Long-term Ship Position Prediction using Automatic Identification System (AIS) and Deep Learning

(AIS データとディープラーニングを用いた船舶の長期的位置予測に関する研究)

Ibadurrahman

D 190785

DISSERTATION

Submitted to the Graduate School of Engineering, Hiroshima University

September 2022

Higashi-Hiroshima

Abstract

Predicting a ship's long-term position can be game-changing in maritime industries. Ship position prediction for maritime situational awareness (MSA), as a critical aspect of maritime safety and security, requires a longer time interval than collision avoidance and maritime traffic monitoring. However, previous studies focused mainly on shorter time-interval predictions ranging from 30 min to 10 h. A longer time-interval ship position prediction is required for MSA and efficient allocation of ships or business objectives. This study used an end-to-end tracking method that inputs the previous position of a vessel to a trained deep learning model to predict its next position with an average 24-h interval, up to the 96-h interval. An AIS dataset with a long-time-interval distribution in nine years for Capesize bulk carriers worldwide was used.

The dissertation is constructed from two main parts of the study: each consists of three experiments followed by constructing discussions. From the former, a deep learning model of the Indian Ocean was examined, and subsequently, the model performance was compared for six different oceans and six primary maritime chokepoints. In the third experiment, a sample location within the Malacca Strait area was selected, and the number of ships was counted daily. The results indicate that the ship position can be predicted accurately with an average time interval of 24 h using deep learning systems with AIS data. From the latter, larger networks of DL models were trained on different observation area sizes with magnified time intervals. Then, more observation size variations were conducted in the next experiment, increasing 4 and 3.2 degrees on longitude and latitude, respectively. Finally, in the last experiment, five DL models were built for five major locations of maritime shipping routes focused on Capesize bulk carriers. The result reveals that the locations where importers of bulk shipments constitute the largest portion appear to be harder to predict than those that constitute bulk export regions.

Keywords: deep learning; AIS; ship position prediction; long-term; end-to-end

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1. The Importance of Long-term Prediction of Ship Position

Maritime transportation is recognized for its central role in the global supply chain considering it accounts for 90% of international trade by volume and 70% by value [1]. The United Nations predicted that the total volume of seaborne trade worldwide would increase by 3.2% from 2019 to 2022 [2]. Therefore, establishing safety and security in maritime transportation is essential. Maritime situational awareness (MSA) is a critical aspect of maritime safety and security that can be achieved through tracking, surveillance, and position prediction of ships [3]. Once a prediction of the ship position is obtained, decision making and action planning can be supported at different information levels. However, ship position prediction for MSA requires a longer time interval than other tasks such as collision avoidance and maritime traffic monitoring, which mainly use a short-term prediction from a high-precision real-time forecast spanning less than one hour [4]. Accordingly, we considered a prediction with a time interval Δt longer than 12 h as a long-term prediction, and a prediction with a time interval between short- and long-term threshold as a medium-term prediction [5].

Studies on the long-term prediction of vessel position remain scarce despite its considerable potential for maritime applications, while almost all previous studies have focused on either near real-time predictions (short-term prediction) or predictions with a time interval lesser than 12 h (medium-term predictions). Long-term ship position prediction is required not only for MSA but also for efficient allocation of ships by shipping companies in accordance with global freight demand. It can be utilized to monitor and assist the fleet, specifically, when communication with a ship operator breaks down owing to poor weather conditions or when a ship is in distress. It can also be implemented by shipping insurance and maritime investigators for investigation purposes. The ability to predict the long-term position of a fleet could prove not merely necessary but vital for strategy formulation in the fast-changing dynamics of maritime industries.

Maximizing the potential of maritime big data is essential for predicting vessel positions. The automatic identification system (AIS) is a self-reporting message system on board a vessel that records its position and condition [6]. Each record of the AIS message contains the static and voyage-related information of the vessel and its dynamic information such as longitude,

latitude, speed over ground, course over ground, and heading. As of July 2008, all commercial vessels above 300 GT serving international routes were regulated to be outfitted with an AIS Class A device by the International Maritime Organization (IMO) [7]. Prior to the widespread use of maritime big data (i.e., AIS data), studies on vessel position prediction were conducted using data from radar or laser sensors, such as in the work of Perera et al. (2012) [8]. Since 2015, studies in the field have started using AIS as a historical data source for ship position information. Czapiewska and Sadowski (2015) conducted position prediction experiments using linear and nonlinear motion functions for location data compression of AIS records [9]. The high volume of AIS data accumulated over the years is a potential asset that needs to be explored, especially in the current age of artificial intelligence. With this big data of vessel position records, it is possible to predict vessel position by applying advanced machine learning (ML) techniques such as deep learning.

1.2. Objective of this Study

This study aims to achieve a 24 h interval and more ship position prediction for MSA. Predicting a ship's long-term position can be game-changing in maritime industries. The application of the long-term position prediction may not be the end goal. The results can be fed to another machine learning system, along with many other signals. Then, a business objective of the machine learning pipeline can be established broadly, such as for ship allocation, shipping investment, or global economic projection. For instance, with the objective of shipping investment, this downstream system will determine whether it is worth investing in a certain area and future time or not by predicting direct components to the measurement of the Baltic Dry Index (BDI), which has a predictive ability for a range of stock markets [10]. The accurate results of the application are critical, as it directly affects revenue. This work provides a valuable benchmark for future studies.

Previous studies focused mainly on short- and medium-term prediction, ranging from 30 min to 10 h intervals as the AIS data were dense (closely packed between short time intervals) but in a limited period or timespan. Naturally, a vessel position prediction with long time intervals (e.g., 24 h) requires data with a long timespan, given that state-of-the-art ML algorithms, such as deep learning, require large amounts of data to make accurate predictions for longer intervals [11]. Accordingly, nine-year AIS data for capesize bulk carriers worldwide and deep learning (DL) were used to accomplish long-term prediction. End-to-end learning refers to training a learning system represented by a single deep network model that supersedes

the preprocessing stages typically present in traditional pipeline designs. By using extensive fleet data and computing power, the model can learn robustness to noise and generalization to data variation; this result confirms that an accurate long-term prediction of vessel position can be generated with the straightforward method. The utilization of nine years of AIS data worldwide and the development of a generalized DL model from an uneven time-interval dataset for the long-term prediction of ship positions constitute the novelty of this study.

1.3. Organization of this Dissertation

Each chapter in this thesis is organized as follows.

Chapter 1 overviews the long-term prediction of ship position and its importance. The objective of this study is also presented.

In **Chapter 2**, studies on vessel position prediction are compiled and described based on the prediction methods. The different characteristics between trajectory-based and motion-based methods are also described. The characteristics of this study are further discussed.

Chapter 3 outlines the basic concept and data used in this study. The AIS and ship data used are described. The basic concept of the deep learning model for position prediction is also explained.

In **Chapter 4**, the first part of the research, a one-day interval of ship position prediction is carried out. In the first experiment, the model performance was demonstrated for the Indian Ocean, followed by a comparison of the performance for six different oceans and six primary maritime chokepoints to investigate the influence of each area. Finally, the model performance for a sample area within the Malacca Strait was analyzed to simulate a practical application.

In **Chapter 5**, as the second part of the research, the time interval prediction was magnified up to several days; a larger network was proposed where the previous model was used as the baseline, and the observation coverage was also enlarged. DL models were trained on different observation area sizes with magnified time intervals in the first experiment. More observation size variations were conducted in the subsequent experiment, increasing 4 and 3.2 degrees on longitude and latitude, respectively. In the last experiment, five DL models were built for five major locations of maritime shipping routes focused on Capesize bulk carriers.

Finally, **Chapter 6** presents conclusions, and some future works are proposed.

The structure of this thesis is depicted in Figure 1-1.

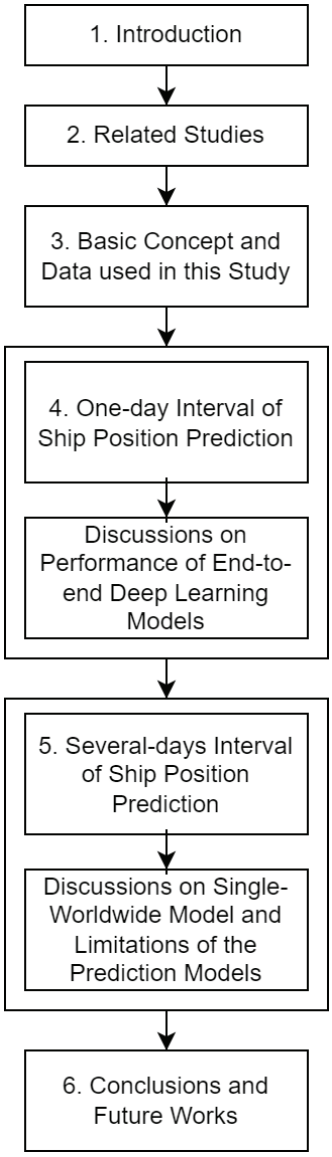


Figure 1-1. Structure of the dissertation

Chapter 2

Related Studies

2.1. Methods of Vessel Position Prediction

The methods for forecasting vessel positions can be classified into three categories: trajectory-based, point-based, and motion-based methods. Overall, studies concerning ship position prediction are summarized in Table 2-1; only studies focused on predicting vessel position are listed. The results of one method are not equivalent to those of another method because each method is specifically developed for the desired objective and the data used.

Table 2-1. Vessel position prediction studies

Prediction Method[1]	Authors	Δt Threshold	Δt Prediction	Objective[2]	AIS Data (Range)	ML Algorithm[3]	Target Vessel[4]	Target Area[5]	Unrestricted Trajectory
T-b	[18]	Medium	8 h	MSA	5 m	-	Cg,Tg,Tk	OW	Yes
T-b	[20]	Medium	10 h	MSA	1 m	-	Cg	OW	No
T-b	[17]	Short	15 min	CA	1 y	-	*nr*	RW	Yes
T-b	[21]	Short	5 min	MTM	2 m	kNN	*nr*	RW	No
T-b	[19]	Medium	1 h	MSA	*nr*	ELM	*nr*	*nr*	Yes
P-b	[23]	Medium	1 h	MTM	1 m	kNN	Fs,Cg,Tk	RW	Yes
P-b	[24]	Short	50 min	MTM	2 y	CNN	Cg, Tk	RW	Yes
P-b	[25]	Short	5 min	MTM	1 m	CNN,LSTM	*nr*	RW	Yes
M-b	[8]	Short	< 1 min	MTM	-	-	*nr*	RW	No
M-b	[9]	Short	5 min	DC	*nr*	-	Cg	RW	No
M-b	[28]	Short	40 min	AD,MSA	1 m	ELM	*nr*	OW	No
M-b	[27]	Short	8 min	MTM	3 m	-	*nr*	RW	Yes
M-b	[30]	Short	3 min	MTM	-	MLP	Tk	RW	No
M-b	[32]	Short	15 min	CA	*nr*	MLP	Fe	*nr*	No
M-b	[31]	Medium	4 h	MSA	*nr*	MLP	Ps	RW	No
M-b	[29]	Short	20 min	CA,MSA	*nr*	MLP	*nr*	RW	Yes
M-b	[33]	Short	10 min	CA	1 y	bLSTM	*nr*	RW	Yes
M-b	[34]	Short	< 1 min	CA	*nr*	LSTM	Fe	RW	No
M-b	this study	Long	24 h	MSA,SA	9 y	MLP	BC	OW	Yes

[1] Prediction methods: T-b, trajectory-based; P-b, point-based; M-b, motion-based
[2] Prediction objectives: CA, collision avoidance; MSA, maritime situational awareness; SA., ship allocation; DC, data compression; MTM, maritime traffic monitoring; AD, anomaly detection.
[3] ML algorithm or DL architecture used for vessel position prediction: SVM, support vector machine; ELM, extreme learning machine; kNN, k-nearest neighbors; MLP, multilayer perceptron; RNN, recurrent neural network; LSTM, long short-term memory; CNN, convolutional neural network; bLSTM, bidirectional long short-term memory.
[4] Target vessel type: Cg, cargo, Tk, tanker, Tg, tugboat, BC, bulk carrier, Fs, fishing vessel, Fe, ferry, Ps, passenger ship.
[5] Target area: OW, open water; RW, restricted water.
nr, not reported.

2.1.1. Trajectory-based Method

In the trajectory-based prediction method, clusters of vessel trajectories must be initially created and classified based on shipping routes from historical voyage patterns [12]. A *trajectory* is defined as a structure of polygon or matrix that forms one shipping route, created from AIS data points of particular ships or vessels entering a coverage area. A waypoint detection is utilized to detect new or existing routes once vessels enter the coverage. Trajectories can also be decomposed from a cluster or classification of AIS data points based on dense points in shipping routes or other criteria from all historical voyage patterns inside an examined area [13]. The route decomposition extracts similar vessel behaviors into a cluster; to accurately estimate vessel trajectory, different features are identified to decompose it into different sub-routes [14]. Figure 2-1 demonstrated a sample of trajectories extracted from a trajectory deonposition of AIS data points in the North Adriatic Sea.

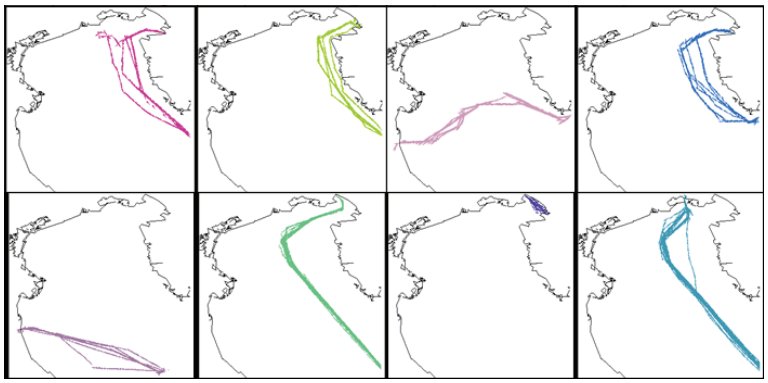


Figure 2-1. An example of decomposed highly dense routes extracted from historical vessel voyages in the North Adriatic Sea [3]

As a set of trajectories creates a trajectory database, each trajectory must have a similar structure. It generally undergoes a trajectory reconstruction such as interpolation or other means (see Figure 2-2 for an example of trajectory reconstruction by interpolation AIS data points). The ship position that is wanted to be predicted is first matched to the database based on its historical trajectory. The reconstruction of trajectories requires intensive preprocessing regarding the varying density of data points, abnormal behavior, trajectory anomaly, noise, missing points, the assumption regarding status, and the trade-off between similarity thresholds [15, 16].

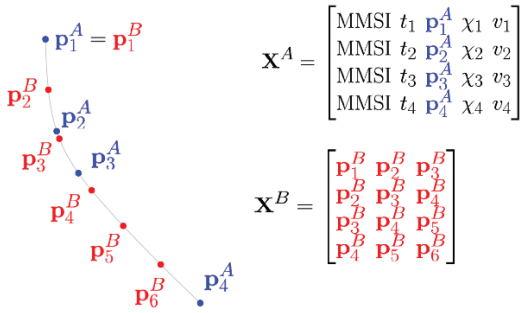


Figure 2-2. An example of short trajectory reconstruction by interpolation represented with both data structures: the original AIS data (blue), and the new interpolated points (red) [17]

A position prediction can then be derived by formulating or interpolating the trajectory structure. Because vessels' entire trajectories (such as port-to-port trajectories) are required in an observed area, this method has been implemented mainly for large areas (e.g., open waters, oceans, and straits).

Several studies have developed unique route extraction algorithms to be used before the vessel position is predicted by deriving from its routes, such as in the work of Pallotta et al. (2014), Tu et al. (2020), and Mazzarella et al. (2015) [18-20]. We considered Δt of Mazarella et al. (2016) to be effective for only 10 h because they merely demonstrated their trajectory-based algorithm on a particular target and route, where the 10 h threshold exhibited the most significant performance compared to those of the other algorithms. Dalsnes et al. (2018) conducted predictions using a Gaussian mixture model combined with the neighbor course distribution method [17]. Using k-nearest neighbor (kNN), Virjonen et al. (2018) compared the trajectories of a new ship and historical ships within the Gulf of Finland [21].

2.1.2. Point-based Method

In the point-based method, the examined area is first transformed into non-overlapping cells or grids. The position of a ship inside the area is converted into a grid reference and defined by the occupied cells. Once a vessel enters the coverage, its movement is captured by these occupied cells. Figure 2-3 illustrates a grid ranges definition to be tested on a selected rectangular shipping area. As a fixed time interval must be maintained inside the observed coverage, the AIS points or trajectories must undergo a trajectory reconstruction such as interpolation or other means. An ML model then calculates the prediction of the next position of the ship, trained from historical movement data inside the grid. Some researchers used CNNs by converting vessel positions from AIS data into a gridded ocean floor model as the input.

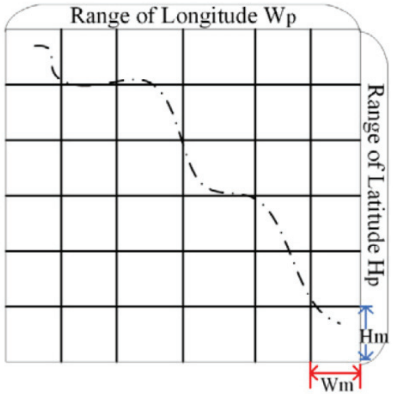


Figure 2-3. An example of a grid ranges definition set on a designated area [22]

Several studies have utilized this method for vessel position prediction for various objectives. Duca et al. (2017) employed a kNN classifier for ship route prediction [23]. Kim et al. (2018) employed a combined hierarchical architecture of a convolutional neural network (CNN) for feature extraction with five separate fully connected neural networks (NNs) to predict the number of ships around the Korean port Yeosu up to 50 min ahead with initial AIS points reconstructed by interpolation [24]. Zhou et al. (2020) used three different architectures, including CNN, long short-term memory (LSTM), and bi-directional LSTM (bLSTM) integrated with CNN, to predict the inflow and outflow of gridded Singapore waters [25].

2.1.3. Motion-based Method

The motion-based method estimates the future vessel position using motion functions from the current geographical position and movements (see Figure 2-4) or a trained ML model using historical information as input. Table 2-2 shows the main differences between trajectory-based and motion-based methods for vessel position prediction. This method is relatively more flexible than the other methods and can be effectively generalized for data variation. Most researchers have used this method for short- and medium-term predictions within a small area (e.g., near ports, waterways, and restricted water areas) and particular trajectories.

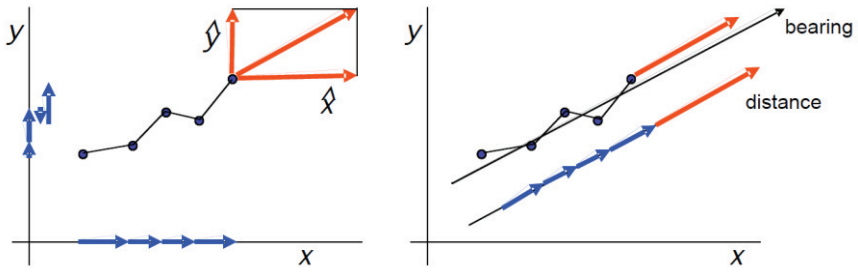


Figure 2-4. An example of two conventional approaches to compute future position from the current geographical position and movements [26]

Several studies have utilized this method with many approaches. Juraszek et al. (2020) used an extended Kalman filter for fast short-term prediction in the distributed processing system Apache Flink [27]. Mao et al. (2017) generated ship position predictions 20 and 40 min ahead for the west coast of the US using an extreme learning machine (ELM) [28]. Duca et al. (2017) employed a kNN classifier to forecast ship movements around Malta [23]. Using an NN, Simsir and Ertugrul (2009) and Zhou et al. (2019) conducted predictions with intervals of 3 and 20 min, respectively, targeting vessels in narrow waterways [29, 30]. Meanwhile, Zissis et al. (2016) attempted predictions in Greek waters up to 4 h ahead, focusing on the regular sea trips of passenger ships around the Aegean Islands [31]. Borkowski (2017) combined AIS location data and sensor data onboard a vessel to predict its subsequent position, aiming at collision avoidance [32]. Gao et al. (2018) used a bidirectional recurrent neural network (bRNN) with an LSTM unit seeking short-term predictions [33], and another study by Gao et al. (2021) combined the TPNet framework and LSTM for multi-step ship trajectory prediction based on four types of navigation stages [34].

2.2. Characteristics of Trajectory-based and Motion-based Methods

As long-term vessel position predictions have not been performed yet, determining a suitable method is vital for achieving long-term prediction. Consequently, the observed location (and dataset) must be considerably larger than the ship distance intervals to accommodate and capture their long-term movements; a small observation area limits long-term movement. The point-based method inevitably involves computing the entire cell information in each computation cycle, and thus the observed area cannot be too large because the computational costs would be exorbitant. Meanwhile, utilization of the trajectory-based method is expected to result in high accuracy for any time-interval threshold and area size; however, this method is not sufficiently flexible for adaptation to factors other than the observed variable. It also involves arduous work on route clustering and classification and trajectory reconstructions, in addition to anomaly detection owing to the partially incomplete and noisy AIS data. Furthermore, although we can create a database for millions of historical trajectories, ship trajectories are not fixed but quickly adapt to the weather and other conditions; ship routes evolve owing to various factors such as climate change and the emergence of new ports and new routes [35]. In contrast, the motion-based method is relatively more flexible, efficient, and practical than the other methods, and, notably, can be effectively generalized for

data variation. Table 2-2 shows the main differences between trajectory-based and motion-based methods for vessel position prediction.

Table 2-2. Primary differences between trajectory-based and motion-based methods for vessel position prediction

	Trajectory-based	Motion-based
Merits	high accuracy for any time-interval threshold and area size	flexible, efficient, and can be generalized for data variation
Demerits	requires arduous work on pre-processing such as trajectory definition, classification, and reconstruction	requires machine learning where developing a model involves a CPU-intensive and specialized expertise

Therefore, from the perspective of long-term position prediction on large international open waters without the consent of any restricted trajectory, a more general and updatable model with generalization to data variation and robustness to noise needs to be constructed that incorporates a fast prediction generation for predicting the long-term position of a vessel. In this study, a motion-based method was used to develop a generalized deep learning model for long-term vessel position prediction to overcome these challenges.

2.3. Characteristics of this Study

In previous studies, such as in Zhang et al. (2020), trajectories with missing records over one day and lasting less than one day were excluded from the training and testing process [35]. However, this study does not execute this pre-processing because eliminating the routes or trajectories with missing data can be applied when trajectories have been defined, only when using other methods (see Table 2-2). Because of the nature of our dataset, we consider an AIS message to be an outlier when its time interval from the previous messages is more than twice the observed time interval. For instance, in a 24-h time interval prediction, AIS data with more than 48 h intervals are considered outliers. These outliers (individual AIS messages, not the entire trajectories) are removed from the training process, but are retained in the testing process.

An end-to-end DL model was developed to supersede a vessel position prediction model that requires arduous preprocessing steps (e.g., trajectory reconstruction) since the time intervals of the AIS data vary considerably. A trajectory reconstruction becomes impractical as the time intervals have an uneven distribution. The asynchronous nature of AIS messages with time intervals of less than 3 min allows the application of a trajectory reconstruction from a

short-interval to a long-interval dataset without a significant error. This satisfactory outcome is because the AIS points can be aligned using a simple method of sampling the nearest point in the time dimension that satisfies the x interval condition as $\lim_{\Delta t \to x} |\Delta t - x| = 0$. Nonetheless, this preprocessing destroys information regarding latent variables [36]; it causes distortion (the trajectory becomes compressed or simplified), causing the acceleration (derived from speed over ground) and rate of turn to become unusable [37].

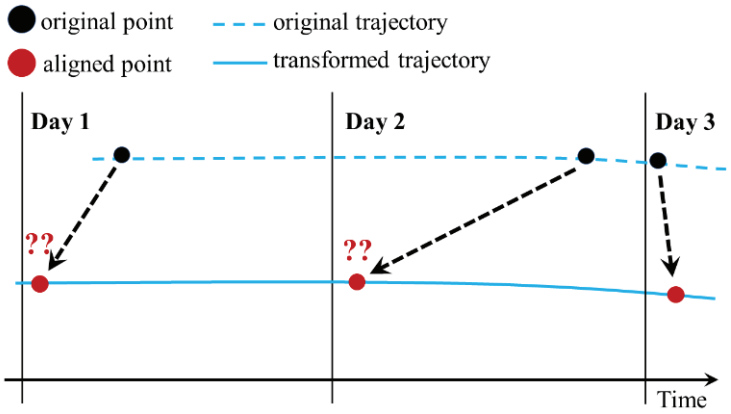


Figure 2-5. Uncertain trajectory reconstruction of long-time-interval data with uneven distribution

However, if the values of the time interval Δt are relatively large with uneven distribution, the interpolation can cause a considerable error in the aligned positions [38]. In this study, the dataset has Δt values ranging between almost zero and 48 h (in the absence of outliers), indicating that although the dataset consists of daily AIS messages (averaging 24 h), their exact time interval can be longer than 24 h. The large gap in time intervals between two consecutive AIS messages signifies that the interpolation to create an evenly-spaced dataset time series would result in large errors. For instance, in a time window of 24 h, a moving ship generally has an average distance interval of more than 400 km. Consequently, interpolating the ship position in this extended time window is expected to result in a significant error in the aligned positions, thus accumulating additional errors when predicting the subsequent position. Figure 2-5 illustrates the uncertainty of trajectory reconstruction of the long-time-interval data with uneven distribution; the aligned data are prone to result in a large error. Therefore, any application or method that requires trajectory reconstruction is considered impractical.

Based on the nature of the AIS dataset used while achieving the objective (which eventually focuses on the long-time interval data and large observation area), performing a trajectory reconstruction is unattainable. Therefore, in this study, deep feed forward network

was employed to solve the long-term position prediction. Previous studies have also attempted to use this basic architecture but have been limited to shallow networks with limited time intervals and area sizes. We aimed for a larger-scale and longer timespan of the AIS dataset, and more importantly, developed a more general and updatable model from a long-time-interval dataset with uneven distribution, which has not been achieved thus far. By leveraging the advantages of deep learning systems supported by extensive fleet data and computing power, the model was developed to learn robustness to noise from extensive data with unrestricted trajectories in international open waters and realize generalization to data variation of time intervals, vessel activities, statuses, and locations.

Chapter 3

Basic Concept and Data Used in This Study

3.1. AIS Data

We processed AIS data from exactEarth [39], considering that it was established as an AIS-data-service company that started deploying AIS receivers on satellites soon after the IMO regulation was implemented. As exactEarth offers the untapped potential of extensive AIS data of SOLAS-compliant ships, we took advantage of the capacity provided by deep learning systems to learn from large datasets to improve the overall performance.

The dataset comprises nine years of archived AIS messages of global capesize bulk carriers, from July 2010 to December 2018. This type of vessel exhibits proper records with less irregularity compared to those of smaller vessel types and more diverse shipping routes than containers and tankers [40]. The dataset alone is composed of over 3.5 million archived AIS messages from 1698 different IMO numbers (vessels).

