18].

Besides its limitations, AIS remains the go-to-source of data for understanding maritime activities and for this several works have focused on identifying the issues[18]–[20] and developing techniques that improve the existing AIS by offering better tracking accuracies and guarantees.

Definition (Incomplete Trajectory): Given the sparse spatial data $p_0(x_0, y_0, t_0)$, $p_1(x_1, y_1, t_1)$, $p_3(x_3, y_3, t_3)$, $p_5(x_5, y_5, t_5)$, $p_7(x_7, y_7, t_7)$, $p_8(x_8, y_8, t_8)$ of a moving ship, consisting of its time stamped locations, the resulting trajectory can be defined as

$$\sum traj_2 = \underline{p_0p_1}, \underline{p_1p_3}, \underline{p_3p_5}, \underline{p_5p_7}, \underline{p_7p_8}$$

It must be noted that $\sum traj_1 \neq \sum traj_2$

The aim of all techniques presented in the following section is to reduce the difference between the two trajectories so that

$$\Delta = \sum traj_1 - \sum traj_2$$

3.3 Density maps

To support improved understanding of this data, density maps are commonly used as an information visualisation method. The added value of density maps is that they support a bird’s eye understanding of vessel traffic, through providing an overview of vessel behaviour either at a regional or global scale.

The term “vessel density” has several connotations and thus is used with several meanings in this domain. Therefore, vessel density can refer to

1. the average number of vessels detected within a defined geographical area (spatial grid) in a given timeframe;
2. the average number of crossings within a defined geographical area (spatial grid) in a given timeframe (often also referred to as “vessel traffic density”);

3. the total vessel presence times within a defined geographical area (spatial grid) in a given timeframe;

There is a considerable difference in the methods used for the creation of density maps according to the definition used, including calculations based on the number of vessel positions detected, the number of vessel tracks, their length crossing a given area and many more variations.

The European Marine Observation and Data Network (EMODnet) enables analysis of this impact by disseminating monthly information on vessel positions on a 1km by 1km grid. It distinguishes between vessel use - cargo, dredging or underwater operations, high-speed craft, fishing, military and law enforcement, passenger, pleasure craft, sailing, service, tanker, tug and towing, other, unknown.

The EMODnet method used for 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.” [14]. In the EMODnet method, density is expressed in hours per square kilometre per month, since vessel trajectories are reconstructed from received position points³. This method assumes that there is a unique identifier available for each vessel. On the other hand, EMSA’s density maps count the number of routes that cross each cell of the same grid.

³ Throughout this text we refer to “density maps” as constructed by applying the EMODnet method, unless otherwise stated.

3.4 Data sources and raster-based spatial analysis

ABSTRACT

A fundamental activity associated with spatial analytics involves the reconciling of multiple data sources to the same or compatible geo-referenced locations[21]. In determining the fitness of a dataset for further analysis, we must homogenise available datasets across a common coordinate space, so as to be able to perform a comparative in-depth analysis of available sources. Towards this direction a common method is raster or grid-based analysis. The introduction of a grid-analysis frame provides a framework for storage and processing of different spatial datasets across a common base. Its base spatial unit is a “cell” defined by the column and row coordinates of an imaginary grid superimposed over an area. The grid cell and is used to statistically characterize[22]:

- points as individual cells,
- lines as connected sets of cells,
- polygons as all cells identifying the edge and interior of discrete parcels
- a surface as all of the cells within a project area with a value assigned to each that indicates the presence by feature type (discrete object) or the relative variable response (continuous gradient).

Raster operations performed on multiple input raster datasets generally output cell values that are the result of computations on a cell-by-cell basis. The value of the output for one cell is usually independent of the value or location of other input or output cells.

The rest of this section describes the steps taken towards creating a common coordinate space (grid), used for the subsequent analysis and the methods used for transforming each dataset into geospatial resources, describing ship detections that can be used for the spatial analysis. Specifically, this section describes the work conducted in “Task 1 Software development and evaluation for assessment of marine waters” and subtasks “[T1.1] Acquisition and processing of AIS data” and “[T1.2] Vessel detection from spaceborne optical and SAR images”.

4. Data sources used in this study

The data used in this study involves

- Automatic Identification System data from terrestrial, satellite and roaming stations. This dataset covers the months from the 5 October 2021 until the 31 of December 2021
- EO imagery optical (Sentinel-2) and SAR (Sentinel-1 images). We outline a process which will convert these images into sets of vessel detections for further analysis.

4.1 Area of interest

The consortium was mandated to create vessel density maps for European waters thus Baltic, Black, Irish, Mediterranean, North East Atlantic and North Seas. The study regions cover both high and low density areas, areas for which Ter AIS, Sat AIS and EO datasets could be obtained.

An extend based on the EU Coastline is used (see Figure 6)⁴. This bounding box is used to select data from the MarineTraffic Global Dataset of Terrestrial and Satellite AIS. Only data contained within this bounding box is then used for the analysis.

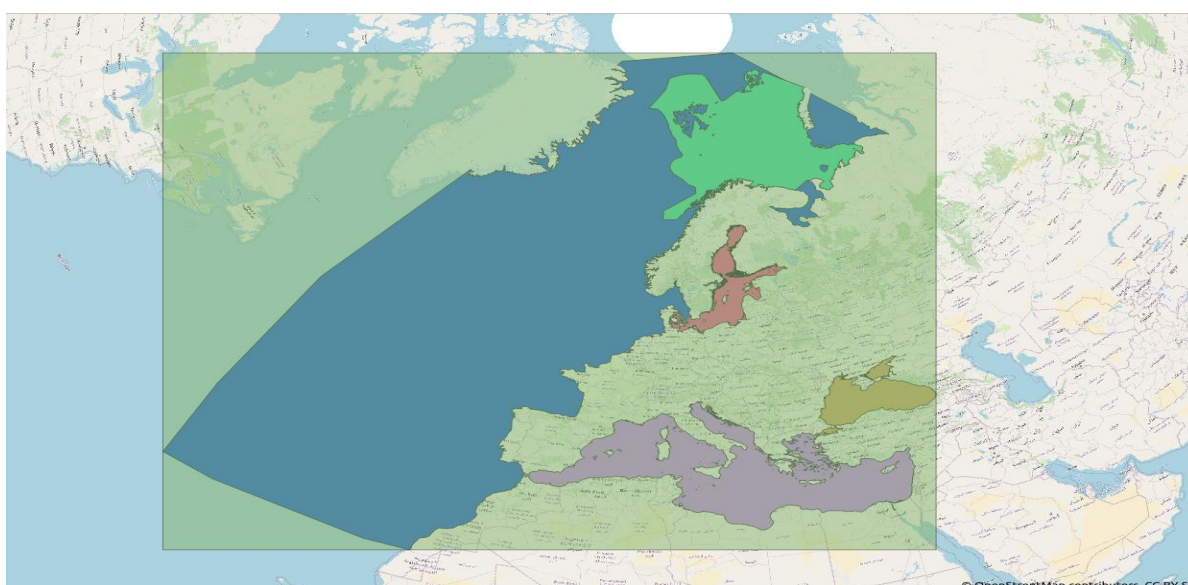


Figure 6. Geometry extent of area of interest. Seas (BAL, BLK, MED, BAR and ATL) are presented using different colors.

⁴ The geometries regarding the EU Coastline are provided by the European Environment Agency and can be found in the following address: <https://www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis-2>].

4.2 Creating a common grid

In order to perform the raster or grid based spatial analysis, the area of interest is partitioned into square cells of equal size (in meters), creating a corresponding grid; each grid is characterised by the edge length of its cells. While we focus on the European seas, we are able to work on the EPSG:3035 projection system [Footnote: see <https://epsg.io/3035> for more information.]. Although this projection system allows for easier navigation through the data and area partition, an additional step of removing all cells that do not include any water areas further improves the performance of our approach. The grids are created in several resolutions. The final grids for most lengths (i.e. 1km, 10km, 100km, 200km and 500km) can be seen in the following figure.

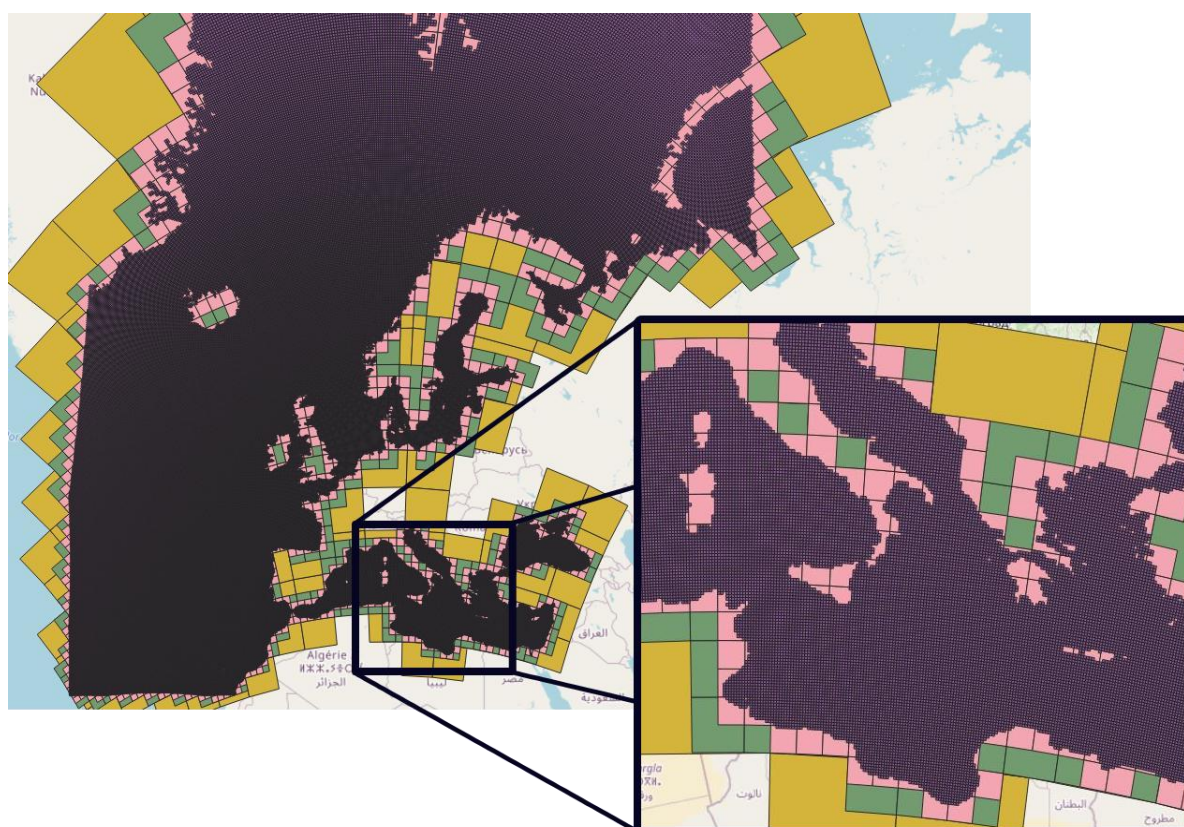


Figure 7. A separate grid is generated according to the size of each cell's edge length. Only cells that cover a sea area are kept in the final grid. Grids for edge length of 10km (purple), 100km (pink), 200km (green) and 500km (yellow) are presented.

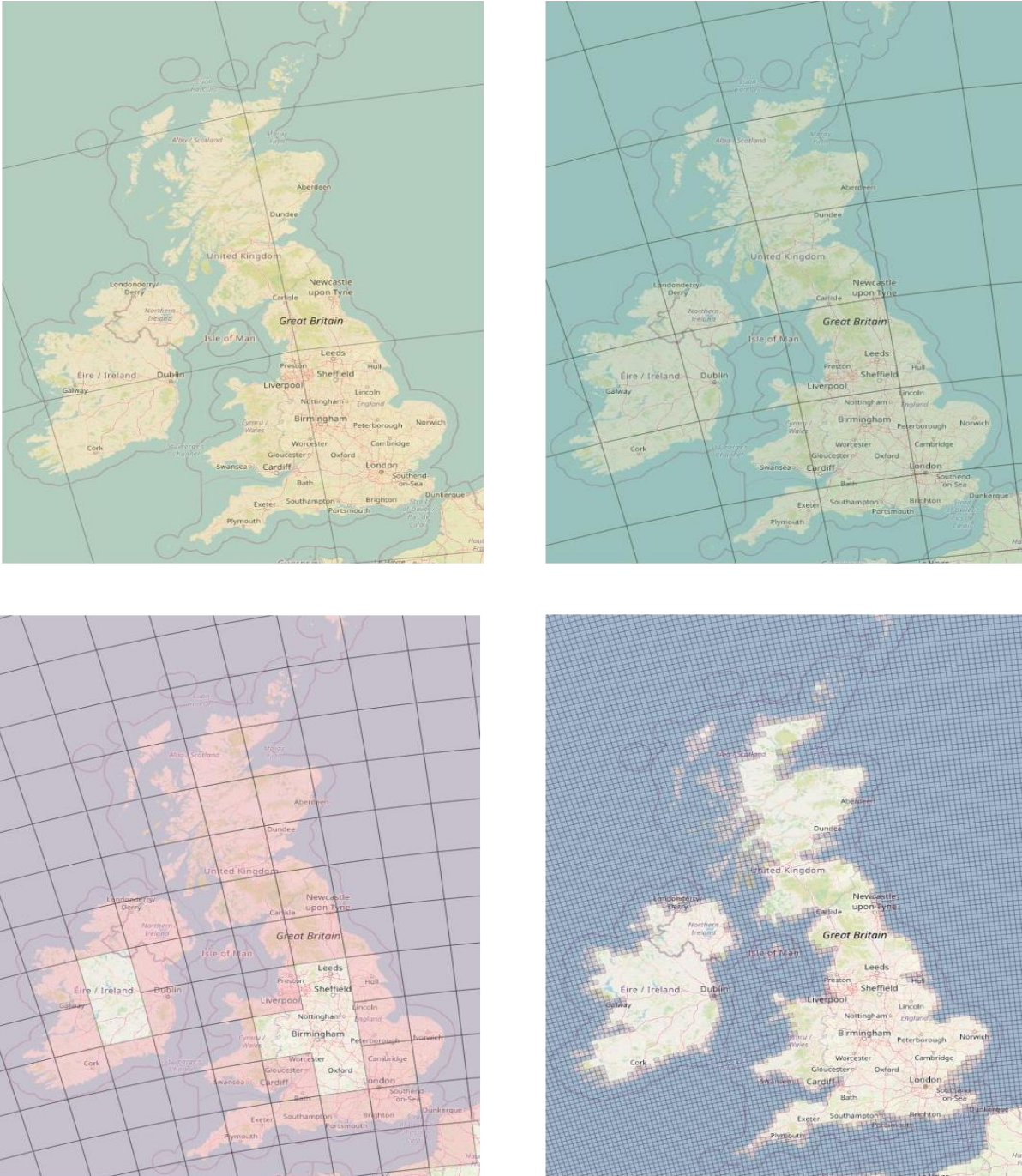


Figure 8. The grid cells that cover the area around the UK; cells of length 500, 200 (top), 100 and 10km (bottom) are presented.

4.3 Transforming the datasets

Transformations are simple methods of spatial analysis that convert the datasets by combining them or comparing them to obtain new datasets and eventually new insights. Transformations use simple geometric, arithmetic, or logical rules, and they include operations that convert raster data to vector data or vice versa. They may also create fields from collections of objects or detect collections of objects in fields[23]. In the following sections we

describe the work conducted towards transforming Ter AIS, Sat AIS and EO into a homogeneous format for further analysis.

4.3.1 AIS Transformation

4.3.1.1 Pre-processing and cleaning

As previously stated, trajectories are never perfectly accurate due to sensor noise and other factors. Removing erroneous and unnecessary messages from AIS datasets is a crucial component for further analysis of maritime movement. Several filters are implemented to smooth the noise and potentially decrease the error in the measurements.

These data come from different sources, either a satellite provider or terrestrial stations, resulting in multiple datasets. A two-step pre-processing is performed on each dataset. First, the positional data are merged with the corresponding static information regarding the vessel in question. Designed for a decoded input, our method matches the AIS positional data with the most recent vessel information needed. Static information appended at the end of each positional message includes the vessel type (e.g., cargo, tanker, passenger) and the vessel class ('A' or 'B').

Secondly, after merging the data, we proceed with the main part of our preprocessing, that follows criteria similar to EMODNET's approach for creating density maps[14]. These criteria refer to both features regarding the positions of the messages, as well as its movement fields. More precisely the filters applied to each input message are the following:

1. Empty fields: messages that monitor movement, like the AIS messages, may include a plethora of features. Besides primary features (positional and temporal features) that denote the exact position of the moving vessel, other fields regarding its characteristics or its current state are usually provided. For the purpose of effectively analysing the input data, we require that each positional message includes non-empty values in the following fields:
 - Positional fields:
 - Vessel Longitude.
 - Vessel Latitude.
 - Timestamp of AIS occurrence (expressed in UNIX time in milliseconds).
 - Movement fields:
 - Vessel Speed-over-Ground (SoG), measured in knots.
 - Vessel Course-over-Ground (CoG), measured in degrees.
2. Invalid movement fields: While most messages include the aforementioned fields regarding a vessel's movement (SoG, CoG), in some instances these fields carry invalid values. In such cases, the messages are characterised as erroneous and are discarded. The thresholds indicating valid movement for each filter is as follows:
 - Speed-over-Ground: real number between 0 and 80 knots,
$$\text{SoG} \in [0.0, 80.0]$$

- Course-over-Ground: real number between 0 and 360 degrees,
CoG $\in [0.0, 360.0)$

3. Invalid vessel ID: With each AIS message referring to a single vessel, a field dedicated to its identification is needed. The Maritime Mobile Service Identity (MMSI) convention is widely used while referring to AIS transmitters (i.e. vessels), with each single entry being a series of (9) characters. Moreover, a list providing some exceptions of invalid MMSI numbers is provided below. Messages whose MMSI does not follow these two criteria are discarded.

MMSI exceptions
000000000, 111111111, 222222222, 333333333, 444444444, 555555555, 666666666, 777777777, 888888888, 999999999
123456789, 0.12345, 1193046
Length more or less than 9 characters

4. Areas of interest: While our approach may be applied regardless of the area in question, defining the space of reference is a crucial part for moving forward for two reasons:
- Removing data that refer to areas outside the purpose of the execution scenario.
 - Removing data that include erroneous coordinates, i.e. not valid longitude / latitude or points on land.
5. Downsampling: Although the frequency where each vessel is transmitting a positional message is usually desired to be as high as possible, having too many messages may result in considerable delays while processing. In order to overcome this issue, downsampling is performed upon the input data. The question at this point is whether we are able to disregard some sample points without sacrificing the quality of the trajectory data required for the target application. For this purpose, the trajectories are filtered so that consecutive messages from the same vessel have at least (3) minutes between them, thus losing as little information as possible while at the same time getting rid of a great amount of data.
6. Time-frame: Restricting the time-frame where AIS messages are to be included in the end result may be useful for creating a custom dataset for analysis and removing messages with erroneous timestamps, due to noise. Additionally, it can be used for excluding messages referring to a time before the dataset's specifications, caused by delays during their transmission.
7. Noise-filter: In some cases, consecutive AIS messages of a single vessel indicate an invalid transition between two points[24], [25]. More precisely, if the distance between consecutive messages is large that it wouldn't be possible for a vessel to cover in the corresponding time frame, this transition is considered noise in our data and the second AIS message is removed. A transition is considered to be improbable if the

calculated speed for the vessel to cover the distance in question exceeds the threshold of 92 km/h (approximately 50 knots).

8. Insignificant tracks: All data regarding vessels that have less than 10 AIS messages after the merging step of our pre-processing are discarded.

For this study, we performed cleaning upon all input data using all the filters mentioned. As a result, for the Terrestrial AIS (Ter AIS) 98.16% of the messages were removed; with 78.07% for the Satellite messages (Sat AIS) accordingly. The effect each filter had on the end results can be seen in following Figure 9:

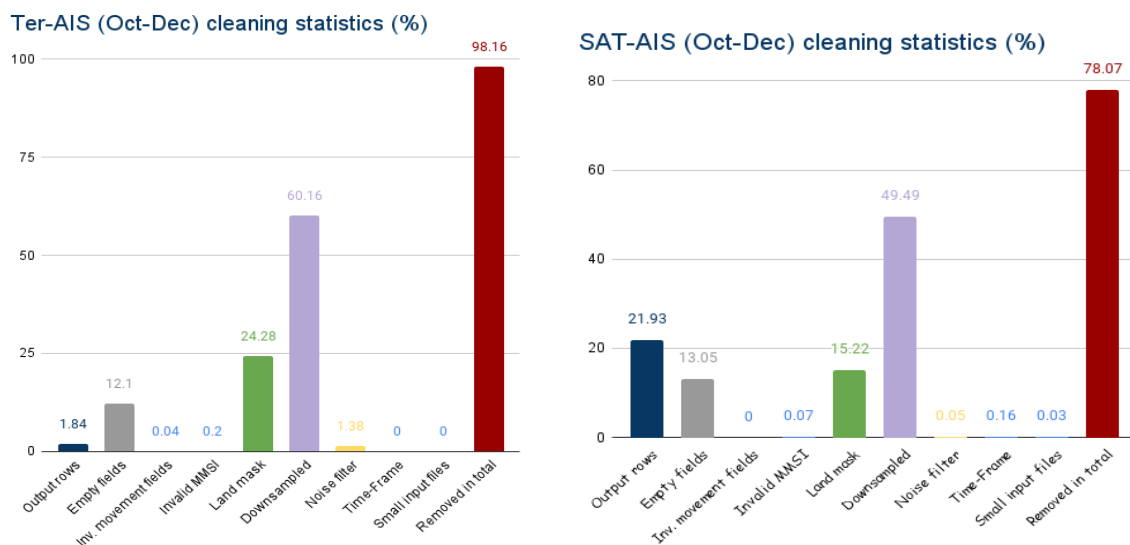


Figure 9. Statistics for data cleaning of Ter AIS and SAT_AIS

Most of the aforementioned criteria require some parameters whose value we have manually selected,. In the public release of the software the user can selected these values arbitrarily. Although the input messages are using the longitude/latitude (4326) projection system, a transformation to the (3035) projection system is also performed for all messages.

4.3.1.2 Output Data

The output of the vessel detection task is a.csv file with the following information

Column	Description and notes
TIMESTAMP	Unix epoch in milliseconds where the AIS message occurred
MMSI	Vessel unique identifier (id)
LON	Vessel longitude
LAT	Vessel latitude

X	Vessel X-positioning - 3035
Y	Vessel Y-positioning - 3035
HEADING	Vessel heading $\in [0, 360)$
COURSE	Vessel course over ground $\in [0, 360)$
SPEED	Vessel speed over ground $\in [0, 80]$
NAVIGATIONAL_STATUS	Navigation status according to AIS Specification , $\in [0, 15]$
TYPE	Vessel type $\in [0, 100]$ (other values indicate unknown/unregistered type)
STATION	Depending on Sat AIS, Ter AIS receivers
CLASS	'A' or 'B', based on AIS type message received

In terms of granularity and accuracy of the Ter AIS the figures below show the number of vessels detected daily for each sea (Figure 10) and the number of total positions received each day (only for CLASS A).

