



Production, Manufacturing and Logistics

A traveling salesman problem with pickups and deliveries and stochastic travel times: An application from chemical shipping

Aurora Smith Elgesem^a, Eline Sophie Skogen^a, Xin Wang^{a,*}, Kjetil Fagerholt^{a,b}^a Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Trondheim, Norway^b SINTEF Ocean, Trondheim, Norway

ARTICLE INFO

Article history:

Received 11 May 2017

Accepted 8 February 2018

Available online 15 February 2018

Keywords:

OR in maritime industry

Traveling salesman problem with pickup and delivery

Stochastic travel time

Monte Carlo simulation

ABSTRACT

This paper introduces a single-ship routing problem with stochastic travel times that is faced by a chemical shipping company in the Port of Houston. We take into explicit consideration the uncertain waiting times associated with the terminals inside the port, and the resulting inefficient transits caused by severe congestion. We show that the problem can be modeled as a stochastic *Traveling Salesman Problem with Pickups and Deliveries* (TSPPD), in which the goal is to find the route within the port with maximized probability that its total length does not exceed a threshold. We show that it is important to properly address the inefficient transits, and that including uncertainty in the travel times can have an impact in the choice of optimal route inside a port. We further show that the layout of the relevant terminals as well as their distances to the anchorage are important drivers of such impact. We conclude with the suggestion that one can use the proposed model and method to find a set of alternative routes, followed by a re-evaluation process since our method encompasses an approximation that underestimates the variation of the route completion time.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Maritime transportation (Christiansen, Fagerholt, Nygreen, & Ronen, 2007) is the backbone of globalization and cross-border transport networks that support supply chains and enable international trade. In 2015, world seaborne trade volumes were estimated to have exceeded 10 billion tons, which accounted for over 80 percent of total world merchandise trade (UNCTAD, 2016). It is customary to divide maritime transportation into several sectors by cargo types, e.g., container shipping, dry bulk (iron ore, coal, grain, etc.) and liquid bulk (petroleum/crude oil, liquefied natural gas, chemicals, etc.). The chemical shipping market belongs to the liquid bulk sector, and includes the carriage of a range of products such as organic and inorganic bulk liquid chemicals, vegetable/animal oils and fats and clean petroleum products. Worldwide, seaborne trade in the chemical sector accounted for 1.82% of the total trade volume at sea and was estimated at 998 millions ton-miles in 2016 (UNCTAD, 2016). One distinct characteristic that separates chemical shipping from other major sectors in the tanker market (characterized by the usage of different types of tankers to provide transport services), such as oil and gas, is that the ship-

ments are relatively small. A typical chemical tanker can have 15 to over 50 segregated tanks allowing the carrier to transport cargoes of different product types on one single tanker, and the cargoes in these tanks can be designated to a number of different charterers (customers). This special characteristic of chemical shipping has given rise to the issue of in-port transits made by chemical tankers as they typically make multiple “terminal calls” during a single port call (a stop or visit at the port to service the cargoes therein). Common to most chemical shipping companies, as a result, are the inefficiencies involved in their port calls and longer time spent in ports.

Odfjell, a leading company in the chemical shipping sector, is no exception. According to Hammer (2013), Odfjell vessels are spending 44% of their available time in port. Therefore, optimizing port call operations and thereby reducing time spent in ports are in their best interest, as more efficient port calls would increase the transport availability of the ships. Arnesen et al. (2017) made one of the first contributions on this topic, considering a single ship servicing numerous terminals and aiming to find an optimized route with shortest time spent in port. The authors modeled the problem as a *Traveling Salesman Problem with Pickups and Deliveries*, *Time Windows and Draft Limits* (draft of a ship is the vertical distance between the waterline and the bottom of its hull, determining the minimum depth of water the ship can safely navigate), or TSPPD-TWDL. Fig. 1 illustrates a small example of this

* Corresponding author.

E-mail addresses: xinwan@ntnu.no, xin.wang@hotmail.co (X. Wang).

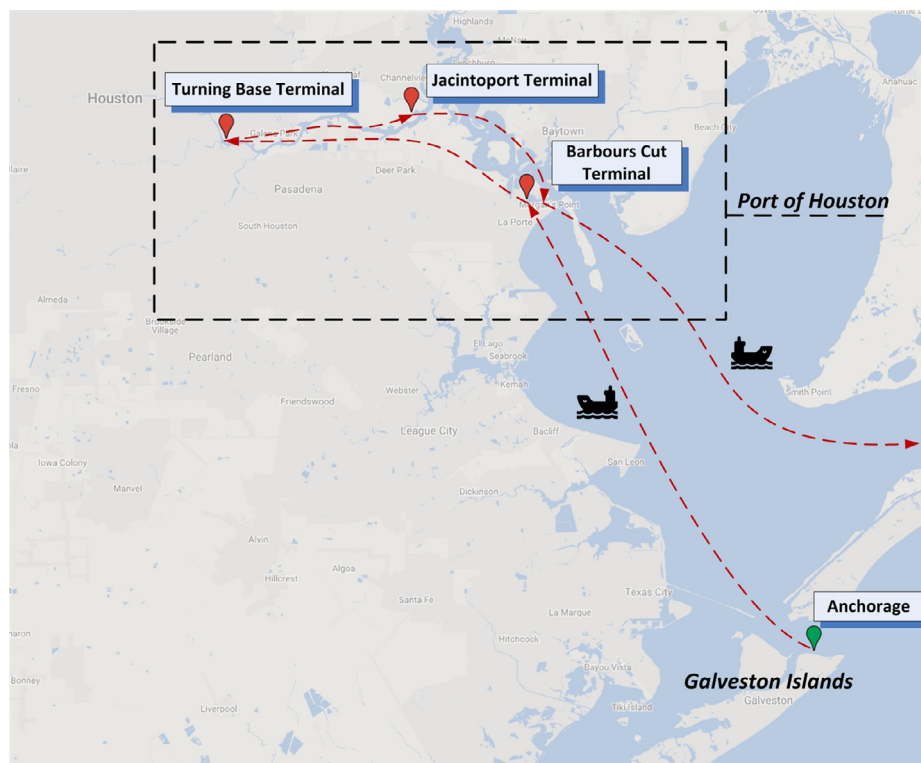


Fig. 1. Illustration of a port call. The ship approaches the Port of Houston from an anchorage point in the Galveston Bay and visits three terminals within the port.

problem in the Port of Houston. In this example the ship approaches the port via an anchorage point, and leaves the port after visiting all three terminals. In reality, as the terminals schedule vessel berthing requests on a first-come, first-serve basis, there are also waiting times at the terminals before they can accept the ship. In Arnesen et al. (2017), however, the authors simplify the problem by assuming deterministic waiting times. Hence the associated transit times between the terminals are also fixed and known.