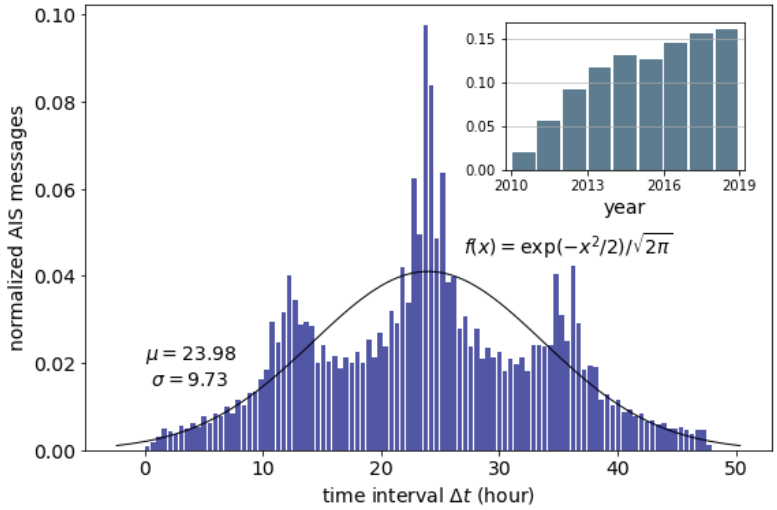


Figure 3-1. Time-interval distribution of the AIS dataset

Daily archived AIS messages have random time intervals Δt between less than 1 h and up to 48 h, averaging 24-h intervals; the time-interval distribution is depicted in Figure 3-1. The fitted probability density function (PDF) of the overall dataset showed a standard Gaussian distribution. The data proliferated from 2010 to 2014 when satellite-based AIS became a standard practice. Although the dataset has an uneven time-interval distribution, it is considered adequate for long-term prediction because it contains years of extensive fleet data worldwide.

3.2. Ship Data

Ship specification data from IHS SeaWeb's fleet data [41] were utilized to organize and validate the capesize bulk carrier vessels from billions of raw exactEarth data records. This dataset was also used to rectify the static information in each AIS message.

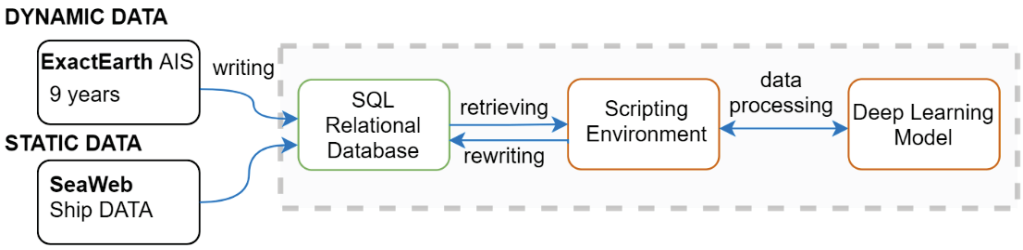


Figure 3-2. Overview of data exploration

Figure 3-2 illustrates the overview of data handling from retrieval to processing for a DL model. All datasets were collected independently from each source and stored in a single SQL relational database for easy access and management. An IMO number identifies every vessel, each of which is related to another piece of information in the dynamic and static data. When a model needs to be generated, a specified chunk of the AIS dataset is retrieved from the relational database and subsequently processed into a DL model. All scripts from preprocessing to postprocessing were written in Python.

3.3. Deep Learning Model for Position Prediction

A DL model refers to the process of training and testing multilayered NNs that can learn complex structures and achieve high levels of abstraction [42]. Building a DL model to perform a specified task satisfactorily requires hours of iterative prototyping; the process is not a trivial task, even for a simple small model. Keras, an open-source high-level API framework capable of running on GPU clusters was used to facilitate rapid model prototyping [43]. The decision regarding the structure of the model and its input-target features was determined through our extensive experimentation and evaluation, focusing on the model objectives: performance and generalization on the observed areas.

In this study, we employed a deep feed-forward Neural Network, or can be called a deep Multi-Layer Perception (MLP). The building block of each neuron or unit is shown in Figure 3-3(a), and the overall network architecture is presented in Figure 3-3(b). For each unit in the hidden layer l, an activation function σ is applied to the input's weighted sum (z). The

information is then propagated through the network up to the output layer L, where the last layer end has no activation (a linear layer).

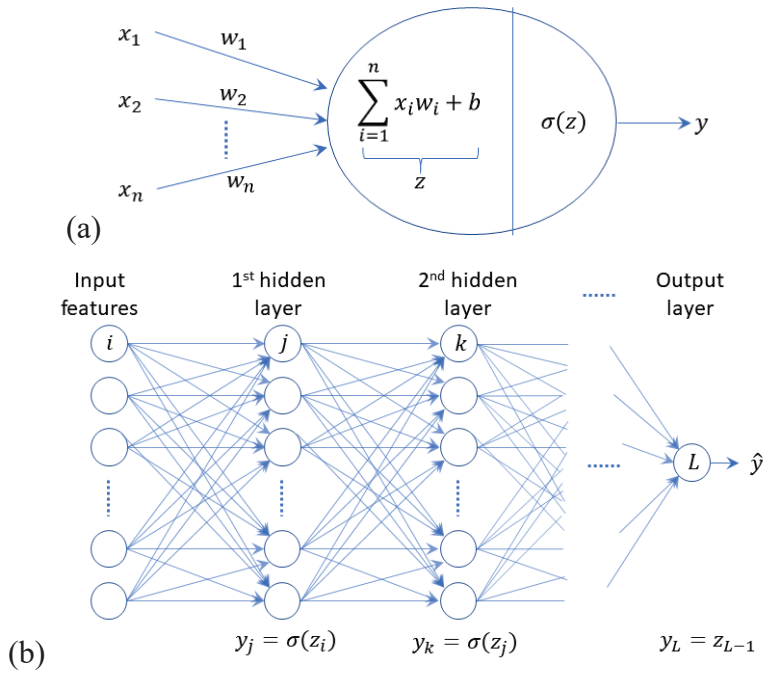


Figure 3-3. (a) The building block of each unit in the MLP (b) The network architecture

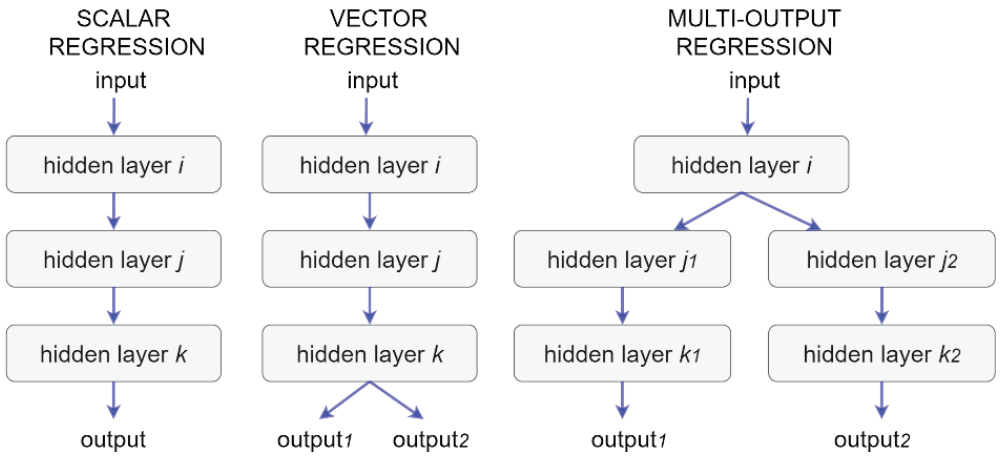


Figure 3-4. MLP architectures for regression

The model was constructed as a DL model to solve regression problems. Compared with the classification problem that treats portions of the covered area as classes, regression problems can achieve a higher location resolution and, therefore, higher prediction accuracy [27]. The option of MLP architecture for regression is depicted in Figure 3-4. Scalar regression model of the MLP architecture outputs one scalar value, where vector regression yields two value at once; multi-output (or multihead) model has more parameters to be computed but can solve the imbalanced loss caused by the different scales of the target. In this investigation, the

target is longitude-interval and latitude-interval value between the current and the next timestep, and hence both vector regression and multihead model would produce similar result.

A straightforward method that takes advantage of the capacity provided by deep learning systems with sufficient data and computing power is proposed to produce long-term ship position prediction. Fast result generation and generalization from the trained model make the proposed method suitable for practical use. The flowchart in Figure 3-5 summarizes the overall method for long-term vessel position prediction. Given a selected location and time range, a sizeable chunk of the AIS dataset is queried from the database. Next, a set of input features X_i and a set of target features y_i comprising vessel positions at the next time step are created. Each set is standardized and then input for training into DL model(s). The model generates a prediction of the displacements of the positions. Then, the results are passed through inverse standardization using the standardization parameters of the target features. Finally, the predicted ship position (i.e., latitude and longitude) is systematically evaluated and analyzed.

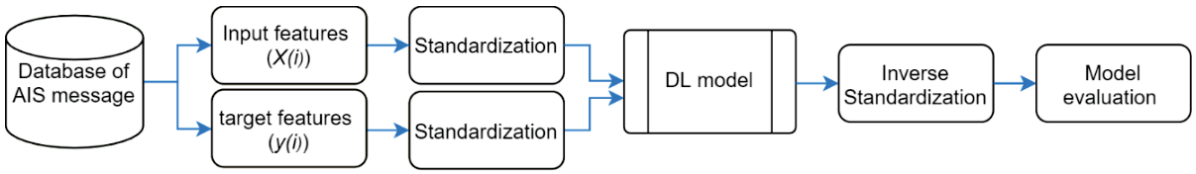


Figure 3-5. Overview of vessel prediction model with AIS and deep learning

3.4. Procedure of This Research

This study explores AIS data and deep learning potential for vessel position prediction by conducting three experiments on 1-d intervals followed by another three on several-day intervals. An overview of the study flow is illustrated in Figure 3-6

In the first part of the research, a model definition was utilized. The model's objectives are to accurately predict the next-1-d position and make a generalization. The first experiment examines a DL model for the Indian Ocean. Subsequently, DL models were assessed for six open oceans and six near-port global primary chokepoints. The target area is the international open waters; we specified a worldwide observation into these twelve areas to avoid building an excessively large model. Finally, a simulation of practical application was carried out inside a sample area of the Malacca Strait.

In the second part of the research, a new model was defined with deeper layers and larger networks. The objectives of this new model are to predict up to the next 4-d position and consequently more extensive coverage with the same procedure as the previous experiments.

The first experiment examines this new model compared with the previous model in four different coverage sizes and time intervals from 1-d to 4-d intervals; each time-interval model has a different coverage size due to the nature of moving ship characteristics. Afterward, the DL model was performed on more detailed size variation to capture the fluctuation of the model performance in more detail in order to find the optimum coverage size requirement for each time interval. Finally, five DL models for five major locations of maritime shipping routes focused on Capesize bulk carriers were assessed to further assess the area's influence.

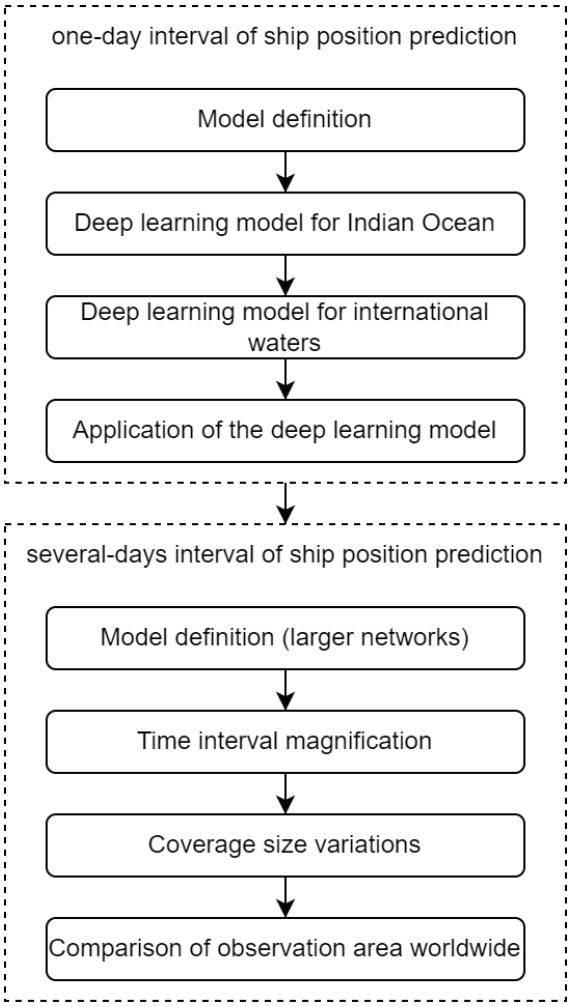


Figure 3-6. Overview of the research's procedure

Chapter 4
One-day Interval of Ship Position Prediction

4.1. Overview of One-day Interval of Ship Position Prediction

Study on long-term prediction of ship position (i.e., more than 12-h interval) is scarce, despite many advantages that can be attained. However, as machine learning algorithm thrives nowadays, supported by abundant maritime big data (i.e., AIS), the lofty aspiration is achievable. Deep learning, one of the most popular machine learning algorithms, craves enormous data for training to effectively produce accurate results. This criterion fits perfectly with maritime transportation as nowadays, AIS data has become mandatory for international shipping.

This research is the first to predict vessel position for around 24-h intervals (on average). The long-term prediction aims to establish additional safety and security in maritime transportation, which benefits all relevant stakeholders. The motion-based method implies building a generalized model that can generate fast results suitable for practical use.

This research consists of three experiments followed by discussions regarding the performance of end-to-end DL models. In the first experiment, a model for the Indian Ocean area was examined, and in the subsequent experiment, 12 models were investigated on open oceans and maritime chokepoints. In the last experiment, a selected sample location within the Malacca Strait area was examined, resembling a simulation of practical application. In the discussion section, the DL model was compared to other end-to-end models with sequence-based architectures (i.e., RNN, LSTM), and then they were tested on a small dataset. Finally, the conclusion of this research is presented.

The detail of the one-day interval of ship position prediction in this dissertation is presented in the following chapter.

4.2. Methodology

The overview of the prediction model is explained in the previous chapter.

We performed the regression task by setting the target as a scalar value for analysis and objective evaluation in the prototyping process at the expense of running it twice; the target features are the displacements of longitude $\Delta\lambda$ and latitude $\Delta\varphi$ directions. Accordingly, the

input and target features were fed into two independent DL models with the same model properties. The model predicts the displacements in longitude and latitude directions that are uncorrelated.

4.2.1. Model Structure and Properties

During the prototyping process, the most salient input features were selected by evaluating permutation feature importance (PFI) based on Fischer et al. (2018) [44]. After a model was trained and the model score s was computed, a corrupted matrix $X_{k,j}$ was generated by permuting feature j in the original matrix X. Subsequently, a new score $s_{k,j}$ was computed based on the prediction of the permuted data $X_{k,j}$. Finally, the PFI score I_j was calculated as follows:

$$I_j = \frac{\frac{1}{K}\sum_{k=1}^{K} s_{k,j} - s}{s} \times 100\% \quad (1)$$

The permuted data $X_{k,j}$ were then returned to the original order. This step was repeated for all the features of the input data.

The PFI evaluation was conducted using the dev set at the prototyping stage to determine the extent to which the model relied on each input feature to generate predictions [45]. All irrelevant and insignificant features with minimum PFI scores were removed, and subsequently, a new feature was extracted and re-evaluated until the final set of features was established. Thus, any timestep-based information added to the final inputs would not improve the model performance; instead, it would overfit the model.

The proposed model uses a deep network with six layers: five hidden layers with a sigmoid activation function and a one-unit linear output layer [64, 64, 32, 32, 16, 1]. The mean absolute error (MAE) was employed as the loss function since it is less sensitive to outliers than the root mean square error (RMSE), and mini-batch gradient descent with a mini-batch size of 64 was used for training. We used an adaptive learning rate method involving the adaptive moment estimation optimization algorithm (Adam), an update to the RMSProp with momentum [39]; the learning rate was automatically reduced by a factor of 0.5 once the validation loss stagnated for 50 epochs. Moreover, to accommodate areas with low AIS data coverage (to avoid overfitting caused by deep networks), a validation-based early stopping was set to interrupt training once the validation stopped improving for 100 epochs. These in-training validation-based methods were solely used for the experiment and were not used in the prototyping stage [40].

A hold-out validation split was conducted to prevent information leakage from the test data. An overview of the hold-out validation split for the dataset is depicted in Figure 4-1. AIS data from July 2010 to December 2017 were randomly shuffled into a training set, and the dataset was disrupted with data from other vessels and timelines. Meanwhile, a dev set was randomly sampled from half of the 2018 data. The remainder was retained as a test set that had never been used for model prototyping. The split was designed to reflect the dev/test set as the recent AIS data. The dev set was used to validate the training set while having the same distribution as the test set. The randomly shuffled data combined with mini-batch training ensured rapid convergence of the gradient descent with minimal disturbances. Moreover, this combination prevented deep networks from conspiring to memorize the chronological sequences of vessel locations.

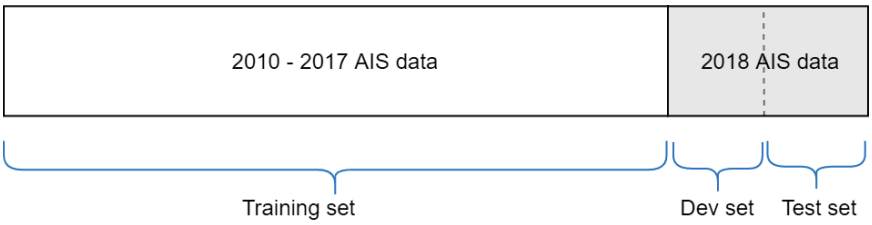


Figure 4-1. Hold-out validation split of the dataset (non-proportional scale)

4.2.2. Input and Target Features

A set of AIS messages was first retrieved from the SQL database, given a boundary condition of location and time. The specified input and output features were then extracted from the dataset with vessel states at the current time step t as the input matrix X_i and its location at the next time step $t + 1$ as the target vector y_i. The input X is composed of 10 features: longitude λ, latitude φ, speed over ground U_{og}, course over ground ψ_{og}, heading ψ_h, time interval Δt, distance d, haversine distance[1] d_h, Manhattan distance[2] d_m, and average speed U_{avg}. Meanwhile, the component of the target y follows the DL model output that targets particularly one of the two variables: the displacements of longitude $\Delta\lambda$ and latitude $\Delta\varphi$

[1] great-circle distance calculation between two points on a sphere given their coordinates.

[2] distance between two points measured along axes at right angles.

between the current time step t and the next time step $t+1$. The input matrix and target vectors are expressed as follows:

$$X_i = \left\{\lambda_t^i, \varphi_t^i, {U_{og}}_t^i, {\psi_{og}}_t^i, {\psi_h}_t^i, \Delta t_t^i, d_t^i, {d_h}_t^i, {d_m}_t^i, {U_{avg}}_t^i\right\} \tag{2}$$

$$y_{i_\lambda} = \left\{\Delta\lambda_t^i\right\};\ y_{i_\varphi} = \left\{\Delta\varphi_t^i\right\} \tag{3}$$

where $i = \{0,1,\dots,m\}$, the subscript t represents the time step, $\Delta\lambda_t = \lambda_{t+1} - \lambda_t$ is the longitude interval, and $\Delta\varphi_t = \varphi_{t+1} - \varphi_t$ is the latitude interval.

Input features of λ, φ, U_{og}, ψ_{og}, and ψ_h are extracted directly from the archived AIS messages, whereas some of the remaining inputs provide implicit information on the position history at the previous time step $t-1$. Further, the target timestamp (at the next time step $t+1$) is represented as a time interval Δt and a rough distance to the target d_t. They are defined as follows:

$$\Delta t_t = T_{t+1} - T_t \tag{4}$$

$$d_t = \Delta t_t {U_{og}}_t \tag{5}$$

$$d_{h_t} = d_{h_{((\varphi_{t-1},\lambda_{t-1}),(\varphi_t,\lambda_t))}} \quad where\ \varphi, \lambda\ in\ radians \tag{6}$$

$$d_{m_t} = d_{h_{((\varphi_{t-1},\lambda_{t-1}),(\varphi_{t-1},\lambda_t))}} + d_{h_{((\varphi_{t-1},\lambda_{t-1}),(\varphi_t,\lambda_{t-1}))}} \tag{7}$$

$${U_{avg}}_t = d_{h_t}/\Delta t_t \tag{8}$$

where T is the timestamp of the AIS message, and T_{t+1} is referred to as the prediction timestamp. We utilized implicit information because it enables the DL model to generate predictions while using less information, rather than employing every possible explicit information directly from the AIS messages (the selection is described in Section 4.1.2). The above calculations (Eq. (3)–(7)) are empirical formulas such that the extraction is performed at a high speed.

After retrieving the input and target, all input features X_i are standardized, similar to the target feature standardization shown in Eq. (9). The standard score of the target feature y_i is calculated as follows:

$$y'_i = std(y_i) = (y_i - \overline{y_i})/\sigma_{y_i} \tag{9}$$

where y_i is the target feature containing y_{i_λ} and y_{i_φ}, $\overline{y_i}$ is the mean of y_i, σ_{y_i} is the standard deviation of y_i, and y'_i is the standard score of y_i. The standardization of the input and output features avoids any imbalance in the network parameters and loss calculations during training.

Two sets of standardized input and target features, (X'_i, y'_{i_λ}) and (X'_i, y'_{i_φ}), are then fed separately into two independent DL models with the same model properties.

4.2.3. Performance Evaluation Metrics

Loss Score

Two separate models were trained for each standardized target feature y'_i of $\Delta\lambda_t$ and $\Delta\varphi_t$ values (scalar regression model). Subsequently, each value was discretely predicted. The predicted vessel position was then generated by inverse standardization using the standardization parameters of the target features as follows:

$$\hat{y}_i = std^{-1}(\hat{y}'_i) \tag{10}$$

where $\hat{y}_i$ is the prediction vector after inverse normalization, and $\hat{y}'_i$ is the prediction vector that combines the outputs of the two separate models ($\hat{y}'_{i_\lambda}$ and $\hat{y}'_{i_\varphi}$). Both results ($\hat{y}_\lambda$ and $\hat{y}_\varphi$) are then transformed into latitude $\hat{\varphi}_{t+1}$ and longitude $\hat{\lambda}_{t+1}$ by the addition of each component in the input (i.e., λ_t, φ_t). Finally, a new loss score for each result is calculated as follows:

$$MAE_\lambda = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\lambda}^i_{t+1} - \lambda^i_{t+1}\right| \tag{11}$$

$$MAE_\varphi = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\varphi}^i_{t+1} - \varphi^i_{t+1}\right| \tag{12}$$

where $\hat{\lambda}^i_{t+1} = \left(\hat{y}_{i_\lambda} + \lambda^i_t\right)$ is the prediction longitude, $\hat{\varphi}^i_{t+1} = \left(\hat{y}_{i_\varphi} + \varphi^i_t\right)$ is the prediction latitude, λ^i_{t+1} is the target longitude, and φ^i_{t+1} is the target latitude.

Metric Score

The metric scores for the combined results (i.e., location) were defined as the mean distance error (MDE) and mean angular error (MaE). The distance error is the haversine distance between the prediction and target (true position) in kilometers, signifying the extent to which the result (prediction) deviates from the true position (target). MDE can be a single evaluation metric for the long-term prediction of vessel location; however, in cases where the distance between previous positions and target positions is remarkably diverse (owing to time-interval variation), another evaluation metric is required.

The MaE as the second evaluation metric was calculated based on the three known points of the base, target, and predicted positions. The angular error between the target and the prediction is a spherical triangle that can be solved using the law of haversines [46]. Because the haversine distance between the prediction and the target is small, the formula is derived in

combination with the spherical law of cosine. Accordingly, these metric scores were calculated as follows:

$$MDE = \frac{1}{m}\sum_{i=1}^{m} d_{h_{(v,w)}}^{i} \tag{13}$$

$$MaE = \frac{1}{m}\sum_{i=1}^{m} C_{(u,v,w)}^{i} \tag{14}$$

$$C_{(u,v,w)} = \arccos\left(\frac{cos\left(d_{h_{(v,w)}}\right) - cos\left(d_{h_{(u,v)}}\right) cos\left(d_{h_{(u,w)}}\right)}{sin\left(d_{h_{(u,v)}}\right) sin\left(d_{h_{(u,w)}}\right)}\right) \tag{15}$$

where u is the base point (φ_t, λ_t), v is the true position $(\varphi_{t+1}, \lambda_{t+1})$, and w is the predicted point $(\hat{\varphi}_{t+1}, \hat{\lambda}_{t+1})$. The predicted position is in a vector space, and consequently, its distance error relative to the current position may not be proportional to its angular error. Figure 4-2 illustrates the relationship between distance and angular errors. Furthermore, the distance and angular errors can reflect the accuracy and precision of the prediction from the target, respectively.

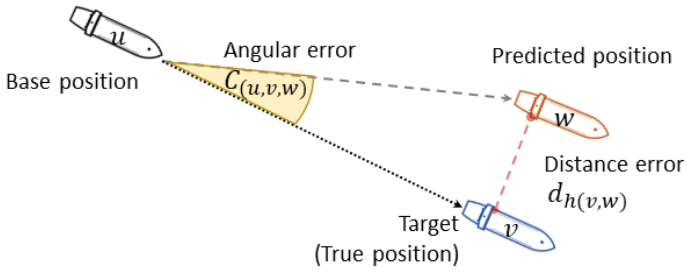


Figure 4-2. Relation between the distance error and the angular error

4.3. Baseline Model

A non-ML baseline and a sanity check were used as model-building guidelines to conduct error analysis on bias and variance, gaining confidence in the overall performance of the DL models. The baseline replicates a human-level performance as a proxy for the Bayes optimal error in the classification problem [47].

The predicted positions were defined as the next timestep positions averaging 24-h intervals, and consequently, a conventional formula for the dead reckoning position[3] of the ship could not be adopted as the baseline. The dead reckoning calculation is too simple for two distant points; while it is accurate in the Euclidian space, it fails to map the Earth as a great circle[4]. An equation for the geodesic on a spherical surface, namely the great circle equation, is more accurate for planning routes [48].

Therefore, the geodesic calculation was established as a baseline model, computed based on the shortest distance on a spherical earth. Given a starting position (φ_t and λ_t), current course ψ_{h_t}, and distance to the target (next position) d_t, the destination point along a (shortest distance) great circle arc is computed as follows:

$$\breve{\varphi}_{t+1} = \mathrm{asin}\big(\sin(\varphi_t)\cos(\delta_t) + \cos(\varphi_t)\sin(\delta_t)\cos\big(\psi_{h_t}\big)\big) \tag{16}$$

$$\breve{\lambda}_{t+1} = \lambda_t + \mathrm{atan2}\big(\sin\big(\psi_{h_t}\big)\sin(\delta_t)\cos(\varphi_t)\,, \cos(\delta_t) - \sin(\varphi_t)\sin(\breve{\varphi}_{t+1})\big) \tag{17}$$

where $\varphi_t, \lambda_t, \psi_{h_t}$ are in radians, $\delta_t = d_t/r_{earth}$ is the angular distance, and r_{earth} is the earth diameter [49]. The results from Eq. (16) and (17) were subjected to normalization to degrees.

