

*Article*

# Maritime Activities Observed Through Open-Access Positioning Data: Moving and Stationary Vessels in the Baltic Sea

Moritz Hütten 

GRID Inc., 3-6-7 Kita-Aoyama, Minato-ku, Tokyo 107-0061, Japan; moritz.huetten@gridpredict.co.jp

## Highlights

### What are the main findings?

- Vessel activity in the Baltic Sea is derived from open-access positioning data.
- Uncertainties from incomplete vessel positioning data can be precisely quantified.

### What are the implications of the main findings?

- Publicly available positioning data allow for effective analysis of vessel activity in coastal waters.

## Abstract

Understanding past and present maritime activity patterns is critical for navigation safety, environmental assessment, and commercial operations. An increasing number of services now openly provide positioning data from the Automatic Identification System (AIS) via ground-based receivers. We show that coastal vessel activity can be reconstructed from open access data with high accuracy, even with limited data quality and incomplete receiver coverage. For three months of open AIS data in the Baltic Sea from August to October 2024, we present (i) cleansing and reconstruction methods to improve the data quality, and (ii) a journey model that converts AIS message data into vessel counts, traffic estimates, and spatially resolved vessel density at a resolution of  $\sim 400$  m. Vessel counts are provided, along with their uncertainties, for both moving and stationary activity. Vessel density maps also enable the identification of port locations, and we infer the most crowded and busiest coastal areas in the Baltic Sea. We find that on average,  $\gtrsim 4000$  vessels simultaneously operate in the Baltic Sea, and more than 300 vessels enter or leave the area each day. Our results agree within 20% with previous studies relying on proprietary data.

**Keywords:** open data; positioning data; data cleansing; Automatic Identification System (AIS); maritime traffic; vessel stop detection

---

## 1. Introduction

Human activities at sea constitute the backbone of the world economy through resource extraction, energy production, fishing, and the transport of most of the global cargo. As such, maritime traffic has increased significantly in recent decades. Today, around 90% of the world trade is transported through the oceans [1]. According to the United Nations Conference on Trade and Development [2], global seaborne trade nearly doubled between 2004 and 2024 with an average annual growth of about 3%, and it is expected to grow further by about 2% yearly between 2025 and 2029. With respect to fishing, the Food and Agriculture Organization of the United Nations [3] reports that over the past two decades, fishing for marine livestock remained largely stagnant. However, aquaculture production in marine areas increased by 50%, accounting for 2024 a third of the global marine food supply. Paolo et al. [4] found that the number of offshore wind turbines is rapidly growing, having surpassed oil platforms globally in 2021.

This rise in maritime activities adds to the pressure on the climate [5–7] and marine ecosystems [8–10], and increases safety [11,12] and security [13] risks along the dense shipping routes. Several

mechanisms are in place to ensure traffic safety and to monitor compliance with local and international regulations in coastal waters and the open seas. Among these are the Vessel Monitoring System, used in fishing, and the Long-Range Identification and Tracking system, LRIT. Unlike these mechanisms, the Automatic Identification System (AIS) allows public tracking of most operating vessels at high temporal resolution, including information such as vessel identity, port of call, and current draft [14]. The International Convention for the Safety of Life at Sea (SOLAS) requires ships of 300 gross tonnage (GT) and above on international voyages (or cargo ships of more than 500 GT on domestic voyages) and all passenger ships to regularly report voyage status information via so-called Class A type AIS transceivers (AIS-A) in intervals of a few seconds to several minutes [15]. Also, other vessels can voluntarily use the AIS via the weaker and less prioritized Class B system (AIS-B). The transmitted data can then be picked up by any suitable radio receiver on land, other vessels (which can relay the signal to land), or by satellites. Although AIS was intended primarily for the immediate safekeeping of maritime traffic, collected AIS data have been used in numerous scientific studies of activity patterns at sea [16–20].

Governmental and private actors with access to space-based systems and dense terrestrial receiver networks can track vessel activity through the AIS with comprehensive coverage, but access to these data is mostly restricted. In contrast, community-based networks and some government authorities increasingly provide open access to their real-time or historical data from terrestrial AIS receivers [21]. Using open access data has recently caught the attention of the research community [22,23]. However, data from open access services are subject to their own distinct limitations. The range of picking up AIS signals from land is usually limited to several tens of nautical miles, restricting terrestrial receiver networks to only track vessels in coastal waters [24]. Moreover, AIS networks may lack receiver stations in certain coastal areas or may be regionally limited. Lastly, community-based networks usually do not apply strict control over the quality or origin of their data, making these networks vulnerable to the injection of accidentally or intentionally incorrect data [25–27].

This work demonstrates that vessel activities can nevertheless be comprehensively studied with data from open-access AIS networks, with accuracy competitive to previous studies based on proprietary data. Accordingly, describing a suitable methodology in this paper will facilitate monitoring, controlling, or studying maritime activities for any public or private stakeholder. In doing so, we present two novel aspects in the analysis of AIS positioning data: first, we provide a rigorous estimate of uncertainties arising from incomplete data. This is of particular importance when data is regionally limited and of inhomogeneous coverage. Second, we provide vessel counts and number densities for moving vessels and stationary activities, such as fishing or berthing in ports, throughout the work in consistent units. This allows us a direct comparison of both contributions to the overall maritime activity, and to infer port locations complementary to previous methods based on clustering algorithms.

We demonstrate the feasibility of providing competitive insights from open-access AIS data using three months of vessel activity in the Baltic Sea in 2024, an area with diverse shipping activity that is mostly, but not exclusively, composed of coastal waters. Moreover, numerous studies based on AIS data have been published for this region [28–35], facilitating the comparison with previous works. We emphasize that we do not address intentional AIS data manipulation, and caution is required when using community data for safety or security purposes. This particularly applies to the Baltic Sea, where cases of illegitimate deactivation of AIS transceivers and AIS spoofing have been repeatedly reported [36].

The remainder of this paper is organized as follows. Section 2 presents our methods. Section 2.1 introduces the study area and describes the AIS data acquisition; Section 2.2 outlines the cleansing applied to the AIS message data; and Section 2.3 outlines our model for converting the AIS data into a spatio-temporal representation of vessel journeys. Section 2.4 describes the derivation of vessel metrics, activity maps, and port locations. In Section 2.5, we quantify the uncertainties in these metrics. Section 3 presents the results of the Baltic Sea analysis. We discuss our findings and conclude in Section 4.

## 2. Methods

### 2.1. Analysis Region and Data Acquisition

The Baltic Sea is one of the busiest maritime areas in the world. Between 8% [37] and 15% [38] of the worldwide cargo is estimated to pass through the region. Cargo shipping through the Baltic Sea increased by about 2% yearly in the decade before 2014 [39] and between 2015 and 2019 [40]. Overall maritime transport volume is expected to continue increasing until 2030 [41,42], driven by both commercial and cruise shipping [32]. Surrounded by nine littoral states, excluding Norway, the Baltic Sea also forms a maritime zone with large geopolitical significance.

Different geological and political definitions exist for the extent of the Baltic Sea. According to the Baltic Marine Environment Commission (Helsinki Commission, HELCOM), the Baltic Sea is the basin between Scandinavia and the European Plain, limited in the west by the latitude of  $\varphi = 57.741^\circ \text{ N}$  between Skagen and Gothenburg at the transition from the Kattegat to the Skagerrak [43]. For this analysis, we choose a larger area bounded in the west by the longitude  $\lambda = 9^\circ \text{ E}$  through the Skagerrak and Limfjord in Denmark, and separated from inland waters by nine additional transit areas, as defined in Table A1 and shown in Figure 1. This region of interest (ROI) covers an area of 440,492 km<sup>2</sup> and two time zones, UTC+1 and UTC+2, which we divide at  $\lambda = 19.5^\circ \text{ E}$  (Figure 1).

For the defined ROI, we subscribed to the open AIS data stream from aisstream.io [44] for 91 days between Monday, 29 July 2024, 00:00:00 UTC and Sunday, 27 October 2024, 23:59:59 UTC, receiving position-report messages from moving vessels (navigational status zero) and ship static data from all AIS-A vessels, collecting a total of 91,111,731 messages from 14,620 different Maritime Mobile Service Identity (MMSI) numbers. We discarded all messages corresponding to coordinates on land. We also do not include Class B vessels in our analysis because of their large seasonal variability, dominantly associated with leisure vessels, and the poorer data quality. However, we separately collected messages from Class B vessels to estimate their average contribution to the overall shipping activity in the uncertainty discussion of Section 2.5.5. We did not collect position reports of different navigational statuses and use the static reports to trace a vessel during a stationary period.

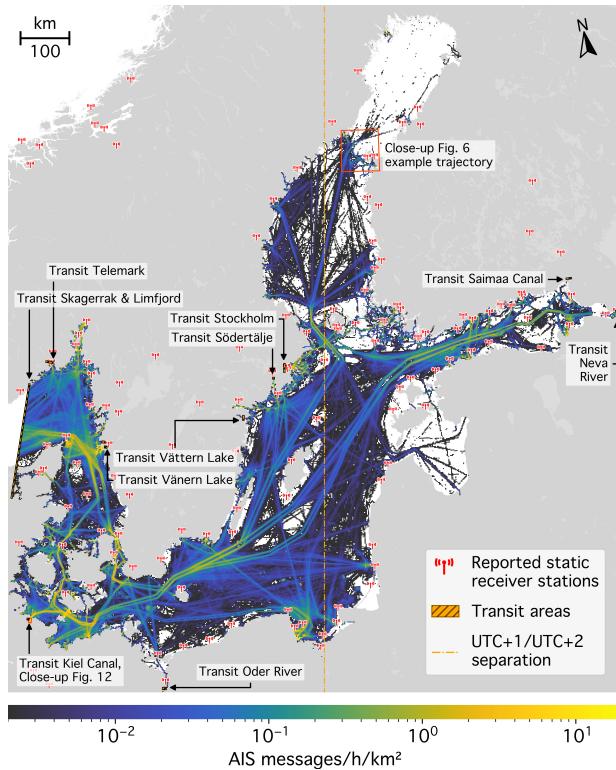
We saved the messages with the reported timing information (one-second precision); latitude and longitude with decimeter precision (six decimal places); the MMSI number; the reported speed over ground with two decimal places (position reports only); and vessel type and destination (static reports only), corresponding to a compressed data size of 1.01 GB (12 bytes per message). No gap larger than 400 s in the received message timestamps was experienced during the analysis period (Figure 2, top panel).

Figure 1 shows the spatial distribution of the vessel positions in the received messages. We also collected status report messages from 147 static base stations of the network in the surrounding area and show their locations in Figure 1. It can be seen that, in particular, the Gulf of Riga and the Gulf of Bothnia (the northern Baltic Sea) are poorly covered by receiver stations, resulting in low message counts in these areas. Additionally, a decrease in message density is observed between the island of Gotland and the Latvian coast, indicating poor reception by the nearest receivers of the messages sent from this area. In Figure 2 (top panel), the black curve shows the distribution of messages according to their reported times. From the message rate associated with stationary vessels (yellow curve, see Section 2.2.4 for details), we infer that we obtained a stable data stream from the subscribed service.

### 2.2. AIS Message Cleansing and Classification

Numerous studies have addressed AIS data cleansing [48,49], with recent works proposing both rule-based [50,51] and machine-learning [52,53] approaches. In this work, we reconstruct vessel movements from the four AIS message variables of reported latitude, longitude, timestamp, and MMSI number to distinguish vessels. We then apply a rule-based algorithm to:

1. Discard messages considered erroneous or irrelevant;
2. Classify messages as corresponding to a vessel being in motion, stationary, or entering or leaving the ROI—we jointly refer to the latter two states as transits;



**Figure 1.** Reported positions of the AIS-A messages received in the Baltic Sea ROI between 29 July and 27 October 2024. The figure spans from 9° E to 32° E in longitude, and from 53° N to 66° N in latitude. The 11 defined transit areas are also shown, along with the static receiver stations reported by the network. The orange dash-dotted line marks the longitude  $\lambda = 19.5^\circ$  E, used to assign messages to the UTC+1 (westwards) and UTC+2 (eastwards) time zones. All maps in this paper are shown in the Transverse Mercator projection.

### 3. Separate the moving and stationary periods.

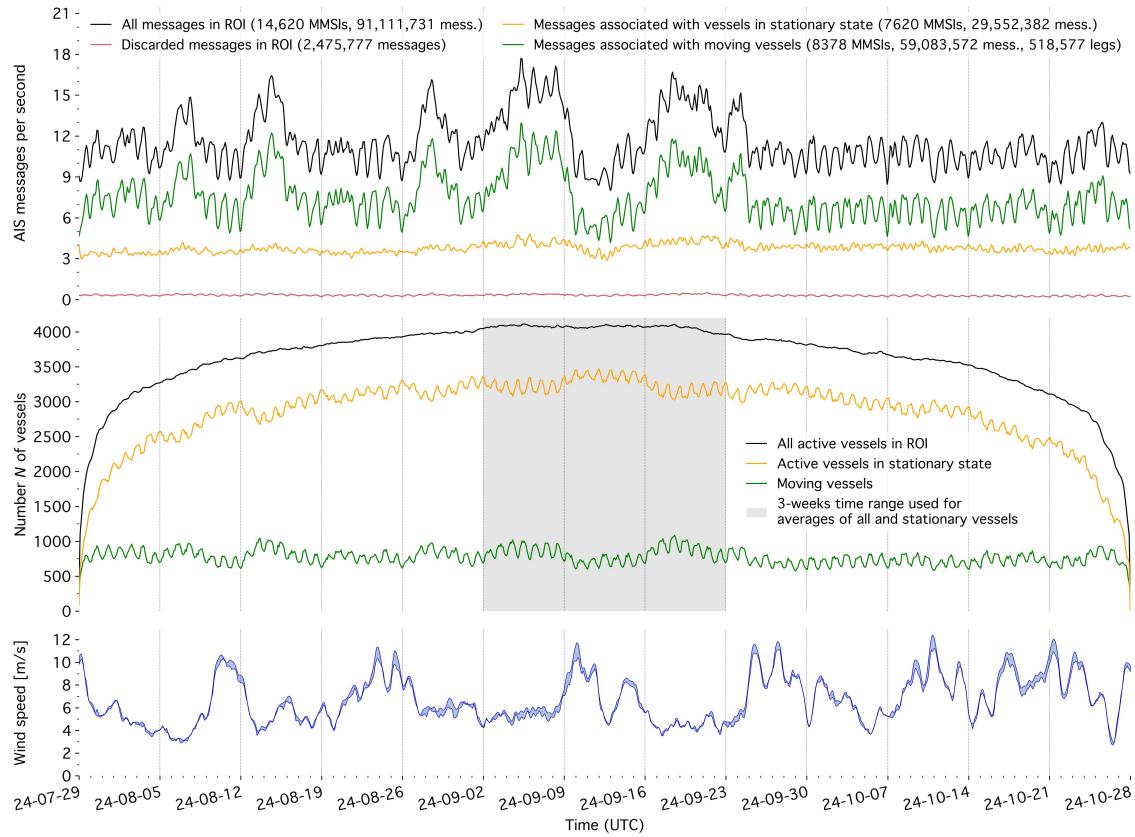
We assume a unique MMSI for each vessel. However, we discuss a possible bias due to MMSI sharing during the uncertainty estimation (Section 2.5.3).