This study addresses a similar in-port routing problem faced by the chemical shipping company Odfjell, and considers a more realistic scenario in the Port of Houston. Not mirrored anywhere else in the world, the Houston ship channel is 80 kilometers long and only 160 meters wide. Because of its layout and the recent boost in demand due to large expansions of U.S. energy and chemical companies, the channel is increasingly congested (Arnsdorf & Murtaugh, 2014). Once a vessel is inside the Galveston Islands, there is literally no place for a vessel to pull over or berth for a short time to accommodate scheduling difficulties, i.e., the vessel cannot wait “outside” or near any terminal (Kruse, 2015). Therefore, as required by the port authority, when a ship begins moving from one terminal, it must go to another terminal, the Bolivar Roads anchorage area (the Anchorage point in Fig. 1), or an offshore anchorage which is typically several miles further out to sea. This has resulted in a large number of *inefficient transits* made by chemical tankers towards layberths or anchorages due to unavailability of terminals, and a lot of time wasted in port. Statistics show that in 2014, around 18% of all transits made by chemical tankers in the Port of Houston are these inefficient transits (Kruse, 2015).

In this study, we explicitly take into account the inefficient transits due to a *special sailing pattern*, i.e., the ship will sail immediately towards the anchorage point after servicing a terminal, and only change its course back to the next terminal upon receiving the notification that the terminal is ready to accept the ship. The unavailability of a terminal is described by an *uncertain waiting time* from the point the ship leaves the previous terminal and

send out the berthing request (known as a NOR, Notice of Readiness), to the point when the next terminal receiving the NOR becomes available. Hence the *travel time* taken from one terminal to the next is also uncertain, depending on the realized waiting time of the next terminal, and also on where the ship is located when the next terminal becomes available, as the ship may still be close by or it may already be “parking” at the anchorage.

In Fig. 2 we illustrate three different cases for a transit between two terminals. In all three cases a ship is leaving Terminal A and is scheduled to service Terminal B next. Assume that the ship will sail at a constant speed and that Terminal B is located on its way back from Terminal A to Anchorage. In Case 1 where the waiting time at Terminal B, w_B , turns out to be 1.5 hours, the ship has not passed Terminal B upon receiving the notification that it is ready to accept the ship. Hence the travel time is 2 hours which corresponds to a direct movement from Terminal A to Terminal B. In Case 2 where $w_B = 3$ hours, the ship has passed Terminal B but not yet reached Anchorage, the resulting travel time is then 4 hours as the ship turns around and spends one more hour retracing part of its route back to Terminal B. Finally in Case 3 where $w_B = 4$ hours, the ship has just arrived at Anchorage when it gets the green light from Terminal B and the resulting travel time becomes 6 hours. This example shows how the transit time (travel time) from one terminal to the next can be uncertain depending on the realized waiting time at the destination terminal, and on the relative locations of the origin, destination terminals and the anchorage point.

This paper presents a stochastic routing model in which the special sailing pattern described above is explicitly addressed and the travel times are uncertain. The goal is to find the route within the port with maximized probability that its total length does not exceed a threshold. In particular, given the initial cargoes on board to be delivered when entering the port, and also a number of cargoes to be picked up during the port call, we model the problem as a stochastic *Traveling Salesman Problem with Pickups and Deliveries* (TSPPD) and aim to find the route that maximizes the probability

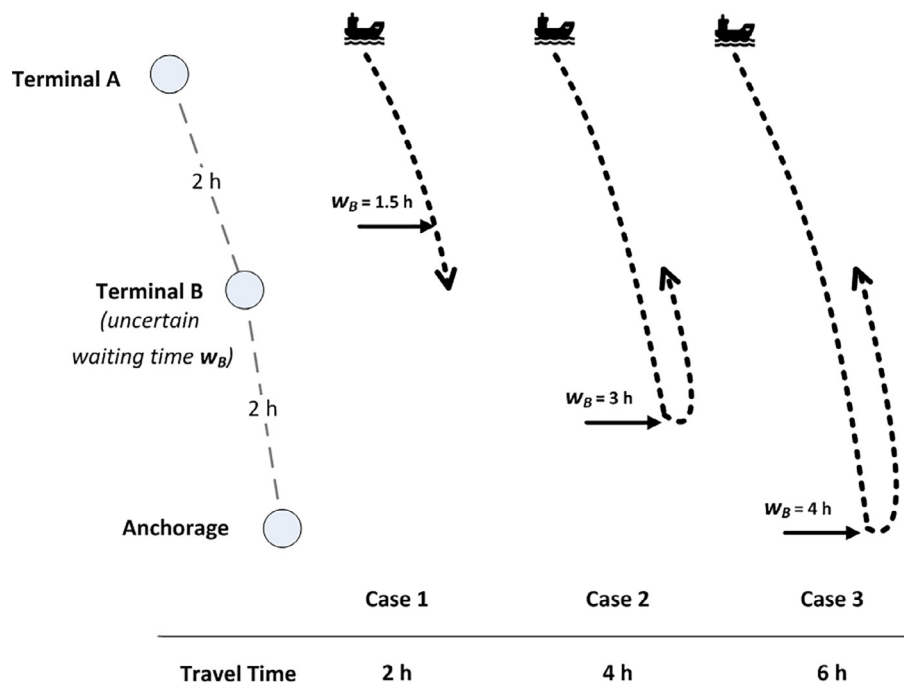


Fig. 2. Example showing three cases with different travel times for a transit from Terminal A to Terminal B.

that the port call can be completed within a threshold time period. This is a practical objective for the company: since there are usually multiple port calls along a ship trade, it benefits the planning of the voyage to the next port if the company knows with higher certainty how long the current port call will take.

In our case study, to better reflect on the realistic situation in the Port of Houston, we derive probability distributions for the uncertain travel times using Monte Carlo simulations based on real port data from Odfjell. We then solve the stochastic TSPPD by adapting the binary search enumeration method based on quasi-convex maximization introduced by Nikolova, Kelner, Brand, and Mitzenmacher (2006), and compare the optimal routes with those stemmed from solving the problem deterministically, e.g., using expected travel times between the terminals. We show that it is important to properly address the special sailing pattern, and that including uncertainty in the travel times can have an impact in the choice of optimal route inside a port. In addition, in response to the question raised by the shipping company, “whether we lose much by not considering uncertainty and using a deterministic model”, we show that in current situations at Port of Houston the numerical benefits of using a stochastic model are limited. However, since new anchorages and layberths closer to the channel are currently under development, we show that including uncertainty in the travel times has a much higher impact when the anchorage point is close to the terminals.