Additionally, the average distance interval of the ships was adopted as a sanity check, assuming that the next position always equals the present state. This approach is adopted to examine the performance evaluation metrics between the current position (φ_t, λ_t) and the next position $(\varphi_{t+1}, \lambda_{t+1})$, calculated as follows:

$$\overline{d_h} = \frac{1}{m}\sum_{i=1}^{m} d_{h^{i}_{((\varphi_t,\lambda_t),(\varphi_{t+1},\lambda_{t+1}))}} \; where\ \varphi, \lambda\ in\ radians \tag{18}$$

where (φ_t, λ_t) is the base point, and $(\varphi_{t+1}, \lambda_{t+1})$ is the next position at 1-d interval.

[3] in seamanship, a position determined by plotting courses and speeds from a known position.

[4] also known as an orthodrome, is the largest circle that can be drawn around a sphere. Earth is not a perfect sphere, but as long as the Arctic and Antarctic Circles are not included, all meridians on Earth can be treated as great circles.

4.4. First Experiment: DL Model for the Indian Ocean

4.4.1. Experimental Setup

In the first experiment, a DL model was built for the Indian Ocean. The observed area is within 60°E–90°E longitude and 24°S–0°N latitude boundaries, covering 8.6 million square kilometers[5]. Compared to all previous studies, the boundaries were relatively larger, adequately capturing the continuous long-term movements of the vessels. Figure 4-3 presents the normalized density distributions of the dataset in the observed area from nine years of AIS messages involving 1531 different IMO numbers (vessels). This distribution was calculated using kernel density estimation (KDE), adopting a Gaussian kernel [50], in which the background geography was visualized using an open-source Python library Matplotlib Basemap toolkit for geospatial data processing [51].

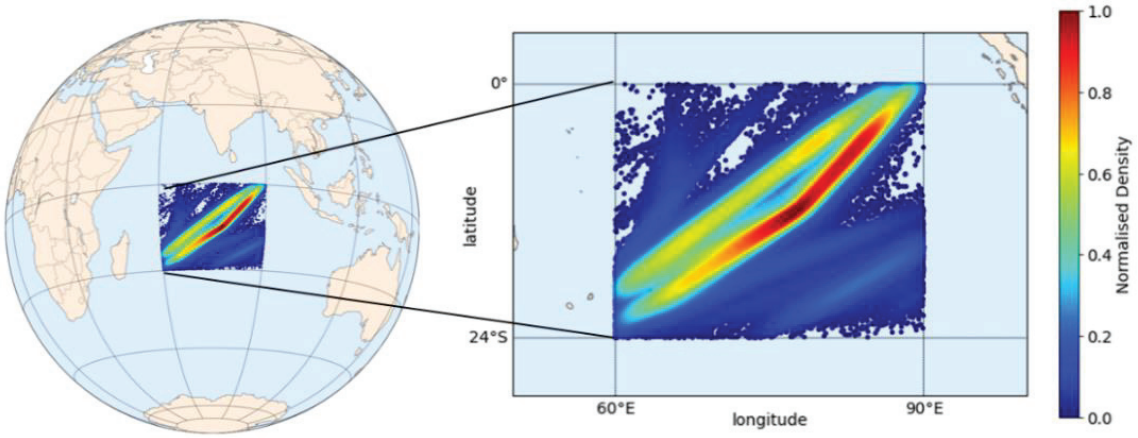


Figure 4-3. Normalized density distribution of AIS data in the observed Indian Ocean area

The dataset was split into a training set and dev-test sets according to the hold-out validation split described in Section 4.1.3. The training set comprised more than 122,000 data records after omitting AIS messages with missing values and abnormalities, whereas each dev and test set contained more than 10,000 data records. The size of these sets was considered sufficiently large, thus providing high confidence in the model performance.

[5] calculated by geodesic area assuming the earth is a perfect sphere with a radius defined in WGS84 as 6,378,137 m.

4.4.2. Error and Sensitivity Analyses

The MAE, as the performance evaluation metric of the model, is shown in Table 4-1; these values correspond to the absolute deviation between the prediction point and the target point in degrees (longitude λ and latitude φ). The results from the geodesic calculations were computed based on the training set.

Interestingly, the DL model performance on the training set and dev set surpassed the geodesic calculation as the baseline model. The model displayed minimal bias, fitting the training and dev sets satisfactorily. In practice, the sole focus of a regression model is not the deviation between the training loss and dev loss (i.e., variance), but rather its prediction performance, that is, the dev loss. Nevertheless, the variance in the model was insignificant, indicating that the model generalized very well to the dev set without overfitting.

Table 4-1. Error analysis of the DL model

Model	MAE_{λ}	MAE_{φ}
Geodesic calculation	0.225	0.226
Training set	0.163	0.145
Dev set	0.169	0.147

Moreover, hyperparameters sensitivity of the DL model was carried out while also applying regularization techniques to confirm further that the model generalized very well to the dev set. All hyperparameters combinations were permutated except the structure of hidden layers and their units to see the effect on prediction performance from hyperparameters' influence. This permutation was repeated for other models with regularization: 20% and 40% dropout on all hidden layers and a model with batch normalization.

Figure 4-4 shows the hyperparameters sensitivity of all models. The selected hyperparameter proved to be the best combination for the standard DL model. The model with no regularization achieved the most optimum error compared to other models with regularization, confirming that the model was already generalized well without overfitting. Applying dropout worsened the model performances than applying batch normalization. Naturally, predictions of longitude values would have accuracy less than latitude values since the longitude values have a distribution twice as large as latitude. This natural divergence appears to negatively affect the model with batch normalization but with the expense of accuracy.

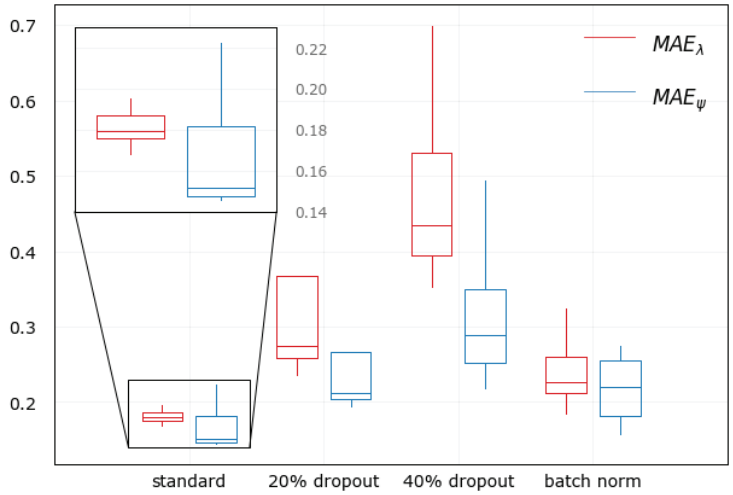


Figure 4-4. Sensitivity analysis of the DL model

Based on these error and sensitivity analyses, we determined that the proposed deep learning model achieved an optimal error with generalization to the dev set and was ready for the evaluation on the unseen test set resembling a practical scenario.

4.4.3. Results and Discussions

The performance evaluation metrics of the DL and baseline models are presented in Table 4-2. All results were computed using the test set. The average distance interval indicates the difference between the current vessel position and the subsequent position. Compared with these average values, the geodesic calculation appears to be reasonably accurate for long-term predictions.

The DL model adequately outperformed the baseline model and almost doubled its performance based on the metric scores. Given that the distance between the current position of a vessel and its position the next day was approximately *486 km*, the geodesic calculation might be accurate for long route planning, yielding an average distance error of *40 km*. However, the DL model generated more accurate and precise predictions with a slight mean distance error of approximately *25 km* and a mean angular error of *1.8°*. If we assume that the evaluation metric (i.e., *MDE*) of the average distance interval created an error of *100%*, simulating a model performance similar to the classification problem, the geodesic calculation and DL model predicted the vessel positions with *8%* and *5%* error, respectively.

Moreover, several ensemble-based ML regressors have been investigated, including random forest, gradient boosted decision trees, and extreme gradient boosting algorithm (XGBoost) [52, 53]. The hyperparameters were tuned using the exhaustive grid search method, and therefore, all hyperparameters were optimized for each ML model in this manner. With the

same input, output, and validation splits, the performances across all the models were broadly comparable. Nevertheless, despite performing more accurately than the geodesic calculation, their performances were not significantly different from each other, among which the DL model still achieved the best performance (*13%* better than the ML models).

Table 4-2. Comparison of the model performance

Methods	Loss		$\boldsymbol{MDE}$	$\boldsymbol{MaE}$
	$\boldsymbol{MAE_\lambda}$	$\boldsymbol{MAE_\varphi}$	(km)	(degree)
Average distance interval	3.553	2.564	486.1	-
Geodesic calculation	0.239	0.239	40.6	3.3
Deep learning	**0.174**	**0.149**	**27.0**	**1.7**
Random forest	0.197	0.174	31.3	2.1
Gradient boosted decision tree	0.198	0.171	31.4	2.3
XGBoost	0.197	0.175	31.6	2.2

Results from the DL model compared with the baseline model can be analyzed further by its distribution, which is plotted in Figure 4-5. The 2D plot represents more than ten thousand test data of vessels' previous position, plus their next position predicted by geodesic calculation and the DL model with axis in degrees longitude and latitude. To demonstrate the accuracy of the predictions, all plotted points were normalized relative to the true position (target) positioned at (0,0). The sparsely distributed plots of vessels' previous position (blue dots) were caused by the time interval of AIS messages, affecting their distance intervals between the previous position and target (next position). Their dense plot past Southwest-Northeast direction was due to the typical pattern of shipping lanes in the Indian Ocean (see Figure 4-3).

Nevertheless, the geodesic calculation and DL model generated precise predictions, at which the model confirmed better accuracy and precision, converging on the target: the standard deviation of the model predictions was slightly lower than the geodesic calculation with 0.40 and 0.43, respectively; *92%* of the predictions were within one standard deviation from the mean, more precise than the geodesic calculation with only *87%*.

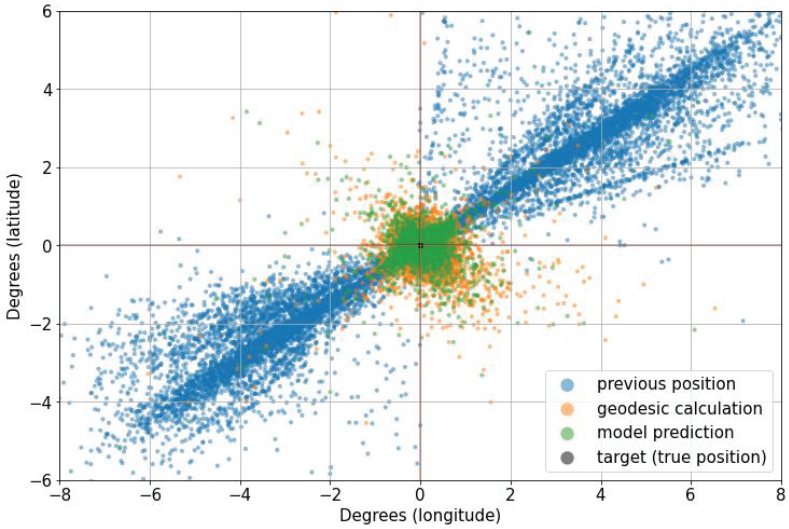


Figure 4-5. Normalized plot distribution of the result in the Indian Ocean area from 10,888 vessel positions

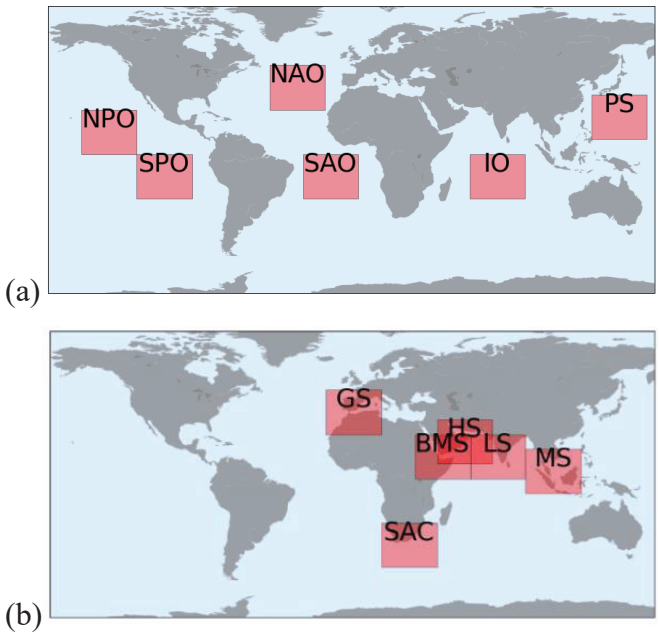
4.5. Second Experiment: DL Models for the International Open Waters

4.5.1. Experimental Setup

In the second experiment, 12 DL models were built for six open oceans and six maritime chokepoints with the same model properties to assess the influence of the respective areas. Open waters of the North Pacific Ocean (NPO), South Pacific Ocean (SPO), North Atlantic Ocean (NAO), South Atlantic Ocean (SAO), Indian Ocean (IO), and Philippine Sea (PS) were selected considering that vessels generally sail across these deep oceans without stopping. According to Rodrigue (2020), the primary maritime shipping chokepoints were selected to represent long-term vessel behavior near the ports and congested (high-traffic) waters, namely, Gibraltar Strait (GS), South Africa Coast (SAC), Bab al-Mandab Strait (BMS), Strait of Hurmuz (HS), Laccadive Sea (LS), and Malacca Strait (MS) [54]. Altogether, the ocean and chokepoint areas correspond to the simple and complex movement of ships, respectively.

Figure 4-6 shows the location and size of each area (red rectangles). The size of the observed maritime chokepoint area was as large as that of the open ocean. Each has the same observation area size: within 30-interval longitude and 24-interval latitude, covering more than eight million square kilometers. The same procedure as in the first experiment was applied to the models: each used the same model properties and followed a similar hold-out validation split. We limited the worldwide observation to these 12 areas to avoid building an excessively large model. A large model is prone to fit noise from inferior-quality data and random vessel

behavior contained in the massive AIS dataset. Regularization methods might be a constructive solution; however, they require an extremely large model to fit all nonlinear problems.



NPO, North Pacific Ocean; SPO, South Pacific Ocean; NAO, North Atlantic Ocean; SAO, South Atlantic Ocean; IO, Indian Ocean; PS, Philippine Sea; GS, Gibraltar Strait; SAC, South Africa Coast; BMS, Bab al-Mandab Strait; HS, Strait of Hurmuz; LS, Laccadive Sea; MS, Malacca Strait.

Figure 4-6. Location of the observed areas: (a) open oceans and (b) maritime chokepoints

4.5.2. Results and Discussions

The performance evaluation metrics of the models for each examined area are listed in Table 4-3. Again, the DL models substantially outperformed the geodesic calculation as the baseline model in all areas, especially in the chokepoint areas. The improvement score (I_S) was the only inference to determine the extent to which the DL model outperformed its counterpart baseline model based on the *MDE* scores. The improvement scores were notably higher in the chokepoint areas than in the open ocean areas, scoring more than a *50%* improvement over the baseline model for almost all the chokepoint areas. The geodesic calculation appears accurate enough to predict moving vessels; however, it fails to predict vessel behavior near the ports and high-traffic areas (i.e., the chokepoint areas). With no information regarding vessel status, historical trajectory, or destination, conventional calculations would fail.

The effect of a high improvement score can be seen by the normalized results distribution plot from Malacca Strait's model (see Figure 4-7), where the DL model was vastly superior to geodesic calculation. Vessel behavior in this area was widely dispersed in terms of voyage direction and time interval. The scattered results of the geodesic calculation from the target

demonstrated the failure of the conventional method to estimate vessel position in this area. Nevertheless, supported by a considerable amount of training data, the model could learn intricate and high-level structures in its deep layers, generating predictions with significant performance than the baseline model.

Table 4-3. Comparison of the loss and metric scores of the DL model and baseline model on each observed area

Area	DL Model				I_s (%)	Geodesic Calculation				Number of Test Data	Average Distance Interval
	Loss		MDE (km)	MaE (deg)		Loss		MDE (km)	MaE (deg)		
	MAE_λ	MAE_φ				MAE_λ	MAE_φ				
NPO	0.18	0.13	27	1.8	27	0.20	0.23	37	3.1	471	485
SPO	0.13	0.11	20	1.5	15	0.13	0.15	24	2.1	499	484
NAO	0.30	0.22	41	3.7	15	0.33	0.28	48	4.5	2474	467
SAO	0.13	0.11	19	1.7	24	0.15	0.15	26	2.5	4672	491
IO	0.17	0.15	27	1.7	33	0.24	0.24	41	3.3	10888	486
PS	0.25	0.31	48	8.0	37	0.48	0.46	77	10.5	9206	453
GS	0.54	0.27	63	21.7	38	0.78	0.53	102	20.2	4263	355
SAC	0.35	0.23	46	15.4	50	0.59	0.54	92	15.0	13337	389
BMS	0.28	0.31	51	4.3	55	0.56	0.75	113	10.9	1251	473
HS	0.34	0.27	52	39.8	58	0.77	0.70	123	31.0	3734	259
LS	0.26	0.33	51	37.4	50	0.57	0.63	102	27.8	5975	273
MS	0.37	0.34	63	16.7	60	0.88	0.93	158	19.5	22745	417

In Figure 4-7, Gaussian KDE was computed to approximate the probability density function of each three different vessel positions: previous position, geodesic calculation, and model prediction. The results are plotted in their normalized density by each latitude and longitude. A 1D plot of each vessel's position is also plotted to explain each distribution. The superior accuracy and precision of the model compared to the geodesic calculation were confirmed by the standard deviation of 0.89 and 2.06 and the population within one standard deviation from the mean of *92%* and *87%*, respectively.

Both the DL model and geodesic calculation showed more accurate predictions for the ocean areas than the chokepoint areas, but not in the average distance intervals between the current position and the next position, indicating that more vessels have shorter distance intervals in the chokepoint areas. Almost all vessels were moving in the ocean areas, whereas in the chokepoint areas, many vessels were idle for berthing or resting at anchorages. These

activities resulted in an extremely short distance interval between two consecutive AIS messages at any time interval.

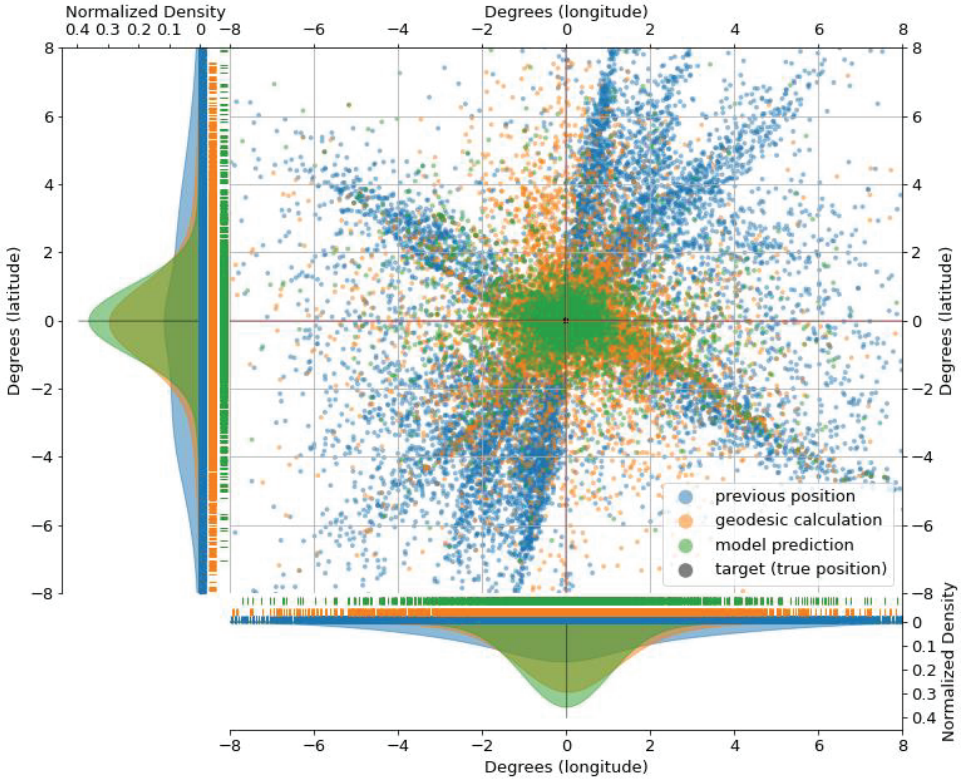


Figure 4-7. Normalized plot distribution of the result in Malacca Strait (MS) from 22,745 vessel positions

Figure 4-8 shows the distance-interval distribution of AIS data on both areas; the distance intervals were calculated from the haversine distance between the base point (φ_t, λ_t) and the next position $(\varphi_{t+1}, \lambda_{t+1})$. Many vessel activities have very low distance intervals in the chokepoint areas but not in the ocean areas. Consequently, the average distance intervals in the chokepoint areas would indicate lower performance evaluation metrics than the ocean areas. The irregularity of idle vessels and activities other than sailing with minimal distance intervals renders the *MaE* unsuitable as a metric score. The angle between the prediction and the target no longer corresponds to the accuracy of the prediction because the distance interval between two consecutive positions is extremely short. Accordingly, for the evaluation of the chokepoint areas, the *MDE* alone suffices as a single metric score.

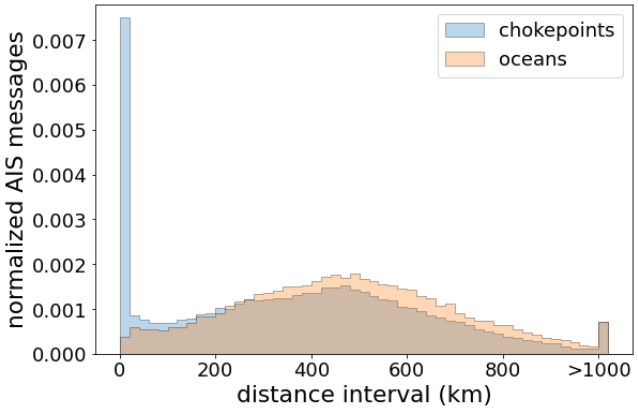


Figure 4-8. Distance-interval distribution of AIS data on the ocean and chokepoint areas

4.5.3. Discussions: Influence of the Area

Figure 4-9 presents the distance error distribution of the predicted positions based on the distance interval, where the distance errors are the haversine distance between the true position $(\varphi_{t+1}, \lambda_{t+1})$ and the predicted point $(\hat{\varphi}_{t+1}, \hat{\lambda}_{t+1})$.

In the ocean areas, as most data have distance intervals between 400 and 600 km, the DL models generated continuous accurate predictions with a distance error below *40 km* for *80%* of the population over the entire test set. For the same *80%* population, the geodesic calculation generated a distance error below *60 km* with a more diverse distribution, where its standard deviation was twice as high as that of the DL models.

In the chokepoint areas, as most data have distance intervals distributed between 400 km and near zero, the DL models constantly generated accurate predictions with a distance error below *65 km* for *80%* of the population. In contrast, the geodesic calculation generated projections with distance errors below *100 km* for only *65%* of the population. Its projections resulted in grave errors when the distance intervals were broad, but the errors became less severe when the distance intervals were near zero. Overall, the geodesic calculation fails to make accurate projections for the complex movements of vessels in chokepoint areas.

Deep learning systems can learn from large datasets to effectively improve their performance to generate accurate predictions by discovering complex variations across all input features [55]. Accordingly, we computed the mean PFI score of the DL models on the ocean areas and chokepoint areas to determine which features play a vital role in improving the overall performance (shown in Figure 4-10). The scores were calculated based on the test set of each area and were grouped by ocean and chokepoint areas and then averaged. The input features were classified into five components: distance, speed, angle, time, and position.

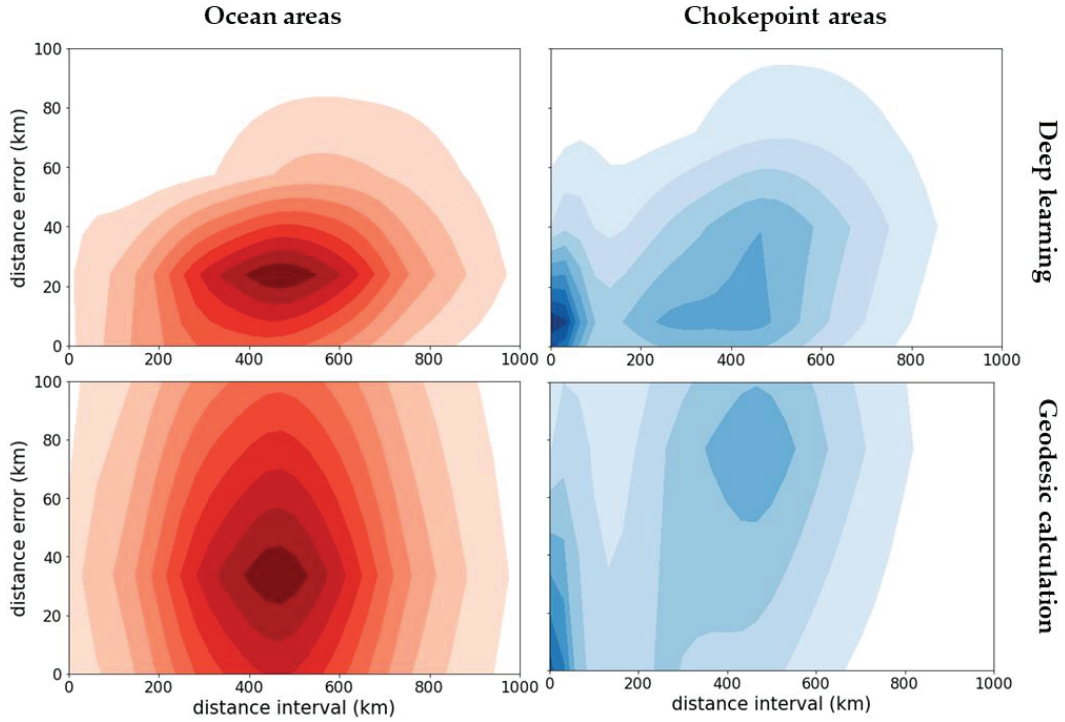


Figure 4-9. Distance error distribution of the predicted positions based on the distance interval on ocean areas (left) and chokepoint areas (right): DL model (top) and geodesic calculation (below).

Based on the results for both the chokepoint and ocean areas, the speed component of the vessel (i.e., U_{og}, U_{avg}) does not appear to be essential because the distance d already represents it, whereas the course over ground ψ_{og} proves vital for making predictions.

