



RESEARCH ARTICLE

# AIS-based near-collision database generation and analysis of real collision avoidance manoeuvres

Arnstein Vestre,<sup>1\*</sup>  Azzeddine Bakdi,<sup>1</sup> Erik Vanem,<sup>1,2</sup> and Øystein Engelhardtson<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Oslo, 0851 Oslo, Norway.

<sup>2</sup>DNV, Veritasveien 1, Høvik, N-1363, Norway.

\*Corresponding author. E-mail: [arnsteinvestre@gmail.com](mailto:arnsteinvestre@gmail.com)

**Received:** 17 July 2020; **Accepted:** 18 March 2021; **First published online:** 25 May 2021

**Keywords:** automatic identification system (AIS), autonomous navigation, closest point of approach (CPA), COLREGs, maritime safety

## Abstract

Economic and technological development has increased the amount, density and complexity of maritime traffic, which has resulted in new challenges. One challenge is conforming to the distinct evasion manoeuvres required by vessels entering into near-collision situations (NCSs). Existing rules are vague and do not precisely dictate which, when and how collision avoidance manoeuvres (CAMs) should be executed. The automatic identification system (AIS) is widely used for vessel monitoring and traffic control. This paper presents an efficient, scalable method for processing large-scale raw AIS data using the closest point of approach (CPA) framework. NCSs are identified to create a database of historical traffic data. Important features describing CAMs are defined, estimated and analysed. Applications on a high-quality real-world data set show promising results for a subset of the identified situations. Future applications may play a significant role in the maritime regulatory framework, navigation protocol compliance evaluation, risk assessment, automatic collision avoidance, and algorithm design and testing for autonomous vessels.

## 1. Introduction

Maritime safety is of utmost importance, but real-world collision data are scarce. This prompts the researcher to ask what can be learned from situations where collision is imminent, but an adverse outcome avoided. Large-scale automatic identification system (AIS) data sets keep track of many such situations, and may be used to create a database of such near-collision situations (NCSs). This may further enable analysis and new insights for researchers and maritime safety practitioners.

With the rising importance of maritime traffic, collision avoidance has already developed as one of the most important concerns for maritime safety (Ozturk and Cicek, 2019). With the emergence of autonomous surface vessels, there is also an emerging need for developing technical standards and methods for autonomous collision avoidance. The pre-eminent role of maritime transport, as well as the continuous strive for safety improvements, has recently led to the development of several theoretical and practical approaches to ensuring higher levels of safety and efficiency in maritime navigation. These methods allow for extensions to AIS applications.

One of these directions focuses on the concept of smart navigation, such as decision support systems based on automatic radar plotting aids as presented by Ozoga and Montewka (2018), AIS-based maritime spatial planning (Le Tixerant et al., 2018) and route planning (Jeong et al., 2019). Furthermore, new methodology has recently been introduced for autonomous navigation (Naeem et al., 2016) and

**Table 1.** Main aspects of a CAM.

| Situation          | Time                 | Space                    | Navigation          |
|--------------------|----------------------|--------------------------|---------------------|
| COLREGs rule       | Manoeuvre start time | Vessel distance at start | Total speed change  |
| Governance         | Situation end time   | Relative speed at start  |                     |
| Manoeuvring vessel | Time to CPA          | Relative course at start | Total course change |
| Type of manoeuvre  | Evacuation time      | Distance to CPA at start |                     |
|                    |                      | Passing distance         |                     |

e-navigation (Kim et al., 2014). A second direction focuses on the concept of navigation safety assessment, as explained by Wang et al. (2019), safety control (Wu et al., 2015), maritime navigation risk identification (Liu et al., 2019), analysis (Chen et al., 2019) and assessment (Ozturk and Cicek, 2019).

Despite the wide range of research methods in the field, few studies address real-world NCSs in a context which may be generalised, and few studies apply robustly identified parameters in their analysis. The proposed frameworks require design, simulation and testing to be applicable and reliable in practice. Hence, identifying aspects of navigation which may be estimated empirically lends itself to extending such methods.

Key aspects of smart navigation, and associated safety analysis, include, among others, collision diameter (Altan, 2019), safety distance (Szlupczynski et al., 2018) and safety regions (Szlupczynski and Szlupczynska, 2017). Szlupczynski and Szlupczynska (2016) identified several collision risk parameters for such situations. Another approach, such as that taken by Zaman et al. (2015), involves using AIS data in concert with the formal safety assessment (FSA) methodology. This approach may be extended to reviewing the choice of risk control options (RCOs) from real-world situations. Existing maritime traffic models are differently described in the literature, and the proposed parameters are not unique (Zhou et al., 2019).

The present work is motivated by the shortcomings of the existing methods, and seek to overcome these by developing a framework for identifying real-world NCSs from high-resolution AIS data. For the methodology to be widely applicable, it needs to be scalable and generalisable, both geographically and over time. An NCS, being distinctly mapped by high-resolution AIS data, may be used to accurately analyse manoeuvre aspects which are relevant for collision avoidance.

Table 1 lists the important aspects describing a collision avoidance manoeuvre (CAM), in the case of an NCS. The interaction between these aspects, and the guidance presented to sailors, is not strictly codified in the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs), which leaves interpretation open to the seafarer.

Rule 16 of COLREGs indicates that the ‘give-way’ vessel shall ‘*so far as possible take early and substantial action*’ (the authors’ emphasis). COLREGs rule 8 requires that ‘any alteration of *course and/or speed* to avoid collision shall, if the circumstances of the case admit, be *large enough* to be readily apparent to the other vessel’. According to rule 17, the stand-on vessel ‘*may take action* to avoid collision’ as soon as it becomes apparent that the give-way vessel ‘is not taking *appropriate action in compliance with these rules*’. It is first when the stand-on vessel ‘finds herself so close that collision cannot be avoided (. . .)’ that the stand-on vessel is put under strict requirements: ‘she *shall* take such action as will best aid to avoid collision’.

In designing and creating an NCS database, estimation and analysis of the listed aspects is made possible, with the aim of describing the de facto interpretation of, and vessel conduct under, different COLREGs rules at sea, such as analysing how early and how substantial actions taken are, and how large course and speed alterations are in observed CAMs.

Historical AIS data are characterised by noise, both in the sampling sense, and in the sense that situations with NCS characteristics may in fact be benign in nature. The primary task in the creation of an NCS database is the detection and identification of credible NCSs.

There exists a body of literature on NCS detection, which in particular relates to real-time applications with the goal of providing increased warning time for sailors for evasion manoeuvres at sea. One method for vessel collision-candidate detection was proposed by Chen et al. (2018), and uses a temporally discrete nonlinear velocity obstacle algorithm, modelling the vessel encounter as a process. Another method proposed by Zhang et al. (2015) describes near-miss collision detection through the use of vessel conflict ranking operators.<sup>1</sup> This method was further improved by Zhang et al. (2016, 2017). The literature also encompasses methods based on simulation models, such as those described by Fang et al. (2018).

The primary drawbacks of the existing NCS detection methods is the lack of automation and the fact that they are not computationally efficient enough to be extended to large-scale data sets. We introduce a framework for handling large AIS data in a computationally feasible way, which allows both for large-scale extension and automation.

An NCS is defined by the vessel pair, which includes the position, pose and speed of the vessels. The method further indicates the risk of collision measured at each timestamp, and allows for inclusion of further characteristics in the analysis, such as vessel type, size and dimensions.

The aggregate analysis at first proceeds by filtering the identified situations in terms of severity. The observed AIS transmissions from the filtered situations form the basis for analysing various aspects of CAM patterns, as presented in Table 1. Second, CAMs must be distinguished from normal route-following and track-keeping actions imposed by traffic separation schemes and natural obstacles. The proposed framework uses speed and course patterns, as well as vessel position, to incorporate the closest point of approach (CPA) algorithm, as described by Sang et al. (2016). This determines an estimated time to CPA and an estimated distance at CPA for every vessel at every timestamp.

The third challenge is detecting critical timestamps during a particular NCS that correspond to when the CAM is initiated and when the NCS is resolved. The change-point detection is difficult owing to the noisy environment and the presence of other route-following actions, which lead to the vessel speed and course to seldom be in a steady state. Other aspects are used to analyse how CAMs are conducted in a practical sense in light of COLREGs rules, vessel characteristics (dimensions, type and pose), which vessel executed a manoeuvre and what manoeuvre was executed. For situations where CAMs are detected, the manoeuvres are described by the magnitude of the steady-state speed and course alteration, relative distance, speed and course at manoeuvre start, as well as the actual passing distance.

Finally, aspects are estimated for all situations in the NCS database, and presented with empirical distributions under various COLREGs conditions, for the subset of vessels in unrestricted waters. The obtained statistics are sensitive to noise in the NCS database as well as the accuracy of the aspect estimators. However, the obtained results have potential applications in future work, such as informing selection of safety limits for risk analysis, ensuring proper tuning of parameters in automatic collision avoidance rule design and as reference benchmarks for comparing and evaluating protocol compliance and safety performance of autonomous vessels.

The rest of this paper is structured as follows. Section 2 presents an overview of the high-resolution AIS data source and presents the raw data that are used in this work. Section 3 presents the suggested framework for identifying NCSs, and discusses situation filtering and manoeuvre identification, as well as outlining various CAM aspects. Section 4 presents examples of situations retrieved by the framework and descriptive statistics for a subset of the identified situations, and discusses challenges and possible applications. The paper concludes in Section 5 with relevant areas for future research and extensions.

## ABBREVIATIONS:

AIS            Automatic identification system

<sup>1</sup> The concept of *near miss* is closely related to *near collision* in that *near collision* is a concept nested within *near miss*. *Near collision* denotes a situation in which two vessels are on track for collision, but avoid an adverse outcome. A *near miss* can be any situation where an adverse outcome is imminent, but avoided, such as a near-grounding event. Both concepts are different from a *collision* in that the adverse outcome is avoided.

|         |  |
|---------|--|
| CAM     | Collision avoidance manoeuvre  |
| COLREGs | Convention on the International Regulations for Preventing Collisions at Sea |
| COG     | Course over ground   |
| CPA     | Closest point of approach  |
| LAT/LON | Latitude/Longitude   |
| NB      | Norwegian baseline   |
| NCS     | Near-collision situation   |
| SOG     | Speed over ground  |

## 2. Data

The present paper uses a high-quality traffic data-source consisting of AIS transmissions from 13 days covering the Norwegian exclusive economic zone (EEZ). AIS data dimensionality varies, but most sources provide traffic data for speed over ground (SOG), course over ground (COG), position (latitude, longitude), as well as unique vessel identifiers. The present data source additionally provides information on vessel type and dimensions. A typical AIS data set of the present quality and geographical coverage contains 18.5 million registered AIS transmissions per day, with a mean temporal resolution of 9.6 s. Rates of transmission vary with vessel speed and degree of course change, allowing for particularly high resolution when vessels are in manoeuvre. This temporal resolution further allows for detailed analysis of the aspects describing manoeuvring patterns, whereas the spatial coverage allows for analysing the manoeuvres over various traffic zones.

