Prediction of vessels locations and maritime traffic using similarity measurement of trajectory

Danial Alizadeh*, Ali Asghar Alesheikh and Mohammad Sharif

*Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran; †Department of Geography, Faculty of Humanities, University of Hormozgan, Bandar Abbas, Iran

ABSTRACT
Maritime traffic prediction is a crucial task for increasing the efficiency of port operations and safety, especially in congested regions. A huge amount of automatic identification system (AIS) data is constantly transmitting from vessels to receivers that contain information about vessels’ movements and characteristics. These historical AIS data can be utilized in movement analyses of vessels. This paper proposes a novel point-based model for location and traffic prediction using vessels’ trajectories adapted from AIS measures. The location prediction procedure is setup based on similarity analysis of historical AIS data. The model is applied to a real dataset of hundreds of vessels’ trajectories in the Strait of Georgia, USA. The correlation results of 0.9976, 0.9887, and 0.9794 for the 10, 20, and 30 minutes, respectively, imply sufficient correspondence between predicted and actual coordinates. The traffic prediction procedure considers the probability of the appearance of new vessels inside an area of interest (AoI) at different time intervals. The Sorensen similarity index (SSI) is used to measure the accuracy of the traffic prediction model. The SSIs for time intervals of 10, 20, and 30 minutes are 70%, 66%, and 59%, respectively, which show the robustness of the model to predict hot spots inside the AoI.

ARTICLE HISTORY
Received 17 February 2020
Accepted 16 October 2020

KEYWORDS
Location prediction; traffic prediction; trajectory; similarity measurement; AIS data

Introduction
Maritime transportation is of critical importance to the global economy because almost 90% capacity of global trade is conducted in this mode. Moreover, it is the most energy-efficient way to transport heavy goods throughout the world since it only uses 7% of all the energy consumed by transport activities (Rodrigue, Comtois, and Slack 2016). Despite the progress in marine navigation systems, an investigation on 177 maritime accident reports revealed that 71% of human error is caused by the lack of situational awareness (SA). Of the SA errors, 8.8% were related to the lack of awareness of the future state or action in the sea (Grech, Horberry, and Smith 2002).

According to the International Maritime Organization (IMO) maritime safety regulations, vessels above a certain size must be equipped with automatic identification systems (AISs) (Solas 2003). AIS data are composed of static information such as vessel name and maritime mobile service identity (MMSI), dynamic information such as vessel position, speed over ground, and course over ground, and voyage information that contains destination and estimated arrival time. Whereas dynamic and static information is reliable, voyage information may consist of incorrect information (Harati-Mokhtari et al. 2007). Dynamic information is sent every 2 to 10 seconds and 3 minutes, respectively, if the vessels are in the underway and anchor positions. Static and voyage information is sent every 6 minutes (Duca, Bacciu, and Marchetti 2017). Various studies have investigated the situational awareness enhancement strategies using AIS data, including anomaly behaviour detection (Sidibé and Shu 2017; Zhao and Shi 2019), waypoints detection (Dobrkovic, Iacob, and Van Hillegersberg 2015), extracting shipping route patterns (Sheng and Yin 2018; Fujino, Clarament, and Boudraa 2018), vessel trajectory prediction (Duca, Bacciu, and Marchetti 2017; Borkowski 2017; Graser, Schmidt, and Widhalm 2018; Pallotta, Vespe, and Bryan 2013; Gao, Shi, and Li 2018; Mao et al. 2018; Wijaya and Nakamura 2013; Alizadeh, Alesheikh, and Sharif 2020), time arrival prediction (Dobrkovic et al. 2016), and traffic prediction (Xiao et al. 2017; Kim and Lee 2018).

Autonomous vessels should be able to predict the next location of their surrounding vessels and adjust their future path to avoid collision with them. Moreover, to increase safety and enhance the efficiency...
of port operations it is essential to predict future areas of dense maritime traffic inside the congested areas and areas close to the ports (Kim and Lee 2018). However, most of the developed methods for location and traffic predictions are based on mathematical models such as Markov models that use probabilistic models to predict the next location of moving objects (Mathew, Raposo, and Martins 2012; Asahara et al. 2011). These models are not suitable for location prediction of vessels due to the nonlinearity of vessel movement and complexity of the maritime environment. They are not functional when nonlinearity and complexity increase (Alizadeh, Alesheikh, and Sharif 2020). In addition, AIS messages are transmitted at irregular time intervals and therefore they cannot be fed into typical introduced Markov models as input (Graser, Schmidt, and Widhalm 2018). Additionally, some of the location and traffic prediction approaches are solely based on the transport network. In other words, moving objects are forced to move only on the transport networks such as road segments. Examples of works following this approach can be found in (Zhu et al. 2018; Simmons et al. 2006). As a result, these approaches are not suitable for moving objects such as vessels that move freely and are not constrained by transport networks (Graser, Schmidt, and Widhalm 2018).