Models in the ocean areas rely primarily on input features of angle and distance components but not on speed and position components; this is similar to the geodesic calculation, which generally proved accurate for position projections of vessels moving in a definite straight direction or making simple movements.

Meanwhile, models in the chokepoint areas rely more on the current position (i.e., λ_t, φ_t), implying that they can be used to develop high-level abstraction to alternatively predict the next position when vessels deviate from a simple trajectory. Furthermore, focusing on the distance component, they also rely more on the distance d as a rough estimate of the distance interval to the next position than the haversine distance d_h and Manhattan distance d_m as the distance interval between the current position and the previous position, suggesting that the previous position of the vessel at $t-1$ is not as important as its current position at t and a rough estimate projection at $t+1$. Nonetheless, the utilization of all input features may not

suffice to cover the entire range of unexpected vessel behavior near ports or with complex movements.

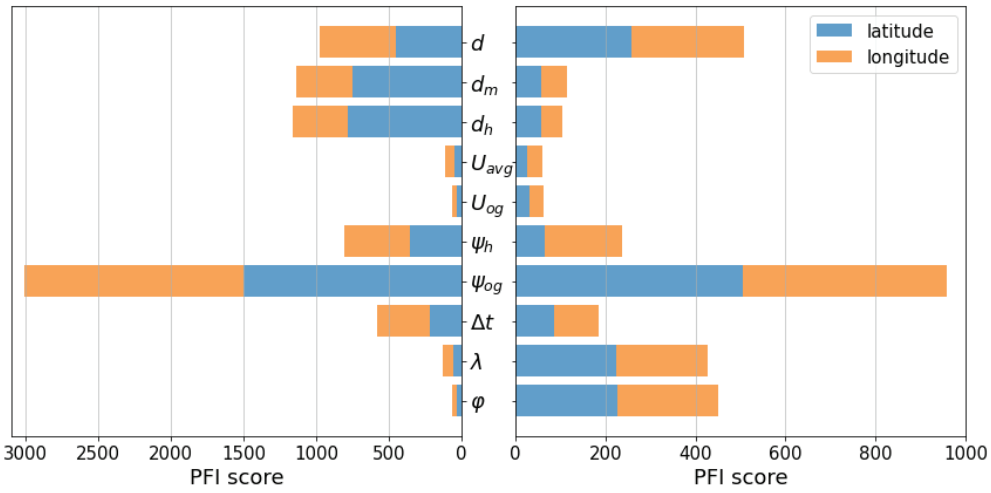


Figure 4-10. Mean PFI score of the DL models on the ocean areas (left) and chokepoint areas (right)

The model performance can be enhanced by incorporating additional information regarding the vessel status, destination, or historical trajectory without a trajectory definition; thus, this approach remains a topic to explore in our future work.

4.5.4. Discussions: Influence of Data Size

The proportion of AIS messages varied widely according to the area location. This variability among different areas is a general characteristic of AIS data: rarely passed areas contributed only a limited amount of data, whereas congested areas contained a substantial quantity of information. Figure 4-11 presents the size distribution of the training and dev-test sets in each observed area based on the same hold-out validation split.

The proportion of the dev-test sets to the training set increased in the areas with the busiest shipping routes; for instance, the Malacca Strait (MS) area encompasses half of South East Asian waters, including the most congested maritime primary chokepoints and three secondary chokepoints in Indonesian waters [54]. The proportion was drastically reduced in the rarely transited areas, such as the Pacific Ocean. Consequently, areas with a small amount of test data would indicate lower confidence in the overall performance of the model than areas with extensive data. The K-fold validation split can solve this problem; nonetheless, it would produce non-comparable models.

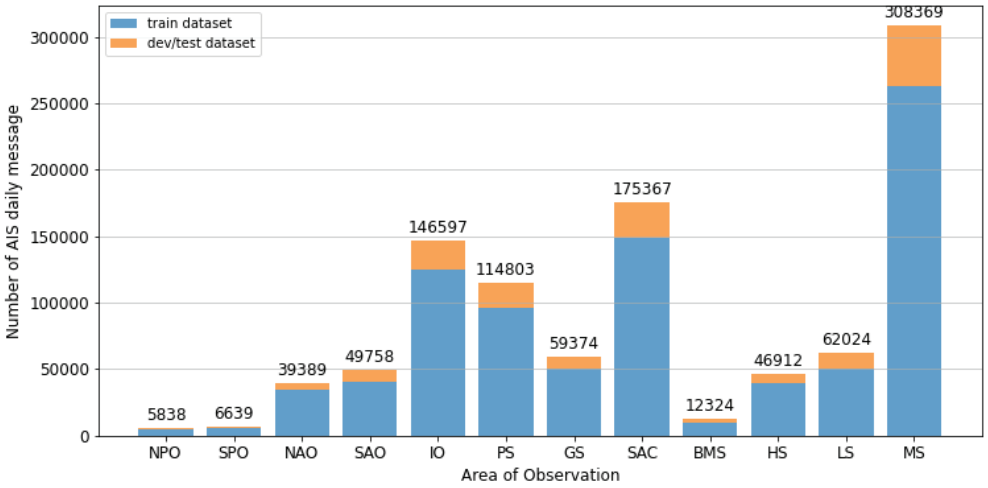


Figure 4-11. Size distribution of the training set and dev-test sets for each area

The dataset size affects deep learning performance, as a small amount of training data may degrade the performance [56]. We recreated the DL model for the Malacca Strait (the area with the most extensive dataset size) for each variation in the training set to show the effect of the data size on the model performance (see Figure 4-12). The most significant improvement in the performance occurred when the training set size exceeded 25,000. The model performance began to stabilize when the training set reached approximately 200,000. This effect may not represent all areas because three areas have a small dataset of less than 25,000 thresholds (i.e., NPO, SPO, BMS); however, all the models were trained with eight years of data, and are thus still capable of producing accurate predictions.

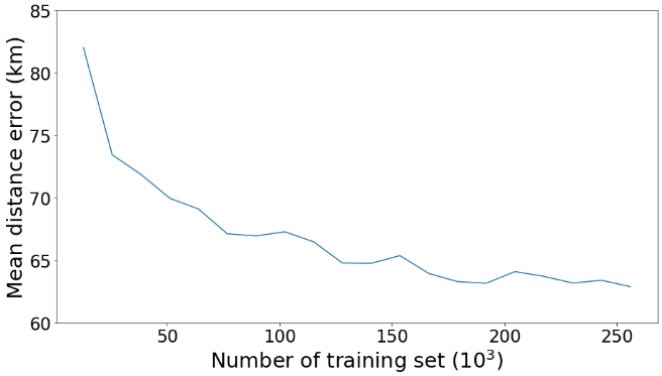


Figure 4-12. Effect of dataset size on the deep learning performance in Malacca Strait (MS)

4.6. Third Experiment: Application of the DL Model

4.6.1. Experimental Setup

In the third experiment, a DL model was built for the Malacca Strait (MS) area, and a sample location was selected in which the number of ships was counted each day consecutively, resembling a simulation of practical application. In the MS area, ships traveled approximately 417 km daily on average (see Table 4-3). Thus, the Malacca Strait itself, between Malaysia and Indonesia (95.2°E–103.2°E longitude and 0.5°N–6°N latitude), was selected as the sample location, covering an area of half a million square kilometers. Figure 4-13 shows the sample locations in the MS area. As the highest density area, Singapore waters were not included in the sample because there would be many outliers in the northeast and southeast portions that enter the observed sample location. Nonetheless, long-term ship behavior near ports can be observed on Malaysia's Teluk Rubiah port inside the location.

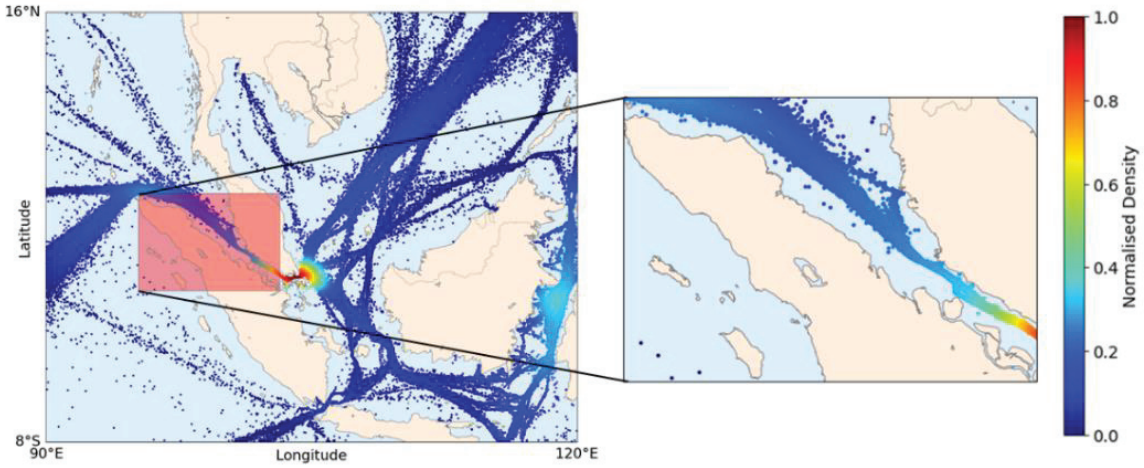


Figure 4-13. Location of the observation area (red-filled square): the Malacca Strait; the normalized density distribution is calculated based on the test set

For the application, uninterrupted AIS data for a period of time were required as the test set. Consequently, the previous hold-out validation split cannot be applied because the test and dev sets are equivalently derived from shuffled 2018 data; in addition, the training and dev sets have to be generated in the same manner as the models in the previous experiment to produce a generalized DL model similar to the previous experiment.

Accordingly, a different scheme for the hold-out validation split was developed. Randomly shuffled AIS data from July 2010 to September 2017 were used as the training set, while randomly shuffled data between October 2017 and September 2018 were used as the dev set. The last three-month data were not randomized and were retained chronologically as the test

set, as shown in Figure 4-14; the predictions were processed sequentially, day by day, rather than instantaneously. Subsequently, the number of ships was counted at the selected locations every day.

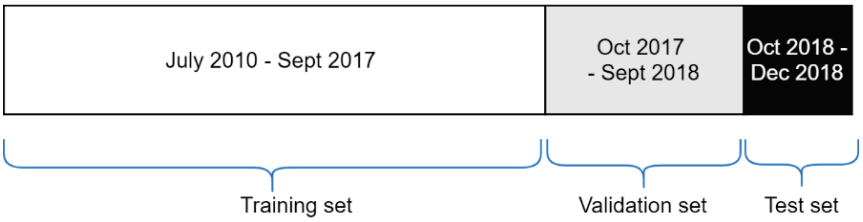
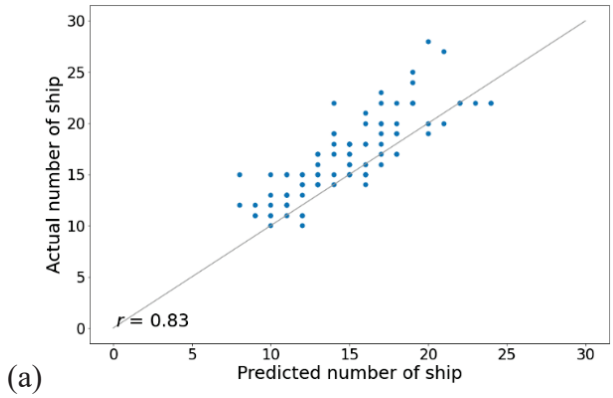
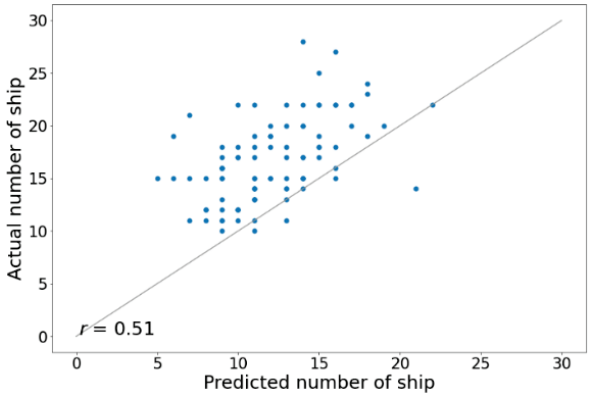


Figure 4-14. Hold-out validation split for the third experiment (non-proportional scale)

4.6.2. Results and Discussions



(a)



(b)

Figure 4-15. Fitted plots of the daily prediction of the number of ships in the Malacca Strait from October to December 2018: (a) DL model and (b) geodesic calculation

The fitted plots of the overall results from the DL models and geodesic calculations were compared to the actual data, as shown in Figure 4-15. The results are the daily prediction of the number of ships inside the Malacca Strait from October to December 2018. The r-value measures the linear relationship between the predictions and actual values. Compared to the actual data, a strong relationship with the r-value of *0.83* from the DL model confirms that

predictions with an average time interval of 24 h using deep learning are possible, which is contrary to the belief that the motion-based method is not effective for long-term vessel location prediction. By using proper DL systems and massive data, this result confirms that an accurate long-term prediction of the vessel position can be generated with the straightforward motion-based method. In contrast, the geodesic calculation showed a weak relationship with an r-value of *0.51*, suggesting that conventional calculations cannot be used for the long-term position projection of a vessel moving in congested waters or near the ports.

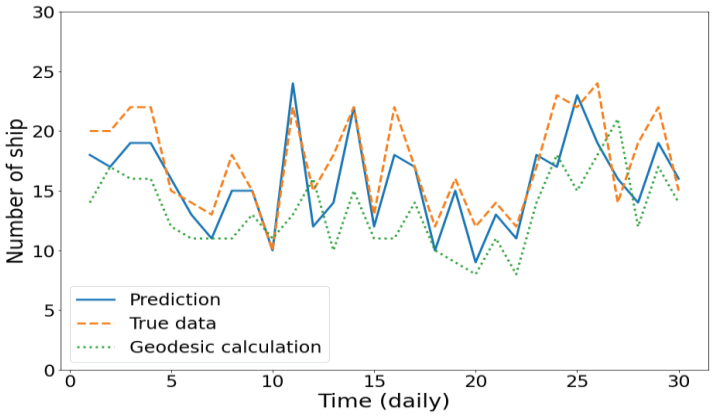


Figure 4-16. Daily prediction of the number of ships in the Malacca strait in November 2018

Moreover, Figure 4-16 shows the daily prediction of the number of ships inside the sample location for the entire month of November 2018; the estimated target values from the DL model and geodesic calculations were compared to the ground truth target values. The DL model proved that it can accurately predict the quantity inside the location for an entire day without additional information regarding historical navigational status. Meanwhile, as the geodesic calculation failed to predict the long-term position of a vessel moving in congested waters or near the ports, it also failed to predict the number of ships.

Detailed ship position predictions inside the sample location are shown in Figure 4-17. The ship position prediction from the DL model and geodesic calculation in the Malacca Strait was compared with the actual position on November 10 and 11, 2018. In this location, most ships were moving, and some were idle near Malaysia's Teluk Rubiah port. Examining the actual data on November 9, there was a line of ships moving from the northwest to the southwest, three of which were separated at the rear. These three ships moved to the southwest the next day, and another line of ships appeared from the northwest, which eventually was the start of a congested line of moving ships that appeared on the 11th. The DL model closely mimicked the position of these ships and the movement of the ground truth targets on November 10 and 11, even for ships idling at the port. The number of ships and their positions

were reliably predicted by the model. From these results, the dl model appears to have a sense of the dimension of the geographic coordinate system that can be or is often passed. In contrast, geodesic calculation failed to predict the position of the ships either in the moving position or idle near ports, and also resulted in a poor estimate of the number of ships inside the area.

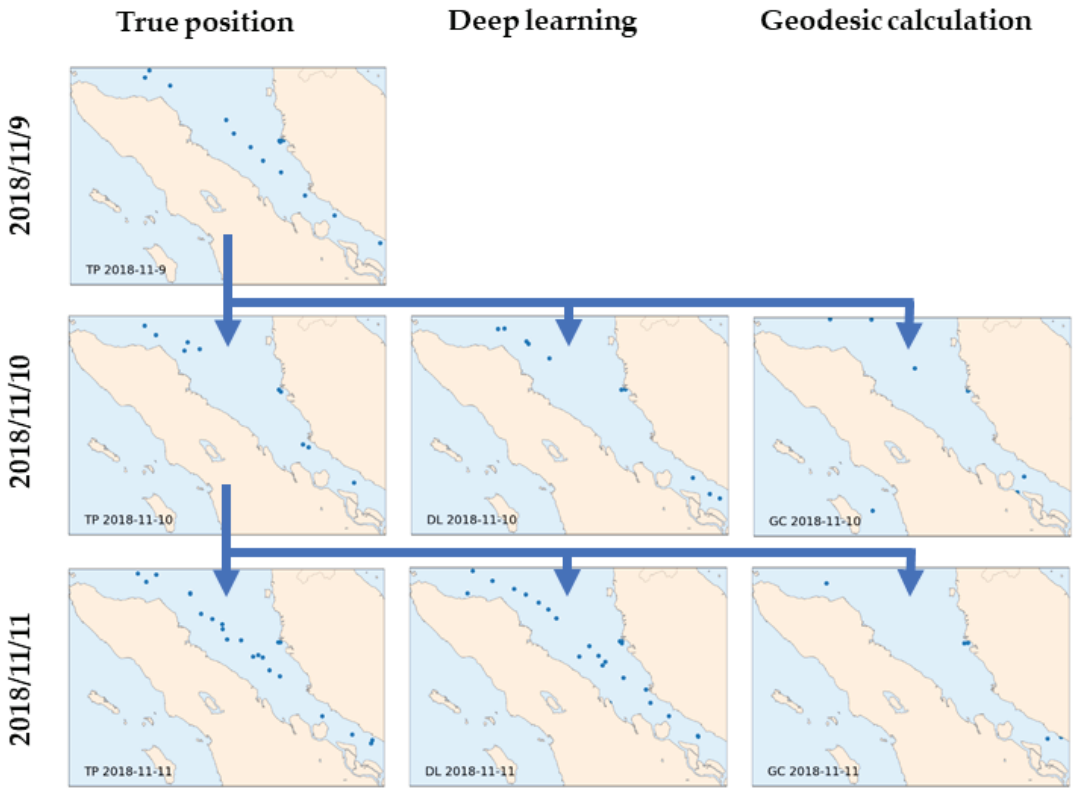


Figure 4-17. Ship positions in the Malacca strait on three consecutive days in November 2018: TP is the true position (left), DL is the deep learning prediction (center), and GC is the geodesic calculation (right)

4.7. Discussions on Performance of End-to-end Deep Learning Models

4.7.1. Comparison of End-to-end DL Models

The MLP model shows a remarkable performance compared to the ensemble-based ML models and conventional approach from the dataset with the long-time-interval distribution. End-to-end means that the approach does not need a trajectory reconstruction and other preprocessing/postprocessing steps for making a prediction.

To further clarify the robustness of end-to-end DL-based models, we built two DL-based models: Deep RNN and LSTM. These sequence-based DL architectures were built with the same input, output, and validation splits to the MLP model so that their performances are broadly comparable (the structure was also constructed as closely as possible to the MLP model

while also tuned to the most optimized). The input does not go through an interpolation or trajectory reconstruction to ensure an equivalent model; it already carries information regarding time interval Δt and a rough distance d_t to the target. The Deep RNN and LSTM models have five hidden recurrent layers with recurrent units of [64, 64, 32, 32, 16] and a one-unit linear output layer; the model was trained in mini-batch with the loss function of mean absolute error (MAE) optimized by Adam. The stacked deep recurrent layers were easily prone to overfit the training set (LSTMs are much easier), which was alleviated using a combination of recurrent dropout and recurrent layer normalization.

The Indian Ocean and Malacca Straits areas similar to the second experiment, representing the simple and complex movement of ships on the ocean and chokepoint areas, respectively, were selected as a definitive test for the DL-based models. The overall prediction methodology of these two recurrent models is similar to the second experiment. The models were also set to generalize on the data at which the training set was shuffled. The number of steps was considered one-time steps: the input only contains information at time step t (the implicit information of $t-1$ is included) to make a prediction at $t+1$, and thus this resembles a multivariate time series with a single time step.

Table 4-4. Comparison of the deep learning model performance

Areas	**Methods**	Loss		$\boldsymbol{MDE}$	$\boldsymbol{MaE}$	**Training**
		$\boldsymbol{MAE_\lambda}$	$\boldsymbol{MAE_\varphi}$	(km)	(degree)	**Time** (s)
Indian Ocean (IO)	MLP	0.174	0.149	27.0	1.7	1,699
	RNN	0.176	0.150	27.4	1.7	2,758
	LSTM	0.177	0.148	27.3	1.7	1,845
Malacca Strait (MS)	MLP	0.376	0.337	62.8	16.9	4,257
	RNN	0.384	0.335	63.6	16.6	6,561
	LSTM	0.381	0.336	63.4	16.2	6,243

The performance of the DL-based models is presented in Table 4-4. The results of the proposed MLP model in the ocean and chokepoint areas are slightly more accurate than the other DL architectures. However, with only a *1%* improvement, the performance of the feed-forward model is considered similar to the recurrent models (RNN and LSTM) that have more trainable parameters; applying more computational cost and complex models appear to negatively improve the performance. The MLP model took the fastest total training times on both areas with the least parameters than the other DL models. These results were obtained on a single Intel i7-9700K processor (8 CPU @3.60 GHz).

The patterns observed result from the small sequences (time steps), at which the regular feed-forward network is sufficient for this task. This is confirmed as the application of the LSTM model did not deliver better performance than the RNN model since LSTMs capability is much better at handling long sequential data than simple RNNs with limited short-term memory. The results correspond to Gers et al. (2002) that tested RNNs and LSTMs for time-series predictions, but the results were poor as simple multilayer perceptrons (MLPs) often outperformed LSTMs when applied to the same data [57].

4.7.2. Comparison of End-to-end DL Models on Small Dataset

A small dataset with short-time intervals near the ports was used to further analyze the robustness of the end-to-end DL models. The dataset was retrieved from the ATD2019 challenge dataset[6]. All AIS training and challenge data with different properties were combined and used to construct a dataset with varying time-interval distribution. Figure 4-18(a) shows the short-time-interval distribution of the small dataset averaging around 10 seconds intervals; most of the data have near-real-time intervals, and the rest varies up to 60 seconds. The dataset consisted of ships movements near ports located around north of Norfolk, Virginia, North America (see Figure 4-18(b)). In this dataset, course over ground ψ_{og} is given but not heading ψ_h, and therefore, to accommodate the same 10 input features as the previous experiments, the heading was copied from the course over ground.

The dataset was split into a training set and dev-test sets according to the hold-out validation split with 80% data for training and the rest 20% for validation and test. The training set comprised more than 90,000 data records, whereas each dev and test set contained more than 10,000 data records. The size of these sets was considered sufficiently large, thus providing high confidence to the overall result.

[6] https://gitlab.com/algorithms-for-threat-detection/2019/atd2019.

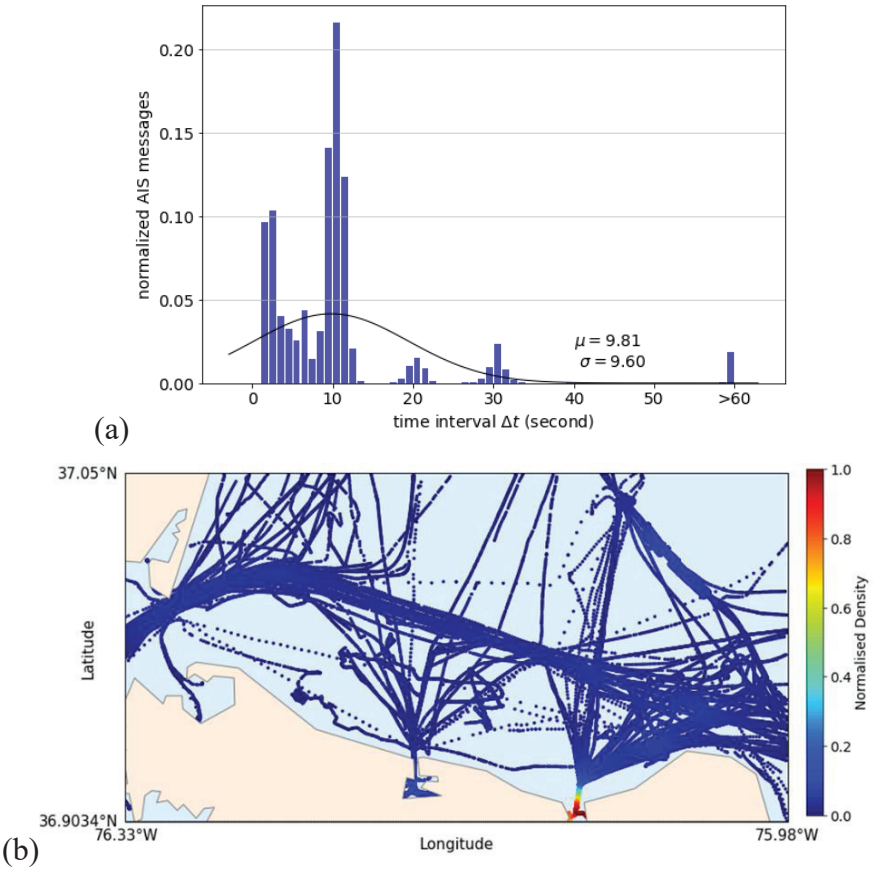


Figure 4-18. The small dataset properties: (a) Time-interval distribution; (b) Normalized density distribution

The results show that compared to the conventional approach, the DL-based models are not as effective as on the long-term prediction task, and the LSTM model delivers the same performance as the feed-forward networks since the time steps are limited (single time steps with 10 dimensions). This limited sequence allows the feed-forward networks to deliver optimum performance but would not maximize the potential of the sequence-based recurrent networks. Moreover, on the dataset with varying time-interval distribution and without trajectory reconstruction, the feed-forward networks generated predictions as accurate as of the LSTM with faster training times and less trainable parameters.

Table 4-5 shows the results from the DL models and geodesic calculation on the small dataset. According to the second experiment results, the pattern of *MaE* on the small dataset confirms that the dataset consists of vessel behavior near ports with short-distance intervals such as berthing, idling, and maneuvering. The average distance interval between two consecutive AIS points is *24.6 meters*. At this small distance, the geodesic calculation is qualitatively similar to the dead reckoning calculation with an error of around *3.5 meters*. The DL models are *8%* better than the geodesic calculation. Naturally, LSTMs can generate the

most accurate short-term predictions of vessel position; however, on this dataset with the uneven distribution and without trajectory reconstruction, the MLP model generated predictions as accurate as of the LSTM with almost half of the training times.

The results show that compared to the conventional approach, the DL-based models are not as effective as on the long-term prediction task, and the LSTM model delivers the same performance as the feed-forward networks since the time steps are limited (single time steps with 10 dimensions). This limited sequence allows the feed-forward networks to deliver optimum performance but would not maximize the potential of the sequence-based recurrent networks. Moreover, on the dataset with varying time-interval distribution and without trajectory reconstruction, the feed-forward networks generated predictions as accurate as of the LSTM with faster training times and less trainable parameters.