#### 2.2.1. Removal of Static Vessels from the Analysis

We remove all vessels whose positions remain within a square of side length of 400 m during the entire 91-day period. This scale is chosen as the analysis's spatial resolution, similar to the previous analysis by HELCOM [33] and matching the available bathymetry data from NCEI ETOPO [54]. This removal defines the term "active vessel" in the remainder; we consider only vessels that move at least once. We choose this approach because, for a vessel appearing static and present only within a limited time, it is impossible to determine whether it is moored at the given position throughout the analysis period or only during the time range of received messages. This removal also eliminates spurious messages with a single, most likely scrambled MMSI number. This cut removes only 1% of all messages but 37% of all observed MMSI numbers.

#### 2.2.2. Correction of Positions in the Ship Static Data

An AIS-A vessel is required to issue a static voyage report every 6 min, for which the used data stream includes the position information of the last position report. We incorporate these static messages into the position reports; however, we correct the position coordinates under the assumption of constant speed and straight vessel movement between the enclosing position reports. This correction affects 18% of all messages (i.e., the percentage of static reports). After this, we compute vessel speeds between all subsequent messages based on the distances and time differences.



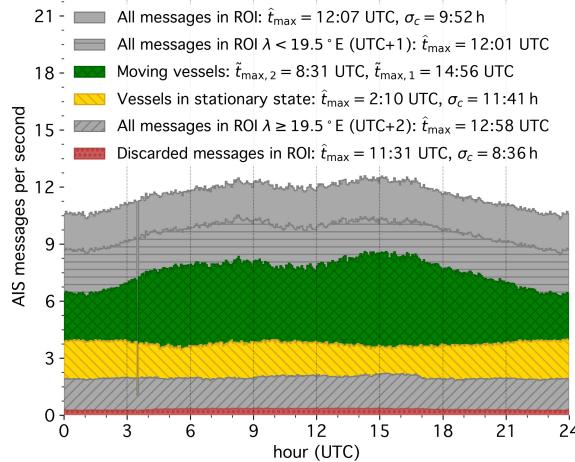
**Figure 2.** Timeline of received AIS-A message counts (**top panel**), inferred number of vessels operating in the ROI at the given time (**middle panel**), and prevalent wind conditions (**bottom panel**) during the analysis period. The vertical dotted lines denote the beginning of each Monday. The black curves in the top and middle panels represent the sum of the other curves. The gray-shaded three-week interval in the middle panel is used to derive time-averaged estimates of the total number of vessels and stationary vessels in the ROI (see Sections 2.4 and 3 for details). Wind conditions are shown as the range of the initial values given by the CMEMS [45] and NCEP GFS [46] forecast models averaged over the ROI area. Figures 3 and 4 show the top and middle panels on a daily cycle.

### 2.2.3. Removal of Duplicate Messages at Low Speeds

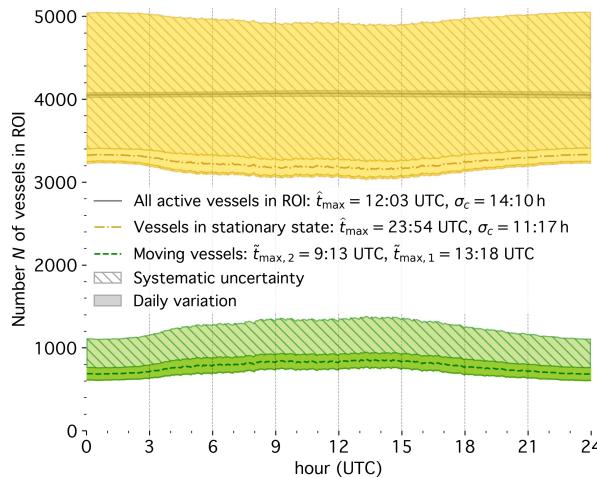
To reduce the amount of data, we also remove messages that report, at high frequency, insignificant vessel movement. We remove messages that occur within 5 s of a previous message, while the speed is below 1 km/h (the threshold for a stationary-vessel state motivated below) and the distance between the messages is less than 1 m. This removes an additional 0.7% of the messages. Messages retained after this step are shown in Figure 5. Figure 5a shows that most messages are broadcast in intervals of 1 min. The static-report interval is also visible, as are its ‘overtones’, indicating occasional missed message broadcast or reception. Figure 5b shows that the most prominent time interval corresponds to a dominant message distance of  $\sim 350$  m, or correspondingly, a dominant speed of 11 kn (Figure 5c). Figure 5d shows the messages grouped by relative acceleration. Note that three neighboring messages are always used to assign an acceleration value to the central one.

### 2.2.4. Movement Segmentation

Various methods have been proposed to identify the state of motion of a vessel and to separate moving periods from vessel stops [51,55–59]. In this work, we adopt a rule-based, iterative splitting algorithm to sort messages into periods of vessel movement and into stationary or absent phases. Stationary phases correspond to vessels being moored at a constant position or slowly drifting. Absent phases correspond to vessels that have left, but re-enter the ROI. We apply the splitting algorithm twice: once before removing messages based on speed and acceleration constraints, and again after.



**Figure 3.** AIS message activity throughout the day, stacked over the 91-day analysis period (Figure 2, top) using a bin width of 4 min. The times  $\hat{t}_{\max}$  and  $\tilde{t}_{\max}$  denote the modes and circular means, and  $\sigma_c$  the circular standard deviations [47]. The sharp gap at 3:30 UTC is a data provider feature.

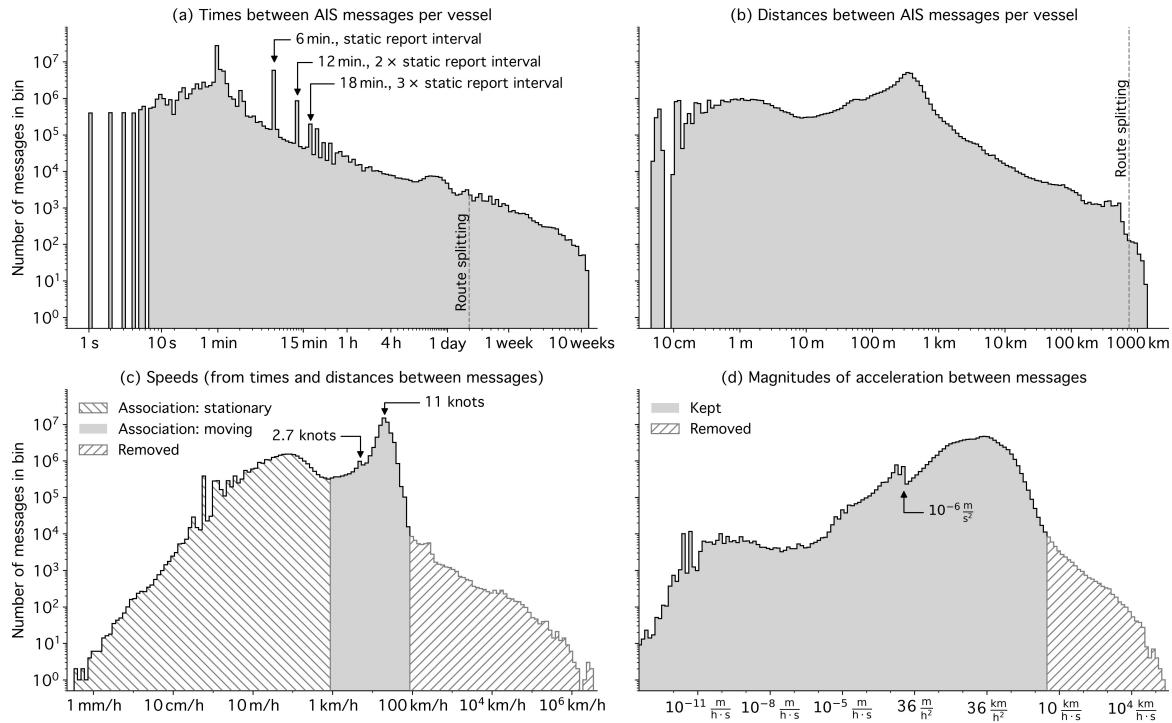


**Figure 4.** Vessel activity throughout the day. Systematic uncertainties (hatched bands) include the analysis uncertainty and estimates for additional untracked and AIS-B vessels, and are upwards dominated by AIS-B vessels.

We do this to disentangle, as far as possible, the segmentation from the cleansing. We split between those messages where at least one of the following conditions is met:

- The speed is lower than 0.5 kn ( $\sim 1$  km/h). This value is commonly acknowledged in the literature as a threshold to discriminate stationary from moving vessels [18,33,60]. Also, it is motivated by our data, where the distribution in Figure 5c shows a prominent minimum in log-space at  $\sim 0.5$  kn.
- The time gap is larger than 48 h (displayed by the dashed vertical line in Figure 5a), an estimate of the longest duration to travel a straight path through the Baltic Sea.
- The distance is greater than 750 km (displayed by the dashed vertical line in Figure 5b), also an upper bound for traversing a clear line of sight through the ROI.
- The vessel track passes through a transit area. These transit areas, defined in Table A1, are located at the ROI boundary and are shown in Figure 1.

If the splitting affects subsequent messages in a row, and only due to the speed condition, those messages are classified as belonging to a stationary period of a vessel. Single messages separated by a time or distance split (i.e., outliers in time or space) are removed. If a vessel track passes through a transit area, messages are split at the largest time difference between two subsequent messages within



**Figure 5.** Messages per vessel retained after the cleaning steps in Sections 2.2.1–2.2.3, sorted into logarithmic intervals based on the time differences between subsequent messages (a), distances (b), speeds (c), and accelerations (d) between those messages. The step in the acceleration histogram around  $10^{-6} \text{ m/s}^2$  is due to the static-report interval multiples and decimeter-length precision.

that area, including the last message before entering and the first after leaving the area. If a movement, enclosed by idle times before and after, entirely falls into a transit area, the messages are removed.

This first pass classifies 31% of the messages in the ROI to belong to stationary vessel states (approximately corresponding to the left hatched area in Figure 5c), removes an additional 0.3% of the messages, and creates 898,222 separate movements.

#### 2.2.5. Removing Outliers by Implausible Speeds and Accelerations

Among the messages classified to belong to vessel movements, we remove erroneous data indicating physically implausible speeds and accelerations between the messages. We remove the second message of any two between which the speed exceeds 50 kn (Figure 5c). This value is above the maximum cruising speed at which current high-speed vessels operate under normal conditions [61]. Also, we remove any message with an associated absolute acceleration larger than  $1 \text{ m/s}^2$  (Figure 5d), acknowledging that under normal conditions, vessels rarely experience horizontal accelerations larger than  $0.1 \text{ g}$ , the gravitational acceleration. Because the speed and acceleration between the remaining messages change after each removal, cleaning is performed iteratively until no more outliers are detected. The cleaning removes an additional 0.4% of the messages, indicated by the right hatched areas in Figure 5c,d. It also removes all but the first of multiple vessels sharing an MMSI. We discard all messages removed by this cleaning for the main analysis. However, we repeat the full analysis pipeline for the events removed in this step to assess the impact of possible physically separate vessels sharing the same MMSI (see Section 2.5.3).

#### 2.2.6. Second Segmentation Pass and Area Filtering

After cleaning from infeasible speeds and accelerations, we apply the splitting algorithm a second time, adding dozens of new movements, thousands of stationary messages, and removing seven messages. We then apply the area cut from Section 2.2.1 to each individual movement, reducing again

the number of movements by 311,665 short-distance tracks. However, this time, we do not remove the corresponding  $\sim 10^6$  messages, but classify them as belonging to stationary vessel periods.

### 2.2.7. Combining Movements

Finally, we iteratively combine all movement periods that are separated by  $\leq 120$  s, twice the most prevalent reporting interval, and remove the first message of the attached movement. This is to undo false segmentation caused by messages whose position coordinates were not updated since the previous reporting time, hence suggesting a short rest of the vessel. It also recombines uninterrupted movements through a transit area that remain within the ROI. This combination is performed in 68,002 cases (12% of the remaining movements), reducing the number of movements and increasing the number of discarded messages by this amount. This finally leaves us with 65% of all messages in the ROI associated with 518,577 movements, 32% of the messages associated with stationary periods in between, and 3% discarded messages. During the various cleansing steps, 6242 MMSI numbers were removed from the original count, leaving us with 8378 vessels for which to reconstruct their journeys.

The top panel of Figure 2 shows the messages associated with vessel movements (green curve) or stationary periods (yellow curve), as well as the removed messages (red curve). It can be seen that the irregularly fluctuating message rate is almost entirely associated with movement periods, while the rate associated with stationary vessels remains largely stable throughout the analysis period. Figure 3 shows the same data on a daily cycle.

## 2.3. Vessel Journey Model

After cleansing and classifying the messages, we build logical entities that allow us to infer the vessel position and state at any time and to compute vessel metrics and spatially resolved activity maps. As part of this process, we simplify the message information to a given precision in position and time. For the remainder of the paper, we use the following terminology: We define a route as the geometric track of a vessel's movement between two endpoints without timing information. A trajectory refers to the movement of a vessel along a route, including timing and speed information. Subsequent trajectories of a particular vessel, separated by stationary periods or times absent from the ROI, form the legs of its journey. Each vessel undertakes exactly one journey within the analysis time range.

### 2.3.1. Route Simplification

We infer the vessel routes from the message positions during the movement periods, reducing the position data to a subset of tracklets, connected at waypoints, where the vessels change course. Dedicated implementations and extensions of the Ramer–Douglas–Peucker (RDP) simplification algorithm [62,63] exist for this task [64–66]. Also, new algorithms have been published for AIS data [67,68]. However, we find that sufficient data compression while maintaining accuracy at any required precision is achieved using the fast implementation of the standard RDP algorithm by Hügel [69]. We express the algorithm's simplification threshold  $\epsilon$  to determine the required route precision in terms of the maximum tolerance distance  $d_{\text{tol}}$  from any original message position as

$$\epsilon = \frac{2 \times d_{\text{tol}}}{60 \text{ nmi}} \times \cos \bar{\varphi}, \quad (1)$$

with  $\bar{\varphi}$  the mean latitude coordinate of the route waypoints. Equation (1) converts metric distance to angular distance (in units of degrees) and applies a less stringent  $\epsilon$  for routes at higher latitudes, where larger longitude ranges represent the same metric distance. The additional factor of 2 assumes that large portions of routes were originally straight paths and ensures that they are reconstructed when positions fluctuate around both sides of the path. In the Baltic Sea at  $\cos(60^\circ) = 0.5$ , both scaling factors mostly cancel out. The performance of this threshold conversion is shown in Figure A1 (middle panel). For this analysis, we set  $d_{\text{tol}} = 100$  m, a quarter of the defined spatial analysis resolution.