The data source has low rates of erroneous or missing records, at below 0.1% for SOG and COG. Approximately 6.0% of records are erroneously registered outside the area of analysis. To focus analysis on commercial vessels, records for which vessel category was not identified, or where vessels were identified as non-commercial, were not considered. In addition, tugs and pilot vessels were not considered, as such vessels by nature engage in close-quarters situations which are difficult to discern from NCSs. The ensuing typical filtered data set consists of 10 million daily records, which entails a sizeable number of vessel-to-vessel interactions to be analysed at each timestamp.

Using AIS data for analysis with uniquely identified vessels further allows for combination with other data sets, such as weather or vessel characteristics. The present paper focuses on suggesting a framework and highlighting some selected manoeuvre aspects, but the versatility of AIS data lends itself to numerous extensions.

### 2.1. A note on computational resources

The algorithm which identifies NCSs, as presented in Section 3, involves calculation of identifying parameters for all close-vessel interactions. Some parameters in the data allow for tuning the balance between temporal resolution and computational efficiency.

The present data source recorded approximately 90% of vessels at least once within each 20 s window, with increased resolution during manoeuvres. The NCS identification algorithm starts by observing all candidate situations once within a moving time window of 20 s, as a substitute for computing on the full-resolution data. In terms of the NCS identification algorithm, this reduces the number of observations by 50% – greatly reducing computational cost – while retaining a satisfactory temporal resolution. The decrease in temporal resolution significantly increases computational efficiency, but does not have a negative effect on the process of identification. The algorithm for pattern analysis of CAMs – based on the list of identified situations – uses the full-resolution data.

Furthermore, as NCSs by definition are confined to vessels within the range of sight of one other, computations are restricted to pairs of vessels within a  $2 \times 2$  km<sup>2</sup> square. Computing distances by such a (box) metric is computationally more efficient than calculating the Euclidean or spheric distance, and

as it is a filtering mechanism for further computations, it does not skew the database, as long as the threshold is set sufficiently high.

The suggested framework is implemented in Python, using the Pandas, Numpy and Numba libraries. Computations scale linearly in time and may be computed in parallel. The computations for the present paper have been run on a MacBook Pro (13-inch, 2016) with a 2 GHz Intel Core i5 processor and 8 GB memory.

### 3. Methodology

The following discusses the identification of real NCSs for an NCS database based on historical traffic data from AIS. The CPA framework is established by Sang et al. (2016) as a method for analysing the collision behaviour of two objects in motion. CPA is defined as the closest point two objects will arrive at if speed and course are unaltered. Distance between vessels at CPA (DCPA) indicates the severity of the hypothetical situation. Time to CPA (TCPA) is the remaining time for the two objects to reach CPA at constant speed and course. Negative TCPA indicates objects moving away from each other. An NCS is defined as a situation where two vessels will, in the near future, come within an unsafe distance of each other. NCSs are identified for each candidate situation by an unsafe DCPA – below a set threshold – and a limited positive TCPA.

AIS data provide positions, speed and course for a vessel  $i$  at timestamp  $t_0$ , and hence provides the information

$$t_0; \quad (x_i(t_0), y_i(t_0)) = (\text{LON}(t_0), \text{LAT}(t_0)); \quad \text{SOG}_i(t_0); \quad \text{COG}_i(t_0)$$

where  $(x_i(t_0), y_i(t_0))$  is given in decimal degrees,  $\text{SOG}_i(t_0)$  in knots and  $\text{COG}_i(t_0)$  in degrees. The velocity vector is given by

$$V_i = \begin{bmatrix} V_{i,x} \\ V_{i,y} \end{bmatrix} = \begin{bmatrix} \text{SOG}_i \sin(\text{COG}_i) \\ \text{SOG}_i \cos(\text{COG}_i) \end{bmatrix} \times \frac{1,852}{3,600} \tag{1}$$

where all variables are evaluated at  $t_0$ . The  $\text{TCPA}_{i,j}(t_0)$  can be determined for a pair of close vessels  $(i, j)$  at time  $t_0$  based on their relative velocity  $\Delta V_{i,j}(t_0)$  and recorded relative distance  $\Delta P_{i,j}^0(t_0)$ . The  $\text{TCPA}_{i,j}(t_0)$  is determined as  $t - t_0$  for a  $t$  such that the future distance  $\|\Delta P_{i,j}(t)\|$  is minimised, for  $t \geq t_0$ . The relative distances are calculated as

$$\Delta P_{i,j}^0 = \begin{bmatrix} (x_i - x_j)\gamma \cos\left(\frac{y_i + y_j}{2}\right) \\ (y_i - y_j)\gamma \end{bmatrix} \tag{2}$$

where  $\gamma = 111,319.9$  is the decimal degree-to-metre conversion rate, longitudes are curvature corrected and all expressions are evaluated at  $t_0$ . Relative velocity is calculated as

$$\Delta V_{i,j} = V_i - V_j \tag{3}$$

This gives  $\text{TCPA}_{i,j}$  as

$$\text{TCPA}_{i,j} = \frac{-(\Delta P_{i,j}^0)^\top \Delta V_{i,j}}{(\Delta V_{i,j})^\top (\Delta V_{i,j})} \quad \text{if } \Delta V_{i,j} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4}$$

If  $0 \leq \text{TCPA}_{i,j}(t_0) \leq \text{TMAX}$ , where the present paper lets  $\text{TMAX} = 1,200$  s, we say vessels  $(i, j)$  will be closest at  $t = t_0 + \text{TCPA}_{i,j}(t_0)$  given present speed and course, and have

$$\Delta P_{i,j}^{\text{TCPA}} = \Delta P_{i,j}^0 + \text{TCPA}_{i,j} \Delta V_{i,j} \tag{5}$$

and, further, the DCPA as

$$DCPA_{i,j} = \sqrt{(\Delta P_{i,j}^{TCPA})^\top (\Delta P_{i,j}^{TCPA})} \tag{6}$$

Otherwise, we define  $DCPA_{i,j} = \infty$ . A version of a function to determine  $TCPA_{i,j}$  for two vessels may be implemented as

**Listing 1.** A simple function to determine TCPA for two vessels.

```

1 def SingleCPA(x1, y1, sog1, cog1, x2, y2, sog2, cog2):
2     gamma = 111319.9
3
4     TMAX = 1200
5     dcpa = inf
6     Kn2ms = 1852 / 3600
7     dVx = (sog1 * sin(cog1) - sog2 * sin(cog2)) * Kn2ms
8     dVy = (sog1 * cos(cog1) - sog2 * cos(cog2)) * Kn2ms
9
10    dx = (x1 - x2) * gamma * cos((y1+y2)/2)
11    dy = (y1 - y2) * gamma
12    tcpa = (dx * dVx + dy * dVy) / (dVx * dVx + dVy * dVy)
13
14    if (0 <= tcpa <= TMAX):
15        dcpa = sqrt((dx + dVx * tcpa)*(dx + dVx * tcpa)
16                  + (dy + dVy * tcpa)*(dy + dVy * tcpa))
17
18    return (tcpa, dcpa)

```

For  $N$  vessels at time  $t_0$ , represented as explained in Section 2, the algorithm applies the function above to every pair of vessels using vectors instead of scalars. The function looks the same, and may be implemented as<sup>2</sup>

**Listing 2.** Vectorised function to determine TCPA for  $N$  vessels.

```

1 def VectorCPA(X1, Y1, SOG1, COG1, X2, Y2, SOG2, COG2):
2     >> Same computations but vectorised
3
4     (...)
5
6     idx = (0 <= TCPA) && (TCPA <= TMAX)
7
8     DCPA[idx] = sqrt((dX[idx] + dVX[idx] * TCPA[idx])**2
9                  + (dY[idx] + dVY[idx] * TCPA[idx])**2)
10
11    return (TCPA, DCPA)

```

where capital letters denote vectors.

The algorithm determines  $TCPA_{i,j}$  and  $DCPA_{i,j}$  for all vessel pairs  $(i, j)$ . An NCS is identified when  $DCPA_{i,j}$  is found to be below a set (unsafe) threshold. The present paper identifies situations where the minimum DCPA is lower than three times the sum of the vessel lengths, i.e.

$$DCPA_{\min}(i, j) \leq 3 \times (L_i + L_j)$$

where  $L_i, L_j$  are the vessel lengths.<sup>3</sup> The algorithm identifies this on the record level for each vessel pair. A database can be constructed in several ways. The present paper lets each NCS between a pair of vessels be identified by the record denoting the lowest  $DCPA_{i,j}(t_0)$  for the vessels.

For every identified situation, full-resolution traffic data are retrieved from the original data set for both vessels at all timestamps preceding and succeeding the timestamp of minimal  $DCPA_{i,j}$  by a sufficient time span, to further analyse the CPA and manoeuvre patterns during the NCS. Choice of time

<sup>2</sup>The algorithm is implemented in Python's Numba library as explained in Section 2. It is presented here in native Python-style code for simplicity.

<sup>3</sup>The choice of such a threshold is important as it decides which NCS candidates are characterised for further inspection. The researcher may opt for a larger threshold to allow for a greater number of NCS candidates for later inspection, but will need to compensate for this in other ways to maintain robustness. The significance of this safety limit can be assessed based on the obtained results for the *passing distance* statistic.

span must be sufficiently long to encompass the situation, and must be informed by the size and type of the vessels. The present paper retrieved records 60 min prior to and following the point of minimal  $DCPA_{i,j}$ .

Transmission times differ, and hence traffic information must be synchronised for each pair of vessels participating in a situation to analyse CAMs. Records are synchronised by a process where each record, with timestamp  $t_{0,i}$ , is coupled with an equivalent  $t_{0,j}$  in the other vessel's timestamp vector such that  $|t_{0,i,k} - t_{0,j,l}| < 5$  and  $t_{0,j} = \arg \min_{t_{0,j} \in T_0} |t_{0,i} - t_{0,j,k}|$ , where  $T_0$  is the vector of timestamps for vessel  $j$ . Records which have no unique equivalent are discarded. The traffic information synchronisation is implemented as

**Listing 3.** Function to synchronise observations.

```

1 def ObservationSynchronizer(t0i, T0j):
2     jt0 = numpy.argmin(abs(t0i - T0j))
3     t0j = T0j[jt0]
4
5     if abs(t0j - t0i) > 5:
6         return
7
8     else:
9         return t0j

```

where the records are organised such that  $T0j$  is the longest timestamp vector of the two.

