



HELCOM Recommendation 33/1-Rev.2

Adopted 6 March 2012 and
amended 7 March 2019 and 7 August 2022
having regard to Article 20, Paragraph 1 b)
of the Helsinki Convention

UNIFIED INTERPRETATION IN RELATION TO ACCESS TO AND USE OF HELCOM AIS DATA

THE COMMISSION,

RECALLING the HELCOM Extraordinary Ministerial Meeting in Copenhagen on 10 September 2001 which decided to establish a Working Group with the purpose of facilitating mutual exchange and deliveries of AIS data, including the construction of the monitoring system for the maritime traffic in the Baltic Sea Area,

RECALLING FURTHER Regulation 9 of Annex IV to the Helsinki Convention on the use of Automatic Identification System (AIS),

ACKNOWLEDGING the successful establishment of the HELCOM Automatic Identification System as regulated by the HELCOM Agreement on Access to AIS Information, and its contribution to the improved safety of navigation and protection of the marine environment in the Baltic Sea area,

APPRECIATING the cooperation among the Baltic Sea countries as well as Norway in maintaining and further developing the HELCOM AIS, taking place within the HELCOM Expert Working Group for Mutual Exchange and Deliveries of AIS data,

HAVING IN MIND the HELCOM Baltic Sea Action Plan, adopted by the HELCOM Ministerial Meeting on 15 November 2007, and the agreement included therein to amend the HELCOM Agreement on Access to AIS Information by 2008 taking into account the proposal elaborated by the 16th Meeting of the HELCOM Expert Working Group for Mutual Exchange and Deliveries of AIS data (HELCOM AIS EWG 16/2007),

OBSERVING the growing interest to use HELCOM AIS data, both within the Baltic Sea countries and Norway as well as in other geographical regions,

RECALLING the successfully completed migration of the HELCOM AIS server from Denmark (Danish Maritime Authority) to Norway (Norwegian Coastal Administration) in January 2017,

RECOGNIZING the need for common interpretation regarding the use of and access to HELCOM AIS data, in light of new services that have been made available to the AIS users, including web based historical data extract, web based real-time AIS display and real-time AIS data exchange,

RECOMMENDS the Governments of the Contracting States as well as Norway to apply the attached Guidelines on unified interpretation in relation to access to and use of HELCOM AIS,

REQUESTS the AIS Expert Working Group to regularly evaluate the implementation of this Recommendation and amend the attached Guidelines as need be.

Attachment

Guidelines on unified interpretation in relation to access to and use of HELCOM AIS data

Definitions

The definitions used by IALA and IMO are used in this document.

Objective

This Recommendation provides for exchange of HELCOM Automatic Identification System (AIS) data free of charge between the countries in accordance with the HELCOM Copenhagen Declaration adopted on 10 September 2001 in Copenhagen. It also provides for the distribution and use of the HELCOM AIS data received from the Participating Parties (HELCOM Contracting States and Norway) with a general aim of enhancing the maritime safety and protection of the Baltic Sea environment as provided for in the Helsinki Convention. Additionally, the Recommendation includes a procedure to deal with all other interests in use of these HELCOM AIS data.

Access to information

Each Participating Party shall make HELCOM AIS data available for access, as decided by the AIS Expert Working Group (AIS EWG) via Internet and as outlined in **Appendix 2**.

HELCOM AIS data from all ships carrying AIS as mandatory equipment shall be made available, taking into account also the provisions of **Appendix 1**.

Distribution and use of *real-time, historical and statistical* HELCOM AIS data

Each Participating party must restrict the distribution of received HELCOM AIS data to the users and uses as expressed in **Appendix 1**.

Each Participating Party is responsible for the use of distributed HELCOM AIS data and must take appropriate actions to ensure that the recipients do not redistribute HELCOM AIS data and that HELCOM AIS data are only used for purposes listed in **Appendix 1**.

Any Participating Party may, however, distribute data to a third person in the event that the participating nation is obliged by law to disclose information according to principles of public access to official records.

In order to deal with interests in distribution and use of HELCOM AIS data beyond the uses and/or users stipulated in **Appendix 1**, the following procedure shall be applied:

1. The HELCOM Secretariat will act as a contact point for receiving applications for granting access to and use of HELCOM AIS data.
2. The Secretariat will verify the information contained in an application for access to HELCOM AIS data.
3. The Secretariat will dispatch via e-mail the verified application to all Participating Parties represented by the Heads of Delegation to HELCOM and contact persons for HELCOM Agreement on Access to AIS Information and the Norwegian Coastal Administration.
4. The decision by the Heads of Delegation to HELCOM and the Norwegian Coastal Administration on granting access to HELCOM AIS data or refraining from doing so will be communicated to the HELCOM Secretariat in writing.
5. Access to HELCOM AIS data will be granted only if such access is approved by all the Participating Parties.
6. The Participating Parties should consider the application and communicate the decision made to the HELCOM Secretariat within 30 days after the dispatch of the request to the Participating Parties. The HELCOM Secretariat should inform the applicant about the decision made within a fortnight.
7. A standard agreement on the use of HELCOM AIS data, providing for possible restrictions in access to HELCOM AIS data, will be signed between the applicant and HELCOM. The standard agreement is provided in **Appendix 3**.

8. Baltic Sea shipping density maps are generated using HELCOM AIS data. The methodology for generating such maps is available as Appendix 4 of this recommendation. The maps are available per year as well as filtered per ship type (table 3 in Appendix 4).

Information concerning private persons

When HELCOM AIS data contain information concerning private persons the Participating Parties undertake to ensure protection of the privacy of these persons in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation (GDPR)) (for EU countries) and also to other international, European Community and national law and regulations as applicable.

The General Data Protection Regulation (GDPR) regulates the processing by an individual, a company or an organisation of personal data relating to individuals in the EU. In the HELCOM AIS dataset, the data received from national servers should be shared when a request to access HELCOM AIS data is approved (cf. Appendix 3). The Secretariat may on a case by case basis reserve its right not to share AIS Class B data, if it concludes that it would violate the GDPR.

The HELCOM AIS data is stored by the HELCOM Secretariat on a dedicated network drive protected by firewall. The HELCOM AIS data contains information about the position of the ships at a given time as well as information related to the voyage (speed and course over ground). It is possible to identify each ships using the Maritime Mobile Service Identity (MMSI).

No warranties

The transmitting or sending party makes no express or implied warranty as to any matter whatsoever, including the availability, accuracy, or reliability of any information or data, whether tangible or intangible, made, developed or supplied under this Recommendation, or the ownership, merchantability, or fitness for a particular purpose of the information, of the data made, developed or supplied.

Force Majeure

Neither Participating party shall be liable for any unforeseeable event beyond its reasonable control not caused by the fault or negligence of such party, which causes such party to be unable to perform its obligations under the Agreement on Access to AIS Information, including, but not limited to, flood, drought, earthquake, storm fire, pestilence, lightning and other natural catastrophes, epidemic, war, riot, civic disturbance or disobedience, strikes, labour dispute, or failure, threat or failure, or sabotage, or any order or injunction made by a court or public agency. In the event of the occurrence of such a Force Majeure event, the party unable to perform shall promptly notify the other party. It shall further use its best efforts to resume performance as quickly as possible and shall suspend performance only for such period of time as is necessary as a result of the Force Majeure event.

Contact persons

A list of names and contact details for contact persons will be kept and updated by the HELCOM Secretariat, with the consent of these contact persons.

Termination of access to AIS-data

Any failure to comply with this Recommendation gives the sending or transmitting party right to terminate access to AIS-data for as long as the receiving party is not complying with the Recommendation.

Appendix 1

APPLICATIONS FOR RETRIEVED HELCOM AIS DATA FROM THE COMMON BALTIC SEA AIS

Real-time HELCOM AIS data in this agreement is defined to be:

- delivered end-to-end non-stop, one vessel report after the other as soon as they are transmitted
- delivered promptly when they are received at the base station
- delivered without any delay (additional latency)
- not sent in blocks
- irrespective of the reporting interval (not all messages received by the national centre are relayed/forwarded to the regional centre. The reporting interval is less frequent)

Type of access:

For each specific case some restrictions – at the discretion of a Participating Party - can be applied, where applicable and justified, such as: time limitation, geographical limitation and/or update rate.

Whenever access is implemented directly from the HELCOM AIS Information Centre, it is given for all Participating Parties' data* and at the update rate in-force.

If an access to a full update rate is needed this should be solved on a bilateral basis.

Type of HELCOM AIS data:

- Real time (R),
- Statistic data (S),
- Historical data (H)
- Web based visualization of AIS real-time data (V)

User	Uses	Type of HELCOM AIS data
HELCOM Secretariat	Aiming at implementation of the Helsinki Convention	S, H, V
National administration, including accident investigation authorities and any research institutes or organizations or their contractors in the Baltic Sea states and Norway acting according to the uses set herein	<ul style="list-style-type: none"> - Pollution preventing and combating - VTS (Vessel Traffic Services) - Port State Control (PSC) - Contingency planning - International Ship and Port Security (ISPS) - Search and Rescue (SAR) - Accident investigation - Traffic planning, efficiency and management, incl. icebreaking services - Mandatory reporting system for HAZMAT reporting requirements - Pilotage - Customs surveillance - Science and research supporting the implementation of the Helsinki Convention and for preparing IMO ships routeing measures 	R, S, H, V
EU institutions Institutions in Russia	Aiming at implementation of the Helsinki Convention, including: <ul style="list-style-type: none"> - Mandatory reporting system for HAZMAT reporting requirements - Traffic planning - Pollution prevention - Maritime safety and security 	R, S, H, V

* Except for real-time streamed data that can be filtered on the basis of the countries' HELCOM AIS data

Appendix 2

Description of connection interface

Purpose

This document describes the interface relevant to the Participating Parties wanting to exchange HELCOM AIS data with the HELCOM server.