The contribution of this paper is twofold. First, it introduces and models the stochastic TSPPD with uncertain travel times based on the real in-port routing problem faced in the Port of Houston, while proposing a simulation method to handle the special sailing pattern. Second, by solving the problem and analyzing the results it demonstrates the usage of the model and the solution method as a decision support tool, providing the shipping company with guidance on how to obtain potentially optimal routes in the port and how they should be evaluated correctly. The remainder of the paper is organized as follows. Section 2 provides a brief literature review. Section 3 presents the mathematical model and introduces the solution method. Section 4 discusses the input data and the instance generation process, in which a simulation-based

approximation scheme is used. A computational study is given in Section 5. We conclude in Section 6.

2. Literature review

This study is most closely related to the work of Arnesen et al. (2017) in which the authors present a *Traveling Salesman Problem with Pickups and Deliveries* (TSPPD, see also Berbeglia, Cordeau, Gribkovskaia, & Laporte, 2007 for a classification for pickup and delivery problems) with draft limits and deterministic travel times between the terminals in the Port of Houston, and propose a dynamic programming method to solve the problem. Malaguti, Martello, and Santini (2018) study a similar TSPPD with draft limits (without time windows) and propose an exact branch-and-cut algorithm, as well as heuristics for obtaining feasible solutions in short computing times. Based on the model of Arnesen et al. (2017), Wang, Arnesen, Fagerholt, Gjestvang, and Thun (2018) propose a combined model taking both in-port routing and the allocation of chemical cargoes to tanks into consideration. Rakke, Christiansen, Fagerholt, and Laporte (2012) have also proposed models and solution methods for a class of TSPs with draft limits that arise in maritime transportation. These studies have not considered any uncertainty in their problems.

There exists a rather limited literature on the TSP and TSPPD with stochastic travel times, examples are Kao (1978), Sniedovich (1981), Carraway, Morin, and Moskowitz (1989), Jula, Dessouky, and Ioannou (2006), Chang, Wan, and OOI (2009), Bertazzi and Maggioni (2015) and Perboli, Gobbato, and Maggioni (2017), wherein travel time distributions are often assumed to be independent and are such that the distribution of the sum of those random variables can be readily computed (e.g., Normal, Poisson and Gamma(a, b) with constant b). In this paper the same assumption of independence and additivity is made using independent normal distributions, while the associated means and variances of the stochastic travel times are obtained through simulation using real data to handle the special sailing pattern illustrated earlier in Fig. 2. To our best knowledge, there are no other studies in the

literature that consider such a sailing pattern which is particularly relevant in the Port of Houston.

Our problem is also related to the *Stochastic Vehicle Routing Problem* (SVRP) which is a generalization of the stochastic TSP. Surveyed in Gendreau, Laporte, and Séguin (1996) and Oyola, Arntzen, and Woodruff (2016), the SVRPs arise when some elements of a vehicle routing problem are uncertain, and most articles consider stochastic demands or (the presence of) customers. There are, however, some efforts made in SVRP with uncertain travel or service times. Among those articles most have assumed quantifiable costs associated with the violation of soft time windows (e.g., Ando & Taniguchi, 2006; Russell & Urban, 2008; Yan, Wang, & Chang, 2014), the overtime of the whole route (e.g., Lei, Laporte, & Guo, 2012; Woensel, Kerbache, Peremans, & Vandaele, 2007) or both (e.g., Li, Tian, & Leung, 2010; Ta, Dellaert, van Woensel, & de Kok, 2013; Zhang, Lam, & Chen, 2013); and others have used probabilistic constraints to ensure the service level related to time windows by means of *chance constrained programs* (CCP) (e.g., Gómez, Marín, Akhavan-Tabatabaei, Medaglia, & Mendoza, 2016; Zhang, Chaovalitwongse, & Zhang, 2012). In practice, however, it is not always straightforward to quantify the costs for route “failure” when some time constraints are violated. This is also the case for Odfjell. Also, it is the company’s experience that the time windows for the customer cargoes (chemicals) in the Port of Houston are usually very wide, we have therefore not considered time windows in this paper.

3. In-port routing with uncertain travel times

We consider a chemical tanker arriving at a port with a given set of pickup and delivery commitments to be fulfilled during the port call, indicating the characteristics of the cargoes and their respective terminals for loading/unloading operations. A pre-planned (nonadaptive) route is to be determined prior to entering the port. Note that it can become adaptive by replanning the route on the fly with updated information. To start the port call, the tanker enters the port via an *anchorage* point to the first terminal in its visiting sequence, carrying a number of contracted *delivery cargoes* that are to be delivered to their respective destination terminals. The tanker then sails a planned route with sequenced terminal visits within the port to unload these cargoes, while also loading a set of *pickup cargoes* from different terminals along the route before finishing the port call. When inside the port, the tanker must sail immediately towards the anchorage after leaving a terminal until receiving notification that the next terminal is ready to accept the ship. All terminals in the port have their respective *draft limits* which are represented in terms of maximum weights the ship can safely carry on board when approaching and leaving the terminals. A tanker may even have to make two stops at the same terminal because of its draft limit – for instance, the water depth may not be sufficient until other cargoes have been discharged. Additionally, the *total capacity* (in deadweight tonnes) of the ship must always be respected, which can be another cause for multiple visits to the same terminal.

This in-port routing problem is formulated as a stochastic TSPPD with draft limits and uncertain travel times, in which the goal is to find the optimal route within the port with maximized probability that its total length does not exceed a threshold. We present its mathematical formulation in Section 3.1, and introduce the exact method used to solve the problem in Section 3.2.

3.1. Mathematical model

The stochastic TSPPD is defined on a directed graph $G = (N, A)$, where $N = \{0, \dots, n+1\}$ is the set of nodes and A the set of admissible arcs, which is a subset of $\{(i, j): i, j \in N, i \neq j\}$. Nodes 0 and

Table 1

Notation of parameters and decision variables.

Parameters	
Q_i	Signed weight of cargo i , positive for pickup cargoes and negative for delivery ones.
Q^+, Q^-	Total tonnage of the pickup and delivery cargoes, respectively. These are both positive values.
D_i	Draft limit for the service of cargo i , represented in terms of maximum weight (in deadweight tonnes) the ship can safely carry on board when approaching and leaving the terminal in which cargo i is located.
K	Total capacity of the ship in deadweight tonnes.
T_{ij}	Stochastic travel time from node i to node j representing the duration of all activities from the start of service for i to the start of service for j . Note that when i and j are located in the same terminal, such duration does not involve actual “traveling” and only consists of the service time at node i .
Decision variables	
x_{ij}	Binary flow variables, equal to 1 if the ship sails directly from node i to node j , and 0 otherwise.
y_{ij}	Total weight on board the ship when sailing from i to j .