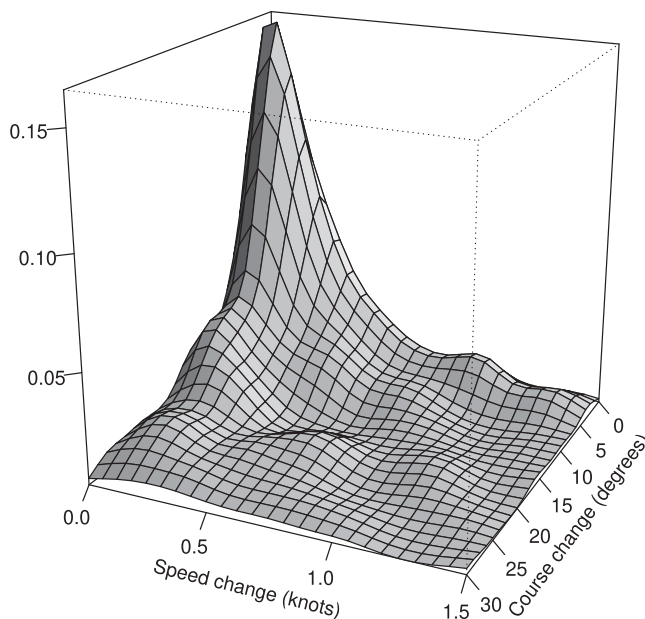
The synchronised time for each record is taken to be the mean timestamp of the coupled records. This introduces imprecision, but the common time scale is not used for computations. At this stage, both vessels are observed at approximately identical timestamps. The synchronised traffic data for each situation are then used to analyse the traffic patterns during CAMs. First, the timestamp when the CAM is initiated is detected as  $t_{1,i,j}$  and the timestamp when the NCS is resolved is detected as  $t_{F,i,j}$ . The evacuation time window is defined as  $[t_{1,i,j}, t_{F,i,j}]$  for each situation, according to the following rules.

1. If  $DCPA_{i,j}$  does not pass below 10 m, the time window is initiated at the minimum  $DCPA_{i,j}$ .<sup>4</sup> This is the normal situation where a CAM is initiated to ensure a safe passing distance by increasing the (predicted) DCPA.
2. If  $DCPA_{i,j}$  passes below 10 m, the time window starts at the initial point where  $DCPA_{i,j}$  passes below 10 m for which  $TCPA_{i,j}$  has not turned negative. For such NCSs, the calculated values are assumed to be subject to noise, and the first crossing of 10 m is considered the global minimum.
3. If  $DCPA_{i,j}$  passes below 10 m, but  $TCPA_{i,j}$  is negative at this point, the time window starts when  $DCPA_{i,j}$  is at its minimum while  $TCPA_{i,j}$  has not turned negative.
4. If  $DCPA_{i,j}$  never passes below a threshold  $DMAX$  metres (in the present paper set as three times the sum of the vessel lengths), as explained previously, the situation is discarded from the NCS database.

If the situation is confirmed as an NCS, it is taken to be resolved at  $t_{F,i,j}$ , the first point after  $t_1$  when  $TCPA_{i,j}$  turns negative. This implies that the two vessels involved in the NCS are no longer approaching one another. The time interval  $t_{F,i,j} - t_{1,i,j}$  is an important manoeuvre aspect as it determines how 'early' (as emphasised in COLREGs) CAMs were initiated, and how the linguistic variable 'early' from COLREGs is interpreted in real situations.

An observed CAM is defined for a vessel if, in a situation, the vessel alters her course or speed by a significant (observable) amount during the evacuation time window. For each situation, we observe either no vessel, one vessel or both vessels engaged in a CAM. In an ideal application, this determines which vessel should take responsibility to give-way and which vessel practised the right-of-way and remained in a stand-on state, as per COLREGs. Furthermore, for each vessel, we may observe no CAM, a course alteration, a speed alteration or both a speed and course alteration. If one of the latter three scenarios is recorded, the vessel is defined as having implemented a CAM. Each of the latter three scenarios defines

<sup>4</sup>The 10 m threshold is introduced to avoid measurement error arising from sampling noise. As AIS data are gathered discretely, choosing the minimum DCPA when this hovers close to zero will lead to late identification of  $t_1$ .



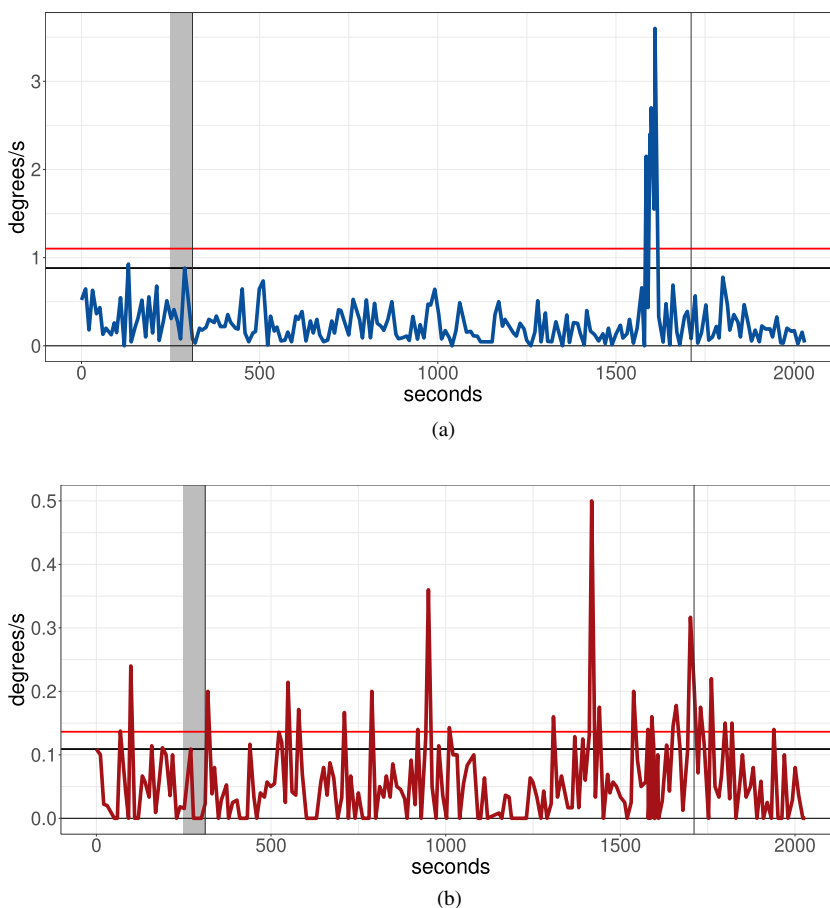
**Figure 1.** Joint distribution of the magnitude of observed steady-state course and speed change for vessels where a course-change CAM is observed. The figure indicates that there is a speed change observed for course-change CAMs.

a type of CAM. Situations where one or more vessels engage in a CAM are the NCSs of interest. Some speed change is also observed during course alteration manoeuvres, as shown in Figure 1. In practice, large vessels more often engage in course-change manoeuvres rather than incremental speed-change manoeuvres. Combining this assumption with Figure 1, one may deduce that a course-change CAM will often entail a detectable speed change and, hence, be registered as a combined manoeuvre. This understanding should enter into the researcher's analysis of the results.

Classification of manoeuvre type is done vessel-by-vessel, by analysing the individual patterns of changes in COG and SOG over the entire CAM time window, as shown in Figure 2. A course-change CAM is classified as such if the time derivative of the COG deviates significantly from the measurement noise inherent in the time series. The same method is used for classifying speed-change CAMs, by exchanging COG with SOG. In this regard, the threshold for 'measurement noise' is defined as 25% above the maximum absolute value of the COG time derivative for the 60 s prior to the start of the evacuation window. If the COG time derivative passes above this measurement noise threshold, the vessel is classified as having engaged in a course-change CAM. The application is illustrated for a representative vessel in Figure 2(a). The figure shows how a threshold is defined based on the pre- $t_1$ -period, and that a course-change CAM is defined for this vessel, as the within-window COG time derivative passes above this threshold. In addition to being useful for analysing the practice of the COLREGs rules, these aspects allow for the analysis of which and how many types of manoeuvres are executed to avoid collision in each situation.

This method is not without its limitations. Measurement noise is amplified when calculating derivatives, and the choice of 25% and 60 s as tuning parameters should be subject to refinement. As seen in Figure 2(b), if the time derivative is consistently and sufficiently close to zero for the duration of the situation, the measurement noise threshold may be set too low, and a CAM may incorrectly be recorded. In particular, misclassifications should be expected for vessels where the total speed change is close to zero. To correct for such misclassifications, vessels for which no observable steady-state course change is recorded are classified as not having performed a course-change CAM.



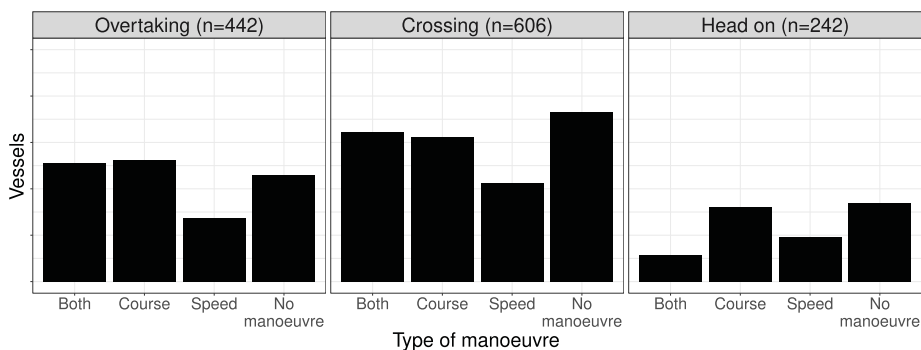


**Figure 2.** COG time derivative for two representative vessels in an overtaking situation: (a) successful classification; (b) faulty initial classification, needs correction. The black line shows the maximum noise recorded prior to the evacuation window. The red line shows the implied noise threshold. For both vessels, a course-change manoeuvre is initially registered. In the first example, this is also the final classification. The grey-shaded field indicates the time window within which the noise threshold is defined. In the second example, our methodology recognises that the total registered course change is too small to indicate an actual course-change manoeuvre, and the final classification is corrected to ‘No manoeuvre’. The time series is taken from the NCS example presented in Figure 7.

The steady-state course change for a vessel is defined as the difference between the mean COG for 11 observations surrounding the start of ( $t_1$ ) and end of ( $t_F$ ) the evacuation time window.<sup>5</sup> We define the observed course change as being observationally zero if it is below  $1.5^\circ$ . This roughly conforms to the lower 30th percentile of the distribution of observed steady-state course change for all vessels. The same approach is undertaken with respect to speed change, but where the threshold for what is considered observationally zero is set at below 0.15 knots of change, which conforms approximately to the lower 30th percentile of the distribution.<sup>6</sup> These corrections result in a change in the course-change CAM status for 6.2% of all NCS-involved vessels, and for 9.6% of all vessels with regards to the detection of a speed-change CAM. The distribution of situation types, differentiated by manoeuvre status, is presented in Figure 3.

<sup>5</sup>For calculations involving COG, means are understood to be circular.

<sup>6</sup>The choice of thresholds for this classification represents a crude instrument for re-classifying obvious misses in the data-driven method. Further research is needed with regards to precise classification of manoeuvre typology.



**Figure 3.** Distribution of registered manoeuvre types for NCSs where at least one CAM is initiated. In the case where both types of manoeuvres are registered, the intuition is not that both manoeuvres were necessarily intended, but rather that a course change often entails a speed change.