Table 4-5. Comparison of the model performance on the small dataset

Methods	Loss		***MDE***	***MaE***	**Training**	**Total**
	$\boldsymbol{MAE_{\lambda}}(10^{-5})$	$\boldsymbol{MAE_{\varphi}}(10^{-5})$	(m)	(degree)	**Time** (s)	**Params**
Average distance int.	19.7	12.1	24.6	-	-	-
Geodesic calculation	2.4	2.1	3.5	29.2	-	-
MLP	2.3	1.8	3.2	31.5	922	8,545
RNN	2.4	1.8	3.3	31.4	2,537	19,041
LSTM	2.3	1.7	3.2	31.4	1,693	76,113

4.8. Conclusions on one-day interval of ship position prediction

This research demonstrated that predictions with an average time interval of 24 h are possible, confirming that the straightforward motion-based method can generate an accurate long-term prediction of the vessel position. The DL models generated more accurate predictions than the geodesic calculation as the baseline model in all areas. Predictions in the open ocean areas yielded higher accuracy than in the chokepoint areas; however, compared to the geodesic calculation, the improvement scores were higher in the chokepoint areas than in the ocean areas since the geodesic calculation failed to predict vessel behavior near the ports and congested waters. The DL model can predict the complex movement of ships near ports and congested routes, whereas conventional calculations fail, with no information regarding vessel status, historical trajectory, or destination. The last experiment demonstrated that the DL model appears to have a sense of the dimension of the geographic coordinate system that can be or is often passed, wherein the chokepoint areas rely more on the input features of the current

latitude and longitude of the vessels. Moreover, on the dataset with varying and uneven time-interval distribution and without a trajectory reconstruction, the proposed MLP model generated predictions as accurate as the LSTM with faster training times.

This research can be a basis for future research for further improvement or even longer time intervals. Additionally, the generated predictions can be directly utilized for any further actions depending on the maximum error threshold the prediction model allows; the final model can then be launched into production. After launching, the model can be trained incrementally with new data on the fly in line with the growth of AIS data over time, further improving the model.

Chapter 5
Several-day Intervals of Ship Position Prediction

5.1. Overview of Several-day Intervals of Ship Position Prediction

The previous chapter demonstrates that a one-day interval prediction of vessel position using the straightforward motion-based method is feasible with deep learning, more accurate than another machine learning algorithm, and much more accurate than the conventional approach. In the previous chapter, the selected observation size is large enough to capture the long-term movement of ships, with the time interval averaging 24 h. However, the optimum or minimum coverage size to capture the long-term ship movement according to the respective time intervals is not yet established.

This research continues the work in the previous chapter, utilizing deep learning and AIS data to predict ship position at four different time intervals: from 24 h to 96 h. This chapter aims to achieve up to a 4-d interval of ship position prediction and resolve the topics regarding optimum coverage size on each magnified time interval. As the distance traveled by moving ships in a straight line is proportional to the time interval, if the time interval is magnified, the observation area size must also be expanded or magnified. Additionally, to cope with the extensive coverage, hence data handled, we utilized a new deep learning model with deeper layers and bigger networks than the model used in the previous chapter.

This research consists of three experiments followed by discussions regarding the single-worldwide model and the limitations of the prediction models. In the first experiment, larger networks of DL models were trained on different observation area sizes with magnified time intervals. We focused more on observation size variations in the next experiment, increasing 4 and 3.2 degrees on longitude and latitude, respectively. In the last experiment, five DL models were built for five major locations of maritime shipping routes focused on Capesize bulk carriers. In the discussion section, a single-worldwide model was executed, and research limitations such as outliers and AIS data uncertainty are discussed. Finally, the research's conclusion is presented at the end of this chapter.

The several-day intervals of ship position prediction in this thesis is detailed in the following chapter.

5.2. Methodology

The overall method for this research is almost similar to the previous chapter. The differences are that the set of input features X_i and target features y_i comprising vessel positions were created based on the next time step $t + t_t$, where t_t is the time-interval target (in days). Moreover, a larger deep learning model was utilized as a vector regression model, generating both the displacements in longitude and latitude directions at the same time.

5.2.1. Model Structure and Properties

Through extensive experimentation and evaluation based on the previous research, we developed a deep feed-forward model with the objectives of performance and generalization on the wider observation area and time intervals. The deep MLP architecture was constructed as a vector regression model to solve regression problems by setting the target as two scalar outputs (vector): the displacements of longitude Δλ and latitude Δφ directions. Compared to the previous research that set the target as a scalar value, hence one model for each target, this model can solve the same problem with half of the running time.

The model used a deeper and bigger network with eleven layers: triangle-shaped ten hidden layers with selu activation function and LeCun initialization and a two-unit linear output layer [98, 98, 88, 58, 44, 43, 32, 28, 26, 16, 2]. The mean absolute error (MAE) was still employed as the loss function since it is less sensitive to outliers, particularly when used in the larger observation area. A larger mini-batch size of 256 was used for training to fasten computation time with GPU as more training data were utilized extensively. Similarly, we used an adaptive learning rate method involving the adaptive moment estimation optimization algorithm (Adam); the learning rate was automatically reduced by a factor of 0.5 once the validation loss stagnated for 20 epochs. A validation-based early stopping was set to interrupt training once the validation stopped improving for 100 epochs. Moreover, a hold-out validation split was applied similar to the previous research.

5.2.2. Input and Target Features

As the overall method of ship position prediction with several-day intervals is similar to the previous research with one-day intervals, the input matrix X_i is still composed of the same 10 features or vessel states at the current time step t. Meanwhile, the target vector y_i is composed of the displacements of longitude $\Delta\lambda$ and latitude $\Delta\varphi$ between the current time step t and the next time step $t + t_t$, following the DL output as a vector regression model. The input

matrix and target vectors are retrieved from the SQL database given a boundary condition of location and time, expressed as follows:

$$X_i = \left\{\lambda_t^i, \varphi_t^i, {U_{og}}_t^i, {\psi_{og}}_t^i, {\psi_h}_t^i, \Delta t_t^i, d_t^i, {d_h}_t^i, {d_m}_t^i, {U_{avg}}_t^i\right\} \tag{1}$$

$$y_i = \left\{\Delta\lambda_t^i, \Delta\varphi_t^i\right\} \tag{2}$$

where $i = \{0,1,\dots,m\}$, $\Delta\lambda_t = \lambda_{t+t_t} - \lambda_t$ is the longitude interval, $\Delta\varphi_t = \varphi_{t+t_t} - \varphi_t$ is the latitude interval, subscript t represents the time step, and subscript t_t represents the target interval in day(s).

Similarly, input features of λ, φ, U_{og}, ψ_{og}, and ψ_h are extracted directly from the archived AIS messages at the current time step t. Some of the remaining inputs provide implicit information on the position history at the previous time step $t - t_t$. The target timestamp (at the next time step $t + t_t$) is represented as a time interval Δt and a rough distance to the target d_t. They are defined as follows:

$$\Delta t_t = T_{t+t_t} - T_t \tag{3}$$

$$d_t = \Delta t_t {U_{og}}_t \tag{4}$$

$$d_{h_t} = d_{h\left((\varphi_{t-t_t},\lambda_{t-t_t}),(\varphi_t,\lambda_t)\right)} \quad where\ \varphi,\lambda\ in\ radians \tag{5}$$

$$d_{m_t} = d_{h\left((\varphi_{t-t_t},\lambda_{t-t_t}),(\varphi_{t-t_t},\lambda_t)\right)} + d_{h\left((\varphi_{t-t_t},\lambda_{t-t_t}),(\varphi_t,\lambda_{t-t_t})\right)} \tag{6}$$

$${U_{avg}}_t = {d_h}_t/\Delta t_t \tag{7}$$

where T is the timestamp of the AIS message, and T_{t+t_t} is referred to as the prediction timestamp.

After extracting the input and target at high speed, all input features X_i and target features y_i are standardized. The standard scores are calculated as follows:

$${X'}_i = std(X_i) = X_i - \bar{X}_\iota/\sigma_{X_i} \tag{8}$$

$${y'}_i = std(y_i) = y_i - \bar{y}_\iota/\sigma_{y_i} \tag{9}$$

where according to the corresponding features, the accent $\bar{}$ is the mean, σ is the standard deviation, and the superscript $'$ is the standard score. The standardized input and target features $({X'}_i, {y'}_i)$ are then fed into a DL model.

5.2.3. Performance Evaluation Metrics

Loss Score

After the DL model generated vessel position, the prediction was inversed using the standardization parameters of the target features as follows:

$$\hat{y}_i = std^{-1}(\hat{y}'_i) \tag{10}$$

where $\hat{y}_i$ is the prediction vector after inverse normalization, and $\hat{y}'_i$ is the prediction vector. The results ($\Delta\hat{\lambda}_t$ and $\Delta\hat{\varphi}_t$) are then transformed into latitude $\hat{\varphi}_{t+t_t}$ and longitude $\hat{\lambda}_{t+t_t}$ by adding each component to the input (i.e., λ_t, φ_t). Finally, a new loss score for each result is calculated as follows:

$$MAE_\lambda = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\lambda}^i_{t+t_t} - \lambda^i_{t+t_t}\right| \tag{11}$$

$$MAE_\varphi = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\varphi}^i_{t+t_t} - \varphi^i_{t+t_t}\right| \tag{12}$$

where $\hat{\lambda}^i_{t+t_t} = \left(\Delta\hat{\lambda}_t + \lambda^i_t\right)$ is the prediction longitude, $\hat{\varphi}^i_{t+t_t} = \left(\Delta\hat{\varphi}_t + \varphi^i_t\right)$ is the prediction latitude, $\lambda^i_{t+t_t}$ is the target longitude, and $\varphi^i_{t+t_t}$ is the target latitude.

Metric Score

We used the same metric scores from the previous research, defined as the mean distance error (MDE) and mean angular error (MaE). The distance error is the haversine distance between the prediction and target (true position) in kilometers, signifying the deviation of the result (prediction) from the true position (target). The MaE as the second evaluation metric was calculated based on the three known points of the base, target, and predicted positions, signifying precision of the prediction from the target. These metric scores were calculated as follows:

$$MDE = \frac{1}{m}\sum_{i=1}^{m} {d_h}^i_{(v,w)} \tag{13}$$

$$MaE = \frac{1}{m}\sum_{i=1}^{m} C^i_{(u,v,w)} \tag{14}$$

$$C_{(u,v,w)} = \arccos\left(\frac{cos\left(d_{h_{(v,w)}}\right) - cos\left(d_{h_{(u,v)}}\right)cos\left(d_{h_{(u,w)}}\right)}{sin\left(d_{h_{(u,v)}}\right)sin\left(d_{h_{(u,w)}}\right)}\right) \tag{15}$$

where u is the base point (φ_t, λ_t), v is the true position $\left(\varphi_{t+t_t}, \lambda_{t+t_t}\right)$, and w is the predicted point $\left(\hat{\varphi}_{t+t_t}, \hat{\lambda}_{t+t_t}\right)$.

5.3. Baseline Model

The previous research is used as a baseline of this research for model-building guidelines especially in the prototyping process. The baseline was defined as a DL model with the same structure and hyperparameters as the previous research, presented in the previous chapter [58]. We also utilized a naive baseline, geodesic calculation, as a non-ML baseline, computed as follows:

$$\breve{\varphi}_{t+t_t} = \text{asin}\left(\sin(\varphi_t)\cos(\delta_t) + \cos(\varphi_t)\sin(\delta_t)\cos\left(\psi_{h_t}\right)\right) \tag{16}$$

$$\breve{\lambda}_{t+t_t} = \lambda_t + \text{atan2}\left(\sin\left(\psi_{h_t}\right)\sin(\delta_t)\cos(\varphi_t), \cos(\delta_t) - \sin(\varphi_t)\sin\left(\breve{\varphi}_{t+t_t}\right)\right) \tag{17}$$

where $\varphi_t, \lambda_t, \psi_{h_t}$ are in radians, $\delta_t = d_t / r_{earth}$ is the angular distance, r_{earth} is the earth diameter, $\breve{\varphi}_{t+t_t}$ and $\breve{\lambda}_{t+t_t}$ are subjected to normalization to degrees [49]. These baseline models were used to conduct error analysis on bias and variance.

Likewise, the average distance interval of the ships was adopted as a sanity check. This approach examines the performance evaluation metrics between the current position and the next position, calculated as follows:

$$\overline{d_h} = \frac{1}{m}\sum_{i=1}^{m} d_h{}^{i}_{\left((\varphi_t, \lambda_t), (\varphi_{t+t_t}, \lambda_{t+t_t})\right)} \; where\ \varphi, \lambda\ in\ radians \tag{18}$$

where (φ_t, λ_t) is the base point, and $\left(\varphi_{t+t_t}, \lambda_{t+t_t}\right)$ is the next position at the designated time interval.

5.4. First Experiment: Time-Interval Magnification

5.4.1. Experimental Setup

In the first experiment, we trained the DL model on four different time intervals prediction: from averaging 24 h to 96 h. The model was trained on four different variations of observation area size at each time interval. Naturally, the distance traveled by moving ships in a straight line is proportional to the time interval. Therefore, the variation in the observation area size is increased twofold as the time interval doubles, and so on. Table 5-1 shows the size variation of the observation area at each time interval.

We focused our observation on Malacca Strait and its surrounding area. The strait is the most congested primary chokepoint, where it is one of the core routes of the main maritime shipping route [59]. The observed area covers from 1 to 204 million square kilometers from the smallest observation area at 24-h average time intervals Δt_t to the largest observation area at 96-h average time intervals, respectively, adequately capturing the continuous long-term movements of the vessels conformed to its time intervals.

Table 5-1. Variation of time intervals and its coverage size

Time intervals (in average)	Index area	Longitude	Latitude	Area size[7] (10^6 km^2)	Train set size (10^3)	Dev/test set size (10^3)
24 h	1	95°E–105°E	0°S–8°N	1	48	5
	2	90°E–110°E	4°S–12°N	4	107	10
	3	85°E–115°E	8°S–16°N	9	172	15
	4	80°E–120°E	12°S–20°N	16	367	33
48 h	1	90°E–110°E	4°S–12°N	4	82	7
	2	80°E–120°E	12°S–20°N	16	316	28
	3	70°E–130°E	20°S–28°N	35	763	77
	4	60°E–140°E	28°S–36°N	60	1352	138
72 h	1	85°E–115°E	8°S–16°N	9	123	11
	2	70°E–130°E	20°S–28°N	35	700	71
	3	55°E–145°E	32°S–40°N	75	1544	158
	4	40°E–160°E	44°S–52°N	126	1892	190
96 h	1	80°E–120°E	12°S–20°N	16	222	20
	2	60°E–140°E	28°S–36°N	60	1280	133
	3	40°E–160°E	44°S–52°N	126	1878	189
	4	20°E–180°E	60°S–68°N	204	2039	204

According to the hold-out validation scheme, the dataset was split into a training set and dev-test sets. In the smallest observation area, the training set comprised more than 48 thousand data records after omitting AIS messages with missing values and abnormalities; each dev and

[7] calculated by geodesic area assuming the earth is a perfect sphere with a radius defined in WGS84.

test set contained more than five thousand data records. Meanwhile, more than 2 million AIS data for training was used in the largest observation area, with 200 thousand more for testing. The size of these sets was considered sufficiently large for each condition, where the smallest area is already above the 25,000 thresholds, thus providing high confidence in the overall model performance [58].

5.4.2 Results and Discussions

Time interval = 24-h average

Figure 5-1 demonstrates four size variations of the observed area with a 24-h average time interval. The normalized density distributions were calculated using Gaussian kernel density estimation (KDE) with the largest observed area dataset from nine years of AIS messages involving 1662 different IMO numbers (vessels). For geospatial data processing, the background geography was visualized using an open-source Python library, the Matplotlib Basemap Toolkit [51]. The smallest observation area captures the densest region, the narrow bottom-end of Malacca Strait and around Singapore Strait. Meanwhile, the largest observation area covers Makassar Strait, a secondary maritime chokepoint or the second densest region.

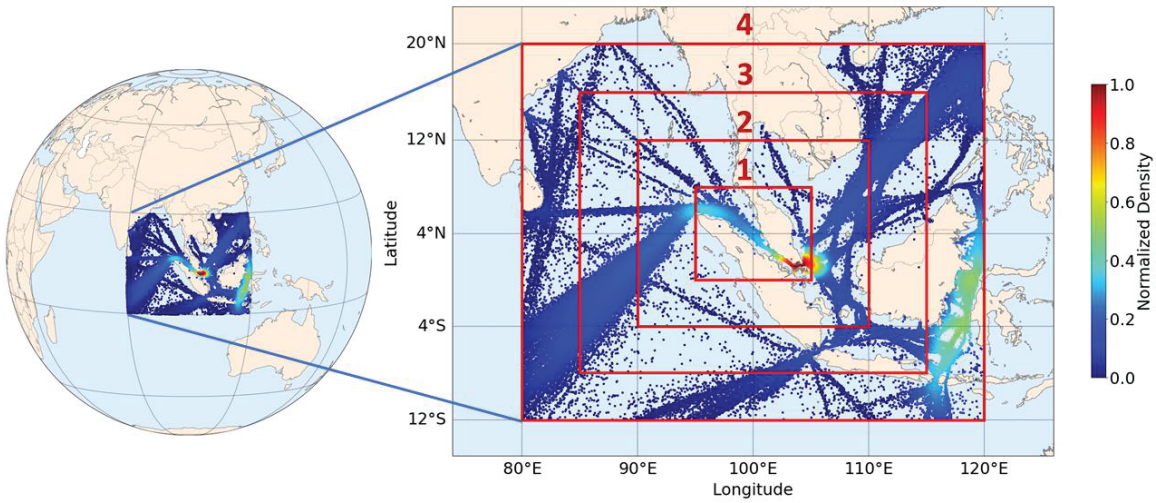


Figure 5-1. Size variations of the observed area with 24-h time-interval average (the normalized density distribution of AIS data is based on the largest area)

The performance evaluation metrics of the DL and baseline models are given in Table 5-2; all results were computed using the test set. The average distance interval indicates the average distance between ships' current and subsequent positions in their corresponding time-interval window. Similar to the geodesic calculation performance on the maritime chokepoint areas from the previous research, the geodesic calculation fails to make predictions of vessel behavior

in this high-traffic area. Both DL models effectively outperformed (accuracy and precision) the non-ML baseline on all area size variations.

Table 5-2. Comparison of model performance in 24-h average time interval

Index area	DL model				Previous model		Geodesic calc.		Average distance interval (km)
	Loss		***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	
	MAE_λ	MAE_φ							
1	0.40	0.26	55	23.7	56	24.5	182	39.7	288
2	0.48	0.35	72	17.9	75	18.8	202	31.3	377
3	0.42	0.31	63	16.1	64	16.4	162	26.0	403
4	0.29	0.27	49	13.3	50	13.2	128	20.3	424

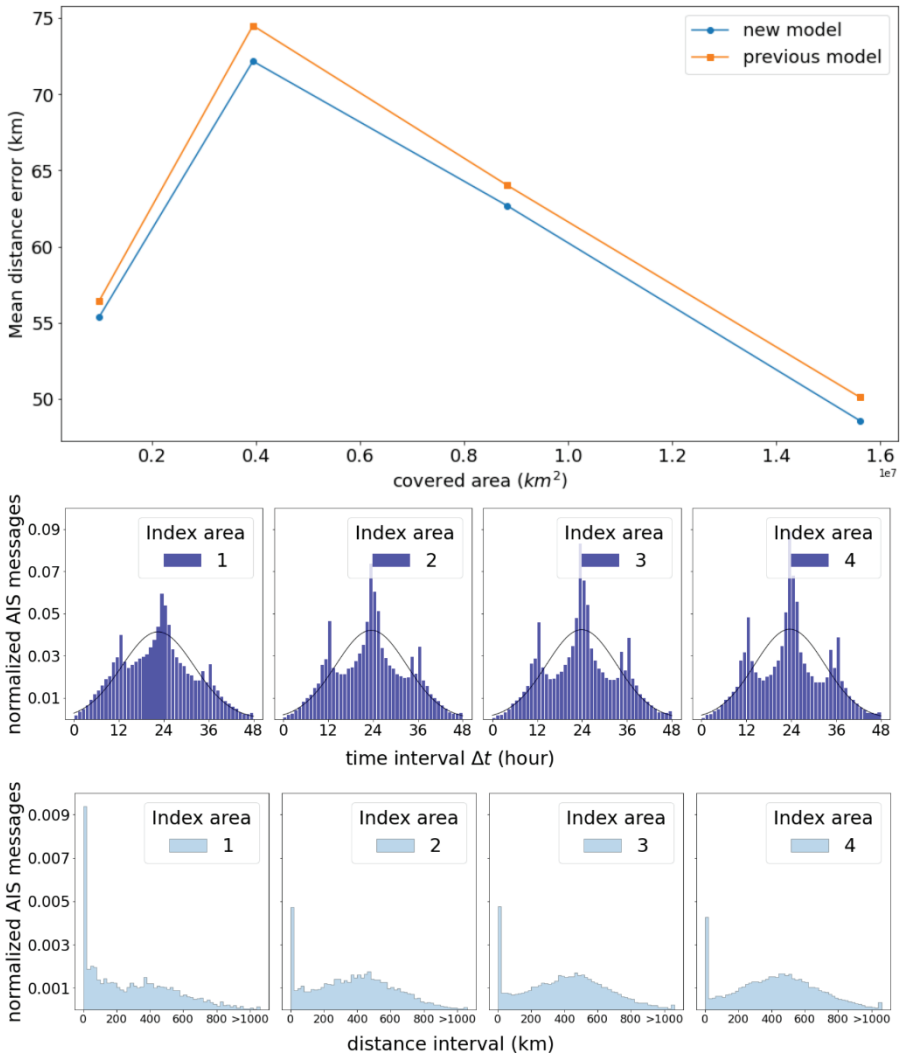


Figure 5-2. Comparison of the DL model performance with the previous model and characteristics of each observed area dataset in 24-h average time interval

Figure 5-2 reveals the mean distance error from the results of the DL model with deeper and larger networks and the previous research's model, along with the dataset characteristics in each area. The DL model constantly generated more accurate and precise predictions than the previous model on all area size variations, which exceeded *2.5%* and *2.3%* on accuracy and precision, respectively. Generally, the models' performance improves with larger covered areas. The irregularity of the smallest area compared to other areas corresponds to the surprisingly higher data related to ships near ports and high-traffic areas in the surrounding Singapore Strait than ships in normal sailing movement, indicated by the high ratio of the low-distance interval.

Time interval = 48-h average

Figure 5-3 depicts four size variations of the observed area with a 48-h average time interval. The normalized density distributions were calculated with the largest observed area dataset from AIS messages involving 1693 vessels over nine years. By the third-largest observation area (index area number 3), the Indian Ocean surrounding Port Hedland, the site of the highest tonnage port in Australia, becomes the densest zone eclipsing Malacca Strait. This transition is caused since the observed AIS data were taken from only Capesize bulk carriers that mainly transport coal, ore, and other commodity raw materials, the largest ships dedicated to ore transportation [60]. This particular fit perfectly with the fact that it is the world's largest bulk export port and a major iron ore exporting port in Western Australia [61]. Following the densest region, the East China Sea surrounding China, South Korea, and Japan as the second place was covered in the largest observation area.

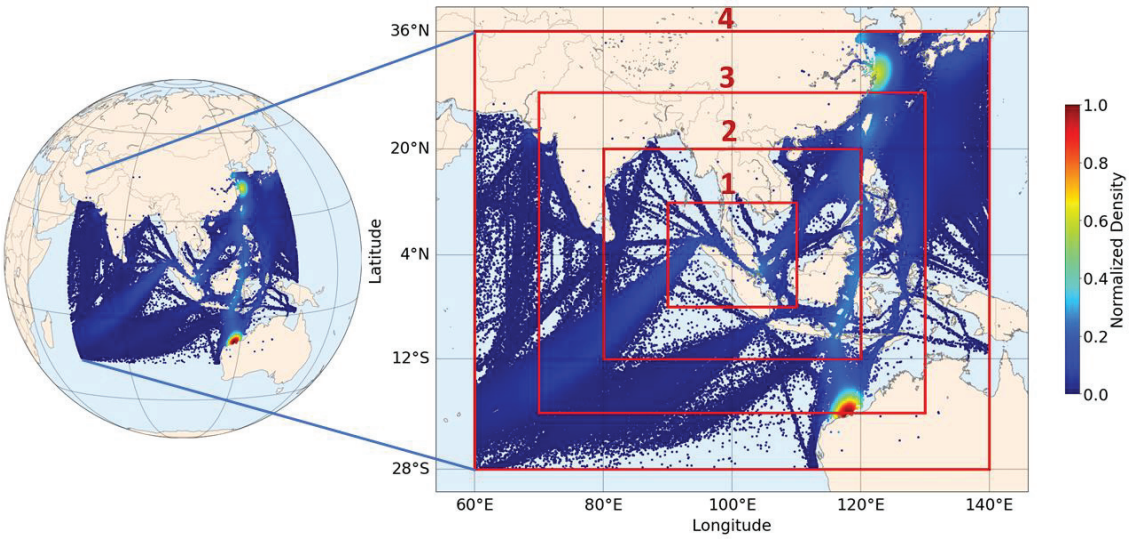


Figure 5-3. Size variations of the observed area with 48-h time-interval average (normalized density distribution of AIS data is based on the largest area)

Table 5-3. Comparison of model performance in 48-h average time interval

Index area	DL model				Previous model		Geodesic calc.		Average distance interval (km)
	Loss		***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	
	MAE_λ	MAE_φ							
1	1.04	0.73	151	16.8	150	16.8	563	42.6	686
2	0.67	0.66	116	14.0	119	15.4	354	26.8	842
3	0.61	0.66	111	14.9	115	15.4	286	23.1	870
4	0.65	0.76	122	27.0	128	27.3	294	34.3	697

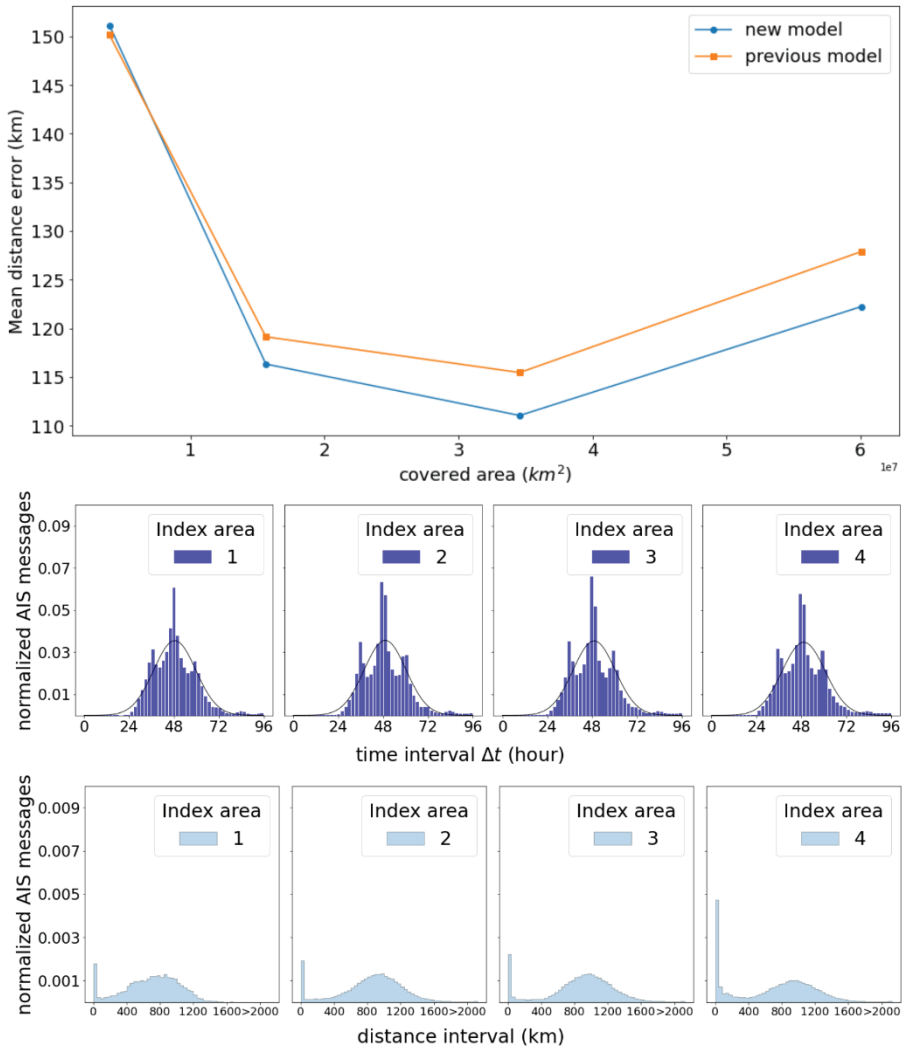


Figure 5-4. Comparison of the DL model performance with the previous model and characteristics of each observed area dataset in 48-h average time interval

The performance evaluation metrics of the DL and baseline models computed using the test set are presented in Table 5-3. Again, the geodesic calculation fails (both in accuracy and

precision) to predict vessel positions in this high-traffic area; both DL models tremendously outperformed the non-ML baseline on all area size variations. The average distance between ships' successive positions was normally increased twofold in this doubled time-interval average window (48-h average time interval with index area 1 and 2 compared to 24-h average time interval with index area 2 and 4, respectively). However, predictions from geodesic calculation grew worse dramatically than the results at the same area size with half time interval, indicating that the geodesic calculation is bound to fail for ship position prediction at this long interval.

