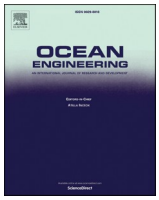
Ocean Engineering 202 (2020) 106790

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Estimate travel time of ships in narrow channel based on AIS data

Xing Wu [a,*], Uttara Roy [a], Maryam Hamidi [b], Brian N. Craig [b]

[a] Department of Civil and Environmental Engineering, Lamar University, USA
[b] Department of Industrial Engineering, Lamar University, USA

ARTICLE INFO

Keywords:
Travel time
Travel time distribution
AIS data
Houston ship channel

ABSTRACT

This paper proposes an AIS-data-based model to accurately estimate vessels' travel time and its distribution between any two points in a narrow channel. It is crucial to know the travel time (plus its distribution) between some critical points for traffic control. The model has two components: (1) to identify a vessel's destination dock, and its arrival and departure times (the key is how to separate a vessel's trips); and (2) to estimate a vessel's travel time from its destination dock (or a specific point) to another specific point, or vice versa. Travel time between two points is called traversal time. The proposed algorithms were then applied to the Houston Ship Channel (HSC) with the AIS data of 11 months in 2017, focusing on the outbound trips in the section above the Beltway-8 Bridge. Through the travel times of cargos and tankers from their destination docks to a point are likely to be normally distributed, their traversal times between two points are more likely to be lognormally distributed. The traversal time distribution is significantly impacted by vessel width, draft and type, but independent from time-of-day and month. The adjacent link traversal times are also highly positively correlated.

1. Introduction

The Port of Houston is the second busiest seaports in the United States in terms of total tonnage according to the data of 2013 (AAPA, 2015). About 200 private and public industrial terminals stretch along the 52-mile-long Houston Ship Channel (HSC), and each year, more than 8200 vessels and 215,000 barges carries more than 247 million tons of cargo through the HSC (Port of Houston, 2019). The HSC plays a critical role in supporting Texas energy and petrochemical industries. Fig. 1 shows the map of the HSC. Broadly speaking, it can be divided into two sections: the lower one goes through Galveston Ship Channel and is constraint by Galveston sea buoys, and the upper one is constraint by actual riverbanks (San Jacinto River and Buffalo Bayou). The HSC is about 45 ft (14 m) deep so it can accommodate big ships.

The HSC is narrow, especially the upper section near Houston (as the "inside channel" marked in Fig. 1). This paper only focuses on this upper section. With so many vessels moving through the channel every day, many vessels suffer from serious delays, caused by various reasons. One of the important reasons comes from big ships as they could block the entire channel when moving through some narrow sections. To better estimate the delay, and further to develop a solution scheme to optimize the vessel flows in the HSC, it is firstly necessary to understand the property of travel time of vessels in different types and sizes. Currently, the knowledge of travel time of vessels in the channel is obtained largely based on pilots' experience and work reports. Such data mainly come from reports.

This paper summarizes a preliminary study in the project on how to optimize the traffic flow in the HSC, with a focus on how to accurately estimate the travel time of vessels from one specific location to another. Such specific points indicate specific sections. For example, the Beltway-8 Bridge in the HSC is an important landmark: in the section of the channel above the bridge, the channel can only hold one big ship with a beam of 106 ft (32 m) or more, and the opposing traffic for big ships has to be interrupted. Even it is a preliminary study, 3 problems are challengeable: (1) for one vessel, how to separate the AIS data of one trip from others; (2) how to tell a vessel's arrival and departure times for different trips based on all AIS points close to the destination dock; and (3) how to precisely estimate a vessel's arrival time at any point in a channel. By solving these questions, the travel time of a vessel in a channel between two points can be found. Then another interesting question is: how are vessels' travel times distributed, and which factors impact the travel time distributions?

Note that though this paper only focuses on big ships (cargos and tankers) in the HSC, the proposed method can be applied to any big ship in inland narrow channels with docks. An inland narrow channel usually is made by dredging a natural riverbed to a certain depth to

* Corresponding author.
E-mail addresses: xing.wu@lamar.edu (X. Wu), uroy@lamar.edu (U. Roy), mhamidi@lamar.edu (M. Hamidi), brian.craig@lamar.edu (B.N. Craig).

https://doi.org/10.1016/j.oceaneng.2019.106790
Received 5 June 2019; Received in revised form 28 November 2019; Accepted 29 November 2019

Fig. 1. Map of houston ship channel.

accommodate big ships. Therefore, the navigation routes of big ships are strictly constrained within a channel including the routes connected to docks. Such a constraint makes it possible to answer the above questions simply based on the navigation data of vessels. For this reason, the proposed method cannot be applied to small vessels such as tugs, because they can be moored anywhere near the bank.

Since 2002, new ships and all sea-going vessels that have 300 gross tons or more, as well as all passenger vessels are required to equip an automatic identification system (AIS) on board. The AIS transmits the sailing status information between vessels, and from vessels to shore or vice versa. The AIS data contains the detailed and consecutive temporal and spatial information of a vessel (vessel ID, name, width, length, type), as well as real maneuvering behavior of a vessel (such as longitude, latitude, speed, course, heading, draft, etc. at every 1 min of a vessel). For this reason, recently AIS data are widely used in waterway transportation research, including the risk analysis of vessel collisions, vessel travel behavior modeling, estimate of vessel trajectory, etc. The AIS data provide us a good and reliable source of data to study the travel time of vessels in the channel. For example, they can be used to estimate the delay suffered by vessels. The delay here is referred to the waiting time of vessels due to various reasons such as navigation restrictions, beam (width) restrictions of vessels, etc. in the channel. Recently, the application of AIS data mainly focuses on two directions: risk analysis of navigation and vessels' travel behavior analysis. To analyze the risk of vessel collision and/or groundings, many studies were based on the concept of ship domain (Horteborn et al., 2019), which is a "surrounding effective waters a ship' navigator want to keep clear of other vessels or fixed objectives" (Goodwin, 1975). Through the analysis of AIS data, researchers studied the frequency of a vessel's ship domain incursion, where the ship domain was defined in various shapes (e.g., see Montewka et al. (2010), Qu et al. (2011), Weng et al. (2014) and Wu et al. (2016)). Other AIS-data-based methods were also proposed. For example, Mou et al. (2010) developed AIS-data based linear regression models to evaluate the risk of vessel collisions off the Port of Rotterdam, based on two indicators: closest point of approach (CPA) and time to closest point of approach (TCPA) between two vessels; Zaman et al. (2014) proposed an AIS-data-based fuzzy FMEA model to evaluate the risk of vessel collisions in the Malacca Strait; and Huang et al. (2019) proposed AIS-data-based crossing-line and Monte Carlo methods to assess the risk of groundings in Kaohsiung Harbor: the former aims to assess the ship grounding probability and the latter is used to simulate ship trajectory so as to evaluate the grounding frequency.

Since they provide the detailed information of vessel status continuously, AIS data become an important source to investigate vessels' travel behavior in critical sections of the channels, in order to better understand the capacity of the channel,[1] as well as to more efficiently arrange vessel traffic in the channel.

Xiao et al. (2015) investigated the travel behavior of vessels, including the distributions of vessel speeds, courses, and arrival time intervals, at two spots (one near the Port of Rotterdam, and another near Shanghai) based on AIS data. Wu et al. (2018) looked at how tankers and cargos pass through the hotspots in the Sabine-Neches Waterways (SNWW), identified in their previous study (Wu et al., 2016). Interestingly, they found that different from highway traffic, the mean travel time of vessels seems to be independent of the traffic density. Such independence was also found at the entrance of the Galveston Ship Channel (Roy and Wu, 2019). Kang et al. (2018) also studied the relationship between mean travel time of vessels and traffic density in 15 legs in the Singapore Strait. However, they found that the travel time is correlated with density. The difference between two studies may be from vessel volumes: the Singapore Strait sees much more vessel volumes than the SNWW, and Galveston Ship Channel. Further, Kang et al. (2019) investigated how ships pass through an L-shaped channel in the Singapore Strait, focusing on ships' acceleration and deceleration pattern. Zhou et al. (2019) proposed an AIS-data based method to classify ships through clustering their travel behavior.

The amount of AIS data is huge. Unavoidably, there exists outliers and missing data. How to efficiently remove outliers and make-up missing data is a big challenge. Various methods were proposed. Zhang et al. (2018) a spline model to reconstruct vessels' trajectories. This model was shown to be able to fit any types of trajectories even with circles, and clearly capture the navigation behavior of vessels. Through trajectory construction, outliers can be identified, and missing data can be made-up. Kang et al. (2019) proposed a series of algorithms to check the longitude and latitude, speed, and course. Wu et al. (2016, 2018), on the other hand, employed statistical method and spatial-relation functions in geographic information system (GIS) packages to remove outliers.

Estimating travel time at highway and roadways has been widely studied in the literature. The California Department of Transportation (Caltrans) suggested that travel time can provide information in various forms to assist managers, traffic engineers, planners, freeway users, and researchers (Choe et al., 2002). A number of studies attempted to develop algorithms for estimating travel time at highways using advanced systems such as video image processing, automatic vehicle identification, cellular phone tracking, and probe vehicles. For example, Turner (1996), Jiyoun et al. (2008) and El Esawey et al. (2011) estimated travel time on a road network by using travel time relationships with neighbor links when the sample data set is small; Nie et al. (2012) estimated each road segment's travel time distribution based on the data from loop detectors and electric toll records for the Chicago Metropolitan road network; and Wang et al. (2014) estimated travel time of any path in a city using map data sources and GPS trajectories. Such models were used for applications such as traffic monitoring (Chawla et al., 2012), taxi dispatching (Yuan et al., 2013), and travel time reliability analysis (Nie et al., 2012; Wu and Nie, 2013; Wu, 2013).

Travel time data were also used for various applications in waterway studies. Using the Houston Pilot transit data, Rahimikelarijani et al. (2018) and Kaneria et al. (2019) used travel times to determine the optimal closure scenario in the HSC (due to the demolition and re-construction of the Beltway-8 Bridge) for minimizing vessel waiting

[1] The capacity of a section of channel refers to the number of vessels able to go through the section of a channel in the unit time. This is the maximum traffic volume considering weather conditions, vessel size, vessel's behavior, speed, safe distance and so on (Liu et al., 2016).

time. For estimating travel time of ships on inland waterways, Asamer and Prandtstetter (2014) developed two methods based on historical and currently observed velocities from the recorded ship trajectories for estimating ship travel times in an inland waterway; DiJoseph and Mitchell (2015) proposed a method to establish baseline travel time statistics using archived AIS position in an inland river system; and Riley et al. (2015) used AIS data and river stage data for analyzing the relationship between vessel transit times and water levels, and for measuring vessel travel times over two years.

But none of these above-mentioned studies is about trip separation and arrival time estimation at any point in a channel (e.g., the study of DiJoseph and Mitchell (2015) is based on the given AIS positions in the channel). An AIS dataset may cover a long time period, in which a vessel could visit the channel of interest more than once. To estimate vessels' travel time based on AIS data, it is important to identify different trips of a vessel. This study aims to build a *more general* method to determine the travel time of vessels in a channel covering any segment considering different vessel types and navigation status (such as the travel time from the departure or the travel time under normal navigation status). It is the major methodological contribution of this paper.

By building the travel time inventory, the travel time reliability of different types of vessels can be obtained and the delay can be quantitatively estimated. This opens the door for optimizing the traffic flow of vessels in a narrow and busy channel, such as the closure and vessel scheduling in the HSC. In order to schedule a closure (e.g., due to dredging tasks or construction) at a specific location of a waterway, the vessel traffic service (VTS) needs to consider vessel travel time to the closed section of the channel to determine the best time for closure. The best closure time minimizes the total waiting time of all arrival and sailing vessels affected by the closure. Although pilots and coast guard constantly work with real-time AIS data, no historical estimations are integrated with their system to optimize the operation. Another application is vessel scheduling. Pilots need to schedule and order vessels in the channel. The complexity of the scheduling arises from navigation restrictions imposed on vessels based on their sizes. For example, vessels with beam (width) greater that 106 ft (32m) are one-way traffic restricted above the Beltway-8 Bridge or liquefied petroleum gas (LPG) tankers with length greater than 560 ft (171 m) are daylight restricted in the HSC. Such limitations increase the congestion in the channel and potentially generate delays for other vessels. Optimizing navigation at the channel considering different navigation rules directly needs travel time estimation.

The reminder of this article is organized as follows. Section 2 gives the methodological framework used to estimate travel time of vessels in the channel; Section 3 describes the numerical examples, where the property of travel time of big ships such as tankers and cargos in the HSC was investigated; and Section 4 concludes the paper.

2. Methodology

2.1. Arrival and departure time at destination dock

As to those questions mentioned at the beginning, this section aims to answer the first two: (1) to determine the destination dock of a big ship, and (2) to determine the arrival and departure times of a big ship at its destination dock.