The observed steady-state change in COG and SOG are also important indicators of how ‘substantial’ the course and/or speed alterations are in real situations. This allows for evaluating the real interpretation of ‘substantial’ in terms of COLREGs compliance for each situation.

The full comprehension of CAMs further requires analysis of the conditions at which the manoeuvres are initiated and their effect on solving the NCS. Several manoeuvre aspects may be of interest. The present paper estimates the relative approach speed, the distance between vessels at  $t_1$  and passing distance, in addition to the evacuation time, and the steady-state course and speed change for different situations in aggregate. The aspects are defined as follows.

- *Relative approach speed* is computed as the absolute value of the relative velocity,  $|V_i - V_j|$ , between the two vessels in the first half of the evacuation time window,  $[t_1, t_M]$  (where  $t_M = (t_1 + t_2)/2$ ). This indicates the speed with which the vessels are approaching each other.
- *Vessel distance at  $t_1$* ,  $|P(t_1) - Q(t_1)|$ , is computed as the actual distance between the two vessels at the start of the CAM,  $t_1$ , and it measures how close vessels are before initiating a CAM.
- *Passing distance*,  $|P(t_F) - Q(t_F)|$ , is computed as the actual distance between vessels at the end of the evacuation time window,  $t_F$ . This measures the effect of the CAM in avoiding close-quarters situations.
- *Evacuation time*,  $t_F - t_1$ , is calculated as the duration from the start of the CAM to the end of the NCS. This indicates how early the manoeuvre was initiated.
- *Steady-state course (respectively speed) change* is calculated for each vessel as the difference between the mean COG (respectively SOG) for 11 observations around the start and end of the evacuation time window.

To analyse situations with respect to practised CAMs, all NCSs are classified based on which COLREGs rule applies for the vessels. In particular, COLREGs rules 13, 14 and 15 stipulate actions to be taken by the participating vessels in situations where a vessel is *overtaking*, *crossing* or in a *head-on* encounter (Ventura, 2005).

COLREGs rule 13(b) specifies that ‘A vessel shall be deemed to be overtaking when coming up with another vessel from a direction more than  $22.5^\circ$  abaft her beam (...)’, whereas COLREGs rule 14(a) specifies that a head-on situation takes place ‘When two power-driven vessels are meeting on reciprocal or nearly reciprocal courses so as to involve risk of collision (...)’ and 14(b) that ‘Such a situation shall be deemed to exist when a vessel sees the other ahead or nearly ahead (...)’. Conditions for when vessels are in a crossing situation (with actions as required by rule 15) are not specified, but it is generally understood to be situations which do not fulfil the criteria of rules 13 and 14.

To classify NCSs in light of COLREGs, it is necessary to mathematically define the terms used in the different rules. As any subsequent alteration of course or bearing should not change the class of a situation, a situation is classified based on its traffic features prior to  $t_1$ , i.e. before the vessels initiate any

CAMs. Data on vessel heading are often recorded as part of AIS, but are in the present data considered to involve too many erroneous and missing records to be used as part of a robust measure. Hence, the heading of a vessel  $i$  at  $t_1$  is approximated as the mean course in the observations preceding  $t_1$ ,

$$\overline{\text{COG}}_i = \text{mean}(\text{COG}_i(t)) \quad \text{s.t. } t_1 - 20 \leq t \leq t_1 \tag{7}$$

Taking the mean over several records makes the measure more robust against noise and contributes to the accuracy of the classification.

In any encounter between two vessels  $i$  and  $j$ , the relative angle of approach, henceforth denoted  $\Delta\text{COG}$ , is determined as the difference, translated to the interval  $[0^\circ, 360^\circ)$ ,

$$\Delta\text{COG} = \text{mod}(\overline{\text{COG}}_i - \overline{\text{COG}}_j, 360^\circ) \tag{8}$$

The relative bearing between vessels  $i$  and  $j$  with respect to  $i$  is calculated as

$$\beta_i = \text{mod}(\text{angle}(\Delta P_{j,i}^{t_1}) - \overline{\text{COG}}_i, 360^\circ) \tag{9}$$

where  $\text{angle}(\Delta P_{j,i}^{t_1})$  is the angle of the relative distance vector starting from the position of vessel  $i$  and ending at the position of vessel  $j$  at  $t_1$ , as defined in Equation (2). This angle is the absolute bearing and is measured with respect to the north in a clockwise direction.

Correspondingly,  $\beta_j$  is the relative bearing with respect to vessel  $j$  and is calculated by switching the indices  $i$  and  $j$  in Equation (9). If vessel  $i$  is own ship and vessel  $j$  is target ship,  $\beta_i$  and  $\beta_j$  are generally understood to be the bearing angle and contact angle, respectively, see Woerner et al. (2019).

NCSs are classified based on  $\beta_i, \beta_j$  and  $\Delta\text{COG}$ , calculated at  $t_1$ . Starting with rule 13(b), a vessel  $j$  is overtaking vessel  $i$  if  $j$  has a relative bearing of more than  $\phi_o$  behind the beam of vessel  $i$ , equivalently  $90^\circ + \phi_o$  off her centreline, where  $\phi_o = 22.5^\circ$  is clearly defined in COLREGs.

Rule 14(a) specifies that two vessels are head-on if  $\Delta\text{COG}$  is in the interval  $180^\circ \pm \phi_h$ . Here  $\Delta\text{COG} = 180^\circ$  indicate reciprocal courses, and  $\phi_h$  is an added tolerance for ‘nearly’ reciprocal. The  $\phi_h$  further accounts for the qualification in rule 14(c) that ‘When a vessel is in any doubt as to whether such a situation exists, she shall assume that it does exist and act accordingly’.

A head-on situation might equivalently be classified using  $\beta_i, \beta_j$  and rule 14(b). In this case, both relative bearings should be close to  $0^\circ$ , such that the vessels see each other ‘ahead or nearly ahead’. The former approach is chosen in the following analysis.

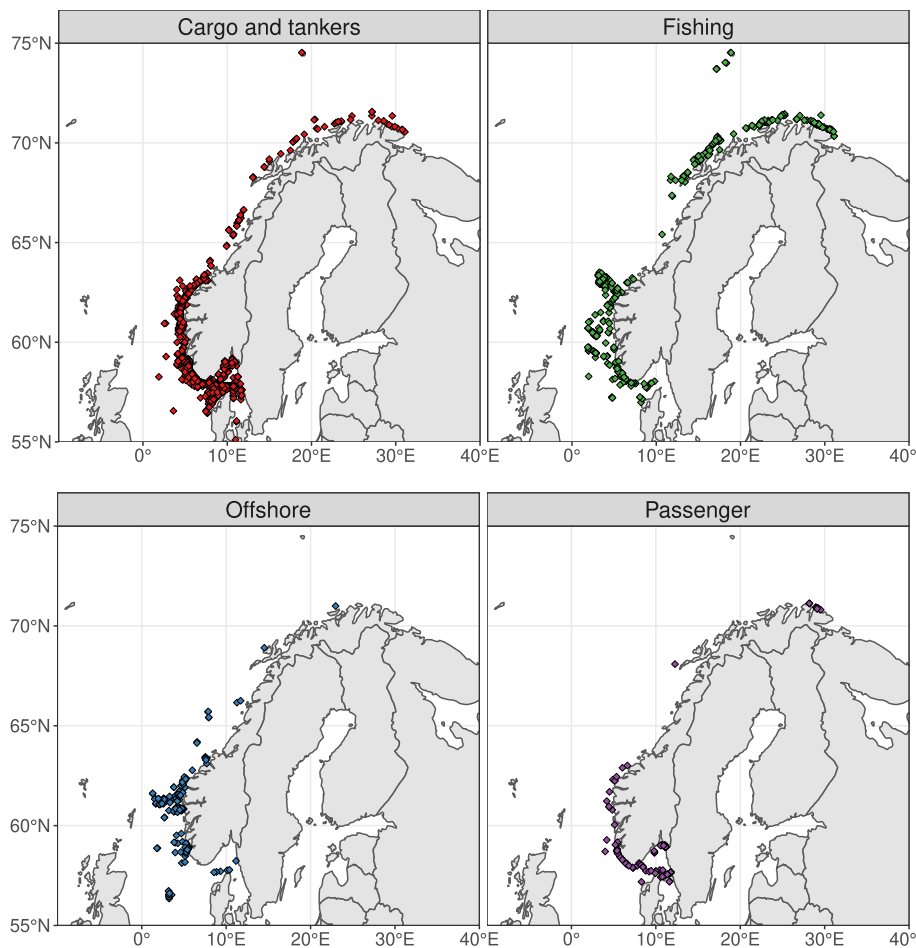
Neither part of rule 14 specifies numerically the magnitude of  $\phi_h$ . Court rulings indicate a convergence towards understanding ‘nearly reciprocal’ as meaning  $\phi_h \in [5, 6]$ . Recent additions to the literature are however not unanimous as to how  $\phi_h$  should be interpreted. Woerner et al. (2019) use  $\phi_h = 13^\circ$  as their default tolerance angle for ‘reciprocal or nearly reciprocal courses’, whereas Wang et al. (2018) apply  $\phi_h = 15^\circ$  when studying obstacle avoidance. At the upper end, Li et al. (2020) and Cho et al. (2020) apply  $\phi_h = 22.5^\circ$  in their analysis. It is outside of the scope of this paper to conclude as to the true value of  $\phi_h$ . However, the noisy properties of AIS data argue against opting for the narrowest definition. In the following,  $\phi_h = 10^\circ$  is applied in the classification algorithm.<sup>7</sup>

A situation is mathematically defined by the following conditions: If a situation is an NCS (where collision risk exists), we have the rule applying as

$$\dots = \begin{cases} j \text{ overtaking } i & \text{if } \cos(\beta_i) \leq \cos(90 + \phi_o) \\ i \text{ overtaking } j & \text{if } \cos(\beta_j) \leq \cos(90 + \phi_o) \\ \text{Head-on} & \text{if } \cos(\Delta\text{COG}) \leq \cos(180 + \phi_h) \\ \text{Crossing} & \text{Otherwise} \end{cases}$$

where cosines are used as they treat angles as equal on each side of  $180^\circ$  and  $0^\circ$ .