### 2.3.2. Speed Model

To express the number densities of moving vessels, it is necessary to know the vessel speed at any moment in time. Therefore, along the simplified routes, we use speed control points describing the vessel motion. We define these points independently of the route waypoints, at irregular spacing, only at significant speed changes. We consider a significant speed change as a 5% variation in vessel speed, estimated at a central message position as the average over the interval between the preceding and following messages. Between speed control points, we interpolate speeds linearly in time, corresponding to constant acceleration. Also, we normalize all speed points such that the full trajectory duration matches the time difference between the first and last message in the vessel movement.

The simplified route waypoints, together with the speed model, form a trajectory model. We compared the vessel positions and times in the model with the original AIS data and found that the model, despite its simplicity, describes the vessel motion over time with an accuracy of about 1%. The detailed comparison of the simplified trajectory model with the AIS data points is provided in Appendix A.

Figures 6 and 7 illustrate this modeling for a single trajectory. In this example, the route never deviates more than 100 m from the original messages (teal diamond dots in Figure 7, bottom left), and when additionally considering timing information, model positions differ at most by 800 m or 80 s from the messages (black circles in the bottom panels of Figure 7). In Figure 7 (top panels), speeds calculated from positions and timestamps (green diamonds) show a larger fluctuation than the values reported with the AIS messages (red circles). This is due to the one-second timestamp precision and position inaccuracies. Nevertheless, the speed model (blue squares) is successfully inferred from the noisy data.

### 2.3.3. Journey Construction

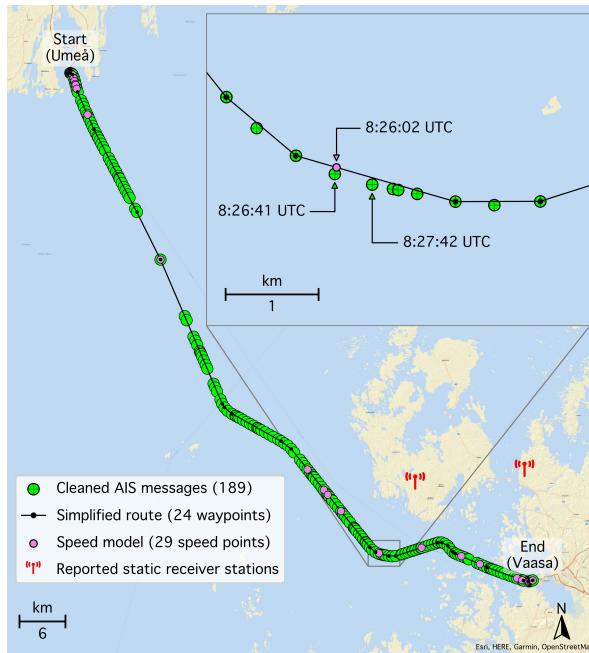
Finally, we combine the trajectories into logical units of vessel journeys, including stationary or absent times between the vessel movements. The challenge of this step is to determine, for each interval between two movements, whether a vessel remains stationary at the position connecting the movements, or whether it is temporarily absent from the ROI. This is because vessels may anchor close to the ROI boundary rather than leave, especially in areas near ports. Conversely, due to imperfect receiver coverage, vessels may leave and re-enter the ROI without messages being detected within a transit area at the ROI boundary. Also, although vessels are expected to transmit position and static reports even when not moving, we observed that this is not always the case. Sometimes, vessels switch off their transceivers while moored, even if not supposed to do so [24]. To correctly classify stationary or absent periods, we apply the following conditions if a movement starts or ends in a transit area:

- 1a. If more than three messages are received from a vessel between two movements, we assume it stays within the ROI at all times between the movements.
- 1b. In the transit areas except the Skagerrak and Kiel Canal, we only consider a vessel as temporarily leaving the ROI if the time between two movement periods exceeds

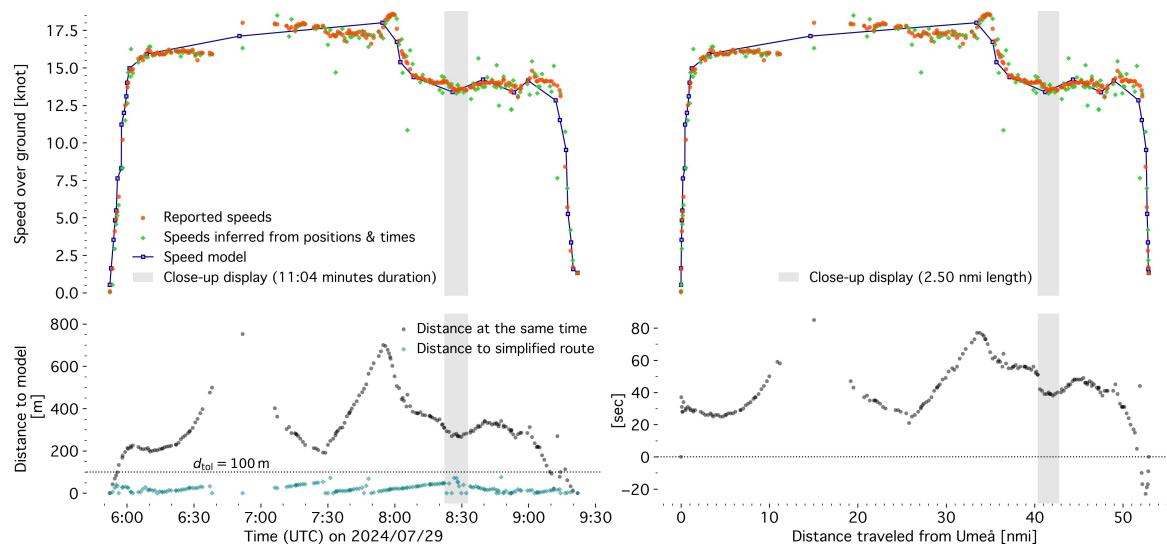
$$t_{\text{thr}} = t_0 \times (\max(v_{\text{exit}}, v_{\text{entry}}) / 10 \text{ kn})^{-4}, \quad (2)$$

where  $v_{\text{exit}}$ ,  $v_{\text{entry}}$  are the last (first) observed speed of the previous (following) leg. Using  $t_0 = 6 \text{ h}$ , we flag vessels at speeds  $> 44 \text{ kn}$  as temporarily absent for message gaps  $< 1 \text{ min}$ , the most common reporting interval (Figure 5a). The exponent in Equation (2) reduces the threshold by more than one order of magnitude for doubled speed and is chosen such that high-speed vessels ( $\gtrsim 40 \text{ kn}$ ) are always flagged as transiting.

- 2a. In the Kiel Canal transit area and within 400 m in front of it (one cell in the later chosen grid), we mark all vessels with idle time  $> 24 \text{ h}$  as absent, overruling condition 1a.
- 2b. In the whole ROI, we only consider a vessel present after receiving the first message in our analysis period, and consider a vessel absent after the last message.



**Figure 6.** An example of the reduction in original AIS messages (large green dots) into a simplified trajectory model. The model is defined by route waypoints (small black dots) and speed control points (violet dots) along the route. This example depicts a ferry trip from Umeå to Vaasa, departing on 2024/07/29 at 7:52:24 UTC+2 and arriving at 12:22:00 UTC+3. The AIS messages and speed points correspond to those in Figure 7, parametrized by time and distance.



**Figure 7.** Ferry speed during the example trajectory shown in Figure 6 over time (**top left**) and distance (**top right**) from the departure port. Both panels compare reported speeds from AIS data (red circles), speeds inferred from the reported positions and timestamps of the AIS messages (green diamonds), and our speed model (blue squares with connecting curves). (**bottom left**) Distances of message positions to the route track (teal diamonds) and to the vessel on the model trajectory at the same time (black circles). (**bottom right**) Time differences between message timestamps and the times when the ferry is at the closest position on the simplified route. Gray-shaded areas indicate the zoomed-in section in Figure 6.

Conditions 1a and 1b eliminate false-positive transit events. Conditions 2a and 2b reduce false-negative transit events, and in particular, remove “forgotten” vessels in front of transit areas. Condition 2b results in an increase in counted stationary vessels at the beginning and a decrease at the end of the analysis period, as seen in the middle panel of Figure 2. For all periods not flagged as absences,

we define the idle position as the central point between the end of the previous and the start of the following vessel movement.

## 2.4. Computation of Maritime Activity Metrics

### 2.4.1. Vessel Count and Transit Rates

From the complete journey model for each of the 8378 vessels, we compute the number of vessels simultaneously present in the ROI, and the rate of entering or leaving vessels per time. To this aim, we divide the analysis period into 32,760 intervals of 4 min each, and for each interval, we count the number of moving or stationary vessels in the ROI, and the first (or last) appearances of moving vessels flagged as absent before (or after). These counts are performed separately for six categories based on the vessels' self-reported type and two categories of vessel size. Table A2 provides the mapping of all AIS vessel codes to these categories. We retrieved the vessel sizes in terms of gross tonnage (GT) from VesselFinder Ltd. [70] for 99% of all observed vessels with an International Maritime Organization identification number ("IMO vessels"). However, IMO vessels only represent 65% of all vessels in the analysis.

The middle panel of Figure 2 shows the count of simultaneously operating vessels in the ROI. The edge effects for the count of stationary vessels (yellow curve) result from the fact that vessels entering from outside operate in the ROI for several weeks. These effects span almost the entire analysis period; therefore, we base our estimate of the average number of stationary vessels (and, by extension, the sum of moving and stationary vessels) on the central 21 days of the analysis period.

### 2.4.2. Spatially Resolved Maps

Based on the vessel journey models, we can express the overall vessel activity in terms of spatially resolved number density. Therefore, we pixelate the ROI into a regular grid with spacing  $(\Delta\varphi, \Delta\lambda) = (15'', 30'')$  in latitude  $\varphi$  and longitude  $\lambda$ , our analysis resolution. In latitude direction, this results in a cell height  $h$  of about 464 m, and in longitude, into a width  $w$  between 559 m at  $\varphi = 53^\circ$  and 378 m at  $\varphi = 66^\circ$ , similar to the resolution from HELCOM [33]. On this grid, for the vessel movements, each time a vessel passes through a grid cell, we perform the following:

1. Count the occurrence of the cell crossing;
2. Compute the average speed during the crossing;
3. Compute the average bearing during the crossing; and
4. Compute the average duration of the crossing.

In order to do so, we sample all vessel trajectories along geodesics between the waypoints in intervals of 100 m, a quarter of the cell width. If a tracklet between two waypoints crosses land at an elevation higher than 2 m above mean sea level according to NCEI ETOPO [54], we exclude the tracklet from the sampling. We choose a threshold of 2 m to avoid excluding too many tracklets in the cleft coastal areas of the Baltic Sea. If several consecutive samplings fall into the same grid cell, we assume the vessel to be constantly present within the cell, and we assign a single crossing, the mean speed, and circular mean bearing for this crossing to the cell. For a computationally feasible estimate of the crossing duration, we compute the crossing time from the average path length and the average vessel speed within the cell. The average path length through a rectangular cell under a constant bearing angle  $\alpha$  is defined by Equation (A4). After sampling an entire trajectory, we normalize the sum of crossing durations to the trajectory duration (with the duration of the excluded tracklets subtracted). Finally, after having sampled all trajectories, we sum up all crossing occurrences and durations in each cell.

For the stationary periods, we compute the mooring durations of each vessel in the corresponding cells. However, we do not spatially resolve stationary periods where the distance between the previous leg end and the following leg start coordinates exceeds 1 km, due to the uncertainty of the actual vessel position during the idle period.

We then sum up the movement times and stationary durations and divide by the analysis period duration (91 days) and cell area to obtain the average number of vessels per unit area.

#### 2.4.3. Finding Port Areas

The combined stationary and moving vessel density allows us to estimate port areas complementary to previous works. Various techniques have been described to identify anchoring and port locations, using spatial clustering algorithms [22,51,57,59,71], a kernel density approach [72], or machine learning on training data [73]. In this work, we infer port areas from the pixelized number density through a two-step map segmentation. First, we find seed locations as the barycenter positions of areas enclosed by a threshold of 0.5 vessels/km<sup>2</sup>, with Gaussian smoothing over 1.5 cells ( $\sim 700$  m) to denoise the density. Using these seed locations, we determine port areas with a watershed algorithm [74], applied to water areas above half the threshold value with the same smoothing, and the port locations as barycenter positions within these areas. The density threshold is the single pre-determined parameter of the algorithm. For the purpose of this study, it is chosen such that we obtain a median area size of about 4 km<sup>2</sup> for a single port. Additionally, we dilate all areas to a size of at least three cells ( $\sim 0.7$  km<sup>2</sup>). From these areas, we also estimate vessel arrivals by counting all journey legs that end within port areas (but do not start within the same areas). For the Kiel port area, containing the Kiel Canal transit area, we exclude vessels that enter the canal, but vessels can arrive in the fjord from the canal.

### 2.5. Uncertainties in the Vessel Metrics

#### 2.5.1. Uncertainty About the Receiver Coverage

Quantifying incomplete and inhomogeneous AIS coverage is generally a challenge for AIS data analysis [16]. In principle, reconstructed trajectories should not depend on message frequency along vessel routes. However, this does not hold when trajectories cannot be constructed at all due to insufficient message data, or when critical waypoints marking course direction changes are missing. Within the ROI, poor data coverage leads to an underestimation of the relative share of moving vessel time in the total vessel count. While we lack comparison data to quantify this uncertainty, we assume the impact on the total vessel count to be minor. We attempt to quantify the impact of poor coverage at the ROI boundaries on vessel counts and transit rates, as described below.

#### 2.5.2. Uncertainty About Leaving or Moored Vessels

If a vessel is falsely flagged as moored within the ROI while it is in reality leaving and reentering (or vice versa), it confounds both the count of stationary vessels and the vessel transits. The movement may have been detected as interrupted by the time cut in Section 2.2.4, but is falsely identified as an idle period instead of an absence.

In Section 2.3.3, we applied rules to reduce such false-positive and false-negative transit counts. To estimate the uncertainty of false-positive and false-negative transits, we create two variations in our default set (case *df*) of vessel journeys: a variation further reducing false positives (case *hi*), and a variation reducing false negatives, but allowing for more false positives (case *low*).

- For the Skagerrak and Kiel Canal, we vary the transit areas in which to check for absence in the journey construction of Section 2.3.3, using a region smaller (case *hi*) and larger (case *low*) than the default. A larger area reduces false-negative transits, but may increase false positives (see Table A1 for all area definitions).
- For all other transit areas, we only create a looser condition on false positives by setting  $t_0 = 1$  h in Equation (2) (case *low*).