$n+1$ represent the origin and destination nodes, respectively, and are both associated with the anchorage. In other words, the route is assumed to be a round trip that both starts from and ends in the same anchorage point. Note that if time spent in port is calculated up to the final terminal visited instead of the anchorage, those routes that end in a terminal deep inside the port channel may be unfairly favored. Let $N^C = \{1, \dots, n\}$ be the set of all cargo nodes, and every cargo is, according to the contracts, either a pickup node or a delivery node. Note that as mentioned earlier, it is not always possible for the ship to service all cargoes located in the same terminal during one terminal visit due to draft and capacity limitations, we thus cannot aggregate the cargo nodes according to their associated terminals. Also note that different nodes may correspond to the same physical terminal in the port. Let N^+ represent the set of pickup nodes and N^- the set of delivery nodes, $N^+ \cup N^- = N^C$. We show in Table 1 the notation of parameters and decision variables used in the model.

A route can then be described by the values of the binary variables x_{ij} , and a feasible route is defined as one that originates and ends in the anchorage and visits every cargo node exactly once, while respecting the total ship capacity and the draft limitations at all times. The total length (time) of a given feasible route can then be computed by $\sum_{(i,j) \in A} T_{ij}x_{ij}$. We aim to find the route which maximizes the probability that its total length does not exceed a threshold H . We therefore solve

$$\max \Pr \left(\sum_{(i,j) \in A} T_{ij}x_{ij} \leq H \right), \quad (1)$$

$$\sum_{j \in N} x_{0j} = 1, \quad (2)$$

$$\sum_{i \in N} x_{i,n+1} = 1. \quad (3)$$

$$\sum_{j \in N | (i,j) \in A} x_{ij} = 1, \quad i \in N^C, \quad (4)$$

$$\sum_{i \in N | (i,j) \in A} x_{ij} = 1, \quad j \in N^C, \quad (5)$$

$$\sum_{j \in N^C} y_{0j} = Q^-, \quad (6)$$

$$\sum_{i \in N | (j,i) \in A} y_{ji} - \sum_{i \in N | (i,j) \in A} y_{ij} = Q_j, \quad j \in N^C, \quad (7)$$

$$y_{0j} \leq Q^- x_{0j}, \quad (0, j) \in A, \quad (8)$$

$$y_{i,n+1} \leq Q^+ x_{i,n+1}, \quad (i, n+1) \in A, \quad (9)$$

$$Q_i x_{ij} \leq y_{ij} \leq (K - Q_j) x_{ij}, \quad (i, j) \in A \mid i, j \in N^+, \quad (10)$$

$$(Q_i - Q_j) x_{ij} \leq y_{ij} \leq K x_{ij}, \quad (i, j) \in A \mid i \in N^+, j \in N^-, \quad (11)$$

$$-Q_j x_{ij} \leq y_{ij} \leq (K + Q_i) x_{ij}, \quad (i, j) \in A \mid i, j \in N^-, \quad (12)$$

$$y_{ij} \leq \min\{K - Q_j, K + Q_i\} x_{ij}, \quad (i, j) \in A \mid i \in N^-, j \in N^+, \quad (13)$$

$$y_{ij} \leq D_j x_{ij}, \quad (i, j) \in A \mid j \in N^-, \quad (14)$$

$$y_{ij} \leq D_i x_{ij}, \quad (i, j) \in A \mid i \in N^+, \quad (15)$$

$$x_{ij} \in \{0, 1\}, \quad (i, j) \in A, \quad (16)$$

$$y_{ij} \geq 0, \quad (i, j) \in A. \quad (17)$$

Constraints (2) and (3) ensure the conservation of flow in the origin and destination nodes, respectively, while constraints (4) and (5) ensure that every cargo within the port is serviced exactly once. Constraint (6) states that the ship leaves from the anchorage carrying the total load that is to be delivered to the terminals within the port. Constraints (7) ensure that the difference between ingoing and outgoing shiploads of each node equals the weight of the cargo serviced at the node. Constraints (8) and (9) connect the x_{ij} and y_{ij} variables out from the origin node and into the destination node. Constraints (10)–(13) ensure that the load on board the ship along the route never exceeds the total capacity of the ship. Constraints (14) and (15) represent the draft limit restrictions when arriving at delivery nodes and departing from pickup nodes, respectively. The variable domains are given by constraints (16) and (17).

3.2. Solution method

Assuming that each T_{ij} is an independent normally distributed random variable, i.e., $T_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$, the probability in objective function (1) can be computed by

$$\Pr \left(\sum_{(i,j) \in A} T_{ij} x_{ij} \leq H \right) = \Phi \left(\frac{H - \sum_{(i,j) \in A} \mu_{ij} x_{ij}}{\sqrt{\sum_{(i,j) \in A} \sigma_{ij}^2 x_{ij}}} \right), \quad (18)$$

where $\Phi(\cdot)$ is the cumulative distribution of the standard normal random variable $N(0, 1)$. Also since Φ is monotonically increasing, its maximization is equivalent to maximizing its argument, i.e., to maximize Eq. (18) one needs to solve

$$\max \frac{H - \sum_{(i,j) \in A} \mu_{ij} x_{ij}}{\sqrt{\sum_{(i,j) \in A} \sigma_{ij}^2 x_{ij}}}. \quad (19)$$

To solve this non-linear and non-convex maximization problem we use the exact method proposed in Nikolova et al. (2006), which is generalized in Nikolova (2010), for stochastic shortest path problems based on quasi-convex maximization. Recall that we denote by A the set of admissible arcs, which is a subset of $\{(i, j) : i, j \in N, i \neq j\}$. By indexing all arcs in A by $1, 2, \dots, m$ we can denote a feasible route by its incidence vector $x = (x_1, \dots, x_m)$ where $x_k = 1$ if arc k is present in the route and $x_k = 0$ otherwise. Denote also the vector of means of all arcs by $\mu = (\mu_1, \dots, \mu_m)$ and the vector of variances by $\sigma^2 = (\sigma_1^2, \dots, \sigma_m^2)$. The set of all feasible routes is then a subset of the vertices of the unit hypercube in m dimensions (the m -dimensional analogue of a 2-dimensional square or a 3-dimensional cube), and we call the convex hull of these feasible vertices the *path polytope* which corresponds to the feasible region given by constraints (2) – (17). Thus, the optimal route in our problem is a solution to

$$\begin{aligned} \max \quad & \frac{H - \mu \cdot x}{\sqrt{\sigma^2 \cdot x}} \\ \text{s.t.} \quad & x \in \text{path polytope} \\ & x \in \{0, 1\}^m. \end{aligned} \quad (20)$$