<sup>7</sup> Upon inspecting the data, setting  $\phi_h = 5$  identifies 185 of the 1055 situations outside the Norwegian baseline as head-on,  $\phi_h = 10$  identifies 303 and  $\phi_h = 15$  identifies 364 situations. Downloaded from <https://www.cambridge.org/core>. IP address: 213.52.102.34, on 09 Sep 2021 at 08:11:53, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/S0373463321000357>



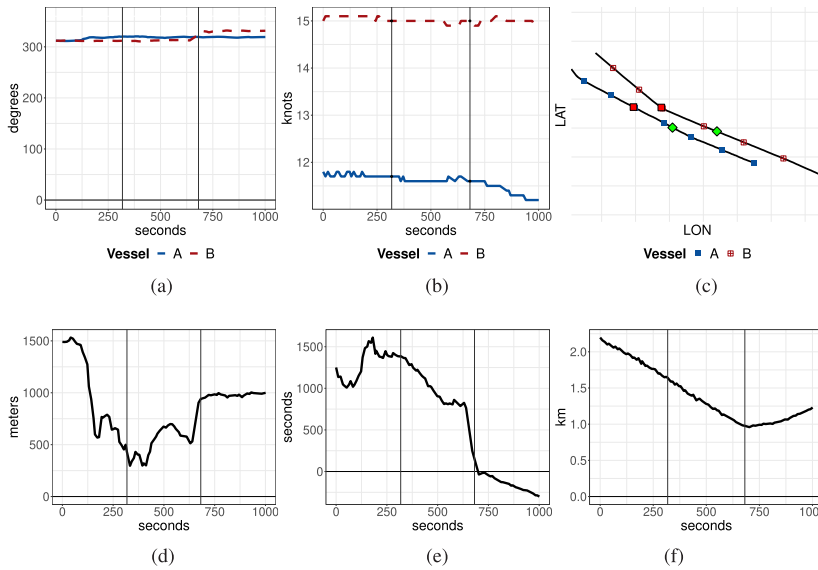
**Figure 4.** Vessels involved in NCSs in Norwegian open waters, vessels identified separately. The figure shows that the observed location of different vessels is localised according to activity.

## 4. Results

This section presents some identified NCSs, and discusses characteristics of such situations, as well as strengths and flaws of the presented framework. Summary statistics for the different classifications are presented, as well as the empirical distributions and mean and median estimates for important manoeuvre aspects. The analysis is focused on situations outside of the Norwegian baseline (NB), which are not subject to fixed obstacles, traffic separation schemes, and shallow and congested waters. The sample of NCSs, as shown in Figure 4, still mostly take place in the Norwegian EEZ, which is more congested than other parts of the open sea. Application on the described data sample identified a total of 9,180 situations of which 1,055 took place outside the NB. There are 2,110 separate vessels involved in such situations. Of these, CAMs were detected in 645 situations .

### 4.1. Representative situations

Figures 5–7 show situations identified by the algorithm which are indeed NCSs, involving two vessels on path for collision.



| Vessel | Type   | Registered manoeuvre | Approach speed | Distance at $t_1$ | Evacuation time | Passing distance | Speed change | Course change |
|--------|--------|----------------------|----------------|-------------------|-----------------|------------------|--------------|---------------|
| A      | Tanker | –                    | 3.8 kn         | 1.59 km           | 6 min           | 0.96 km          | 0.07 kn      | 0.56 °        |
| B      | Tanker | Course change        |                |                   |                 |                  | 0.05 kn      | 10.89 °       |

(g)

Figure 5. Representative ‘overtaking’ situation: (a) COG; (b) SOG; (c) track; (d) DCPA; (e) TCPA; (f) vessel distance; (g) manoeuvre aspects.

Figure 5 shows a correctly classified situation where two tankers meet in an overtaking-type encounter. As shown by Figure 5(a), 5(c) and 5(f), vessel A and vessel B are on the same course, with vessel B closing in on vessel A from behind at a higher SOG. During the evacuation time window, vessel B engages in a course-change CAM. As shown by Figure 5(d), this increases the distance at CPA, and the vessels clear the NCS. TCPA becomes negative when the situation is resolved. The manoeuvre aspects presented in Figure 5(g) underline these observations. Vessels pass within one kilometre of each other when the situation is resolved.

Figure 6 shows a situation where a cargo vessel (A) and an offshore vessel (B) meet under the crossing rule. Vessel A makes a course-change manoeuvre, leaves the NCS and returns to its prior course. As shown by Figure 6(d), the course alteration by B increases DCPA substantially during the CAM, while, as Figure 6(e) shows, TCPA decreases. As seen in Figure 6(g), the detected course change for vessel A is somewhat smaller than what the track plot indicated. This is likely owing to the vessel correcting back to its initial course directly after having cleared the situation. Vessel B makes a small course adjustment during the evacuation window, which is recorded by the algorithm.<sup>8</sup>

Figure 7 depicts a situation where a cargo vessel (A) and a tanker (B) meet under head-on rules. Figure 7(b) shows that there is substantial difference in speed between the vessels. Vessel A engages in a course-change manoeuvre, increasing DCPA and decreasing TCPA. As Figure 7(d) shows, DCPA varies substantially when the vessels are still far apart. In this instance, it may be argued that both vessels engage in a manoeuvre, although the manoeuvre of vessel A is more pronounced. This likely stems from the fact that the measurement of the course change is smaller than what the track plot indicates, as the vessels correct back to their original course directly after clearing the NCS.

<sup>8</sup>The extent to which this correction is to be regarded as a CAM, or as holding the course, is up for discussion. Under the present framework, it is registered as a CAM.

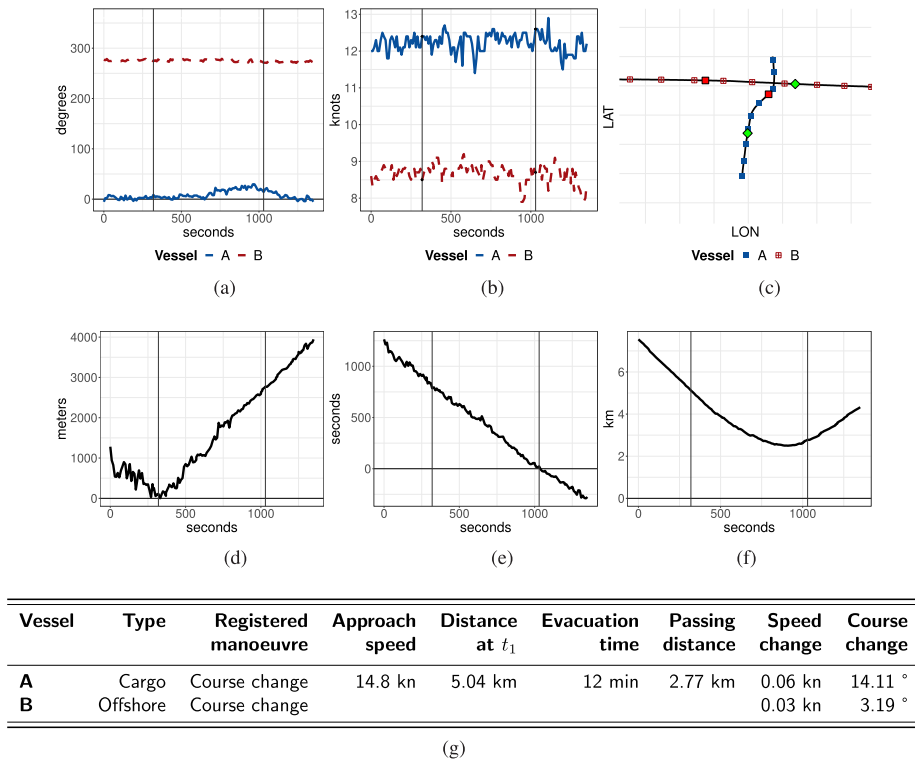


Figure 6. Representative ‘crossing’ situation: (a) COG; (b) SOG; (c) track; (d) DCPA; (e) TCPA; (f) vessel distance; (g) manoeuvre aspects.

### 4.2. Manoeuvre aspects

The suggested CAM aspects encompass time, space, situational and navigational aspects. The spatial and temporal aspects may be classified as either pre- or post-manoevr aspects. In the following, these aspects, as well as navigational aspects, are presented for the subset of situations where a CAM was detected in an NCS outside of the NB.

Figure 8 presents the empirical distribution of relative approach speed and Figure 9 presents the vessel distance at  $t_1$ , which are both pre-manoevr aspects, for situations governed by different COLREGs rules. The relative approach speed estimates are as expected, relative to both the COLREGs rule and vessel type, although there is some particular clustering of passenger vessels in head-on situations.

The average vessel distance at  $t_1$  is different in NCSs of different encounter types, as Figure 9 shows. Overtaking-type NCSs on average have a lower vessel distance at  $t_1$  than crossing situations, which again on average have a lower vessel distance at  $t_1$  than head-on-type situations. Figure 9 further shows that the distributions of vessel distance at  $t_1$  differ between the different vessel categories, conceivably owing to the vessel length.

Figure 10 presents the empirical distributions for the passing distance and Figure 11 for the evacuation time, which are both post-manoevr aspects, for various NCSs. As Figure 10 shows, there are only minor differences in the mean and median when comparing crossing-type situations to overtaking and head-on situations. It is, however, evident that the distributions vary with vessel type, and that passenger vessels in particular, and to some extent fishing vessels, skew lower than cargo vessels in terms of the observed passing distance, as is expected.

Evacuation time is an aspect that indicates how early the CAM was initiated prior to the situation being resolved. In terms of the average evacuation time, Figure 11 shows that although there is little difference in the mean and median between the different overtaking and crossing situations, both statistics are

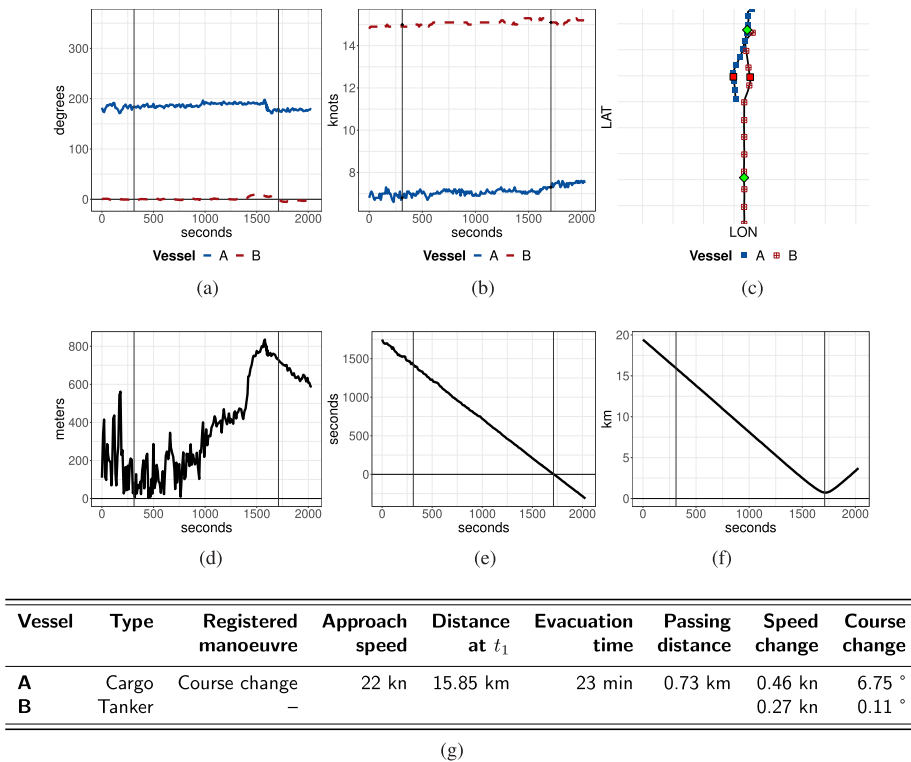


Figure 7. Representative ‘head-on’ situation: (a) COG; (b) SOG; (c) track; (d) DCPA; (e) TCPA; (f) vessel distance; (g) manoeuvre aspects.

substantially lower for head-on situations. For all NCS types, the distributions of evacuation time have long upper tails, which is related to the vessel speed and available space.