Case *low* results in lower stationary-vessel numbers and higher transit rates, and vice versa for case *hi* (detailed in Table A3). We include this variation in our overall uncertainty estimate individually for each vessel category, with uncertainties up to 17% (cargo vessels) for the vessel counts, and up

to 55% (Kiel Canal) for the all-vessel transit rates. Due to low statistics, the transit-rate uncertainty becomes 100% for some categories.

### 2.5.3. Uncertainty About MMSI Sharing

Multiple simultaneously operating vessels may share the same MMSI (e.g., [48,75]). In our analysis, all of those except the one with the first detected message would be discarded in the cleansing. To assess the possibility of having removed valid vessels with duplicate MMSI numbers, we repeated the full analysis pipeline using only the messages removed in the cleansing step, Section 2.2.5. After scrutinizing the obtained vessel journeys, we found that all checked trajectories correspond to the same as in the retained data, and that they are built from messages rejected due to time jitters, etc.

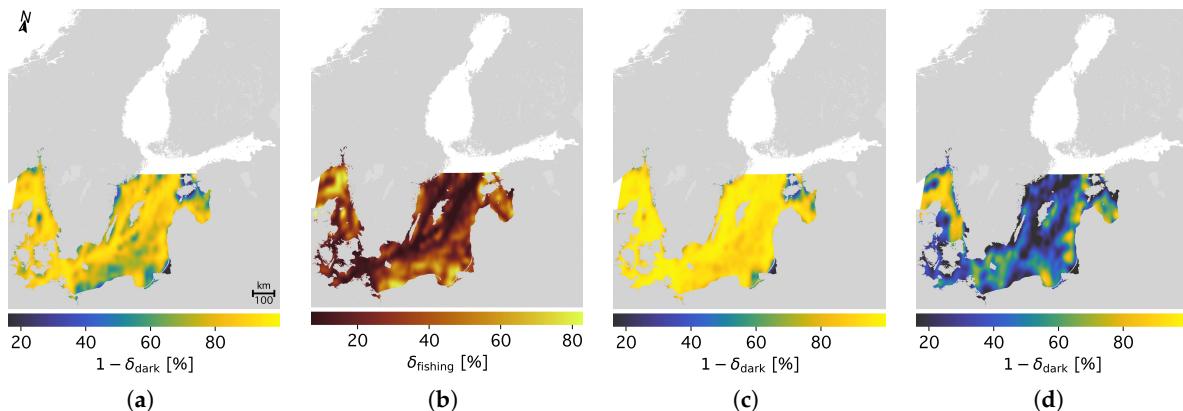
For an estimate of the largest possible error, we obtain from the analysis of rejected messages a total travel time of 7104 d on valid tracklets versus 60,577 d in the main analysis. With this, we would underestimate vessel activity by at most 10% due to rejecting vessels with duplicate MMSI. If we assumed the same message frequency for all vessels, this value would reduce to at most the 3% rejected messages. In summary, we assume that the error due to MMSI sharing is negligible, consistent with the analysis from the supplementary material in Kroodsma et al. [17], and do not include it in our combined uncertainty estimate.

### 2.5.4. Uncertainty About Vessels Not Using the AIS

Some vessels do not submit any AIS signals, either because they are not required, miss having their transceiver switched on, or intentionally, legitimately or illegitimately, prefer to operate unnoticed [24]. Paolo et al. [4] conducted a comprehensive correlation of worldwide satellite imagery with reported AIS positions in coastal waters. In Figure 8, we reproduce their public data at  $(\Delta\varphi, \Delta\lambda) = (0.2^\circ, 0.2^\circ)$  resolution. From their data (Figure 8a), we read

$$\delta_{\text{dark}} = \begin{cases} 21\% & (\text{full ROI}) \\ 13\% & (\text{Skagerrak at } 9^\circ \text{ E}) \\ 0\% & (\text{Kiel Canal at Holtenau}) \end{cases} \quad (3)$$

the fraction of untracked vessels in the Baltic Sea and the Skagerrak Strait. Paolo et al. [4] do not provide data in the direct vicinity of shorelines, including the Kiel Fjord, and we assume that all vessels passing the Kiel Canal correctly use the AIS. While Paolo et al. [4] find that the usage of the AIS is particularly poor among fishing vessels (Figure 8d), we include a global constant fraction, defined by Equation (3), in our combined uncertainty estimate for simplicity.



**Figure 8.** Fraction of vessels tracked by the AIS and the fraction of fishing vessels in the Baltic Sea between 2017 and 2021, based on data from Paolo et al. [4]. Paolo et al. only analyze the region  $\varphi \leq 59.1^\circ \text{ N}$  within the Baltic Sea. (a) Fraction of tracked vessels among all vessels. (b) The fraction of fishing vessels among all vessels. (c) The fraction of tracked vessels among non-fishing vessels. (d) The fraction of tracked vessels among fishing vessels.

### 2.5.5. Uncertainty About AIS-B Vessels

As stated in Section 2.1, our analysis excludes vessels using the AIS-B. We find from our data that in the summer months, about twice as many AIS-B vessels ( $\gtrsim 40\%$  of all vessels) are operating in the Baltic Sea as in the winter months ( $\sim 20\%$ ), primarily attributed to leisure activity. On a yearly average, we assume that a fraction of 30% of all vessels using the AIS use Class B transceivers. For simplicity, we assume the same fractions for moving and stationary vessels.

We find that the fraction of all vessels that enter or leave the Baltic Sea and use the AIS-B is lower, at 21%, than their share among the activities within the ROI. The percentage of AIS-B vessels is slightly larger (22%) among the traffic through the Skagerrak than through the Kiel Canal (20%). The percentages for each vessel category used in the uncertainty estimates for vessel numbers and transit rates are provided in Table A4.

### 2.5.6. Combined Uncertainty of Vessel Counts

We include in our uncertainty estimates the bracketing of false-positive/negative transits and the underestimation of vessels due to missing untracked and AIS-B vessels. We build upper (+) and lower (−) bounds  $\Delta N_{\text{syst}}$  and  $\Delta \dot{N}_{\text{syst}}$  on vessel numbers  $N$  and transit rates  $\dot{N}$  as

$$\begin{aligned} (\Delta N_{\text{syst}}^+)^2 &= (N_{\text{df}} - N_{\text{hi}})^2 + (\tilde{\delta}_{\text{dark}}^2 + \tilde{\delta}_{\text{ais-b}}^2) N_{\text{df}}^2, \\ \Delta N_{\text{syst}}^- &= N_{\text{df}} - N_{\text{low}}, \\ (\Delta \dot{N}_{\text{syst}}^+)^2 &= (\dot{N}_{\text{df}} - \dot{N}_{\text{low}})^2 + (\tilde{\delta}_{\text{dark}}^2 + \tilde{\delta}_{\text{ais-b}}^2) \dot{N}_{\text{df}}^2, \\ \Delta \dot{N}_{\text{syst}}^- &= \dot{N}_{\text{df}} - \dot{N}_{\text{hi}}, \end{aligned} \quad (4)$$

with  $N$ ,  $\dot{N}$ , and  $\delta_{\text{ais-b}}$  according to each category under scrutiny, and  $\delta_{\text{dark}} = 21\%$  except for the transit rates  $\dot{N}$  through the Skagerrak and Kiel Canal according to Equation (3), and  $\tilde{\delta} = \delta/(1-\delta)$ . In all tables and figures, systematic uncertainties are provided according to Equation (4).

## 3. Results

### 3.1. Vessel Count and Density Within the Baltic Sea

We find an average total of 4061 vessels operating in the Baltic Sea at any moment in time between August and October 2024, divided into 774 moving vessels and 3287 stationary vessels (Table 1). Stationary vessels include ships with negligible speed over ground, such as those moored or drifting while engaged in fishing activities. Compared to previous estimates, this number is about double the figure reported by Storgard et al. [76], who claim “nearly 2000 ships operating at any given moment [in the Baltic Sea]” and is 7% higher compared to the more recent estimate by Czermański [77], albeit in our case, on a larger ROI extending into the Skagerrak. For fishing vessels alone, Kroodsma et al. [17] find an average of 729 vessels in our analysis ROI between 2012 and 2016, compatible with our count of 1028 vessels in the category “Others including fishing” (Table 1). Restricting fishing activity to self-reported AIS code 30 (codes 1002 and 1003 were not observed) results in an average of 378 active fishing vessels in our analysis. Figure 2 (middle panel) shows these numbers over the 91-day analysis period. The total number of vessels (black curve) remains largely constant over both the course of a day and week, with the relative contribution of moving (green curve) and stationary vessels (yellow curve) strongly fluctuating during the day. Additionally, the movement state correlates with the average wind speed in the Baltic Sea (Figure 2, bottom panel). For moving vessels (green curve in Figure 2, middle panel), we do not observe any significant long-term variation on our 91-day timescale.

Figure 4 maps the same vessel count shown in Figure 2 onto a daily scale. The average total vessel count changes by only 0.4% throughout the day, peaking around noon UTC and being slightly lower at night (gray solid line). However, the proportion of moving vessels (green dashed line) varies by more than 20%, with modal peak activity occurring shortly after 1 p.m. UTC (2 p.m. in Central European Summer Time, 3 p.m. Eastern European Summer Time). A shallower secondary peak appears in the

**Table 1.** Average number  $N$  of vessels operating in the Baltic Sea (August–October 2024). The first error terms represent statistical fluctuations based on day and time. The second error terms account for systematic uncertainty regarding the ROI transits (both all vessels and stationary vessels) and the estimates on AIS-B and untracked vessels (positive terms only). The table continues in Table 2.

Vessel Type	All Active Vessels	Moving Vessels	Stationary Vessels	Percentage (All)	Percentage (Moving)	Percentage (Stationary)
All vessels in ROI	$4061 \pm 4^{+1639}_{-22}$	$774 \pm 57^{+389}_{-0}$	$3287 \pm 62^{+1639}_{-22}$	100%	19%	81%
Passenger, high-speed	$642 \pm 2^{+145}_{-1}$	$137 \pm 38^{+40}_{-0}$	$505 \pm 37^{+145}_{-1}$	16%	18%	15%
Law enforcement, military	$211 \pm 0.3^{+71}_{-0}$	$15 \pm 5^{+5}_{-0}$	$196 \pm 5^{+71}_{-0}$	5%	2%	6%
Cargo	$939 \pm 1^{+230}_{-10}$	$338 \pm 9^{+102}_{-0}$	$601 \pm 8^{+230}_{-10}$	23%	44%	18%
Pilot, tug, rescue, diving/dredging	$837 \pm 1^{+267}_{-5}$	$54 \pm 10^{+20}_{-0}$	$783 \pm 11^{+267}_{-5}$	21%	7%	24%
Tanker	$404 \pm 1^{+82}_{-4}$	$126 \pm 1^{+33}_{-0}$	$278 \pm 1^{+82}_{-4}$	10%	16%	9%
Others including fishing	$1028 \pm 1^{+1152}_{-3}$	$104 \pm 14^{+129}_{-0}$	$924 \pm 18^{+1152}_{-3}$	25%	13%	28%

**Table 2.** Average number  $N$  of IMO vessels in the Baltic Sea and further categorized by gross tonnage (GT, August–October 2024). GT data are available for 99% of observed IMO vessels.

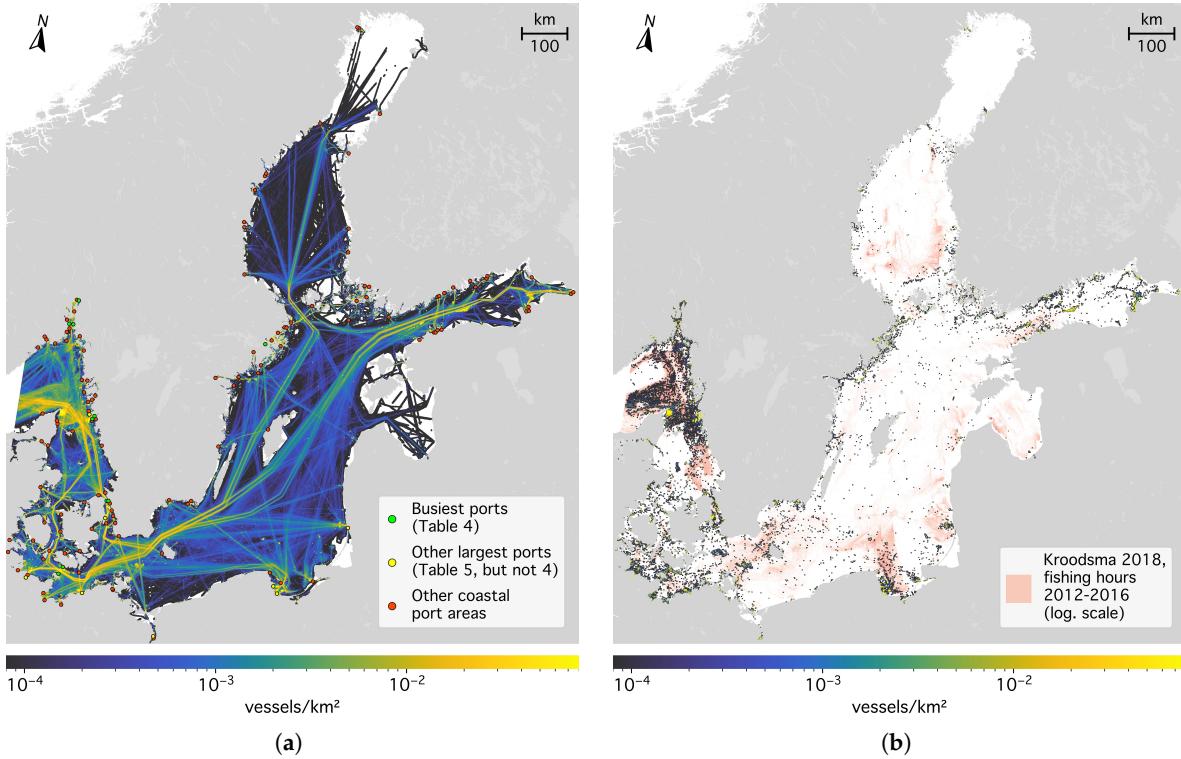
	All IMO Vessels	GT < 10,000	GT $\geq$ 10,000
All active vessels	$2265 \pm 3^{+816}_{-19}$	$1672 \pm 2^{+811}_{-17}$	$571 \pm 1^{+86}_{-3}$
Moving vessels	$648 \pm 22^{+323}_{-0}$	$399 \pm 30^{+252}_{-0}$	$243 \pm 11^{+64}_{-0}$
Stationary vessels	$1587 \pm 23^{+816}_{-19}$	$1273 \pm 31^{+811}_{-17}$	$328 \pm 11^{+86}_{-3}$
Percentage (all active vessels)	56%	$\geq 41\%$	$\geq 14\%$
Percentage (moving vessels)	84%	$\geq 52\%$	$\geq 31\%$
Percentage (stationary vessels)	48%	$\geq 39\%$	$\geq 10\%$

morning around 9 a.m. UTC. We find that most vessel activity occurs in the western Baltic (accounting for 83% of all vessel activity in terms of messages), as does the fluctuation in daytime activity (Figure 3). The message rate from stationary vessels (yellow upper diagonally hatched distribution in Figure 3) peaks at  $\hat{t}_{\max} = 2:10$  UTC. Similarly, the number of stationary vessels (dashed yellow curve in Figure 4) is largest at  $\hat{t}_{\max} = 23:54$  UTC, 6% larger than at daytime. The count of stationary and moving vessels only slightly varies between different days ( $\sim 12\%$  or  $\sim 90$  vessels; Figure 4, yellow and green solid bands). Uncertainties considered in Section 2.5 result in a wide range of possible vessel counts. This combined uncertainty range is represented by the hatched bands in Figure 4. Therefore, vessel counts depend strongly on whether AIS-B vessels or even ships lacking AIS tracking devices are included in the definition.