Interface Description

This section describes the interface used when supplying data to and receiving data from the HELCOM server.

Data format

The data must adhere to the latest edition of the IEC 61162-1 standard (Maritime Navigation and Radio communication Equipment and Systems Part 1: Single talker and single listeners") as implemented for SafeSeaNet, i.e. AIS messages are wrapped in the VDM or VDO sentences, resembling the output on the Presentation Interface of a Base Station.

When supplying data to the HELCOM server the Participating Parties will supply a TCP/IP socket connection (i.e. like a telnet connection, defined by a fixed IP address and port number) where data from their AIS system is available.

Similarly, a TCP/IP socket connection will be available for retrieving data from the HELCOM Server.

Reporting of failures

The Regional server of HELCOM will notify automatically when the national has system stopped transmitting AIS data.

Participating Parties should contact the Norwegian Coastal Administration (NCA) to give an estimation of the expected down time. When the problem is solved, Participating Parties should provide NCA with information regarding:

- description of the problem; and
- period of the down time.

Contact to NCA

Email to AIS support: ais@kystverket.no

Phone to AIS support (24/7): +47 95 05 60 52

Appendix 3

FORM FOR A STANDARD AGREEMENT ON ACCESS TO AND USE OF HELCOM AIS DATA

Preamble

This agreement by and between HELCOM and the Recipient establishes the terms of access to specified HELCOM AIS data.

§ 1 Parties

The Parties to this agreement are:

The Baltic Marine Environment Protection Commission
Helsinki Commission
Katajanokanlaituri 6B
FI-00160 Helsinki

hereinafter referred to as HELCOM

and

[Insert name and contact details of Applicant]

hereinafter referred to as the Recipient.

§ 2 Background

The HELCOM AIS system consists of a network of national AIS base stations situated in the Baltic Sea Countries and Norway.

The HELCOM AIS network is technically enabled by a server hosted by the Norwegian Coastal Administration.

The historic data resulting from the data exchange within this network is stored in databases of the Helsinki Convention Contracting Parties, the Norwegian Coastal Administration as well as at the HELCOM Secretariat.

§ 3 Warranties by the Recipient

The Recipient assures that its use of the HELCOM AIS data is restricted to [insert a purpose of use as applied for] as stated in the application and cf. § 4.

The Recipient further assures that the granted HELCOM AIS data and any derivatives thereof will not at any time be redistributed to or used by any third party except [insert end users of the results and/or derivatives thereof]. The recipient also assures not to share any information with which it is possible to identify individual ships and/or personal data.

The Recipient agrees to share the results of its use of HELCOM AIS data with the relevant HELCOM fora upon request.

Moreover, the Recipient assures that it by all necessary means will protect the granted HELCOM AIS data from any unauthorized use by the Recipient or a third party.

The Recipient agrees to delete the HELCOM AIS data which is the subject of this agreement by termination of the agreement which will take place latest by the closure of the [insert a purpose of use as applied for]

§ 4 Granted HELCOM AIS data

As the intended use of the HELCOM AIS data, as specified by the Recipient, has been identified to fall within the uses specified within Appendix 1 of HELCOM Recommendation 33-1, the access to the following HELCOM AIS data is granted:

1) Period covered: [insert date from and to]

2) Data: [mark as appropriate]

- a) Historical data
- b) Statistical data
- c) On-line access to real-time data:
 - Web access
 - Streamed data access
 - Geographical scope:
 - Update rate:

3) Other restrictions to be applied: [describe as appropriate]

§ 5 Duties of HELCOM

HELCOM ensures that the appropriate HELCOM AIS data is available from the HELCOM AIS Information Centre accessible via a secured connection on the Internet and provides information on the log-on procedure including username and password.

§ 6 Disclaimer on behalf of HELCOM or any other third party

HELCOM or any other third party that provided AIS data makes no explicit or implied warranty as to any matter whatsoever, including the availability, accuracy, or reliability of any information or data, whether tangible or intangible, made, developed or supplied under this agreement, or the ownership, merchantability, or fitness for a particular purpose of the information, of the data made, developed or supplied.

HELCOM does not accept any liability in the event of unavailability of HELCOM AIS data from the system.

§ 7 Entry into force and right of termination of the agreement

This agreement enters into force at the final day of signing, cf. § 9.

The agreement is valid for an unlimited period of time, however, the access to HELCOM AIS data is restricted according to the abovementioned limit in time, cf. § 4.

Both Parties are entitled the right of terminating this agreement upon a formal written notice on the intention of terminating this agreement in total or in part. Such notification on termination must be received by the other party in order to take effect.

The term of notice is six (6) months.

Any failure to fulfil any of the conditions in this agreement entitles HELCOM the right to terminate the Recipient's access to HELCOM AIS data without further notice.

The restrictions for the use and distribution of the HELCOM AIS data are valid after termination of the agreement.

§ 8 Applicable law and dispute settlement

Without prejudice to both the applicable European law, the Grant Agreement and the Russian Federation law, this Agreement shall be governed by Finnish law, being the law of the country of the coordinator.

The coordinator and the partner sign this Agreement with the intention to cooperate amicably. Should a dispute arise between the coordinator and the partner or between the partners, the parties concerned are obliged to work towards a mutually acceptable settlement.

If efforts to achieve an amicable settlement should fail, the parties concerned are obliged to seek an arbitration procedure. Any dispute, controversy or claim arising out of or relating to this Agreement, or the breach, termination or validity thereof shall be submitted to the District Court of Helsinki, at the seat of the coordinator.

§ 9 Signature

This agreement has been drawn up in two (2) identical originals; one for each party.

On behalf of HELCOM

[Insert place and date]

[Insert name and title]

On behalf of the Recipient

[Insert place and date]

[Insert name and title]

Appendix 4

Methodology to create statistics and density maps from HELCOM AIS data

This methodology was developed by the HELCOM Secretariat to generate maps and statistics for the HELCOM Maritime Assessment 2018 (BSEP No.152). The following text is the Annex 1 of this same publication. The Methodology was welcomed by the AIS EWG meeting 27-2016. All scripts are available in GitHub (<https://github.com/helcomsecretariat>) and the detailed methodology is available in Appendix 4. The shipping density maps are available on the HELCOM Map and Data service (for example: <http://maps.helcom.fi/website/mapservice/?datasetID=95c5098e-3a38-48ee-ab16-b80a99f50fef>). The maps are available per year as well as filtered per ship type.

A METHOD DIVIDED IN THREE STEPS

The method we used to produce statistics and density maps from AIS data is divided in three parts:

1. First, data should be processed from raw data to a human-readable format and to be harmonized: the chapter AIS Data Preparation.
2. Once the data is prepared in an appropriate format, we must process it to produce statistics based on events: chapter Methods to produce statistics
3. The result dataset to produce statistics is the base to make the shipping density maps: chapter Methods to produce density maps.

1. AIS DATA PREPARATION

All AIS signals received by each country in the Baltic Sea since 2005 has been stored in a centralized HELCOM AIS database, until recently hosted by Denmark (DMA) but currently in the process of being migrated to Norway (Norwegian Coastal Administration). Taken as a whole this regional AIS dataset is so big (Figure 1) and complicated to handle that several preparatory steps are needed to make the data usable for creating maps or statistics. The whole process to prepare the AIS data is summarized figure 2.