The presented framework, which is based on similarity search, not only can predict the couple next future locations of nearby vessels but also reveals the future status of congested areas with vessels. The major contributions of this paper are twofold: (1) proposing a novel model to predict the future locations of target vessels in Euclidean space; and (2) proposing a novel method for predicting the future location of areas with a high density of vessels. This is fulfilled by defining a new concept called an expanded area that considers the probability of the advent of a vessel within the area of interest (AoI).

The remainder of this paper is organized as follows. Following this introduction, a review of the previous studies in location and trajectory prediction is provided. Our proposed model is described in afterwards. Next, the proposed model is experimentally evaluated by applying it on a real AIS dataset. In conclusion, the main findings are outlined and promising research topics are suggested for future work.

Related work
Analysing trajectories for prediction purposes is normally conducted by point-based or trajectory-based methods (Alizadeh, Alesheikh, and Sharif 2020). In the point-based methods, AoI is divided into independent cells. The probability of vessel entering to each cell is determined based on the vessel’s status and the extracted knowledge from the historical AIS data. Duca, Bacciu, and Marchetti (2017) divided the AoI into \( m \times n \) cells. In their proposed model, which was based on the K-nearest neighbour (kNN) classifier for training and prediction, the status of the vessel at the \( t \) time is inserted into the model as an input. Next, the probability matrices of vessel attendance after 30, 45, and 60 minutes were measured. Xiao et al. (2017) prediction model is based on the extracted pattern through lattice-assisted mining, and traffic was modelled by Kernel Density Estimation (KDE) technique. For extracting patterns, they performed the density-based spatial clustering of applications with noise (DBSCAN) algorithm (Ester et al. 1996) only on lattices that were not empty. The result showed improvements in the efficiency of knowledge extraction, storage, and retrieval operations. Wijaya and Nakamura (2013) retrieved the future location of a vessel based on the kNN algorithm that fulfilled the following requirements: (1) it had the same destination and nationality as the target vessel, and (2) it had similar course and speed to the target vessel. Then, the prediction of the next location was carried out based on the assumption that the distance between the target vessel and the nearest vessel to it was the same at the \( t \) and \( t + \Delta t \) times. Graser, Schmidt, and Widhalm (2018) applied three approaches of linear prediction, learned statistical model, and similarity search to predict vessels’ trajectories. The results indicated that the similarity search had better accuracy than the two other approaches. This research utilized the location prediction presented in (Wijaya and Nakamura 2013) to perform similarity search means. Kim and Lee (2018) introduced a deep neural network framework termed STENet for prediction of medium-term traffic and long-term traffic of the caution area. Both vessel movement data and vessel attribute data, which were extracted from AIS data, were inserted into the prediction model as the input. This model predicts the number of vessels that are located in the caution area in the future. Point-based approaches presume that the AIS messages are independent. Therefore, these methods ignore the possible increase in performance, which can be achieved by considering the spatial dependence between AIS messages related to each vessel. However, independent assumptions simplify the algorithm used to predict and analyse maritime traffic (Cazzanti and Pallotta 2015).

Trajectory-based approaches take the sequence of the AIS messages into account. To compose the trajectory of each vessel, messages with the same MMSIs are sorted in chronological order. Trajectory-based
approaches are normally implemented in four steps. In the first step, historical AIS data are clustered to extract the shipping route pattern. In the second step, the new trajectory is attributed to one of the obtained clusters based on the previous step. In the third step, the representative trajectory is designated for each cluster. Finally, the prediction is done according to the representative trajectory found in the previous step (Dalsnes et al. 2018). Gan et al. (2016) first grouped the historical AIS data by K-Means. Then, the clustering result obtained from the previous step is used for training an artificial neural network (ANN) to predict the vessel’s trajectory. They achieved 70% accuracy using this model. Nishizaki et al. (2018) predicted the next course of the vessels that are exited from Tokyo bay by training the support vector machine (SVM) model with the AIS data related to 5 and 10 minutes before exiting the bay. The achieved prediction accuracy for 5 and 10 minutes was 80.3% and 77%, respectively. Trajectory-based approaches can provide a richer knowledge by considering the spatial correlation between the AIS messages. Despite this advantage, they require more complex operations and precise bookkeeping in the first stages of the analysis. In other words, the techniques that are being used for extracting information from data are computationally expensive and need more time when dealing with trajectory-based models compared to the point-based models (Cazzanti and Pallotta 2015).