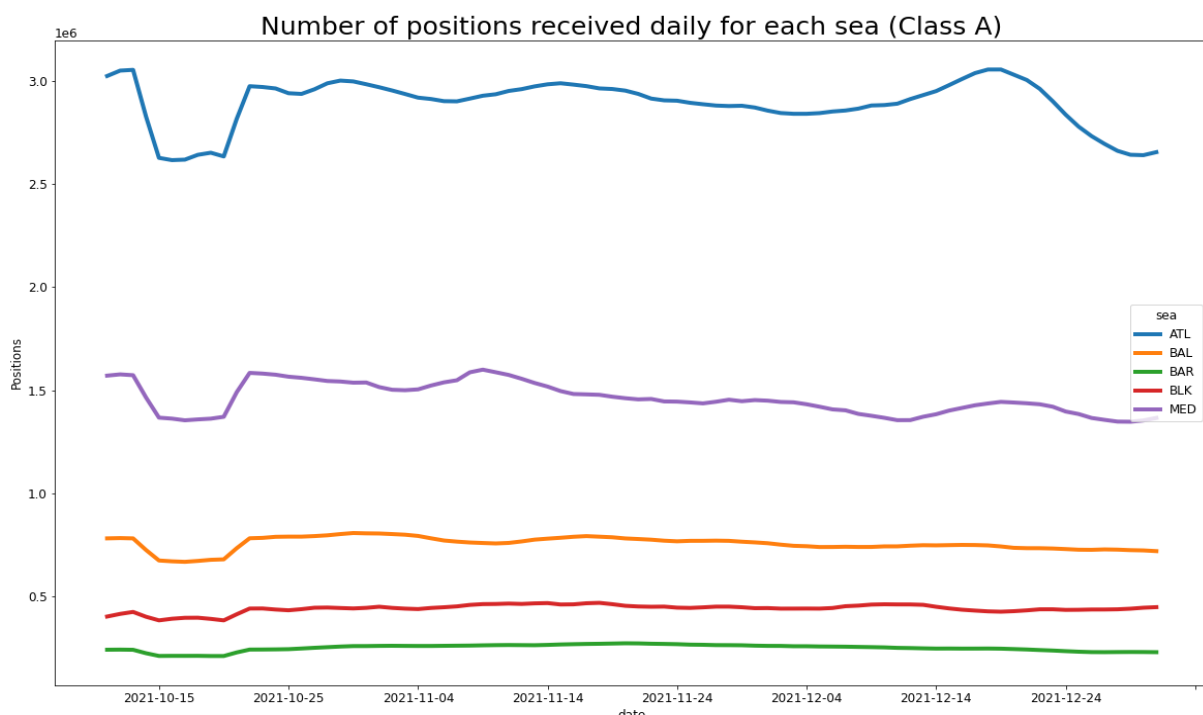


Figure 10. Number of unique ships detected daily for each sea (7-days moving average).

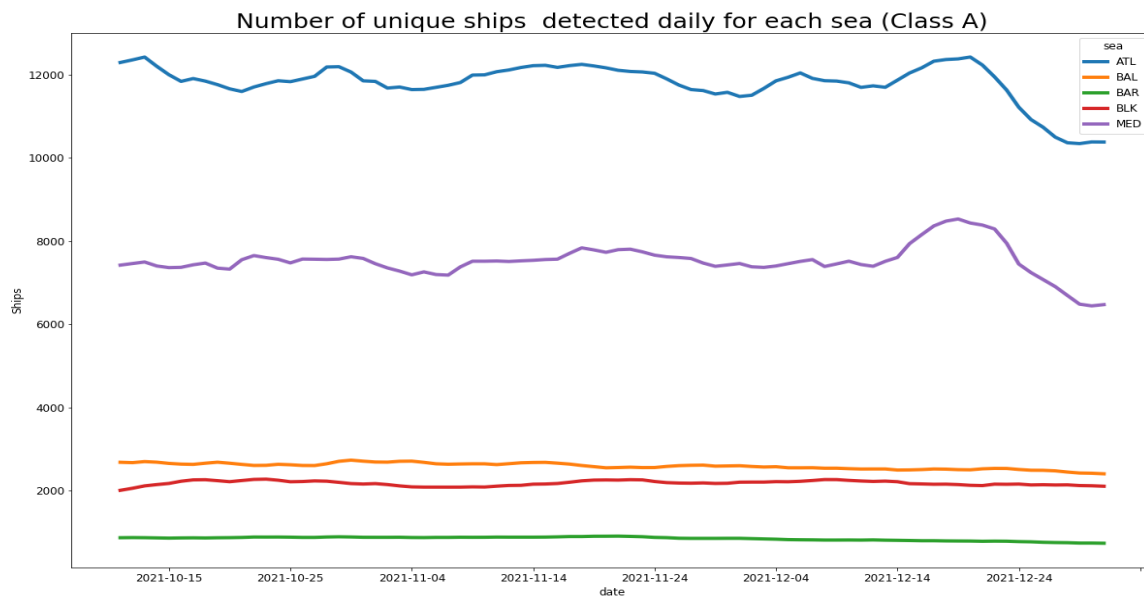


Figure 11. Number of positions received per day for each sea (7-days moving average).

4.3.2 Transformation of EO images

This task involves the operations of acquiring EO imagery, processing and detecting annotated vessels.

In the related literature, the problem of vessel detection in satellite imagery has been the focus of many academic and industrial research activities. These activities can be classified into two broad categories: (a) The approaches that are based on the employment of thresholding-based algorithms, such as the CFAR algorithm[26] (especially for SAR imagery, such as the JRC tool SUMO and others[27]), and (b) AI-based approaches that employ Neural Networks in order to detect vessels based on trained models[28]. [29] presents a detailed, up-to-date overview of approaches for vessel detection in optical imagery covering both categories. The processing chain that we used to perform vessel detection in satellite imagery for the needs of this project is based on our previous work, that belongs to the AI-based category and is partially covered in [30]. However, in order to address the known scalability problems that come with the use of NNs in large volumes of satellite data[31], we further extended our approach described in [30] and developed a hybrid approach that employs both threshold-based mechanisms and CNNs, in order to get the best of both worlds: High accuracy results provided by CNNs while addressing scalability issues by using thresholding to filter out redundant image tiles. Part of the work described in this section is covered in a recent publication [32].

A high level overview of the technology stack developed for vessel detection in satellite imagery is given below

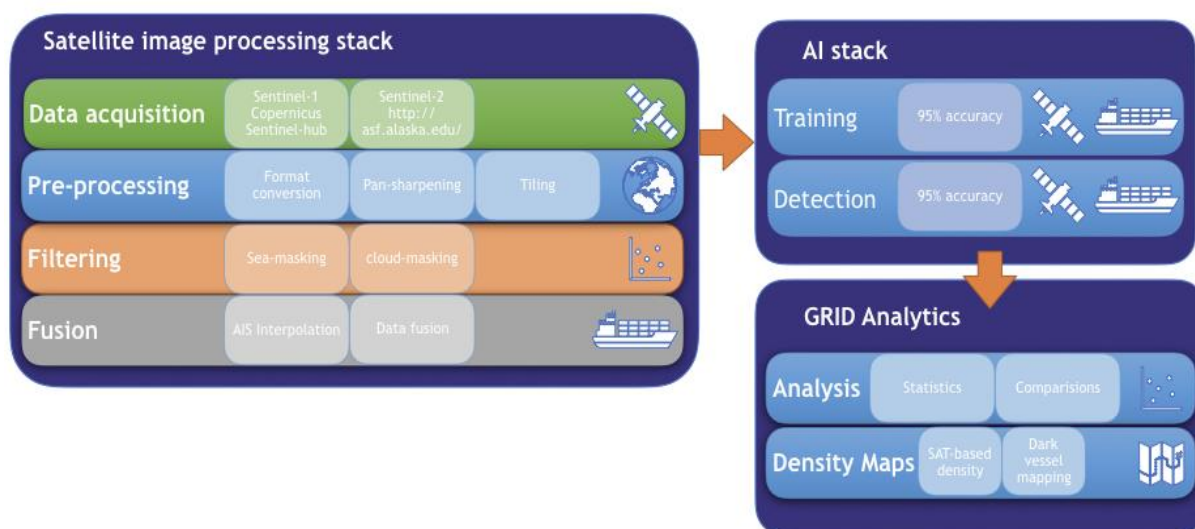


Figure 12. The EO detection high level workflow

In the rest of this section, we describe our approach for vessel detection in optical (Sentinel-2) and SAR (Sentinel-1) images. A high-level overview of our system that automatically detects vessels in satellite imagery is depicted in the figure provided above. It consists of the following components:

4.3.2.1 Data acquisition

We acquire Sentinel-1 and Sentinel-2 data from the Copernicus open access hub using the SentinelSat python API⁵. We also use the Alaska Satellite Facility repositories for Sentinel-1 imagery⁶ as a back-up repository. We download Sentinel-1 and Sentinel-2 data per month in a rolling way, starting from October 2021. For Sentinel-1 data, we download IW GRD products coming from both S1A and S1B satellites, and for Sentinel-2 data, we download all bands available for S2A and S2B products.

4.3.2.2 Pre-processing

After we download the satellite images, we have to preprocess them. Since optical images are very different from SAR images, we perform different pre-processing steps for each kind of imagery. For SAR imagery, we perform corrections, such as converting SAR geometries to geo-referenced geometries using GDAL tools in python (e.g., gdalwarp). SAR images come as single band images, while Sentinel-2 images come in RGB bands together with an infrared (IR) band, in three different resolutions: 10m, 20m and 60m. For Sentinel-2 images, we stack the RGB and the SWIR bands together and we compose a panchromatic image in the GeoTIFF format, using 16-bit encoding, and we perform pancharpening to increase its resolution.

⁵ <https://scihub.copernicus.eu>

⁶ <https://asf.alaska.edu>

The panchromatic image has been created by computing the average of all the 10m resolution bands, that is RGB and NIR in order to obtain a single high-resolution image. The aim of the pan-sharpening technique is to fuse the higher spatial information from the panchromatic image and the spectral information from a lower spatial information multispectral image.

A true colour image (TCI) in 10m is readily available in a Sentinel-2 image directory, however, we prefer to compose the panchromatic pan sharpened image using the individual bands since the TCI image comes in 8bit encoding, which is lossy. Maintaining a 16-bit encoding will be crucial for the filter stage which follows.

TILING

After we have pre-processed the images, we divide them into smaller tiles in order to be further processed more easily. The dimensions of the tiles are currently 256x256.

FILTERING

Since we use an AI-based approach for detecting vessels in satellite imagery, i.e., we use a neural network, feeding thousands of satellite images which correspond to several TBs of data into the network can turn it into a serious bottleneck and it will compromise the performance and applicability of our approach in a density maps use case, in which thousands of satellite images need to be processed. Thus, we filter out the image tiles which are redundant, i.e., for which we have indications that they do not contain vessels before we feed them to the neural network. For example, a Sentinel-1 tile that is totally black depicts the sea and no object appears in it. For Sentinel-1 image tiles, we use statistics and thresholding (i.e., amount of black/white pixels), while for Sentinel-2 image tiles, we use thresholds that are based on the difference of pixel values between the red band and the infrared band (R-SWIR). We also use the ACL as a mask to filter out clouds.

Eventually, the filtering step results in image tiles sizing up to at least one order to magnitude less than the size of the original image.

4.3.2.3 AI-based vessel detection

For the vessel detection task, we trained a neural network based on the state-of-the-art vessel detection framework YOLOv4, with vessels detected in satellite images, using AIS data as ground truth. Since we also want to detect the precise location of vessels, we used an object detection framework instead of a simple CNN. The object detection framework deploys a CNN per image pixel and determines the bounding box of the detected object. Our trained model is able to detect vessels with 92% accuracy as shown in the experiment results provided in Table 2.

Metric	Value
Precision	92%
Recall	93%

F1-score	92%
True Positive (TP)	80%
False Positive (FP)	7%
False Negative (FN)	6%

Table 2. Vessel detection experiment results using 400 semi-manually labelled sentinel-1 and sentinel-2 satellite imagery. 80% of images in the benchmark dataset were used as training set and 20% of the images were used as test set.

4.3.2.4 Post-processing

The output of the vessel detection task is a.csv file with the following information: (a) the bounding box of the detected vessel (with respect to the centre of the image), (b) the class of the detected object (e.g., vessel, tanker, etc.), (c) and a confidence value that indicates the possibility of the detected object belonging to the detected class. In a post-processing task, we (i) retain only the detections with high-confidence (> 0.6), and we geo-reference the coordinates of the detected objects by transforming them from topological coordinates to geo-referenced coordinates, so that they can be correlated with the respective AIS coordinates. The output of this task is a.csv file per image that contains the image tile that the detected object belongs to, its geo-referenced location (the geo-referenced centre of its bounding box), and the timestamp, which is the acquisition time of the image. We produce one.csv file for every detected vessel and one.csv file per vessel type (e.g., Tanker, Cargo and Tug for Sentinel 1 imagery). All .csv files that are produced as output of this step follow the format described in Table 3.

Column	Description
path	The path of the actual image tile in which the object (vessel) was detected
confidence	Confidence of detection
lon	Vessel longitude (calculated)
lat	Vessel latitude (calculated)
timestamp	Textual datetime representation of detection time in UTC (acquisition time of the satellite image that was used to detect the vessel)

Table 3. Format of post-processed satellite image detections

4.3.2.5 Data fusion

After we detect vessels in satellite images, we correlate this dataset with the respective AIS positions acquired by MarineTraffic. The correlation task involves the following steps:

- Spatio-temporal filtering. The temporal resolution of satellite images is significantly lower than the temporal resolution of AIS messages (i.e., the revisit time of Sentinel satellites is 2-3 days in high coverage areas whereas vessels with AIS transponders transmit AIS messages every few seconds or minutes, depending on their navigational status and speed). In order to be able to correlate these two data sources, for every image, we extract all AIS positions that are located into the area covered by the image during a 1-hour time window, spanning 30 minutes before and after the image acquisition time. For the filter, we create a temporal index on the geo-dataframe where we load all AIS positions, filtering out all positions that fall out of the time window and then we perform spatial joins that retains only the positions that are covered by the spatial extent of the image.
- Interpolation. Then, we create trajectories for each vessel contained in the dataset. For each trajectory, we retrieve the position of the vessel at the time the image was acquired by interpolating the vessel's locations before and after the acquisition time of the image. The output of this step is a snapshot of all AIS vessel positions which spatio-temporally intersect with the image. More specifically, we produce a.csv file for each image storing the position of every vessel located in the spatial extent of the image footprint at the time the image was acquired.
- Fusion. The fusion task matches the interpolated AIS-positions of the previous step with the vessels detected in the image that are the output of the vessel detection step and the post-processing step. Since the AIS dataset contains an identifier for each vessel, the fusion task assigns each detection to a vessel position. We perform a KNN-join between the two datasets in the following way: For each vessel detected in a satellite image, we search for the nearest AIS neighbor (i.e., nearest interpolated position of a vessel). To achieve this, we store the interpolated positions that are the output of the previous step in a KD-Tree in order to speed up the distance joins. Vessels detected in satellite images that do not have matching AIS detections, given a distance threshold, are considered as "dark vessels", i.e., they are either located in a low coverage area or they have intentionally switched off its transponder. For each image, the fusion step creates the following files:
 - a) A.csv file that contains the "dark vessels", i.e., the vessels detected only in the satellite image
 - b) A.csv file that contains the "matched vessels", i.e., the vessels that were detected both in AIS and satellite imagery.
 - c) A.csv file that contains the "AIS-only vessels", i.e., the vessels that were only detected in AIS. A vessel can only be visible via AIS and not in a satellite image due do the following reasons:
 - I. The resolution of Sentinel-1 and Sentinel-2 imagery does not allow for the detection of small vessels with high confidence.
 - II. The AIS position of the vessel might be wrong. Since AIS is a collaborative maritime reporting system, a vessel's crew might alter the GPS position of the vessel when transmitting the AIS messages (e.g., spoofing).

- III. Interpolation error: When the AIS messages that we have around the acquisition time of an image transmitted by a vessel are not enough, and the vessel has changed its navigational status in the meantime (e.g., a vessel suddenly stops or it accelerates and changes heading), the estimated position of the vessel in the interpolation step might differ considerably from the actual position of the vessel.

Vessel detection examples

In the following, we provide some examples of the vessel detection workflow in Sentinel-1 and Sentinel-2 images of October 2021. Each image tile is annotated with the confidence of vessel detection. In a post-processing step, we only keep the detections with threshold > 0.6 . The model was trained on 2019 data and it is able to distinguish vessels from (i) clouds (in the case of Sentinel-2), waves (ii), and (iii) land parcels. In the first row of the set of images that are provided below, we show some positive examples. In the second row, we show some examples that are either true negative (no detection) or false positive, but below the confidence threshold that we have set. One of the advantages of the DL-based methods is that the model can be trained to detect vessels regardless of the background in contrast to the threshold-based methods. This is mostly highlighted in the optical (Sentinel-2) images.

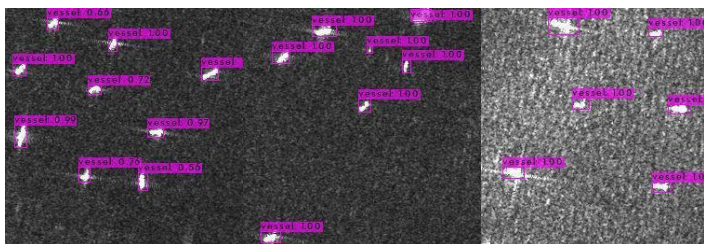


Figure 13. Positive examples of vessel detection in Sentinel-1 tiles of October 2021 with high confidence (in most cases)

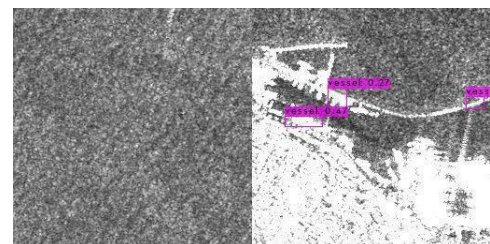


Figure 14. True negative examples (a) and detections below threshold (< 0.6) (b)



Figure 15. Sentinel-2 vessel detections with high confidence

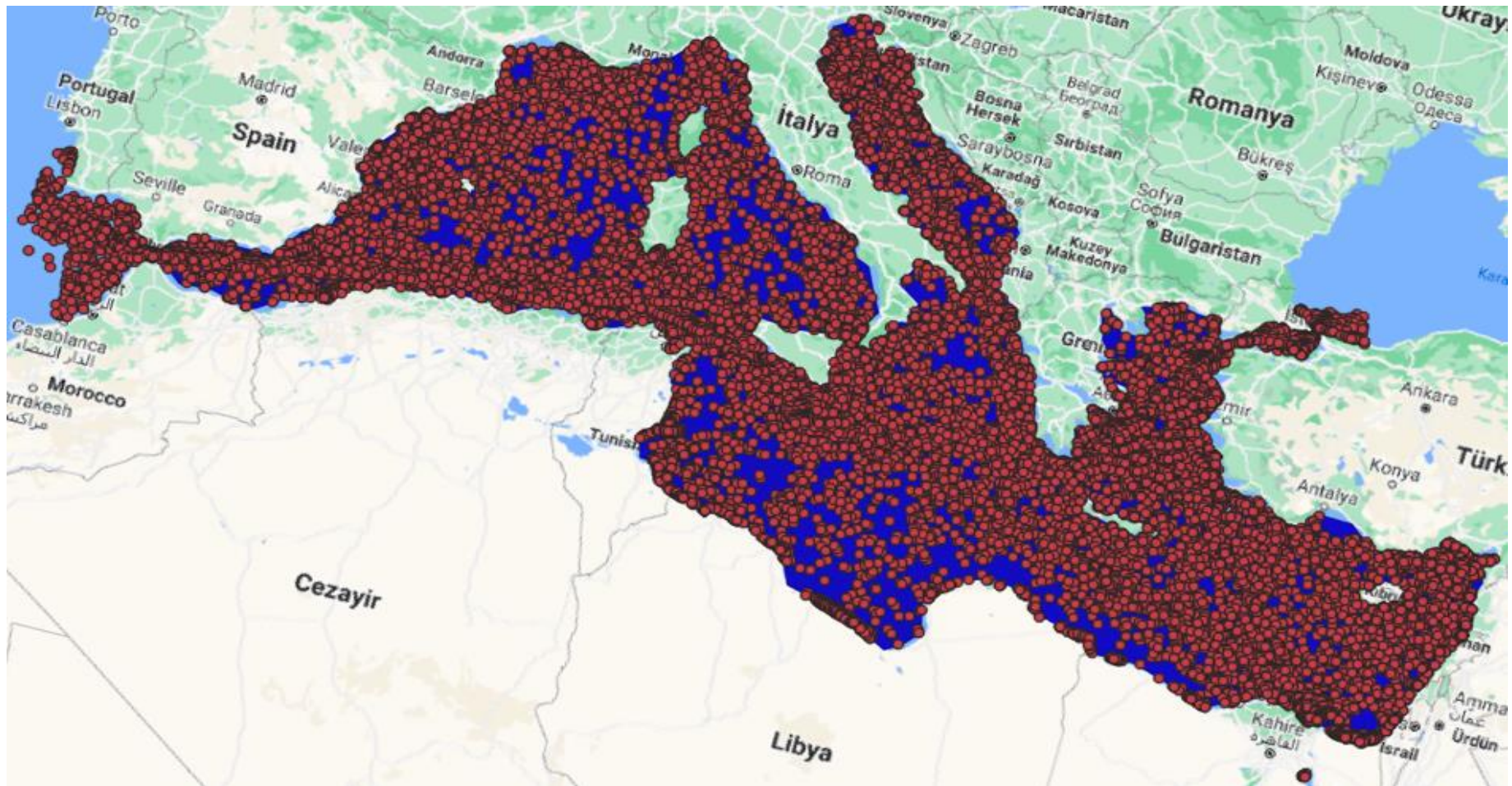


Figure 16. Sentinel-1 Detections in MED sea during October 2021

The figure provided above shows the vessels detected in the Mediterranean Sea during October 2021 using Sentinel-1 imagery.

5. Spatial Analysis and Comparisons

ABSTRACT

The objective of this task is to use the outputs of Task 1 and perform a qualitative and quantitative analysis in order to compare the computed vessel density per grid cell for each data source used. In general, raster data analysis is based on cells and rasters. Raster data analysis can be performed at the level of individual cells, or groups of cells, or cells within an entire raster. In our study we use both approaches so as to measure differences per cell but also in the entire region.

This section reports on the results from “Task 2: Analysis” and specifically the work performed in the tasks: [T2.1] Comparison between vessel density results from Sat AIS and SAT imagery processing; [T2.2] Comparison between different AIS Receivers on satellites and [T2.4] Comparison with receivers mounted on board vessels and other offshore facilities.

5.1 Comparison between different Ter AIS and Sat AIS

In this section we assess the capabilities of Sat AIS and Ter AIS vessel detection systems, both quantitatively and qualitatively.

Where possible we compare performance measuring,

1. the number of unique vessels received in a defined time interval, both at a regional scale and across EU waters
2. the average distance between two sequential AIS messages of a vessel for the three Sat AIS providers per day.
3. the average distance between two sequential AIS messages of a vessel for Sat AIS and Ter AIS per day.

These indicators specify in particular each platform’s capabilities in terms of capacity to detect the correct number of vessels in an area, but also the granularity of messages or detections. We must note, though, that due to differences in coverage it is not always possible to provide a 1-1 comparison for the entire study space, where possible we perform a per cell comparison. Additionally, we must note that comparisons are performed only on CLASS A messages. The capability of Sat AIS to identify a vessel strongly depends on the class of the device used. As seen in the table below, Sat AIS strongly underperforms in its capacity to detect CLASS B AIS.