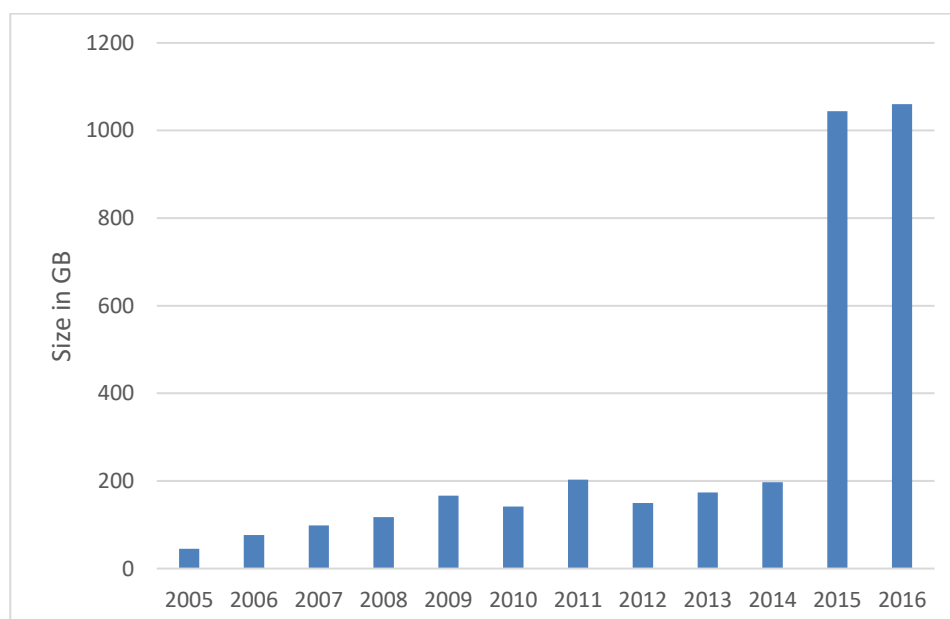


Figure 1: Size of HELCOM AIS data files according to year

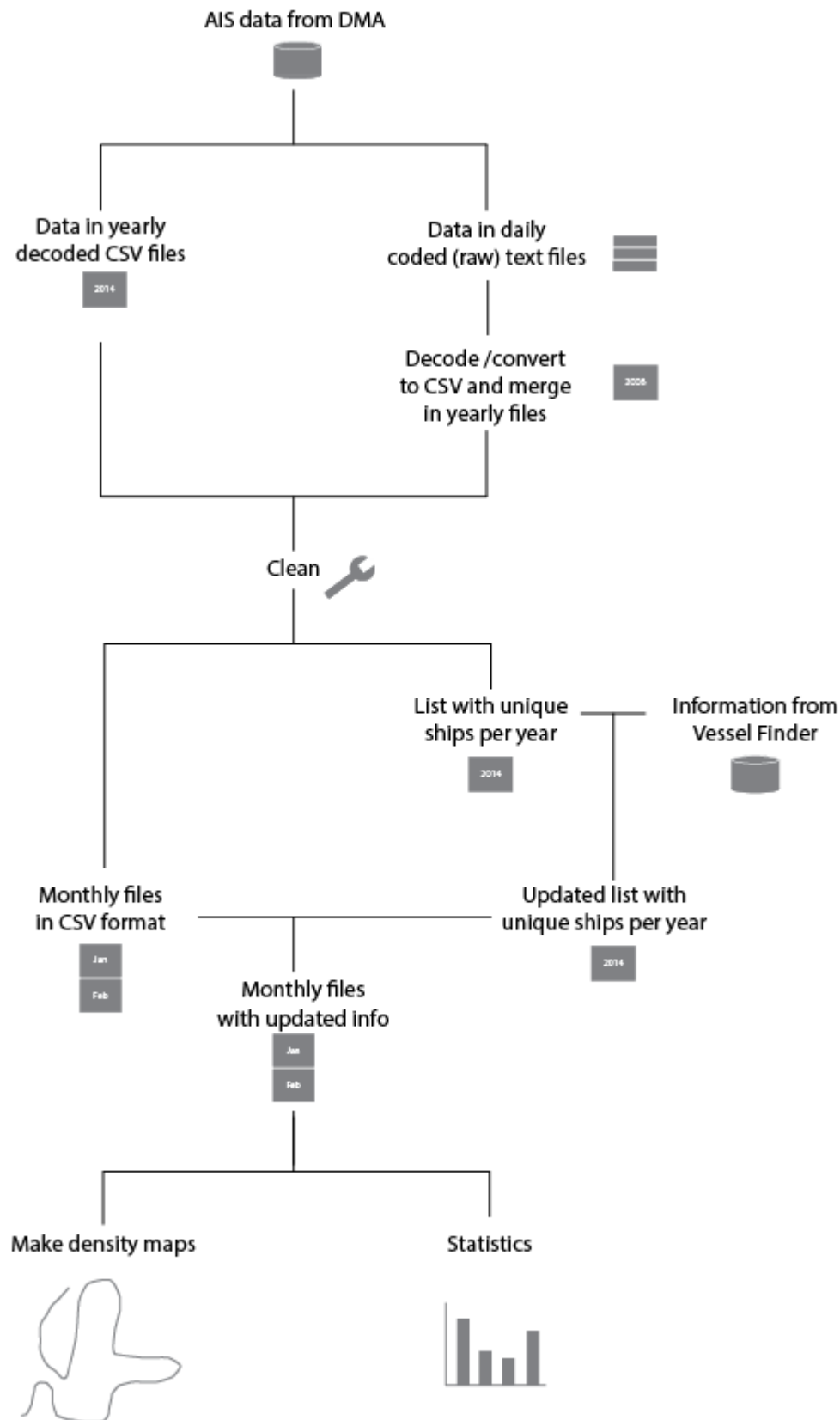


Figure 2: Overview flowchart of processing AIS data

BEFORE STARTING: HARDWARE SET UP

Following interviews with AIS data users we decided to work with the data on a dedicated server available by remote access with the following specifications: Intel Xeon E5-2630 0 @ 2,30GHz 10 cores with 48 GB RAM. This server allows several persons working at the same time.

The process consists of four parts:

1. First, we convert the files that are in raw data into CSV files.
2. Second, we clean the data from erroneous signals and duplicates
3. Third, we sort the selected AIS data per month
4. Finally, we add or update the ship information

CONVERTING RAW DATA

The AIS data was delivered by the 2005-2016 HELCOM AIS data host Danish Maritime Authority (DMA) in both decoded and raw data (NMEA sentences). Decoded data is human readable tables but raw data is not and should be converted before working with it.

Depending on the year the HELCOM Secretariat received from the data host both decoded and raw contain AIS messages (position reports from ships, base stations reports, etc.) received from all the national AIS base stations that are part of the HELCOM AIS network.

Table 1 Year and format of data received

Year	Format of AIS data received from DMA
2005	Yearly file (.CSV), decoded
2006	
2007	Daily AIS raw strings (.txt)
2008	
2009	Yearly file (.CSV), decoded
2010	
2011	
2012	
2013	
2014	Daily AIS raw strings (.txt)
2015	
2016	

The decoded AIS data in CSV files are human-readable tables containing several parameters (columns) such as the date and time when the signal was issued, the identification of the AIS message, the identification number of the AIS transmitter, etc.

The data for 2007, 2008, 2015 and 2016 were received in raw data divided into daily files. This data was in the globally standardized NMEA sentence format which is a set of data strings preceded by an encapsulated tag. These tags, in our case beginning with the characters “\$PGHP” (Gatehouse proprietary format), contain the information related to the date and time when the signal was issued. The NMEA sentences contain the rest of information: the identification of the message issued, the identification of the AIS antenna, etc.

In order to be harmonized with the rest of the data the material in NMEA sentences had to be decoded to convert them to human-readable CSV files (see Figure 3 below). These monthly files were then decoded with a decoder called AIS2CSV, a free software available online developed in 2015 by DMA. The application decodes each NMEA sentences with its encapsulated tag and generates CSV files.

In order to make the process faster the daily raw data files were merged into monthly files. The output was a monthly decoded files in CSV that were merged into a yearly file.

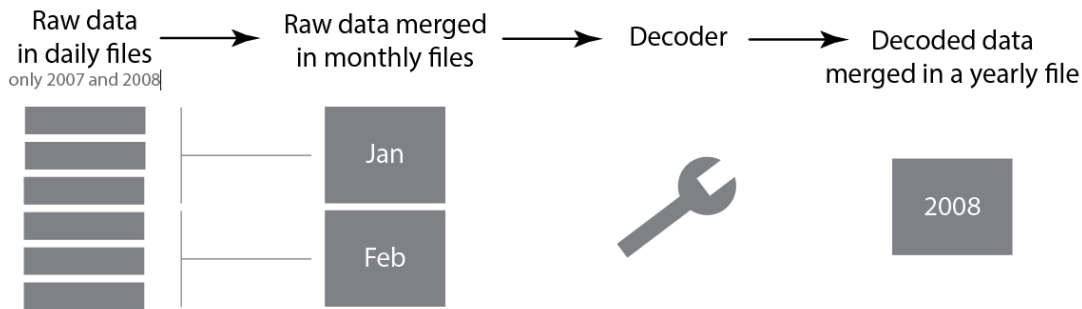


Figure 3: Raw data is merged in monthly files and decoded

DATA CLEANING

Once we had all years in yearly CSV file, the next step was to clean the data to produce monthly CSV files which will be used to create maps and statistics. The data cleaning is necessary to remove erroneous signals and duplicates (fig 4)

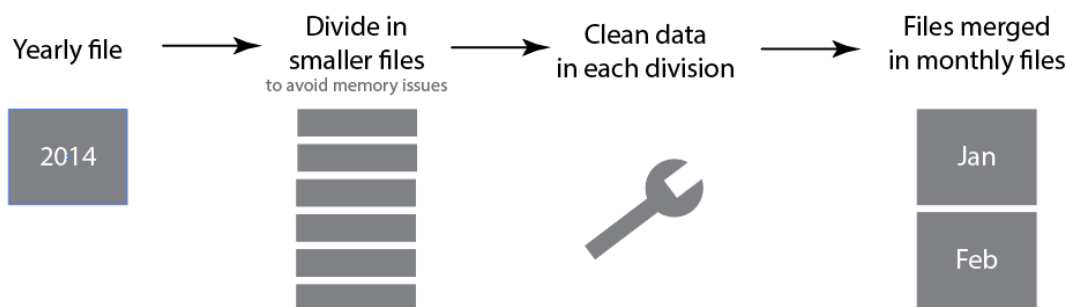


Figure 4: Time to clean the data

The inputs of the preprocessing steps are yearly CSV files of AIS data containing all messages. The outputs are monthly CSV files that contain the positions of the ships in the Baltic Sea (position reports).