Fig. 3 illustrates the projection of the hypercube vertices (representing all incidence vectors) and the path vertices (representing the feasible incidence vectors or feasible routes) onto the (μ, σ^2) -plane. The projection of the convex hull of the path vertices, i.e., the shadow of the path polytope, is also shown, with its dominant represented using (red) solid lines. Then, for example, computing the leftmost and lowermost vertices of the path polytope shadow correspond to finding the shortest path, in terms of arc means and arc variances, respectively. It is shown in Nikolova et al. (2006) that objective function (20) has a special structure such that, as long as the deadline H is no less than the mean of the smallest-mean path (i.e., the leftmost vertex of the path polytope shadow), objective function (20) is quasi-convex and its maximum always lies at one of the extreme points on the dominant of the path polytope shadow. It is further shown in Nikolova et al. (2006) that there is a *one-to-one* correspondence between the extreme points on the dominant of the path polytope shadow on the (μ, σ^2) -plane and the breakpoints of the parametric shortest path problem with edge weights $\mu + \lambda \sigma^2$, $\lambda \in [0, 1]$; and that each such extreme point thus corresponds to solving the *minimization* problem (notice that a lower route mean or a lower route variance helps objective function (20) to achieve a higher value) with a linear objective of the following form for some $\alpha \in [0, 1]$:

$$\min \quad \alpha \mu \cdot x + (1 - \alpha) \sigma^2 \cdot x. \quad (21)$$

We use the exact binary search enumeration method suggested in Nikolova (2010) to find all breakpoints of the parametric

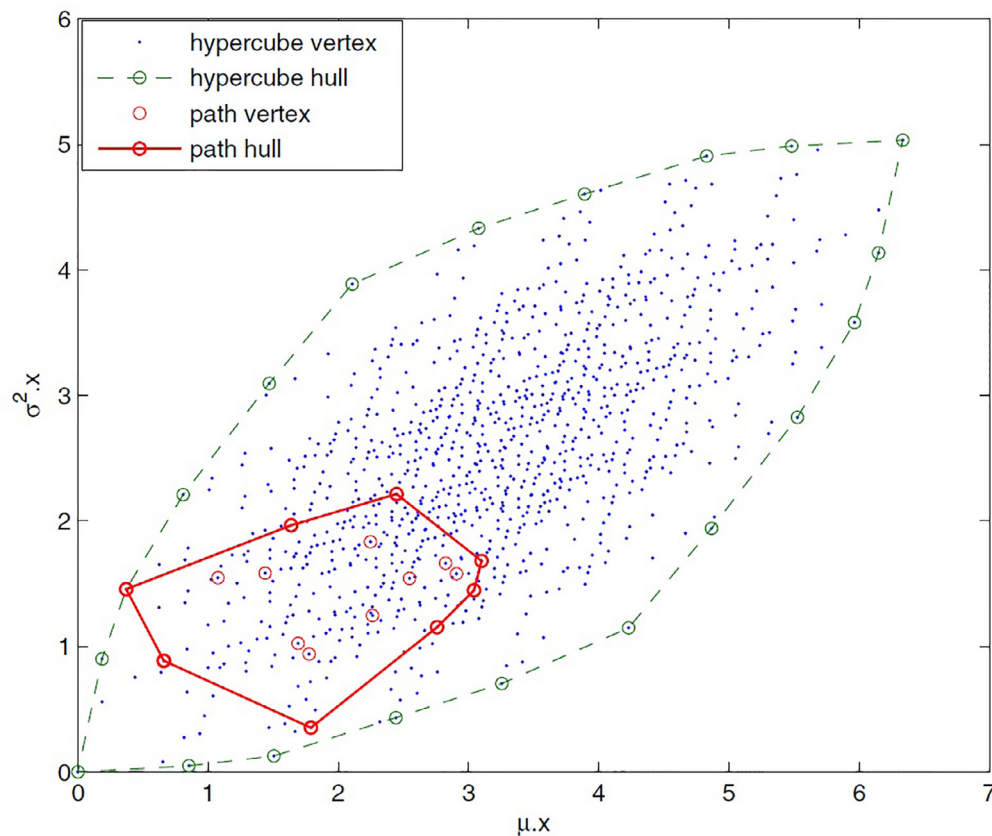


Fig. 3. Projection of the hypercube vertices, path vertices and their respective convex hulls onto the (μ, σ^2) -plane (Nikolova et al., 2006).

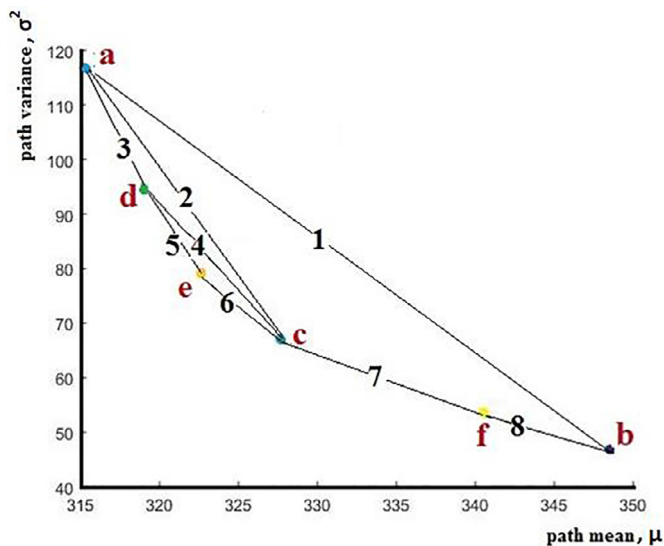


Fig. 4. Implementation of the binary search method on an example problem showing the order in which new points (routes) are found.

shortest path problem, i.e., all the extreme points on the shadow dominant. We show in Fig. 4 its implementation using an example. The method starts with setting $\alpha = 1$ and then $\alpha = 0$, solves the two corresponding deterministic routing problems by replacing the objective function (1) with (21) and using, e.g., a commercial solver for general mixed integer linear programs (MILPs). The resulting two solutions, represented by points **a** and **b** on the (μ, σ^2) -plane in Fig. 4, correspond to the route with the smallest mean and the route with the smallest variance, respectively. We next set α so

that the slope of the linear objective is the same as the slope of the line connecting points **a** and **b**. We solve the deterministic problem with the updated α , and denote the resulting route by point **c** (see Fig. 4). The search then continues similarly to find a route between **a** and **c**, and between **c** and **b**, etc., until no further new routes are found. In the example shown in Fig. 4, after point **c** we first search to its left (using updated α corresponding to line 2), finding point **d**. We search to the left, again, of point **d**, however finding that there exists no new point in that direction. We then turn to the right of point **d**, to find new points between points **d** and **c**, and between points **c** and **b**, and so on. In Fig. 4 the points **a**, **b**, ..., **f** therefore indicate the order in which feasible points are found, and the numbers 1, 2, ..., 8 represent the order in which the slopes of the corresponding edges are used to update α in the linear objective. Note that in this way, by finding all the points minimizing some positive linear combination of the mean and variance, we obtain all extreme points of the shadow dominant, regardless of the sequence in which they are found (which depends on the direction to which we perform the search after finding a new point). Then, since the optimal point always lies in this set of extreme points, we may compare the obtained solutions in terms of objective function (20) and the optimal route is the one that maximizes this objective.