Thus, throughout the rest of the analysis we focus only on CLASS A AIS devices, as this is a known issue for Sat AIS, that otherwise would strongly affect the results.

	sea	ATL	BAL		BAR		BLK		MED		
	metric	positions	ships	positions	ships	positions	ships	positions	ships	positions	ships
datasource	ais_class										
all-ais	%A	81.03	66.22	87.78	78.26	58.90	42.07	68.80	61.71	81.65	73.07
	%B	18.97	33.78	12.22	21.74	41.10	57.93	31.20	38.29	18.35	26.93
sat-ais-all	%A	98.29	85.89	100.00	99.91	96.88	52.33	99.98	99.89	99.99	99.81
	%B	1.71	14.11	0.00	0.09	3.12	47.67	0.02	0.11	0.01	0.19
ter-ais	%A	79.42	65.85	87.69	78.23	53.08	40.96	68.56	60.28	80.82	71.66
	%B	20.58	34.15	12.31	21.77	46.92	59.04	31.44	39.72	19.18	28.34

Figure 17. Number of positions and ships detected for all sea by AIS class from 5th October 2021 until 31 of December 2021. We present for each sea (ALT,BAL,BAR,BLK,MED) and for each provider type (all-ais, sat-all-ais, ter-ais) the distribution of AIS vessel classes for the received AIS position reports and the number unique vessels detected. For example the distribution of vessel classes for all messages received, regardless the provider (all-ais), is 81.03% class A and 18.9% class B in terms of position reports and 66.22% class A and 33.78% class B in terms of unique vessel detected for the Atlantic Ocean area(ATL).

The results of the computations are displayed in Figure 18 to Figure 24. The observed differences in the performance are closely linked to the different system concepts and design capabilities.

In sum,

- The comparison clearly indicates that terrestrial AIS detects a higher number of unique vessels across the given area with the exception of Barents Sea
- In several cases unique vessels detected from Ter AIS are several orders of magnitude more.
- The granularity and resolution of the TER_AIS dataset is much higher than that of Sat AIS both on a temporal and spatial scale.

In sum, where Ter AIS is available, Sat AIS is a subset of the Ter AIS dataset. Since the results are heavily affected by the Ter AIS coverage, it is important to examine the data at a cell level, so as to better understand the differences. This is performed in Section 6, with the detection difference map.

The brief interpretation of the data exemplifies the added value of combining datasets from both sources.

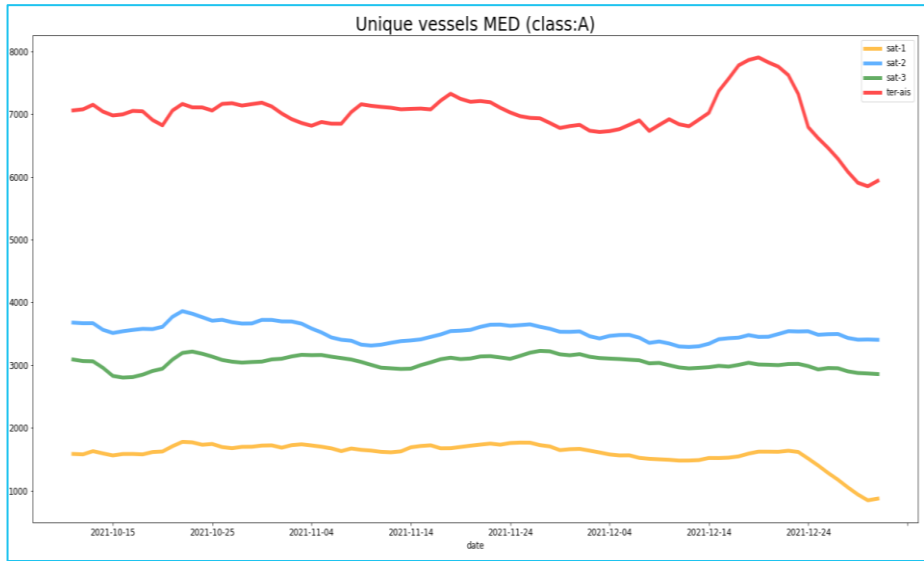


Figure 18. Detailed view of the 7 days average number of unique ships detected from each AIS provider for the Mediterranean Sea from 5th October 2021 until 31 of December 2021

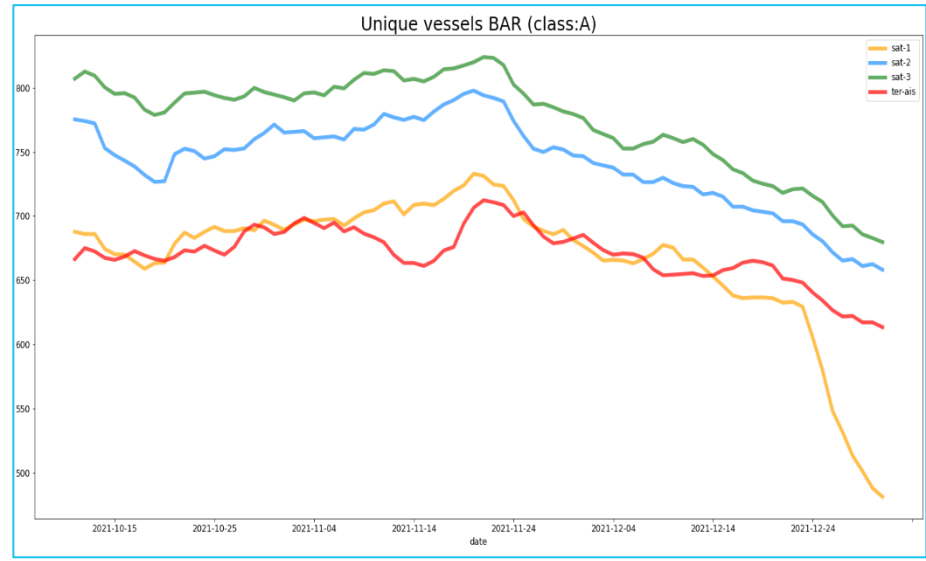


Figure 19. Detailed view of the 7 days average number of unique ships detected from each AIS provider for the Black Sea from 5th October 2021 until 31 of December 2021

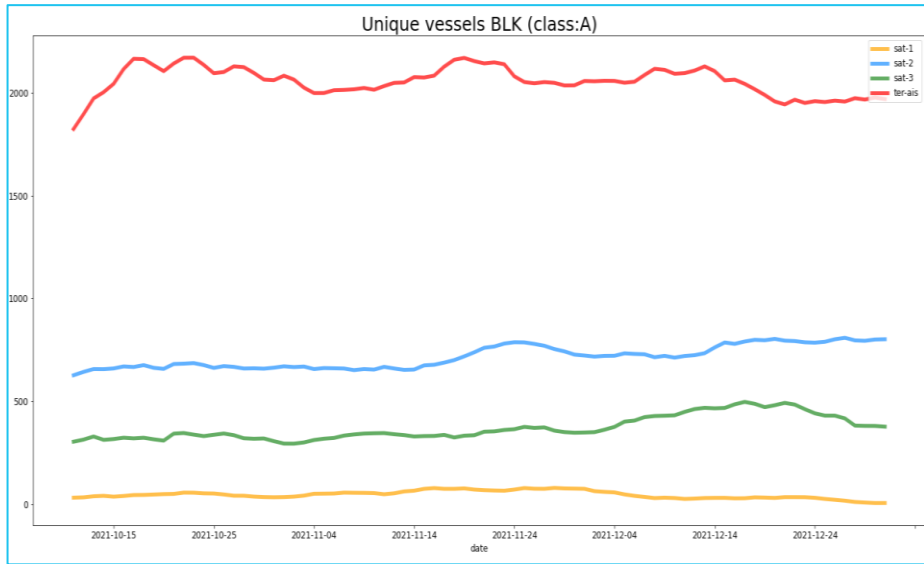


Figure 20. Detailed view of the 7 days average number of unique ships detected from each AIS provider for the Barents Sea from 5th October 2021 until 31 of December 2021

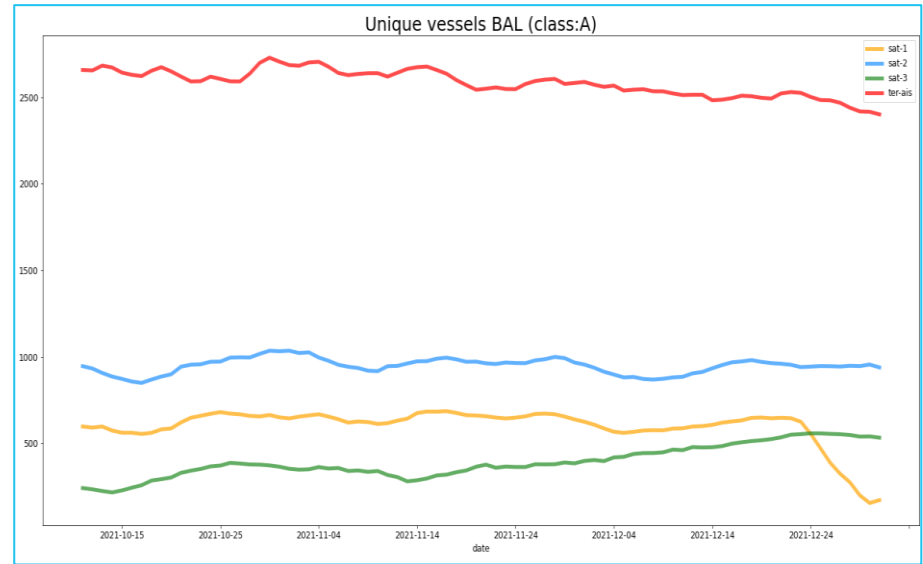


Figure 21. Detailed view of the 7 days average number of unique ships detected from each AIS provider for the Baltic Sea from 5th October 2021 until 31 of December 2021

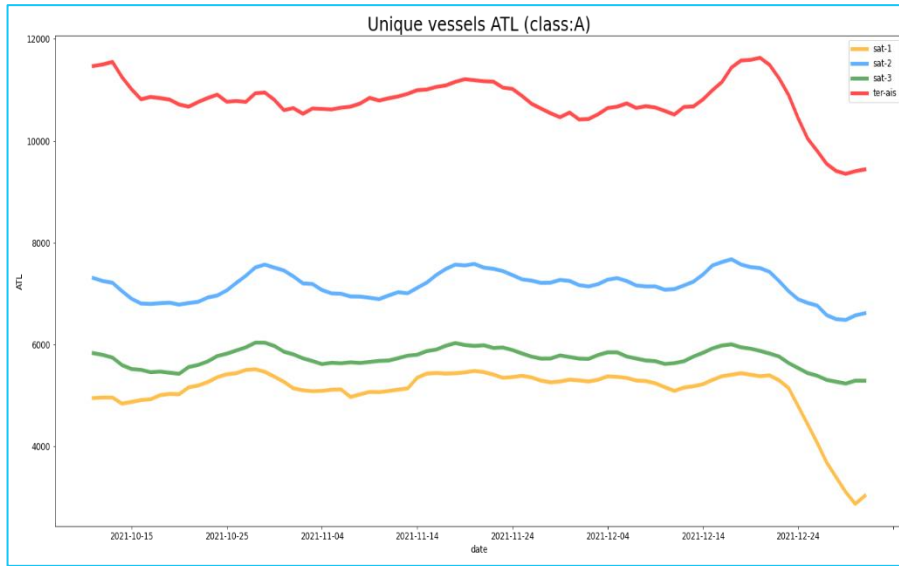


Figure 22. Detailed view of the 7 days average number of unique ships detected from each AIS provider for the seas of North East Atlantic, Irish Sea and North from 5th October 2021 until 31 of December 2021

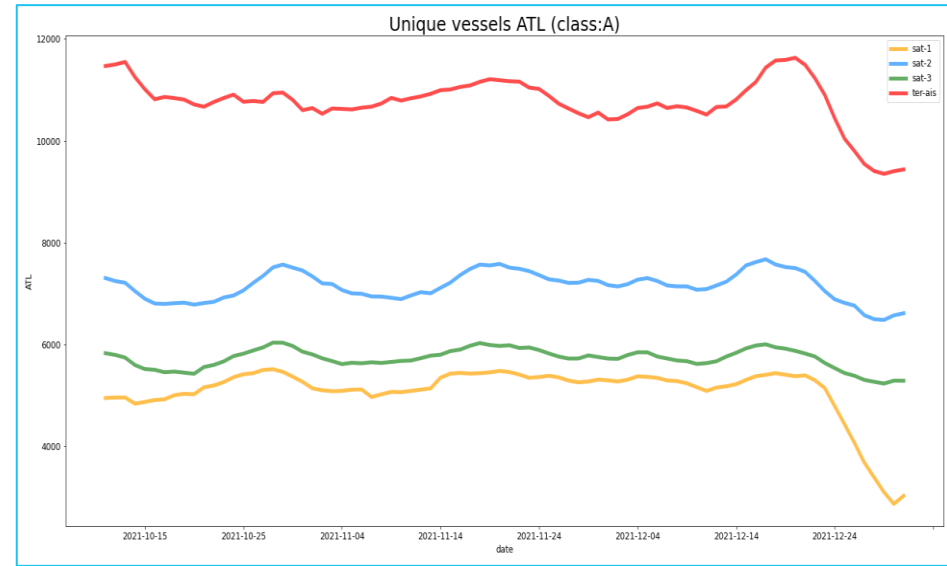


Figure 23. Comparison of average distance between two sequential AIS messages of a vessel for the three Sat AIS providers per day, time-series represent a 7 days moving average window.

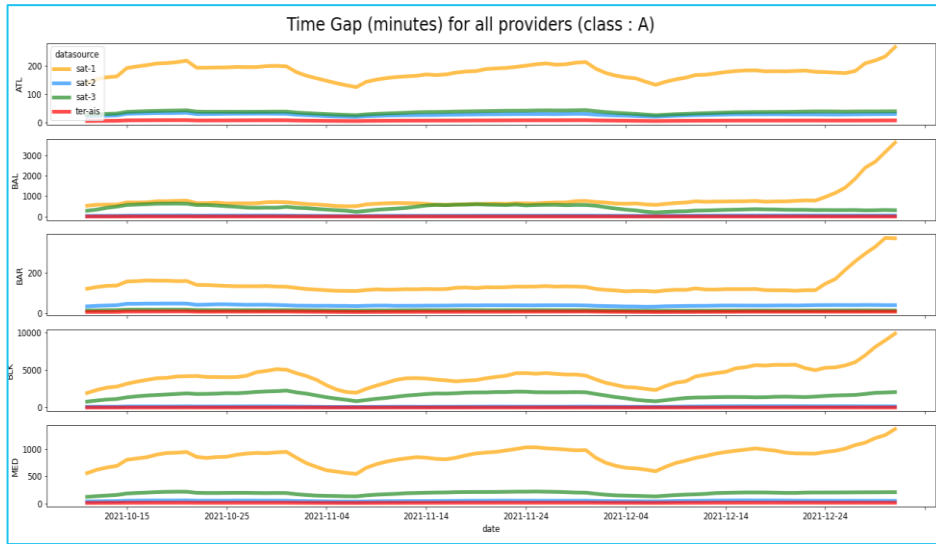


Figure 24. Comparison of average time gap between two sequential AIS messages of a vessel for the three Sat AIS providers per day, time-series represent a 7 days moving average window.

5.2 Comparison between Sat AIS, Ter AIS and SAT EO

The results of the computation are displayed in Figure 25 to Figure 27.

The following figure shows a comparison between the number of vessels identified by Sat AIS providers and the number of vessels identified by Sentinel-1 for October 2021. For each Sentinel-1 image snapshot, we consider only the Sat AIS positions that were received 30 minutes before and 30 minutes after the image was acquired and that their location was within the footprint of the image. We observe that the Sentinel-1 image detections outnumber the respective Sat AIS detections. A possible explanation for that is a combination of the following reasons: (i) Sat AIS messages may have been lost due to packet collisions, (ii) the detection of smaller vessels, that is a limitation of satellite images, is also a limitation for Sat AIS since smaller vessels that cannot be captured by Sentinel satellites typically bear class B transponders. During the interim review, we showed the results of our preliminary analysis that only covered the Mediterranean sea. In this final version of the report, we extended the analysis to cover the whole area of interest and the results are in good agreement to our initial analysis:

Sentinel-1 detects more vessels than Sat AIS (due to class B) but less that Ter AIS.

However, this is not always the case. Interestingly there are days when Sentinel-1 detects more vessels than terrestrial-AIS as well. This depends on the orbit of the satellite itself: Sometimes it visits places where terrestrial coverage is low

We compared the coverage for each one of the two sources in the following way:

For each image snapshot t , we compute the number of vessels that are detected by Sat AIS, $N_{\text{SAT-AIS}(t)}$, and the number of vessels that are detected by satellite images $N_{\text{EO}(t)}$, as well as the number of vessels that are seen by both satellites $N_{\text{SAT-AIS-EO}(t)}$. Then, the total number of vessels for snapshot t , denoted as N_T , where $N_T = N_{\text{SAT-AIS}(t)} + N_{\text{EO}(t)} - N_{\text{SAT-AIS-EO}(t)}$

We must note that a one-to-one comparison of the performance of all sensors can be misleading, due to coverage areas selected or traffic density in the region. Due to the orbits of the satellites and the temporal “snapshots” used for the analysis, it is not always possible to provide a 100% reference ground truth for comparisons.

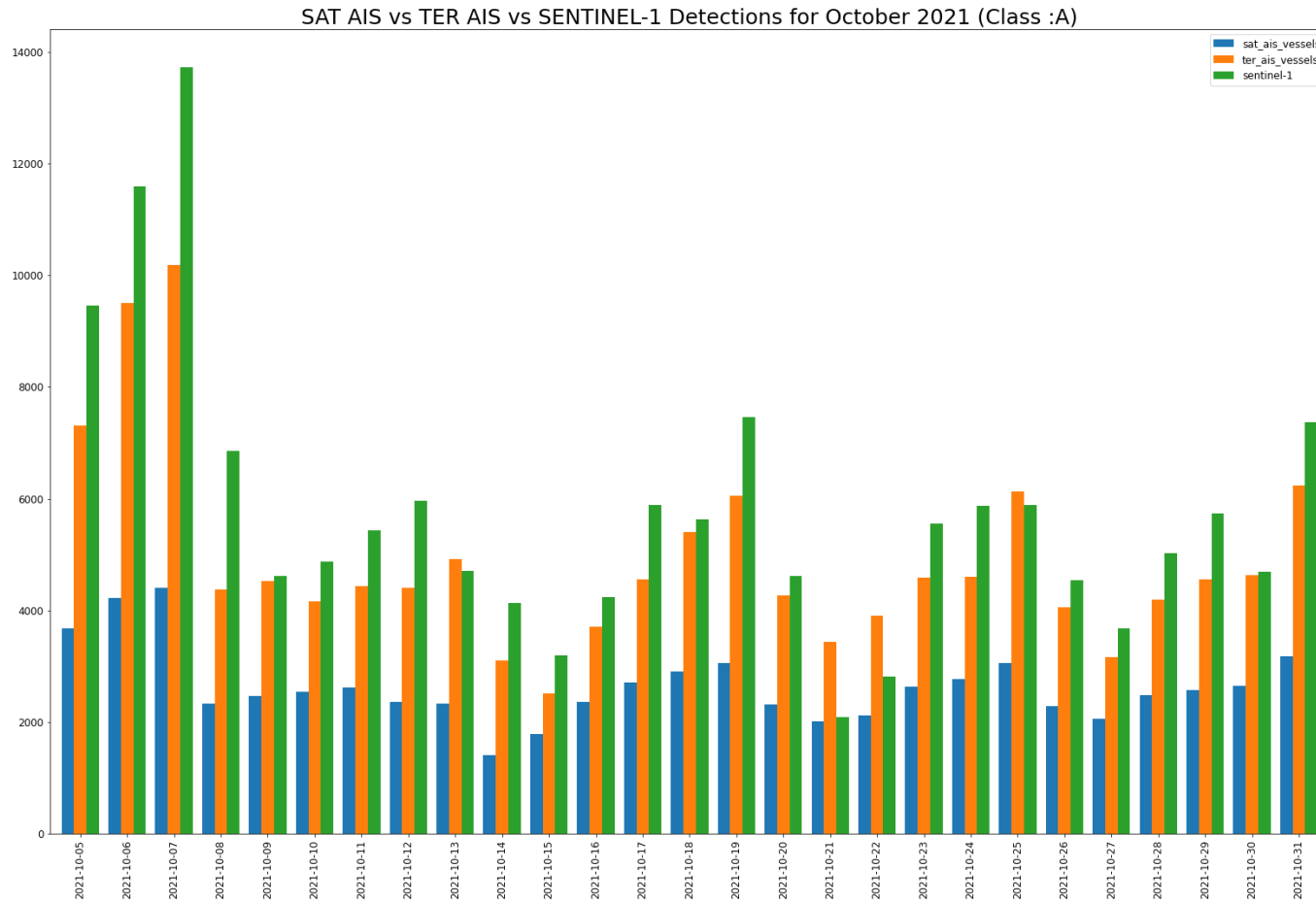


Figure 25. Number of vessels detected via Sat AIS vs number of vessels detected using satellite images per day vs Ter AIS for October 2021 in the area of the Mediterranean sea. For the comparison, we only considered the Sat AIS and Ter AIS positions that were received 30 minutes before and after the acquisition time of the image and were located within the image extent.