The data cleaning was done using R language with the RStudio interface, the same script was applied to the yearly files one by one.

The yearly file was divided in smaller files of 1 000 000 rows to avoid running out of memory. For each division, a process is going through each AIS signals to select the relevant data and to remove erroneous signals:

- Removal of the signals that are not from the selected year.
- Removal of the duplicated signals.
- Selection of AIS messages relevant for assessing shipping activities (1, 2, 3, 18 and 19).
- Removal of wrong MMSI signals. A list that can be updated (i.e. less or more than 9 digits or equal to 000000000, 111111111, 222222222, 333333333, 444444444, 555555555, 666666666, 777777777, 888888888, 999999999, 123456789, 0, 12345, 1193046).
- Correction of wrong IMO numbers: each signal with an erroneous IMO number (not seven digits) is replaced with "NA".

- Add the Maritime Identification Digits (MID) and the flag of the ships for each signal. The MID is the three initial digits of the MMSI. This action is also removing MMSI numbers that do not have a MID (erroneous MMSI).
- Removal of special characters in all the division.
- Addition of two columns: one for the week number and one for the month.
- Selection of the signals within the planning area. A polygon was drawn manually around the planning area and only the signals within this polygon were kept.
- Removal of the signals with erroneous SOG (Speed Over Ground): negative values or more than 80 knots.
- Removal of the signals with erroneous COG (Course Over Ground): negative values or more than 360°.
- Selection of parameters to generate data products.

All key parameters were kept for all of the signals. This introduced a lot of redundancy but, because of the processing time to create the final files, we decided to avoid deleting information.

Table 2 Parameters in the preprocessed AIS data

Parameters	Description
timestamp_pretty	time in format dd/mm/yyyy hh:mm:ss
timestamp	Unix time stamp (seconds since 01/01/1970 00:00:00)
msgid	The AIS message the signal was issued
targetType	AIS type A or B
mmsi	MMSI number of the ship
lat	Latitude in decimal format
	Longitude in decimal format
posacc	Position accuracy
SOG	Speed Over Ground in 0.1 knot
COG	Course Over Ground in 0.1°
shipType	Ship type of the vessel
dimBow	The dimension between the AIS transmitter and the bow of the ship in meters
draught	Draught of the ship in 0.1 meter
dimPort	The dimension between the AIS transmitter and the port side (left) of the boat in meters
dimStarboard	The dimension between the AIS transmitter and the starboard side (right) of the boat in meters
dimStern	The dimension between the AIS transmitter and the stern of the ship in meters
month	Month the signal was issued (between 1 and 12)
week	Week number the signal was issued
imo	IMO number of the ship
country	Flag of the ship

Each division was saved as a CSV file. For each division, we created a file with the amount of signals kept after removing the erroneous signals—duplicated signals, wrong MMSI, etc. We call this file a report.

SORTING THE SELECTED AIS DATA BY MONTH

This final step of data handling goes through each division and creates a new CSV file for all the signals from a given month. The column “month” is used to sort the data into the final files. In total file sizes for a complete year ranged from about 15 GB to almost 80 GB (Fig 5)

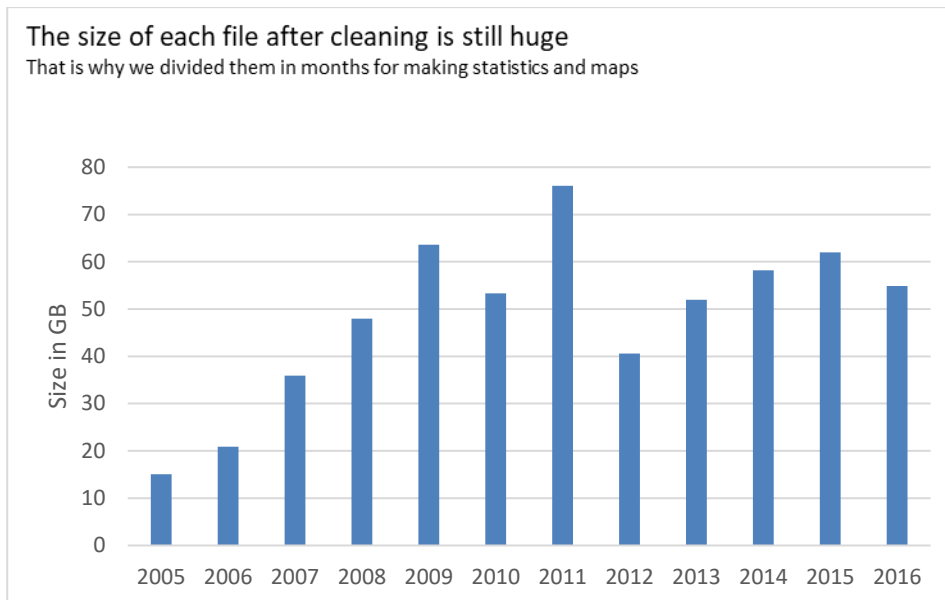


Figure 5: The size of the files are huge

UPDATING THE SHIP INFORMATION

Finally, we need to update some information from each ship like dimensions or ship type. AIS signals include this data in the AIS Message 5. However, it is not reliable because it is not obligatory.

We purchased a ship database from a data provider called Vessel Finder with up-to-date information of each ship. We did it as follows:

- First, we made a ship list for each year
- We removed the duplicate ships from the resulting list
- Finally, we created a categorization of ship types

A SHIP LIST FOR EACH YEAR

A ship list based on AIS data static information was generated for each year. The lists include all ships (IMO and non-IMO registered ships) and contain the following parameters:

- MMSI (Maritime Mobile Service Identity), a standardised series of nine digits which uniquely identify ships or other transmitting stations
- IMO number
- Name of the ship
- Callsign
- Country
- Target type
- Ship type
- DimBow
- Draught
- DimPort

- DimStarboard
- DimStern

REMOVING DUPLICATES FROM THE SHIP LIST

The yearly lists of ships were merged to have a unique list of all ships that operated in the Baltic Sea during the studied period. The ship-related information was purchased from Vessel Finder. The IMO numbers were used to identify the ships and the following parameters were provided:

- IMO number
- Name
- Ship type
- Gross Tonnage
- Net Tonnage
- Length
- Width
- Draught

Each yearly ship lists were edited using the new information. When the information was available, the ship information from the AIS data (i.e. ship type) was replaced by the new information from the provider (only for IMO registered ships). When the information was not available from the provider, the original data (from AIS) was kept. At the end of this step, a total of 120 ship types were available in the ship list.

MAKING THE SHIP TYPES CATEGORIES

Finally, two levels of ship type categories were created to use with full potential the 120 available ship types. The first level, the gross ship type, gives broad information about the ship. The second level, the detailed ship type, gives more precise information, for example about the type of cargo or tanker. The table below describes the gross and detailed:

Table 3 Ship type categorisation used to update ship information

Gross ship categorisation	ship type	Detail ship categorisation	ship type
Cargo		Bulk cargo	
		General cargo	
		Other cargo	
Tanker		Chemical tanker	
		Oil product tanker	
		Gas tanker	
		Crude oil tanker	
		Other tanker	

Container	Container
Passenger	Cruise
	Ferry
	ROPAX
Other	Dredger
	Other
	Tug
	Yacht
Fishing	Fishing
Service	Service
Vehicle carrier / RORO	Vehicle carrier
Cargo	RORO Cargo

These ship types were chosen following the current work about emissions from shipping in the Baltic Sea done by the Finnish Meteorological Institute (cf. Information document 4-4 Emissions from Baltic Sea shipping in 2014 submitted by Finland for the HELCOM MARITIME 15-2015 meeting).

2. METHODS TO PRODUCE STATISTICS

After the AIS data is cleaned and in CSV format, we want to prepare the files to create statistics and density maps for IMO registered ships.

To do this, we need to identify events occurring in four different areas that we created:

Event	Description	Area to identify the event
Exit	Ships exiting the Baltic Sea	Exiting area
Enter	Ships entering the Baltic Sea	Entering area
Trip	Ships going from port to port	between ports
Stop	Ships stopping in a port	in ports

Statistics can be made based on these events (number of visits per ports, distance sailed, ships entering or entering the Baltic Sea, etc.). Density maps can be produced using the information about the trips at sea: each trip will be assigned to a unique identification number that will be used to generate density maps.

The method outputs two files:

- A CSV file describing the ships movements (events). This file is used to create statistics on how many visits per port, how many ships entering or exiting the Baltic, etc. There is one CSV file per year.
- A CSV file with a trip ID numbers. Each AIS signal is assigned a trip ID which identifies where the ship is going from port to port, from outside the Baltic Sea, etc. This file is used to produce density maps. There is one CSV file per month.

THE EXIT AND ENTER EVENTS

This first section will explain how to find ships exiting and entering the Baltic Sea. The output of this analysis are the exit and enter events table. It is a list describing which ships are leaving or entering the Baltic Sea, at what time and where.