Based on the above information, our proposed location prediction method mitigates the computational cost of the previous methods and handles the nonlinearity of vessel movement. It is based on the assumption that if two moving objects have had similar pattern in the past, they will follow similar patterns in the future. Similar to the previous researches, we use movement parameters such as course and speed to measure the similarities between two moving objects. In contrast, we do not consider the vessel’s nationality and destination because these parameters are determined by the operators and can increase the probability of error. Moreover, our proposed framework consists of the location prediction phase to predict the next location of vessels in the Euclidean space and traffic prediction phase to predict the next locations of areas with a high density of vessels based on a new concept called expanded area.

Methodology

In this section, firstly, the proposed algorithms for location and traffic predictions are presented. As Figure 1 illustrates, after pre-processing raw AIS data, they are grouped based on their MMSI similarities for trajectory constitution. We have eliminated the trajectories that have less than 50 points because they are less informative in this work. The most similar point of the trajectory of the vessels to the target vessel is extracted from our constructed database. The location prediction algorithm utilizes the extracted point to predict future locations. Secondly, the AoI is expanded according to the prediction duration and maximum speed of the vessel. Furthermore, by applying the location prediction algorithm for vessels that are located inside the expanded area, traffic within the AoI is predicted.

Location prediction

When vessels enter or exit a harbour, they must follow a particular path according to their characteristics such

![Figure 1. The proposed location and traffic prediction models.](image-url)
as speed over ground (SOG), course over ground (COG), draft, and current location (Wijaya and Nakamura 2013). For instance, different vessels that are constrained by their draft require sufficient depth of water to move safely. Therefore, vessels with similar draft are highly likely to follow similar moving patterns in the specific maritime area. By analysing the behaviour of the most similar vessel to the target vessel, the future behaviour of the target vessel can be predicted. In the location prediction model, firstly, the most analogous vessel with the target vessel in terms of the elements outlined in the similarity section is retrieved from the databases. Secondly, the Euclidean distance between the target vessel and the extracted vessel is calculated. Thirdly, the next location of the target vessel is predicted by considering the location of the extracted vessel after duration prediction and assuming that the Euclidean distance remains constant.

**Similarity measurement**

For the similarity measure of trajectories, the distances between the elements of the first trajectory to the correspondent elements of the second trajectory are measured (Sharif and Alesheikh 2017; Sharif, Alesheikh, and Tashayo 2019). The elements are attributed to the movement data, movement parameters, and context information (e.g., latitude, longitude, altitude, mode of transportation, speed and heading, and environmental conditions) (Sharif and Alesheikh 2018) (Kaffash-Charandabi, Alesheikh, and Sharif 2019). To predict the next location, the most analogous vessel to the target vessel based on the selected elements is chosen. The selected elements for similarity measurement consist of latitude, longitude, ground speed, and ground course that shows the actual direction of navigation between the previous transmission and current transmission (Tsou 2019). The chosen elements are extracted from the AIS data. Spatial parameters of latitude and longitude are chosen as the criterion of similarity measurement according to the first law of geography that declares ‘everything is related to everything else, but near things are more related than distant things’ (Miller 2004), and the third law of geography that declares ‘the more similar geographic configurations of two points (areas), the more similar the values (processes) of the target variable at these two points (areas)’ (Zhu et al. 2018). Movement parameters of speed and course are selected under the logical assumption that the vessels that have similar courses and speeds are likely to reach similar locations after the duration prediction (Wijaya and Nakamura 2013). Three distance functions, namely, spatial distance measurement, directional distance measurement, and speed distance measurement are defined as follows.

- **Spatial distance measurement**: All coordinates inside the constructed database are converted into the UTM projection system to calculate the spatial distance between coordinates of a particular target vessel and the entire coordinates inside the database using the Euclidean distance function (Equation (1)).

  \[ D_s = \sqrt{(X_{TP} - X_{DB})^2 + (Y_{TP} - Y_{DB})^2} \]  

  where \( X_{TP} \) and \( Y_{TP} \) are the UTM coordinates of the target vessel, and \( X_{DB} \) and \( Y_{DB} \) are the UTM coordinates of each AIS data stored in the database. \( D_s \) represents the spatial distance matrix, where each element of the matrix indicates the spatial distance between a particular target vessel with each coordinate of the AIS data in the database. The smaller each element of the \( D_s \), the higher the spatial similarity between the certain target vessel and the AIS coordinate is.