2.1.1. Model formulation

To investigate the travel time inventory of vessels in a busy narrow channel, it is necessary to find the destination docks of a vessel, as well as its arrival time at this dock and departure time from this dock, because the speeds of a vessel near its destination dock will be very different from those when a vessel is in normal navigation status. Also, to build the travel time inventory, it is necessary to find the destination dock of a vessel as its destination and origin.

In the dataset on vessel information, the destination of this vessel is reported. However, such destination only refers to the destination port. A port usually has many docks, as shown in Fig. 2 (a). There are totally 81 docks along the HSC. Given a narrow channel like the HSC, which have a large number of big ships such as tankers and cargos, the water depth near docks are much larger than the depth near the bank without dock. Therefore, for a big ship, if its AIS points are close to one dock (as shown in Fig. 2), it implies that this dock is the destination dock of this ship.

Given a set of AIS data, for one vessel, if it had only one destination dock, and visited this destination dock only once, then it would be easy to determine this vessel's destination dock and the corresponding arrival and departure times, respectively – we can get the information by simply checking all of this vessel's AIS data which are close to docks, e.g., within the buffer of this dock, see Fig. 2(b). However, a given AIS dataset usually covers the whole month. It is possible that a vessel may visit more than one destination dock, and/or visit a dock more than once. For this reason, it is necessary to develop a general algorithm to identify the destination dock, and the arrival and departure times, respectively.

When applying to the AIS data of the HSC, the buffer radius is 150 ft (46 m) to determine the destination dock of a vessel. Only the AIS data inside the buffer are analyzed to study the arrival and departure time at the destination dock. By doing this, the AIS data with spatial error are excluded for further analysis.

Consider an inland waterway transportation system with a series of

(a) AIS data of one vessel (b) Detailed view of AIS near the dock

Fig. 2. Example of identifying the destination dock of a vessel based on its AIS data.
(a) AIS data of one vessel. (b) Detailed view of AIS near the dock
Note: the brown line segments on the banks of the channel refers to the locations of docks. The area around a dock line refers to the buffer of this dock, with a radius of 150 ft (about 46 m). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

docks: $j = 1, 2, \cdots, n_d$, and vessels $i = 1, 2, \cdots, n_v$, where n_d and n_v refer to the numbers of unique docks and vessels in this system, respectively. Then, denote $\delta_{ij} = 1$ if vessel i was moored at dock j; and $\delta_{ij} = 0$, otherwise. For vessel i, an AIS point contains the following information at one timestamp t: longitude, latitude, speed over ground (SOG) and course over ground (COG) (in the following text, "speed" and "course" are used, referring to SOG and COG, respectively). These types of information for vessel i are denoted as: x_i^t, y_i^t, s_i^t,and c_i^t, respectively. Given a timestamp, the AIS point of a vessel can be uniquely identified. It implies that given t_i, x_i^t, y_i^t, s_i^t,and c_i^t can be identified.

Moreover, the timestamp series of two different vessels are not the same. For example, for one vessel, the timestamp series are 2018-08-01 00:01:00, 2018-08-01 00:02:02, 2018-08-01 00:03:04, …; while another one's timestamp series can be 2018-08-01 00:01:15, 2018-08-01 00:02:16, 2018-08-01 00:03:14, …Such inconsistency of timestamps increases the difficulty of AIS data analysis.

On the other hand, if $\delta_{ij} = 1$, let λ_{ij}^k and μ_{ij}^k be the arrival and departure times of vessel $i$at dock j in its k^{th}visit, respectively, where $j = 1, 2, \cdots, n_{ij}$, and n_{ij} refers to the number of visits of vessel $i$to dock j. To get λ_{ij}^k and τ_{ij}^k, we need to define a subset of the AIS points of vessel i: the AIS points in this subset should be close enough to the dock, implying the moored status or the status close to the moored status. Such a subset can be obtained through a GIS tool, such as function "Buffer" (available in popular GIS packages, such as ArcGIS and PostGIS). The purpose of using a buffer is to filter data for the next step analysis, so as to avoid scanning the entire dataset. As aforementioned, the buffer radius is set to be 150 ft (about 46 m) based on the geometric property of the HSC. As shown in Fig. 2(b), vessel i's AIS points within the buffer of dock j form the aforementioned subset, denoted as Γ_{ij}. To reduce the complexity of notations, in the following text, for vessel i, we simply use $\forall s_i^t \in \Gamma_{ij}$(or $\forall c_i^t \in \Gamma_{ij}$) to indicate the speed (or the course) of an AIS point that is inside the buffer of dock j. As to timestamps, we denote T_{ij} as the subset of vessel i's timestamps, at which the AIS points are inside the buffer of dock j; and $\overline{T}_{ij}$ as the complimentary of T_{ij}, i.e., the subset of vessel i's timestamps, at which the AIS points are outside the buffer of dock j. Please refer to Appendix B at the end of the paper for the full list of notations.

Then, the arrival and departure times of vessel i at dock $j$from its k^{th} visit, i.e., λ_{ij}^k and μ_{ij}^k, can be determined based on the following two formulations, respectively. Note that at the arrival or departure time, a vessel's speed (SOG) is 0, as shown in Conditions (2) and (4), respectively.

$$\lambda_{ij}^k = \min_{\forall t_i > \mu_{ij}^{k-1},\ t_i \in T_{ij}} t_i \tag{1}$$

$$\text{s.t.} \quad s_i^t = 0,\ \forall s_i^t \in \Gamma_{ij}, \tag{2}$$

$$k = 2, \cdots, n_{ij}$$

$$\mu_{ij}^k = \max_{\forall t_i < \zeta_{ij}^k,\ t_i \in T_{ij}} t_i \tag{3}$$

$$\text{s.t.} \quad s_i^t = 0,\ \forall s_i^t \in \Gamma_{ij} \tag{4}$$

$$\zeta_{ij}^k = \min_{\forall t_i > \lambda_{ij}^k,\ t_i \in \overline{T}_{ij}} t_i,\ \ k = 1, \cdots, n_{ij} \tag{5}$$

where t_i is the timestamp of vessel i(associated with an AIS point); and ζ_{ij}^k in Condition (5) is an auxiliary variable, referring to the earliest timestamp after vessel i left the buffer of dock $j$after its k^{th} visit to dock j.

To solve λ_{ij}^k and μ_{ij}^k, respectively, the following boundary conditions are needed:

$$\begin{cases} \lambda_{ij}^1 = \min_{\forall t_i \in T_{ij}} t_i,\ \text{where}\ \ s_i^t = 0,\ \forall s_i^t \in \Gamma_{ij} \\ \mu_{ij}^{n_{ij}} = \max_{\forall t_i \in T_{ij}} t_i,\ \text{where}\ \ s_i^t = 0,\ \forall s_i^t \in \Gamma_{ij} \\ n_{ij} = 1,\ \text{if}\ \ \forall t_i\ in\ \lambda_{ij}^1 \le t_i \le \mu_{ij}^{n_{ij}},\ \text{we have}\ t_i \in T_{ij} \end{cases} \tag{6}$$

where λ_{ij}^1 and $\mu_{ij}^{n_{ij}}$ refer to the arrival time of the first visit ($k = 1$) and the departure time of the last visit ($k = n_{ij}$), respectively; n_{ij} is the number of visits made by vessel i at dock j; and t_i refers to a timestamp associated with an AIS point of vessel i.

The last boundary condition implies that for vessel i, there is only one visit (i.e., $n_{ij} = 1$), if no AIS point between the arrival time of its first visit and the departure time of its last visit, is outside the buffer of dock j. This idea of determining the arrival and departure times is also illustrated in Fig. 3.

As shown in Formulations (1–6), the arrival time is regarded as the earliest time when a vessel's speed is 0 after it enters the buffer of a dock; and the departure time is regarded as the latest time when its speed is 0 before leaving the buffer. Note that the berthing and un-berthing process for a big ship to a dock is complicated. However, since the focus of this paper is an inland channel made by dredging the riverbed to a certain depth to accommodate big ships, a big ship's route is strictly limited within the channel of a waterway/river. Therefore, when a big ship is found close to a dock, it can be regarded in the process of berthing or unberthing. Moreover, the constraints on speed and timestamp can exclude the error data. Given a large data set (e.g., 11 months' AIS data were considered in this paper), the impact of such error could be marginal (see the details in Section 3).

2.1.2. Solution algorithm

To solve Formations (1–6), one key is to separate trips made by vessel i at dock j, i.e., determining k. It can be solved through an iterative procedure. After determining the arrival and departure times of the first visit ($k = 1$), starting from $k = 2$, the arrival time of the k^{th} visit must be the earliest time with speed of 0 (i.e., $s_i^t = 0$, where $t = \lambda_{ij}^k$), larger than the departure time of the $(k-1)^{th}$ visit. Once the arrival time of the k^{th} visit, i.e., λ_{ij}^k, is determined, the trip of the k^{th} visit is thus separated. By repeating this procedure, the trip of each visit can be separated until we cannot find an arrival time later than the departure time of the last visit.

The following gives the details of this iterative algorithm based on

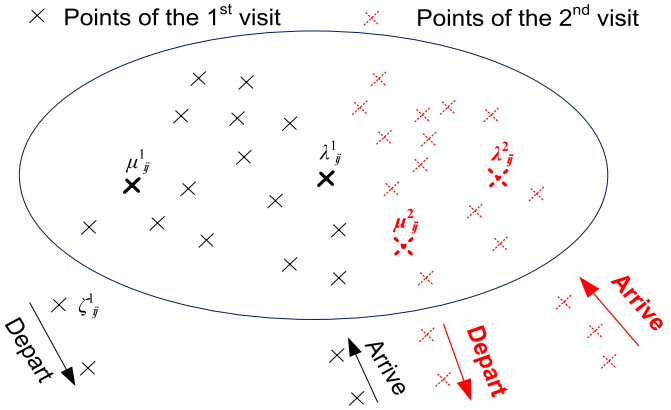


Fig. 3. Illustration of the solution idea to identify the arrival and departure time for each visit.

Note: the dots in the oval refer to the subset Γ_{ij}; the points associated with λ_{ij}^1 and λ_{ij}^2, μ_{ij}^1 and μ_{ij}^2 refer to the points with the arrival and departure times of two visits, respectively; and ζ_{ij}^1 refers to the point with the earliest timestamp after the first visit; and in the real world, the AIS points from multiple trips are mixed together, but in this case, for better illustration, the AIS points from two trips are marked separately.

vessel i's AIS data.

Input: AIS dataset of vessel i and the geometric location of dock j.
Step 1: For dock j, define a buffer. Then based on the AIS dataset of vessel i, employ GIS function "Buffer" to define a set of AIS data: Γ_{ij}.
Step 2: Based on Γ_{ij}, find $\lambda_{ij}^1, \mu_{ij}^{n_{ij}}$, respectively. If λ_{ij}^1 and/or $\mu_{ij}^{n_{ij}}$ cannot be found, it implies that vessel $i$may not be moored at dock j(it may just pass by with a small distance to the dock – it could happen to small vessels, such as tugs), then let $\delta_{ij} = 0,$and go to vessel $i+1$to restart this procedure; otherwise, let $\delta_{ij} = 1,$and go to Step 3.
Step 3: If $n_{ij} = 1$ based on the last condition of Eq. (6), implying that there is only one visit. Let $\mu_{ij}^1 = \mu_{ij}^{n_{ij}}$, and then record the λ_{ij}^1 and μ_{ij}^1, respectively. Go to the next vessel $i+1$and restart. Otherwise if $n_{ij} > 1,$go to Step 4.
Step 4: At $k = 1$, determine ζ_{ij}^k according to Eq. (5). Then based on ζ_{ij}^k, determine departure time μ_{ij}^k according to Formations (3–4). Go to Step 5.
Step 5: At $k = k+1$, solve arrival time λ_{ij}^k based on μ_{ij}^{k-1} according to Formulations (1–2), and then determine ζ_{ij}^k. Based on ζ_{ij}^k, determine departure time μ_{ij}^k. Repeat this step until $\mu_{ij}^k = \mu_{ij}^{n_{ij}}$.

For many large ships, such as tankers and cargos, the number of visits at one dock in a month is no more than 1; and only a few have 2 visits. But for small vessels, such as tugs or harbor vessels, the number of visits in a month can be more than 10. Actually, it is found that the aforementioned method can accurately determine the arrival and departure times of big ships (see the next section on numerical experiments). However for small vessels, some errors exist, because as mentioned before, these small vessels can be moored near any bank due to their small drafts. In this paper, only big ships are focused on.

2.2. Timestamp of passing by a dock

Now, we aim to answer the last question – how to precisely estimate a vessel's arrival time at any point in a channel. The model and algorithm developed in the following can be applied to any point in a channel.