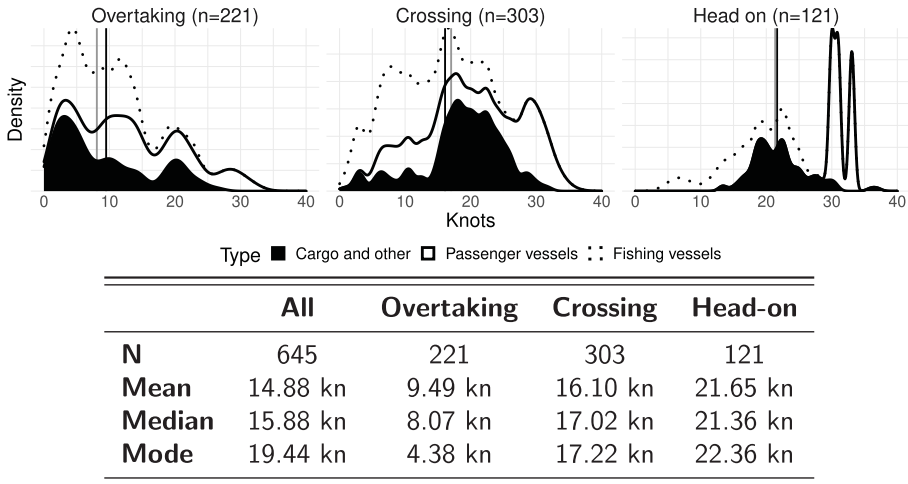
Figures 12 and 13 present the empirical distributions of steady-state course change and steady-state speed change, respectively, which are navigational aspects of CAMs. These aspects indicate the magnitude of implemented manoeuvres, which measure how ‘large’ and ‘substantial’ (as per COLREGs) are these alterations. The aspect distributions are presented for different vessel types, and only for vessels which engage in the respective manoeuvre types. Figure 12 shows that the observed average course-change CAM is on average equally significant for vessels in overtaking and crossing situations, which both are more significant than in head-on situations. The distribution of the magnitude of course-change CAMs has a considerable spread in overtaking and crossing situations.

Figure 13 shows the distribution of the absolute value of speed change during CAMs in various NCSs. The observed average speed-change is on average larger in overtaking situations than in crossing situations, which again is larger than in head-on situations. There are some differences with respect to the tails of the distributions when disaggregating by vessel type.

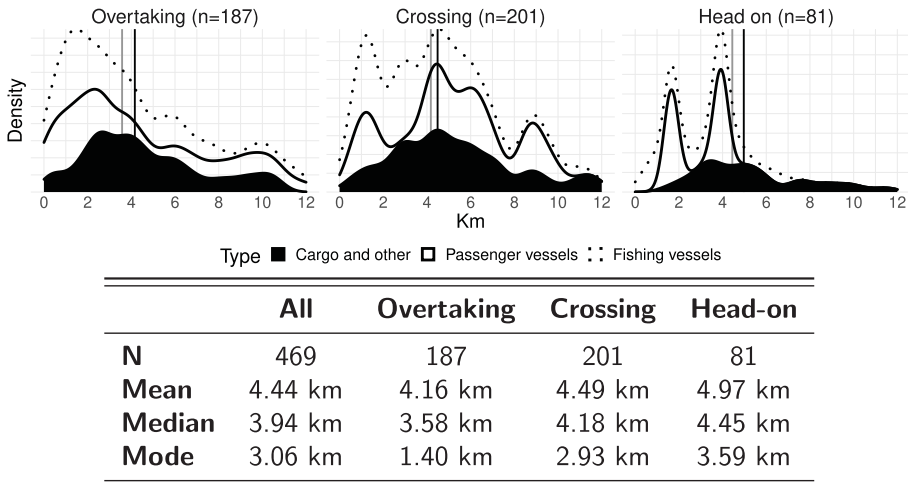
### 4.3. Effect of pre-manoeuve aspects on navigational outcomes

Estimating manoeuvre aspects allow for presenting the bivariate distributions of the aspects. In particular, the effect of pre-manoeuve aspects, such as approach speed and vessel distance at  $t_1$ , on navigational aspects are of interest, as these may conceivably affect the intended actions on the bridge and, hence, also the observed aspects in the data.

Figures 14 and 15 show the distribution of steady-state course and speed change, contingent on the distribution of approach speeds. Figure 14 shows that vessels undertaking a course-change CAM



**Figure 8.** Distribution of relative approach speed for vessels in a situation where a manoeuvre is implemented. The numbers indicate the speed with which two vessels approach each other prior to a CAM being initiated.

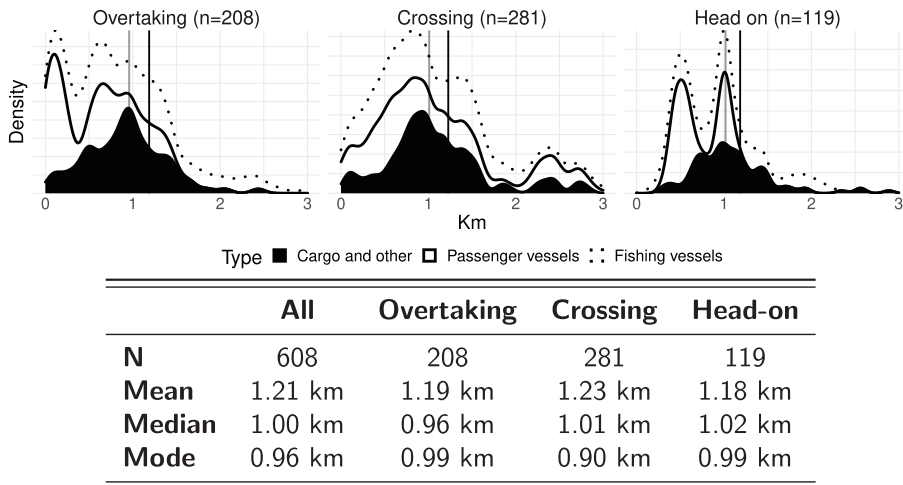


**Figure 9.** Distribution of vessel distance at  $t_1$  for vessels in a situation where a CAM is implemented. The numbers indicate how far away vessels are at the point where a CAM is initiated.

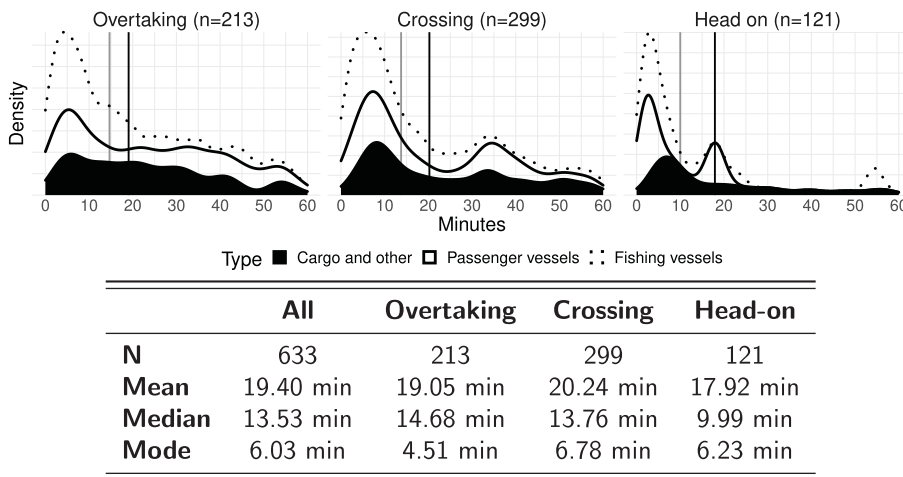
undertake course changes of different magnitude contingent on the approach speed involved in the situation. Figure 15 show that the same result applies for the speed change.

Figures 16 and 17 show the distribution of steady-state course and speed change, respectively, contingent on the distribution of vessel distance at  $t_1$ . The distribution of steady-state course change increases with vessel distance at  $t_1$  for all vessels undertaking a course-change manoeuvre. The same take-away message applies with regards to steady-state speed change.





**Figure 10.** Distribution of passing distance for vessels in a situation where a manoeuvre is implemented. The numbers indicate the distance between vessels at the point where a CAM is concluded and vessels are no longer in an NCS.



**Figure 11.** Distribution of evacuation time for vessels in a situation where a manoeuvre is implemented. The numbers indicate the time spent between when a CAM is initiated and its conclusion.

#### 4.4. Challenges of the proposed framework

##### 4.4.1. Misclassification and the importance of tuning parameters

Identifying NCSs and classifying CAMs is not straightforward and, in particular, both the NCS identification algorithm and the CAM classification algorithm depend on tuning parameters.

The main tuning parameters for the CAM classification algorithm are the padding of the measurement noise threshold (in the present paper, taken at 25% above the observed pre- $t_1$  maximum observation) and the length of the window in which the pre- $t_1$  maximum observation is estimated. With regards to the NCS identification algorithm, the choice of methodology for pinpointing  $t_1$  is crucial, as is the size of the window for estimating incoming and outgoing COG and SOG.

A wrong choice with regards to the former may lead to a later misclassification of the CAM. A wrong choice with regards to the latter may lead to situations being misclassified or defined as being

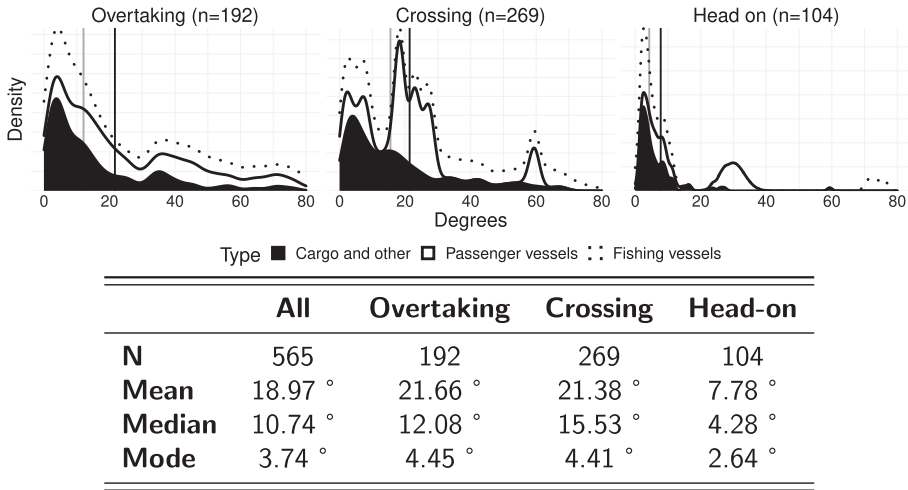


Figure 12. Distribution of steady-state course change for vessels classified as implementing a course-change manoeuvre. The numbers indicate the change in COG undertaken by vessels during a CAM.