Tables 1 and 2 provide vessel counts categorized by type and their relative contributions. We find that cargo and tanker vessels (Table 1) spend most of their time moving, as do large vessels (Table 2). Conversely, vessel types such as pilot or tugboats, rescue vessels, or fishing vessels are mostly found moored or drifting.

Figure 9 presents the spatial distribution of average vessel counts for all vessels (Figure 9a) and for the subset of stationary vessels (Figure 9b). Because moving vessels are not represented on incorrectly reconstructed tracklets, and stationary vessels are omitted if their position cannot be determined with at least 1 km accuracy (Section 2.4.2), these figures represent only 55% of all observed vessels and 49% of stationary vessels.

Figure 9a shows the limitations of our AIS dataset in accurately capturing vessel activity in the Gulf of Riga and Bothnia. Additionally, large distances to the next receivers between Gotland and the Latvian coast result in poorly reconstructed tracklets. This is indicated by the seemingly high vessel density close to the southeast coast of Gotland, which is an artifact of the incomplete data. In Figure 9b, a large number of stationary vessels can be seen in the northern Kattegat and Skagerrak regions. This aligns with the high fishing activity in these areas, as indicated by the red-shaded area from Kroodsma et al. [17]; see also HELCOM [78] and the reproduced result from Paolo et al. [4] in Figure 8b.



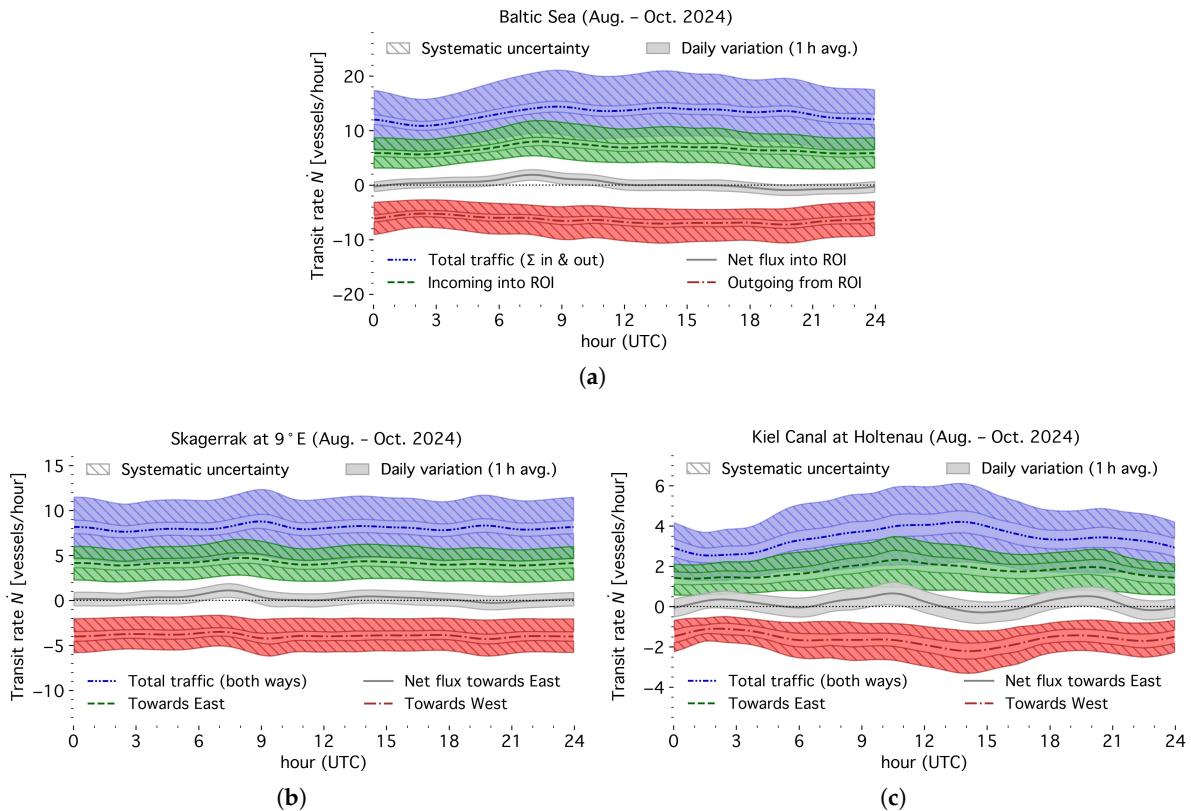
**Figure 9.** Vessel density in the Baltic Sea, average from August to October 2024. Densities are shown for all vessel types combined. The left panel also shows inferred coastal port areas (Section 3.3), with the busiest (Table 4) marked in green, and additional large ports (Table 5) in yellow. (a) Combined density of all moving and stationary vessels (55.2% of the average 4061 simultaneous vessels shown). (b) Stationary vessels with positions found at 1 km accuracy (48.6% of 3287 vessels shown). Fishing data from Kroodsma et al. [17].

### 3.2. Traffic Within, into, and from the Baltic Sea

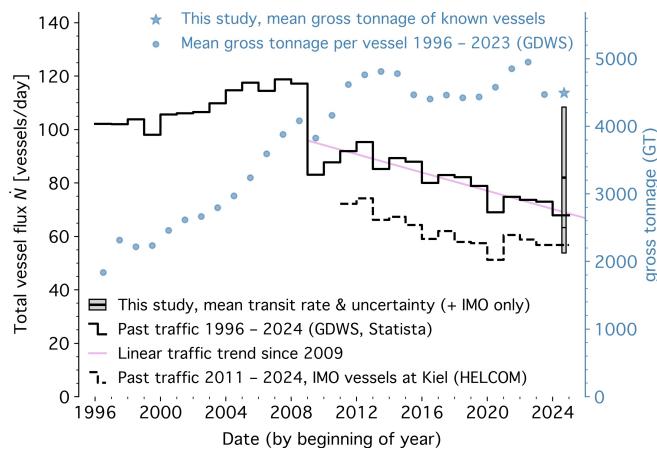
Figure 10a–c resolve the overall transit rate and the rates through the most prominent transit areas over the course of the day. We observe a slight net inflow of vessels into the Baltic Sea in the morning and net outflow in the afternoon, primarily driven by traffic through the Skagerrak (black solid lines in Figure 10a,b). Notably, we observe an approximately eight-hour period of net traffic direction change at Kiel Holtenau (black solid line in Figure 10c). The transit rates through these areas are listed in Table 3.

**Table 3.** Transit rates  $\dot{N}$ : Vessels entering or leaving the Baltic Sea each day (August–October 2024).

Vessel Type	All Transit Areas		Skagerrak at 9°E		Kiel Canal at Holtenau	
All vessels	$313 \pm 23^{+121}_{-98}$	100%	$194 \pm 14^{+61}_{-67}$	100%	$82 \pm 6^{+26}_{-28}$	100%
Passenger, high-speed	$32 \pm 2^{+17}_{-31}$	10%	$17 \pm 1^{+3}_{-1}$	9%	$0.7 \pm 0.05^{+0.2}_{-0.7}$	1%
Law enforcement, military	$3.0 \pm 0.2^{+1.2}_{-1.8}$	1%	$1.7 \pm 0.1^{+0.5}_{-0.8}$	1%	$0.4 \pm 0.03^{+0.1}_{-0.4}$	<1%
Cargo	$144 \pm 11^{+44}_{-65}$	46%	$97 \pm 7^{+21}_{-34}$	50%	$34 \pm 3^{+5}_{-28}$	41%
Pilot, tug, rescue, diving/dredging	$37 \pm 3^{+18}_{-29}$	12%	$3.0 \pm 0.2^{+0.8}_{-1.4}$	2%	$33 \pm 2^{+14}_{-28}$	40%
Tanker	$57 \pm 4^{+15}_{-21}$	18%	$46 \pm 3^{+7}_{-16}$	24%	$7.9 \pm 0.6^{+0.3}_{-4.8}$	10%
Others including fishing	$39 \pm 3^{+50}_{-17}$	12%	$29 \pm 2^{+37}_{-14}$	15%	$6.2 \pm 0.4^{+8}_{-3}$	8%
IMO vessels	$266 \pm 19^{+123}_{-99}$	85%	$176 \pm 13^{+59}_{-59}$	91%	$63 \pm 5^{+33}_{-37}$	77%
Vessels with GT < 10,000	$173 \pm 12^{+110}_{-73}$	$\geq 55\%$	$92 \pm 7^{+55}_{-37}$	$\geq 47\%$	$57 \pm 4^{+34}_{-34}$	$\geq 70\%$
Vessels with GT $\geq 10,000$	$95 \pm 7^{+26}_{-26}$	$\geq 30\%$	$82 \pm 6^{+12}_{-22}$	$\geq 42\%$	$8.6 \pm 0.6^{+0.5}_{-2.9}$	$\geq 10\%$



**Figure 10.** Vessels per hour entering or leaving the Baltic Sea ROI through all transit areas (**top panel**), only the Skagerrak Strait (**bottom left panel**), or the Kiel Canal (**bottom right panel**). No systematic uncertainty is shown for the net flux for simplicity. (a) Vessels entering or leaving the Baltic Sea during the day. Shown are net traffic (gray curve and band without systematic uncertainties); inbound (green dashed curve and bands), outbound (red dash-dotted curve and bands), and total traffic crossing the ROI boundary (blue dash-double-dotted curve and bands). (b) Part of the traffic shown in Figure 10 passes through the Skagerrak. (c) Part of the traffic shown in Figure 10a passes through the Kiel Canal. In the bottom panels, a course direction towards the east corresponds to entering vessels, and towards the west to leaving vessels.



**Figure 11.** Timeline of historic traffic (full and section passages; left black scale) and average GT per vessel (right blue scale) through the Kiel Canal, compared with our results. For our results, the gray box marks the uncertainty range of the transit rate into/from the Canal, with the thick horizontal line the mean all-vessel transit rate, and the thin horizontal line the rate for IMO vessels only. Our mean estimated GT is marked by the blue star. Results are compared against the data from [80–83] (black curve and blue dots) and HELCOM [84] (black dashed curve).

Figure 11 compares our results of traffic into or from the Kiel Canal at the Holtenau lock with historic data from other sources. The Kiel Canal is the world's busiest artificial waterway in terms

of passages [79]. On average, around 80 vessels have traveled each day through the full canal or a section of it over the last decade (black solid line based on data from [80–83]). However, traffic has declined since 2009 (violet line). Our 2024 estimate (gray box) is consistent with previous years and the overall trend within uncertainties. Our mean all-vessel transit rate (thick black horizontal line inside the box) is 20% larger than the reported all-year traffic in 2024. While the historic reference data in principle include sectional passages, it may not account for all short-term trips from or into the canal by pilot, tug, or dredging vessels, which are included in our analysis. Additionally, Figure 11 displays the vessel transits at Holtenau as inferred from the HELCOM data [84] (black dashed curve), which can be compared with our result for IMO vessels only (thin black horizontal line inside the gray box). Here, too, our result is slightly (11%) higher than the all-year data from 2024.

While fewer vessels have passed the Kiel Canal over the last decades, the average vessel gross tonnage on full and sectional passages increased in the years prior to 2023, but dropped in 2023 (blue dots in Figure 11; no all-passage GT values have been officially published for 2024). Based on the available gross tonnage information for IMO vessels, our three-month window suggests a value lower than before 2023, also in 2024 (GT = 4493, blue star in Figure 11). This indicates that not only has the number of vessels decreased, a long-term trend towards fewer, but larger vessels since the 1960s [82], but also the total volume of goods transported through the Kiel Canal has recently declined. Interestingly, such a net decline in transported volume may also be connected to the trend towards larger vessels, which may exceed the capacity of the Kiel Canal and are, therefore, routed via the Skagerrak.

### 3.3. Port Areas in the Baltic Sea

Using a map segmentation of the vessel density (Section 2.4.3), we find 145 overdensities in the ROI with  $3.9 \text{ km}^2$ ,  $5.6 \text{ km}^2$ , and  $33.0 \text{ km}^2$  in median, mean, and maximum size, respectively. In total, 95% of those areas have a barycenter location within 2.7 km from the coast. Among these coastal port areas, we identify 44% of 143 Baltic Sea ports in the World Port Index (WPI, [85]) classified as small or larger and 85% of the large ports, with Riga, Sillamae, and Oulu missing among the latter. Conversely, 67% of the identified overdensities correspond to a WPI-listed port. Our algorithm also detects several short-distance ferry connections without major port infrastructure. Figure 9a shows the port areas located within 2.7 km of the shoreline.

Table 4 lists the ten busiest identified ports in terms of arrivals. These ports include the following pairs of ferry lines in the Baltic Sea with the highest frequency: the Helsingør–Helsingborg ferry between Denmark in Sweden serving every 20 min during daytime (first two rows), the Moss–Horten ferry in the Oslo Fjord (Norway; forth and eighth rows), and the Rødby–Puttgarden connection between Denmark and Germany (seventh and ninth rows). For all these ports, except Moss, we obtain an arrival rate equal to or greater than expected from the corresponding ferry timetables. For Moss, we inferred only 53.7 arrivals per day, while we expect 54 arrivals from the ferry connection alone. Our arrival rates are about a factor of ten larger than the daily AIS-inferred port calls published by Arslanap et al. [86]. Table 5 ranks the ten largest ports by the number of vessels moored in the areas found in our analysis. We find that Gothenburg is the largest port, also by area, with almost two vessels per  $\text{km}^2$  on an area larger than  $30 \text{ km}^2$ . The remaining ports in the table belong to major urban areas in the Baltic Sea region, among them six of the ten ports found by Synak-Miłosz and Rozmarynowska-Mrozek [87] with the largest cargo turnover in the first half of 2023: Gdańsk, Gothenburg, Szczecin, Rostock, Klaipéda, and Gdynia.