4. The input and test instances

The input data used for our computational study is based on real operating data of Odffjell in the Port of Houston. In Section 4.1 we present the geography of the port and the layout of the terminals. We discuss in Section 4.2 the relation between uncertain waiting and travel times, and we propose the simulation-based approximation method to handle the special sailing pattern. Finally in Section 4.3, we provide an overview of the randomly

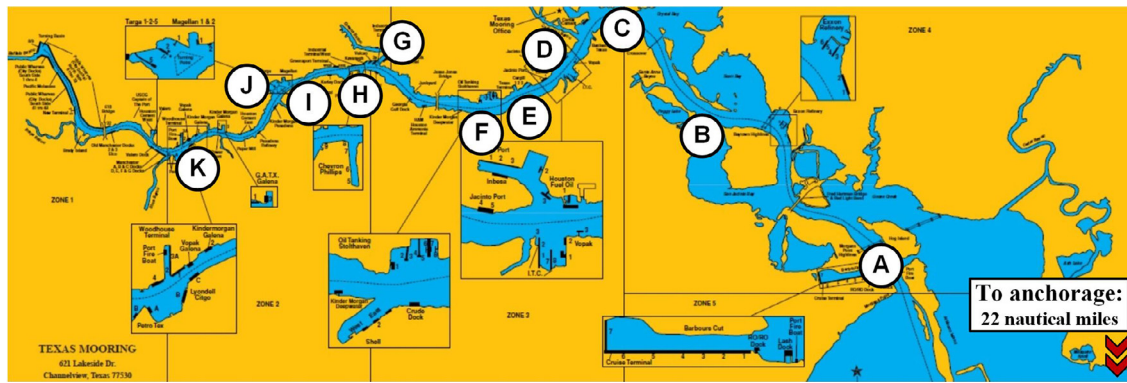


Fig. 5. The 11 terminals in the Port of Houston that are considered in the tests.



Fig. 6. The terminals A–K and the anchorage (point 0) seen as points distributed on a straight line. The numbers on the line indicate the direct sailing times in hours between the adjacent points.

generated instances. These instances may be downloaded from: <https://doi.org/10.17632/mx6cpkwdt9.1>.

4.1. Geography

We consider in total 11 terminals in the Port of Houston shown as Terminals A–K in Fig. 5. The anchorage is located outside the port, some 22 nautical miles from Terminal A. During a port call, Odfjell's chemical tankers normally visit a subset of these terminals and may service several cargoes at each terminal. Since all the terminals are located along the narrow port channel, without much loss of accuracy, we can see them as being distributed along a straight line (as shown in Fig. 6) in order to facilitate our tests and analysis. Fig. 6 also shows the approximated direct sailing times (in hours) between any two adjacent terminals, as well as between the anchorage and Terminal A. These are calculated by assuming a constant sailing speed of 4.3 knots in the port.

Depending on the locations of the service commitments, the tanker usually visits a different set of terminals each time when calling the port, and that seldom includes all 11 terminals shown above. In this paper, we consider the most common scenario based on the company's daily operations and hence always include six terminals when constructing our test instances. However, to diversify the instances we differentiate among three types of layout of the relevant terminals in the port, namely *Even*, *Far* and *Split*. The *Even* layout indicates that all terminals are relatively spread out along the ship channel, while the *Far* layout represents a situation where all terminals are located deep inside the channel and thus relatively far from the anchorage. Finally in the *Split* layout, the terminals are split into two groups, one deep inside the channel and the other located close to the outlet of the channel.

In our computational study we use three different sets of (six) terminals to approximate the three types of layout (see Fig. 7): Terminals A, B, D, G, I, K for the *Even* layout; Terminals F, G, H, I, J, K for the *Far* layout; and Terminals A, B, H, I, J, K for the *Split* layout.

4.2. Uncertain waiting and travel times

Recall that in our model we assume each travel time T_{ij} , defined as the duration of all activities from the start of service for node i to the start of service for node j , is an independent normally distributed random variable, i.e., $T_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$. The travel time T_{ij}

consists of two major components: a deterministic component, denoted L_i , that consists in the service duration of cargo i , including its loading/unloading time as well as tank washing time if i is a delivery cargo, which are fixed and known a priori; and a stochastic component, T_{OD} – the actual sailing time to move from the origin terminal (associated with node i) to the destination terminal (associated with node j). As illustrated in Fig. 2, the actual sailing time does not solely depend on the direct sailing distance between the two terminals, but also on the waiting time for the destination terminal which is defined as the time from the point the ship leaves the origin terminal to the point when the destination terminal (or one of its berths) is ready to receive the ship.

The dependency between waiting time and actual sailing time can be described as follows. Consider the move from Terminal O to Terminal D. Let t_{OA} and t_{DA} be the (known) time taken to sail from Terminal O and Terminal D, respectively, to Anchorage. For ease of presentation, we use t_{OD} to represent the direct movement time between Terminal O and Terminal D, i.e., $t_{OD} = |t_{OA} - t_{DA}|$. Further let w be the waiting time for Terminal D. Then, if Terminal D is located between Terminal O and Anchorage, the actual sailing time from Terminal O to Terminal D, T_{OD} , can be expressed with the following equations:

$$T_{OD} = \begin{cases} t_{OD} & \text{if } w < t_{OD}, \\ 2w - t_{OD} & \text{if } t_{OD} \leq w < t_{OA}, \\ w + t_{DA} & \text{if } w \geq t_{OA}. \end{cases} \quad (22)$$