As mentioned in the first section, in a narrow channel where many docks are located, a big vessel's movement can affect the navigation of many others which need to leave from these docks. Therefore, it is important to know the vessels' travel time from one dock to another, so as to better arrange big vessels' traffic. For instance, when a wide-body vessel is moving to its destination dock, and a sailing vessel is waiting at a dock above the destination dock of that wide-body one, then the sailing vessel could start sailing only in the condition that it cannot pass the destination dock of that arriving wide-body vessel until this wide-body one arrives at its destination. Thus, in this case, the dock-to-dock travel time (from the dock where the sailing vessel is moored to the destination dock of the wide-body vessel) is an important factor for scheduling the departure time of a vessel.

2.2.1. Dock line

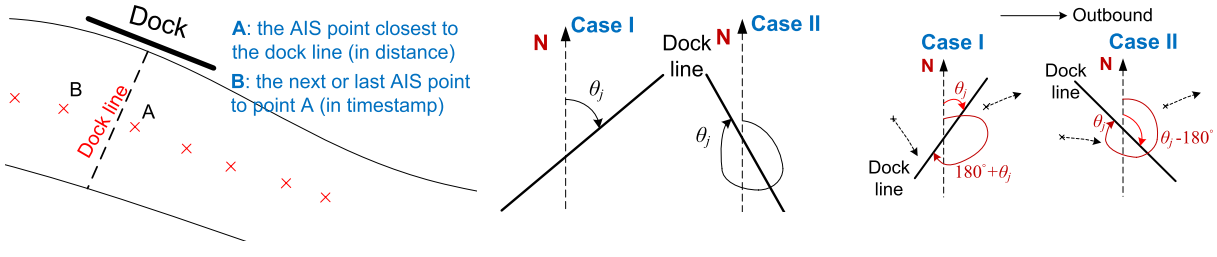
To determine the dock-to-dock travel time of a vessel, it is necessary to determine the timestamp when a vessel passed by a dock (not stop there). As shown in Fig. 2(a), a dock is a polyline along the waterway. For consistency, we define a "*dock line*" based on the centroid of a dock (also see Fig. 4(a)). Such a dock line crosses the channel, so the time when a vessel passing by a dock can be regarded as the time when arriving at the dock line of this dock. Each dock has a dock line, so in the following, we directly call dock j's dock line as dock line j. We call this time as the "arrival time" of a vessel at a dock line. Note that this "arrival time" is different from the arrival time of a vessel at a dock, as discussed on Section 2.1.

2.2.2. Point closest to dock line

Since a vessel's AIS points are discrete, intuitively, such time can be determined via *interpolation* based on the timestamps of two AIS points that are closest to the dock line (one at each side of the dock line, as points A and B shown in Fig. 4(a)). In application, however, there are two challenges (1) if a vessel had multiple trips (i.e., went through a dock line back and forth more than once): for the period covered by the given AIS dataset, then how to extract these two points for each trip? (2) In a few cases, two points that closest to a dock line are on the same side of the dock line, then how to find alternative one on the other side of the dock line?

To increase the search efficiency, like the method in Section 2.1, we define a subset of one vessel's AIS points which are close enough to a dock line. This can be also achieved by using function "Buffer" provided by GIS package. Since some small vessels may have a speed up to 20 knots in the channel, and time interval between two consecutive AIS points are usually 1 min. In this case, the radius of this buffer is set to be 300 ft (about 91.5 m). Note that in last section, the buffer's radius is set to be 150 ft only for determining a vessel's destination dock. The reason is that under the berthing and unberthing process, vessels are moving with low speeds. If such radius is too large, lots of points under the normal navigation status could be included, making the computational load unnecessarily large. On the other hand, under the normal navigation status, vessels have much higher speeds. In order to ensure to capture two points on each side of a dock line (i.e., A and B as shown in Fig. 4 (a)), the radius of this type of buffer needs to be large enough.

The AIS points of vessel i covered by the buffer of dock line j form a subset, denoted as Π_{ij}. For simplicity, T_{ij} is still defined as the subset of timestamps of vessel i's AIS points within the buffer of dock line j. Note that the AIS points within Π_{ij} may be from several trips. Therefore, it is



(a) Two points for interpolation: A and B

(b) Defining the angle of dock line, θ_j

(c) Determining the moving direction

Fig. 4. Demonstration of points for interpolation and the angle of dock lines in two cases. (a) Two points for interpolation: A and B. (b) Defining the angle of dock line, θ_j. (c) Determining the moving direction.
Note: the points represent AIS points, and the arrows at points in (c) represent the courses of navigation.

important to separate the data of one trip from others.

Given that a vessel visited the dock(s) in the channel (from the Ocean) more than once within the time period covered by an AIS dataset, if moving inbound, the time when a vessel passed through a dock line should be earlier (less) than its arrival time at its destination dock for this visit (λ_i^k, as this is the k^{th} trip), but later (larger) than the departure time of its last visit (μ_i^{k-1}); on the other hand, if moving outbound, the time when a vessel passed through a dock line should be later (larger) than the departure time of this visit (μ_i^k), but earlier (less) than the arrival time of its next visit (λ_i^{k+1}). Note that λ_i^k and μ_i^k are different from the notations of the arrival and departure times of vessel i at dock j used in Section 2.1.2, because its destination dock may vary in different visits to the channel. However, they can still be determined through the solution procedure described in Section 2.1.2 by sorting the arrival time of vessel i for all of its destination docks.

This logic can be used to identify AIS points A and B in Fig. 4(a). As mentioned above, for the convenience of using notations, we denote λ_i^k and μ_i^k, $k = 1, 2, \cdots, n_i$, as the arrival and departure time (to and from the destination dock) for the k^{th} visit of vessel i, where n_i refers to the total number of visits made by vessel i in the given AIS dataset. λ_i^k and μ_i^k are used to limit the search range.

Let d_{ij}^t be the distance of the AIS point of vessel i at timestamp t, to dock line j. Note that for one vessel, one timestamp uniquely determines one AIS point. Such a distance can be obtained via function "ST_Distance" in PostGIS, which calculates the distance between two geometric objects if they are defined in the same spatial reference system. Therefore, within a pool of AIS points, the point closest to dock line j in the k^{th} visit can be identified iteratively using the following formulations for inbound and outbound trips, respectively. The entire iterative solution algorithm is given at the end of this section (Section 2.2.5).

$$\min\_d_{ij}^k = \min_{\forall\, t_i \in T_{ij}} d_{ij}^t \tag{7}$$

For outbound trips

$$\text{s.t.} \quad \mu_i^k < t_i < \lambda_i^{k+1}, \quad k = 1, \cdots, \quad n_i \tag{8}$$

$$\lambda_i^{n_i+1} = \max t_i \tag{9}$$

For inbound trips

$$\text{s.t.} \quad \mu_i^{k-1} < t_i < \lambda_i^k, \quad k = 1, \cdots, \quad n_i \tag{10}$$

$$\mu_i^0 = \min t_i \tag{11}$$

where $\min\_d_{ij}^k$ refers to vessel i's minimum distance to dock line j, based on the AIS points in subset Π_{ij} and considering the AIS points from its k^{th} visit only; $\max t_i$ and $\min t_i$ refer to the earliest and latest timestamp of vessel i in the given dataset; μ_i^k and λ_i^k refer to the arrival and departure times (to and from the destination dock) of vessel i in its k^{th} visit, where $k = 1, 2, \cdots, n_i$. Note that Eqs. (9) and (11) are just two boundary conditions, not actually indicating the arrival or departure time. For example, for vessel i having two visits, if dock line j is above the destination dock of the second visit but below the destination dock of the first visit, we may not be able to find $\min\_d_{ij}^2$ for any t later than μ_i^2, i.e., the departure time of the second visit of vessel i, because Condition (8) or (10) makes t_i out of T_{ij} .

For the purpose of interpolation, the first AIS point needed (as point A in Fig. 4(a)) is the one closest to the dock line of interest, and its timestamp is denoted as $\text{dmin}\_t_{ij}^k$. Another (i.e., point B in Fig. 4(a)) is the one last of or next to the first point. It is not easy to get this "last" or "next" point. Before answering this question, it is needed to determine the moving direction of a vessel based on its AIS data.

2.2.3. Moving direction

To answer the question of how to identify two points for interpolation, first, it is necessary to extract the AIS points from a vessel's one-direction movement, such as inbound or outbound movement. To achieve this purpose, we can compare the course of an AIS point, c_i^t, with the direction of dock line j (denoted as θ_j). In this paper, the direction of a dock line is defined as the angle from the true north (N) in the clockwise direction: θ_j is between 0° and 90° (case I) or between 270° and 360° (case II), as shown in Fig. 4(b).

Once the direction of a dock line is determined, the course of each AIS point can be used to determine the vessel's moving direction. As shown in Fig. 4(c), the rules are as follows. Without generality, the direction to the right is regarded as outbound.

Note:

(1) c_i^t refers to the course of vessel i at time t; and θ_j is the direction of dock link of dock j.
(2) A course is regarded error if it is greater than 360° or less than 0°.
(3) Case I or II depends on the direction of the dock line (see Fig. 4 (b)).

Vessel traffic in a channel is not as regular as highway traffic. Some small vessels, such as tugs or fish boats may stop frequently, and make a U-turn anywhere. Also, big ships make U-turn at some specific locations. However, this paper aims to get the travel time information in a channel based on "normal navigation", so the travel times extracted from the vessels not under the "normal" navigation status, such as making U-turns, will be marked as outliers, and excluded in the analysis.

2.2.4. Next or last point

As mentioned before, two points are needed for interpolation: the first one is the point (from a given trip) closest to dock line j; and the second is the *last or next* point of the first one in terms of their timestamps. As shown in Fig. 4(c), given an outbound trip, if the point closest to the dock line is on the left side of the dock line, then its next point is needed; otherwise, it is needed to identify its last point. The rule is similar for an inbound trip. Denote $\text{nl}\_t_{ij}^k$ as the timestamp of this desired "next" or "last" point. This section will discuss how to get $\text{nl}\_t_{ij}^k$ for vessel i when it passed through dock line j in its k^{th} visit in the channel.

As described above, given the AIS point which is closest to the dock line (for vessel $i$to dock j's dock line), first it is necessary to judge if this point is on the right or left side of a given dock line. For this purpose, we define φ_{ij}^t as the direction from the true north (N) to the line from the centroid of the dock line to this point (suppose that the timestamp of this point is t), as shown in Fig. 5.

From Fig. 5, it is not difficult to see such rules for two cases, respectively:

If $\varphi_{ij}^t = \theta_j$, this point is exactly on the dock line, i.e., $\min\_d_{ij}^k = 0$. Therefore, the timestamp of this point, $\text{dmin}\_t_{ij}^k$, is exactly the time when vessel i crossed the dock line of dock j.

Otherwise, go to the following formulations to find $\text{nl}\_t_{ij}^k$ and then conduct the interpolation.

For brevity, the AIS point closest to the dock line is called the *closest AIS point* or *closest point*. Formulations (12–27) were developed to determine $\text{nl}\_t_{ij}^k$, considering 2 cases for an inbound or outbound trip, respectively. For brevity, the scenarios that the closest point is on the right or left side of the dock line, are put together. Formulations (12a - 14a, 15), (16a - 18a, 19), (20a - 22a, 23) and (24a - 26a, 27) are for the scenario that the closest point is on the right side of the dock line; and formulations (12b–14b, 15), (16b–18b, 19), (20b–22b, 23) and (24b–26b, 27) are for the scenario of the closest point on the left side.

For instance, in the case of outbound trips, if the closest point is on the right side of the dock line, we aim to find its last point, which must be on the left side of this dock line, and its timestamp must be the latest (i.

- Case I: the point is from an outbound trip if $\theta_j < c_i^t < 180^\circ + \theta_j$, or inbound otherwise.
- Case II: the point is from an outbound trip if $c_i^t > \theta_j$or $c_i^t < \theta_j - 180^\circ$, or inbound otherwise.

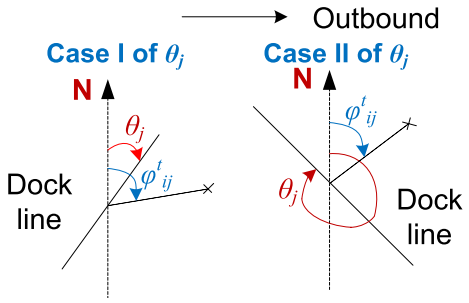


Fig. 5. Demonstration of the angle φ_{ij}^t from the centroid of the dock line to the AIS point.
Note: no matter which case of θ_j, φ_{ij}^t is always measured in the clockwise direction from the true north (N) to the connected line from the centroid of the dock line to the AIS point of interest.

e., largest) timestamp in all of this vessel's timestamps earlier than the timestamp of the closest point, i.e. $\text{nl\_}t_{ij}^k < \text{dmin\_}t_{ij}^k$; on the other hand, if the closet point is on the left side of the dock line, we aim to find its next point, which must be on the right side of this dock line, and its timestamp must be the earliest (i.e., smallest) timestamp among all of this vessel's timestamps later than the timestamp of the closest point, i.e., $\text{nl\_}t_{ij}^k > \text{dmin\_}t_{ij}^k$.