We also list the most common destination names provided by the AIS in these tables. No standard format exists for these data. Sometimes, the arrival port is provided either by UN/LOCODE or real name, sometimes both the last departure and destination ports are provided, and sometimes the field is used to provide information about the carrier. Additionally, spelling mistakes were frequently observed in our data. This renders, in particular, the human-inputted AIS information challenging for automated data analysis [24,88,89].

**Table 4.** The ten busiest port areas in the Baltic Sea (August–October 2024) in terms of arrivals.

Rank	$\varphi$ [°]	$\lambda$ [°]	Name	Most Common AIS Destination Name	Arrivals per Day	Vessels in Port	Port Area [km²]	Density [Vessels/km²]
1	56.040	12.679	Helsingborg (SE)	FERRY DK	74.4	10.0	8.9	1.1
2	56.035	12.612	Helsingør (DK)	FERRY DK	71.9	6.8	7.7	0.9
3	57.673	11.854	Gothenburg (SE)	STYRSOBOLAGET FRAKT	68.8	55.8	33.0	1.7
4	59.419	10.479	Horten (NO)	HORTEN MOSS	63.8	12.1	9.0	1.3
5	59.898	10.729	Oslo (NO)	NOOSL	62.0	18.7	9.5	2.0
6	54.644	11.354	Rødbyhavn (DK)	DKROD	57.0	20.3	17.0	1.2
7	59.419	10.638	Moss (NO)	HORTEN MOSS	53.7	4.3	4.4	1.0
8	54.502	11.229	Puttgarden (DE)	DKROD	47.4	6.4	6.3	1.0
9	57.610	11.788	Southern Gothenburg archipelago (SE)	STYRSOBOLAGET FRAKT	37.4	6.3	6.9	0.9
10	59.319	18.104	Stockholm (SE)	SL LINJE	35.0	48.4	8.1	5.9

**Table 5.** The ten largest port areas in the Baltic Sea (August–October 2024) in terms of average number of vessels in port.

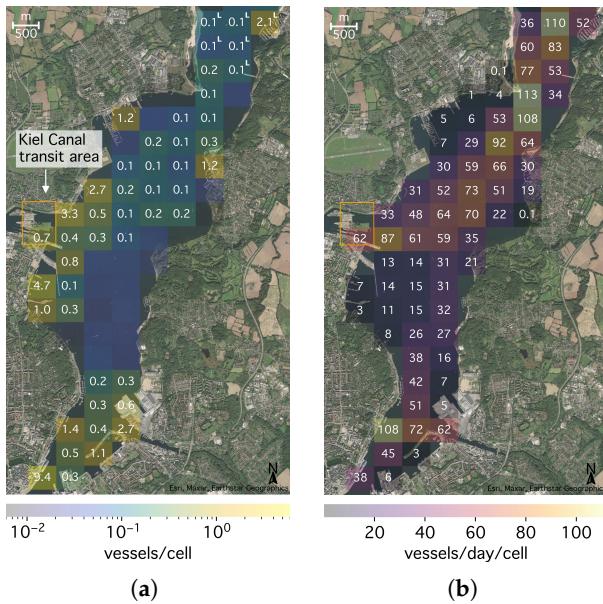
Rank	$\varphi$ [°]	$\lambda$ [°]	Name	Most Common AIS Destination Name	Arrivals per Day	Vessels in Port	Port Area [km²]	Density [Vessels/km²]
1	57.673	11.854	Gothenburg (SE)	STYRSOBOLAGET FRAKT	68.8	55.8	33.0	1.7
2	59.319	18.104	Stockholm (SE)	SL LINJE	35.0	48.4	8.1	5.9
3	54.390	18.671	Gdańsk (PL)	GDANSK ANCHORAGE	18.8	43.5	23.6	1.8
4	55.681	21.129	Klaipėda (LT)	KLAIPEDA LITHUANIA	14.3	41.5	11.4	3.6
5	54.531	18.546	Gdynia (PL)	GDYNIA POLAND	14.3	40.7	18.8	2.2
6	54.177	12.096	Rostock (DE)	DERSK HANSESAIL	33.9	38.9	17.4	2.2
7	54.348	10.154	Kiel Fjord without Laboe (DE)	KIEL PILOT	32.1	37.1	16.1	2.3
8	53.444	14.588	Szczecin (PL)	SZCZECIN PLSZZ	9.5	29.2	9.5	3.1
9	55.694	12.621	Kopenhagen (DK)	DKCPH	13.4	23.8	23.8	1.0
10	57.694	11.662	Northern Gothenburg archipelago (SE)	BJORKO	7.8	22.8	14.1	1.6

Finally, Figure 12 zooms into the Kiel Fjord port area (ranked 7 in Table 5), showing spatially resolved vessel density (Figure 12a) and traffic in terms of cell crossings (Figure 12b). This port area also contains the Kiel Canal transit area, marked by the orange box. Cells belonging to the separate area of Laboe port are marked by the letter 'L' in Figure 12a. We verified the average densities (Figure 12a) by comparing against momentary live data from VesselFinder Ltd. [70] and found consistent results.

#### 4. Discussion

In this work, we have presented a method to infer maritime activities from open-access AIS positioning data. By analyzing three months of vessel activity in the Baltic Sea, we have shown that available open data is of sufficient quality to study coastal regions as large as 1000 km in size, and provide insights that are competitive with those derived from proprietary data. Our estimates of both the average vessel count in the area and of the traffic into and from the Kiel Canal agree within 20% with previous studies, after we have assessed multiple sources of systematic uncertainty and have justified their quantification. Using a simple stop-identification algorithm and density thresholding, we have identified half of the 143 major WPI-listed ports in the region, and we obtained metrics for these port areas consistent with other sources. This shows that reliable statistics on coastal-water vessel activity can be obtained from publicly available and community data, supporting research in economic, environmental, and transportation sciences.

A major challenge in the limited time-period analysis of regional data is the treatment of edge effects in space and time. At the spatial boundaries, this involves properly identifying transiting or



**Figure 12.** Vessel density and traffic in the port areas of the Kiel Fjord (Kieler Förde) and part of Laboe (marked by an L in the left panel). No densities are indicated for less than 0.05 vessels per cell. **(a)** Average vessel density of moving and stationary vessels. **(b)** Daily cell crossings by moving vessels.

idle vessels. If data is available, expanding the analysis region beyond the actual ROI can mitigate this; for example, to infer the Kiel Canal traffic with larger precision, as presented in this analysis. Also, if training data is provided, machine-learning methods are well suited for such a discrimination task. However, extended or reference data may not always be available. In a separate work, we will show that the rule-based cleansing and discrimination techniques presented in this paper can also be robustly applied at higher spatial resolution to AIS data only covering single port areas and from single receivers with fluctuating and fading coverage. Temporally, we have found the chosen three-month interval for the given ROI to be considerably affected by edge effects. For longer periods or continuous analysis, techniques like split-apply-combine strategies [90] may be indispensable to manage memory and to efficiently parallelize computation.

In this study, we did not address the detection of intentionally manipulated AIS data. Vessels that temporarily deactivate their transceivers for illegal activities are not distinguished from those omitting position signals unintentionally [24,88] or due to other data gaps. During such gaps, vessel routes are always reconstructed as the most plausible, valid shortest paths or interpreted as mooring states. Revealing the true positions of temporarily dark vessels, or of those not using AIS at all, requires complementary data sources such as satellite imagery [4]. Nevertheless, several approaches have been proposed to detect voluntary transmission gaps in AIS data streams alone [91,92]. The same applies to identifying and removing intentionally injected but otherwise plausible false AIS messages [25–27]. Further research is needed to determine whether these detection methods can be reliably applied to community data and whether open-access AIS data can be effectively protected against spoofing attacks. Additional improvements to the analysis include incorporating AIS-B data and, in general, applying machine-learning techniques for error correction of false or patchy data. This is left for future work.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original data presented in the study and research outputs are openly available at <https://dx.doi.org/10.6084/m9.figshare.29062715>.

**Acknowledgments:** This work relied on the GeographicLib [93] and n-vector [94] software for geodesic calculations, SciPy [95], scikit-image, GDAL, QGIS; and the Basemap, Colortet, cmocean, Palletable, and Font Awesome packages

for visualization. The author would like to thank aisstream.io [44] and the network's receiver stations for providing the AIS data, and VesselFinder Ltd. [70] for the kind permission to use their gross tonnage data. Michael Dziomba, Henrik Hanssen, Marcel Strzys, Hiroki Tsuda, Thomas Zwetyenga, and the two anonymous reviewers provided valuable comments that helped to improve the quality of the manuscript.

**Conflicts of Interest:** The author declares no conflicts of interest. Author Moritz Hütten is employed by the company GRID Inc. The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. GRID Inc. approved this manuscript for publication; however, it had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

AIS	Automatic Identification System
GT	Gross Tonnage
HELCOM	Baltic Marine Environment Protection Commission (Helsinki Commission)
IMO	International Maritime Organization
MMSI	Maritime Mobile Service Identity (number)
RDP	Ramer-Douglas-Peucker (simplification algorithm)
ROI	Region Of Interest
UTC	Coordinated Universal Time

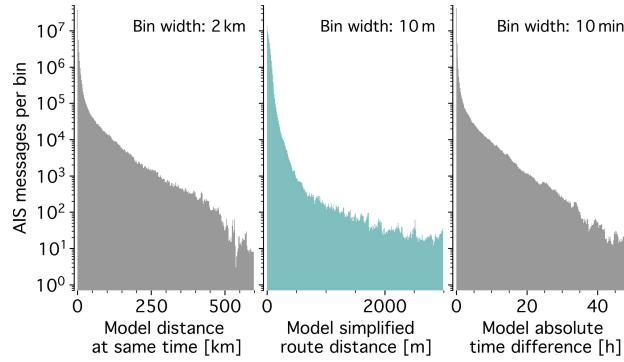
## Appendix A. Accuracy of the Trajectory Model

We validated the accuracy of the trajectory model (Sections 2.3.1 and 2.3.2 of the main paper) by comparing all 518,577 vessel movements with the positions and times of the  $5.9 \times 10^7$  cleaned AIS messages associated with these movements. Namely, we compared the following:

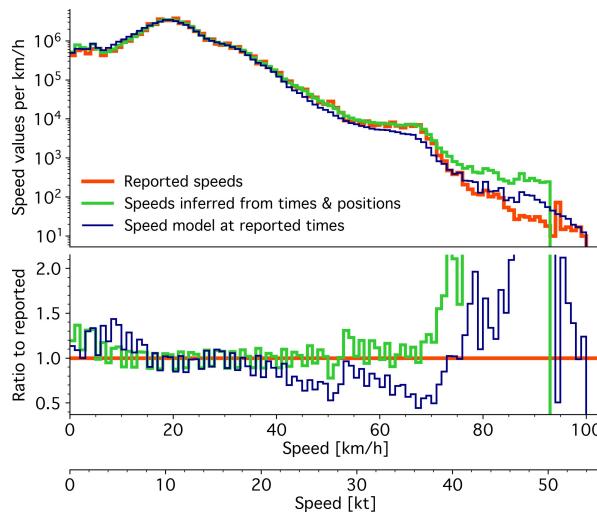
- The predicted and message positions at the same times as the original messages;
- The distance of the simplified route from the original messages (without timing information); and
- The predicted times at the positions where the simplified routes pass closest to the original AIS message positions.

These comparisons take into account that routes may pass close to the same location multiple times, e.g., when a vessel is moving in circles, by applying an adaptive time window search. Figure A1 presents these comparisons. The median distance between the trajectories' vessel positions and message positions at reported time is 585 m, with 90% of deviations being below 8 km (Figure A1, left panel). Ignoring timing information, the middle panel of Figure A1 (teal histogram) shows the impact of the route simplification by the RDP algorithm alone. The original messages' median distance from the routes is 23 m, with 3% of the distances exceeding the 100 m threshold. Large outliers occur on long-distance tracklets due to the fact that the RDP algorithm uses Euclidean geometry. The model's median time deviation is 112 s, with 90% of cases being below 25 min (Figure A1, right panel). Relatively, these figures translate into a median deviation of vessel positions of 1.2% of the route length, and also a median relative timing deviation of 1.2% of the trajectory duration.

We also compared the speed values from all position reports delivered with the AIS messages (thick red curve in Figure A2) with the speeds inferred from positions and times used to derive speed control points in the analysis (green curves), and with the speeds resulting from the trajectory model at the message times (thin blue curves). For speeds  $\lesssim 20$  kn (40 km/h), all speed descriptions align well. At higher speeds, corresponding to about 2% of all messages, the trajectory model underestimates speeds by up to 50% and overestimates the highest speeds. This is attributed to the model's inability to capture rapid speed changes and peaks (seen, e.g., around 8:00 (mile 36) in the example trajectory of Figure 7).



**Figure A1.** Comparison of original AIS message positions and times with the trajectory model. **Left panel:** Distances between message positions and vessel positions in the model at the corresponding message time. **Middle panel:** Distances between the message positions and simplified routes. **Right panel:** Time differences between message times and the times at which the trajectory model has the smallest distance to a message.



**Figure A2.** Vessel speeds from the AIS position-report messages (thick red curves), inferred from reported positions and times (green curves), and from the trajectory model at the message times (thin blue curves). The bottom panel shows the relative deviation from the reported values. The green curve shows the same data as in Figure 5c, except omitting in this figure the static reports. Also, values are grouped in bins of constant width of 1 km/h in this figure.

## Appendix B. Mean Segment Length of Lines Crossing a Rectangle Under a Fixed Angle

The mean segment length  $\overline{\Delta y}$  through an area  $A$  along an axis  $y$  is given by integration over the orthogonal axis  $x$ ,

$$\overline{\Delta y} = \frac{1}{\Delta x} \int_{x_{\min}}^{x_{\max}} \frac{dA(x)}{dx} dx, \quad (\text{A1})$$

with  $\Delta x = x_{\max} - x_{\min}$ , the extent of the rectangle in the coordinate system  $(x, y)$ . For a rectangle, the area is always

$$A = \int_{x_{\min}}^{x_{\max}} \frac{dA(x)}{dx} dx = hw, \quad (\text{A2})$$

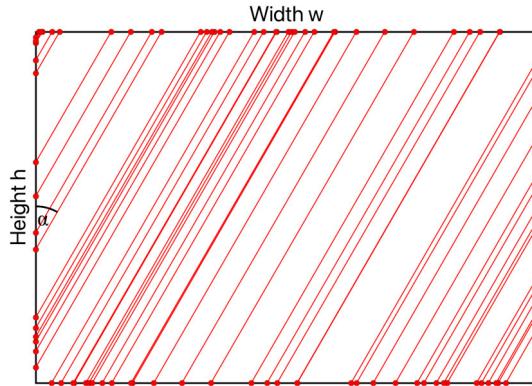
the product of height  $h$  and width  $w$ , whatever orthogonal coordinates  $x, y$  are chosen, and it is

$$\overline{\Delta y} = \frac{hw}{\Delta x} \quad (\text{A3})$$

in all rotated orthogonal coordinate systems. Rotating the coordinate system by an angle  $\alpha$  gives a projected extent  $\Delta x' = w |\cos \alpha| + h |\sin \alpha|$  of a rectangle upright in the system  $(x, y)$ . Thus, the average segment length  $\bar{\Delta}y'$  in the direction orthogonal to  $x'$ , with  $\alpha$  measured clockwise from  $y$ , the vertical edge  $h$ , is:

$$\bar{\Delta}y' = \frac{r}{r |\cos \alpha| + |\sin \alpha|} h =: \bar{d}(\alpha, r) \quad (\text{A4})$$

with  $r = w/h$ . (See also <https://math.stackexchange.com/q/3636050> (version 2020-04-21 accessed on 20 November 2025). Figure A3 shows an illustration of 100 random lines with  $\alpha = 30^\circ$ , uniformly distributed along  $x'$ .