- **Speed distance measurement**: Each element of the speed matrix represents the difference between the speed of a certain target vessel and the speed of each AIS data within the database. The speed matrix is calculated by Equation (2),

  \[ D_v = |Sog_{TP} - Sog_{DB}| \]  

  where \( Sog_{TP} \) is the speed of the last point of the target trajectory and \( Sog_{DB} \) is the speed of each point of trajectories within our constructed database. \( D_v \) demonstrates the speed matrix that each element of the matrix is computed from the absolute difference between the speed of a certain target vessel and the speed of a particular AIS data in the database. The smaller the element of \( D_v \), the higher the speed similarity between the certain target vessel and the AIS speed is.

- **Course distance measurement**: AIS data are comprised of the COG and heading columns. Heading implies where the vessel is pointing, whereas COG is the actual direction of the vessel compared to the north. In this paper, we utilize COG instead of heading because it indicates the actual direction that a vessel has navigated. The course matrix is determined by Equation (3),

  \[ D_d = |Cog_{TP} - Cog_{DB}| \]  

  where \( Cog_{TP} \) is the course of a specific target vessel, and \( Cog_{DB} \) is the course of each AIS data in the database. \( D_d \) represents the course matrix that each element of the matrix is computed from the absolute difference between the course of a specific target vessel and course of a particular AIS data inside the database. The bigger
the element of $D_d$, the lower course similarity between the specific target vessel and the AIS course is $D_c$, $D_s$, and $D_d$ distances are normalized to the scale of 1 using Equation (4),

$$D_{\text{norm}} = \frac{D - D_{\min}}{D_{\max} - D_{\min}} \quad (4)$$

where $D$ is the normalized distance, the $D_{\max}$ and $D_{\min}$ are, respectively, the maximum and minimum values of each spatial, directional, and speed similarity distances. Consequently, the final similarity measurement function is achieved by using the overall weighted distance measurement of the three spatial, speed, and course functions using Equation (5).

$$DSM = Ws \times Dns + Wv \times Dnv + Wd \times Dnd \quad (5)$$

where $D_{\text{ns}}$, $D_{\text{nv}}$, and $D_{\text{nd}}$ stand for the normalized spatial, speed, and course distances, respectively, $W_s$, $W_v$, and $W_d$ are their weights, and $W_s + W_v + W_d = 1$. $DSM$ represents the total distance between a specific target point and each AIS data in the database in terms of spatial, speed, and course elements. The smaller the $DSM$ the higher similarity between the target vessel and the particular AIS data is.

**Location prediction model**

After retrieving the most similar AIS data within the database to the target vessel, the retrieved point, which belongs to the most similar vessel, is utilized to predict the future location of the target vessel as is illustrated in Figure 2.

In Figure 2, $A0$ stands for the location of the target point, and $B0$ represents the retrieved point from the AIS database through Equation (5), which is the most similar to the target point in terms of spatial, speed, and course distances. $d$ is the Euclidean distance between the target vessel and the extracted point. $A1$ and $B1$, respectively, illustrate the future locations of the target point and extracted point within prediction duration. $g_{B0A0}$ indicates the bearing angle that is calculated using the coordinates of $A0$ and $B0$ (Equation (6)).

$$g_{B0A0} = \tan^{-1} \frac{|x_{t0} - x_{r0}|}{|y_{t0} - y_{r0}|} \quad (6)$$

where $g_{B0A0}$ is the bearing angle between two points of $A0$ and $B0$ as illustrated in Figure 2. $(x_{r0}, y_{r0})$ and $(x_{t0}, y_{t0})$ represent the coordinates of $A0$ and $B0$ at time $t_0$, respectively. $G$ denotes the Azimuth angle that is defined as the angle between a specific direction and the grid north and is calculated using the bearing angle in Equation (7).

$$G_{B0A0} = \begin{cases} g_{B0A0} & \text{if } \Delta X > 0 \text{ and } \Delta Y > 0 \\ 180 - g_{B0A0} & \text{if } \Delta X > 0 \text{ and } \Delta Y < 0 \\ 360 - g_{B0A0} & \text{if } \Delta X < 0 \text{ and } \Delta Y > 0 \\ 180 + g_{B0A0} & \text{if } \Delta X < 0 \text{ and } \Delta Y < 0 \end{cases} \quad (7)$$

where $G_{B0A0}$ stands for the Azimuth angle of $B0A0$ direction at time $t_0$. After calculating $d$ and $G$, the future location of the target vessel is calculated using Equation (8) based on the assumption that $G$ and $d$ remain constant within prediction duration. The location of the target vessel ($A1$) at time $t_1$ will be predicted as follows.