DEFINITION OF INSIDE, OUTSIDE AND EXIT AREAS

The very first step is to find the AIS signals from ships that could be leaving or entering the Baltic Sea. The signals are by definition the position reports from ships. The latitude and longitude are the two parameters used for this spatial analysis.

To start the process we defined five “exit areas”, four “outside areas” and one “inside area”. *Figure 6: Location of outside and exit areas, example of the Kattegat Sea* shows an example in the Kattegat area:



Figure 6: Location of outside and exit areas, example of the Kattegat Sea

The 5 “exit areas” are the borders between outside and inside the Baltic Sea where the ships have to go through to leave the Baltic Sea either towards the North Sea or to major lakes in the Baltic Sea Region (Lakes Vänern, Ladoga and Saimaa). These areas are between 2,4 and 16,6 km wide.

Name of the exit area	Location
Skagen	Between the port of Skagen (Denmark) and the village of Kärna located 15 km north of Gothenburg (Sweden).
Goteborg	On the Göta Canal next to the city of Kungälv (Sweden). This location marks the entrance and exit to the Vänern lake in central Sweden.

Kiel Canal	On the Kiel Canal on the Western side of the port of Kiel (Germany). This location marks a canal between the Baltic Sea and the North Sea.
Neva River	Along the Neva river, after the city of Saint-Petersburg (Russia). This location is used to assess the ships going to or coming from the lake Ladoga in Russia.
Lappenrenta	On the Saimaa Canal, between the Gulf of Finland the Saimaa lake.

We created a new column called “exit areas” in the table (position reports from ships). For each signal that is in one of the 5 polygons, the name of the exit area was added to this new column.

The “outside areas” are 4 polygons located next to each of the “exit areas”.

Name of the outside area	Location
Skagen / Vänerm	The area of the North Sea (outside the Kattegat Sea towards the North Sea) and the Vänerm lake.
Kiel	The Kiel Canal.
Saimaa	The Saimaa lake
Neva River	The Neva River and the Ladoga lake

We added a new column to the table to identify the signals that are inside these “outside areas”. The signals that are not in the “exit area” or in the “outside area” are assigned to be in the “inside area”. They come from ships in ports or traveling in the Baltic Sea.

IDENTIFICATION OF THE SEQUENCES OF SIGNALS IN THE EXIT AREAS OF THE BALTIC SEA

After we have defined the areas, we must now find the location of each AIS signals: in an exit area, outside or in the Baltic Sea. We need to create sequences of signals entering or exiting the Baltic, then we need to find the direction of the ship.

A sequence is a group of signals from a ship describing a movement in the exit areas. It is identified by sorting the data by IMO number and by time. Each one has a unique identification number: the signals are gathered into groups of signals that match the same sequence.

In the *Figure 7*, the points X_6 , X_7 and X_8 get an identification number for the same sequence. The points from X_{218} to X_{220} get another unique identification number for this other sequence. But the points from X_{105} to X_{108} do not have a sequence identification number. The ship is not leaving the Baltic Sea since there is no point in the outside area. This process is applied for all of the signals found in the exit areas.

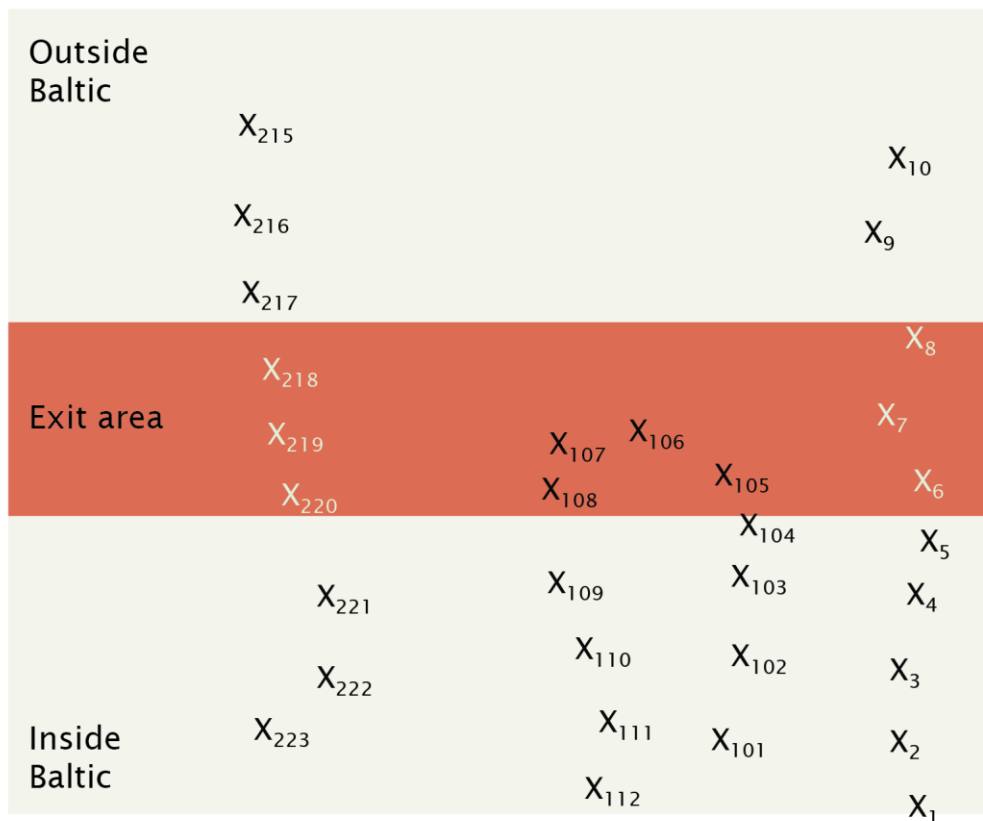
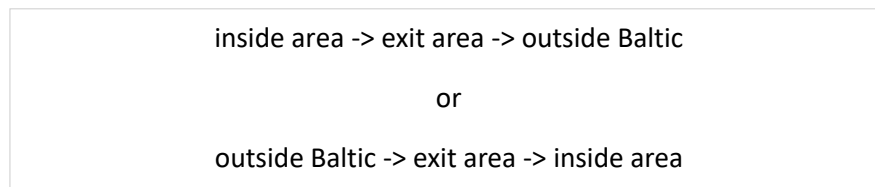


Figure 7: generating the sequences in an exit area

Therefore, we give the sequence identification number to the ships following two kind of paths:



For the ships that are returning to the same area (outside or inside) than they came from, there is no sequence identification number added to the signals.

Now that we have identified the ships that are in the exit area we need to know which ones are actually exiting or entering. We do it thanks to the COG parameter.

DEFINITION OF THE SEQUENCES ENTERING OR EXITING THE BALTIC SEA

Each of the signals of AIS data have a parameter called “COG” (Course Over Ground) which represents the direction of the vessel from 0° to 360°. Following the COG of the signals within the 5 “exit areas”, we could identify if the ships were entering or leaving the Baltic Sea.

For each signal in the exit area that has a sequence identification number, a new value with the direction of the boat was assigned: 0 for traveling north, 90 for traveling east, 180 for traveling south and 270 for traveling west. For example, if the boat is traveling between 90° and 270° in an exit area, the value 180 is added. This new column in the dataset is called COG₂.

The next step was to identify all the sequences leaving or entering the Baltic Sea through the exit areas. A function available with the software R (under the package *dplyr*) was used to summarize the COG₂ for each of the unique sequences and therefore the direction of the ships within the exiting area. The average of the COG₂ (\bar{x}_{COG_2}) was computed for each of the unique sequences. We made these calculations of COG₂ and \bar{x}_{COG_2} to avoid some unexpected ship movements in the exit areas.

Finally, we selected the signals from ships going outside or inside the Baltic Sea for each exits area following the name of the location of these unique sequences:

Exit area	Value of \bar{x}_{COG_2} for exiting the Baltic Sea	Value of \bar{x}_{COG_2} for entering the Baltic Sea
Skagen	$\bar{x}_{\text{COG}_2} = 0$ (north)	$\bar{x}_{\text{COG}_2} = 180$ (south)
Goteborg	$\bar{x}_{\text{COG}_2} = 180$	$\bar{x}_{\text{COG}_2} = 0$
Kiel Canal	$\bar{x}_{\text{COG}_2} = 270$ (west)	$\bar{x}_{\text{COG}_2} = 90$ (east)
Neva River	$\bar{x}_{\text{COG}_2} = 90$	$\bar{x}_{\text{COG}_2} = 180$
Lappenrenta	$\bar{x}_{\text{COG}_2} = 0$	$\bar{x}_{\text{COG}_2} = 180$

After this step, we could know if the ships were exiting or entering the Baltic Sea for each exit area and for each sequence.

CREATION OF THE EXIT/ENTER TABLE

The next step was to create the exit and enter events table for each of the unique sequence identification number. An event is the information about the exits and enters of the Baltic Sea:

Parameters created for the exit and enter events	Definition
IMO number	IMO number of the ship
MMSI number	MMSI number of the ship
MinTime	Date and time of the first signal in the exit area
MaxTime	Date and time of the last signal in the exit area
Event	Exit or Enter
Location	the name of the exit area
Sequence number	The unique sequence of the exit or enter

After the events are identified, we could know the time of exiting or entering the Baltic Sea through the 5 exit areas for each ship. This new data was stored in a temporary file and was used during the last step of the analysis.