Conditions (14a, 14b), (18a, 18b), (22a, 22b) and (26a, 26b) indicate where the closest AIS point is located: on the right or left side of the dock line; and conditions (15), (19), (23) and (27) indicate if it is an inbound or outbound trip, respectively, as discussed in Section 2.2.3. Readers are referred to Appendix B at the end of the paper for the list of all notations.

Case I of dock line (for outbound trips):

$$\text{nl\_}t_{ij}^k = \max_{\forall\ t_i \in T_{ij}} t_i \tag{12a}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k < \text{dmin\_}t_{ij}^k \tag{13a}$$

$$\theta_j < \varphi_{ij}^t < 180^\circ + \theta_j \tag{14a}$$

Or

$$\text{nl\_}t_{ij}^k = \min_{\forall\ t_i \in T_{ij}} t_i \tag{12b}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k > \text{\_}t_{ij}^k \tag{13b}$$

$$\varphi_{ij}^t > 180^\circ + \theta_j \text{ or } \varphi_{ij}^t < \theta_j \tag{14b}$$

$$\theta_j < c_i^t < 180^\circ + \theta_j,\quad t = \text{dmin\_}t_{ij}^k \tag{15}$$

Case I of dock line (for inbound trips):

$$\text{nl\_}t_{ij}^k = \min_{\forall\ t_i \in T_{ij}} t_i \tag{16a}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k > \text{dmin\_}t_{ij}^k \tag{17a}$$

$$\theta_j < \varphi_{ij}^t < 180^\circ + \theta_j \tag{18a}$$

Or

$$\text{nl\_}t_{ij}^k = \max_{\forall\ t_i \in T_{ij}} t_i \tag{16b}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k < \text{dmin\_}t_{ij}^k \tag{17b}$$

$$\varphi_{ij}^t > 180^\circ + \theta_j \text{ or } \varphi_{ij}^t < \theta_j \tag{18b}$$

$$c_i^t > 180^\circ + \theta_j,\ \text{ or } c_i^t < \theta_j,\ \ t = \text{dmin\_}t_{ij}^k \tag{19}$$

Case II of dock line (for outbound trips):

$$\text{nl\_}t_{ij}^k = \max_{\forall\ t_i \in T_{ij}} t_i \tag{20a}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k < \text{dmin\_}t_{ij}^k \tag{21a}$$

$$\varphi_{ij}^t > \theta_j \ \text{ or } \ \varphi_{ij}^t < \theta_j - 180^\circ \tag{22a}$$

Or

$$\text{nl\_}t_{ij}^k = \min_{\forall\ t_i \in T_{ij}} t_i \tag{20b}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k > \text{dmin\_}t_{ij}^k \tag{21b}$$

$$\theta_j - 180^\circ < \varphi_{ij}^t < \theta_j \tag{22b}$$

$$c_i^t > \theta_j \ \text{ or } \ c_i^t < \theta_j - 180^\circ,\ \ t = \text{dmin\_}t_{ij}^k \tag{23}$$

Case II of dock line (for inbound trips):

$$\text{nl\_}t_{ij}^k = \min_{\forall\ t_i \in T_{ij}} t_i \tag{24a}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k > \text{dmin\_}t_{ij}^k \tag{25a}$$

$$\varphi_{ij}^t > \theta_j \ \text{ or } \ \varphi_{ij}^t < \theta_j - 180^\circ \tag{26a}$$

Or

$$\text{nl\_}t_{ij}^k = \max_{\forall\ t_i \in T_{ij}} t_i \tag{24b}$$

$$\text{s.t.}\quad \text{nl\_}t_{ij}^k < \text{dmin\_}t_{ij}^k \tag{25b}$$

$$\theta_j - 180^\circ < \varphi_{ij}^t < \theta_j \tag{26b}$$

$$\theta_j - 180^\circ < c_i^t < \theta_j,\ \ t = \text{dmin\_}t_{ij}^k \tag{27}$$

Note that $\text{dmin\_}t_{ij}^k$ can be determined based on Formulation (7–11). Formulations (12–27) provide a theoretical base to calculate the

- **Case I**: the point is on the right side of the dock line, if $\theta_j < \varphi_{ij}^t < 180^\circ + \theta_j$, or on the left side if $\varphi_{ij}^t > 180^\circ + \theta_j$ or $\varphi_{ij}^t < \theta_j$;
- **Case II**: the point is on the right side of the dock line if $\varphi_{ij}^t > \theta_j$ or $\varphi_{ij}^t < \theta_j - 180^\circ$, or on the left side if $\theta_j - 180^\circ < \varphi_{ij}^t < \theta_j$.
- **Note:** Case I or II depends on the direction of the dock line (see Fig. 4(b)).

timestamp (i.e., $\text{nl\_}t_{ij}^k$) of the AIS point last or next of the closest point. Once $\text{nl\_}t_{ij}^k$ is found, the distance from the corresponding AIS point to the dock line can be easily found via the aforementioned GIS function "Distance". Let $\text{nl\_}d_{ij}^k$ denote such a distance. Then the time when vessel i crossed the dock line of dock j during its k^{th} visit, denoted as τ_{ij}^k, can be determined through interpolation, as Eq. (28).

$$\tau_{ij}^k = \text{nl\_}d_{ij}^k \cdot \frac{\text{dmin\_}t_{ij}^k}{\text{min\_}d_{ij}^k} \tag{28}$$

Based on Eq. (28), given visit k of vessel i, the travel time between dock j and dock $j+1$, denoted as tt_{jj+1}^{ik}, can be determined as:

For outbound trips

$$tt_{jj+1}^{ik} = \tau_{ij+1}^k - \tau_{ij}^k \tag{29a}$$

For inbound trips

$$tt_{jj+1}^{ik} = \tau_{ij}^k - \tau_{ij+1}^k \tag{29b}$$

where τ_{ij}^k refers to the timestamp when vessel $i$crossed the dock line of dock j during its k^{th} visit.

2.2.5. Solution algorithm

Sections 2.2.2 -2.2.4 discuss the theoretical method of how to determine the time when a vessel crossed a dock line. These formulations can be solved iteratively. In this section, a more detailed solution procedure is given.

Given vessel i, based on the method described in Section 2.1, its arrival and departure times can be determined. As defined before, μ_i^k and λ_i^k refer to the departure and arrival times of vessel $i$in its k^{th} visit, and μ_i^{k+1} the departure time of vessel i in its k+1st visit. We have $\mu_i^{k-1} < \lambda_i^k < \mu_i^k$, i.e., for vessel i, the departure time of the k^{th} trip should be later than the arrival time of this visit, and the arrival time of the k^{th} visit should be later than the departure time of the $k-1$st visit. Note that if $k = 1$, $\mu_i^{k-1} = \min t_i$, i.e., the earliest timestamp of vessel $i$in the dataset; or if $k = n_i$, i.e., the last visit, $\lambda_i^{k+1} = \max t_i$, i.e., the latest timestamp of vessel i in the dataset. μ_i^k and λ_i^k can be determined by using the procedure described in Section 2.1.2.

Input: AIS dataset of vessel i and the geometric location of the dock line of dock j (in the following, we called dock line j for simplicity).
Step 0: Based on the solution algorithm (procedure) described in Section 2.1.2, identify the departure and arrival time of each trip, i.e., μ_i^k and λ_i^k, $k = 1, 2, \cdots, n_i$.
Step 1: Define a buffer based on the geometric location of dock line j. Then define the AIS point set Π_{ij} based on this buffer. If Π_{ij} is empty, it implies that vessel i never crossed dock line j. Stop. Otherwise, determine the direction of dock line (i.e., θ_j, see Fig. 4(b)), and go to Step 2.
Step 2: Extract AIS points in Π_{ij} based on the condition $\mu_i^k < t < \lambda_i^k$, for $k = 1, 2, \cdots, n_i$. Given visit k, if no data available, go to Step 2.6; otherwise, go to Step 2.1.
Step 2.1: The AIS data between μ_i^k and λ_i^k could cover the navigation data in both inbound and outbound directions. According to the research needs, extract the data in the outbound or inbound direction only based on the condition described in Section 2.2.3.
Step 2.2: Based on the dataset extracted between μ_i^k and λ_i^k for one direction, determine the AIS point that is closest to the dock line, i.e., the point with $\text{min\_}d_{ij}^k$, and the corresponding timestamp $\text{dmin\_}t_{ij}^k$.
Step 2.3: Determine φ_{ij}, so as to determine if we need to find the next or last point of the AIS point that is closest to the dock line, for the purpose of interpolation.
Step 2.4: Determine the desired last or next point based on the formulations described in Section 2.2.4. If such last or next point does not exist, it indicates that vessel $i$never crossed dock line j when moving inbound or outbound between time μ_i^k and λ_i^k, so go to Step 2.6; otherwise, go to Step 2.5.
Step 2.5: Based on these two points, conduct the interpolation to determine the time when vessel i across dock line j, i.e., τ_{ij}^k.
Step 2.6: Go to the k+1$^{\text{st}}$ trip, and repeat Steps 2.1–2.5 until all trips made by vessel $i$is studied. Stop.

Remarks:

(1) In Step 1, Π_{ij} can be obtained easily using GIS function "Buffer", which is available in most popular GIS packages, such as ArcGIS and PostgreGIS.
(2) If $k = 1$, $\mu_i^{k-1} = \min t_i$, i.e., the earliest timestamp of vessel i in the dataset; or if $k = n_i$, i.e., the last trip, $\lambda_i^{k+1} = \max t_i$, i.e., the latest timestamp of vessel $i$in the dataset.
(3) In case that no $\text{dmin\_}t_{ij}^k$ is available based on formulations (7–11), it implies that vessel $i$did not cross dock link j in its k^{th} visit.

In the following, the AIS data in the HSC from 11 months of 2017 were used to test the algorithms described above to determine the travel time, as well as to analyze the properties of travel time in the HSC.

3. Application to HSC

3.1. Review of AIS data and testbed

For testing the algorithms described above, we downloaded the AIS data from marinecadastre.gov (developed by the U.S. National Oceanic and Administration Office and Bureau of Ocean Energy Management). The data of 2017 excluding February were employed (errors were reported when importing the data of February to the database). For 11 months, there are totally 102,187,704 records in the upper section of the HSC (see Fig. 1), and each month has 9,289,791 records on average. All data were imported to a PostgreSQL database, and the solution algorithms in Sections 2.1.2 and 2.2.5 were coded in PL/pgSQL.

Though the data covers the whole upper section of the HSC, this paper particularly looks at the section above the Beltway-8 Bridge (see Fig. 7). In this section of the HSC, the first dock above the Bridge is dock 48, and the last one is dock 107. The dock lines of docks 55 and 54 are in a branch of the HSC (not in the main channel). The AIS data from the 11

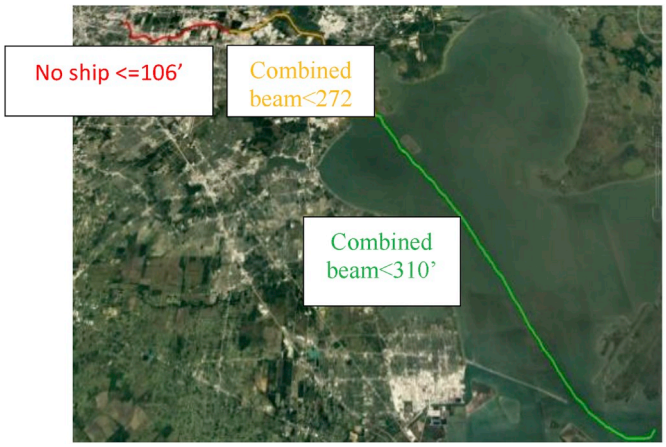


Fig. 6. Beam restrictions along the HSC.

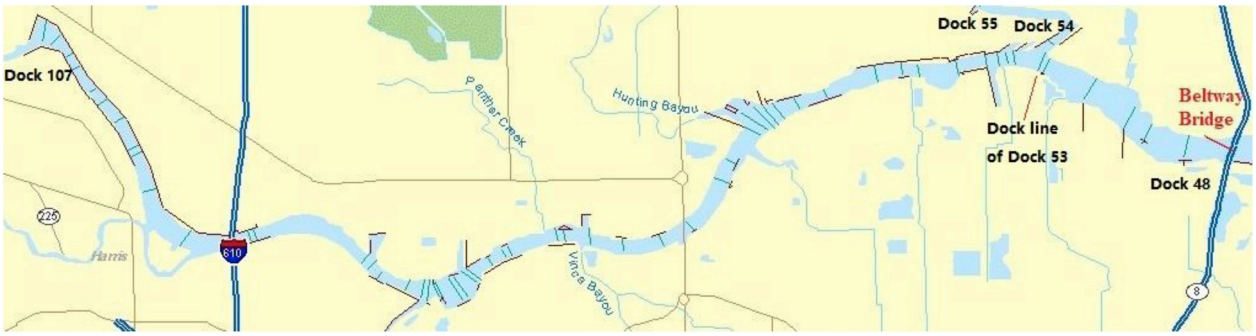


Fig. 7. Section of Houston Ship Channel above Beltway-8 Bridge.
Note: docks are marked along the channel, and each dock's dock line is also marked as a line across the channel.

months of 2017 shows that no tanker and cargo used these two docks.