**Figure 2.** The proposed location prediction model.
\[ \begin{align*}
X_{t_1} &= X_{t_0} + d \times \sin G_{BA1} \\
Y_{t_1} &= Y_{t_0} + d \times \cos G_{BA1}
\end{align*} \tag{8} \]

where \((X_{t_1}, Y_{t_1})\) denotes the future coordinate of the target vessel \((A1)\) at time \(t_1\) and \((X_{t_1}, Y_{t_1})\) represents the future coordinate of the extracted vessel \((B1)\) at time \(t_1\). \(G_{BA1}\), which will be equal to \(G_{BAO}\), stands for the Azimuth angle of \(BA1\) direction, and \(d\) is the Euclidean distance.

**Traffic prediction**

Marine traffic congestion in straits and main harbours varies in particular times. To avoid collision and to improve mobility management, it is crucial to predict future traffic in these areas. Changes in the volume of traffic are mainly caused by three factors. First, the vessels that are currently located in the Aol and will navigate towards the Aol during the prediction time. Second, the vessels that are currently located inside Aol and will exit from the Aol during the prediction time. Third, the vessels that are not currently located in the Aol; however, they are highly likely to enter the Aol. By applying the proposed location prediction method on vessels that are currently located inside the Aol, only maritime traffic obtained from the first and second factors will be achieved. Therefore, it is essential to define a bigger area around the Aol, which is named an expanded area, to take the traffic affected by the third factor into account. In other words, by applying the location prediction method on vessels that are located in the expanded area instead of the Aol, the possibility of entering new vessels inside the Aol (third factor) will be considered. The size of the expanded area depends on the maximum speed of the vessel and the prediction duration. For example, for traffic prediction in the Aol after 10 minutes and under the assumption that the maximum speed of the vessels will be 20 knots, the Aol area will be expanded to 6174 metres based on Equation (9),

\[ \Delta x = \Delta t \times v \tag{9} \]

where \(\Delta x\) is the maximum movement, \(\Delta t\) represents the prediction duration, and \(v\) shows the vessel’s maximum speed. The maximum movement that a vessel can travel during the prediction time determines the dimension of the expanded area. Therefore, the maximum speed of the

**Figure 3.** ABCD is the defined polygon as Aol, EFGH is the defined polygon for \(\Delta t = 10\) minutes, IJKL is the defined polygon for \(\Delta t = 20\) minutes, and MNOP is the defined polygon for \(\Delta t = 30\) minutes.
vessel has been used in Equation (9). By the location prediction of the vessels that are located in the expanded area for Δt = 10 minutes, the probability of entering vessels that are around the AoI is taken into account. Likewise, the expanded area is created around the AoI for prediction duration (Δt) as depicted in Figure 3. To predict the traffic inside the AoI after Δt, it is required to predict the location of vessels that are located within the corresponding expanded area.

Model evaluation

Location prediction evaluation

To evaluate the location prediction model, scatter plots with R² values are generated to show to what degree the predicted coordinates are close to the actual coordinates. In this context, R² is calculated for the Xs and Ys, and the coordinates of predicted values and the actual values. R² is between the 0 and 1, where 1 shows that the predicted and actual values are the same and 0 shows that there is no correspondence between the actual and predicted values.

Traffic prediction evaluation

To evaluate the traffic prediction model, the AoI has been divided into cells with equal size. Then, the number of vessels that will be predicted to be located in each cell will be count. In other words, the number of vessels that will be located inside each cell of the AoI after specific time intervals will be predicted by performing the proposed location prediction model on vessels that are located in the expanded area. Eventually, the actual number of vessels in each cell will be compared with the predicted number of vessels in each cell as the evaluation of our proposed traffic prediction model.