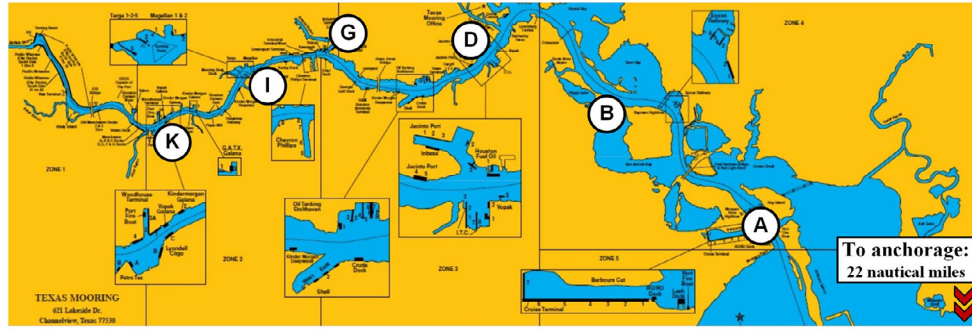
Note that the three cases in Eq. (22) correspond to the following three scenarios of where the ship is when the waiting time has elapsed (much like the cases illustrated in Fig. 2), respectively: the ship has not passed Terminal D; the ship has passed Terminal D but not arrived at Anchorage yet; the ship is already parked at Anchorage.

Similarly, if Terminal O lies between Terminal D and Anchorage, then

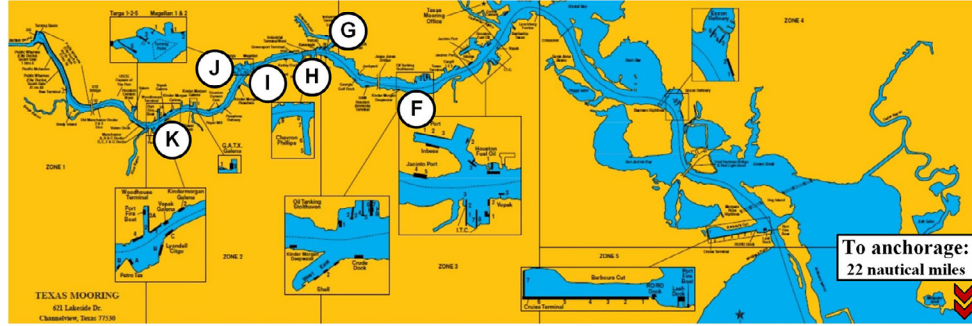
$$T_{OD} = \begin{cases} 2w + t_{OD} & \text{if } w < t_{OA}, \\ w + t_{DA} & \text{if } w \geq t_{OA}. \end{cases} \quad (23)$$

Also, if Terminal O and Terminal D are in fact the same, i.e., if nodes i and j are located in the same terminal, then $T_{OD} = 0$.

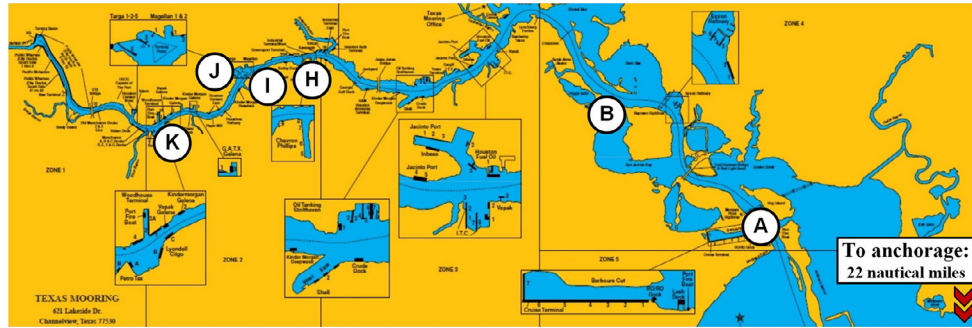
Clearly, to better represent the stochasticity of travel times $T_{ij} = L_i + T_{OD}$ (where terminals O and D correspond to nodes i and j , respectively), it is useful to start with the reality of waiting times for



(a) Even layout



(b) Far layout



(c) Split layout

Fig. 7. Three different types of layout of terminals.

the terminals. We have therefore acquired from Odfjell a data set containing nearly 400 waiting times experienced by their tankers in the Port of Houston. This set of data contains a small portion of data points that we consider to be obviously erroneous such as waiting times that are excessively long. Also, some berth visits lack the specifics of the associated terminal names, mainly because the collection of this type of data has not been part of their operational routine. However, this set of data can give us a general picture of the uncertainty faced by the company. We have therefore not tried to fit some distribution to the data points due to the noise therein, and instead only make use of the fact that the *median* of all waiting times is around 5 hours. We thus denote the *waiting time* for each cargo $j \in N^C$ by w_j and assume it follows a normal distribution, i.e., $w_j \sim N(m_j, s_j^2)$. Then, we let all m_j be independent random variables that are uniformly distributed over a support $[0, 10]$, and then set the standard deviations to 50% of their respective means, i.e., $s_j = 0.5 \times m_j$ for all $j \in N^C$.

We then use Monte Carlo simulation to derive the parameters for the stochastic travel times T_{ij} . Recall that in our model we assume $T_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$, and that $T_{ij} = L_i + T_{OD}$, where terminals O and D correspond to nodes i and j , respectively. For every pair of i and j , we can then run a Monte Carlo simulation to observe the

outcomes for T_{OD} , based on sampling the values for $w_j \sim N(m_j, s_j^2)$, and subsequently computing T_{OD} with Eqs. (22) and (23). The mean and variance of these outcomes, $E[T_{OD}]$ and $VAR[T_{OD}]$, are then used to estimate the parameters for T_{ij} , i.e., $\mu_{ij} = L_i + E[T_{OD}]$, and $\sigma_{ij}^2 = VAR[T_{OD}]$.

We illustrate in more detail the process of simulating outcomes for T_{OD} with two examples. Consider the moves from Terminal C to Terminal E, and from Terminal G to Terminal E (see Figs. 5 and 6 for their locations), and $w_j \sim N(5, 2.5^2)$ in both cases. For the move from Terminal C to Terminal E, we can draw 10,000 realizations of w_j , and then based on each of them calculate the realized sailing time T_{OD} using Eq. (23). In this way we obtain 10,000 realized sailing times, which are aggregated and shown as a histogram in Fig. 8a. Similarly, using Eq. (22), we can obtain 10,000 realized sailing times for the move from Terminal G to Terminal E, shown in Fig. 8b.