This section of the channel is narrow. From the Beltway-8 Bridge until dock 107 (end of the channel), there is a one-way traffic restriction. The restriction starts from the Beltway-8 Bridge, and the first dock above the Bridge is 48. According to this restriction, there cannot be any meeting of ships greater than 106 feet (32 m). Below the Bridge, there is combined beam restrictions. Fig. 6 shows three different restrictions on the width (beam) of vessels for different narrow sections of the HSC.

Large ships, such as tankers and cargos in this section of the HSC are significantly impacted by wide-body vessels (with beam of 106 ft or more, i.e., 32 m): a wide-body vessel can occupy the whole width of the channel when moving in this section of the HSC. However, as mentioned above, most tugs are not impacted as their drafts are small enough, so that they can be moored anywhere near the bank. Therefore, the algorithm described in Section 2.1.2 may not well capture a tug's "real" destination, as well as its arrival and departure times as it stops frequently in the channel, and not necessarily at a dock.

For these reasons, only big ships' travel times were studied. The big ships in this study are considered as those having draft greater than 16 feet (4.87 m). In the HSC, only two types of big ships are considered: tankers and cargos. The number of passenger ship can be neglectable. Fig. 8(a) reports the monthly distributions of different vessel types in the HSC recorded in 11 months of 2017. It is seen that except tugs, tankers and cargos were the major vessels. Therefore, the travel times of only tankers and cargos were studied in this paper.

By using the algorithm described in Section 2.1.1, the trips made by tankers and cargos in the section above the Bridge in each month (except February) of 2017 were investigated, and the results were reported in Fig. 8(b). It is seen that tankers play an important role in the HSC: in some months of 2017, tankers made nearly 200 trips to the docks located above the Bridge, implying that on average there were over 5 tankers in each day moving in this section of the HSC.

3.2. Distribution of travel time/traversal time

Based on the AIS data of 11 months of 2017, we firstly calculated destination dock, departure and arrival times of each tanker and cargo according to the algorithm in Section 2.1.2. Then, based on the revealed trips, the draft of each trip was determined from the raw AIS dataset. Finally, based on the revealed trips, focusing on *outbound trips ONLY* (the procedure for inbound trips is the same), i.e., trips heading to the Gulf, we calculated the travel time of each tanker and cargo, respectively, from its destination dock to the Bridge (see Fig. 7), based on the algorithm described in Section 2.2.5.

A big ship's starting process from the stationery state is time-consuming, for increasing its speed and changing the heading direction (a U-turn may be needed). Therefore, another type of travel time – *traversal time* was considered in this paper. A vessel's traversal time from one dock line to another is the time spent by this vessel in moving from that dock line to another under its normal navigation status. Since only outbound trips were considered, for convenience, if vessel *i*started from dock j, its traversal time was calculated from dock line $j-3, j-4, \cdots, 48$, respectively, to the Bridge. Note that as shown in Fig. 7, the next of dock line 56 is dock line 53, and dock line 48 is the last one above the Bridge. For simplicity, these two types of travel times are called (1) *travel time* (from the destination dock) and (2) *traversal time*, respectively in the following text.

For each dock (or dock line), a series of travel times (or traversal times) were calculated, respectively. Due to the data error and some special navigation status (e.g., U-turn), some travel time data are particularly large. Since we focus on the regular navigation status, to avoid generating some bias results, it is necessary to remove these large travel/traversal time data. Note that unlike traffic in highway, vessel navigation in a waterway is fully controlled by a pilot and the dispatch of a vessel follows the plan. A record is regarded to be an outlier if it is less than the value of the first quartile minus the 1.5 times IQR or greater than the value of the third quartile plus the 1.5 times IQR, where IQR (interquartile range) is the distance between the third and first quartiles of the data (Montgomery et al., 2010). Actually, it is found that the number of outliers is quite small: on average only 5.97% (for the travel time dataset) and 1.88% (for the traversal time dataset) are identified as outliers. More travel time data were found to be outliers because a vessel may make a U-turn when departing from the dock (the data error is also more likely to occur when a vessel starts from the stationery state).

The information of travel time and traversal time in a narrow channel is very useful for helping better managing vessel traffic in a narrow channel. For example, the reliability of travel time for different type of vessels can be obtained; the delay caused by wide-body vessels can be estimated; moreover, such information can help build a vessel dispatch scheme, so as to minimize the travel delay. This paper, however, aims to study the properties of these two types of travel time only.

Apparently, the number of traversal time data is much larger than the number of travel time data for those dock lines close to the Bridge (as we only looked at outbound trips). For the purpose of statistical analysis, at one dock (or dock line), if the sample size (the number of travel times or traversal times of tankers and cargos) was less than 20 (for travel time data) or 100 (for traversal time data), this dock was skipped for analysis. The statistical summary of the two types of travel times were reported in Table A1 in Appendix A at the end of this paper.

As expected (a large amount of time is required in the starting process), it is seen that the mean travel time from one dock to the Bridge is much larger than the corresponding traversal time from the dock line of the same dock to the Bridge. To better understand the properties of these two types of travel time, their distributions were estimated using two methods: the Kolmogorov-Smirnov (K–S) and chi-square goodness-of-fit tests (via MATLAB), respectively.

Under two types of tests, the P-values under four most commonly used distributions: gamma, lognormal, normal and Weibull distributions were calculated, respectively, with or without outliers. Based on the P-values reported by each test, the number of datasets estimated to follow

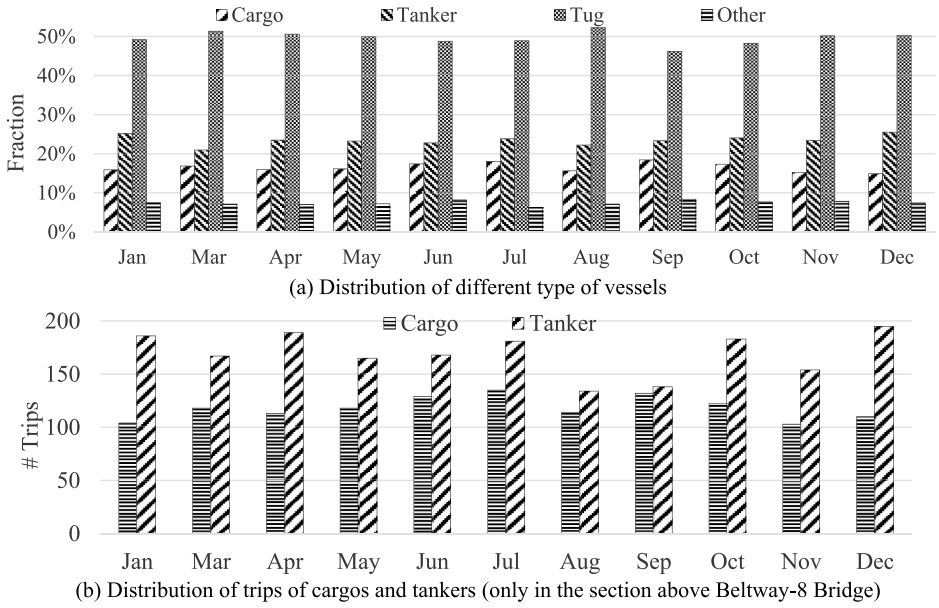


Fig. 8. Distributions of vessel types and trips in 11 months of 2017.
(a) Distribution of different type of vessels. (b) Distribution of trips of cargos and tankers (only in the section above Beltway-8 Bridge).

Table 1
Number of datasets estimated to follow some distributions under two tests ($\alpha = 0.05$).

Distribution	Traversal time (47 datasets)				Travel time (32 datasets)			
	Outliers removed		Outliers not removed		Outliers removed		Outliers not removed	
	K–S	Chi-square	K–S	Chi-square	K–S	Chi-square	K–S	Chi-square
Gamma	47	31	15	13	31	22	14	9
Lognormal	47	36	15	16	31	23	16	9
Normal	27	5	14	4	31	20	14	8
Weibull	2	0	0	0	30	15	11	5

Note: "K–S" refers to the Kolmogorov-Smirnov test; "Chi-square" refers to the chi-square goodness-of-fit test. A dataset is estimated to follow a type of distribution under a test if its P-value is greater than α, by default, use 5%.

each type of distribution was reported in Table 1. Note that the default significance level is 0.05: if the P-value of a test is no more than 0.05, the null hypothesis that the samples are from a desired distributed population, can be rejected; otherwise, we cannot reject the hypothesis that the data follow a certain type of distribution. From Table 1, it is seen that (1) outliers have significant impact on type of distribution estimate; (2) the results under two tests are close though the chi-square test seems to be stricter; and (3) few seem to be Weibull distributed (so in the following, we focus on the first three types of distributions). Through Monte-Carlo tests, Wang (2009) concluded that given the significance level of 0.05, the K–S test seems to be more effective than the chi-square test because the former exhibits smaller type I error and larger power except the case that the null hypothesis has a known mean. Therefore, we just reported the P-values of the K–S tests in Appendix A.1. Other methods were also employed in the literature. For example, Kang et al. (2018, 2019) employed the normal probability plot to check the normality of travel time data in the Strait of Singapore.

In the literature, traversal times of vehicles in an uncongested highway section are sometimes assumed to be normally distributed (Lo et al., 2006; Shao et al., 2006); for vessel traffic in the channel, some studies also found that the traversal times are normally distributed (Xiao et al., 2015; Kang et al., 2018, 2019). However, interestingly, it is seen from Table A1 that even though most travel times from destination docks to the Bridge are normally distributed (P-values > 0.05), most traversal times are NOT.

Instead, it is seen that the traversal time from each dock line to the Bridge is more likely to follow a *lognormal distribution* (P-values > 0.05), even excluding outliers. Note that for traversal time datasets, on average 85% outliers are on the right tail. It implies that the traversal times of different types and sizes of vessels may have different distributions. This postulation will be studied in the next section.

As to the travel time from destination docks to the Bridge, since the time spent in starting process is large, it entails that such time for starting process is similar across different types and sizes of ships.

3.3. Factors impacting traversal time

To better understand why traversal time is not normally distributed, we aimed to investigate which factors could impact the traversal time in the HSC. The following 5 factors were considered:

(a) Type of a vessel: tanker or cargo
(b) Width of a vessel: a vessel is wide-body (WB) if its width is 32 m (106 ft) or larger; non-WB otherwise. Such information was provided by the pilots of the HSC.
(c) Draft of a vessel: a vessel is in draft class 1 (D1) if its draft is at most 9 m; in draft class 3 (D3) if its draft is more than 11 m; and in draft class 2 (D2) otherwise.
(d) Time of day: daytime if the time arriving at the Bridge is later than 6:00 but earlier than 19:30 in May, June, July and August, or later than 6:30 but earlier than 18:30 in March, April, September and October (note that it is under daylight saving time), or later than 7:30 but earlier than 17:00 in January, November and December (no AIS data in February); nighttime otherwise.

(e) Month: January, March, April, …December (no AIS data in February).

Note that among these 5 factors, two are temporal factors – time of day and month; two are related with vessel itself – type and width); and one is related with both vessel and voyage – draft.

To investigate the impacts of these factors, a series of the analysis of variance (ANOVA) were conducted. Note that ANOVA requires the sample under each category to be normally distributed. Therefore, the K–S test was first used to test the normality of the samples in the category of each factor. For example, the traversal time of wide-body vessels (tankers and cargos) from dock line 59 to the Bridge was found not normally distributed, so the ANOVA for the factor of vessel width could not be conducted for this dataset (traversal time from dock line 59).

Under each factor, the sample size varies. For example, there were 819 wide-body tankers but only 340 non-wide-body tankers (for all 11 months of 2017). Therefore, two-way ANOVA with unequal sample size (considering one temporal factor and one vessel feature) were used in MATLAB, and the P-value in each test was reported in Table A2 (in Appendix A). The P-values show that both temporal factors (time-of-day and month) did NOT have significant impact on traversal time from each dock line to the Bridge. This finding is somewhat different from our experience on highway traffic which have peak hours and off-peak hours and link travel time distributions are usually time-dependent (Wu, 2013). The reason probably is that the traffic of big vessels in a narrow and busy channel, such as the HSC, is fully monitored by the port authority and big vessels are fully controlled by pilots assigned by the port authority. On the other hand, the factors of vessel features, especially width and draft were found to significantly impact the traversal time from all dock lines, and vessel type had significant impact on the traversal time from most dock lines.