THE TRIP AND STOP EVENTS

After the previous step we know which ships are entering and exiting the Baltic. In the next step we are going to explain how to find the stops and the trips inside the Baltic Sea.

THE PORTS OF THE BALTIC SEA

Before starting we need to create the ports. We created 245 polygons around ports in the 9 countries of the Baltic Sea by following these steps:

- We filtered some monthly AIS data keeping only the signals where the speed of the boat was equal to 0.
- We used Open Street Map, Google Maps as background images.
- Shipping lines produced with AIS data helped to see ship movements
- The publication Baltic Ports List of 2012 helped to identify ports.

Once the ports polygons were ready we wanted to check if the ships where inside or outside those areas. We plotted the AIS signals in R software and overlaid them on top of the ports polygons and the exit areas.

PREPARATION OF THE SIGNALS DEFINING A STOP IN A PORT

To locate the ships that are stopped in a port we used the speed of the ship which is available as the parameter SOG (Speed Over Ground) for each position reports.

In theory, just filtering those ships with SOG equals zero would be enough. However, because of the accuracy of the AIS transmitter onboard the ships, it is possible that the SOG is not equal to 0 even when the ship is at berth in a port. To avoid this issue, the SOG of a ship at port with less than 0.5 knot was replaced by 0 knot.

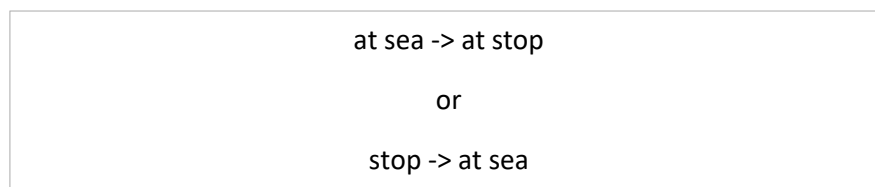
When the speed was equal or higher than 0.5 knot, we kept the original SOG—the boat could be moving during the maneuver along the berth.

PRODUCING THE EVENTS STOPS IN A PORT

The next step was to generate the stop events: a group of signals with a unique sequence identification number. Each signal that was qualified as a stop from the previous step would be assigned this number.

To find the stops as sequences we sorted the data by IMO number and by time. The table would show when a ship is at sea and coming into a port and is stopping (when the SOG value is equal to 0).

The sequence identification number is given to the ships following two kind of paths:



Finally, we created the following parameters by summarizing the information by each unique identification number:

Parameters created for the stops sequences	Definition
IMO number	IMO number of the ship
MMSI number	MMSI number of the ship
MinTime	Date and time of the first signal with SOG = 0
MaxTime	Date and time of the last signal with SOG = 0
Event	Stop
Location	the name of the port
Sequence number	The unique sequence of stop

Once we calculated the Mintime and Maxtime, it was possible to know the duration of the stops. Thereafter, we made a subset of the stops longer than 10 minutes is done which was stored as a temporary file to be used later.

PRODUCING THE EVENTS TRIPS AT SEA

After the stops at port were created we proceeded to calculate the trip of each trip at sea.

Naturally, all the signals outside the 245 port polygons and in the Baltic Sea were selected as trips.

To find the trips between two stops, between an enter and a stop and between a stop and an exit, we sorted the data the data by ship (IMO number) and by time

The *Figure 8* below shows three types of situation: the signals as trips between an exit area and a stop (situation 1), between two ports (situations 2) or between a port and an exit area (situation 3).

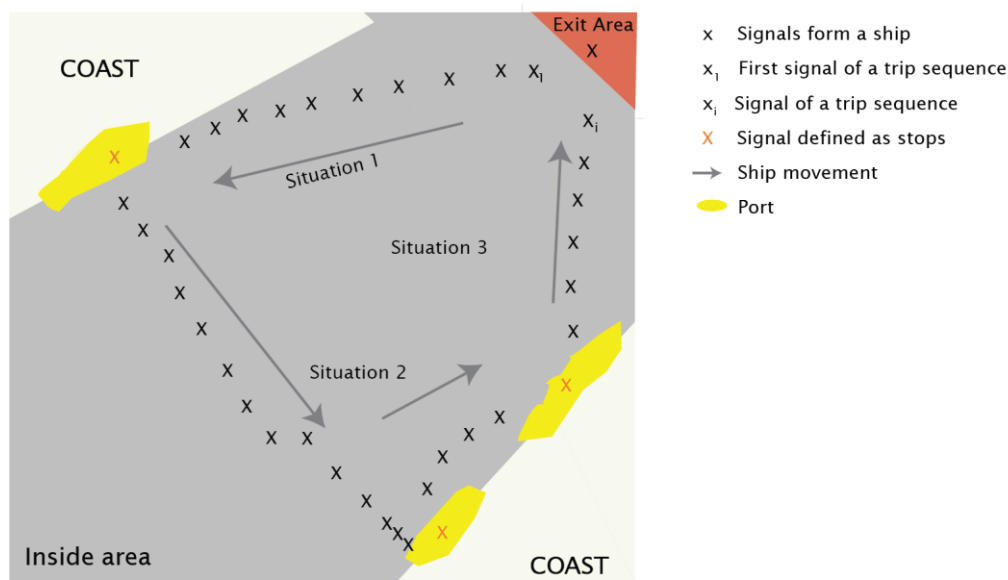


Figure 8 The 3 types of situations for the trips sequences

We assigned a unique sequence number, called the trip Id, to each movements between the locations.

Every signals of the same sequence will have assigned the same sequence number, called trip Id (*Figure 9*). This is the input needed to produce lines and shipping density maps. The same process was applied to all of the ships.

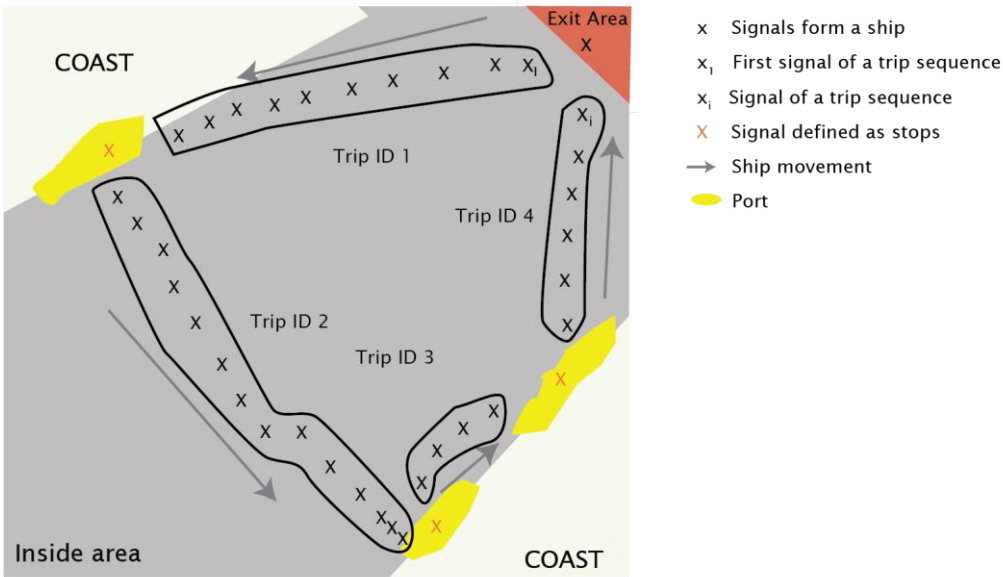


Figure 9 Assigning the trip ID value to the sequences at sea

These are all the parameters that we calculated:

Information collected for the trip sequences	Definition
IMO number	IMO number of the ship
MMSI number	MMSI number of the ship
MinTime	Date and time of the first signal of the trip
MaxTime	Date and time of the last signal of the trip
Distance sailed	Distance sailed
Event	Trip
Sequence number	The trip Id value

To calculate the duration of each trips we could use Mintime and Maxtime.

The distance sailed was calculated during the summarizing step—it is the cumulative distance between each signals of the same trip Id value. This table is saved as temporary file and will be used for the next step.

OUTPUTS OF THE WHOLE PROCESS

The output of the whole process are two CSV files.

- The first one is used to produce statistics of the events: they have to be harmonized to be relevant for further analysis.
- The second one is the file with the position reports from ships with a trip Id assigned to each AIS signals. These monthly files are used to produce shipping density maps.

Output to produce statistics

Three files per month:

- A table with the exit / enter sequences information
- A table with the stop sequences information
- A table with the trip sequences information

These three table are merged and sorted by ship (IMO number) and time.

However, the files have to harmonize: for the same ship, two stop sequences or visit sequences cannot be consecutive. If some of them are consecutive, they are merged together and the duration of the events and the distance sailed at sea is corrected.

The monthly files are finally merged as yearly files to generate statistics on a yearly basis.

Output to produce density maps

Monthly files with a trip ID assigned to each AIS position report with which we can produce density maps.