The AoI is considered as a matrix such that each cell of AoI corresponds to an element of the matrix, each element of the matrix represents the number of located vessels in its corresponding cell in the AoI. Therefore, each prediction duration has two matrices. One matrix is the prediction matrix that each element demonstrates the predicted number of the located vessels within its corresponding cell in the AoI. The other matrix is the actual matrix that each element shows the predicted number of the located vessels inside its corresponding cell in the AoI. The model is evaluated by using the Sorenson similarity index (Yan et al. 2014). The value of this index is between 0 and 1. The similarity index of 1 shows that both predicted and actual matrices are the same. SSI has been used to measure to what degree the elements of two matrices are similar (Equation (10)) (Kang et al. 2015).

\begin{equation}
SSI = \frac{2 \sum_{ij} \min(T_{ij}^{data}, T_{ij}^{model})}{\sum_{ij} T_{ij}^{data} + \sum_{ij} T_{ij}^{model}}
\end{equation}

where \( T_{ij}^{data} \) represents the actual matrix whose elements show the actual number of vessels that are located in each cell of the AoI. \( T_{ij}^{model} \) represents the predicted matrix whose elements indicate the numbers of vessels that have been predicted to be located in each cell of the AoI. In Equation (10), \( \sum_{ij} T_{ij}^{data} \) and \( \sum_{ij} T_{ij}^{model} \) are the total values of elements of actual and predicted matrices, respectively.

Analysis and results

To evaluate the applicability of the proposed location and traffic prediction models, they are separately implemented on a real AIS dataset. Regression analysis and Sorenson similarity index (SSI) are used to evaluate the proposed location and traffic prediction models, respectively.

Case study

The Strait of Georgia, as one the most important and congested harbours in the west of the United States, is

Figure 4. The study area: Strait of Georgia, USA.
selected as the study area. The AIS dataset belongs to UTM Zone 10 N and dated back to February 2017. To predict traffic in the Aol, data with x values between 460,000 and 500,000 metres and y values between 5,400,000 and 5,500,000 metres are extracted from the whole dataset (Figure 4).

**Data pre-processing**

The whole dataset contains 3,354 number of trajectories (Marinecadastre), which are composed of 26,745,371 sampling points. The vessels with zero speed have been eliminated from the dataset. In addition, AIS data are composed of various columns that only MMSI, Timestamp, Latitude, Longitude, SOG, and COS columns are preserved. The number of sampling points is reduced to 9,888,562, after the pre-processing step.

**Results**

To demonstrate the distribution and hot spots (locations of the area with a high density of vessels) inside the Aol, the region is divided into cells with equal sizes (1379 × 3448 metres). The location prediction method is performed on vessels that are located in the expanded area and the next locations of them have been predicted. The number of predicted vessels that will be located within each cell in the future time has been used as an indicator to show the density of each cell. Figure 5 illustrates the results of the traffic prediction procedure in the Aol. As can be seen, the volume of traffic over time is changing. The major congested area within the Aol has been located in the North East of the Aol where is adjacent to Vancouver harbour. In addition, hot spots are shifted within a certain period. As Figure 5 shows, the predicted hot spots during the prediction time intervals tend to move towards the west direction of the Aol where the existence of Vancouver harbour is located. The Aol is divided into cells that have been coloured based on the number of vessels in each cell so that the greater the number of vessels in the cells results in brighter the colour of the cells. Yellow colour represents a high density of vessels in the cell and the dark blue represents no vessel in the cell. The result shows the feasibility of the proposed models in hot spot prediction that can lead to better monitoring and management of vessels in any Aol.

**Evaluation**

The applicability of the proposed location and traffic prediction models are assessed in this section.

![Figure 5](image-url) Traffic prediction within the Aol. (a) actual traffic at 6:00 AM, (b) predicted traffic at 6:10, (c) predicted traffic at 6:20, (d) predicted traffic at 6:30 on 18 February 2017.
**Location prediction evaluation**

The whole vessels that are located inside the expanded areas that correspond to a particular prediction duration are predicted. Table 1 indicates that as the prediction duration increases, the values of the $R^2$ decreases. In addition, the value of $R^2$ in the Y direction is greater than the X direction for all the prediction time intervals. The reason is that vessels were moving more in the X direction rather than the Y direction within the AoI.

In every six diagrams in Figure 6, the horizontal axis represents the actual values and the vertical axis represents the predicted values. The plotted red line is the $Y = X$ diagram that represents a diagonal line. The closest points to the plotted red line denote more accuracy of prediction result.