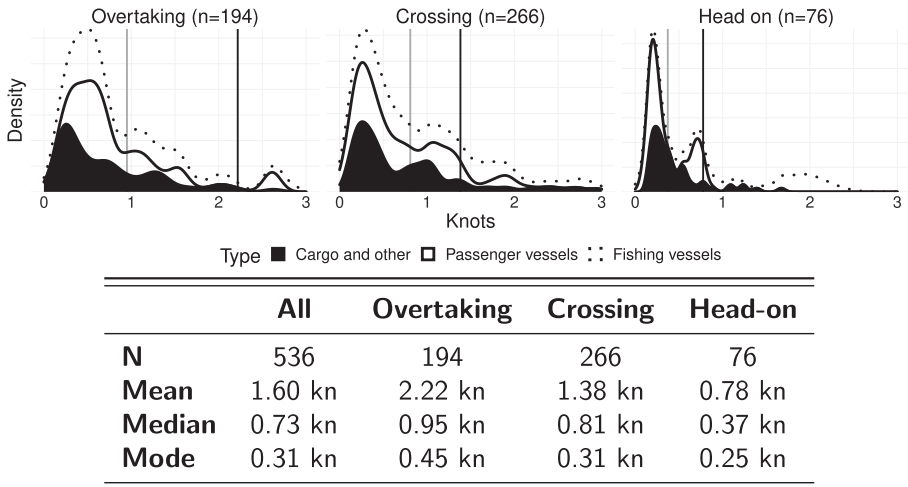
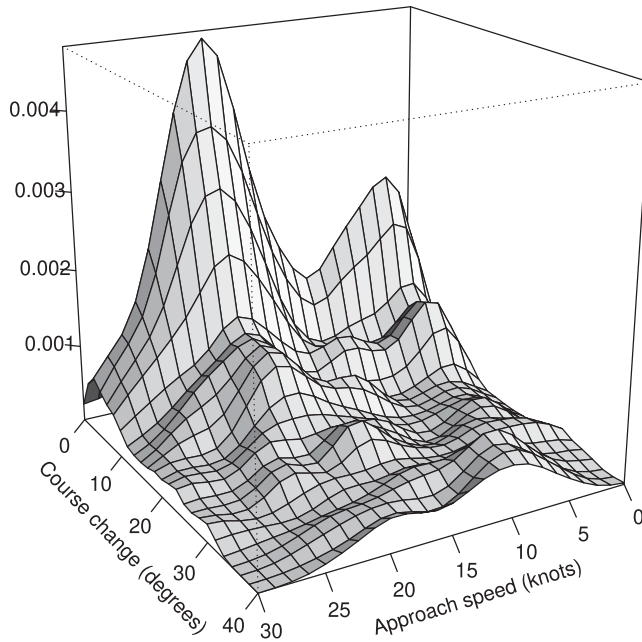


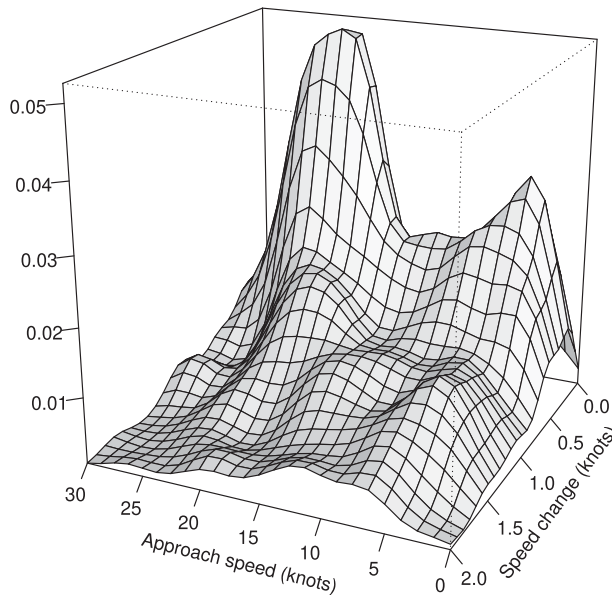
Figure 13. Distribution of steady-state speed change for vessels classified as implementing a speed-change manoeuvre. The numbers indicate the change in SOG undertaken by vessels during a CAM.

under different COLREGs rules if the vessel undertakes a manoeuvre just prior to  $t_1$ , which will produce wrong estimates for the  $\beta_i$ ,  $\beta_j$  and  $\Delta\text{COG}$ , and possibly incorrect COLREGs rule classification.

One example of a misidentification of  $t_1$  is given in Figure 18, which shows a situation where two vessels approach each other under a crossing regime (in this case, correctly classified). Vessel B undertakes a course-change manoeuvre, clears the situation and returns to its former course. Instead of correctly identifying  $t_1$ , the algorithm wrongly identifies  $t_1$  just after the CAM is concluded and instead classifies B as not having undertaken any CAM. The situation still enters the data because the algorithm misclassifies a course-following manoeuvre by vessel A as a CAM. Misclassifications such as this may be avoided by developing the algorithm for defining  $t_1$ .



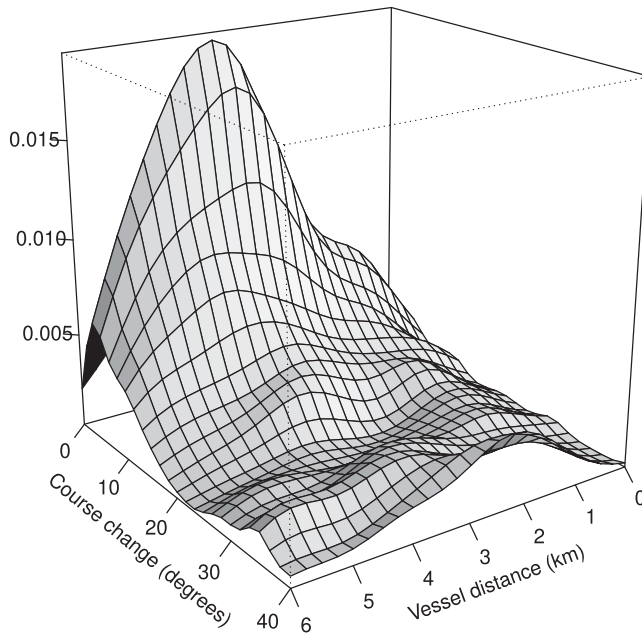
**Figure 14.** Bivariate distribution of approach speed and steady-state course change for vessels which undertake a course-change CAM.



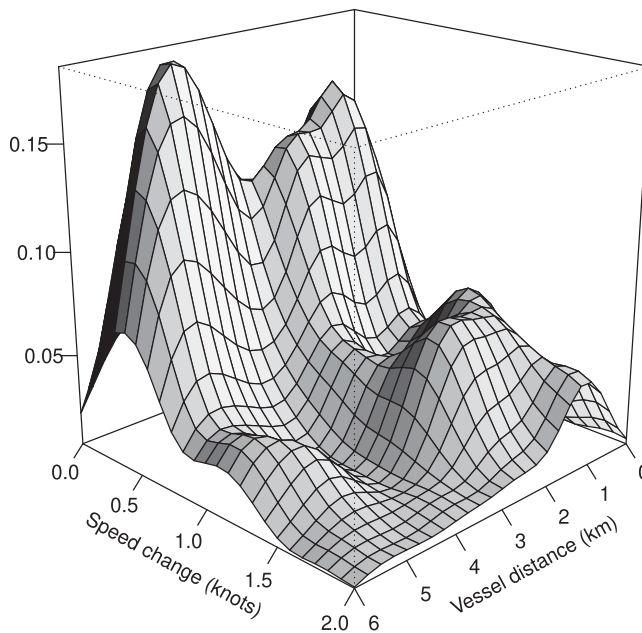
**Figure 15.** Bivariate distribution of approach speed and steady-state speed change for vessels which undertake a speed-change CAM.

4.4.2. Difference between observed and intended CAMs

The ideal framework for NCS and CAM analysis would allow the analyst to record the actions and intentions of the crew and captain on the bridge of every vessel. Such a framework would allow for the analysis of choices and could incorporate situational analysis. The AIS-based researcher is restricted to what may be observed from the data.

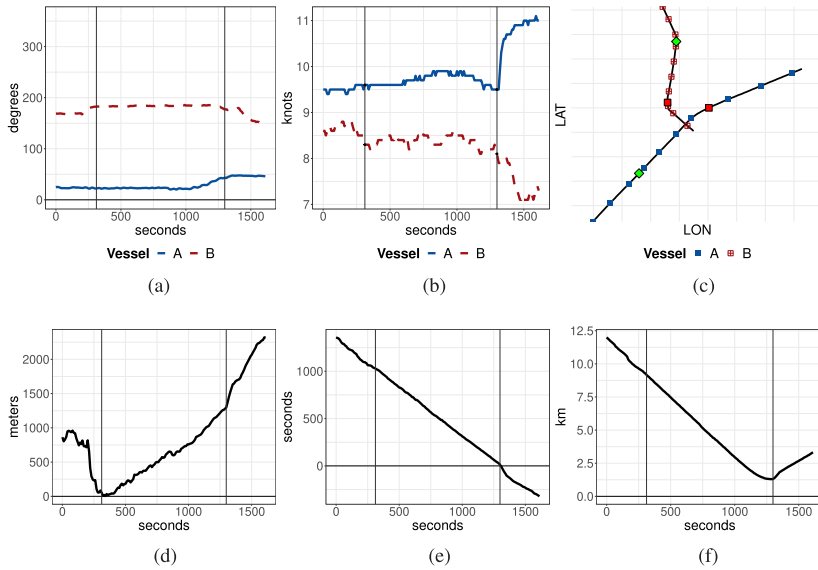


**Figure 16.** Bivariate distribution of vessel distance at  $t_1$  and steady-state course change for vessels which undertake a course-change CAM.



**Figure 17.** Bivariate distribution of vessel distance at  $t_1$  and steady-state speed change for vessels which undertake a speed-change CAM.

AIS data are a strong source of information, in that they are abundant and omnipresent. The proposed framework suggests algorithms which may enable the analyst to learn from these data. However, it may not inform about which actions are taken on the bridge. Furthermore, there may be other restricting



| Vessel | Type     | Registered manoeuvre | Approach speed | Distance at $t_1$ | Evacuation time | Passing distance | Speed change | Course change |
|--------|----------|----------------------|----------------|-------------------|-----------------|------------------|--------------|---------------|
| A      | Offshore | Course change        | 17.8 kn        | 9.08 km           | 17 min          | 1.39 km          | 0.23 kn      | 21 °          |
| B      | Cargo    | –                    |                |                   |                 |                  | 0.32 kn      | 3.55 °        |

(g)

**Figure 18.** A wrongly classified ‘crossing’ situation: (a) COG; (b) SOG; (c) track; (d) DCPA; (e) TCPA; (f) vessel distance; (g) manoeuvre aspects. In this situation, the algorithm defines  $t_1$  too late in the situation, which leads the algorithm to overlook the actual manoeuvre (taken by vessel B prior to the recorded  $t_1$ ), and instead designates a course-following manoeuvre taken by vessel A as the CAM. The table gives the mis-estimated CAM aspects.

factors at play. The present framework does not take account of sailing patterns and intended destination and route.