**Figure A3.** 100 straight lines randomly intersecting at  $\alpha = 30^\circ$  a rectangle with  $r = 10/7$ . The observed mean is  $\bar{d} = 0.78 h$ , and the expected mean, Equation (A4), is  $\bar{d}(\alpha = 30^\circ, r = 10/7) = 0.82 h$ .

## Appendix C. Supplementary Tables

**Table A1.** Definition of transit areas. All areas are defined as rectangular boxes bound by lines of constant latitude and longitude. The Vänern Lake transit area is split into two separate boxes for the Nordre älv (1) and Göta älv (2) arms of the Trollhättan Canal.

Name	$\varphi_{\min}$ [°]	$\varphi_{\max}$ [°]	$\lambda_{\min}$ [°]	$\lambda_{\max}$ [°]
Skagerrak (small)	57.050	58.667	9.0	9.02
Skagerrak (default)	57.050	58.667	9.0	9.05
Skagerrak (large)	57.050	58.667	9.0	9.10
Kiel Canal (small)	54.3636	54.371	10.140	10.145
Kiel Canal (default)	54.3636	54.371	10.140	10.150
Kiel Canal (large)	54.3636	54.371	10.140	10.160
Limfjord	56.535	57.050	9.0	9.05
Oder River	53.343	53.385	14.493	14.621
Telemark	59.097	59.130	9.480	9.747
Vänern Lake 1	57.766	57.806	11.805	11.905
Vänern Lake 2	57.677	57.719	11.902	12.002
Vättern Lake	58.384	58.476	16.620	16.680
Söderåsen	59.163	59.200	17.631	17.708
Stockholm	59.2918	59.459	18.025	18.084
Saimaa Canal	60.700	60.730	28.619	28.830
Neva River	59.857	60.008	30.259	30.312

**Table A2.** Mapping of vessel type codes in the AIS-A ship static messages to the categories in which the results are presented in this work. No codes above 100 were observed in the analysis period.

Vessel Type Category	Vessel Type Codes
Passenger, high-speed	20, 23–29, 40–49, 60–69
Law enforcement, military	35, 55
Cargo	70–79
Pilot, tug, rescue, diving/dredging	21, 22, 31–34, 50–58
Tanker	80–89
Others, including fishing	all other codes or no code

**Table A3.** Average number of vessels present in the ROI ( $N$ ) and total inbound and outbound traffic through all transit areas,  $\dot{N}$ , for the default ( $df$ ) parameters and variations  $hi$  and  $low$  to suppress false-positive and false-negative transit events. The bottom rows show the relative difference to the default case.

	Vessels in ROI, $N$	Transits/Day, $\dot{N}$
Case $low$	4038.9	347.1
Default (case $df$ )	4061.1	313.1
Case $hi$	4256.2	214.6
$\delta_{low}$	−0.5%	+ 10%
$\delta_{hi}$	+4.8%	− 31%

**Table A4.** Fraction for AIS-B vessels,  $\delta_{ais-b}$ , among all vessels in the Baltic Sea. The percentages for the first five categories are estimated from the October 2024 Baltic Sea AIS data. The percentages in the vessel categories “Others including fishing” and  $GT < 10,000$  are normalized to the yearly total average using the category results from Tables 1 and 2.

Vessel Type	$\delta_{ais-b}$
All vessels (assumed yearly average)	30%
Passenger, high-speed	11%
Law enforcement, military	20%
Cargo	13%
Pilot, tug, rescue, diving/dredging	18%
Tanker	<1%
Others including fishing (derived)	55%
$GT < 10,000$ (derived)	37%
$GT \geq 10,000$ (assumed)	0%

## References

- Christiansen, M.; Hellsten, E.; Pisinger, D.; Sacramento, D.; Vilhelmsen, C. Liner shipping network design. *Eur. J. Oper. Res.* **2020**, *286*, 1–20. <https://doi.org/10.1016/j.ejor.2019.09.057>.
- United Nations Conference on Trade and Development. Review of Maritime Transport. 2024. Available online: <https://unctad.org/publication/review-maritime-transport-2024> (accessed on 24 December 2024).
- Food and Agriculture Organization of the United Nations. *The State of World Fisheries and Aquaculture 2024*; FAO: Rome, Italy, 2024. <https://doi.org/10.4060/cd0683en>.
- Paolo, F.S.; Kroodsma, D.; Raynor, J.; Hochberg, T.; Davis, P.; Cleary, J.; Marsaglia, L.; Orofino, S.; Thomas, C.; Halpin, P. Satellite mapping reveals extensive industrial activity at sea. *Nature* **2024**, *625*, 85–91. <https://doi.org/10.1038/s41586-023-06825-8>.

5. Clarke, D.; Chan, P.; Dequeljoe, M.; Kim, Y.; Barahona, S. *CO<sub>2</sub> Emissions from Global Shipping: A New Experimental Database*; OECD Statistics Working Papers; OECD Publishing: Paris, France, 2023. <https://doi.org/10.1787/bc2f7599-en>.
6. Deng, S.; Mi, Z. A review on carbon emissions of global shipping. *Mar. Dev.* **2023**, *1*, 4. <https://doi.org/10.1007/s44312-023-00001-2>.
7. Llabrés Pohl, I. The State of Shipping and Oceans. *Understanding the Impact of Global Shipping on Climate, the Ocean and Human Health*; Consultancy Report for Seas at Risk; The State of Shipping and Oceans: Brussels, Belgium, 2023. Available online: <https://seas-at-risk.org/publications/the-state-of-shipping-oceans-report/> (accessed on 28 December 2024).
8. Jägerbrand, A.K.; Brutemark, A.; Barthel Svedén, J.; Gren, I.M. A review on the environmental impacts of shipping on aquatic and nearshore ecosystems. *Sci. Total Environ.* **2019**, *695*, 133637. <https://doi.org/10.1016/j.scitotenv.2019.133637>.
9. Marappan, S.; Stokke, R.; Malinovsky, M.P.; Taylor, A. Assessment of the impacts of the offshore oil and gas industry on the marine environment. In *OSPAR 2023: The 2023 Quality Status Report for the North-East Atlantic OSPAR Commission*; OSPAR: London, UK, 2022. Available online: <https://oap.ospar.org/en/ospar-assessments/quality-status-reports/qsr-2023/other-assessments/impacts-offshore-oil-and-gas-industry/> (accessed on 24 December 2024).
10. Clovis, H.I.A.; Simon, A.M. Understanding Overfishing: A Literature Review. *Asian J. Fish. Aquat. Res.* **2024**, *26*, 61–71. <https://doi.org/10.9734/ajfar/2024/v26i1727>.
11. Eliopoulou, E.; Alissafaki, A.; Papanikolaou, A. Statistical Analysis of Accidents and Review of Safety Level of Passenger Ships. *J. Mar. Sci. Eng.* **2023**, *11*, 410. <https://doi.org/10.3390/jmse11020410>.
12. Allianz Commercial. Safety and Shipping Review 2024: An Annual Review of Trends and Developments in Shipping Losses and Safety. 2024. Available online: <https://commercial.allianz.com/news-and-insights/reports/shipping-safety.html> (accessed on 24 December 2024).
13. Bueger, C.; Edmunds, T.; Stockbruegger, J. *Securing the Seas. A Comprehensive Assessment of Global Maritime Security*; United Nations Institute for Disarmament Research: Geneva, Switzerland, 2024. Available online: <https://unidir.org/publication/securing-the-seas-a-comprehensive-assessment-of-global-maritime-security/> (accessed on 24 December 2024).
14. Shepperson, J.L.; Hintzen, N.T.; Szostek, C.L.; Bell, E.; Murray, L.G.; Kaiser, M.J. A comparison of VMS and AIS data: The effect of data coverage and vessel position recording frequency on estimates of fishing footprints. *ICES J. Mar. Sci.* **2018**, *75*, 988–998. <https://doi.org/10.1093/icesjms/fsx230>.
15. IMO. *International Convention for the Safety of Life at Sea (SOLAS)*; United Nations International Maritime Organization: London, UK, 1974; Chapter V, 19, 2.4.
16. Natale, F.; Gibin, M.; Alessandrini, A.; Vespe, M.; Paulrud, A. Mapping Fishing Effort through AIS Data. *PLoS ONE* **2015**, *10*, e0130746. <https://doi.org/10.1371/journal.pone.0130746>.
17. Kroodsma, D.A.; Mayorga, J.; Hochberg, T.; Miller, N.A.; Boerder, K.; Ferretti, F.; Wilson, A.; Bergman, B.; White, T.D.; Block, B.A.; et al. Tracking the global footprint of fisheries. *Science* **2018**, *359*, 904–908. <https://doi.org/10.1126/science.aoa5646>.
18. Cerdeiro, D.; Komaromi, A.; Liu, Y.; Saeed, M. *World Seaborne Trade in Real Time: A Proof of Concept for Building AIS-based Nowcasts from Scratch*; IMF Working Papers; International Monetary Fund: Washington, DC, USA, 2020; 20/57. <https://doi.org/10.5089/9781513544106.001>.
19. Yan, Z.; Xiao, Y.; Cheng, L.; He, R.; Ruan, X.; Zhou, X.; Li, M.; Bin, R. Exploring AIS data for intelligent maritime routes extraction. *Appl. Ocean Res.* **2020**, *101*, 102271. <https://doi.org/10.1016/j.apor.2020.102271>.
20. Visky, G.; Rohl, A.; Katsikas, S.; Maennel, O. AIS Data Analysis: Reality in the Sea of Echoes. In Proceedings of the 2024 IEEE 49th Conference on Local Computer Networks (LCN), Normandy, France, 8–10 October 2024. <https://doi.org/10.1109/lcn60385.2024.10639765>.
21. Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E.; Huang, G.B. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1559–1582. <https://doi.org/10.1109/TITS.2017.2724551>.
22. Hadjipieris, A.; Dimitriou, N.; Arandjelović, O. Unsupervised Port Berth Localization from Automatic Identification System Data. *Sensors* **2025**, *25*, 6845. <https://doi.org/10.3390/s25226845>.

23. Chen, S.; Zhao, R.; Yang, J.B.; Huang, Y. AISStream-MCP: A Real-Time Memory-Augmented Question-Answering System for Maritime Operations. *J. Mar. Sci. Eng.* **2025**, *13*, 1754. <https://doi.org/10.3390/jmse13091754>.
24. Emmens, T.; Amrit, C.; Abdi, A.; Ghosh, M. The promises and perils of Automatic Identification System data. *Expert Syst. Appl.* **2021**, *178*, 114975. <https://doi.org/10.1016/j.eswa.2021.114975>.
25. Androjna, A.; Pavić, I.; Gucma, L.; Vidmar, P.; Perkovič, M. AIS Data Manipulation in the Illicit Global Oil Trade. *J. Mar. Sci. Eng.* **2023**, *12*, 6. <https://doi.org/10.3390/jmse12010006>.
26. Kessler, G.C.; Zorri, D.M. AIS Spoofing: A Tutorial for Researchers. In Proceedings of the 49th Conference on Local Computer Networks (LCN), Normandy, France, 8–10 October 2024; pp. 1–7. <https://doi.org/10.1109/LCN60385.2024.10639747>.
27. Louart, M.; Szkolnik, J.J.; Boudraa, A.O.; Le Lann, J.C.; Le Roy, F. An approach to detect identity spoofing in AIS messages. *Expert Syst. Appl.* **2024**, *252*, 124257. <https://doi.org/10.1016/j.eswa.2024.124257>.
28. Jalkanen, J.P.; Johansson, L.; Kukkonen, J. A Comprehensive Inventory of the Ship Traffic Exhaust Emissions in the Baltic Sea from 2006 to 2009. *AMBIO* **2014**, *43*, 311–324. <https://doi.org/10.1007/s13280-013-0389-3>.
29. Löptien, U.; Axell, L. Ice and AIS: Ship speed data and sea ice forecasts in the Baltic Sea. *Cryosphere* **2014**, *8*, 2409–2418. <https://doi.org/10.5194/tc-8-2409-2014>.
30. Goerlandt, F.; Goite, H.; Valdez Banda, O.A.; Höglund, A.; Ahonen-Rainio, P.; Lensu, M. An analysis of wintertime navigational accidents in the Northern Baltic Sea. *Saf. Sci.* **2017**, *92*, 66–84. <https://doi.org/10.1016/j.ssci.2016.09.011>.
31. Lensu, M.; Goerlandt, F. Big maritime data for the Baltic Sea with a focus on the winter navigation system. *Mar. Policy* **2019**, *104*, 53–65. <https://doi.org/10.1016/j.marpol.2019.02.038>.
32. Fernandez Arguedas, V.; Pallotta, G.; Vespe, M. Maritime Traffic Networks: From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 722–732. <https://doi.org/10.1109/TITS.2017.2699635>.
33. HELCOM. *HELCOM Assessment on Maritime Activities in the Baltic Sea 2018*; Baltic Sea Environment Proceedings 152; Helsinki Commission: Helsinki, Finland, 2018. Available online: [https://helcom.fi/post\\_type\\_publ/bsep152/](https://helcom.fi/post_type_publ/bsep152/) (accessed on 27 July 2024).
34. Kulkarni, K.; Li, F.; Liu, C.; Musharraf, M.; Kujala, P. System-Level Simulation of Maritime Traffic in Northern Baltic Sea. In Proceedings of the 2022 Winter Simulation Conference (WSC), Singapore, 11–14 December 2022; pp. 1923–1934. <https://doi.org/10.1109/WSC57314.2022.10015257>.
35. Dettner, F.; Hilpert, S. Emission Inventory for Maritime Shipping Emissions in the North and Baltic Sea. *Data* **2023**, *8*, 85. <https://doi.org/10.3390/data8050085>.
36. Caprile, A.; Leclerc, G. Russia’s ‘Shadow Fleet’: Bringing the Threat to Light. European Parliamentary Research Service. 2024; PE 766.242. Available online: [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI\(2024\)766242](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)766242) (accessed on 2 February 2025).
37. Serry, A. Dynamics of maritime transport in the Baltic Sea: Regionalisation and multimodal integration. In Proceedings of the 6th Maritime Transport Conference 2014, Barcelona, Spain, 25–27 June 2014. Available online: <https://www.researchgate.net/publication/263653508> (accessed on 3 February 2025).
38. Posti, A.; Häkkinen, J.M. *Survey of Transportation of Liquid Bulk Chemicals in the Baltic Sea*; The Centre for Maritime Studies, University of Turku: Turku, Finland, 2012; A 60. Available online: <https://www.researchgate.net/publication/264889915> (accessed on 3 February 2025).
39. Boteler, B.; Tröltzsch, J.; Abhol, K. *Drivers for the Shipping Sector*; Policy for Sustainable Shipping in the Baltic Sea Region (SHEBA) project; BONUS Secretariat (EEIG): Brussels, Belgium, 2015. Available online: <https://www.ecologic.eu/13804> (accessed on 10 February 2025).
40. Statista. Volume of Cargo Traffic in Baltic Sea Ports in Central and Eastern Europe from 2015 to 2019, by cargo type, Available online: <https://www.statista.com/statistics/1183151> (accessed on 20 November 2025).
41. Matczak, M.; Meyer, N.; Ooms, E.; Schröder, L.; Vološina, M.; Warmelink, H.; Zaucha, J.; Czermański, E.; Coornaert, C.; Koch, A. Quo Vadis—Exploring the future of shipping in the Baltic Sea. In *European Maritime Spatial Planning Platform*; European Commission: Brussels, Belgium, 2018. Available online: <https://maritime-spatial-planning.ec.europa.eu/practices/quo-vadis-exploring-future-shipping-baltic-sea> (accessed on 10 February 2025).
42. Friman, E. *Transport and Logistics in the Baltic Sea Region by 2030 A Foresight Study*; Turku School of Economics, University of Turku: Turku, Finland, 2019. <https://doi.org/10.13140/RG.2.2.31747.66084>.