Figure 5-4 shows the performance of the DL model and the previous model and the data characteristics in each area. The DL model generated more accurate and precise predictions than the previous model on almost all area size variations, with more *than 2.5%* accuracy and *3.1%* precision on average. The performance improves relative to the size of the area until drastically altered when the covered area at the largest. The patterns observed in the largest area may be caused by the surprisingly high low-distance interval ratio, leading to the low average distance interval compared to other areas.

Time interval = 72-h average

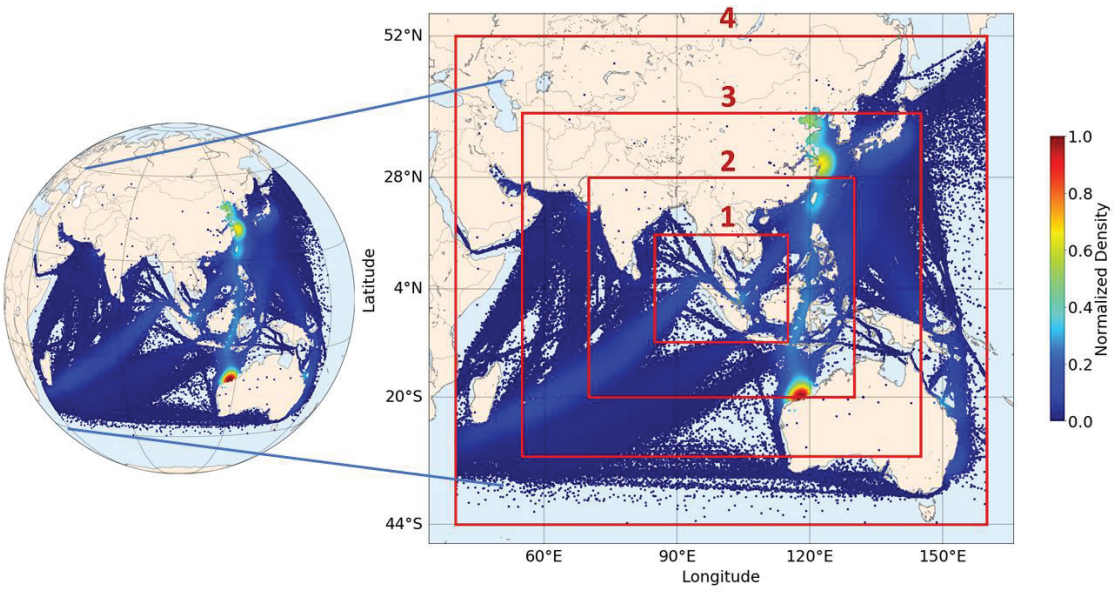


Figure 5-5. Size variations of the observed area with 72-h time-interval average (normalized density distribution of AIS data is based on the largest area)

Figure 5-5 reveals the area size variations with a 72-h average time interval. The normalized density distributions were calculated with the largest observed area dataset involving the same number of vessels at the largest area with a 48-h average time interval. The

Indian Ocean was captured completely by the largest observation area, including waters surrounding Australia, Japan, China, and Indonesia. The East China Sea and the Yellow Sea surrounding China were the second densest regions after Northwest Australia waters.

Table 5-4. Comparison of model performance in 72-h average time interval

Index area	DL model				Previous model		Geodesic calc.		Average distance interval (km)
	Loss		***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	
	MAE_λ	MAE_φ							
1	1.58	1.18	235	18.2	237	18.0	890	45.1	1081
2	1.02	1.12	188	16.0	195	16.0	522	27.1	1283
3	1.11	1.40	218	28.4	227	29.1	544	38.9	1024
4	1.10	1.35	211	28.3	225	29.2	521	38.2	1020

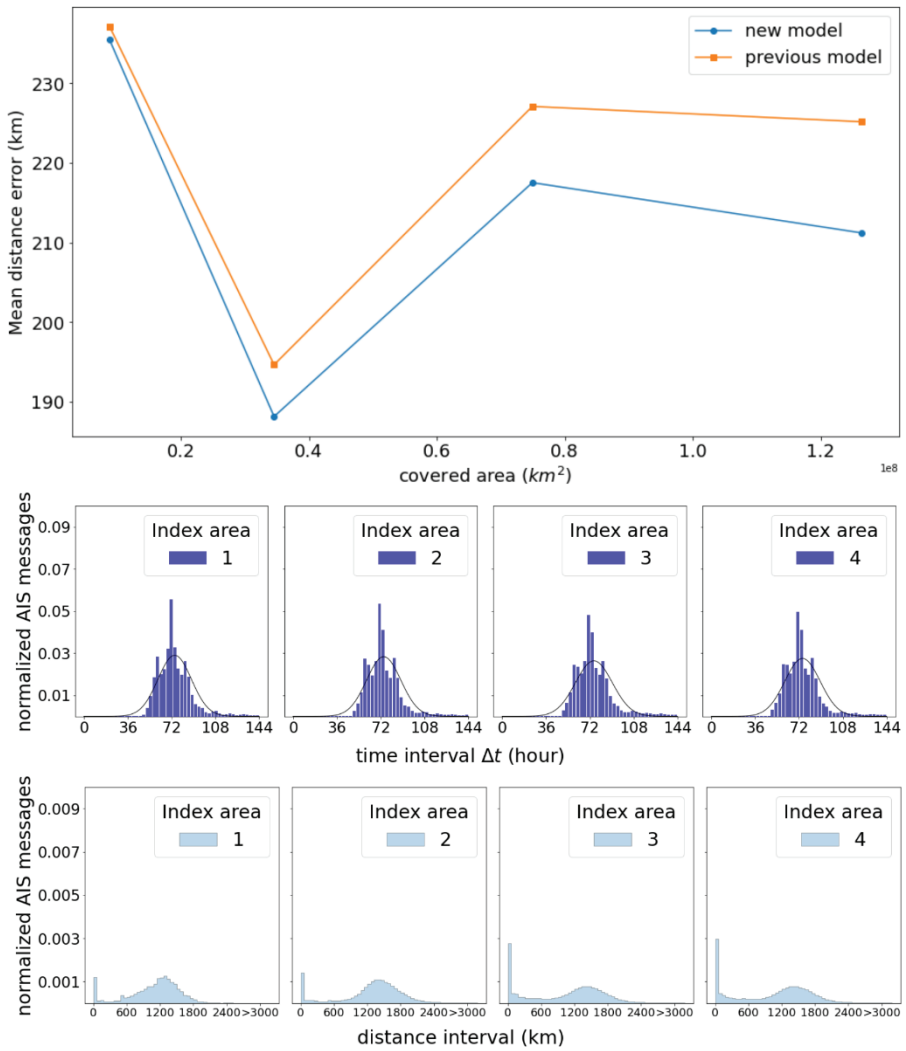


Figure 5-6. Comparison of the DL model performance with the previous model and characteristics of each observed area dataset in 72-h average time interval

Table 5-4 gives an overview of the performance of the DL and baseline models. Similar performance of geodesic calculation was generated, which is bound to fail at time intervals longer than 24 h. The DL models far exceeded the non-ML baseline performance, although naturally, the model's performance decreased as the time interval increased. As the average distance interval reached more than one thousand kilometers at this large time interval, the model performance degraded to more than one-degree error in both longitude and latitude directions.

Figure 5-6 demonstrates the performance comparison of the DL models, including the dataset characteristics in each area. Again, the DL model generated more accuracy and precision with *3.6%* and *1.1%* on average than the previous model on all area size variations, respectively. At this large time interval, the performance appears to accumulate the total errors as the coverage area expanded, indicated by the distancing performance of both models. Consequently, the larger networks improve steadily relative to the size of the area.

Time interval = 96-h average

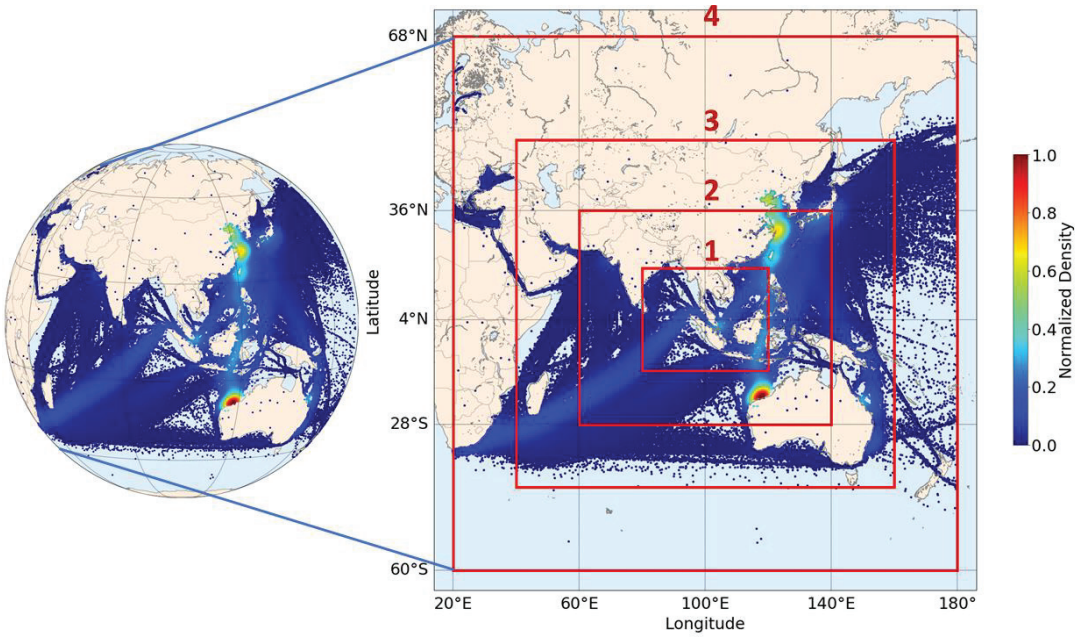


Figure 5-7. Size variations of the observed area with 96-h time-interval average (normalized density distribution of AIS data is based on the largest area)

Figure 5-7 depicts the variations of the coverage area with a 96-h average time interval. Almost all capsize bulk carriers were captured with 1695 vessels over nine years and used to calculate density distributions by the largest observation area. Almost all regions in the eastern hemisphere were covered at this most extensive coverage, including 6 out of 8 primary

chokepoints and 8 out of 13 secondary chokepoints in main maritime shipping routes [59]. Eastern Parts of the North Pacific and South Pacific Oceans were covered, including Southern parts of the Mediterranean Sea.

Table 5-5. Comparison of model performance in 96-h average time interval

Index area	DL model				Previous model		Geodesic calc.		Average distance interval (km)
	Loss		***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	***MDE*** (km)	***MaE*** (deg)	
	MAE_λ	MAE_φ							
1	1.77	1.61	293	18.1	296	18.6	1024	39.5	1484
2	1.51	1.95	304	27.1	314	27.1	813	40.6	1364
3	1.55	2.02	309	27.9	328	28.9	793	40.8	1350
4	1.57	1.99	308	28.0	322	28.0	767	39.4	1367

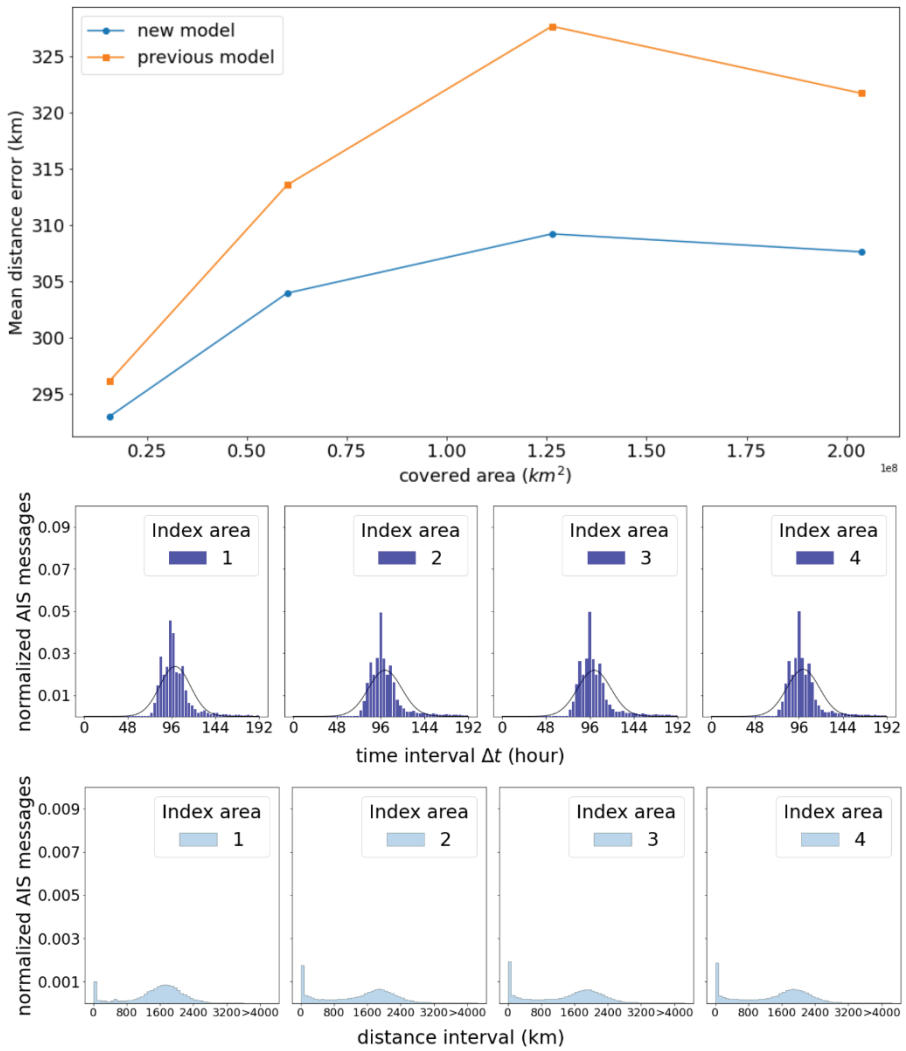


Figure 5-8. Comparison of the DL model performance with the previous model and characteristics of each observed area dataset in 96-h average time interval

Table 5-5 presents the performance evaluation metrics of the DL and baseline models using the test set. At this longest time interval, based on the total average distance interval between two AIS points, the DL models still delivered good performance, especially compared to the complete failure of the non-ML baseline. However, this result may not be reasonably acceptable since the performance degraded steadily relative to the time interval magnification.

Figure 5-8 demonstrates the comparison of DL models' performance and the dataset characteristics of each area. The model generated a steady improvement from the previous model with *3.5%* more accuracy and *1.5%* more precision on average. Additionally, the accumulation of the prediction performance in the large areas appears to distance the total average performance of both models.

5.5. Second Experiment: Coverage Size Variations

5.5.1. Experimental Setup

In the previous experiment, the coverage area variation was different between time interval dimensions since the dimension naturally modulated the extent of vessel movement, which is directly related to the minimal coverage area to observe the objects completely. Thus, the results from each time interval dimension could not be compared. However, from the previous experiment, the geodesic calculation as a non-ML baseline failed when the time interval average was beyond 24 h, and the performance of the previous DL model as the baseline was getting worse in large coverage areas compared to the new larger networks. The DL model delivered a satisfactory performance compared to the baseline models and succeeded in responding to the task of an increasingly large time interval dimension and thus coverage area; thus, the larger networks generalized better. As the optimization was carried out with no explicit regularization such as dropout or batch normalization, this phenomenon is allegedly caused by the optimization algorithm that remains largely unexplained [62].

In the second experiment, we trained only the proven DL model on observation size variations based on the first experiment. We focused more on the coverage area, increasing 4 and 3.2 degrees on longitude and latitude, respectively. The coverage variation was from 90°E–110°E to 60°E–140°E longitude and from 4°S–12°N to 28°S–36°N latitude based on the first experiment where the fluctuation occurred before the models' performance settled down. The more detailed size variation was expected to capture the fluctuation of the model performance in more detail. The model was then trained on each of four different time interval Δt_t dimensions.

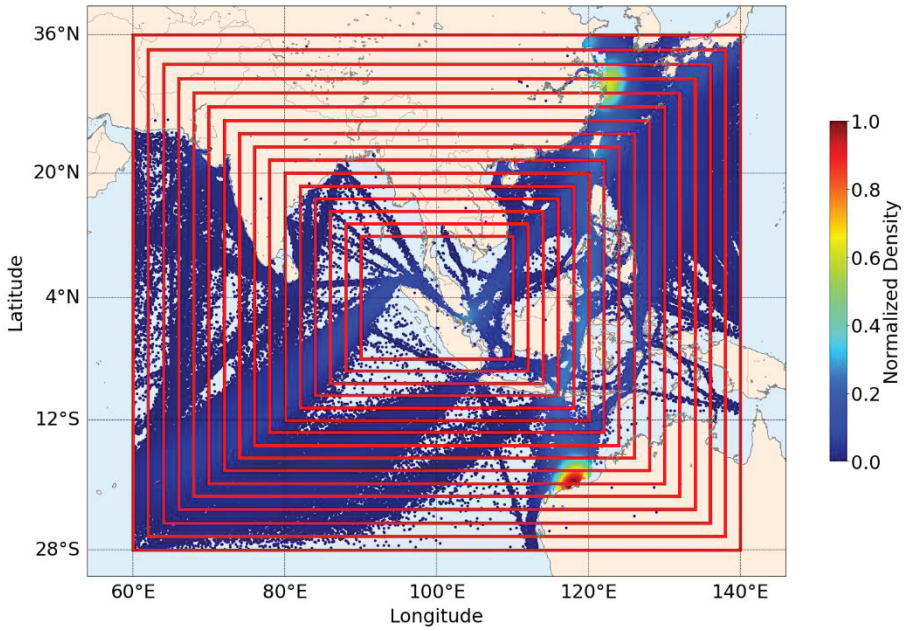


Figure 5-9. Size variations of the observed area for four-different time intervals (the normalized density distribution of AIS data is based on the largest area by the 24-h average time interval)

Figure 5-9 reveals the size variations of the observed area for four-different time interval dimensions. The normalized density distributions were calculated using Gaussian kernel density estimation (KDE) with the dataset at the largest observation area by the 24-h average time interval dimension involving 1693 different IMO numbers. Meanwhile, the smallest observation area at 96-h time interval dimension covered 1590 vessels. The most extensive coverage was similar to the largest observation area at a 48-h time interval in the second experiment. The Indian Ocean surrounding Northwest Australia, followed by the East China Sea surrounding China, South Korea, and Japan, became the densest zone eclipsing Malacca Strait.

Table 5-6 gives the overview of size variation of the observation area. The observation area covers from 4 to 60 million square kilometers. The dataset was split into training and dev-test sets with the same hold-out validation scheme as the first experiment. Separate training and dev-test sets were prepared due to the differences in time interval dimensions. The longest time interval suffers from the size limitation because the starting point at t and the target point at $t_t = 96\ h$ have to be in the coverage area, at which vessels traveled by more than 1300 km on average (see Table 5-5). Still, the size of these sets was sufficient where the smallest area at the largest time interval was already above the 25,000 thresholds, thus providing high confidence in the overall model performance.

Table 5-6. Size variation of observation coverage

Index area	Longitude	Latitude	Area size (10^6 km^2)	Training; dev/test size (10^3) based on average time intervals			
				24 h	48 h	72 h	96 h
1	90°E–110°E	4.0°S–12.0°N	4	107; 10	82; 7	61; 6	43; 4
2	88°E–112°E	5.6°S–13.6°N	6	132; 12	107; 9	84; 7	64; 6
3	86°E–114°E	7.2°S–15.2°N	8	157; 14	132; 12	108; 10	86; 7
4	84°E–116°E	8.8°S–16.8°N	10	185; 16	161; 14	137; 12	114; 10
5	82°E–118°E	10.4°S–18.4°N	13	235; 21	194; 17	169; 15	146; 13
6	80°E–120°E	12.0°S–20.0°N	16	367; 33	316; 28	271; 25	222; 20
7	78°E–122°E	13.6°S–21.6°N	19	516; 47	457; 41	399; 37	342; 32
8	76°E–124°E	15.2°S–23.2°N	22	585; 54	522; 48	465; 44	410; 39
9	74°E–126°E	16.8°S–24.8°N	26	663; 63	593; 56	527; 50	471; 46
10	72°E–128°E	18.4°S–26.4°N	30	761; 76	692; 69	626; 64	563; 58
11	70°E–130°E	20.0°S–28.0°N	35	832; 83	763; 77	700; 71	641; 67
12	68°E–132°E	21.6°S–29.6°N	39	1044; 101	997; 96	952; 92	908; 89
13	66°E–134°E	23.2°S–31.2°N	44	1147; 113	1098; 109	1053; 105	1013; 101
14	64°E–136°E	24.8°S–32.8°N	49	1223; 123	1176; 118	1131; 115	1091; 111
15	62°E–138°E	26.4°S–34.4°N	54	1278; 129	1230; 125	1184; 121	1142; 118
16	60°E–140°E	28.0°S–36.0°N	60	1393; 142	1352; 138	1314; 135	1280; 133

5.5.2. Results

Table 5-7 presents the performance evaluation metrics of the DL models across area size variations at four-different time intervals; results were computed using each test set by their respective trained models. On average, across all coverage size variations, the accuracy difference (based on MDE) between DL models with different long-time intervals was directly proportional to the difference in average distance interval. Each has a performance baseline that is almost equidistant from the others as the time intervals (thus the distance intervals) have uniform magnification. Meanwhile, models with shorter time intervals were just more precise (based on MaE) than longer-time interval models in general without any correlation.

Figure 5-10 gives an overview of the model performance across the coverage variations for each time interval dimension. All models generated comparable performances based on the time intervals, fluctuating from the initial coverage size until reaching around 20 million square kilometers. The fluctuation appears to be more intense for models with longer intervals. At this point, the performances improved from the initial coverage size and became stable. Then, a weak fluctuation was generated between this point and 30 million square kilometers of

coverage for models with the longest-time intervals (i.e., 72 h, 96 h). Eventually, all performances experienced a gradual increase on over 40 million square kilometers of coverage.

Table 5-7. Comparison of model performance in all area size and time interval variations

Index	***MDE*** **(km)**				***MaE*** **(deg)**				**Average distance interval (km)**			
area	Δt_t=1d	Δt_t=2d	Δt_t=3d	Δt_t=4d	Δt_t=1d	Δt_t=2d	Δt_t=3d	Δt_t=4d	Δt_t=1d	Δt_t=2d	Δt_t=3d	Δt_t=4d
1	73	149	211	247	18.2	17.3	19.2	18.6	377	686	912	1078
2	68	155	230	281	16.8	16.0	18.0	18.4	396	745	991	1192
3	65	154	243	307	14.7	15.6	17.7	17.9	411	783	1070	1300
4	60	145	240	327	15.7	16.3	17.7	19.2	411	797	1108	1356
5	55	135	235	332	16.2	18.0	19.1	20.9	395	788	1111	1371
6	49	116	199	293	13.2	14.3	15.8	18.3	424	842	1208	1484
7	45	105	184	271	14.7	15.4	16.8	17.5	422	839	1226	1558
8	46	113	193	284	15.7	16.6	18.0	18.7	419	835	1229	1581
9	46	113	197	287	15.4	16.7	17.6	18.6	420	841	1240	1604
10	46	111	187	276	13.8	15.3	16.6	16.8	432	867	1282	1656
11	45	111	188	278	14.1	14.7	15.6	17.4	430	870	1283	1658
12	46	110	185	276	19.6	18.9	19.8	20.9	391	783	1161	1521
13	47	113	191	285	23.3	22.4	22.0	23.6	373	747	1114	1459
14	49	116	197	290	25.5	24.8	25.3	24.8	364	727	1076	1416
15	50	119	201	295	25.7	25.1	26.2	25.8	361	721	1070	1403
16	50	123	206	303	28.4	27.0	27.4	26.6	347	697	1037	1364

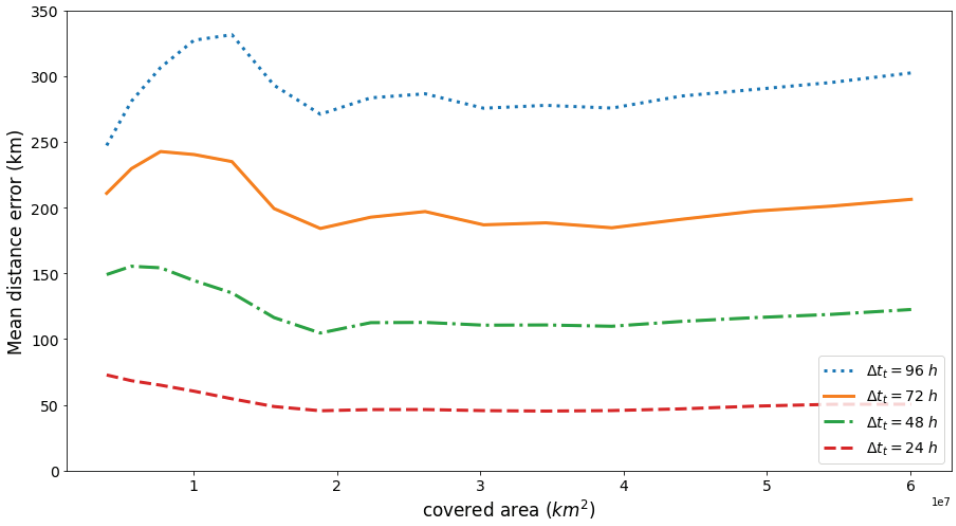


Figure 5-10. The DL model performance on area size variations in four-different time interval

5.5.3. Discussion: coverage size expansion

From the experiment result, the performances across long time intervals became stable after the coverage size reached 20 million square kilometers but increased gradually on 40 million square kilometers of coverage. After that range, the performance tendency leads to uncertainty about whether to remain in a gradual increase or become stable.