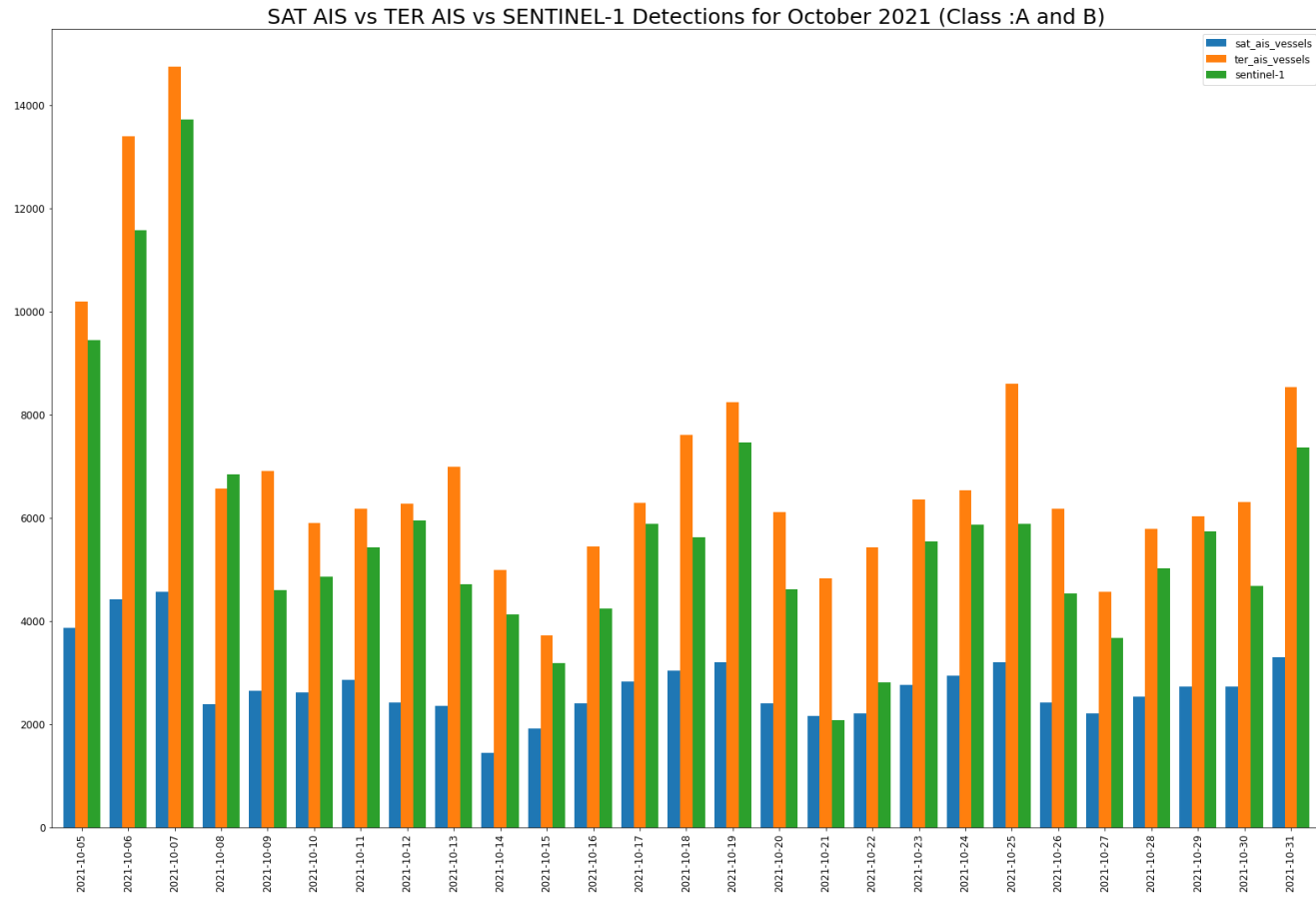


Figure 26. Comparison of unique number of vessels detected via Sentinel-1 imagery vs Sat AIS vs Ter AIS (only AIS positions contained in the image temporal and spatial footprint were considered)

The figure above shows the number of vessels detected via Sentinel-2 compared to the number of unique class A vessels detected via Ter AIS and Sat AIS and Figure 27 performs the same analysis considering both class A and class B vessels. By comparing these two sets of figures to the previous one (about the Sentinel-1 analysis), we can make the following observations: First of all, it is obvious that the total number of vessels captured by Sentinel-2 is considerably smaller than the number of vessels captured by Sentinel-1 during a one-month period. The number of vessels detected via AIS located in the footprint of the respective images is also considerably smaller. The reason for this is that (i) we only considered Sentinel-1 images with no more than 10% cloud coverage, and (ii) filtered out all image tiles that contained clouds, so the Sentinel-2 cloud-less footprint is considerably smaller, leading to a smaller number of vessels detected by all sensors. However, despite the inherent cloud coverage issues of Sentinel-2, it can pick up more vessels compared to Ter AIS, unlike Sentinel-1. This is because (i) Sentinel-2 offers images at better spatial resolution than Sentinel-1 and (ii) Sentinel-2 images are optical while Sentinel-1 images are radar images. These inherent features of Sentinel-2 imagery enables it to capture smaller vessels (e.g., high-speed vessels, pleasure crafts) that might not bear Class A transponders or no AIS transponders at all. This assumption is strengthened by comparing the results of the comparisons shown in and Figure 27. The difference between satellite detections and Ter AIS detections is far smaller when both Class A and Class B vessels are considered: This demonstrates the ability of Sentinel-2 to pick up class-B vessels in many cases better than Ter AIS.

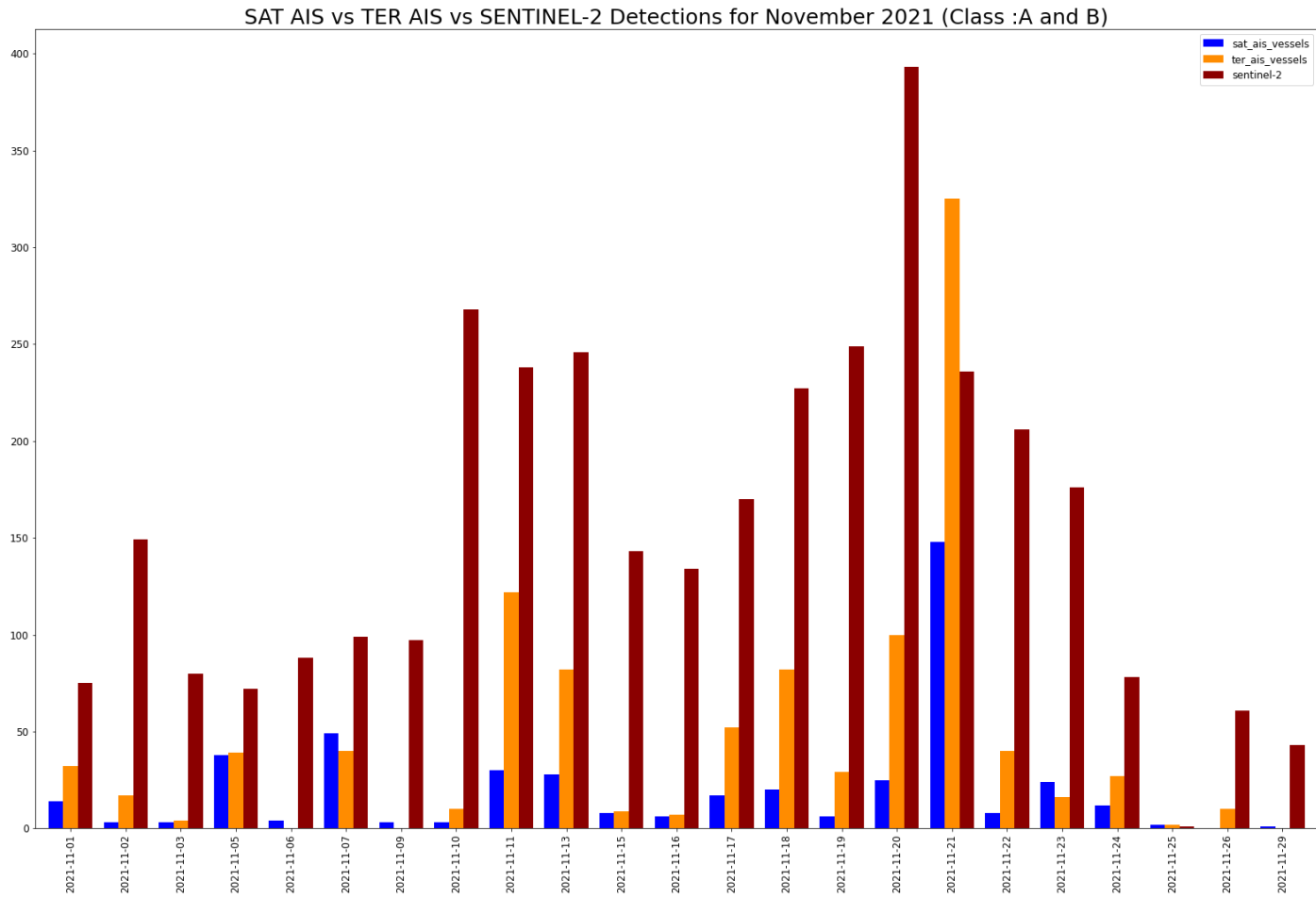


Figure 27. Class A and B vessels detected by Sentinel-2 vs Sat -AIS vs Ter AIS per day for November 2021

Summary

The findings of the above analysis can be summarised as follows. First, Sentinel-2 imagery is a valuable source of information to capture smaller vessels (e.g., class B). In several conditions, both Sentinel-1 and Sentinel-2 can pick up more vessels than Sat AIS, due to Sat AIS known AIS packet collision problems in congested areas, as well as its limited capacity to monitor class B vessels.

For the same reasons, Sentinel-2 often detects more vessels than Ter AIS, while Ter AIS outperforms Sentinel-1 in most cases. However, cloud coverage is a major limitation for this use case, as the cloud-free coverage that is offered by Sentinel-2 optical imagery is too limited to enable us to estimate the vessel traffic density at global scale, as we will see later on in this document.

6. Creating new sets of digital density maps

ABSTRACT

In this section we report on performed activities for the creation of the digital map outputs, including the trajectory reconstruction methods applied for the improvement of the end results.

Overall map outputs are created in five different spatial resolutions and with different approaches to highlight different aspects of the analysis.

This section reports on the results from “Task 3: Creating new sets of digital density maps” and specifically the work conducted in the subtasks “[T3.1] Creating density maps used in EMODnet”; “[T3.2] Vessel density maps using vessel density information acquired from satellite image processing”; “[T3.3] Corrected density maps with Sat AIS message collision elimination”; “[T3.4] Density maps showing the probability to avoid detection of a vessel” and “[T3.5] Corrected trajectories density digital maps”.

6.1 Density maps using AIS

For the map creation, we based our implementation on the current EMODnet method. Specifically, this process outputs a different density map for 14 vessel types, included in the EMODnet density maps, for the given areas of study and in five different spatial resolutions (1, 10, 100, 200 and 500km). Table 4 presents the vessel category types, according to the code included in the AIS messages; each vessel corresponds to a single type.

Vessel Type Categories	
Category	AIS vessel type codes
All	all
Cargo	between and including 70 and 79
Tanker	between and including 80 and 89
Dredging (Dredging or underwater operations)	33
HSC (High-speed crafts)	between and including 40 and 49
Fishing	30
Military_Law	35 and 55

(Military and law enforcement)	
Passenger	between and including 60 and 69
Pleasure (Pleasure crafts)	37
Sailing	36
Service	50, 51, 53, 54, 58
Tug (Tug and towing)	31, 32, 52
Unknown	in case a vessel type code is not provided
Other	otherwise

Table 4. Vessel type categories.

The density is extracted by calculating the cumulative interval that the vessels in question spent within each cell. For this purpose, trajectory segments are created from the cleaned AIS data; a segment is created for each pair of consecutive messages of the same vessel. During these segments the vessel is considered to follow a linear path. As already mentioned, this approach may be considered accurate for relatively small time intervals and short distances but cannot be used universally. Thus, segments that originate from AIS intervals larger than 6 hours or have a length larger than 30km are labelled “lossy” and annotated for the reconstruction process. Following this, each segment is split according to the grid and the respective time for each cell is computed. Figure 28 provides an example of this process for the trajectory of a vessel.

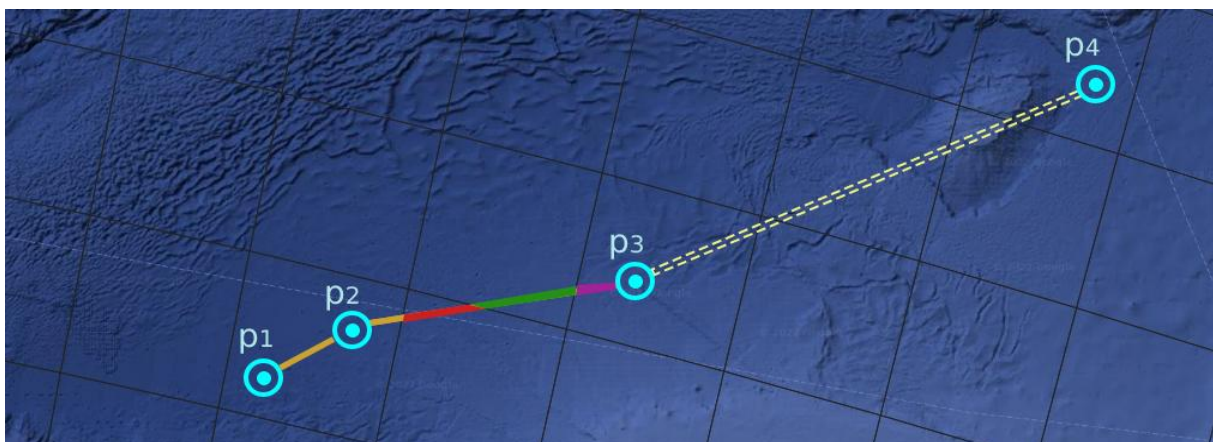


Figure 28. The density calculated from a vessel’s positions ($p_1 \Rightarrow p_2 \Rightarrow p_3 \Rightarrow p_4$). For the first segment ($p_1 \Rightarrow p_2$) the full time-interval is assigned to a single cell, since the vessel stays within its borders. For the second segment ($p_2 \Rightarrow p_3$), a line split is performed so that the corresponding sub-segments are assigned to four cells in total. Lastly, the third segment ($p_3 \Rightarrow p_4$) does not provide any additional measurements since the distance between its confines is too large.

EMODNET’s colour palette was used for the creation of the final maps, with the density thresholds (hours per square km per month) adjusted for sizes other than 1km, as seen in Table 5. The end results are saved as TIFF files.

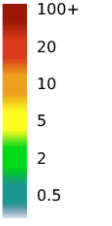
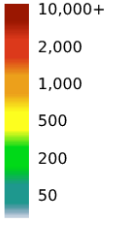
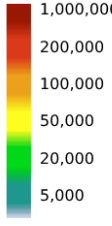
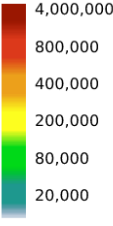
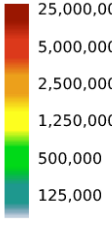
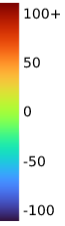
Density Maps Colours						Difference Maps
Cell size	1km	10km	100km	200km	500km	
Colours (h / cell per month)						

Table 5. Density maps colour palette.

Considering all five resolutions, the two categories of data sources (Ter AIS and Sat AIS) and the 14 different vessel types, 420 TIF files were produced for all three months (October, November and December 2021). An indicative sample is presented in Figure 29 to Figure 36.

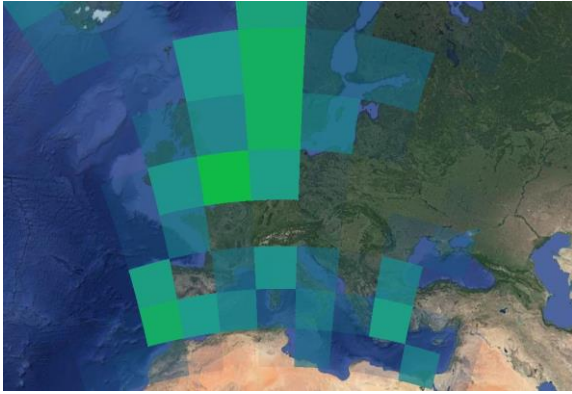


Figure 29. Density map based on Sat AIS, for all vessel types, using a resolution of 500km, for the month of October.

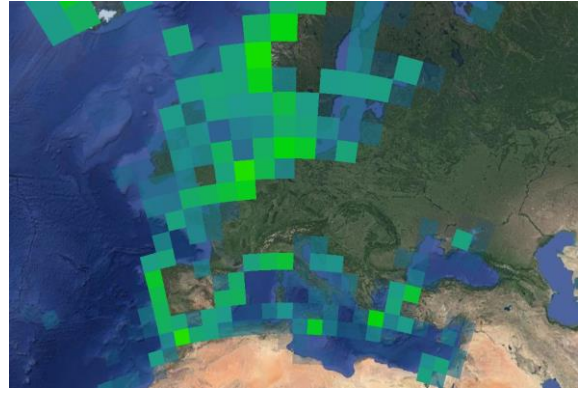


Figure 30. Density map based on Sat AIS, for all vessel types, using a resolution of 200km, for the month of October.

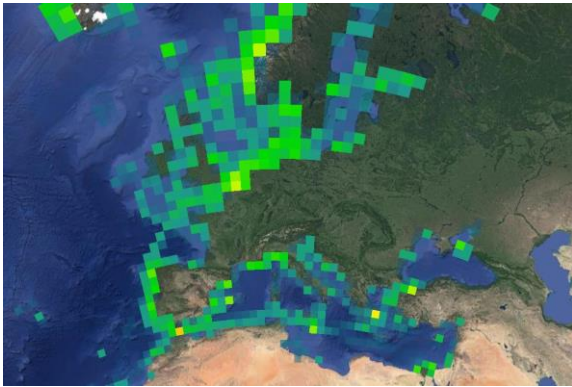


Figure 31. Density map based on Sat AIS, for all vessel types, using a resolution of 100km, for the month of October.

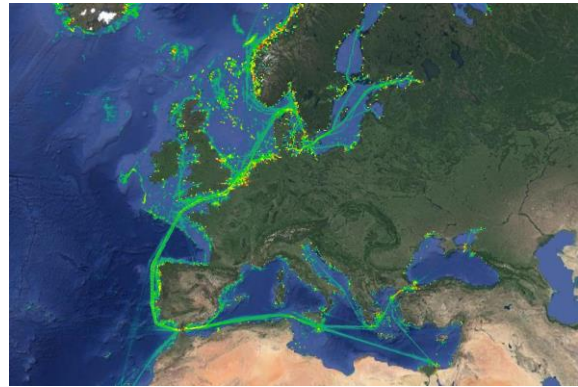


Figure 32. Density map based on Sat AIS, for all vessel types, using a resolution of 10km, for the month of October.

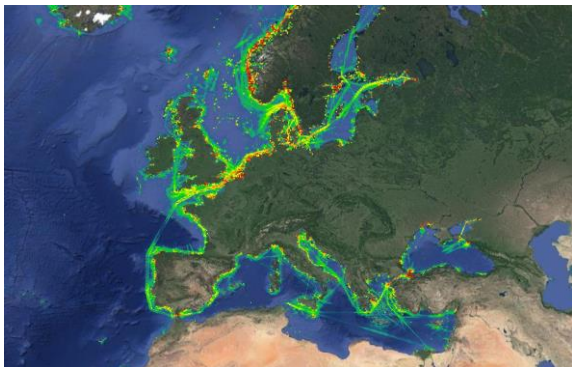


Figure 33. Density map based on Ter AIS, for all vessel types, using a resolution of 10km, for the month of November.

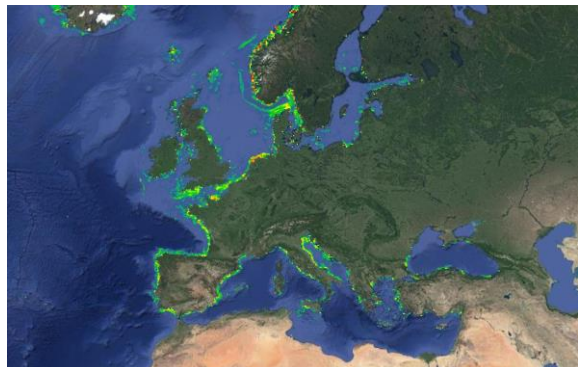


Figure 34. Density map based on Ter AIS, for fishing vessels, using a resolution of 10km, for the month of November.

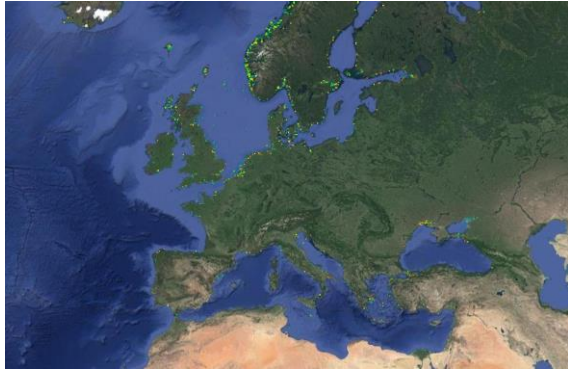


Figure 35. Density map based on Ter AIS, for tug and towing vessels, using a resolution of 10km, for the month of November.



Figure 36. Density map based on Ter AIS, for military and law enforcement vessels, using a resolution of 10km, for the month of November.

Furthermore, in order to visualise the difference between satellite and terrestrial AIS messages we created maps that carry such information upon each grid cell (for all grid edge lengths and vessel types). More precisely, in each grid cell we calculated the difference percentage between the alternative sources according to the following formula:

$$dperc = 100 * \frac{density_{TER} - density_{SAT}}{density_{TER} + 1}$$

where, $density_{TER}$ and $density_{SAT}$ are the density values, calculated according to previous steps, for the terrestrial and satellite AIS datasets accordingly. Note that, we added a single millisecond in the denominator so that no infinite values were extracted in case terrestrial data didn't occur in a single cell.

Afterwards, we categorise each cell according to its $dperc$ and generate the map of difference, using the colours depicted in Table 6. In these maps blue and purple colours indicate areas where Sat AIS have larger density values; yellow and red point to larger Ter AIS densities, while green indicates equal values from both sources. Some examples of the end results using a resolution of 10km are presented in the following Figures.

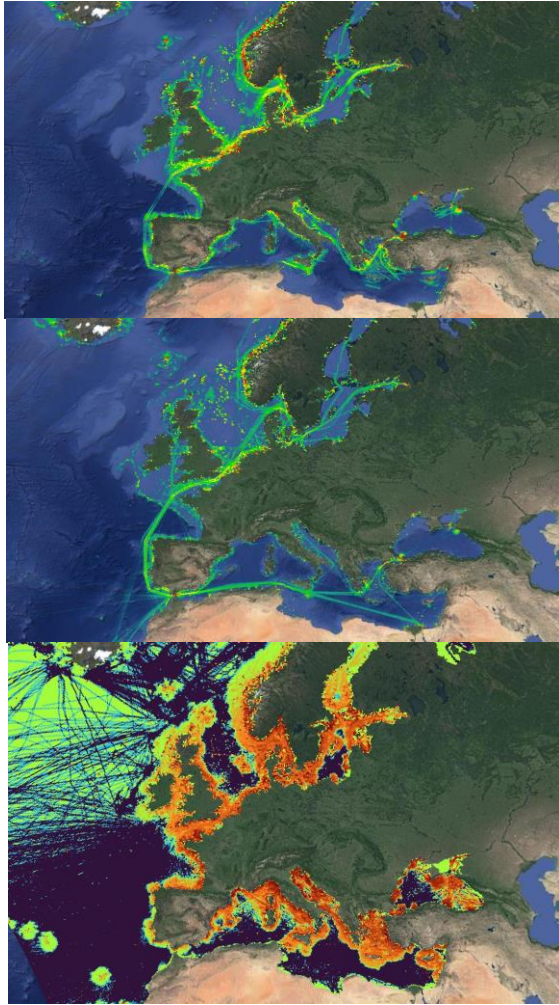


Figure 37. Density map based on Ter AIS (top), based on Sat AIS and the corresponding difference map, for all vessel types, using a resolution of 10km, for the month of December.

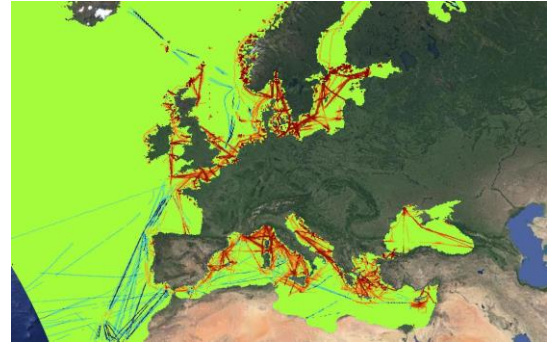


Figure 38. Difference map, for passenger vessels, using a resolution of 10km, for the month of December.