43. HELCOM. *Convention on the Protection of the Marine Environment of the Baltic Sea Area*; Helsinki Commission: Helsinki, Finland, 1992. Available online: <https://helcom.fi/about-us/convention/> (accessed on 20 November 2025).
44. aisstream.io. Real-Time Data Stream WebSocket API. 2025. Available online: <https://aisstream.io/> (accessed on 30 May 2025).
45. CMEMS. *Global Ocean Hourly Sea Surface Wind and Stress from Scatterometer and Model*. E.U. Copernicus Marine Service Information (CMEMS); Marine Data Store; Mercator Ocean International: Toulouse, France, 2024. <https://doi.org/10.48670/moi-00305>.
46. NCEP GFS. *0.25 Degree Global Forecast System Historical Archive*; NSF National Center for Atmospheric Research: Boulder, Colorado, 2015. <https://doi.org/10.5065/D65D8PKW>.
47. Mardia, K.V.; Jupp, P.E. *Directional Statistics*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 1999; pp. 2013–2014. <https://doi.org/10.1002/9780470316979>.
48. Zhao, L.; Shi, G.; Yang, J. Ship Trajectories Pre-processing Based on AIS Data. *J. Navig.* **2018**, *71*, 1210–1230. <https://doi.org/10.1017/S0373463318000188>.
49. Wolsing, K.; Roepert, L.; Bauer, J.; Wehrle, K. Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches. *J. Mar. Sci. Eng.* **2022**, *10*, 112. <https://doi.org/10.3390/jmse10010112>.
50. Lv, T.; Tang, P.; Zhang, J. A Real-Time AIS Data Cleaning and Indicator Analysis Algorithm Based on Stream Computing. *Sci. Program.* **2023**, *8345603*. <https://doi.org/10.1155/2023/8345603>.
51. Asian Development Bank. *Methodological Framework for Unlocking Maritime Insights Using Automatic Identification System Data: A Special Supplement of Key Indicators for Asia and the Pacific 2023*; Asian Development Bank: Mandaluyong City, Philippines, 2023; pp. 1–90. <https://doi.org/10.22617/FLS230401-2>.
52. Zhang, W.; Jiang, W.; Liu, Q.; Wang, W. AIS data repair model based on generative adversarial network. *Reliab. Eng. Syst. Saf.* **2023**, *240*, 109572. <https://doi.org/10.1016/j.ress.2023.109572>.
53. Yang, Y.; Liu, Y.; Li, G.; Zhang, Z.; Liu, Y. Harnessing the power of Machine learning for AIS Data-Driven maritime Research: A comprehensive review. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *183*, 103426. <https://doi.org/10.1016/j.tre.2024.103426>.
54. NCEIETOPO. *15 Arc-Second Global Relief Model*; NOAA National Centers for Environmental Information: Asheville, NC, USA, 2022. <https://doi.org/10.25921/fd45-g74>.
55. Van Der Wielen, N.; McGurk, J.; Barrett, L. Optimising port arrival statistics: Enhancing timeliness through Automatic Identification System (AIS) data. *Stat. J. IAOS* **2024**, *40*, 421–434. <https://doi.org/10.3233/SJI-230100>.
56. Wijaya, W.M.; Nakamura, Y. Port performance indicators construction based on the AIS-generated trajectory segmentation and classification. *Int. J. Data Sci. Anal.* **2024**, *20*, 2473–2492. <https://doi.org/10.1007/s41060-024-00614-w>.
57. Iphar, C.; Le Berre, I.; Foulquier, É.; Napoli, A. Port call extraction from vessel location data for characterising harbour traffic. *Ocean Eng.* **2024**, *293*, 116771. <https://doi.org/10.1016/j.oceaneng.2024.116771>.
58. Guo, Z.; Qiang, H.; Xie, S.; Peng, X. Unsupervised knowledge discovery framework: From AIS data processing to maritime traffic networks generating. *Appl. Ocean Res.* **2024**, *146*, 103924. <https://doi.org/10.1016/j.apor.2024.103924>.
59. Qiang, H.; Guo, Z.; Peng, X.; Jia, C. FDBR: Ultra-fast and data-efficient behavior recognition of port vessels using a statistical framework. *Ocean Eng.* **2025**, *315*, 119737. <https://doi.org/10.1016/j.oceaneng.2024.119737>.
60. Li, G.; Liu, M.; Zhang, X.; Wang, C.; Lai, K.H.; Qian, W. Semantic Recognition of Ship Motion Patterns Entering and Leaving Port Based on Topic Model. *J. Mar. Sci. Eng.* **2022**, *10*, 2012. <https://doi.org/10.3390/jmse10122012>.
61. Yun, L.; Bliault, A. *High Performance Marine Vessels*; Springer US: Boston, MA, USA 2012. <https://doi.org/10.1007/978-1-4614-0869-7>.
62. Ramer, U. An iterative procedure for the polygonal approximation of plane curves. *Comput. Graph. Image Process.* **1972**, *1*, 244–256. [https://doi.org/10.1016/S0146-664X\(72\)80017-0](https://doi.org/10.1016/S0146-664X(72)80017-0).
63. Douglas, D.H.; Peucker, T.K. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *Cartographica* **1973**, *10*, 112–122. <https://doi.org/10.3138/FM57-6770-U75U-7727>.

64. Pallero, J. Robust line simplification on the surface of the sphere. *Comput. Geosci.* **2015**, *83*, 146–152. <https://doi.org/10.1016/j.cageo.2015.07.011>.
65. Tang, C.; Wang, H.; Zhao, J.; Tang, Y.; Yan, H.; Xiao, Y. A method for compressing AIS trajectory data based on the adaptive-threshold Douglas-Peucker algorithm. *Ocean Eng.* **2021**, *232*, 109041. <https://doi.org/10.1016/j.oceaneng.2021.109041>.
66. Mehri, S.; Hooshangi, N.; Mahdizadeh Gharakhanlou, N. A Novel Context-Aware Douglas-Peucker (CADP) Trajectory Compression Method. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 58. <https://doi.org/10.3390/ijgi14020058>.
67. Onyango, S.O.; Owiredu, S.A.; Kim, K.I.; Yoo, S.L. A Quasi-Intelligent Maritime Route Extraction from AIS Data. *Sensors* **2022**, *22*, 8639. <https://doi.org/10.3390/s22228639>.
68. Zhang, T.; Wang, Z.; Wang, P. A method for compressing AIS trajectory based on the adaptive core threshold difference Douglas-Peucker algorithm. *Sci. Rep.* **2024**, *14*, 21408. <https://doi.org/10.1038/s41598-024-71779-4>.
69. Hügel, S. Simplification. v0.7.13. 2021. <https://doi.org/10.5281/zenodo.5774852>.
70. VesselFinder Ltd. (Sliven, Bulgaria). Real-Time and Historical AIS Data. 2025. Available online: <https://www.vesselfinder.com> (accessed on 12 February 2025).
71. Liu, Z.; Gao, H.; Zhang, M.; Yan, R.; Liu, J. A data mining method to extract traffic network for maritime transport management. *Ocean Coast. Manag.* **2023**, *239*, 106622. <https://doi.org/10.1016/j.ocemoaman.2023.106622>.
72. Millefiori, L.M.; Zissis, D.; Cazzanti, L.; Arcieri, G. A distributed approach to estimating sea port operational regions from lots of AIS data. In Proceedings of the 2016 IEEE International Conference on Big Data, Big Data 2016, Washington DC, USA, 5–8 December 2016; pp. 1627–1632. <https://doi.org/10.1109/BigData.2016.7840774>.
73. Yan, Z.; Cheng, L.; He, R.; Yang, H. Extracting ship stopping information from AIS data. *Ocean Eng.* **2022**, *250*, 111004. <https://doi.org/10.1016/j.oceaneng.2022.111004>.
74. van der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T.; the scikit-image contributors. scikit-image: Image processing in Python. *PeerJ* **2014**, *2*, e453. <https://doi.org/10.7717/peerj.453>.
75. Pallotta, G.; Vespe, M.; Bryan, K. Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction. *Entropy* **2013**, *15*, 2218–2245. <https://doi.org/10.3390/e15062218>.
76. Storgard, J.; Lappalainen, J.; Wahlstrom, I.; Kajander, S. *Scenarios for the Development of Maritime Safety and Security in the Baltic Sea Region*; The Centre For Maritime Studies, University of Turku: Turku, Finland, 2012. Available online: <https://www.dma.dk/Media/637743958502717512/Scenarios%20for%20the%20Development%20of%20Maritime%20Safety%20and%20Security%20in%20the%20Baltic%20Sea%20Region.pdf> (accessed on 20 November 2025).
77. Czermański, E. Baltic Shipping Development in Maritime Spatial Planning Aspect. *Studia I Mater. Inst. Transp. I Handlu Morskiego* **2017**, *14*, 48–64. Available online: <https://www.researchgate.net/publication/322815667> (accessed on 10 February 2025).
78. HELCOM. Fishing Intensity and Effort 2016–2021 (VMS Data). HELCOM Map and Data Service. (version 2023-05-31). 2023. Available online: [https://maps.helcom.fi/website/mapservice/?datasetID=human\\_activities162](https://maps.helcom.fi/website/mapservice/?datasetID=human_activities162) (accessed on 21 February 2025).
79. Heitmann, N.; Rehdanz, K.; Schmidt, U. Determining optimal transit charges: The Kiel Canal in Germany. *J. Transp. Geogr.* **2013**, *26*, 29–42. <https://doi.org/10.1016/j.jtrangeo.2012.08.005>.
80. Statista. Total gross tonnage of ships in the Kiel Canal from 1996 to 2018, 2019. Available online: <https://www.statista.com/statistics/270171> (accessed on 13 January 2025).
81. Statista. Number of ships in the Kiel Canal from 2007 to 2024, 2025. Available online: <https://www.statista.com/statistics/270173> (accessed on 13 January 2025).
82. GDWS. Der Nord-Ostsee-Kanal. German Federal Waterways and Shipping Administration. 2024. Available online: [https://www.gdws.wsv.bund.de/SharedDocs/Downloads/DE/Publikationen/\\_GDWS/Wasserstrassen/NOK.pdf](https://www.gdws.wsv.bund.de/SharedDocs/Downloads/DE/Publikationen/_GDWS/Wasserstrassen/NOK.pdf) (accessed on 13 January 2025).
83. GDWS. Press releases of the German Federal Waterways and Shipping Administration. 2025. Available online: <https://www.gdws.wsv.bund.de/DE/presse/pressemitteilungen/pressemitteilungen-node.html> (accessed on 13 January 2025).
84. HELCOM. AIS Shipping Density 2011–2024, All Vessel Types. HELCOM Map and Data Service. (version 2025-11-19). 2025. Available online: <https://maps.helcom.fi/website/mapservice/?datasetID=shipping225> (accessed on 21 November 2025).

85. World Port Index. US National Geospatial-Intelligence Agency. 2025. Available online: <https://msi.nga.mil/Publications/WPI> (accessed on 27 March 2025).
86. Arslanalp, S.; Koepke, R.; Verschuur, J. Tracking Trade from Space: An Application to Pacific Island Countries. IMF Working Papers. 2021. 21/225. Available online: <https://portwatch.imf.org> (accessed on 10 April 2025).
87. Synak-Miłosz, E.; Rozmarynowska-Mrozek, M. Results of Top 10 Baltic Ports in the First Half of 2023. Actia Forum. Port Monitor. 2023. Available online: [https://actiaforum.pl/wp-content/uploads/2023/10/BalticPortsH12023\\_total\\_containers.pdf](https://actiaforum.pl/wp-content/uploads/2023/10/BalticPortsH12023_total_containers.pdf) (accessed on 20 November 2025).
88. Harati-Mokhtari, A.; Wall, A.; Brooks, P.; Wang, J. Automatic identification system (AIS): Data reliability and human error implications. *J. Navig.* **2007**, *60*, 373–389. <https://doi.org/10.1017/S0373463307004298>.
89. Sun, R.; Abouarghoub, W.; Demir, E. Enhancing data quality in maritime transportation: A practical method for imputing missing ship static data. *Ocean Eng.* **2025**, *315*, 119722. <https://doi.org/10.1016/j.oceaneng.2024.119722>.
90. Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* **2011**, *40*, 1–29. <https://doi.org/10.18637/jss.v040.i01>.
91. Sharma, A.; Ghosh, S.; Shekhar, S. Physics-Based Abnormal Trajectory Gap Detection. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 106. <https://doi.org/10.1145/3673235>.
92. Bernabe, P.; Gotlieb, A.; Legeard, B.; Marijan, D.; Sem-Jacobsen, F.O.; Spieker, H. Detecting Intentional AIS Shutdown in Open Sea Maritime Surveillance Using Self-Supervised Deep Learning. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 1166–1177. <https://doi.org/10.1109/TITS.2023.3322690>.
93. Karney, C.F.F. Algorithms for geodesics. *J. Geod.* **2013**, *87*, 43–55. <https://doi.org/10.1007/s00190-012-0578-z>.
94. Gade, K. A Non-singular Horizontal Position Representation. *J. Navig.* **2010**, *63*, 395–417. <https://doi.org/10.1017/S0373463309990415>.
95. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.