An example of a situation where it is difficult to ascertain whether a situation is in fact an NCS is presented in Figure 19. In this situation, a cargo vessel (A) and a fishing vessel (B) approach each other under crossing rules. Vessel B adjusts its speed and course to clear the situation. Nevertheless, vessel B does not heed the direction under crossing rules of passing due starboard of the vessel A. This may be a finding in and of itself, but truly understanding the factors which drive such actions necessitates extending the framework further to incorporate weather, water depth, sailing patterns and sea lanes.

#### 4.5. AIS NCS framework as a platform for CAM analysis and further applications

The NCS identification algorithm for creating a reference NCS database provides a useful platform for intelligent and safe maritime transport applications and analysis.

The presented results are useful with respect to the regulatory framework, to analyse how navigation rules, such as COLREGs, are practised in real situations, e.g. whether rules are followed, as well as which manoeuvres are taken by vessels involved in an NCS. The descriptive statistics allow for evaluating the interpretation of linguistic variables in navigation rules such as ‘early’, ‘substantial’, ‘far’, ‘large’ and ‘safe distance’. Providing technical standards or amendments to such rules is crucially dependent on fully comprehending the degree of compliance with current rules and which actions are taken by vessels in situations governed by different sets of rules.

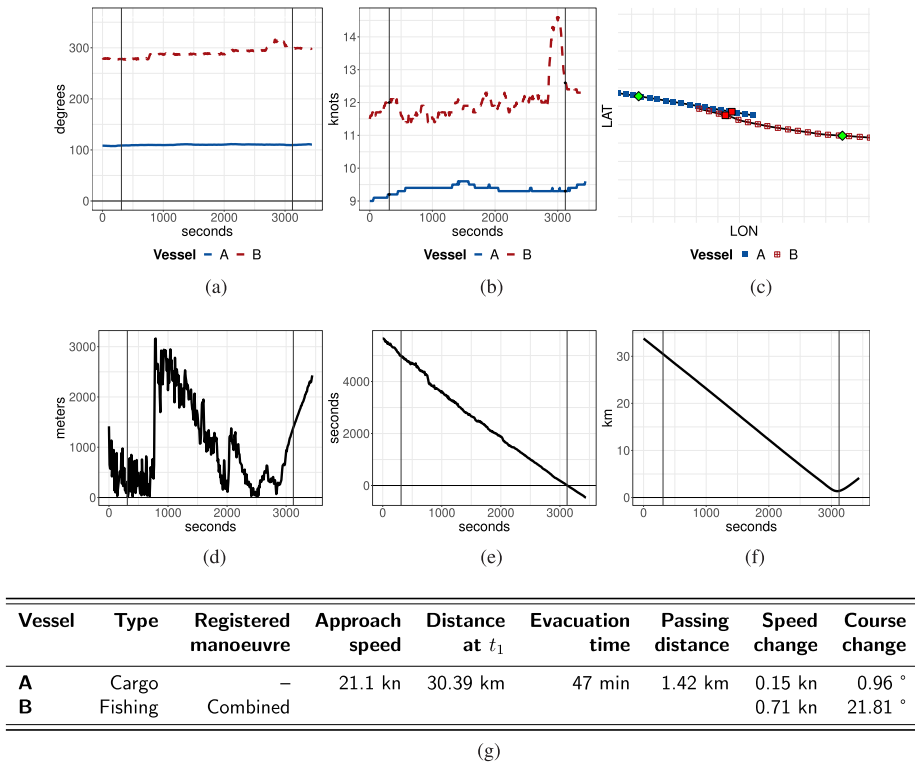


Figure 19. Representative ‘head-on’ situation: (a) COG; (b) SOG; (c) track; (d) DCPA; (e) TCPA; (f) vessel distance; (g) manoeuvre aspects.

For the application in risk analysis, the detection of risky manoeuvres can be made easier in comparison with the identified NCS database, and the evaluation of collision risk in real time can be improved using the provided statistics for common CAMs. For example, the presented statistics for evacuation time and passing distance may inform the selection of safety limits, with confidence intervals, for time to CPA and distance at CPA quantities, respectively.

The presented statistics provide useful inputs for the design of automatic collision avoidance algorithms, which are based on fixed rules designed for decision making to determine which actions should be taken, as well as when and how they should be executed in a particular situation. This has applications with regards to RCOs and safety assessment. In addition to risk evaluation, applications in performance evaluation of maritime autonomous surface vessels, which considers protocol compliance and human–robot interactions between unmanned and manned vessels, are relevant. Both challenges require explicit descriptions of ‘standard’ collision avoidance protocols and the expected behaviour of manned vessels in NCSs.

**5. Conclusion**

This paper has presented a framework for identifying NCSs using widely available AIS traffic data and the CPA algorithm.

The paper proves that the framework may be implemented in a simple, scalable and computationally efficient way, and that identification of such situations and creation of an NCS database may be used to analyse the execution of CAMs.

Future work will consider further refinements necessary for implementation in real-world applications. In particular, the framework needs to be adapted for use in restricted waters to accommodate for several vessels and stationary obstacles.

By the nature of geographically identified data, the framework lends itself to a range of further extensions, such as classification by weather aspects, in particular wave height, wind and sea surface currents, as well as the presence of restricted traffic zones, natural obstacles, recommended routes and traffic separation schemes. The framework may also be extended to include more detailed vessel characteristics.

**Acknowledgements.** We thank the three anonymous reviewers for thorough comments and feedback. This research was funded by the Norwegian Research Council research-based innovation center BigInsight, project no 237718.

## References

- Altan, Y. C. (2019). Collision diameter for maritime accidents considering the drifting of vessels. *Ocean Engineering*, **187**, 106158.
- Chen, P., Huang, Y., Mou, J. and van Gelder, P. H. A. J. M. (2018). Ship collision candidate detection method: A velocity obstacle approach. *Ocean Engineering*, **170**, 186–198.
- Chen, P., Huang, Y., Mou, J. and van Gelder, P. H. A. J. M. (2019). Probabilistic risk analysis for ship-ship collision: State-of-the-art. *Safety Science*, **117**, 108–122.
- Cho, Y., Han, J. and Kim, J. (2020). Efficient COLREG-compliant collision avoidance in multi-ship encounter situations. *IEEE Transactions on Intelligent Transportation Systems*, 1–13. doi:10.1109/TITS.2020.3029279.
- Fang, M.-C., Tsai, K.-Y. and Fang, C.-C. (2018). A simplified simulation model of ship navigation for safety and collision avoidance in heavy traffic areas. *The Journal of Navigation*, **71**(4), 837–860.
- Jeong, M.-G., Lee, E.-B., Lee, M. and Jung, J.-Y. (2019). Multi-criteria route planning with risk contour map for smart navigation. *Ocean Engineering*, **172**, 72–85.
- Kim, K.-I., Jeong, J. S. and Park, G.-K. (2014). Development of a gridded maritime traffic DB for e-navigation. *International Journal of e-Navigation and Maritime Economy*, **1**, 39–47.
- Le Tixerant, M., Le Guyader, D., Gourmelon, F. and Queffelec, B. (2018). How can Automatic Identification System (AIS) data be used for maritime spatial planning? *Ocean & Coastal Management*, **166**, 18–30.
- Li, J., Zhang, G., Liu, C. and Zhang, W. (2020). COLREGs-constrained adaptive fuzzy event-triggered control for underactuated surface vessels with the actuator failures. *IEEE Transactions on Fuzzy Systems*, 1–11. doi:10.1109/TFUZZ.2020.3028907.
- Liu, Z., Wu, Z. and Zheng, Z. (2019). A novel framework for regional collision risk identification based on AIS data. *Applied Ocean Research*, **89**, 261–272.
- Naeem, W., Henrique, S. C. and Hu, L. (2016). A reactive COLREGs-compliant navigation strategy for autonomous maritime navigation. *IFAC-PapersOnLine*, **49**(23), 207–213.
- Ożoga, B. and Montewka, J. (2018). Towards a decision support system for maritime navigation on heavily trafficked basins. *Ocean Engineering*, **159**, 88–97.
- Ozturk, U. and Cicek, K. (2019). Individual collision risk assessment in ship navigation: A systematic literature review. *Ocean Engineering*, **180**, 130–143.
- Sang, L.-z., Yan, X.-p., Wall, A., Wang, J. and Mao, Z. (2016). CPA calculation method based on AIS position prediction. *The Journal of Navigation*, **69**(6), 1409–1426.
- Szlapczynski, R. and Szlapczynska, J. (2016). An analysis of domain-based ship collision risk parameters. *Ocean Engineering*, **126**, 47–56.
- Szlapczynski, R. and Szlapczynska, J. (2017). Review of ship safety domains: Models and applications. *Ocean Engineering*, **145**, 277–289.
- Szlapczynski, R., Krata, P. and Szlapczynska, J. (2018). Ship domain applied to determining distances for collision avoidance manoeuvres in give-way situations. *Ocean Engineering*, **165**, 43–54.
- Ventura, M. (2005). COLREGS - International regulations for preventing collisions at sea. *Lloyd's Register Rulefinder 2005*, p. 74.
- Wang, Y., Yu, X., Liang, X. and Li, B. (2018). A COLREGs-based obstacle avoidance approach for unmanned surface vehicles. *Ocean Engineering*, **169**, 110–124.
- Wang, L., Liu, Q., Dong, S. and Guedes Soares, C. (2019). Effectiveness assessment of ship navigation safety countermeasures using fuzzy cognitive maps. *Safety Science*, **117**, 352–364.
- Woerner, K., Benjamin, M. R., Novitzky, M. and Leonard, J. J. (2019). Quantifying protocol evaluation for autonomous collision avoidance. *Autonomous Robots*, **43**(4), 967–991.
- Wu, B., Wang, Y., Zhang, J., Savan, E. E. and Yan, X. (2015). Effectiveness of maritime safety control in different navigation zones using a spatial sequential DEA model: Yangtze River case. *Accident Analysis & Prevention*, **81**, 232–242.
- Zaman, M. B., Santoso, A., Kobayashi, E., Wakabayashi, D. and Maimun, A. (2015). Formal safety assessment (FSA) for analysis of ship collision using AIS data. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, **9**(1), 67–72.
- Zhang, W., Goerlandt, F., Montewka, J. and Kujala, P. (2015). A method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, **107**, 60–69.

- Zhang, W., Goerlandt, F., Kujala, P. and Wang, Y.** (2016). An advanced method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, **124**, 141–156.
- Zhang, W., Kopca, C., Tang, J., Ma, D. and Wang, Y.** (2017). A systematic approach for collision risk analysis based on AIS data. *The Journal of Navigation*, **70**(5), 1117–1132.
- Zhou, Y., Daamen, W., Vellinga, T. and Hoogendoorn, S.** (2019). Review of maritime traffic models from vessel behavior modeling perspective. *Transportation Research Part C: Emerging Technologies*, **105**, 323–345.