**Traffic prediction evaluation**

Table 2 demonstrates that SSIs' descending trends over time. In addition, SSIs are affected by cell sizes. The relationship between the SSI values and the cell sizes is illustrated in Figure 7. The horizontal axis represents the time and the vertical axis represents the SSI in percentage. The breaking point occurred after approximately 20 minutes, which shows that by increasing the prediction time intervals, the impact of cell size on accuracy decreases. It can be inferred that the large cell size of AoI will increase the accuracy obtained by SSI.

### Table 1. The values of $R^2$.

| $\Delta t$ (min) | $R^2$ (X) | $R^2$ (y) | $R^2$ (Coordinates) |
|------------------|-----------|-----------|---------------------|
| 10               | 0.9957    | 0.9996    | 0.9976              |
| 20               | 0.9808    | 0.9967    | 0.9887              |
| 30               | 0.9663    | 0.9924    | 0.9794              |

**Figure 6.** Scatter plots show the correlation between the predicted and actual values. (a), (b), and (c) show the correlation of X after 10, 20, and 30 minutes, respectively. (d), (e), and (f) show the correlation of Y after 10, 20, and 30 minutes, respectively.
From the accuracy perspective, the Euclidean distance has been calculated as the prediction error between the predicted and the actual locations. Comparing to the similar researches (Wijaya and Nakamura 2013; Graser, Schmidt, and Widhalm 2018) that used point-based approaches, this study conducted a longer-term prediction experiment (i.e., 20 and 30 minutes), which is required in vessel monitoring and collision avoidance and anomaly detection to get more time for decision-makers to control activities on vessels movement. Moreover, the proposed framework can predict the next location of areas with a high density of vessels, which is needed for increasing the efficiency of port operations. In addition, the experimental results show that our proposed model has improved the prediction accuracy by almost 320 metres for the prediction time intervals of 10 minutes compared to the previous studies.

It is evident that comparing larger numbers of trajectories may result in more computational cost. In order to reduce the computation cost, we provided an approach to consider the most effective sampling points and their corresponding parameters in trajectory analysis and location prediction. In this context, the most similar point to the target point is retrieved from the database and the proposed model only applies to the last point of trajectories. This approach certainly reduces the amount of computation. In terms of computational load, the experiments were run on a computer with the characteristics of 64-bit operating system, 6 GB RAM, and 2.30 GHz CPU. The next location (10 minutes) of a given vessel was achieved less than one minute. Such fast computation helps the decision-makers to have sufficient time to make their decision consciously and avoid any collision.

**Conclusion**

This paper addressed a novel point-based model for discovering vessels’ locations and predicting maritime traffic based on movement data and parameters obtained from streaming AIS messages. Unlike the conventional traffic prediction models that are probabilistic and are based on the network of communication routes, the traffic prediction in this study resulted from the location prediction of vessels that was located within the respective expanded area of the prediction duration. The expanded areas, which were designed based on the maximum speed of the vessels and prediction duration, enabled considering the vessels that were outside the AoI and were likely to enter inside the AoI within the prediction duration. High accuracy in the location and

### Table 2. Values of the Sorenson similarity index.

| Cell size | Δt = 10 | Δt = 20 | Δt = 30 |
|-----------|---------|---------|---------|
| 1379 × 3448 | 72%    | 57%    | 48%    |
| 2105 × 5263 | 70%    | 66%    | 59%    |
| 2857 × 7142 | 79%    | 68%    | 68%    |

![Figure 7](image-url) The relationship between SSI% and cell size over time.
traffic prediction results demonstrated the robustness of these models in predicting the vessels in congested areas such as straits and ports.

Although the proposed models can effectively recognize vessels’ locations and examine maritime traffic, it assumed that the distance between the target vessel and the most similar vessel to it is constant. Meanwhile, the angle between distance directions and north direction was preserved constant too. As future works, we suggest executing trajectory-based similarity measures to achieve a comprehensive conception about the commonalities of the target vessel to other vessels. In addition, a promising research study can consider the effective environmental contexts in the sea (e.g., wind and wave) along with the AIS data in movement prediction. It is hoped that the findings of this research improve the safety of maritime navigation and securities of the seas, promote the port’s efficiency, and avoid vessel accident.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Ali Asghar Alesheikh https://orcid.org/0000-0001-9537-9401
Mohammad Sharif https://orcid.org/0000-0002-1241-007X

References

Alizadeh, D., A. A. Alesheikh, and M. Sharif. 2020. “Vessel Trajectory Prediction Using Historical Automatic Identification System Data.” Journal of Navigation 1–19. doi:10.1017/S0373463320000442.