Therefore, we conclude that the traversal time is largely impacted by vessel features such as its width, draft and type. To better understand such impact, the comparison across different vessel features were further investigated.

First, the traversal times of wide-body-and-large-draft (draft class 3, denoted as WB-D3) tankers and cargos were compared with those of non-wide-body-and-small-draft (draft class 1, denoted as NWB-D1) ones from each dock line (if the sample size was large enough). It was found that from all dock lines, the former is always greater than the latter, i.e., larger ships ran slower. Fig. 9(a) reports the distribution of the mean and median differences between WB-D3 and NWB-D1 ships, respectively. For example, from the same dock line to the Bridge, nearly 43% of WB-D3 cargos saw the mean traversal time 10% longer than NWB-D1 cargos; while almost 28% of WB-D3 tankers found the mean travel time 15% longer than NWB-D1 tankers. Such difference is significant. Further, Fig. 9(a) also shows that the difference between big and small tankers is larger than the difference between large and small cargos.

Then, the difference between tankers and cargos were investigated when they were both WB-D3 or NWB-D1. The distribution of the mean and median differences was reported in Fig. 9(b), respectively. For example, if both were big ships (WB-D3), as to mean traversal time, nearly 20% of tankers saw 8% longer than cargos; however, if both were small (NWB-D1), over 50% of tankers saw only 2% more than cargos. It well implies that big tankers moved slower than big cargos, but the difference between small tankers and cargos were not clear.

3.4. Spatial correlation

In a highway network, if one link suffers from traffic congestions, its adjacent upstream[2] link is very likely to be congested too. This is called spatial correlation. To the best knowledge of the authors, such spatial correlation was not studied for vessel traffic in waterways. For narrow and busy channels, such correlation is likely to impact travel time of vessels significantly. The revealed traversal times from one dock lines to another opens the door for answering this question.

In last section, it was already found that vessel features have significant impact on traversal times. For this reason, the spatial correlation was studied in different categories in terms of vessel type, width, and draft, such as WB-D3 tankers or cargos, NWB-D1 tankers or cargos, etc.

Considering the sample size (as at least 100 samples were preferred from one dock line), we picked a section of the channel from dock line 70 to the Bridge, so totally 21 links were covered (note that the next of dock line 56 is dock line 53). For one vessel, its traversal time from one dock line to the next dock line was calculated as link traversal times (see Eq. (29)). As we already calculated the traversal time from one dock line to the Bridge, it would be easy to get such link traversal time of one vessel. One observation of a series of link traversal times came from one vessel when it passed through the channel from dock line 70 to the Bridge. Since wide-body tankers/cargos in draft class 1 (WB-D1), and non-wide-body tankers/cars in draft class 3 (NWB-D3) were unusual (small sample size), only 4 categories of vessels were considered here: WB-D3, WB-D2, NWB-D2, NWB-D1 for tankers and cargos, respectively, so 8 correlation coefficient matrixes were calculated, respectively (4 types of tankers and cargos, respectively). Each matrix has a size of 20×20, as there are 21 links. Fig. 10 summarizes such correlation coefficient matrixes based on the distance between two links: two links are adjacent links if they are just separated by 1 dock line; and since there are totally 21 dock lines, the farthest two links are separated by 20 dock lines.

From Fig. 10, we can clearly see that the correlation coefficient between two adjacent links are close to 1 in most cases (only two cases of cargos where the coefficient is below 0.7, which can be regarded as outliers). Also, when two links are separated by more dock lines, the correlation becomes weaker. However, even if two links are separated by 10 dock lines, the correlation coefficients are still about 0.5. These findings well indicate that the link traversal time between links in the HSC are highly positively correlated.

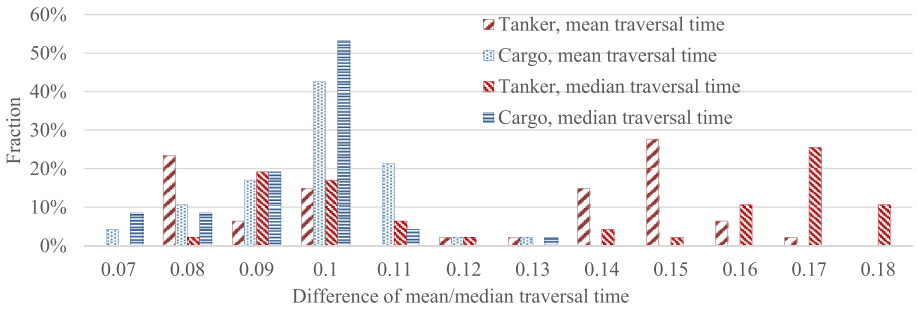
4. Discussion and conclusions

This paper proposes a novel AIS-data-based model and the corresponding algorithms to estimate the travel time of vessels between any two points in narrow and busy channel. The model has two components: (1) to determine the destination dock of a vessel, as well as its arrival and departure times from this dock; and (2) to determine the timestamp when a vessel passed a landmark of interest in the channel. This model can help us easily determine the travel time of a vessel from one point to another in the channel, including the travel time from its destination dock to a specific landmark (or vice versa) in the channel.

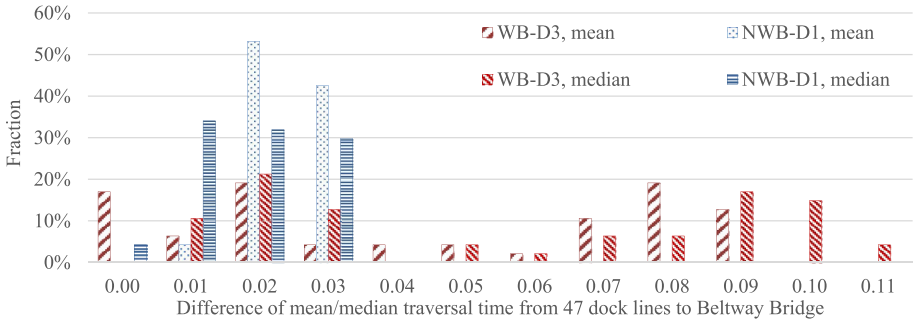
Based on the proposed model, this paper employed the AIS data of 11 months in 2017, collected in the HSC to study the property of travel times of tankers and cargos, focusing on the section of channel above the Beltway-8 Bridge. This narrow section has lots of docks, making traffic condition in this section critical. Considering outbound trips only, two types of travel times were defined: (1) the travel time from a vessel's destination dock to the Bridge (starting from the stationery state), and (2) the traversal time from one dock line to the Bridge when a vessel went through this distance under the normal navigation status. Apparently, from the same dock, the travel time is much longer than the traversal time as the starting process is time-consuming. The findings of travel time properties were summarized as follows.

(f) If considering all tankers and cargos together (with various size and draft), the traversal time from a dock line to the Bridge is likely to be lognormally distributed, not normally distributed. On the other hand, in most cases, the travel time from a destination dock to the Bridge is normally distributed.

[2] In the HSC, the upstream direction refers to the direction from the Gulf of Mexico to the Port of Houston.

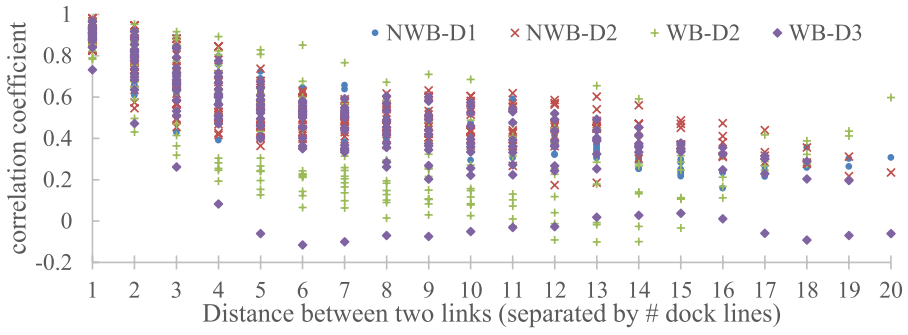


(a) wide-body-large-draft tankers/cargos v.s. non-wide-body-small-draft tankers/cargos

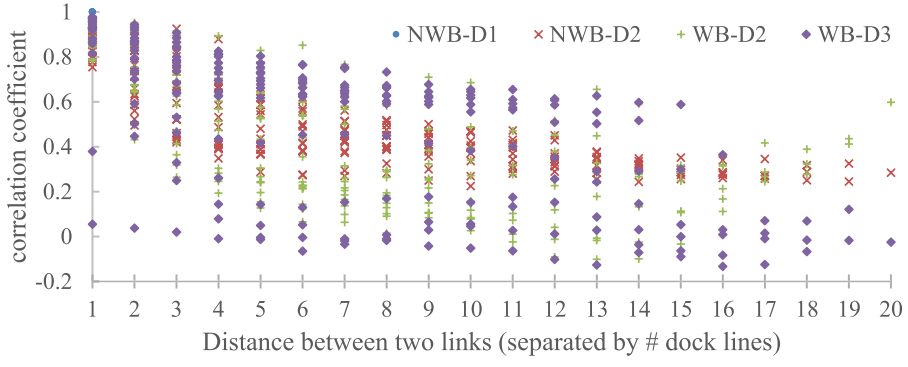


(b) tankers v.s. cargo (in the same width and draft class)

Fig. 9. Distribution of the mean/median traversal time gaps between ships in different size and type. (a) wide-body-large-draft tankers/cargos v. s. non-wide-body-small-draft tankers/cargos. (b) tankers v. s. cargo (in the same width and draft class).
Note: (1) WB-D3 refers to wide-body tankers or cargos (beam $\geq$ 32m, draft $\geq$ 11m); NWB-D1 refers to non-wide-body tankers or cargos (draft $\leq$ 9m). (2) the difference of xv.s. $y$is calculated as $(x - y)/y$, so it is a unitless value.



(a) Tanker



(b) Cargo

Fig. 10. Correlation coefficients for two links separated by a certain number of dock lines. (a) Tanker. (b) Cargo.
Note: If two links are separated by 1 dock line, they are adjacent links. WB-D3 or WB-D2 refers to wide-body (WB) tankers or cargos in draft class 3 (D3) or draft class 2 (D2); NWB-D1 or NWB-D2 refers to non-wide-body (NWB) tankers or cargos in draft class 1 (D1) or draft class 2 (D2). The number of WB-D1 and NWB-D3 tankers or cargos are small, so their correlation coefficients were not studied.

(g) Time-of-day (daytime or nighttime) and month have insignificant impact on the distribution of traversal time from each dock line to the Bridge; on the other hand, vessel width, draft and vessel types were found have significant impact, especially the width and draft.

(h) Larger ships moved slower, and the difference is not small. On average, the mean traversal time of wide-body (WB: $\geq$32 m) tankers in draft class 3 (D3: >11 m) was 12% longer than that of non-wide-body (NWB) tankers in draft class 1 (D1: $\leq$9 m); while for cargos, such difference was a little bit smaller: 9% on average. On the other hand, the difference on median traversal time was

even larger: WB-D3 tankers experienced averagely 13% longer median traversal time than NWB-D1 tankers; and while for cargos, such difference was also about 9% on average.

(i) Tankers generally moved slower than cargos, especially for larger ships. On average, the mean traversal time of WB-D3 tankers was 4% larger than WB-D3 cargos; while only 2% more if comparing between NWB-D1 tankers with NWB-D1 cargos. On the other hand, WB-D3 tankers saw averagely 5% longer median traversal time than WB-D3 cargos; but the median traversal times between NWB-D1 tankers and cargos were very close: only 1% more on average.

(j) The traversal times between two adjacent links in the HSC were significantly positively correlated (the correlation coefficient is close to 1), implying if one link suffers from congestion, the upstream adjacent link is very likely to see congestion too. Such correlation decreases when one link is farther away from another. But even if two links are separated by 10 dock lines, the correlation coefficient is still about 0.5.

These findings are well consistent with the real experience. It shows that the proposed models and solution algorithms produce reasonable travel time estimates. By building an inventory of travel times between any two points in the channel, we will know the travel time distributions, and it opens the door for many studies, including the estimate of the vessel delay caused by congestion in the channel, the optimization of dispatching vessels in the channel, etc. The research team has started the research on these topics. Even if the aforementioned models and solution algorithms were developed based on the research needs of the HSC, they can be easily applied to any other ship channels.

Statement of author contributions

Xing Wu developed the models for estimating the arrival and departure times of vessels at their destination docks, cleaned the AIS data of 2017, developed the codes based on the model, and calculated the results based on the AIS data of 2017. He drafted the initial version of the paper. In the first-round revision, he calculated the new results based on the suggestions of reviewers. He finalized the draft of the Response to the Reviewers, and the draft of the revised manuscript in the both first- and second-round revisions.