On that account, we trained the DL model on the expanded variations of the coverage size. The coverage area was increased in size, continuing from beyond 60 million to 204 million square kilometers as the widest coverage area in the first experiment (from 56°E–144°E to 20°E–180°E longitude and 31.2°S–39.2°N to 60°S–68°N latitude). As the performance was expected to be steady at this more extensive coverage, the size interval was doubled: expanded by 8 and 6.4 degrees on longitude and latitude, respectively.

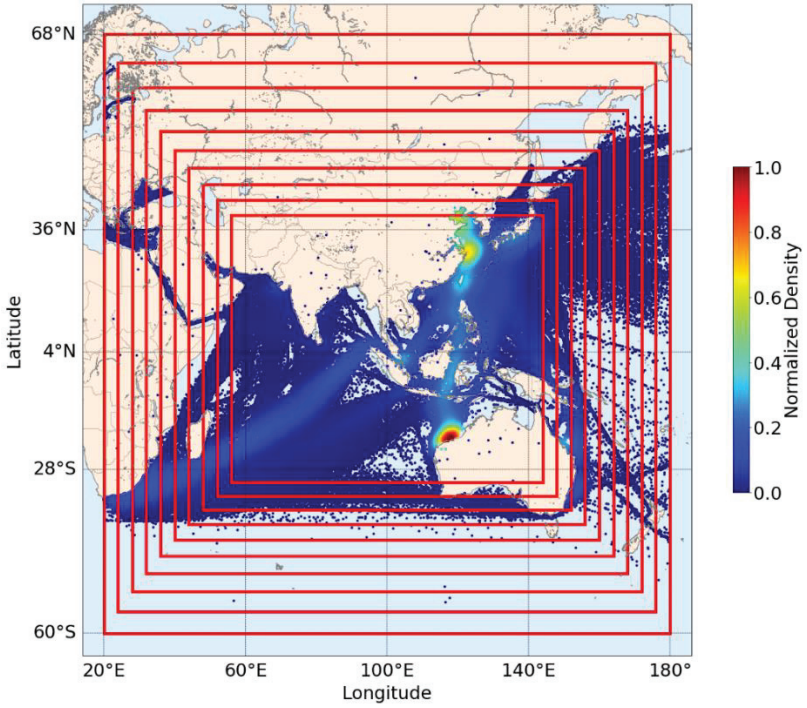


Figure 5-11. Expanded size variations of the coverage area with doubled size interval (the normalized density distribution of AIS data is based on the largest area)

Figure 5-11 portrays the expanded variations of the coverage area, whereby the largest observation area, almost all 1695 capsize bulk carriers, was captured over nine years and calculated for the density distribution. Identical to the most extensive coverage in the first experiment, almost all regions in the earth's eastern hemisphere were covered.

Figure 5-12 reveals the model performance across the expanded coverage variations for each time interval dimension. Overall, DL models across multiple time intervals deliver stable

performance and generate a proportional magnitude to the time intervals (predicted distance intervals). The performances remained steady from the initial to the end of the observed area. The observed performance stability across the size range may result from no high-density low-distance-interval region dominating beyond the initial coverage.

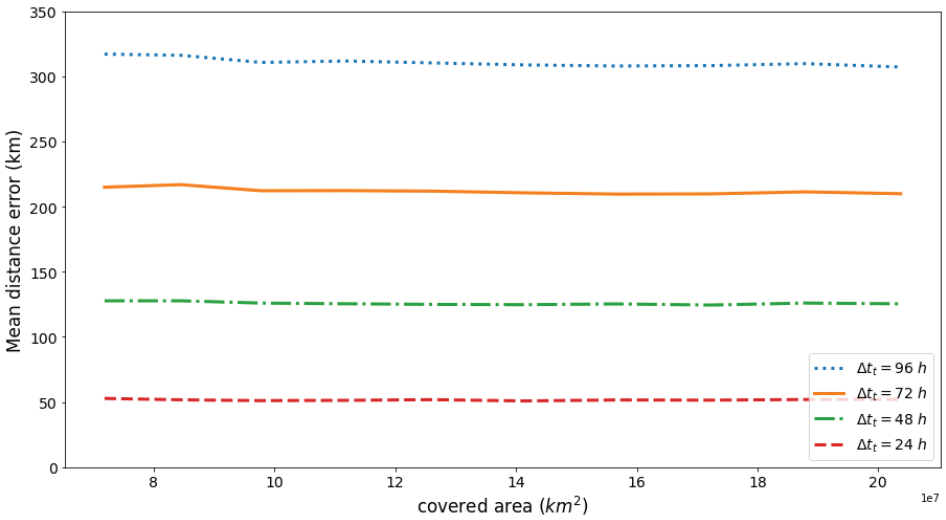


Figure 5-12. The DL model performance on the expanded coverage variations in four-different time interval

5.5.4. Discussion: the optimum coverage size for long-term position prediction

As the stability is different in each time-interval model, the observed performance patterns may not result from only enlarging the coverage area. The shortest-interval model produced a direct result, but models with time intervals longer than 24 h produced different patterns, although the amount of training data is roughly equivalent. The optimum coverage size needed to be answered whether it is influenced more by its size or possibly the dataset characteristics of the area.

Accordingly, we compare the dataset properties used to train the models across the time interval dimension and coverage size. The dispersion for continuous data can be measured by standard deviation, but comparing the variability when the coverage size and time intervals are different is tricky [63]. Therefore, we compare the number of training sets and AIS data characteristics on each observed area size and time interval represented by the dispersion of vessels distance interval (the haversine distance between two AIS messages). We calculated the coefficient of variation (CV), a measure of relative variability or dispersion of frequency distribution, to compare the distribution of ships' distance intervals across different time intervals as follows [64]:

$$CV_{d_h} = \frac{\sigma_{d_h}}{\overline{d_h}} \times 100 \quad (19)$$

where σ_{d_h} and $\overline{d_h}$ are the standard deviation and the mean of ships' distance interval between the base and true positions, respectively (see Equation 18).

Figure 5-13 compares the model performance and dataset characteristics across observed time intervals and covered area size variations. The dataset size may not directly influence the performance since the observed area is already large and contains sufficient data for the machine learning algorithm. The coverage size affects the data size until the coverage reaches the furthest of the northern and southern hemispheres. As the coverage size reaches 100 million square kilometers, AIS data is no longer significantly captured.

Based on the coefficient of variation, the dataset with a time interval of 24 h average has a similar pattern but different magnitude from other datasets with magnified time intervals. Its magnitude is relatively larger than the others across the coverage size, indicating that it has a greater variability to the ships' average distance interval despite having an almost equal amount of data at each coverage. This profound distinction may result from using the daily archived AIS messages with natural time intervals averaging 24-h. The dataset sampling in the time dimension satisfies the $\Delta t \to t_t$ interval condition as $\Delta t \in (0, 2t_t)$, destroying information regarding latent variables and causing distortion, indicated by the low dispersion compared to the raw AIS dataset with daily interval (i.e., averaging 24 h). Consequently, the transformed dataset became unbalanced or skewed, hindering the machine learning model's performance [65]. The unbalanced dataset may also be caused by a new formation of outliers that are not strictly defined when the time interval was magnified.

The model performance with 24 h time interval is already stable at 19 million square kilometers. Malacca Strait is the densest region in this coverage, followed by the Makassar strait. Meanwhile, the performance of magnified-interval models fluctuated until 39 million square kilometers of coverage. In this coverage, the Northwest Australian waters as the densest region are completely captured, at which the CV soared to around 60 percent. The sharp increase of CV corresponds to capturing the densest region, indicating that dataset characteristics influenced these model performances more than the observed area size. However, after this point, the model performances increase gradually and become stable when the coverage reaches around 72 million square kilometers, corresponding to dispersion rate stability. This coverage completely captured the waters surrounding China, the second densest region.

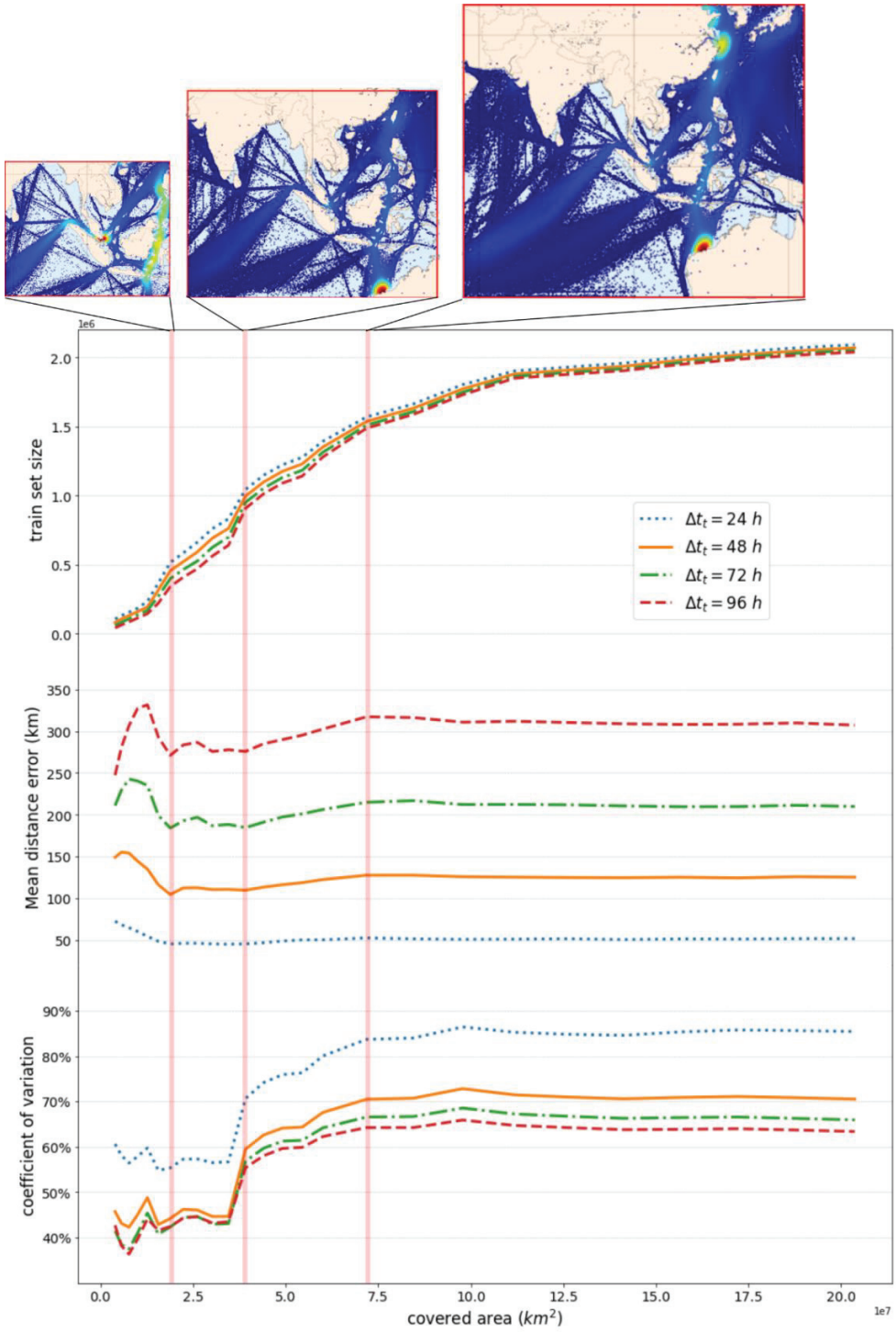


Figure 5-13. DL model performance and dataset characteristics comparison across different time intervals and coverage size variation

Therefore, according to the observed location, 19 million kilometers of coverage is optimal for the prediction model with 24-h intervals but not for the other DL models with magnified-time intervals. Meanwhile, the optimum coverage size for all models is 39 million square kilometers, corresponding to the dataset's high dispersion rate. The subsequent gradual increase

observed from the longest-time interval models appears insignificant, resulting from capturing the second densest region degrees by degrees. Thus, by capturing the densest region on the observed location, which leads to a proportional ratio of the low-distance to the high-distance intervals, optimum coverage for long-term position prediction can be achieved. Enlarging the coverage area size may eventually solve the prerequisite of partial coverage.

5.6. Third Experiment: Comparison of Observation Area Worldwide

5.6.1. Experimental Setup

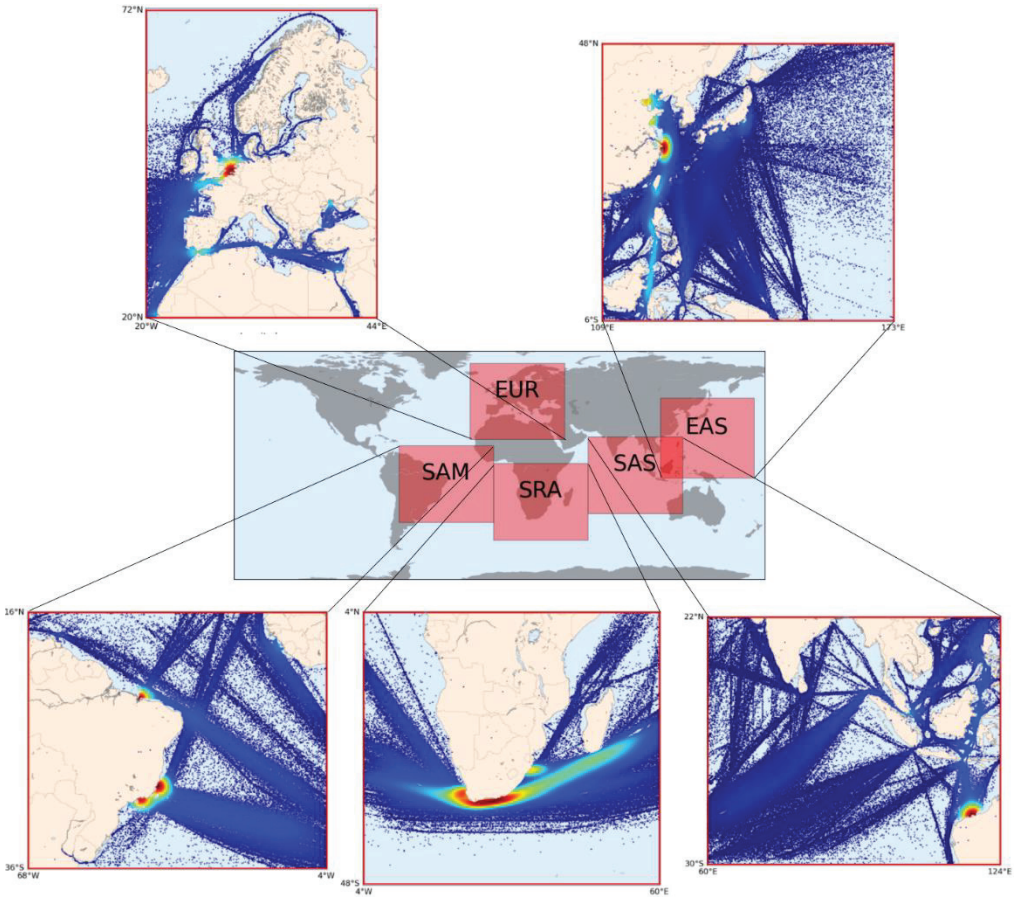
Predictive performance is directly proportional to the time interval dimension based on performance evaluation metrics. Meanwhile, the observation area size is not only the main determinant in optimizing prediction performance but also a dataset characteristic shaped based on its location. To further assess the influence of the area, in the third experiment, five DL models were built for five major locations of maritime shipping routes focused on Capesize bulk carriers. The same procedure was applied to the models as in the previous experiments: each used the same model properties and followed a similar hold-out validation split.

The observed locations shared the same longitude and latitude length based on the optimum coverage size from the previous experiment: within 64-interval longitude and 52-interval latitude. Each area's size varies despite having the same geographical boundaries since the geographic coordinate measurements are angles and are not on a planar surface [66]. Each was named after the crossed region of the continent due to its large size, namely, Europe (EUR), South America (SAM), Southern Region of Africa (SRA), South and Southeast Asia (SAS), and East Asia (EAS).

Figure 5-14 illustrates the location and size of each area (red rectangles) based on the equidistant cylindrical projection, which neither covers equal-area nor conformal, a standard for imaging the relationship between the observed location on the map and its corresponding geographic location on Earth [67]. Each observed location is further depicted based on the Mercator projection, which inflates the area size away from the equator [68][8]. The normalized

[8] Despite large distortion at high latitude, it is commonly used in nautical and aerial use and is practical for visual mapping the Earth’s globe (without polar areas).

density distribution on each observed location reveals the complete capture of the densest region, not partial, complying with the recommendations for the optimum coverage of long-term position prediction.



Europe (EUR), South America (SAM), Southern Region of Africa (SRA),
South and Southeast Asia (SAS), East Asia (EAS)

Figure 5-14. Location of the observed area

Capesize bulkcarriers are one-way cargo transport, where round trips between the countries are a common pattern. China, Australia, and Brazil are the top players with Capesize ships' iron ore shipments [69]. From the normalized density distribution, the densest region in East Asia is located between Shanghai Port and Port of Ningbo in China; combined, they are the world's biggest importers of iron ore and bulk shipments [70]. Meanwhile, the densest region in South and South East Asia is Northwest Australia, where the nearest Port Hedland is the world's largest bulk export port [71]. In South America, the densest region is dry bulk cargo ports around Ponta da Madeira in North Brazil and Sepetiba and Tubarao in South Brazil, exporting mainly iron ore [72]. In Europe, the densest region is around the port of Rotterdam in the Netherlands, an important transit point for transport bulk between the European continent

and other parts of the world [73]. Further, the densest region in the Southern Region of Africa stretches around the Cape of Good Hope as ships sail between the Pacific and Atlantic oceans, plus additional traffic of iron ore exports around Port of Saldanha on the south-western coast of South Africa [74].

5.6.2. Results and Discussions

The performance evaluation metrics of the models for each examined area are shown in Figure 5-15. Based on the mean angle error (MaE), model performance in the Southern Region of Africa (SRA) generated the most precision across time intervals, followed by models in South and Southeast Asia and South America. Their precisions gradually fade with the time-interval magnification scheme. Meanwhile, model performance in Europe followed by East Asia produced the worse precision, to which magnifying time intervals barely affected them.

All models produced a similar performance pattern based on the mean distance error, where the accuracy was inversely proportional to the time intervals. At 24-h average intervals, their magnitude appeared not substantially different from each other. As the time interval was magnified, the model's distance error in Europe and East Asia closely soared to their highest levels of more than 300 km at 96-h average intervals, much higher than the models in other observed locations. Meanwhile, models on South America, the Southern Region of Africa, and South and Southeast Asia generated a more accurate result and similar trends, with a gradual increase over the time-interval magnification.

As ships' behavior in each observed location has different observable properties, the error ratio demonstrates the generated error (MDE) relative to the average movement (i.e., distance interval) of the predicted objects per location and time-interval scheme, calculated as follows:

$$error\ ratio = \frac{MDE}{\overline{d_h}} \times 100 = \sum_{i=1}^{m} \frac{d_{h_{(v,w)}}^{\ i}}{d_{h_{(u,v)}}^{\ i}} \times 100 \tag{20}$$

where u is the previous or base point (φ_t, λ_t), and v and w are the true $\left(\varphi_{t+t_t}, \lambda_{t+t_t}\right)$ and predicted positions $\left(\hat{\varphi}_{t+t_t}, \hat{\lambda}_{t+t_t}\right)$ at designated time intervals.

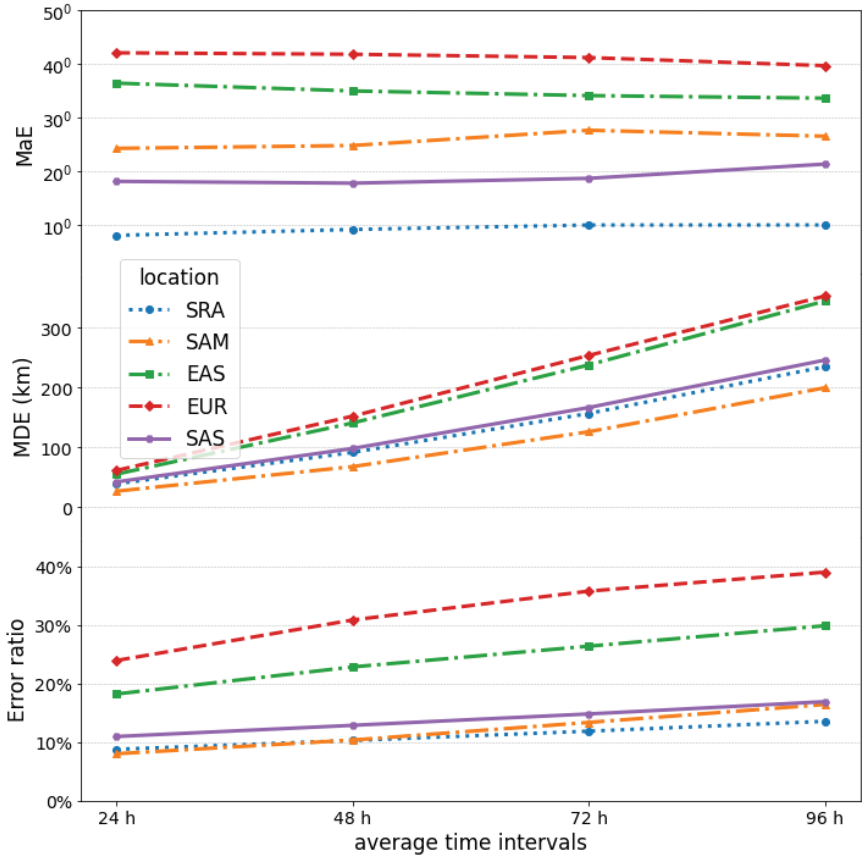


Figure 5-15. Comparison of DL model performance across observed location and time intervals

Based on the error ratio, model performances in Europe and East Asia produced poor results across the time interval magnification compared to the model performances in other observed locations. By 24-h and 96-h average intervals, their error ratio worsened significantly from around 24% and 18% to 30% and 39%, respectively, double the average of the other model performances at the same time intervals. Although their mean distance error is not statistically significant to the other models, especially at low magnification time intervals, their error ratio appears to have a large gap. At 24-h and 96-h average intervals, the error ratio of South America, the Southern Region of Africa, and South and Southeast Asia averaged 9% and 15%, respectively. For the case like the Southern Region of Africa, where the dominant AIS characteristics formed from the common sailing movement of bulk fleet stretching across the Cape of Good Hope, an ideal condition for a position prediction model, the DL model generated the lowest error ratio with the smallest increase over the longest time intervals.

Therefore, the deep learning models generated worse predictions in Europe and East Asia than in the other observed locations (i.e., South America, the Southern Region of Africa, and South and Southeast Asia). The locations where importers of bulk shipments constitute the

largest portion appears to be harder to predict than those that constitute bulk export regions, which corresponds with the fact that Capesize dry bulk carriers' unloading time takes longer with a wider approximate range than loading time. Based on IHS MARKIT/TheTradeNet Market Intelligence Network (MINT) data from Muzhoffar, D.A.F., et al. (2022), the average port-staying time of worldwide Capesize bulkcarriers when loading is around 3 days ± 2 days, and when unloading is around 5 days ± 3 days [75]. The unloading operations require more complex procedures in cargo handling treatment than loading operations, which may lead to more lengthy delays or other uncertainties [76]. More importantly, the divergent state through unloading conditions is considered unsuitable for long-term position predictions of ships as the processed input is only their last movement and geographical position.

5.7. Discussions on Single-Worldwide Model and Limitations of the Prediction Models

5.7.1. Single-Worldwide DL Model

The more accurate performance of the larger networks than the previous DL model can mitigate the larger observation area, thus a bigger and wider range of data. The larger model also performs better as the time interval was magnified longer than 24 h average, a more complex problem than the 24-h interval model. To further assess performances from a much larger model, a single worldwide model, we created a DL model trained from the whole worldwide data, having the same model properties and hold-out validation split. Model performances from the third experiment provide a baseline for this process.

Figure 5-16 reveals performances from the single-worldwide model across the magnified time intervals. Compared to the location-based models in the third experiment, the worldwide model generated poor performance based on the MDE evaluation metric, similar to the inferior performance of the models with import-dominated regions of bulk shipments (i.e., Europe, East Asia). However, based on the error ratio, the worldwide model demonstrated output equivalent to the average between the poor performance models of the import-dominated location and strong performance models of the export-dominated locations, indicating that the worldwide model poses more generalized networks. The bigger and wider range of training data of the worldwide model is a way to generalize position prediction from the combinations of specific ships' behavior that may be only associated with specific locations.

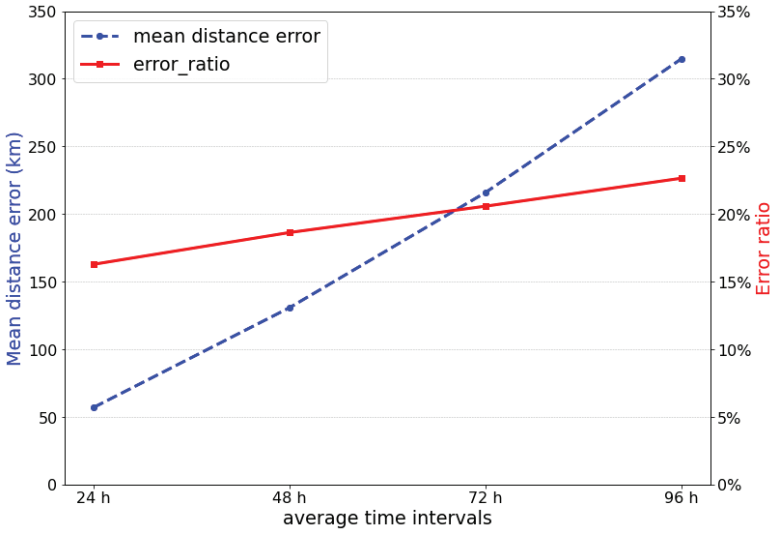


Figure 5-16. Single-worldwide model performance across time intervals: mean distance error (left y-axis) and error ratio (right y-axis)

The worldwide model may be suitable for predicting general responses; however, it will limit accuracy in certain circumstances or sites previously covered by the less-generalized location-based models that perform better on the designated observation area [77]. Moreover, this model contains extensive data that can naturally lead to more spurious correlations being fit undetected across many spots worldwide; consequently, more input features are required to generate optimal results [78]. Finally, the networks can be either overparameterized (e.g., large networks trained on small data tasks) or underparmeterized, which are too small for the enormous and complex world-scale model [79].