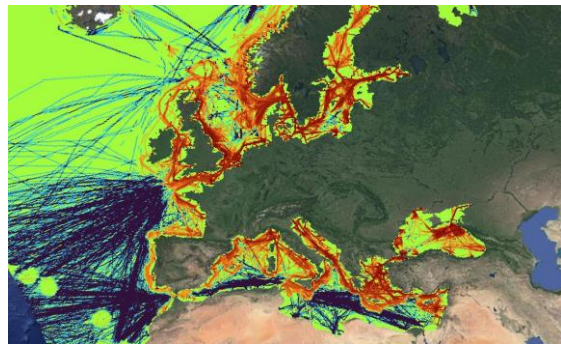


Figure 39. Difference map, for tanker vessels, using a resolution of 10km, for the month of December.

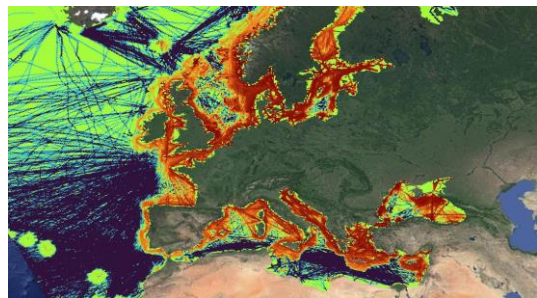


Figure 40. Difference map, for cargo vessels, using a resolution of 10km, for the month of December.

6.2 Density maps using EO data

In this task we create density maps based on the EO detections. More specifically, we execute the following steps:

1. We map the vessels detected in satellite imagery for each month to a grid. We use grids of different resolutions: 1m, 10m, 100m, 200m, and 500m resolution. We create a set of density maps for each grid, so we compute the number of vessels located in each grid for each one of the aforementioned resolutions.
2. We create two different kinds of density maps: First, we visualise the density of all vessels detected in satellite images. Then, we also visualise the density of all dark vessels detected in satellite images, i.e., vessels that were only detected in satellite images without matching AIS positions. This set of density maps will help us identify “dark” areas (e.g., grid cells) with increased traffic of “dark” vessels, which can either be areas of low AIS coverage or areas associated with illegal activities (e.g., sanctioned areas).

Using Sentinel-1 and Sentinel-2 vessel detections, we produced two sets of satellite-based density maps for the following resolutions: 10km, 100km, 200km, and 500km (all included in the deliverables). Some examples are provided below. Figure 41 shows the Sentinel-1 density map for October 2021 (10km cell size). Figure 42 shows the density map using Sentinel-2 data for the same resolution. The cloud coverage problem is obvious in the case of Sentinel-2, which makes this data source unsuitable as a single source of vessel tracking data. However, it can be used as a complementary source.

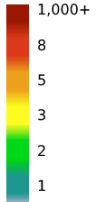
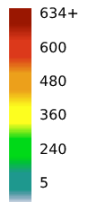
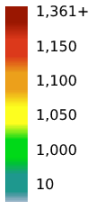
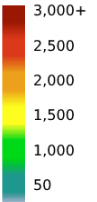
Density Maps Colours				
Cell size	10km	100km	200km	500km
Colours (number of detections)				

Table 6. Density maps colour palette (satellite imagery detections).

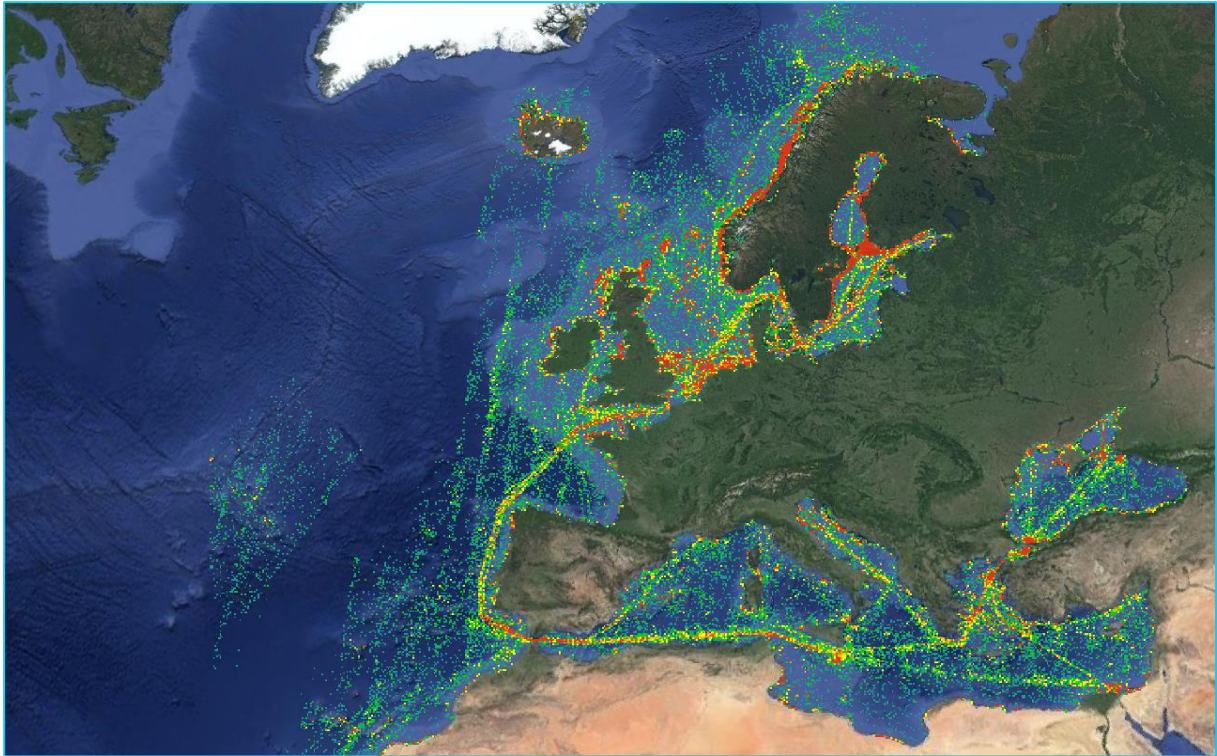


Figure 41. Vessel traffic density map using Sentinel-1 image detections for October 2021 (density metric: number of vessels)

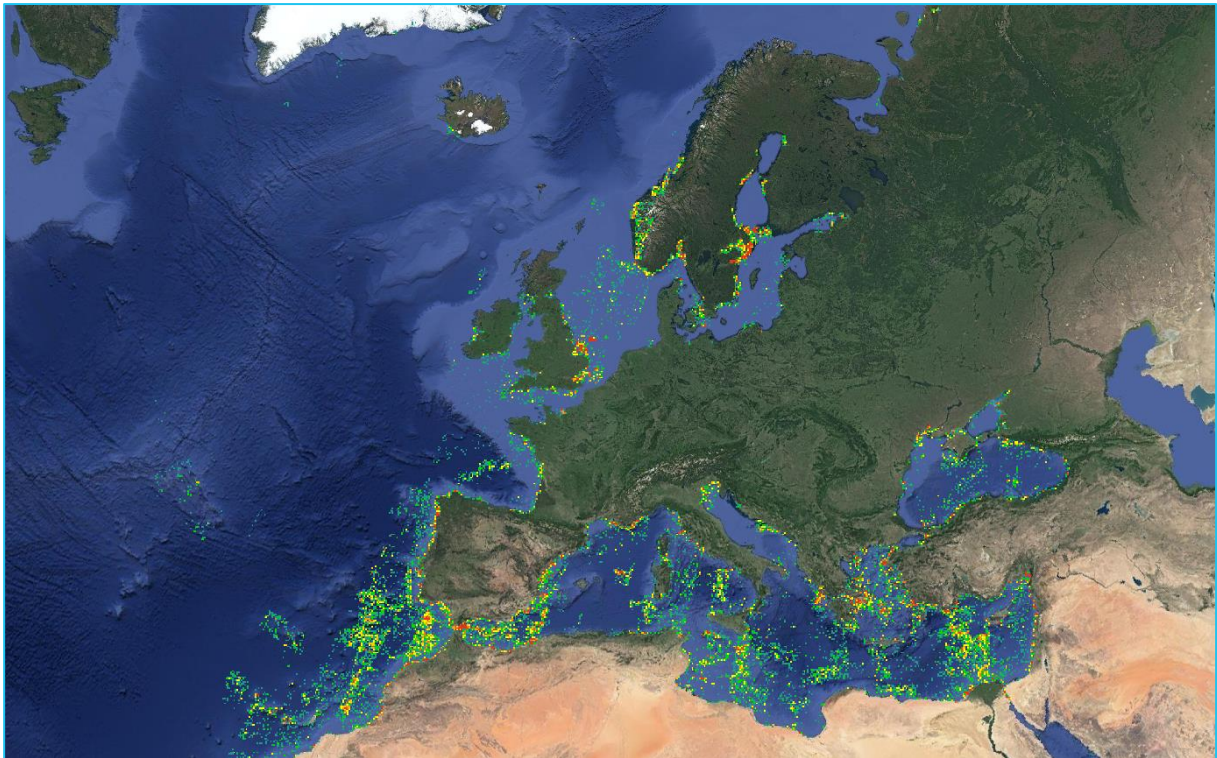


Figure 42. Density map (10 km) of vessels detected using sentinel-2 imagery for November 2021

We can observe that even in Sentinel-1 there are significant gaps in coverage, especially in areas far away from the coastline. This is a known issue of both ESA and also some commercial satellites. For example, the Sentinel-1 October orbits can be seen in Figure 43 and Figure 44 demonstrating limited coverage in the high seas.

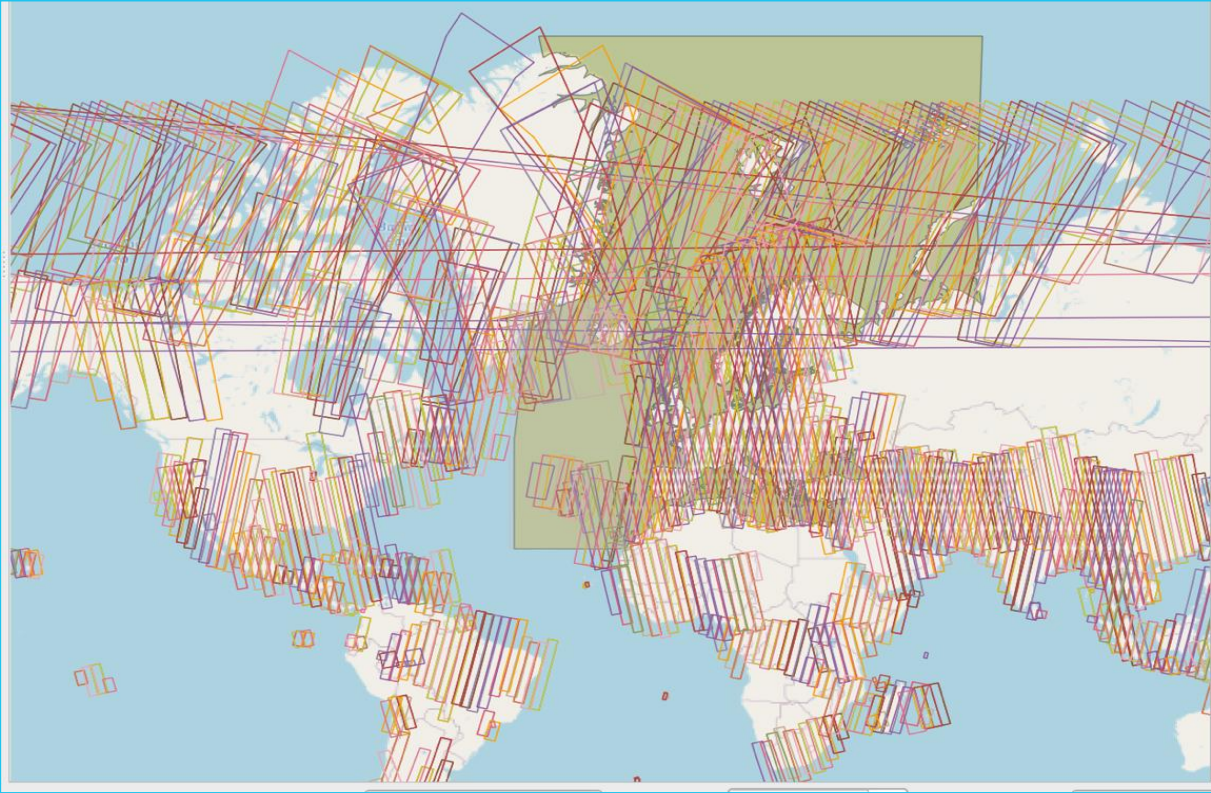


Figure 43. Sentinel-1 orbits in October 2021

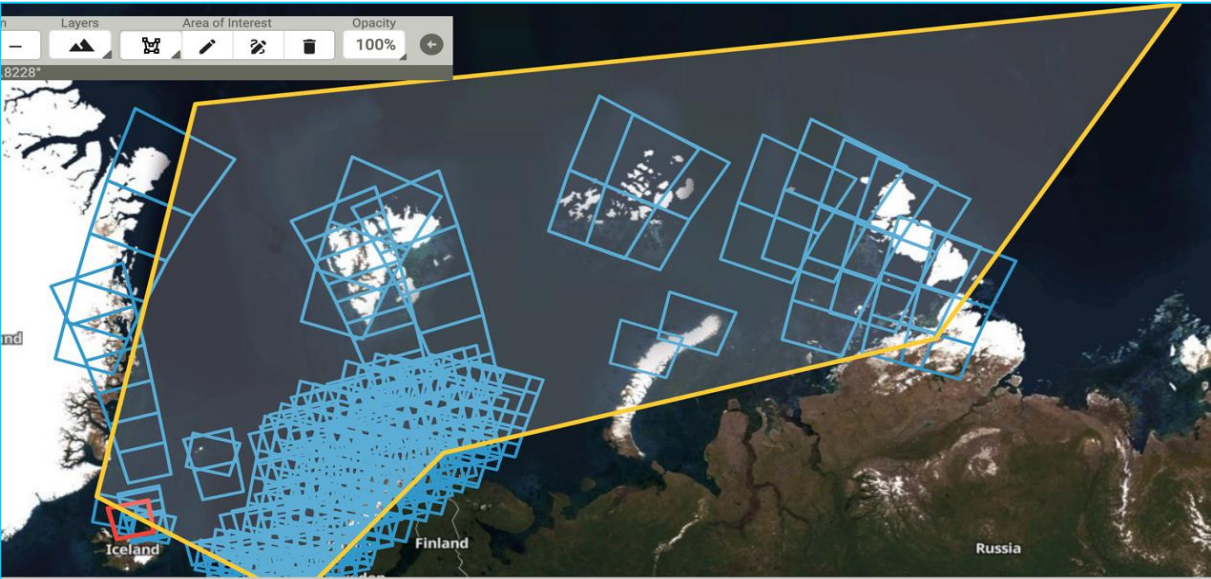


Figure 44. Sentinel-1 arctic coverage

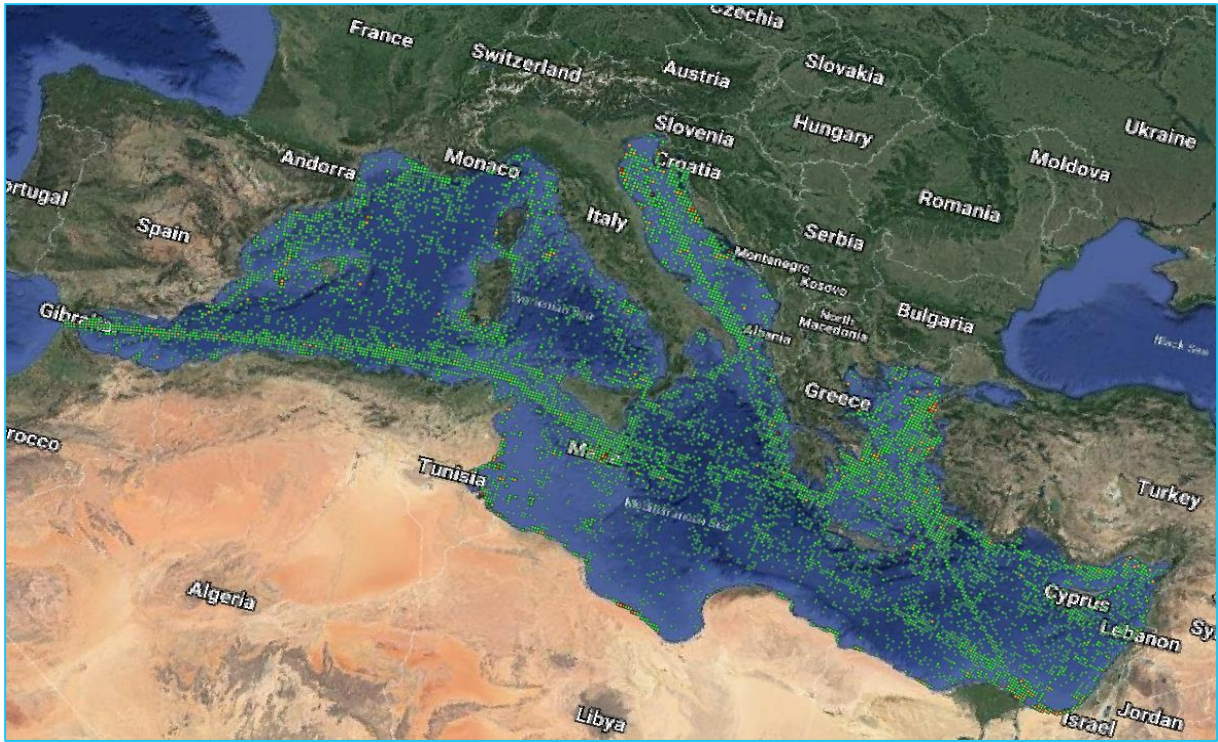


Figure 45. Dark vessel density map using Sentinel-1 images from October 2021 in the MED area (1:10000 grid resolution)

6.3 Density maps of dark areas

Figure 46 shows the density of dark vessels in the Mediterranean sea using Sentinel-1 data correlated with AIS data. Darker red cells correspond to higher density of dark vessels. We observe that many of these dark red cells correspond to higher vessel density in general, so the dark vessel density might be due to AIS packet collision

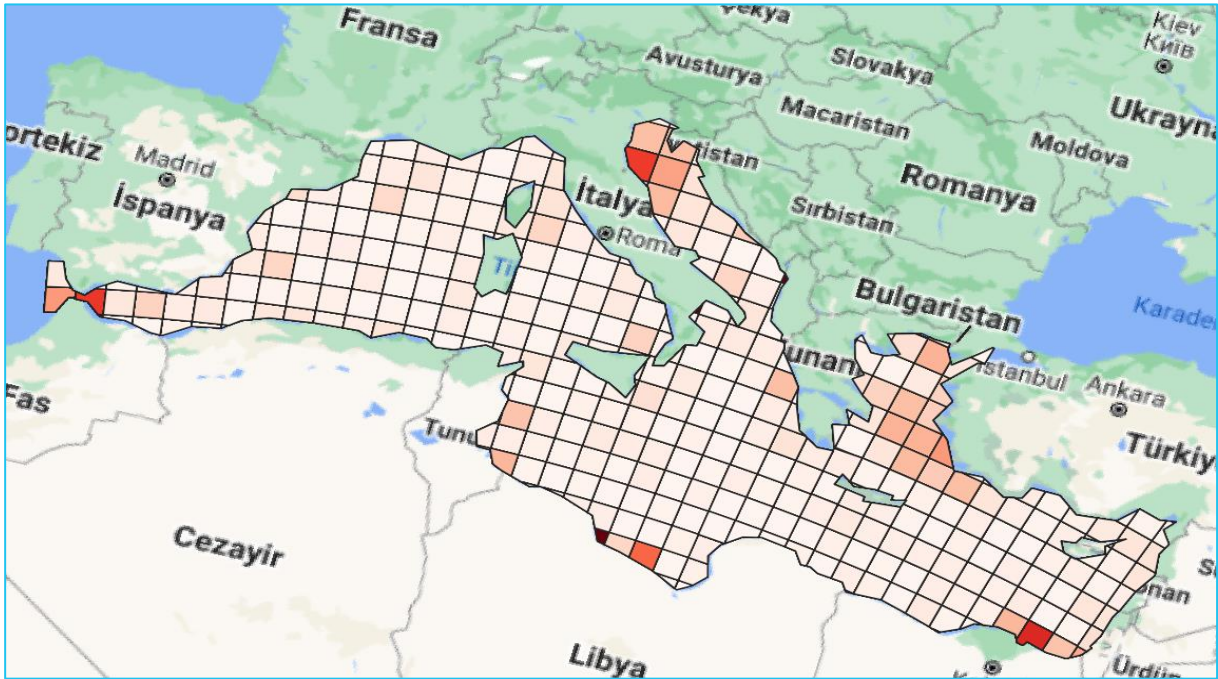


Figure 46. Dark vessel density maps for the Mediterranean sea based on Sentinel-1 vessel detection (Grid cell: 10 km)

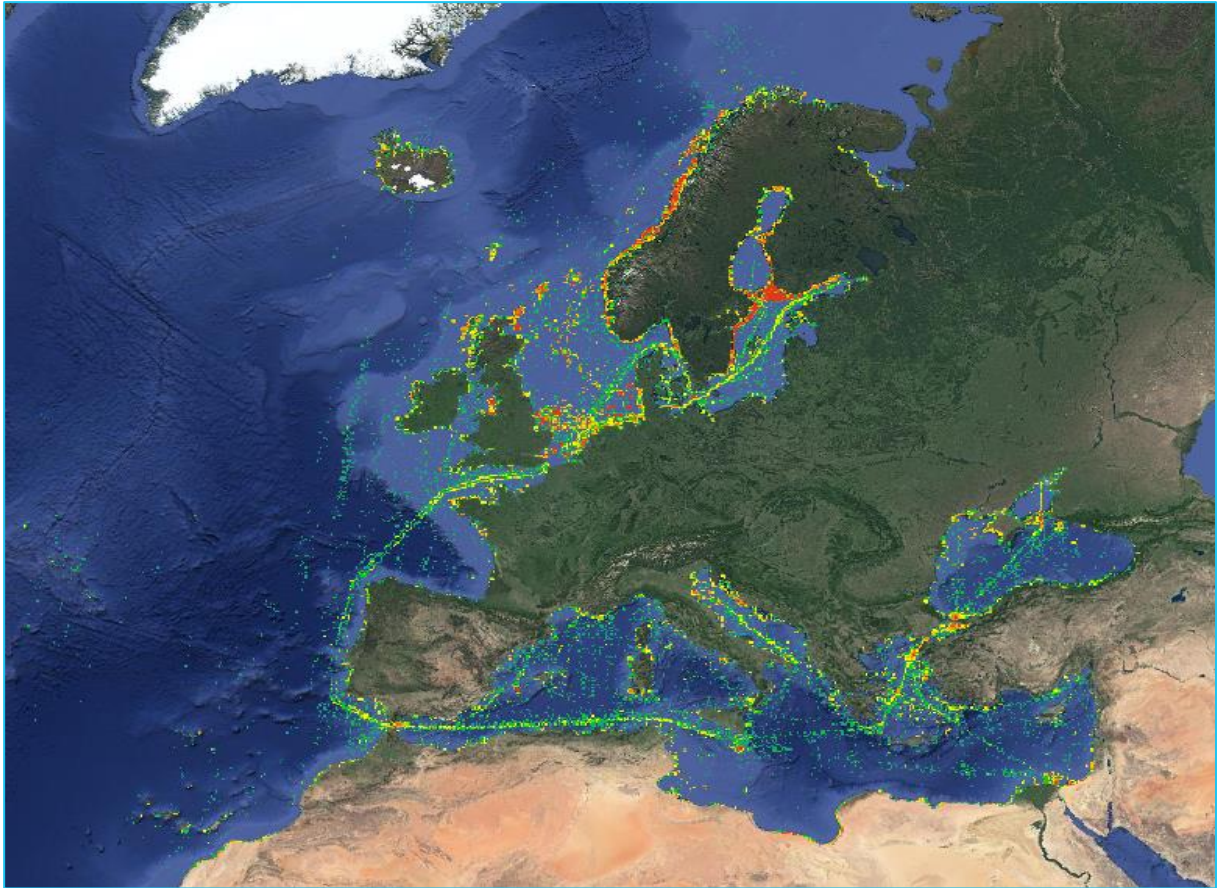


Figure 47. Dark vessels during October 2021 detected via Sentinel 1 imagery correlated with Sat AIS data

To differentiate between low AIS coverage due to packet collision and intentional switch-off of AIS transponders, we performed an analysis described in this section, which is based on identifying areas (e.g., grid cells) where vessels are in close distance, thus increasing the risk of losing AIS packets due to congestion.