Uttara Roy helped identified the locations of docks along the Houston Ship Channel (HSC), and built the shape files of these docks, as well as the shape files of dock lines used for determining the travel time. When preparing the revised draft, she helped update the literature review to reflect the existing studies on travel time. She also corrected all grammar errors and typos, and prepared the initial draft of the Responses to the Reviewers in the first-round revision.

Maryam Hamidi provided the information of the HSC, such as the location of docks. She also addressed the significance of the study. When preparing the revised draft in the first round, she helped answer the questions on the significance of the study on ships' travel times in a channel, and the possible applications in the maritime engineering, such as the applications related with the vessel traffic service (VTS) system, and the current research needs of the HSC. She made significant contributions in revising the paper and preparing the Response to Reviewers.

Brian Craig helped address the significance of this study, and offered the information of the HSC. When preparing the revised manuscript and the Response to Reviewers, he helped address the questions of the reviewers on the significance and motivation of the study. He also helped update the data of the HSC. He made significant contributions to the explanation of the motivation and significance of the study in the revised manuscripts.

Acknowledgement

This study is sponsored by the Center for Advances in Port Management of Lamar University (LU) with Projects 220879 and 220890. The Houston Pilots offered great help to provide us the valuable information on the features of the HSC. The authors would like to thank three anonymous reviewers for their valuable comments. The authors resume sole responsibilities for the content expressed.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oceaneng.2019.106790.

APPENDIX A

Table A.1
Statistical summary of traversal and travel times from departure dock to the Beltway-8 Bridge

Dock	Traversal time* from a dock line to the Bridge							Travel time† from departure dock to the Bridge						
	Num**	Std min.	Mean min.	Median min.	P-value of K–S Test			Num‡	Std min.	Mean min.	Median min.	P-value of K–S Test		
					Gamma	LN	Normal					Gamma	LN	Normal
48	1580	0.41	2.64	2.62	0.73	0.71	0.09	–	–	–	–	–	–	–
49	1481	0.73	5.14	5.10	0.93	0.52	0.17	–	–	–	–	–	–	–
50	1307	0.90	6.56	6.50	0.49	0.17	0.26	–	–	–	–	–	–	–
51	1194	1.24	9.47	9.40	0.88	0.66	0.14	88	5.85	27.20	25.72	0.30	0.48	0.10
52	1193	1.54	12.43	12.30	0.64	0.88	0.06	–	–	–	–	–	–	–
53	1167	1.58	12.80	12.67	0.50	0.82	0.04•	45	4.38	38.26	38.09	0.98	0.94	0.97
54	–	–	–	–	–	–	–	42	6.34	28.88	28.12	0.88	0.93	0.67
56	1169	1.73	14.39	14.21	0.27	0.65	0.02	30	10.04	33.08	29.80	0.34	0.47	0.18
57	1187	1.78	14.85	14.66	0.32	0.73	0.02	20	6.22	26.74	25.84	0.98	0.97	0.91
58	1227	1.80	15.43	15.27	0.22	0.57	0.01	36	5.17	27.53	27.51	0.85	0.86	0.93
59	1228	1.88	16.16	15.99	0.32	0.73	0.02	–	–	–	–	–	–	–
60	1197	1.98	17.84	17.67	0.45	0.85	0.05	26	3.31	25.42	26.06	0.75	0.70	0.88
61	1193	2.21	20.22	20.02	0.25	0.51	0.04	58	4.57	28.24	27.56	0.25	0.34	0.12
62	1136	2.27	21.14	20.94	0.27	0.54	0.03	–	–	–	–	–	–	–
63	1087	2.40	22.66	22.38	0.16	0.41	0.02	94	4.57	35.66	35.23	0.94	0.96	0.72

(continued on next page)

Table A.1 (*continued*)

Dock	Traversal time* from a dock line to the Bridge							Travel time† from departure dock to the Bridge						
	Num**	Std min.	Mean min.	Median min.	P-value of K–S Test			Num‡	Std min.	Mean min.	Median min.	P-value of K–S Test		
					Gamma	LN	Normal					Gamma	LN	Normal
64	1084	2.73	26.14	25.86	0.16	0.37	0.01	–	–	–	–	–	–	–
65	1069	2.95	28.12	27.79	0.09	0.18	0.01	48	5.71	42.51	41.75	0.84	0.91	0.74
66	1019	3.00	29.37	29.04	0.07	0.14	0.01	47	7.48	46.26	45.52	0.96	0.98	0.92
67	997	2.92	29.84	29.57	0.19	0.34	0.05	–	–	–	–	–	–	–
68	997	2.96	30.21	29.93	0.17	0.31	0.05	–	–	–	–	–	–	–
69	996	3.05	30.78	30.47	0.14	0.29	0.02	39	13.77	45.52	42.28	0.50	0.70	0.16
70	994	3.12	31.43	31.12	0.11	0.26	0.02	25	5.19	48.57	48.45	0.98	0.99	0.97
71	975	3.29	33.94	33.53	0.05	0.13	0.01	–	–	–	–	–	–	–
72	920	3.25	35.09	34.66	0.05	0.13	0.01	–	–	–	–	–	–	–
73	864	3.43	38.73	38.36	0.08	0.19	0.01	–	–	–	–	–	–	–
74	852	3.65	40.83	40.38	0.07	0.15	0.01	26	19.01	68.45	68.65	0.43	0.47	0.49
75	827	3.72	43.15	42.77	0.11	0.23	0.02	54	5.57	61.62	60.87	0.90	0.87	0.93
76	785	3.81	45.21	44.77	0.22	0.39	0.05	47	6.78	65.48	65.12	0.76	0.69	0.80
77	784	3.92	46.75	46.38	0.26	0.44	0.06	–	–	–	–	–	–	–
78	776	3.97	47.25	46.85	0.37	0.59	0.10	25	5.38	67.02	66.03	0.90	0.86	0.89
79	770	4.21	50.34	50.04	0.36	0.58	0.10	37	9.91	63.53	63.04	0.81	0.89	0.68
80	727	4.26	50.70	50.32	0.26	0.43	0.07	–	–	–	–	–	–	–
81	723	4.32	51.71	51.34	0.31	0.50	0.09	–	–	–	–	–	–	–
82	693	4.12	51.49	51.18	0.59	0.79	0.25	–	–	–	–	–	–	–
83	689	4.23	52.85	52.56	0.64	0.83	0.25	36	6.89	66.93	65.90	0.48	0.55	0.37
84	666	4.23	53.15	52.88	0.72	0.88	0.34	–	–	–	–	–	–	–
85	660	4.33	54.01	53.73	0.85	0.98	0.41	36	11.48	78.58	76.04	0.73	0.82	0.61
86	656	4.49	54.50	54.24	0.71	0.89	0.35	–	–	–	–	–	–	–
87	624	4.70	55.76	55.35	0.44	0.65	0.14	24	6.78	67.26	66.81	0.89	0.91	0.78
88	542	4.49	55.98	55.65	0.58	0.76	0.27	–	–	–	–	–	–	–
89	542	4.51	56.45	56.15	0.66	0.82	0.32	–	–	–	–	–	–	–
90	543	4.55	56.87	56.55	0.65	0.83	0.29	45	7.58	66.91	66.56	0.90	0.92	0.92
91	526	4.59	57.80	57.49	0.53	0.71	0.23	81	9.26	71.46	70.75	0.99	1.00	0.85
92	433	4.74	59.99	59.80	0.73	0.87	0.40	–	–	–	–	–	–	–
93	398	4.75	60.91	60.80	0.86	0.96	0.56	–	–	–	–	–	–	–
94	348	4.83	61.30	61.30	0.96	0.97	0.87	–	–	–	–	–	–	–
95	242	5.42	68.77	68.62	0.77	0.88	0.54	92	12.38	80.62	75.69	0.02	0.03	0.01
96	234	5.59	69.17	68.89	0.67	0.80	0.44	40	15.89	88.98	85.74	0.50	0.60	0.40
97	–	–	–	–	–	–	–	44	12.03	97.30	98.61	0.53	0.44	0.72
98	–	–	–	–	–	–	–	102	8.59	96.85	96.69	0.87	0.89	0.77
100	–	–	–	–	–	–	–	60	13.07	98.36	97.08	0.99	1.00	0.93
101	–	–	–	–	–	–	–	44	7.97	91.15	90.81	0.98	0.99	0.91
102	–	–	–	–	–	–	–	26	17.53	103.58	99.74	0.83	0.86	0.73
103	–	–	–	–	–	–	–	39	10.24	97.45	95.85	0.68	0.73	0.63
104	–	–	–	–	–	–	–	41	10.95	98.25	98.92	0.57	0.62	0.58

*__Note__: the traversal time refers to the travel time of a vessel from a dock line to a landmark (Beltway-8 Bridge) under the normal navigation status. If a vessel departed dock *j*(as its destination dock) and moves outbound, we record its traversal time from the dock line of docks *j*-3, *j*-4, …, as when it just started to move, the speeds were low.

†: the travel time from a departure dock to a landmark (Beltway-8 Bridge) records the travel time of a vessel from its departure time from its dock (speed = 0) to the arrival time at a landmark (the time when it crossed this landmark).

**: all outliers have been removed before the analysis; and if less than 100 vessels passed at one dock line, the analysis of the traversal time from this dock-line to the Bridge is skipped.

‡: if the number of outbound trips departing from one dock is less than 20, the analysis of the travel time from this dock to the Bridge is skipped.

•: since 0.05 is used as the significance level, if the P-value is less than 0.05. it indicates the null hypothesis can be rejected.

Table A.2
P-Values of two-way ANOVA for traversal times from one dock line to Beltway-8 Bridge

From*	Draft**-Month		Type-Month		WB-Month		Draft-Daytime		Type-Daytime		WB-Daytime	
	Draft	Month	Type	Month	WB†	Month	Daytime	Draft	Daytime	Type	Daytime	WB
48	0.000•	0.299	0.000	0.590	0.000	0.276	0.708	0.000	0.543	0.000	0.535	0.000
49	0.000	0.312	0.000	0.695	0.000	0.523	0.661	0.000	0.554	0.000	0.533	0.000
50	0.000	0.337	0.000	0.765	0.000	0.601	0.252	0.000	0.475	0.000	0.270	0.000
51	0.000	0.079	0.000	0.284	0.000	0.220	0.519	0.000	0.764	0.000	0.564	0.000
52	0.000	0.068	0.000	0.227	0.000	0.218	0.466	0.000	0.684	0.000	0.487	0.000
53	0.000	0.103	0.000	0.268	0.000	0.250	–‡	–	–	–	–	–
56	0.000	0.133	0.000	0.444	0.000	0.226	–	–	–	–	–	–
57	0.000	0.117	0.000	0.396	0.000	0.233	–	–	–	–	–	–
58	0.000	0.201	0.000	0.662	0.000	0.292	–	–	–	–	–	–
59	0.000	0.137	0.001	0.569	–	–	–	–	–	–	–	–
60	0.000	0.159	0.000	0.689	0.000	0.222	0.509	0.000	0.711	0.000	0.344	0.000
61	0.000	0.120	0.000	0.589	0.000	0.163	0.193	0.000	0.344	0.000	0.132	0.000
62	–‡	–	0.000	0.565	0.000	0.115	–	–	0.311	0.000	0.110	0.000
63	–	–	0.000	0.521	0.000	0.093	–	–	0.210	0.000	0.070	0.000

(*continued on next page*)

Table A.2 (*continued*)