5.7.2. Limitations: Outliers and AIS Data Uncertainty

Despite a thorough understanding of the data and a deep learning ability that is robust to outliers, outliers identification and appropriate analytical technique are still hard to be carried out in preprocessing operations [80]. As our observation covers a relatively larger area and time range of AIS data with no restrictions on shipping trajectory than other position prediction studies, defining all outliers would be tremendously difficult had we examined everything at the micro level. Point or global outliers naturally inhere within AIS data are one thing, but detecting contextual or conditional outliers related to geographical position requires a monumental work that needs to specify every existing geographical boundary (e.g., polygon) and practicable definition of the geographical position [81, 82]. This preprocessing, nonetheless, is not essential in the motion-based method of vessel position prediction studies, where generalization plays an important role. Additionally, the AIS data used does not require

a preprocessing of anomalous motion patterns or trajectory outliers found naturally in the low time-interval data sets [83].

Global outliers from the functions of movement and geographical position explicitly from AIS data were already omitted earlier in the database handling process, along with the absence and abnormality of AIS message components. Following the preprocessing, to rationalize on a huge scale, we further considered the outliers based on the time-interval functions and all their derivatives, which may affect prediction performances in particular scenarios. Considering that a formation of outliers possibly contributed to data sets when the time interval was magnified due to the nature of the AIS data used. This problem intensifies for the worldwide model as the CV dropped significantly compared to the location-based models when the time interval was magnified (see Figure 5-17 (a)). In this extensive coverage, the dispersion rate across time intervals is affected mainly by data set characteristics, not their size, since the number of AIS data has no correlation to the magnitude of CV on the respected location (see Figure 5-17 (b)).

Conditional outliers related to ships' navigational status also play notable roles which may impact the data set's balance. Their identification analysis would reveal the balance of the data set; ML experts can then choose the most suitable set to be trained so that the prediction results will be more sensitive and accurate, but only in the specified situation [84]. Instead, we generalized the deep learning model since we focused only on the last geographical movements as the input, and it is interesting to incorporate the statuses in future studies.

Moreover, other limitations inherent to the nature of AIS data may add certain deviations to the data set characteristics, affecting the deep learning performance. The development of AIS technology and mandatory installation regulation took years to fully implement, leading to differences in data volume and signal coverage across the observed years (see Figure 3-1) [85]. Some loss reported where it is difficult for satellites to capture all messages emitted, attributed to the congestion in some spots, such as in China [86, 87]. Nevertheless, these concerns are trivial as the observation covers a more extensive zone and time coverage.

The archived and compressed AIS messages with random sampling intervals also inheres major limitations to the study [88]. Despite being compressed within a one-day interval, the random sampling renders the data set distribution; having a clear Gaussian distribution will benefit machine learning building [89]. However, the issue has been minimalized as the time-interval targets are not strictly defined. The DL model is also improved by replacing sigmoid functions that work most naturally with normally distributed data with selu activation [90]. For the same time windows within 24-h intervals, the deep learning generated remarkable good

performance; however, at the time-intervals magnification, it is possible that the issue intensifies, especially for the single-worldwide scenario.

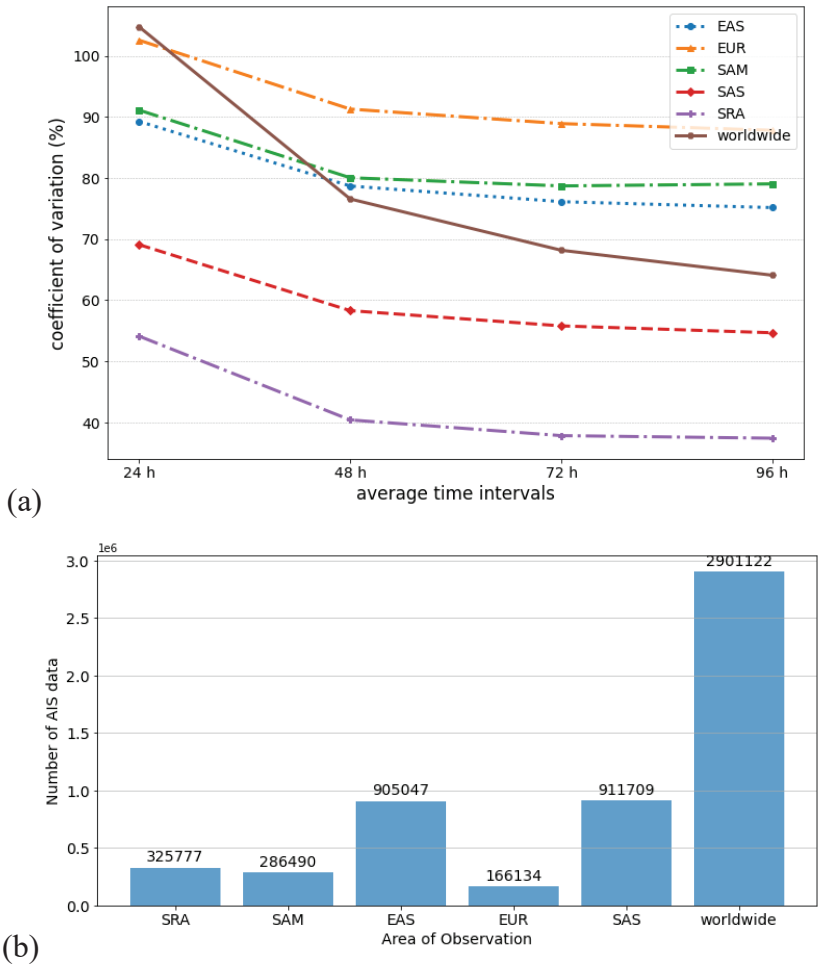


Figure 5-17. Dataset characteristics comparison of the location-based and worldwide models: coefficient of variation across time interval magnification (a) and number of training data (b)

5.8. Conclusions on several-day intervals of ship position prediction

Results from this research show that the geodesic calculation as a non-ML baseline failed when the time interval average was magnified beyond 24 h, and the performance of the previous DL model as the baseline became worse in large coverage areas compared to the new larger networks. The vector regression model is better than the scalar regression model; thus, the larger networks generalized better. The DL model delivered a satisfactory performance compared to the baseline models and succeeded in responding to the task of an increasingly large time interval dimension and coverage area.

Meanwhile, the optimum coverage size for all time-interval models is 39 million square kilometers, corresponding to the dataset's high dispersion rate. By capturing the densest region

on the observed location, which leads to a proportional ratio of the low-distance to the high-distance intervals, optimum coverage for long-term position prediction can be achieved. The deep learning models generated worse predictions in Europe and East Asia than in the other observed locations (i.e., South America, the Southern Region of Africa, and South and Southeast Asia). The locations where importers of bulk shipments constitute the largest portion appears to be harder to predict than those that constitute bulk export regions, which corresponds with the fact that Capesize dry bulk carriers' unloading time takes longer with a wider approximate range than loading time.

Moreover, the worldwide model is a way to generalize position prediction and may be suitable for predicting general responses, but it will limit accuracy in certain circumstances or areas where location-based models perform better on them. The deep learning model was trained based on the last geographical movements; thus, this limitation intensifies at the time-intervals magnification.

Chapter 6
Conclusions and Future Works

6.1. Conclusions

This study employed DL models with experimentally defined properties using exactEarth AIS daily data. There are two main chapters: each consist of three experiments and discussions. In the first experiment of the former, a model for the Indian Ocean area was examined, and in the subsequent experiment, 12 models were investigated on open oceans and maritime chokepoints. Then, a selected sample location within the Malacca Strait area was examined, resembling a simulation of practical application. In the first experiment of the latter, larger networks of DL models were trained on different observation area sizes with magnified time intervals. We focused more on observation size variations in the next experiment, increasing 4 and 3.2 degrees on longitude and latitude, respectively. Finally, five DL models were built for five major locations of maritime shipping routes focused on Capesize bulk carriers.

The conclusions of this dissertation are summarized as follows:

1. Ship position prediction with an average time interval of 24 h was possible with the straightforward motion-based method, more accurate than the geodesic calculation as the baseline model in all areas.
2. Predictions in the open ocean areas yielded higher accuracy than in the chokepoint areas; the improvement compared to the geodesic calculation was higher in the chokepoint areas than in the ocean areas as the conventional calculation failed to predict vessel behavior near the ports and congested waters.
3. The DL model can predict the complex movement of ships near ports and congested routes with no information regarding vessel status, historical trajectory, or destination. The DL model appears to have a sense of the dimension of the geographic coordinate system that can be or is often passed, wherein the chokepoint areas rely more on the input features of the current latitude and longitude of the vessels.
4. The proposed MLP model generated predictions as accurate as of the LSTM with faster training times on the dataset with varying and uneven time-interval distribution and without a trajectory reconstruction.
5. From the research of several-day intervals of ship position prediction, the DL model delivered a satisfactory performance compared to the baseline models in increasingly

large time interval dimensions and coverage size; thus, the larger networks generalized better. The geodesic calculation as a non-ML baseline failed when the time interval average was beyond 24 h, and the performance of the previous DL model as the baseline became worse in large coverage areas.

6. The optimum coverage size for all time-interval models is 39 million square kilometers, corresponding to the dataset's high dispersion rate. To achieve the optimum coverage for long-term position prediction, the densest region on the observed location must be captured, leading to a proportional ratio of the low-distance to the high-distance intervals.
7. The locations where importers of bulk shipments constitute the largest portion appears to be harder to predict than those that constitute bulk export regions, which corresponds with the fact that Capesize dry bulk carriers' unloading time takes longer with a wider approximate range than loading time.
8. The worldwide model is a way to generalize position prediction and may be suitable for predicting general responses, but it will limit accuracy in certain circumstances or areas where location-based models perform better. The deep learning model was trained based on the last geographical movements; thus, this limitation intensifies at the time-intervals magnification.

6.2. Future Works

This paper is the first study to explore long-term ship position prediction using an AIS dataset in nine years for Capesize bulk carriers worldwide, providing a valuable benchmark for future studies. As this is the first study on long-term position prediction, we focused only on the basics of AIS data as the input features, that is, the last geographical position and movement. Therefore, it is very possible to incorporate other valuable input features and utilize other techniques in future studies to improve the prediction performance, such as:

- Adding new essential input features regarding the vessel status, destination, or longer historical trajectory.
- Adding feature-engineered input processed from AIS data and other data (e.g., Port data, Country data, etc.).
- Adding input from other data (e.g., Weather data, operation data, etc.).

- Exploring new thrived architecture specialized for the objective (sequence-based time-series problems): The neural ordinary differential equation model and Transformer Neural Networks for regression.

In future studies, we intend to improve the model performance by utilizing these techniques and incorporating other types of vessels, thereby increasing the complexity of the model.

References

1. Rodrigue, J.-P., C. Comtois, and B. Slack, *The geography of transport systems*. 2016: Routledge.
2. Taylor, I. and K. Smith, *United Nations Conference on Trade and Development (UNCTAD)*. 2007: Routledge.
3. Pallotta, G., M. Vespe, and K. Bryan, *Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction.* Entropy, 2013. **15**(6).
4. Statheros, T., G. Howells, and K.M. Maier, *Autonomous Ship Collision Avoidance Navigation Concepts, Technologies and Techniques.* Journal of Navigation, 2008. **61**(1): p. 129-142.
5. Murray, B. and L.P. Perera, *A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data.* Ocean Engineering, 2020. **209**: p. 107478.
6. Sheng, P. and J. Yin, *Extracting Shipping Route Patterns by Trajectory Clustering Model Based on Automatic Identification System Data.* Sustainability, 2018. **10**(7): p. 2327.
7. Robards, M., et al., *Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)-A review.* Bulletin of Marine Science -Miami-, 2016. **92**: p. 75-103.
8. Perera, L.P., P. Oliveira, and C.G. Soares, *Maritime Traffic Monitoring Based on Vessel Detection, Tracking, State Estimation, and Trajectory Prediction.* IEEE Transactions on Intelligent Transportation Systems, 2012. **13**(3): p. 1188-1200.
9. Czapiewska, A. and J. Sadowski, *Algorithms for Ship Movement Prediction for Location Data Compression.* TransNav: International Journal on Marine Navigation and Safety of Sea Transportation, 2015. **9**(1).
10. Bakshi, G., G. Panayotov, and G. Skoulakis. *The Baltic Dry Index as a predictor of global stock returns, commodity returns, and global economic activity*. in *AFA 2012 Chicago Meetings Paper*. 2011.
11. Chen, X. and X. Lin, *Big Data Deep Learning: Challenges and Perspectives.* IEEE Access, 2014. **2**: p. 514-525.
12. Liu, J., G. Shi, and K. Zhu, *Vessel Trajectory Prediction Model Based on AIS Sensor Data and Adaptive Chaos Differential Evolution Support Vector Regression (ACDE-SVR).* Applied Sciences, 2019. **9**(15).

13. Qi, L. and Z. Zheng, *Trajectory prediction of vessels based on data mining and machine learning.* J. Digit. Inf. Manage, 2016. **14**(1): p. 33-40.
14. Arguedas, V.F., G. Pallotta, and M. Vespe, *Maritime traffic networks: From historical positioning data to unsupervised maritime traffic monitoring.* IEEE Transactions on Intelligent Transportation Systems, 2017. **19**(3): p. 722-732.
15. Nguyen, D., et al. *A multi-task deep learning architecture for maritime surveillance using AIS data streams*. in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. 2018. IEEE.
16. Dobrkovic, A., M.-E. Iacob, and J. van Hillegersberg, *Maritime pattern extraction and route reconstruction from incomplete AIS data.* International journal of Data science and Analytics, 2018. **5**(2): p. 111-136.
17. Dalsnes, B.R., et al. *The Neighbor Course Distribution Method with Gaussian Mixture Models for AIS-Based Vessel Trajectory Prediction*. in *2018 21st International Conference on Information Fusion (FUSION)*. 2018.
18. Pallotta, G., et al. *Context-enhanced vessel prediction based on Ornstein-Uhlenbeck processes using historical AIS traffic patterns: Real-world experimental results*. in *17th International Conference on Information Fusion (FUSION)*. 2014.
19. Tu, E., et al., *Modeling Historical AIS Data For Vessel Path Prediction: A Comprehensive Treatment.* arXiv e-prints, 2020: p. arXiv:2001.01592.
20. Mazzarella, F., V.F. Arguedas, and M. Vespe. *Knowledge-based vessel position prediction using historical AIS data*. in *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. 2015.
21. Virjonen, P., et al. *Ship Movement Prediction Using k-NN Method*. in *2018 Baltic Geodetic Congress (BGC Geomatics)*. 2018.
22. Chen, X., et al., *A ship movement classification based on Automatic Identification System (AIS) data using Convolutional Neural Network.* Ocean Engineering, 2020. **218**: p. 108182.
23. Duca, A.L., C. Bacciu, and A. Marchetti. *A K-nearest neighbor classifier for ship route prediction*. in *OCEANS 2017 - Aberdeen*. 2017.
24. Kim, K.-I. and K.M. Lee, *Deep Learning-Based Caution Area Traffic Prediction with Automatic Identification System Sensor Data.* Sensors, 2018. **18**(9).
25. Zhou, X., et al., *Using Deep Learning to Forecast Maritime Vessel Flows.* Sensors, 2020. **20**(6).

26. Palacios, R., A. Doshi, and A. Gupta, *Computing aircraft position prediction.* The Open Transportation Journal, 2008. **2**(1).
27. Juraszek, K., et al. *Extended Kalman Filter for Large Scale Vessels Trajectory Tracking in Distributed Stream Processing Systems*. 2020. Cham: Springer International Publishing.
28. Mao, S., et al., *An automatic identification system (AIS) database for maritime trajectory prediction and data mining*, in *Proceedings of ELM-2016*. 2018, Springer. p. 241-257.
29. Zhou, H., Y. Chen, and S. Zhang, *Ship Trajectory Prediction Based on BP Neural Network.* Journal of Artificial Intelligence, 2019. **1**(1): p. 29-36.
30. Simsir, U. and S. Ertugrul, *Prediction of manually controlled vessels' position and course navigating in narrow waterways using Artificial Neural Networks.* Applied Soft Computing, 2009. **9**(4): p. 1217-1224.
31. Zissis, D., E.K. Xidias, and D. Lekkas, *Real-time vessel behavior prediction.* Evolving Systems, 2016. **7**(1): p. 29-40.
32. Borkowski, P., *The Ship Movement Trajectory Prediction Algorithm Using Navigational Data Fusion.* Sensors, 2017. **17**(6).
33. Gao, M., G. Shi, and S. Li, *Online Prediction of Ship Behavior with Automatic Identification System Sensor Data Using Bidirectional Long Short-Term Memory Recurrent Neural Network.* Sensors, 2018. **18**(12).
34. Gao, D.-w., et al., *A novel MP-LSTM method for ship trajectory prediction based on AIS data.* Ocean Engineering, 2021. **228**: p. 108956.
35. Zhang, C., et al., *AIS data driven general vessel destination prediction: A random forest based approach.* Transportation Research Part C: Emerging Technologies, 2020. **118**: p. 102729.
36. Rubanova, Y.a.C.R.T.Q.a.D.D.K., *Latent Ordinary Differential Equations for Irregularly-Sampled Time Series*, H.W.a.H.L.a.A.B.a.F.d.t.A.-B.a.E.F.a.R. Garnett, Editor. 2019.
37. Bonham, C., et al., *Analysing port and shipping operations using big data.* Data Science Campus, ONS, 2018.
38. Shickel, B. and P. Rashidi, *Sequential Interpretability: Methods, Applications, and Future Direction for Understanding Deep Learning Models in the Context of Sequential Data.* arXiv preprint arXiv:2004.12524, 2020.
39. ExactEarth. *exactAIS Archive™.* [cited 2020 20 March].

40. Gl, D.N.V., *MARITIME FORECAST TO 2050—Energy Transition Outlook 2019*. 2018, Technical report. Available online: DNVGL. com (accessed on 10 October 2019).
41. IHSMarkit. *Sea-web Ships*. [cited 2020 13 April].
42. Vieira, S., W.H.L. Pinaya, and A. Mechelli, *Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications.* Neuroscience & Biobehavioral Reviews, 2017. **74**: p. 58-75.
43. Chollet, F., *Keras*. 2015, GitHub. p. https://github.com/fchollet/keras.
44. Fisher, A., C. Rudin, and F. Dominici, *All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously.* Journal of Machine Learning Research, 2019. **20**(177): p. 1-81.
45. Molnar, C., *Interpretable machine learning*. 2020: Lulu. com.
46. Robusto, C.C., *The cosine-haversine formula.* The American Mathematical Monthly, 1957. **64**(1): p. 38-40.
47. Tumer, K. and J. Ghosh. *Estimating the Bayes error rate through classifier combining*. in *Proceedings of 13th International Conference on Pattern Recognition*. 1996.
48. Tseng, W.-K. and H.-S. Lee, *The Vector Function for Distance Travelled in Great Circle Navigation.* Journal of Navigation, 2007. **60**(1): p. 158-164.
49. Veness, C. *Calculate distance, bearing and more between Latitude/Longitude points*. 2019 [cited 2020 13th December].
50. Botev, Z.I., J.F. Grotowski, and D.P. Kroese, *Kernel density estimation via diffusion.* Ann. Statist., 2010. **38**(5): p. 2916-2957.
51. Whitaker, J., *Basemap*. 2011, Github. p. https://github.com/matplotlib/basemap.
52. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python.* the Journal of machine Learning research, 2011. **12**: p. 2825-2830.
53. Chen, T. and C. Guestrin. *Xgboost: A scalable tree boosting system*. in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.
54. Rodrigue, J.-P. *Maritime transportation: drivers for the shipping and port industries*. in *International Transport Forum*. 2010.
55. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-444.
56. Wan, R., et al. *Towards Making Deep Transfer Learning Never Hurt*. in *2019 IEEE International Conference on Data Mining (ICDM)*. 2019.

57. Gers, F.A., D. Eck, and J. Schmidhuber, *Applying LSTM to time series predictable through time-window approaches*, in *Neural Nets WIRN Vietri-01*. 2002, Springer. p. 193-200.
58. Ibadurrahman, et al., *Long-Term Ship Position Prediction Using Automatic Identification System (AIS) Data and End-to-End Deep Learning.* Sensors, 2021. **21**(21): p. 7169.
59. Rodrigue, J.-P., *The geography of transport systems*. 2020: Routledge.
60. No, I., *30: Modern ship size definitions.* Lloyd's Register Foundation, 2014.
61. *Western Australia Iron Ore Industry Profile*. 2016, Government of Western Australia Department of State Development.
62. Neyshabur, B., et al., *Exploring generalization in deep learning.* Advances in neural information processing systems, 2017. **30**.
63. Zwillinger, D., *CRC standard mathematical tables and formulas*. 2018: chapman and hall/CRC.
64. Watters, P. and S. Boslaugh, *Statistics in a nutshell*. 2008: O'Reilly Media, Incorporated.
65. Mani, I. and I. Zhang. *kNN approach to unbalanced data distributions: a case study involving information extraction*. in *Proceedings of workshop on learning from imbalanced datasets*. 2003. ICML.
66. Chang, K.-T., *Introduction to geographic information systems*. Vol. 4. 2008: McGraw-Hill Boston.
67. Eysn, L., et al., *A Practical Approach for Extracting Tree Models in Forest Environments Based on Equirectangular Projections of Terrestrial Laser Scans.* Remote Sensing, 2013. **5**(11): p. 5424-5448.
68. Smetanová, D., et al., *Mercator's Projection–a Breakthrough in Maritime Navigation.* NAŠE MORE: znanstveni časopis za more i pomorstvo, 2016. **63**(3 Special Issue): p. 182-184.
69. Association, W.S., *World steel in figures 2019.* World Steel Association, 2018.
70. (SISI), S.I.S.I., *Global Port Development Report (2019)*. 2020, Shanghai Pujiang Education Press Shanghai: http://sisi.gstta.org/index.php?c=article&id=18943.
71. Curtis, B., *Channel optimisation and risk management through technology at the world's largest bulk export port.* Australasian Coasts & Ports, 2017: p. 310-315.
72. (AAPA), A.A.o.P.A., *Port Statistics Database*. 2017: http://www.aapa-ports.org/unifying/content.aspx?ItemNumber=21048.

73. Shi, H. and Y. Xing, *The development of Rotterdam port enlightenment to the ports of Jiangsu.* Journal of Coastal Research, 2015(73 (10073)): p. 116-118.
74. International, D.C., *South Africa Regional Report 2012*. 2012: https://www.drycargomag.com/south-africa-regional-report-2012.
75. Muzhoffar, D.A.F., et al., *Basic Ship-Planning Support System Using Big Data in Maritime Logistics for Simulating Demand Generation.* Journal of Marine Science and Engineering, 2022. **10**(2): p. 186.
76. Beghin, D., *Safety of bulk carriers-prime concern of the maritime community.* WIT Transactions on The Built Environment, 1970. **1**.
77. Kawaguchi, K., L.P. Kaelbling, and Y. Bengio, *Generalization in deep learning.* arXiv preprint arXiv:1710.05468, 2017.
78. Arjovsky, M., et al., *Invariant risk minimization.* arXiv preprint arXiv:1907.02893, 2019.
79. Nakkiran, P., B. Neyshabur, and H. Sedghi, *The deep bootstrap framework: Good online learners are good offline generalizers.* arXiv preprint arXiv:2010.08127, 2020.
80. Kuhn, M. and K. Johnson, *Applied predictive modeling*. Vol. 26. 2013: Springer.
81. Zhaojin Yan and Liang Cheng and Rong He and Hui, Y., *Extracting ship stopping information from AIS data.* Ocean Engineering, 2022. **250**: p. 111004.
82. Cheng, L., et al., *Using big data to track marine oil transportation along the 21st-century Maritime Silk Road.* Science China Technological Sciences, 2019. **62**(4): p. 677-686.
83. C, H.R.a.A.P.T.a., *Data mining approach to shipping route characterization and anomaly detection based on AIS data.* Ocean Engineering, 2020. **198**: p. 106936.
84. Oh, J., H.-J. Kim, and S. Park. *Research on the Automatic Classification of Ship's Navigational Status*. 2020. Singapore: Springer Singapore.
85. Renli Chen and Xiaoqing Wu and Baijing Liu and Yueqi Wang and Zhiqiang, G., *Mapping coastal fishing grounds and assessing the effectiveness of fishery regulation measures with AIS data: A case study of the sea area around the Bohai Strait, China.* Ocean & Coastal Management, 2022. **223**: p. 106136.
86. exactEarth, *An exactEarth Technical White Paper on Satellite AIS*. 2012: http://main.exactearth.com/whitepaper-first-pass-detection.
87. Jia, H.a.P.V.a.S.T., *Estimating vessel payloads in bulk shipping using AIS data.* International Journal of Shipping and Transport Logistics, 2019. **11**: p. 25.

88. He, W., et al., *A Visual Analysis Approach to Understand and Explore Quality Problems of AIS Data.* Journal of Marine Science and Engineering, 2021. **9**(2): p. 198.
89. Vabalas, A., et al., *Machine learning algorithm validation with a limited sample size.* PLOS ONE, 2019. **14**(11): p. e0224365.
90. Berkson, J., *Application of the logistic function to bio-assay.* Journal of the American statistical association, 1944. **39**(227): p. 357-365.

Acknowledgements

The research motivation was formulated together with my supervisor, Professor Kunihiro Hamada. His valuable suggestion and advice have smoothed out the flow of my research. I am also thankful for his support in scholarship and administrative matters.

This work was impossible without the supervision and support of Prof. Hamada, Prof. Hirata, and Dr. Wada. Also, I am very much grateful to the reviewers of this thesis, Prof. Mitsuru Kitamura and Prof. Eiji Shintaku, for their suggestions and comments.

I am also grateful to my family for supporting me during these three years of hard times: my long-distance parents, my beloved wife, and my ludicrously entertaining children. At the same time, I am lucky to be together with them.

Lastly, thanks to my only discussional lab mates, Dimas and Ede, who share the roller-skating experiences for these years, along with the other Laboratory members of Transportation System Innovation, the office's staff of the Graduate School of Engineering, friends, and colleagues I cannot mention. I am benefitted from discussing and debating issues with them. Principally, I am grateful to all.