To achieve this, we applied the following formula to different snapshots of Ter AIS data (we used the same snapshots of the respective Sentinel-1 images used to identify dark “cells”):

$$\underline{NND} = \frac{\sum NND}{n}$$

NND: Nearest neighbour distance

N: Number of vessels

For each grid cell, we calculated the mean nearest neighbour distance, i.e., the mean distance between a vessel and its neighbours. This metric indicates that the lower the mean distance between neighbouring vessels is, the higher the risk of congestion. Thus, we define the “detectability” maps as the maps that showcase congested areas that might lead to detectability issues. We constructed detectability maps for the Mediterranean Sea and the whole area of interest (Figure 50).

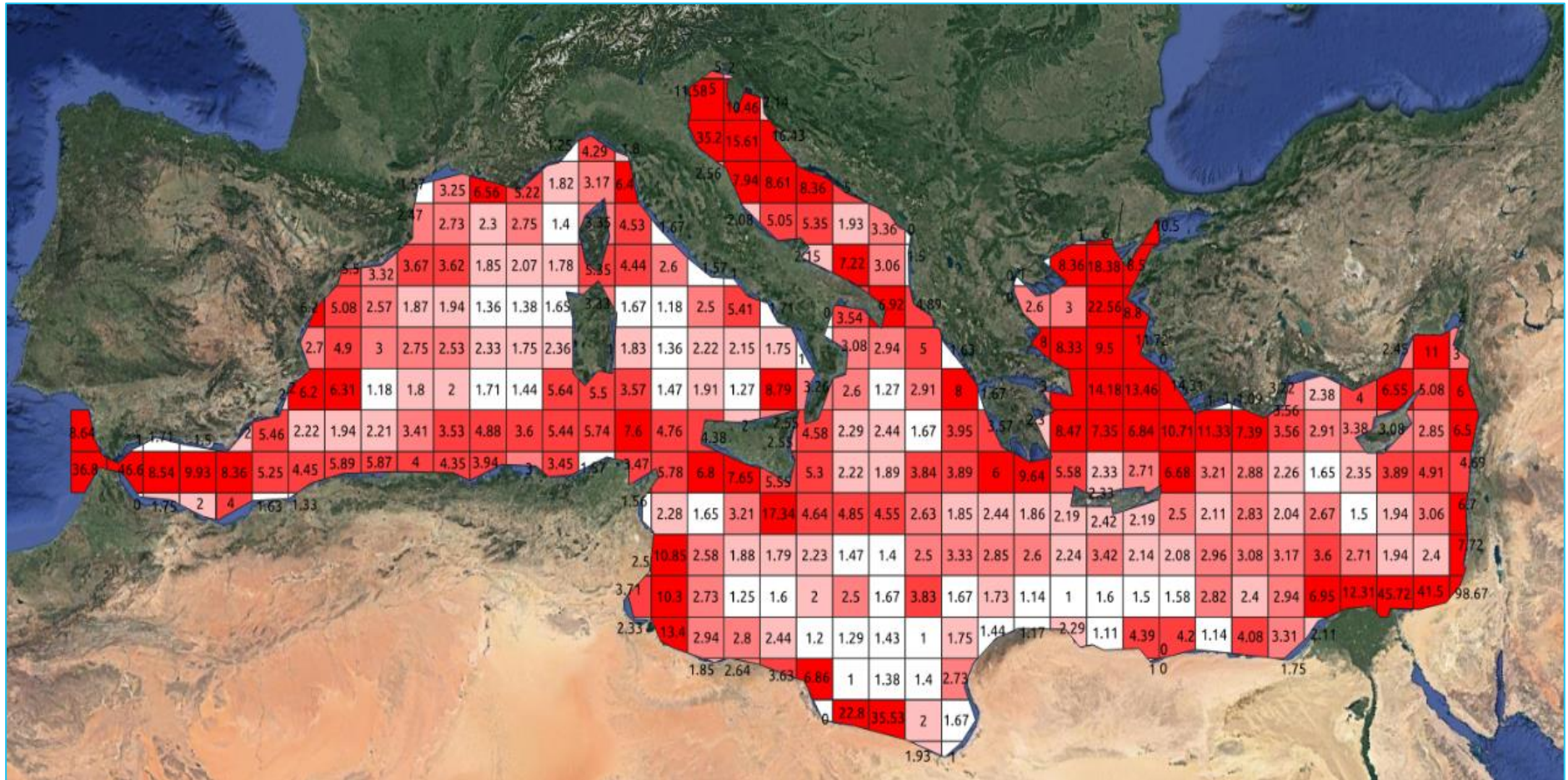


Figure 48. Detectability map for MED using October 2021 Ter AIS data

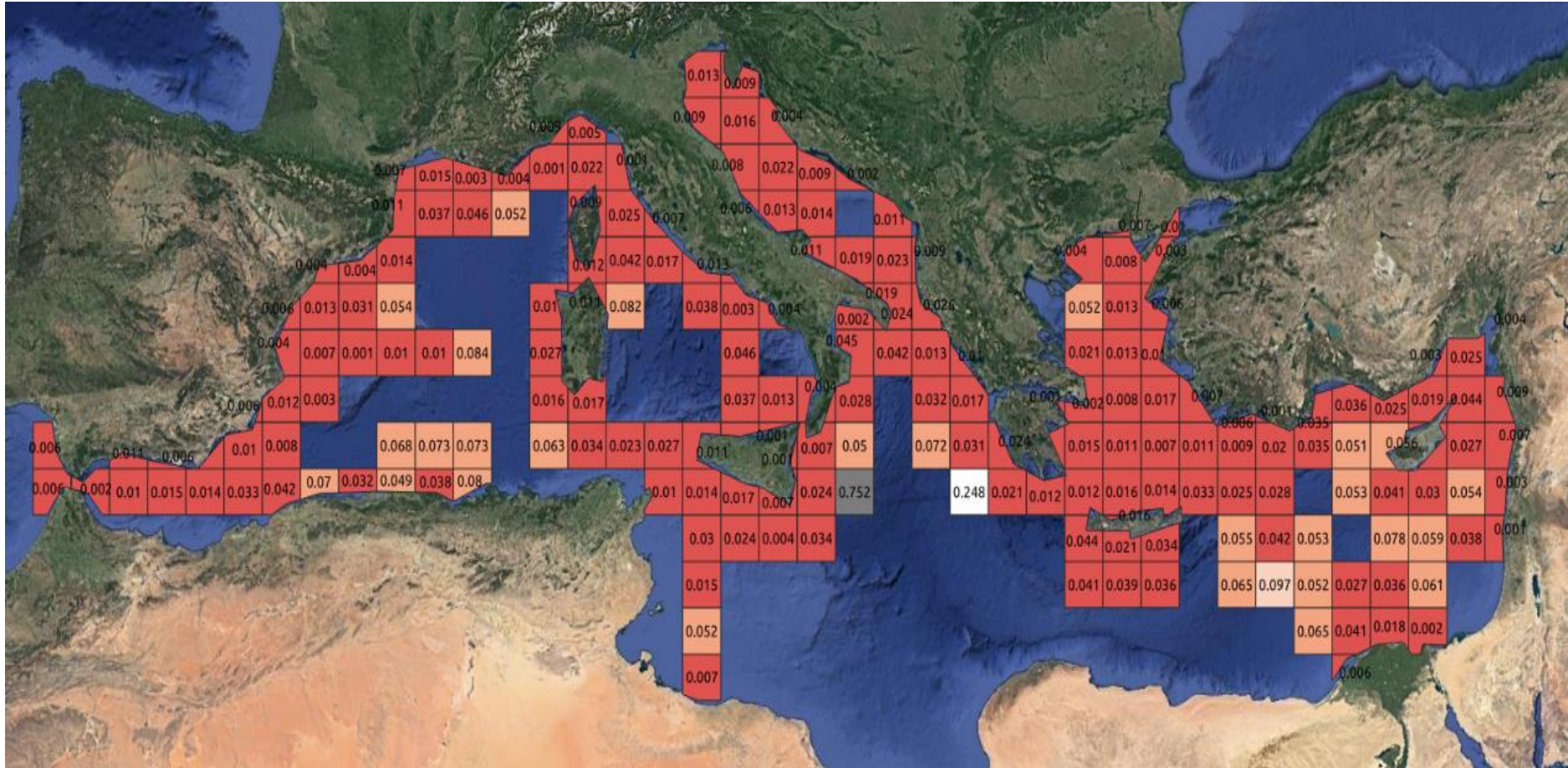


Figure 49. Detectability map for the Mediterranean

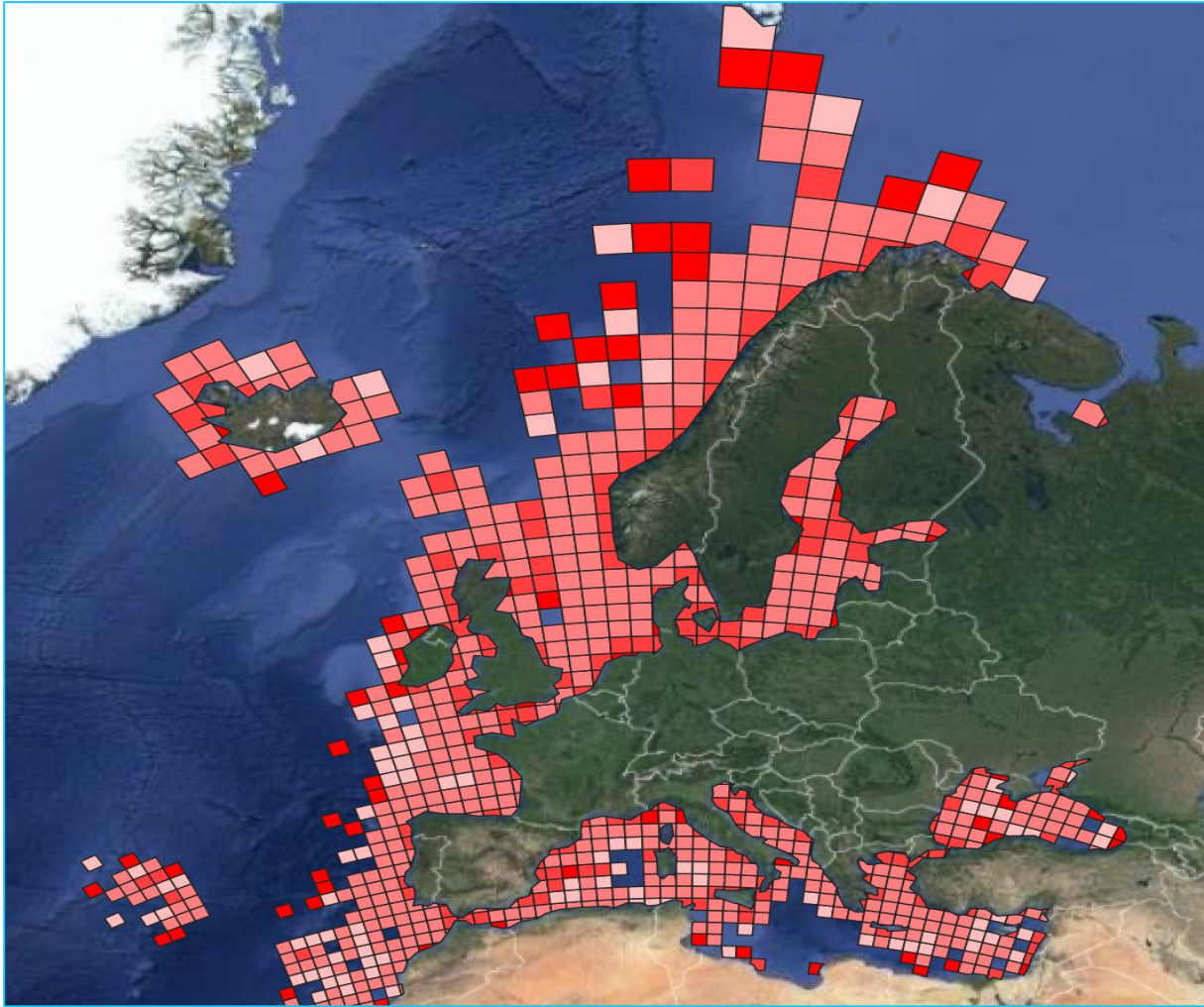


Figure 50. Detectability map for the European seas using Ter AIS October 2021 data. Darker colours correspond to higher probability to avoid detection.

6.4 Density maps combining Sat AIS & EO data

After producing the dark vessel density maps, as explained in the previous section, we used this information to complement the Sat AIS based density maps. Figure 51 shows the density map for October 2021 produced by using only Sat AIS data (number of positions/cell) and Figure 52 shows the respective density map of dark vessels for the same period. We consider only the dark vessels, and not all vessels detected by satellites for duplicate elimination (so that we don't count the same vessel position twice) and using this combination of data sources (Sat AIS and Sentinel-1) we produce the combined density map shown in Figure 53. Reasonably, the combined density map resembles the Sat AIS density map shown in Figure 51, as the limited revisit time of Sentinel satellites compared to the AIS message frequency do not allow for significant differences to be highlighted. However, looking at the combined density map more carefully, we can see that some areas that appear in orange/red in Figure 51 appear slightly darker in Figure 53 (e.g., Gibraltar, Norway, Iceland, English Channel, etc.). This means that areas that are known to be dense using Sat AIS, are even more dense in reality.

Areas that are both of high Sat AIS density and dark vessel density are prone to Sat AIS packet collisions and although ESA satellites are limited by their low revisit time and coverage, they can still be used to highlight these areas, although the differences of the combined density map and the Sat AIS density map are not significant.

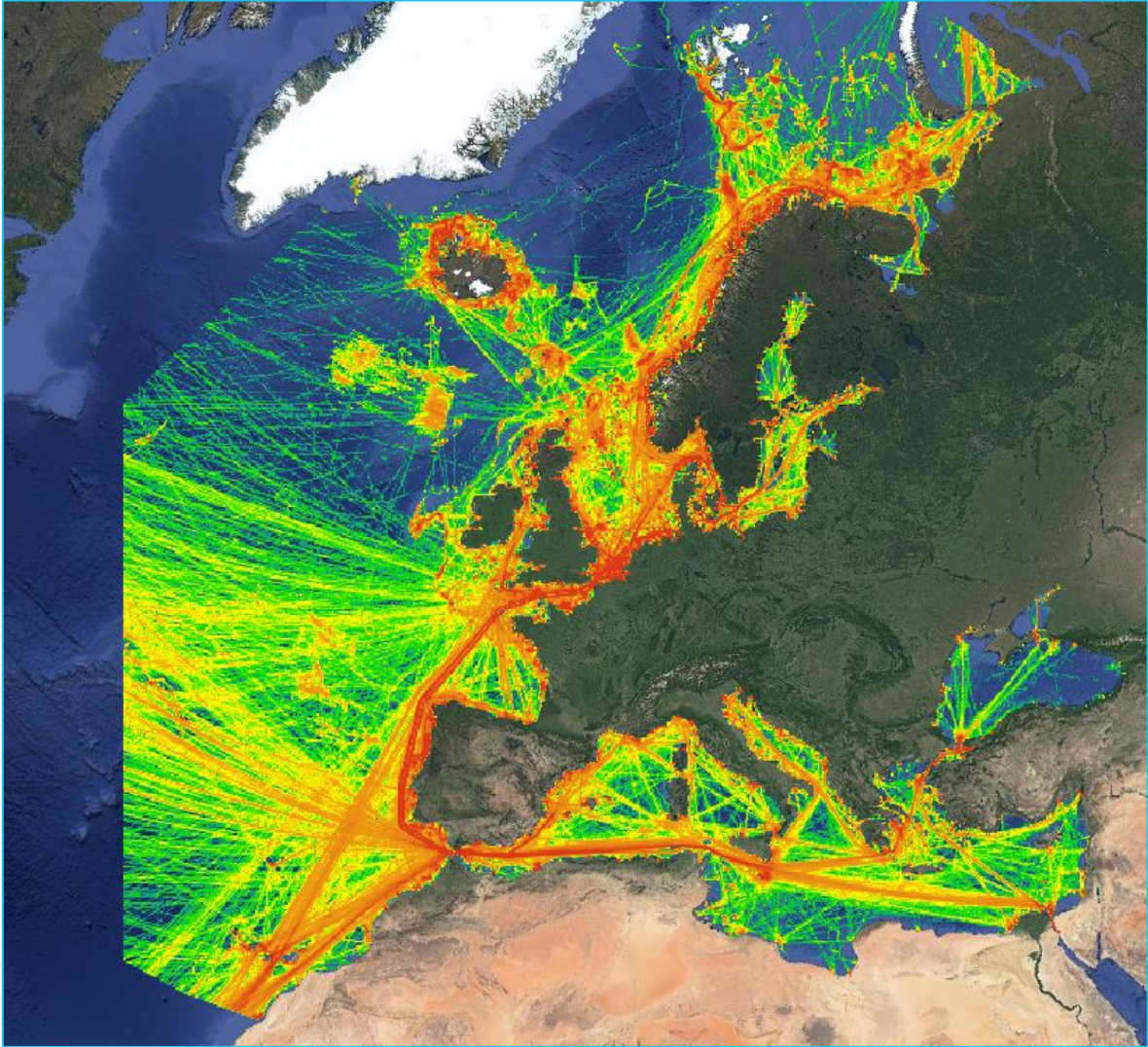


Figure 51. Sat AIS density map (number of detections/cell) for October 2021

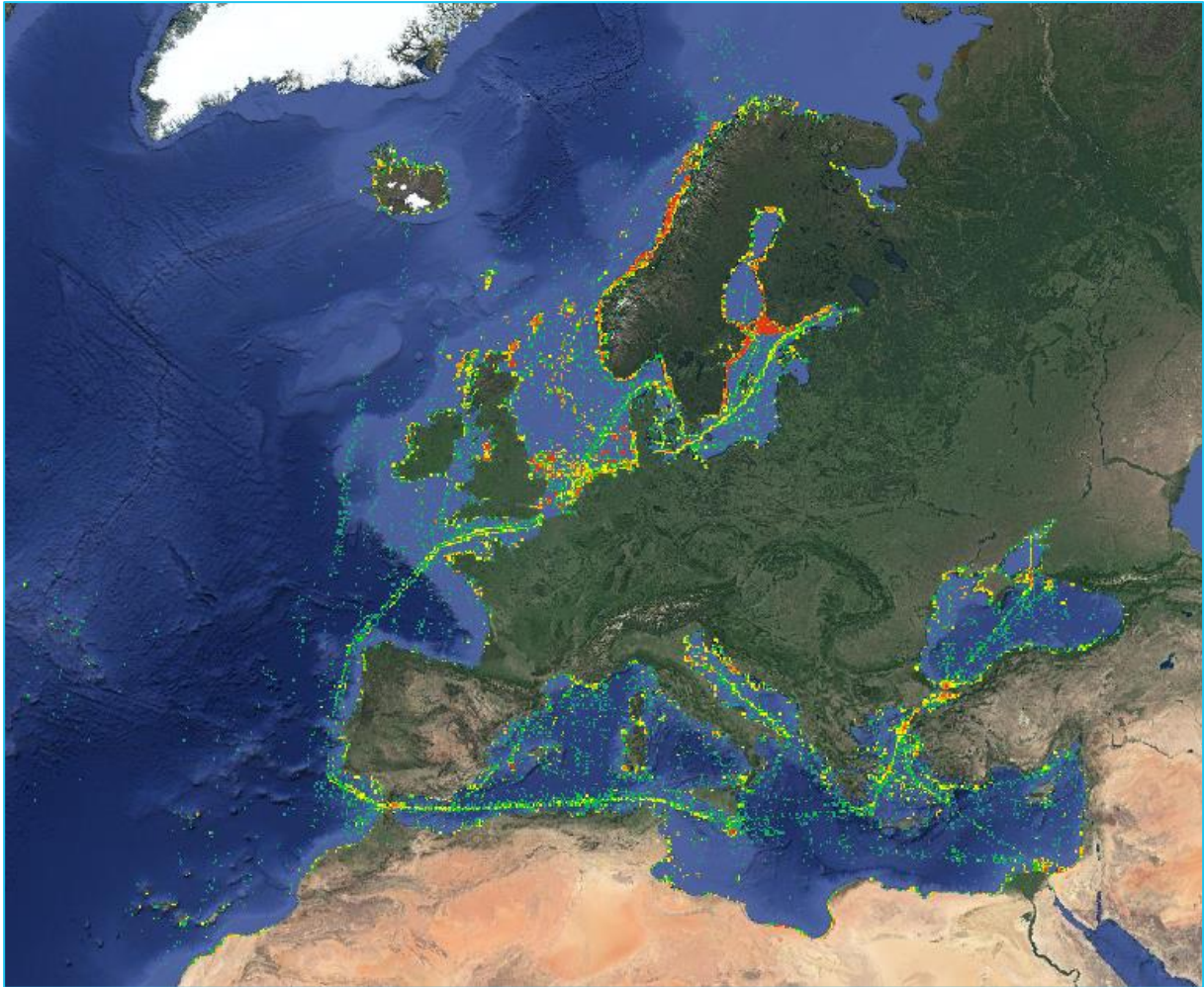


Figure 52. Dark vessels density map (number of detections/cell) for October 2021

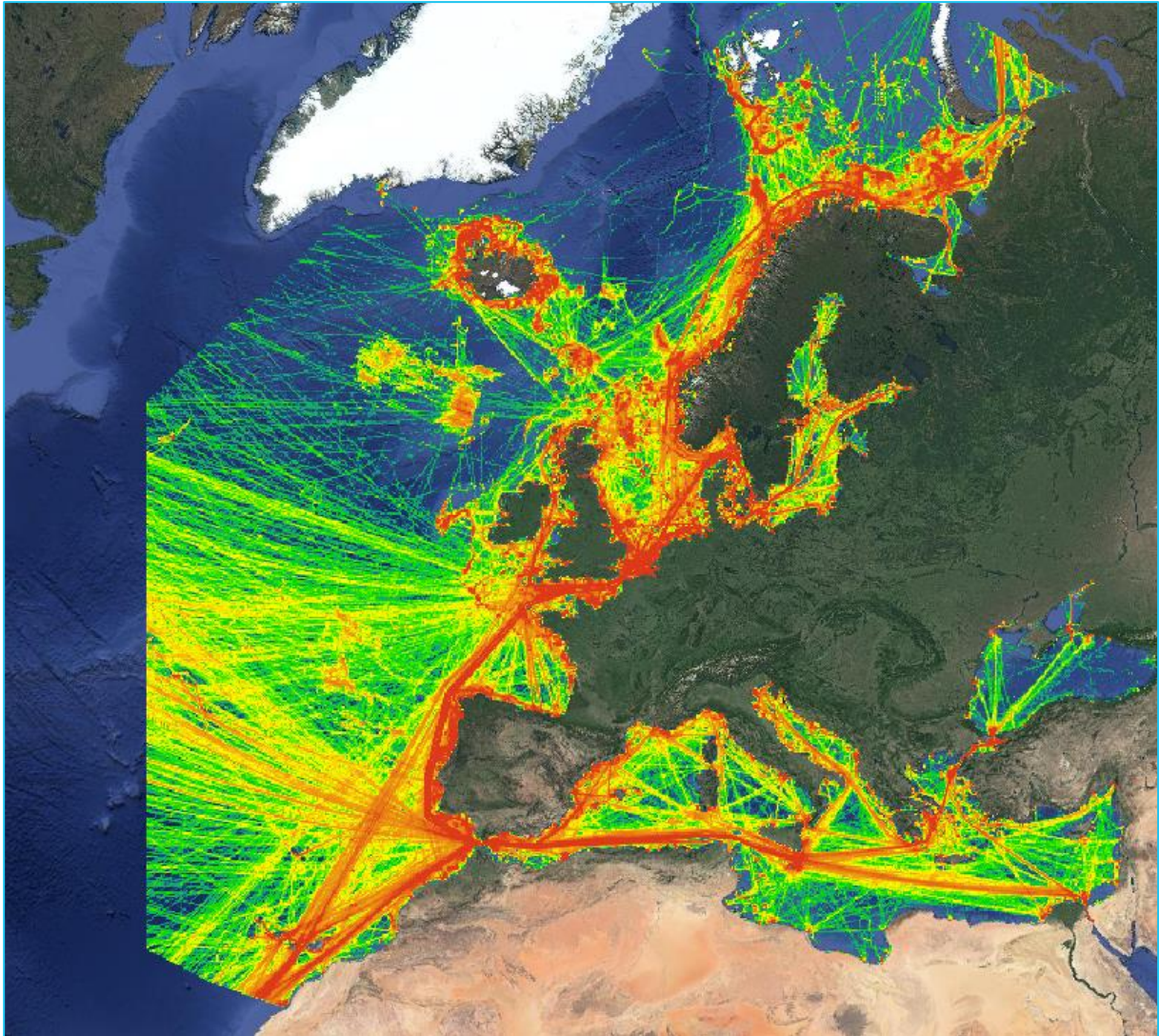


Figure 53. Sat AIS & Sentinel-1 density map for October 2021

6.5 Corrected trajectories density digital maps

6.5.1 Trajectory Reconstruction Methods

As previously noted, sparse AIS data are problematic for vessel movement analysis, since the reconstructed trajectories taken from the actual positional messages do not represent the vessels' real paths. Tasks such as the creation of line-based density maps depend on the validity of the data and may in turn suffer in accuracy.