Asahara, A., K. Maruyama, A. Sato, and K. Seto. 2011. “Pedestrian-movement Prediction Based on Mixed Markov-chain Model.” Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems, Chicago, IL, USA. 25–33.

Borkowski, P. 2017. “The Ship Movement Trajectory Prediction Algorithm Using Navigational Data Fusion.” Sensors 17 (6): 1432. doi:10.3390/s17061432.

Cazzanti, L., and G. Pallotta. 2015. “Mining Maritime Vessel Traffic: Promises, Challenges, Techniques.” OCEANS 2015-Genova. Genoa, Italy: IEEE.

Dalsnes, B. R., S. Hexeberg, A. L. Flåten, B.-O. H. Eriksen, and E. F. Brekke 2018. “The Neighbor Course Distribution Method with Gaussian Mixture Models for AIS-based Vessel Trajectory Prediction.” 21st International Conference on Information Fusion (FUSION). Cambridge, UK: IEEE.

Dobrkovic, A., M.-E. Iacob, and J. Van Hillegersberg 2015. “Using Machine Learning for Unsupervised Maritime Waypoint Discovery from Streaming AIS Data.” Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business. Graz, Austria: ACM.

Dobrkovic, A., M.-E. Iacob, J. Van Hillegersberg, M. R. Mes, and M. Glandrup 2016. “Towards an Approach for Long Term AIS-based Prediction of Vessel Arrival Times.” In: Zijm H., Klumpp M., Clausen U., Hompel M. (eds.), Logistics and Supply Chain Innovation. Lecture Notes in Logistics. Springer, Cham.

Duca, A. L., C. Bacciu, and A. Marchetti. 2017. “A K-nearest Neighbor Classifier for Ship Route Prediction.” OCEANS 2017-Aberdeen. Aberdeen, UK: IEEE.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” Kdd 96: 226–231.

Fujino, I., C. Claramunt, and A.-O. Boudraa. 2018. “Extracting Courses of Vessels from AIS Data and Real-Time Warning against Off-Course.” Proceedings of the 2nd International Conference on Big Data Research. Weihai, China: ACM.

Gan, S., S. Liang, K. Li, J. Deng, and T. Cheng. 2016. “Ship Trajectory Prediction for Intelligent Traffic Management Using Clustering and ANN.” UKACC 11th International Conference on Control (CONTROL). Belfast, Northern Ireland: IEEE.

Gao, M., G. Shi, and S. Li. 2018. “Online Prediction of Ship Behavior with Automatic Identification System Sensor Data Using Bidirectional Long Short-Term Memory Recurrent Neural Network.” Sensors 18 (12): 4211. doi:10.3390/s18122411.

Graser, A., J. Schmidt, and P. Widhalm. 2018. “Predicting Trajectories with Probabilistic Time Geography and Massive Unconstrained Movement Data.” Workshop on Analysis of Movement Data AMD2018 in conjunction with GIScience 2018. Melbourne, Australia.

Grech, M. R., T. Horberry, and A. Smith. 2002. “Human Error in Maritime Operations: Analyses of Accident Reports Using the Leximancer Tool.” Proceedings of the human factors and ergonomics society annual meeting. Sage Publications Sage CA: Los Angeles, CA.

Harati-Mokhtari, A., A. Wall, P. Brooks, and J. Wang. 2007. “Automatic Identification System (AIS): Data Reliability and Human Error Implications.” The Journal of Navigation 60 (3): 373–389. doi:10.1017/S0373463307004298.

Kaffash-Charandabi, N., A. A. Alesheikh, and M. Sharif. 2019. “A Ubiquitous Asthma Monitoring Framework Based on Ambient Air Pollutants and Individuals’ Contexts.” Environmental Science and Pollution Research 26 (8): 7525–7539. doi:10.1007/s11356-019-04185-3.

Kang, C., Y. Liu, D. Guo, and K. Qin. 2015. “A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint.” PloS One 10 (11): e0143500. doi:10.1371/journal.pone.0143500.

Kim, K.-I., and K. Lee. 2018. “Deep Learning-based Caution Area Traffic Prediction with Automatic Identification System Sensor Data.” Sensors 18 (9): 3172. doi:10.3390/s18093172.

Mao, S., E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang 2018. “An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining.” Proceedings of ELM-2016. Springer. Marinecadastre. Available: www.marinecadastre.gov 20 December 2019.

Mathew, W., R. Raposo, and B. Martins. 2012. “Predicting Future Locations with Hidden Markov Models.” In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA 911–918.