3. METHODS TO PRODUCE DENSITY MAPS

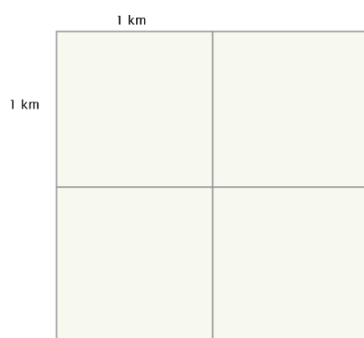
WHAT IS A SHIPPING DENSITY MAP?

A shipping density map represents the intensity of shipping traffic in certain time period.

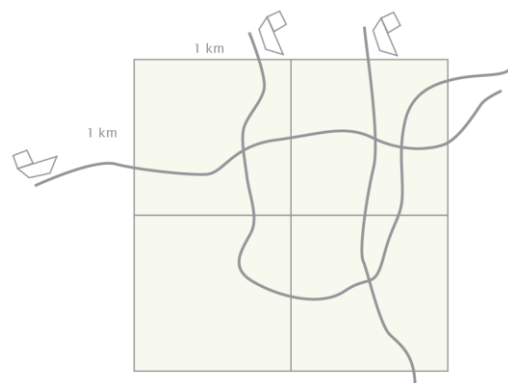
There is no standard definition or method to create shipping density maps. We have developed a method that answers a basic question: Assuming we have a grid of cells and the trips of ships from port to port, how many times do those trips cross each cell of the grid?

The process can be explained in four steps:

1 We need a grid with 1km*1km cells



2 We overlap the trips of each ship onto the grid



3 We count how many lines are crossing each cell

4 We apply the style—darker color means more density

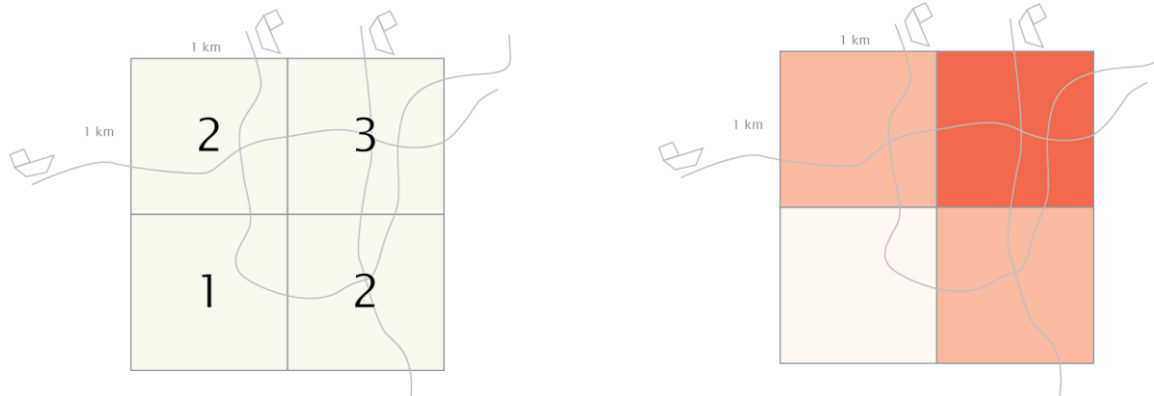


Figure 10

WHAT WE NEED BEFORE CREATING THE DENSITY MAPS

Before we begin to produce shipping density maps we need:

- A GIS software
- Monthly CSV files with all the AIS signals
- A grid to overlap the lines and create the maps
- A ship list: a csv file containing all ships per year

WHAT SOFTWARE WE NEED TO CREATE SHIPPING DENSITY MAPS:

We used ArcGIS 10.4 for Desktop advanced license with Spatial Analyst for creating raster layers. All scripts were written in python scripting language version 2.7 using ArcPy, a Python ArcGIS scripting module.

MONTHLY CSV FILES

We have to be sure we create beforehand the following folders under each year. O1_trips is created in a previous step (see 2. METHODS TO PRODUCE STATISTICS). It contains a CSV file per month:



Each monthly file contains the following columns:

date_time	timestamp	mmsi	lat	long	trip_id	visit	distance	new_trip_id
24.4.2015 14.2	1429874645	205114000	57,736862	10,843556	1	0	0	1
24.4.2015 14.2	1429874826	205114000	57,731323	10,85466	1	0	1	1
24.4.2015 14.3	1429875015	205114000	57,725517	10,866167	1	0	1	1
24.4.2015 14.3	1429875205	205114000	57,71953	10,877648	1	0	1	1
24.4.2015 14.3	1429875396	205114000	57,713375	10,889172	1	0	1	1
24.4.2015 14.3	1429875595	205114000	57,707054	10,900722	1	0	1	1
24.4.2015 14.4	1429875785	205114000	57,70108	10,911857	1	0	1	1
24.4.2015 14.4	1429875975	205114000	57,695354	10,922934	1	0	1	1

Figure 11

WHAT GRID?

We need a grid, a shapefile with square cells, in order to put the lines on top. We then count how many of them are crossing each cell.

The grid file was [downloaded from the European Environment Agency \(EEA\)](#) and it is based on recommendations made at the 1st European Workshop on Reference Grids in 2003 and later from INSPIRE geographical grid systems. This standard grid was recommended to facilitate the management and analyses of spatial information for a variety of applications.

EEA offers a grid in shapefile format for each country in three scales: 1km, 10km and 100km. We chose the 1km grid to produce high resolution maps.

After we downloaded each Baltic Sea country we joined them all in one file. The resulting merged grid had more than four millions cells. We then delete the cells on land to save space and make the file easier to manage. The result was a file with about 400,000 cells.

THREE STEPS FOR CREATING DENSITY MAPS

Once we have the monthly csv files with the AIS signals, the folders and the grid we can start the process to create a density map.

We followed three steps to create density maps:

1. Create lines
2. Divide the lines in ship types (cargo, tanker, passenger...)
3. Make the density map overlapping lines onto the grid

There is a python script for each step. They all use python's multiprocessing module. This module allows us to submit multiple processes that can run independently from each other in order to make best use of our CPU cores. We used a machine with 10 cores.

The final result of processing one year is a density map in raster format (TIFF) per month and per ship type. That makes 96 maps. It takes about 40 hours to complete the process on average per year.

CREATE LINES

NAME OF THE SCRIPT: TrackBuilderFromCSV_multiprocessing.py

PURPOSE: Creates lines representing trips of ships from port to port.

DURATION: three hours per year with multiprocessing.

PRECONDITIONS: Before creating lines we need:

- A year folder under which there are csv files containing all trips per month
- The file name of the monthly file must have the year and the month name in English
- The monthly csv file must have the following columns:
 - MMSI
 - Trip_id: Number of trips per MMSI (per ship)

HOW THE SCRIPT WORKS: This script reads each monthly csv file and converts coordinates to lines

OUTPUT: the output of this process is a folder with lines shapefiles for each month

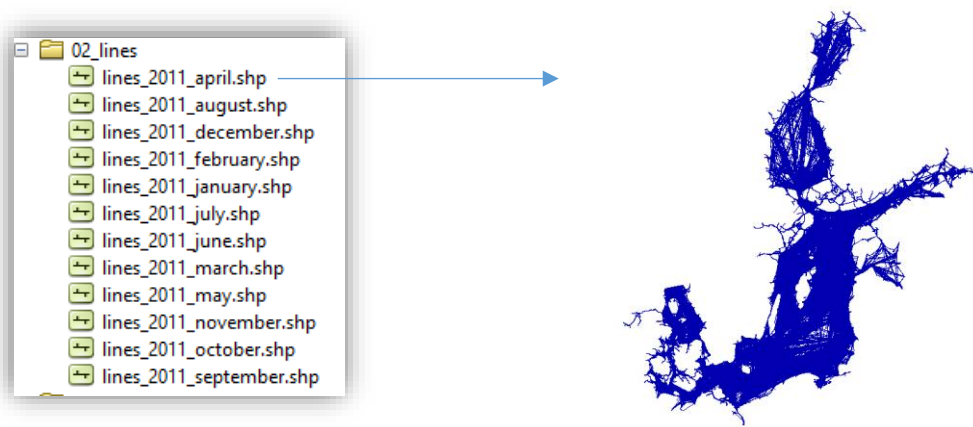


Figure 12

SPLIT LINES BY SHIP TYPE

NAME OF THE SCRIPT: SplitTracksByShipType_multiprocessing.py

PURPOSE: Once we have the lines shapefiles we proceed to divide each file into different ship types

DURATION: two hours per year with multiprocessing.