Several works have proposed strategies for dealing with the imperfections of AIS data and specifically for completing or filling the gaps in trajectories. The majority of these works assume that at least in theory, the trajectory of a ship can be approximated as a straight line for a short time interval. Therefore, linear interpolation is the most widely used gap filling method [33]–[36]. Linear interpolation interpolates the latitude and longitude coordinates of the ship at a given interpolation time. However, this method is only suitable for small forecasting windows (e.g., several minutes), high frequency data, and in situations when the vessel is expected to follow uniform linear motion. Additional interpolation methods have been proposed, such as polynomial, cubic spline interpolation, Lagrange and Hermit interpolation methods which take into consideration additional features such as direction, heading and speed to reconstruct trajectories with curves [35], [37]. Unfortunately, though, the above methods do not take into account the environment or prior information regarding a ship travel area, while also showing degraded accuracy in scenarios where a ship conducts manoeuvres often (e.g. river).

Several works have attempted to combine the methods above with the ship's navigational status to improve results. Zhang et al. proposed a trajectory reconstruction approach considering the navigation states, namely hoteling, manoeuvring, and normal-speed sailing [38]. [39] applied both linear interpolation and cubic spline interpolation for straight and curve sub-trajectories respectively to reconstruct a new smooth trajectory. The majority of these approaches remain geometry-based approaches, unfortunately not making use of the kinematic information of the ship or historical information of the area, thus the accuracy of the results is limited [40].

Recently several Machine Learning and AI driven techniques have appeared in the literature [35], [41]–[47]. The main idea being that historical data can be used to train a deep learning model which can then forecast future positions. In [42] convolutional neural networks are used to reconstruct AIS trajectories, while the proposed method is tested and compared with cubic spline interpolation. The results show that this method is capable of higher accuracy than the cubic spline interpolation, especially when the trajectories are curved and have a high loss rate. Unfortunately, the test results cover only a small geographical area. In [41] a Long Short-Term Memory (LSTM)-based supervised learning method is used to reconstruct the vessel trajectories, achieving good results in short term forecasts, without taking into account the environmental factors of ship sailing. Authors in [48] apply a deep learning method based on Bi-directional Long Short-Term Memory Recurrent Neural Networks

(BLSTM-RNNs) for trajectory restoration. Although in complex waterways such as the Yangtze River the method demonstrates higher accuracy than linear interpolation methods, the deep learning method comes with a higher computational cost although validated in a small region two distinct reaches (Chongqing and Wuhan) in the Yangtze River.

In [46] a novel sequence-to-sequence vessel trajectory prediction model, based on encoder-decoder recurrent neural networks (RNNs) that are trained on historical trajectory data to predict future trajectory samples given previous observations, is proposed. The suggested method showcases superior accuracy than baseline methods but relies heavily upon historical data of paths belonging to the same motion pattern of the test trajectory, i.e., a large and representative sample of data from the domain. The authors mention that this potentially limits the method’s applicability since it may be sensitive to the training dataset, in particular, the number of ship trajectories available and domain representativeness. It is important to note that the majority of these approaches focus on short term forecasting (e.g., several minutes) in a regional context and not full route forecasting at a global scale.

Most current approaches to density map generation will simply remove interpolated trajectories with large missing parts due to the reconstruction difficulty[14], [41].

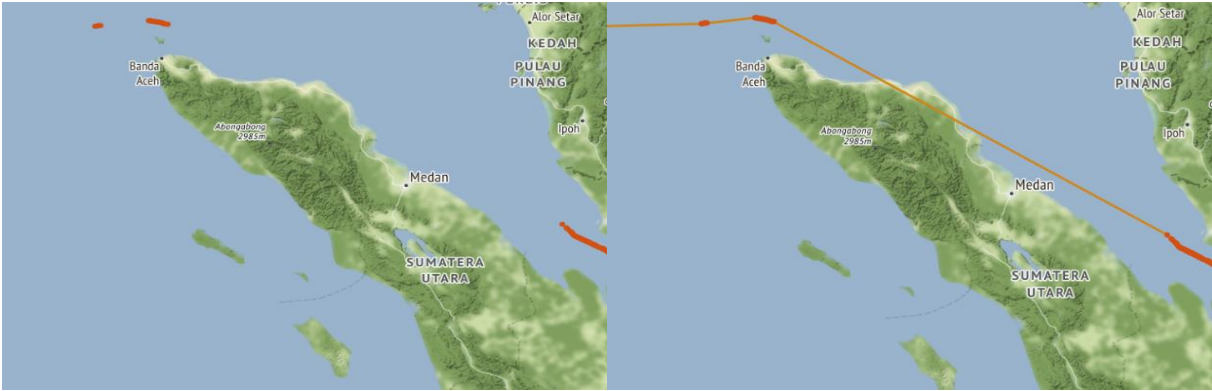


Figure 54. Vessel movement close to Medan, Malaysia. When significant time gaps occur on the raw data (left), a simple interpolation may result in erroneous trajectories (right).

The EMODnet method used for density maps, reconstructs “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. If the distance between two consecutive positions of a vessel was longer than 30 km or if the time interval was longer than 6 hours, no line was created.”. The interpolation method adopted, assumes that a straight line can be used to restore a trajectory when the course is steady, and the vessel is in linear navigation.

sea	Average Distance	Maximum Distance	50th percentile	75th percentile	90th percentile	95th percentile
ATL	20.2	43881.4	6	9.1	19.9	32

BAL	49.1	41747.5	12.7	26.1	53	84.1
BAR	11.3	34614.7	4.5	8.2	16.1	25.8
BLK	161.6	37348.8	28	73	194.1	458.3
MED	41.3	43146.5	9	18	43.1	85.5

Table 7. Monthly averaged statistics for class-A vessels for the period October to December 2021 regarding the spatial and temporal gaps, from the cleaned Sat AIS data, per sea.

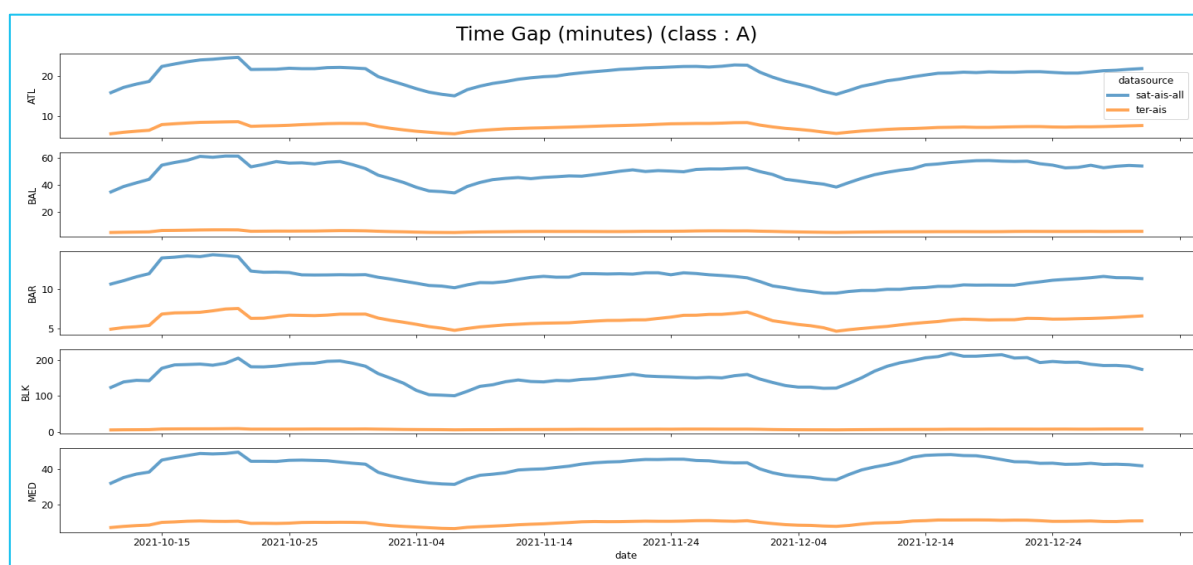


Figure 55. Average time gap between two sequential AIS messages of a vessel for Sat AIS and Ter AIS per day, time-series represent a 7 days moving average window.

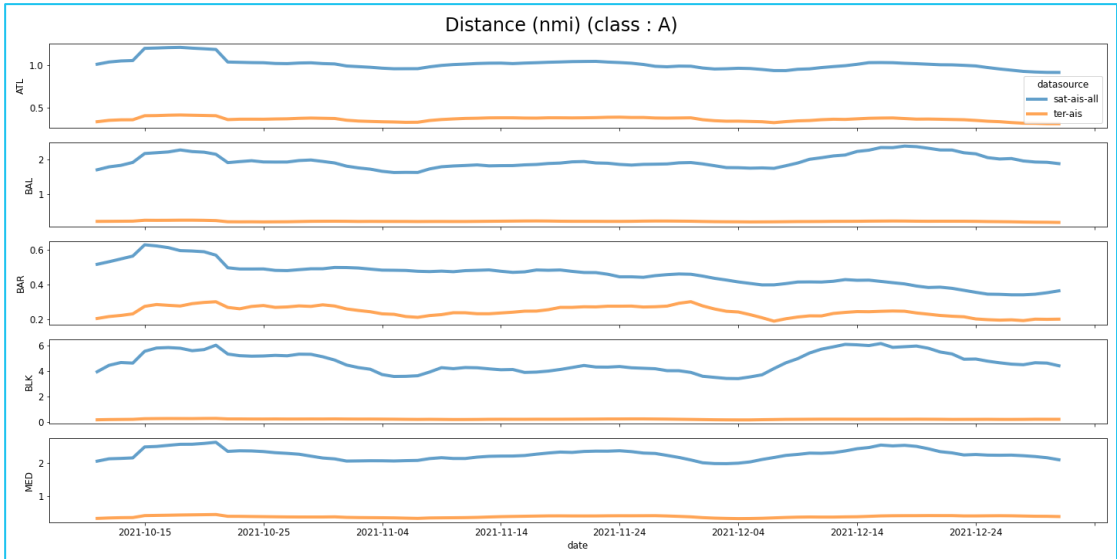


Figure 56. Average distance between two sequential AIS messages of a vessel for Sat AIS and Ter AIS per day, time-series represent a 7 days moving average window.

6.5.2 Reconstructing incomplete trajectories based on historical information

We have designed and developed a novel data driven approach for accurately reconstructing trajectories with gaps, specifically designed for restoring trajectories from Sat AIS or sparse AIS datasets. The implemented technique takes advantage of historical data to discover frequently followed paths using graph theory methods. Firstly, the movement space, i.e., the sea areas of interest, is partitioned and transformed into a grid. For the purposes of accuracy, we selected a grid cell edge length equal to 1km. Furthermore, movement of historical data is modelled using vessel transitions between neighbouring cells. Cells are considered to be neighbouring in case they have no other cells between them, i.e., each cell has at most (8) neighbours (Figure 57).

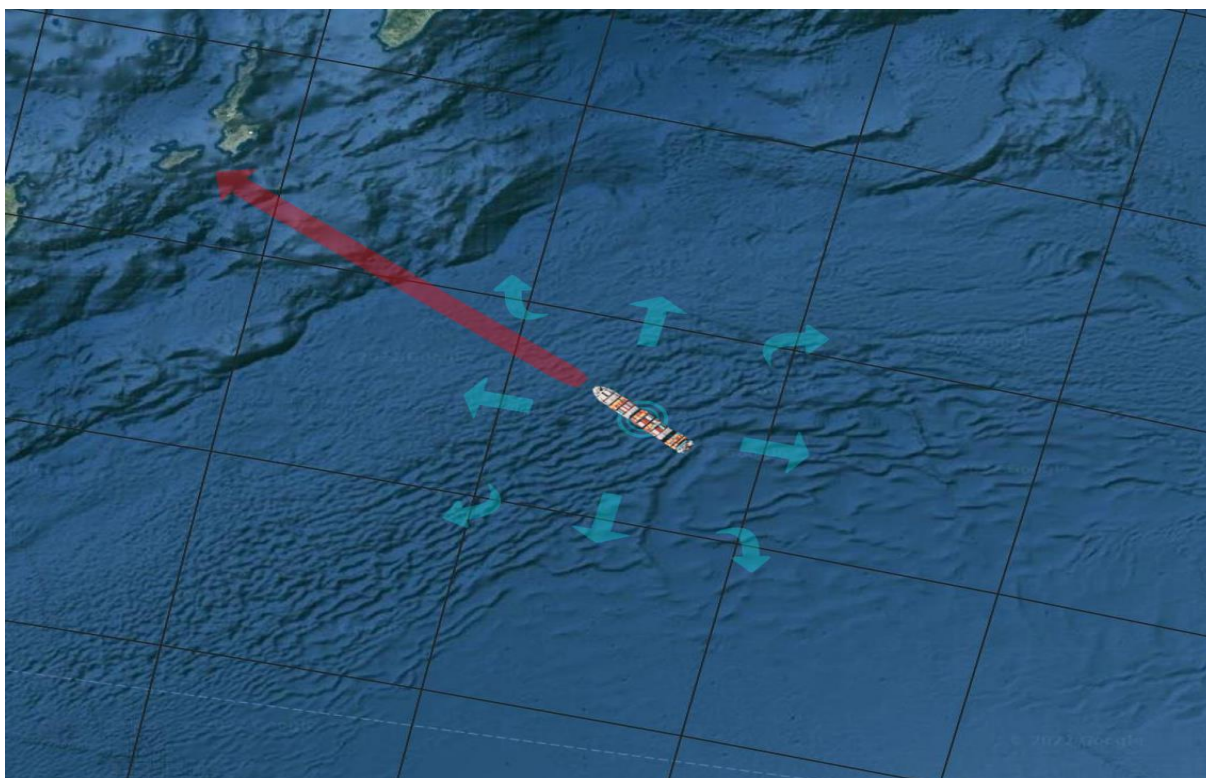


Figure 57. Each grid cell has (8) neighbours; the light blue arrows indicate all the possible transitions that may be followed by the vessel, while there can not be a single transition to the cell with the island (represented by the red arrow).

After extracting all transitions between the cells, we transform the grid into a directed weighted graph. Each graph node represents a grid cell, while the edges of the graph are used to model transitions between neighbouring cells. The weight of each edge is calculated as the (Euclidian) distance between the centres of the cells. Additionally, a second weight-penalty, modelling the patterns of vessel movement, is assigned to each edge using the formula:

$$p_h(n_i, n_j) = \begin{cases} P_E, & \text{if } N(c_i, c_j) = 0 \\ 1 - \frac{N(c_i, c_j)}{\sum_k N(c_i, c_k)}, & \text{otherwise} \end{cases}$$

where $p_h(n_i, n_j)$ is the second weight between nodes n_i and n_j , (c_i, c_j) are the corresponding neighbouring cells, $N(c_i, c_j)$ is the number of transitions between these cells and $\sum_k N(c_i, c_k)$, is the total number of outgoing transitions for cell c_i . Finally, for neighbouring cells with no transitions observed between them a weight-penalty equal to P_E is assigned. This penalty should be a real number greater than or equal to 1, indicating that a larger price needs to be paid for such a path. Note that all other weight-penalties can be bounded by 0 and 1, with smaller values assigned to more frequent transitions.

This process extracts useful information regarding vessel movement in the area that can be used to generate possible paths during AIS gaps. The process of estimating such missing paths consists of the following steps:

- I. Identify the AIS gap within a trajectory and the two cells where the gap started and ended.
- II. Use the aforementioned graph to extract a path between these cells, through an enhanced A* algorithm.
- III. Transform the returned path, expressed as a series of grid cells, to a series of real coordinates.
- IV. Assign a timestamp for each generated point, based on the length of the resulting path and the overall interval. Incorporate the resulting coordinate-timestamp pairs in the original AIS trajectory.

As mentioned, a modified version of the A-star algorithm (A*)[49] is used to generate the possible path the vessel followed during its communication gap. Designed for obtaining the shortest path between two graph nodes in a directed (weighted) graph, it uses two major components:

- A. the weighted graph
- B. a heuristic function that returns an estimation of the distance required to travel from a current to a target node.

Beginning from the 'start' node, at each step of the algorithm, the neighbours of an already visited node are investigated. For each neighbour, A* combines information from both (A) and (B) in order to update a calculated cost, in case it is part of the shortest graph. After assigning an estimated total cost to each neighbour, the current node is discarded. Then the algorithm moves on examining the one with the smallest total cost of all previously visited nodes. The A* algorithm terminates when the end node is reached, resulting in the shortest path in terms

of the graph's weights. For the calculation of the estimated total cost of each neighbour the following formula is used:

$$f(n_i) = [g(n_c) + w(n_c, n_i)] + h(n_i, n_t)$$

where $g(n_c)$ is the cumulative total cost from the start node to the current node n_c , $w(n_c, n_i)$ is the graph weight between the current node and the neighbour in question (n_i) and $h(n_i, n_t)$ is the result of the heuristic function that estimates the cost from the neighbour to the target node (n_t).

In our approach we alter the A* algorithm by incorporating the knowledge extracted from the historical data in the aforementioned cost functions. More precisely, a penalty function $p(\cdot)$ is added to the calculation of the total cost:

$$f(n_i) = [g(n_c) + w(n_c, n_i) + p(n_c, n_i)] + h(n_i, n_t)$$

where:

$$p(n_c, n_i) = w(n_c, n_i) * W_p * p_h(n_c, n_i)$$

with $w(n_c, n_i)$ being the weight (distance) between the two nodes as above, $p_h(n_c, n_i)$ being the weight-penalty based on historical data and W_p is a factor that determines the importance of the historical information during the gap-filling, ranging from 0 to 1. Additionally, since both $p_h(\cdot)$ and W_p are bounded by 0 and 1 (with the exception of a P_E penalty larger than one in the case of no transitions were found in the historical data), the $p(n_c, n_i)$ function is in turn bounded by 0 and the real distance between the cells, thus providing a penalty normalised to the original graph.

For the purpose of an experimental evaluation of this approach, a dataset with synthetic AIS gaps was created. More precisely, focusing on the eastern part of the Mediterranean Sea, i.e., the Aegean and Levantine Sea, a sample of (3000) trajectories was extracted. The combined dataset from both Ter AIS and Sat AIS from the month of October was used as a starting point. The extracted trajectories were continuous vessel paths, with no temporal gap larger than 15 minutes between consecutive messages. Additionally, in order to avoid including vessel stops at port areas, segments shorter than 30 minutes where the vessel remained stopped/idle were removed from the trajectories. These segments were substituted by a single transition, from the start of the stop till the end, with the corresponding time interval deducted from the total trip duration. Moreover, for each trajectory a gap was formed using the starting and ending points.

For each synthetic gap a reconstructed trajectory was exported using the proposed mechanism. Afterwards, the results of this process, as well as the results of a straight-line interpolation between the gap points were evaluated compared to the true path the trajectory followed. Since the resulting trajectories do not have the same number of points the FastDTW (Fast Dynamic Time Warping) [50] metric was used to determine the approach's accuracy. As depicted in Table 8 the aforementioned approach provides comparable results to the straight-line interpolation for gaps that range from 1 to 3 hours. A significant improvement can be seen when the temporal gaps have a duration greater than three hours, with an improvement of over 16% and 26% for gaps of 3-6 and 6-24 hours respectively.

FastDTW (km)						
Trajectory gap (hours)	1-3	Impr. (%)	3-6	Impr. (%)	6-24	Impr. (%)
Number of trajectories	1655	-	664	-	681	-
Straight line interpolation	1.94	-	5.38	-	16.45	-
A* hist. ($W_p = 0.1$)	1.91	1.67	4.5	16.34	12.16	26.08
A* hist. ($W_p = 0.3$)	1.91	1.44	4.47	16.91	12.11	26.37
A* hist. ($W_p = 0.5$)	1.92	1.19	4.52	15.83	12.12	26.29
A* hist. ($W_p = 0.7$)	1.94	0.3	4.61	14.21	12.55	23.67

Table 8. Accuracy of the gap filling method (*A* hist.*), compared to the vessel's true path, using the FastDTW algorithm. The Improvement (*Imprv.*) percentage is calculated with regard to the straight-line interpolation results as: $Impr = 100 * \frac{Sli - Ash}{Sli}$, where *Sli* is the FastDTW for the straight line interpolation and *Ash* the accuracy of the proposed method.