From*	Draft**-Month		Type-Month		WB-Month		Draft-Daytime		Type-Daytime		WB-Daytime	
	Draft	Month	Type	Month	WB†	Month	Daytime	Draft	Daytime	Type	Daytime	WB
64	–	–	0.000	0.454	0.000	0.057	–	–	0.208	0.000	0.075	0.000
65	–	–	0.000	0.480	–	–	–	–	–	–	–	–
66	–	–	0.000	0.520	–	–	–	–	–	–	–	–
67	–	–	0.000	0.596	0.000	0.287	–	–	0.263	0.000	0.195	0.000
68	–	–	0.000	0.626	0.000	0.300	–	–	0.248	0.000	0.171	0.000
69	–	–	0.000	0.664	0.000	0.404	–	–	0.177	0.000	0.127	0.000
70	–	–	0.000	0.613	0.000	0.340	–	–	–	–	–	–
71	0.000	0.172	–	–	–	–	–	–	–	–	–	–
72	0.000	0.172	–	–	–	–	–	–	–	–	–	–
73	0.000	0.100	–	–	–	–	–	–	–	–	–	–
74	0.000	0.099	–	–	–	–	–	–	–	–	–	–
75	0.000	0.500	0.750	0.351	–	–	–	–	–	–	–	–
76	0.000	0.527	0.550	0.511	–	–	0.729	0.000	0.899	0.268	–	–
77	0.000	0.578	0.515	0.676	–	–	0.664	0.000	0.798	0.216	–	–
78	0.000	0.612	0.636	0.627	–	–	0.694	0.000	0.855	0.313	–	–
79	0.000	0.740	0.487	0.743	0.000	0.930	0.710	0.000	0.875	0.180	0.812	0.000
80	0.000	0.701	0.365	0.690	–	–	0.780	0.000	0.857	0.141	–	–
81	0.000	0.754	0.215	0.780	0.000	0.880	0.808	0.000	0.831	0.067	0.891	0.000
82	0.000	0.598	0.715	0.667	0.000	0.817	0.491	0.000	0.543	0.202	0.478	0.000
83	0.000	0.634	0.964	0.594	0.000	0.845	0.375	0.000	0.547	0.371	0.393	0.000
84	0.000	0.602	0.375	0.434	0.000	0.834	0.615	0.000	0.765	0.094	0.644	0.000
85	–	–	0.381	0.513	0.000	0.826	–	–	0.655	0.089	0.491	0.000
86	–	–	0.453	0.620	0.000	0.768	–	–	0.641	0.055	0.420	0.000
87	0.000	0.888	0.023	0.590	0.000	0.857	0.300	0.000	0.648	0.002	0.608	0.000
88	0.000	0.348	0.005	0.196	0.000	0.451	0.529	0.000	0.595	0.000	0.710	0.000
89	0.000	0.227	0.006	0.164	0.000	0.392	0.371	0.000	0.421	0.000	0.585	0.000
90	0.000	0.220	0.005	0.158	0.000	0.376	0.350	0.000	0.404	0.000	0.567	0.000
91	0.000	0.205	0.004	0.195	0.000	0.337	0.298	0.000	0.364	0.000	0.470	0.000
92	0.000	0.328	0.142	0.409	0.000	0.570	0.404	0.000	0.961	0.008	0.892	0.000
93	0.000	0.242	0.129	0.323	0.000	0.431	0.577	0.000	0.856	0.006	0.916	0.000
94	0.000	0.133	0.236	0.348	0.010	0.495	0.323	0.000	0.784	0.017	0.610	0.001
95	0.001	0.891	0.357	0.904	–	–	0.604	0.000	0.596	0.034	–	–
96	0.001	0.926	0.323	0.927	–	–	0.793	0.000	0.553	0.036	–	–

*__Note__: travel time from this dock to the Bridge in the HSC.

the draft is categorized into 3 classes: **D1: draft no more than 9 m, **D2:** draft no more than 11 m but larger than 9 m, **D3:** draft larger than 11 m.

†: wide-body vessel (with beam no less than 32 m) v.s. non-wide-body vessel.

‡if the data in this category is found not normally distributed, the two-way ANOVA is not conducted because ANOVA requires the data to be tested must follow normal distributions.

•: since 0.05 is used as the significance level, if the P-value is less than 0.05. it indicates the null hypothesis can be rejected.

APPENDIX B. LIST OF NOTATIONS

Information of Vessel	
t_i	Timestamp of vessel i, based on available AIS data of vessel i
s_i^t	Speed over ground of vessel $i$at timestamp t_i
c_i^t	Course over ground of vessel $i$at timestamp t_i
x_i^t, y_i^t	Longitude and latitude of vessel $i$at timestamp t_i, respectively
n_v	Number of unique vessels (based on vessel's ID) in the given set of AIS data
n_i	Number of trips vessel $i$in the section of channel of interest
$\max t_i$	Latest timestamp of vessel $i$in the given set of AIS data
$\min t_i$	Earliest timestamp of vessel $i$in the given set of AIS data

Information of Dock	
θ_j	Angle of the dock line of dock j(in the clockwise direction from the true north)
n_d	Number of docks in the channel of interest

Spatial Relationship of Vessel $i$and Dock j	
δ_{ij}	Indicator: $\delta_{ij} = 1$if vessel $i$has been moored at dock j; $\delta_{ij} = 0$otherwise
n_{ij}	Number of visits of vessel $i$at dock j
λ_i^k, λ_{ij}^k	Arrival time of vessel i(at dock j, if specified) in the k^{th}visit (dock $j$may not be specified)
μ_i^k, μ_{ij}^k	Departure time of vessel i(from dock j, if specified) in the k^{th}visit (dock $j$may not be specified)
Γ_{ij}	Set of vessel i's AIS points within the buffer of dock j(with radius of 150 ft, about 46 m)
Π_{ij}	Set of vessel i's AIS points within the buffer of the dock line of dock j(with radius of 300 ft)
T_{ij}	Set of the timestamps of vessel i's AIS points in Γ_{ij}or Π_{ij}
$\overline{T}_{ij}$	Set of the timestamps of vessel i's AIS points outside Γ_{ij}or Π_{ij}
ζ_{ij}^k	Earliest timestamp of vessel $i$after it leaves the buffer of dock $j$after the k^{th}visit
φ_{ij}^t	Direction from the true north (N) to the line from the centroid of dock j's dock line to the AIS point of vessel $i$at timestamp t(in the clockwise direction)

(*continued on next page*)

(*continued*)

Information of Vessel	
d_{ij}^{t}	Vessel i's distance to dock line $j$at timestamp t
$\min\_d_{ij}^{k}$	Vessel i's minimum distance to dock line j, based on the AIS points in subset Π_{ij}, considering the AIS points from its k^{th}visit only
$\text{dmin}\_t_{ij}^{k}$	Timestamp of vessel i's AIS point that is closet to dock j's dock line, considering the AIS points from vessel i's k^{th}visit only
$\text{nl}\_t_{ij}^{k}$	For vessel i's AIS points in its k^{th}visit only, timestamp of the AIS point next or last to the AIS point that is closest to dock j's dock line (this closest AIS point is from vessel $i$in its k^{th}visit only)
$\text{nl}\_d_{ij}^{k}$	For vessel i's AIS points in its k^{th}visit only, the distance from the AIS point with timestamp $\text{nl}_t_{ij}^{k}$to dock j's dock line
τ_{ij}^{k}	Timestamp when vessel $i$crossed the dock line of dock $j$in vessel i's k^{th}visit
tt_{jj+1}^{ik}	In its k^{th}visit, vessel i's travel time from dock line $j$to dock line $j+1$

References

American Association of Port Authorities (Aapa), 2015. U.S Port Ranking by Cargo Volume.

Asamer, J., Prandtstetter, M., 2014. Estimating ship travel times on inland waterways. In: Transportation Research Board 93rd Annual Meeting. National Research Council, Washington DC.

Chawla, S., Zheng, Y., Hu, J., 2012. Inferring the root cause in road traffic anomalies. Proc. of IEEE ICDM 2012.

Choe, T., Skabardonis, A., Varaiya, P., 2002. Freeway performance measurement system (PeMS): an operational analysis tool. In: Transportation Research Board 81st Annual Meeting. National Research Council, Washington DC.

DiJoseph, P., Mitchell, K., 2015. Estimating vessel travel time statistics for inland waterways with automatic identification system data. Transportation Research Board 94th Annual Meeting. National Research Council, Washington DC.

El Esawey, M., Sayed, T., 2011. Travel time estimation in urban networks using limited probes data. Can. J. Civ. Eng. 38 (3), 305–318.

Goodwin, M., 1975. A statistical study of ship domains. J. Navig. 28, 328–344.

Horteborn, A., Rinsgsberg, J., Svanberg, M., Holm, H., 2019. A revisit of the definition of the ship domain based on AIS analysis. J. Navig. 72 (3), 777–794.

Huang, J., Nieh, C., Kuo, H., 2019. Risk assessment of ship maneuvering in an approaching channel based on AIS data. Ocean. Eng. 173, 399–414.

Jiyoun, Y., Elefteriadou, L., Lawphongpanich, S., 2008. Travel time estimation on a freeway using Discrete Time Markov chains. Transp. Res. Part B Methodol. 42 (4), 325–338.

Kaneria, A., Hamidi, M., Zhu, W., Craig, B., 2019. Traffic simulation of Houston Ship Channel for assessing the impact of waterway closures on vessel waiting time. J. Waterw. Port, Coast. Ocean Eng. 145 (4), 04019014.

Kang, L., Meng, Q., Liu, Q., 2018. Fundamental diagram of ship traffic in the Singapore Strait. Ocean. Eng. 147, 340–354.

Kang, L., Meng, Q., Zhou, C., Gao, S., 2019. How do ships pass through L-shaped turnings in the Singapore strait? Ocean. Eng. 182, 329–342.

Liu, J., Zhou, F., Li, Z., Wang, M., Liu, R., 2016. Dynamic ship domain models for capacity analysis of restricted water channels. J. Navig. 69, 481–503.

Lo, H., Luo, X., Siu, B., 2006. Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transport. Res. Part B 40 (9), 792–806.

Montewka, J., Hinz, T., Kujala, P., Matusiak, J., 2010. Probability modeling of vessel collisions. Reliab. Eng. Syst. Saf. 95, 573–589.

Montgomery, D., Runger, G., Hubele, N., 2010. Engineering Statistics, fifth ed. Wiley and Sons, Inc.

Mou, J.M., van der Tak, C., Lighteringen, H., 2010. Study on collision avoidance in busy waterways by using AIS data. Ocean. Eng. 37, 483–490.

Nie, Y., Wu, X., Nelson, P., Dillenburg, J., 2012. Providing reliable route guidance: a case study using Chicago data. Transport. Res. Part A 46, 403–419.

Port of Houston, 2019. Port of Houston (2019) overview. http://porthouston.com/about-us/. (Accessed 15 September 2019).

Qu, X., Meng, Q., Suyi, L., 2011. Ship collision risk assessment for the Singapore Strait. Accid. Anal. Prev. 43, 2030–2036.

Rahimikelarijani, B., Abedi, A., Hamidi, M., Cho, J., 2018. Simulation modeling of Houston Ship Channel vessel traffic for optimal closure scheduling. Simulat. Model. Pract. Theor. 80, 89–103.

Riley, I.L., Mitchell, K.N., DiJoseph, P.K., Whalin, R.W., Wang, F., 2015. Analysis of vessel travel times and river stages. In: Transportation Research Board 94th Annual Meeting. National Research Council, Washington DC.

Roy, U., Wu, X., 2019. AIS-data based vessel traffic's characteristics and travel behavior analysis: a case study at Houston Ship Channel. J. Ocean Technol. 14 (4), 58–74.

Shao, H., Meng, Q., Tam, M., 2006. Demand-driven traffic assignment problem based on travel time reliability. Transport. Res. Rec. 1985, 220–230.

Turner, S., 1996. Advanced techniques for travel time data collection.. In: Transportation Research Board 75th Annual Meeting. National Research Council, Washington DC.

Wang, H., 2009. Comparison of the goodness-of-fit tests: the Pearson chi-square and Kolmogorov-Smirnov tests. Journal of Econometric Management 6, 57–64. https://pdfs.semanticscholar.org/0c3a/736708f50897690875b6e578e62657df437b.pdf.

Wang, Y., Zheng, Y., Xue, Y., 2014. Travel time estimation of a path using sparse trajectories. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2014.

Weng, J., Meng, Q., Qu, X., 2014. Vessel collision frequency estimation in the Singapore Strait. J. Navig. 65, 1–15.

Wu, X., 2013. Finding reliable shortest paths in dynamic stochastic network. Transport. Res. Rec. 2334 (1), 80–90.

Wu, X., Nie, Y., 2013. Solving multi-class percentile user equilibrium traffic assignment problem: a computational study. Transport. Res. Rec. 2334, 75–83.

Wu, X., Mehta, A., Zaloom, V., Craig, B., 2016. Analysis of waterway transportation in Southeast Texas waterway based on AIS data. Ocean. Eng. 121, 196–209.

Wu, X., Rahman, A., Zaloom, V., 2018. Study of travel behavior of vessels in narrow waterways using AIS data – a case study in Sabine-Neches Waterways. Ocean. Eng. 147, 399–413.

Xiao, F., Ligteringen, H., van Gulijk, C., Ale, B., 2015. Comparison study on AIS data of ship traffic behavior. Ocean. Eng. 95, 84–93.

Yuan, N.J., Zheng, Y., Zhang, L., Xie, X., 2013. T-finder: a recommender system for finding passengers and vacant taxis. IEEE Trans. Knowl. Data Eng. 25 (10), 2390–2403.

Zaman, M., Wakabayashi, N., Khanfir, S., Maimum, T., 2014. Fuzzy FEMA model for risk evaluation of ship collisions in the Malacca Strait: based on AIS data. J. Simulat. 8, 91–104.

Zhang, L., Meng, Q., Xiao, Z., Fu, X., 2018. A novel ship trajectory reconstruction approach using AIS data. Ocean. Eng. 159, 165–174.

Zhou, Y., Daamen, W., Vellinga, T., Hoogendoorn, S.P., 2019. Ship classification based on ship behavior clustering from AIS data. Ocean. Eng. 175, 176–187.