PRECONDITIONS: Before splitting lines by ship type we need:

- A folder with monthly lines shapefiles
- The file name of the monthly file must have the year and the month in English
- A ship list: a csv file with unique number of ships per year created in a previous step (see A SHIP LIST FOR EACH YEAR). This file must include at least the following fields:
 - IMO number
 - HELCOM_Gr: HELCOM Gross ship type—a wide classification of ship types
 - HELCOM_De: HELCOM Detail ship type—a detail classification

HOW THE SCRIPT WORKS: This script divides the monthly line files into the HELCOM gross classification ship types (see *Table 3 Ship type categorisation used to update ship information*)

OUTPUT: a folder per ship type under which there are the monthly lines shapefiles

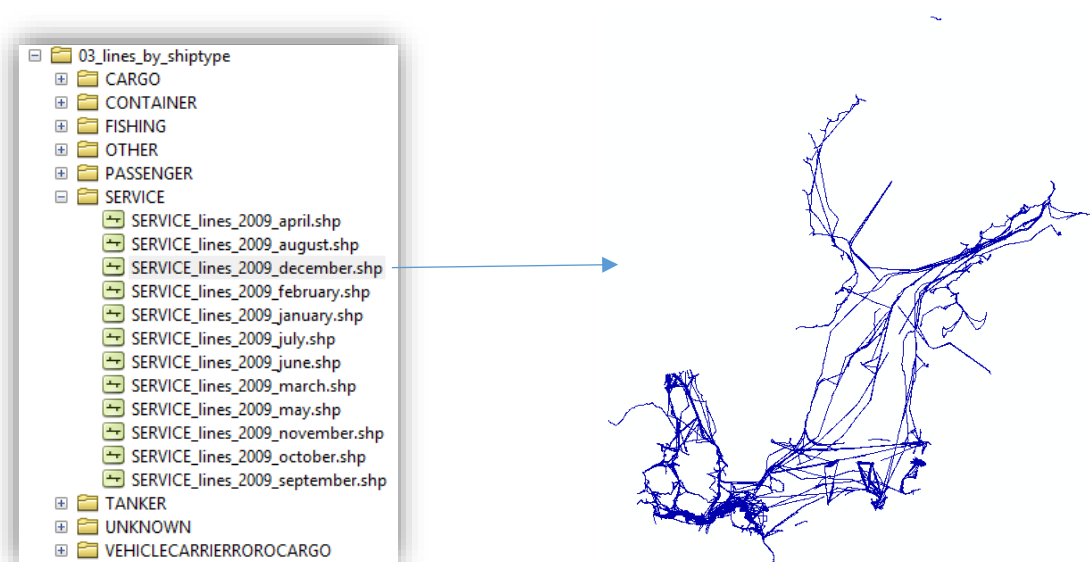


Figure 13

MAKE THE DENSITY MAP

NAME OF THE SCRIPT: CreateRastersYear_multiprocessing.py

PURPOSE: Create a density map by overlapping the lines onto a grid and counting the number of lines crossing each cell

DURATION: It depends on the ship type. We estimate an average of 5h per ship type. For cargo, the ship type with most signals, it takes about 10 hours. A ship type with less signals as service can take about 2h

PRECONDITIONS: Before creating the density map we need:

- A folder with monthly lines shapefiles divided in ship types
- The file name of the file must have the year and the month in English
- A grid shapefile (see WHAT GRID?)

HOW THE SCRIPT WORKS:

The process to create a map is divided into six steps, illustrated in the figure below. It uses python's multiprocessing module to run six months simultaneously:

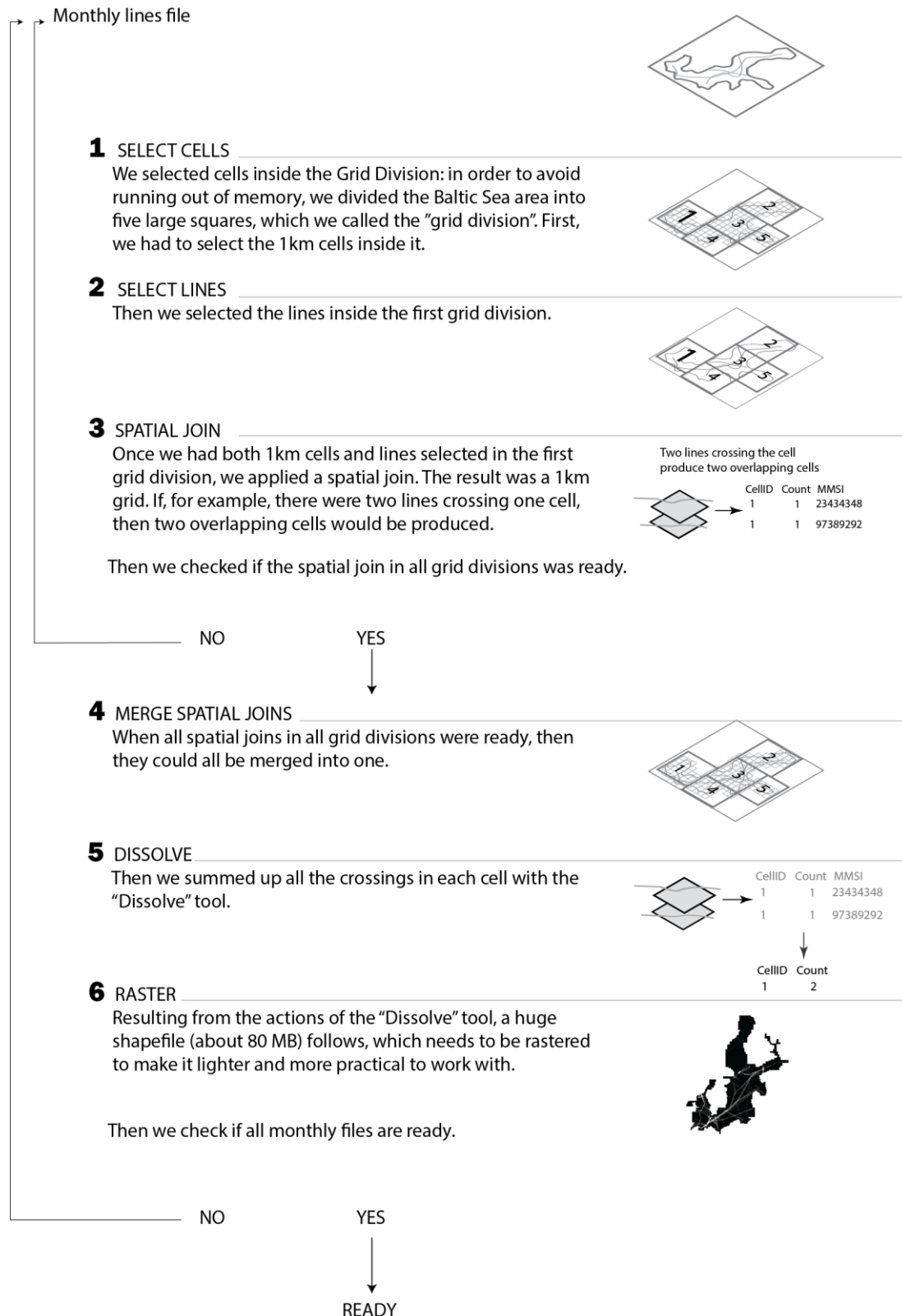


Figure 14

OUTPUT: a folder, 04_rasters, with subfolders for each ship type. Under each subfolder there is a raster file in TIFF format for each month. There is also a yearly file (CARGO_2009_Year_Raster.tif) with the sum of all monthly raster files

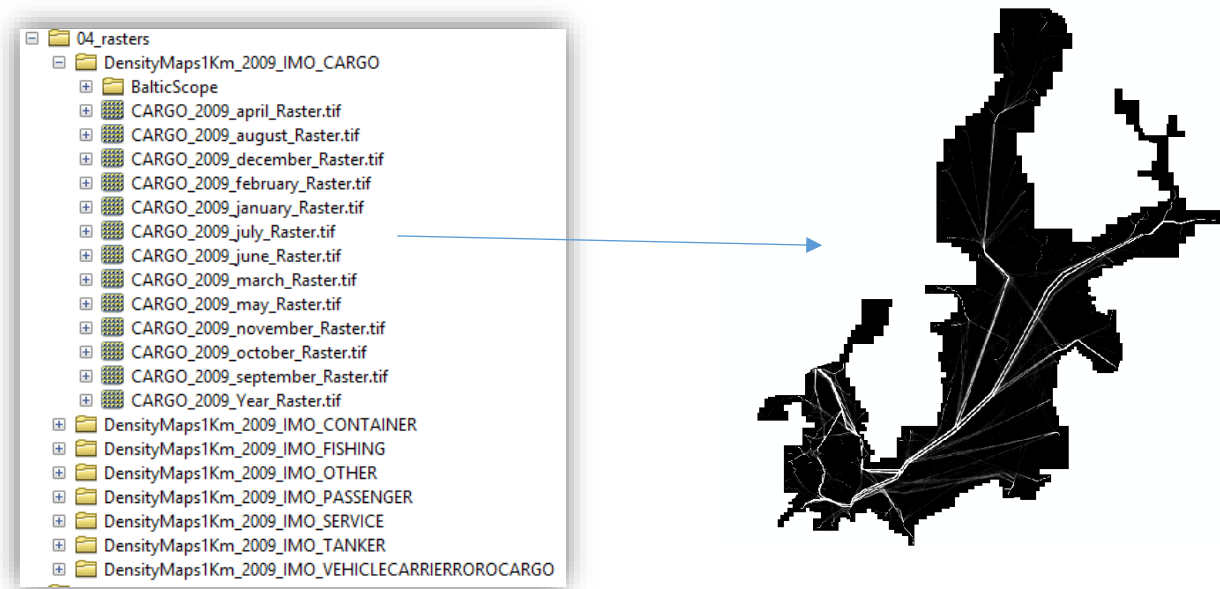


Figure 15