6.6 Corrected Density maps

This gap-filling mechanism is then used in order to reconstruct trajectories when large temporal gaps between AIS messages occur (see 0). The inclusion of these trajectories allows for a more accurate density calculation, throughout the area of interest.

This process was applied upon the cleaned AIS data for the month of October, focusing on the Aegean and Levantine Sea. More precisely, the dataset from the combined Ter AIS and Sat AIS was used, as well as the dataset for all Sat AIS for comparison. For this analysis, gaps solely from class A vessels were reconstructed, since smaller vessels (class B) tend not to follow common routes. During the gap filling process, the gaps with duration from 30 minutes to 24 hours were reconstructed. Also, a W_h factor of 0.3 was selected since it provided the best accuracy during the evaluation, while the resolution of the end density map results was chosen to be 10km. For the purposes of better visualising the contribution upon the resulting density, some difference maps were generated, similarly with the Ter AIS and Sat AIS comparisons (Figure 59 and Figure 60). In these maps, the green colour indicates no density change from the gap filling process, while blue and purple points to an increase because of it.



Figure 58. Examples from the gap-filling process. The original trajectory is depicted in green, while the reconstructed segments during the gaps are in orange.

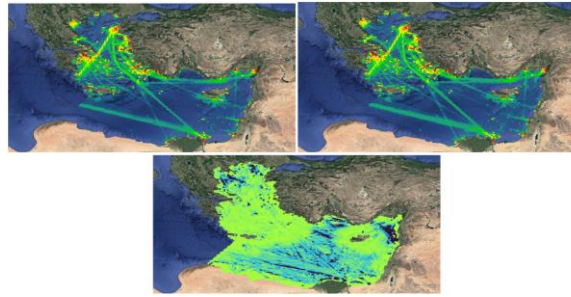


Figure 59. Density maps for the month of October for all vessel types, from all AIS data (top), for an improved dataset using the gap filling process (middle) and the difference map.

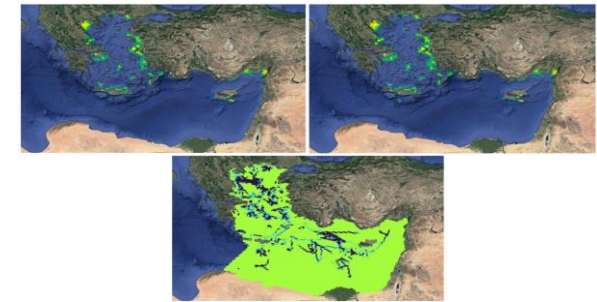


Figure 60. Density maps for the month of October for Fishing vessels, from all AIS data (top), for an improved dataset using the gap filling process (middle) and the difference map.

As depicted in the above figures, the gap filling process provides useful information regarding the movement of vessels during these gaps. In more detail, it can be seen that some main corridors of movement are presented as more complete while also some new pathways for specific vessel types were discovered and enhanced, e.g., for the fishing vessels. Such conclusions are more obvious when processing datasets composed only of Sat AIS data, where AIS gaps are much more common.



Figure 61. Density maps for the month of October for all vessel types, from Sat AIS data (top), for an improved dataset using the gap filling process (middle) and the difference map



Figure 62. Density maps for the month of October for Cargo vessels, from Sat AIS data (top), for an improved dataset using the gap filling process (middle) and the difference map.

7. Conclusions and future work

7.1 Conclusion

The study performed by MarineTraffic (in collaboration with Ubitech) has provided a more precise insights regarding the accurate generation of vessel density maps.

The project has gone through all the stages required for this kind of analysis:

- State of the art analysis
- Collection and transformation of the data sources
- Raster based spatial analysis and performance comparisons
- Proposed potential corrections and improvements

The overall goal of this study to investigate the capabilities and capacities of AIS datasets for accurate density maps generation, and especially Satellite AIS capabilities and limitations, while exploring its complementarity with other sources (such as Terrestrial AIS and EO images). The results obtained from the analysis performed in this study clearly indicate the feasibility of these systems to capture AIS signals from space, outperforming Ter AIS in terms of coverage but underperforming in terms of accuracy and reliability.

Regarding the methodology applied for assessing the capabilities of all vessel detection systems, an approach for quantitatively and qualitatively measuring performance was adopted.

Performance measurements, counted

1. the number of unique vessels received in a defined time interval, both at a regional scale and across EU waters
2. the average distance between two sequential AIS messages of a vessel for the three Sat AIS providers per day.
3. Average distance between two sequential AIS messages of a vessel for Sat AIS and Ter AIS per day.

These indicators specify in particular each platforms capabilities in terms of capacity to detect the correct number of vessels in an area, but also the granularity of messages or detections.

It must be noted that the capability of Sat AIS to identify a vessel strongly depends on the class of the AIS device used, the number of ships in the surrounding area and the density of traffic. Due to several reasons analysed in this study, Sat AIS strongly underperforms when detecting CLASS B AIS devices: thus, smaller ships. Overall, the performance test results returned promising results over low-density shipping areas for Sat AIS, while in areas of high density

the performance of Sat AIS deteriorated (due to saturation). Interestingly with the exception of Barents Sea, Ter AIS mostly outperformed Sat AIS in EU waters. Following this we are able to highlight areas and conditions of potentially low detection from Sat AIS, using Ter AIS as the ground truth for further analysis.

Overall, the performance test results returned promising results for all sensor platforms with interesting results returned for the vessel detection from EO observation images. Especially Sentinel-2 imagery proves to be a valuable source of information and in certain conditions can even capture small vessels (e.g., class B), although limited by the cloud coverage combined with low revisit time of ESA satellites. For density-maps use case, though, Sentinel-1 proves to be better due to the fact that SAR products are not affected by weather conditions. Sentinel-1 helped us uncover areas where density is higher than the one perceived via AIS, as well as dark areas and/or areas prone to AIS packet collisions. Although Sentinel-1 revisit time is still worse than the frequency of AIS messages, it still outperforms Sat AIS. As a result of this study, we would argue that all sensors should be used in order to obtain a complete picture of the actual vessel density of an area.

We must note that a one-to-one comparison of the performance of all sensors can be misleading, due to coverage areas selected or traffic density in the region.

Due to the orbits of the satellites and the temporal “snapshots” used for the analysis, it is not always possible to provide a 100% reference ground truth for comparisons.

Based on the findings the following overall conclusions can be drafted:

- Sat AIS and Ter AIS are strongly complementary. While Ter AIS may have higher granularity and accuracy in the majority of cases, Sat AIS can provide data outside the Ter AIS coverage.
- Areas and conditions where Set AIS underperform can be mapped and used for reference
- EO imagery highly complements the “image” above by detecting vessels often unseen by the other sources (interference, device malfunction, congestions etc.)
- Incomplete trajectories can be accurately repaired using historical data and should not be discarded from the datasets.

We presented a novel trajectory reconstruction method that can be used for density map generation. This makes use of historical trajectory data to reconstruct a lossy trajectory as accurately as possible; thus not disregarding any data. We demonstrate that this leads to much more accurate density maps.

The main outcomes of the project are:

- Performance analysis highlights the strong complementarity of all above sources
- EO detection fusion pipelines implemented have proved promising results
- The proposed novel trajectory reconstruction method based on historical data provides an accurate method for improving vessel density maps

- The lack of freely available software for this kind of work led to the public release of a toolbox providing a number of modules to support handling AIS data while improving their transformation into actionable visualisations such as density maps

Overall, this study will provide valuable inputs for improving the knowledge in vessel activities in EU waters.



The MarineTraffic Toolbox is published under the Creative Commons license (Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0))

It is available online at

<https://www.marinetraffic.com/research/the-marinetraffic-ais-toolbox/>

7.2 Nest steps

Regarding future activities, the project identified the following issues to be promoted

- This type of analysis and data fusion have proved promising results. But several questions remain open in relation to the EO detection capacity. This approach should be further developed by setting up more demonstration projects, which would explore several EO platforms and their capabilities, and also fine tuning part of this process. In this study we used Sentinel 1 and Sentinel 2, it might be worthwhile to consider a similar approach using commercial providers.
- During the course of the project, the value of historical information was highlighted for trajectory reconstruction. Such data holds insights for several important projects, such as MSP, changing patterns of life detection and similar. A specific mechanism and funding at an EU level would be worthwhile exploring to consider such potential.
- Another issue is related to “dark vessels”. Throughout the duration of the project, we detected a number of vessels that due to interference, malfunctioning devices or other reasons were not transmitting their AIS messages although within coverage. It is proposed to further investigate the AIS signal behavior and specifically “dark vessel” detection methods through information fusion as this can be of exceptional importance in cases of sanctions avoidance, illegal fishing and others.

References

- [1] "Technical Note 4.1 Vessel Density Mapping. Preparatory Action for Assessment of the Capacity of Spaceborne Automatic Identification System Receivers to Support EU Maritime Policy. DG MARE Service Contract : No MARE/2008/06/ SI2.517298," 2010. [Online]. Available: https://webgate.ec.europa.eu/maritimeforum/system/files/6039_PASTA%20MARE_GH_TN-004-1_Density_Plot_I4.pdf
- [2] R. Challamel, T. Calmettes, and C. N. Gigot, "A European hybrid high performance Satellite-AIS system," in *2012 6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC)*, Sep. 2012, pp. 246–252. doi: 10.1109/ASMS-SPSC.2012.6333084.
- [3] F. Clazzer, F. Lázaro, and S. Plass, "Enhanced AIS receiver design for satellite reception," *CEAS Space J*, vol. 8, no. 4, pp. 257–268, Dec. 2016, doi: 10.1007/s12567-016-0122-8.
- [4] Poļevskis, J.; Krastiņš, M.; Korāts, G.; Skorodumovs, A.; Trokšs, J., "Methods for Processing and Interpretation of AIS Signals Corrupted by Noise and Packet Collisions.," *Latvian Journal of Physics & Technical Sciences*. Jan2012, Vol. 49 Issue 3, p25-31. 7p., Accessed: Jun. 09, 2022. [Online]. Available: <https://web.p.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authType=crawler&jrnl=08688257&AN=84555587&h=eCG09akcCIDP47wwca%2bBHR8iCEX7V7Flx7%2bDuDcrKB2rwmSWjdNOr40F4KgKrrdxGk0vdsELvoI9jKKHT7erQ%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authType%3dcrawler%26jrnl%3d08688257%26AN%3d84555587>
- [5] K. Metcalfe et al., "Using satellite AIS to improve our understanding of shipping and fill gaps in ocean observation data to support marine spatial planning," *Journal of Applied Ecology*, vol. 55, no. 4, pp. 1834–1845, 2018, doi: 10.1111/1365-2664.13139.
- [6] *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE, THE COMMITTEE OF THE REGIONS AND THE EUROPEAN INVESTMENT BANK A Clean Planet for all A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy*. 2018. Accessed: Jun. 07, 2022. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0773>
- [7] "Paris Agreement." https://ec.europa.eu/clima/eu-action/international-action-climate-change/climate-negotiations/paris-agreement_en (accessed Jun. 07, 2022).
- [8] "Marine Renewable Energy - an overview | ScienceDirect Topics." <https://www.sciencedirect.com/topics/engineering/marine-renewable-energy> (accessed Jun. 07, 2022).
- [9] M. R. MacLeod and W. M. Wardrop, "Operational Analysis at Combined Maritime Forces," p. 14.
- [10] Zissis D., Chatzikokolakis K., Spiliopoulos G. and Vodas M, "A distributed spatial method for modeling maritime routes," *IEEE Access*, doi: 10.1109/ACCESS.2020.2979612.

- [11] ITU, "Rec. ITU-R M.1371-5:Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band," ITU, Rec. ITU-R M.1371-5, 2014.
- [12] ORBCOMM, "WHITE PAPER: UNDERSTANDING SATELLITE AIS AND THE SDPOB ADVANTAGE."
- [13] M. Hodkiewicz and N. Montgomery, "Data fitness for purpose: assessing the quality of industrial data for use in mathematical models," Apr. 2014.
- [14] Luigi Falco, Alessandro Pittito, William Adnams, Nick Earwaker, Harm Greidanus, "EU Vessel density map-Detailed method." [Online]. Available: https://www.emodnet-humanactivities.eu/documents/Vessel%20density%20maps_method_v1.5.pdf
- [15] M. Vespe, M. Gibin, A. Alessandrini, F. Natale, F. Mazzarella, and G. C. Osio, "Mapping EU fishing activities using ship tracking data," *Journal of Maps*, vol. 12, no. sup1, pp. 520–525, Aug. 2016, doi: 10.1080/17445647.2016.1195299.
- [16] L. Wu, Y. Xu, Q. Wang, F. Wang, and Z. Xu, "Mapping Global Shipping Density from AIS Data," *The Journal of Navigation*, vol. 70, no. 1, pp. 67–81, Jan. 2017, doi: 10.1017/S0373463316000345.
- [17] R. Wawrzaszek, M. Waraksa, M. Kalarus, G. Juchnikowski, and T. Górski, "Detection and Decoding of AIS Navigation Messages by a Low Earth Orbit Satellite," in *GeoPlanet: Earth and Planetary Sciences*, 2019, pp. 45–62. doi: 10.1007/978-3-319-94517-0_4.
- [18] P. Last, M. Hering-Bertram, and L. Linsen, "How automatic identification system (AIS) antenna setup affects AIS signal quality," *Ocean Engineering*, vol. 100, pp. 83–89, May 2015, doi: 10.1016/j.oceaneng.2015.03.017.
- [19] P. Last, C. Bahlke, M. Hering-Bertram, and L. Linsen, "Comprehensive Analysis of Automatic Identification System (AIS) Data in Regard to Vessel Movement Prediction," *The Journal of Navigation*, vol. 67, no. 5, pp. 791–809, Sep. 2014, doi: 10.1017/S0373463314000253.
- [20] M. Redoutey, E. Scotti, C. Jensen, C. Ray, and C. Claramunt, "Efficient Vessel Tracking with Accuracy Guarantees," in *Web and Wireless Geographical Information Systems*, Berlin, Heidelberg, 2008, pp. 140–151. doi: 10.1007/978-3-540-89903-7_13.
- [21] "Introduction to Geospatial Resources and Formats." https://www.loc.gov/preservation/digital/formats/content/gis_intro.shtml (accessed Jun. 10, 2022).
- [22] "Grid-based Map Analysis in GIS Curriculum." http://www.innovativegis.com/basis/present/GIS07_Grid_teaching/Grid_teaching07.htm (accessed Jun. 10, 2022).
- [23] "Geographic Information Science and Systems, 4th Edition | Wiley," *Wiley.com*. <https://www.wiley.com/en-us/Geographic+Information+Science+and+Systems%2C+4th+Edition-p-9781119031307> (accessed Jun. 10, 2022).
- [24] E. d’Afflisio, P. Braca, and P. Willett, "Malicious AIS Spoofing and Abnormal Stealth Deviations: A Comprehensive Statistical Framework for Maritime Anomaly Detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 4, pp. 2093–2108, Aug. 2021, doi: 10.1109/TAES.2021.3083466.
- [25] A. Androjna, M. Perkovič, I. Pavic, and J. Mišković, "AIS Data Vulnerability Indicated by a Spoofing Case-Study," *Applied Sciences*, vol. 11, no. 11, Art. no. 11, Jan. 2021, doi: 10.3390/app11115015.
- [26] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, 1st edition. Reading, Massachusetts Wokingham Amsterdam: Pearson, 1991.
- [27] H. Greidanus, M. Alvarez, C. Santamaria, F.-X. Thoorens, N. Kourti, and P. Argentieri, "The SUMO Ship Detector Algorithm for Satellite Radar Images," *Remote Sensing*, vol. 9, no. 3, Art. no. 3, Mar. 2017, doi: 10.3390/rs9030246.

- [28] C. Corbane, L. Najman, E. Pecoul, L. Demagistri, and M. Petit, "A complete processing chain for ship detection using optical satellite imagery," *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5837–5854, Dec. 2010, doi: 10.1080/01431161.2010.512310.
- [29] U. Kanjir, H. Greidanus, and K. Oštir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sensing of Environment*, vol. 207, pp. 1–26, Mar. 2018, doi: 10.1016/j.rse.2017.12.033.
- [30] K. Bereta, R. Grasso, and D. Zissis, "Vessel Detection using Image Processing and Neural Networks," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, Sep. 2020, pp. 2276–2279. doi: 10.1109/IGARSS39084.2020.9323883.
- [31] "Object-based image analysis approach for vessel detection on optical and radar images." <https://www.spiedigitallibrary.org/journals/journal-of-applied-remote-sensing/volume-13/issue-01/014502/Object-based-image-analysis-approach-for-vessel-detection-on-optical/10.1117/1.JRS.13.014502.full?SSO=1> (accessed Jun. 13, 2022).
- [32] K. Bereta, G. Karantaidis, and D. Zissis, "Vessel Traffic Density Maps based on Vessel Detection in Satellite Imagery."
- [33] Wang, C, X. Wang, J. Zhu, and G. Wang, "A Survey of Radar and AIS Information Fusion," presented at the Commnad Control & Simulation, 2009. Accessed: May 28, 2022. [Online]. Available: http://en.cnki.com.cn/Article_en/CJFDTOTAL-QBZH200902004.htm
- [34] G. Scaglia, E. Serrano, A. Rosales, and P. Albertos, "Linear interpolation based controller design for trajectory tracking under uncertainties: Application to mobile robots," *Control Engineering Practice*, vol. 45, pp. 123–132, Dec. 2015, doi: 10.1016/j.conengprac.2015.09.010.
- [35] X. Jie, W. Chaozhong, C. Zhijun, and C. Xiaoxuan, "A novel estimation algorithm for interpolating ship motion," in *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, Dec. 2017, pp. 557–562. doi: 10.1109/ICTIS.2017.8047821.
- [36] B. Brandoli *et al.*, "From multiple aspect trajectories to predictive analysis: a case study on fishing vessels in the Northern Adriatic sea," *Geoinformatica*, Mar. 2022, doi: 10.1007/s10707-022-00463-4.
- [37] P. Kolendo and R. Smierzchalski, "Experimental Comparison of Straight Lines and Polynomial Interpolation Modeling Methods in Ship Evolutionary Trajectory Planning Problem," vol. 386, pp. 331–340, Aug. 2016, doi: 10.1007/978-3-319-23180-8_24.
- [38] L. Zhang, Q. Meng, Z. Xiao, and X. Fu, "A novel ship trajectory reconstruction approach using AIS data," *Ocean Engineering*, vol. 159, pp. 165–174, Jul. 2018, doi: 10.1016/j.oceaneng.2018.03.085.
- [39] X. Zhang, Y. He, R. Tang, J. Mou, and S. Gong, "A Novel Method for Reconstruct Ship Trajectory Using Raw AIS Data," in *2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, Sep. 2018, pp. 192–198. doi: 10.1109/ICITE.2018.8492619.
- [40] S. Guo, J. Mou, L. Chen, and P. Chen, "Improved kinematic interpolation for AIS trajectory reconstruction," *Ocean Engineering*, vol. 234, p. 109256, Aug. 2021, doi: 10.1016/j.oceaneng.2021.109256.
- [41] M. Liang, R. W. Liu, Q. Zhong, J. Liu, and J. Zhang, "Neural Network-Based Automatic Reconstruction of Missing Vessel Trajectory Data," in *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, Mar. 2019, pp. 426–430. doi: 10.1109/ICBDA.2019.8713215.
- [42] S. Li, M. Liang, X. Wu, Z. Liu, and R. W. Liu, "AIS-Based Vessel Trajectory Reconstruction with U-Net Convolutional Networks," in *2020 IEEE 5th International Conference on Cloud*

- Computing and Big Data Analytics (ICCCBDA)*, Apr. 2020, pp. 157–161. doi: 10.1109/ICCCBDA49378.2020.9095616.
- [43] X. Chen *et al.*, “Ship Trajectory Reconstruction from AIS Sensory Data via Data Quality Control and Prediction,” *Mathematical Problems in Engineering*, vol. 2020, p. e7191296, Aug. 2020, doi: 10.1155/2020/7191296.
- [44] D. Zissis, E. K. Xidias, and D. Lekkas, “Real-time vessel behavior prediction,” *Evolving Systems*, vol. 7, no. 1, pp. 29–40, Mar. 2016, doi: 10.1007/s12530-015-9133-5.
- [45] D. Zissis, E. K. Xidias, and D. Lekkas, “A cloud based architecture capable of perceiving and predicting multiple vessel behaviour,” *Applied Soft Computing*, vol. 35, pp. 652–661, Oct. 2015, doi: 10.1016/j.asoc.2015.07.002.
- [46] S. Capobianco, L. M. Millefiori, N. Forti, P. Braca, and P. Willett, “Deep Learning Methods for Vessel Trajectory Prediction Based on Recurrent Neural Networks,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 6, pp. 4329–4346, Sep. 2021, doi: 10.1109/TAES.2021.3096873.
- [47] H. Georgiou *et al.*, “Future Location and Trajectory Prediction,” in *Big Data Analytics for Time-Critical Mobility Forecasting: From Raw Data to Trajectory-Oriented Mobility Analytics in the Aviation and Maritime Domains*, G. A. Vouros, G. Andrienko, C. Doulkeridis, N. Pelekis, A. Artikis, A.-L. Joussetme, C. Ray, J. M. Cordero, and D. Scarlatti, Eds. Cham: Springer International Publishing, 2020, pp. 215–254. doi: 10.1007/978-3-030-45164-6_8.
- [48] C. Zhong, Z. Jiang, X. Chu, and L. Liu, “Inland Ship Trajectory Restoration by Recurrent Neural Network,” *The Journal of Navigation*, vol. 72, no. 6, pp. 1359–1377, Nov. 2019, doi: 10.1017/S0373463319000316.
- [49] P. E. Hart, N. J. Nilsson, and B. Raphael, “A Formal Basis for the Heuristic Determination of Minimum Cost Paths,” *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, Jul. 1968, doi: 10.1109/TSSC.1968.300136.
- [50] S. Salvador and P. Chan, “Toward accurate dynamic time warping in linear time and space,” *Intell. Data Anal.*, vol. 11, no. 5, pp. 561–580, Jul. 2007.

Appendix

Technical Specifications

The presented methods were developed and tested with the following technical specifications.

Operating System	GNU/Linux Ubuntu 20.04.3 LTS (x86_64)
Programming Language	Python 3.8

The (primary) required Python libraries and tools for our implementation are listed in the following table.

Python Tool / Library	Version	Primary Purpose
json	json-c 0.15	Data file handling
jsonschema	4.5.1	Data file handling
sys	-	Data file handling
glob	-	Data file handling
fiona	1.8.21	Data file handling
pickleshare	0.7.5	Data file handling
csv	-	Data handling
bz	bzip2 1.0.8	Data handling
pandas	1.4.2	Data handling
geopandas	0.10.2 pygeos=0.12.0	Data and geometry handling
pyproj	3.3.1	Data transforming
scipy	1.7.3	Data transforming
numpy	1.22.4	Data transforming

fiona	1.8.20	Geometry file handling
shapely	1.8.2	Geometry handling
haversine	2.5.1	Geometry data handling
osgeo	-	Map (TIFF) creation
networkx	2.8.2	Graph queries for trajectory reconstruction
asttokens (ast)	2.0.5	Data storing
math	-	Mathematics manipulations
concurrent.futures	-	Process parallelization
fastdtw	0.3.4	Trajectory comparison
Other tools	Version	Primary Purpose
GDAL (gdal_rasterize, gdaldem)	2.2.3	Map (TIFF) creation