One may notice the tall bars at the leftmost of the histograms in Fig. 8. In Fig. 8a, for instance, the tall bar represents the occurrence of actual sailing times within the range $[2.5, 3.0)$. This is because when drawing realizations for the normally distributed w_j we set those negative values (which happens with a probability of 2.28% when the standard deviation is 50% of the mean) to 0 for

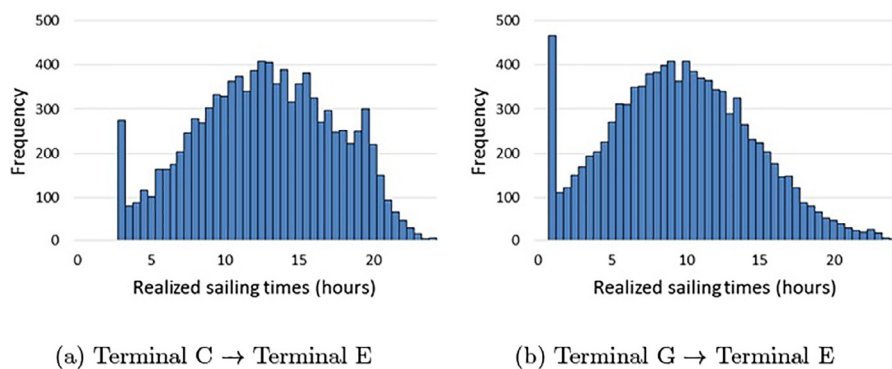


Fig. 8. Two examples of Monte Carlo simulation for actual sailing time where the waiting time of the destination terminal is subject to $N(5, 2.5^2)$: Terminal C → Terminal E for Example (a) and Terminal G → Terminal E for Example (b).

Table 2

The results of measuring the mean, variance and skewness of two sets of realized sailing times shown in Fig. 8.

Sample measures	(a) C → E	(b) G → E
Mean	12.4	9.4
Variance	4.7 ²	4.8 ²
Skewness	−0.1 (right skewed)	0.2 (left skewed)

practical purposes, and because the sailing time should be at least 2.54 hours which is the direct distance (in hours) between the two terminals.

We further show the mean, variance and skewness of the set of realized sailing times for the two examples in Table 2. In the first case, for example, where the ship moves from node i in Terminal C to node j in Terminal E, the simulation yields 12.4 and 4.7² as $E[T_{OD}]$ and $VAR[T_{OD}]$, respectively. Then, given the deterministic service time L_i of node i , the simulated stochastic travel time from i to j is such that $T_{ij} \sim N(L_i + 12.4, 4.7^2)$.

In this way we can obtain the Monte Carlo estimates for the μ 's and σ^2 's of all T_{ij} , given the specifics of the nodes (location, deterministic service time, etc.) and their respective waiting time distributions, by running a simulation for each pair of nodes $(i, j) \in A$. Note that there are two special situations where no simulation is required for node pair (i, j) : (1) when i and j are located in the same terminal, in which case the service of j will carry out immediately after that of i and the corresponding sailing time is 0; (2) when $j = n + 1$ representing the final transit back to the anchorage after servicing all cargoes, the corresponding μ_{ij} is then L_i plus the distance (in hours) from i to the anchorage, and the corresponding σ_{ij}^2 is 0.

In addition, to ensure the validity using Monte Carlo estimation, we also check the goodness of fit of the finishing times of a complete route to the normal distribution by running whole-route simulations (detailed in Section 5.4). It will be shown later that the normality assumption for a complete route holds, but disregarding the skewness and the “leftmost bars” may lead to underestimation of route variance.

4.3. Generating the test instances

The test instances used in our computational study are generated based on real data and a randomized process. We first give an overview of the input data used to generate the instances in Table 3 before explaining in more detail. The ranges in the “Values” column indicate that the associated input parameters are chosen uniformly at random from the given intervals. As mentioned earlier in Section 4.1, every instance considers six terminals according

Table 3

Summary of input data used to generate the instances.

Input data	Values
Layout	<i>Even, Far, Split</i>
Number of terminals	6
Number of cargoes	20
Number of pickup cargoes	8–12
Total pickup/delivery load as % of total ship capacity	80–90%
Draft limits as % of total ship capacity	50–90%
Mean of waiting time	0–10 hours
SD as % of mean	50%
Loading/unloading time	0.5–14 hours
Tank cleaning time	3–15 hours

to a chosen layout, *Even*, *Far* or *Split*, which determines which six terminals to include.

The chemical tankers used by Odfjell have capacities ranging from 4000 to 50,000 deadweight tonnes (*dwt*). In our computational study, we only consider one type of ship with 19,805 *dwt* capacity for all test instances. However, we let draft limits and cargo sizes be expressed as percentages of the total ship capacity when generating the instances, so that the instances are representative for more than one type of ship.

We consider for every instance 20 cargoes, each of them is either a pickup or a delivery cargo. In each instance, we randomly choose 8 to 12 to be pickup cargoes and the rest are delivery cargoes. The total load to be picked up and the total load to be delivered are both randomly set to between 80% and 90% of the total ship capacity. The size of each cargo is then assigned at random (with a lower bound) such that the sums of all pickup and delivery cargoes equal the total pickup load and total delivery load, respectively.

The draft limits at the terminals determine the maximum weight on board the ship when approaching and leaving every terminal along the route, and, as shown in Arnesen et al. (2017), they can be restrictive and significantly affect the routing decisions. For our problem to be feasible, at least one terminal must have large enough draft limit for the ship to enter the port with total delivery load on board, and also at least one terminal for the ship to leave with total pickup load on board. Therefore, for every instance we always draw one or two terminals at random and set their draft limits to 120% of total ship capacity to increase the likelihood of feasibility of the problem. The draft limits for the remaining terminals are then chosen randomly between 50% and 90% of the total ship capacity.

To obtain the means and variances for the travel times T_{ij} in our model, as explained in Section 4.2, we first generate the distribution for the waiting time $w_j \sim N(m_j, s_j^2)$ associated with each

cargo node j . Also, since at Port of Houston the uncertainty of the waiting times mainly come from queuing for available berths at the terminals and not the loading/unloading preparation of the cargoes, we generate the waiting time distributions based on terminals. Therefore, with each terminal we associate a mean waiting time randomly chosen between 0 and 10, and then let all nodes located in this terminal have this same mean. After the mean waiting times (m_j 's) are chosen for all cargo nodes, we let the standard deviations (s_j 's) be 50% of their respective means. Furthermore, the deterministic components of the travel times are generated as follows to represent real-life situations: the loading/unloading time of a cargo is chosen uniformly at random between 0.5 and 14 hours; and the tank cleaning time associated with a delivery cargo is random between 3 and 15 hours.

5. Computational study

For the computational study we start with creating three sets of instances based on the terminal layout: the *Even* set, the *Far* set and the *Split* set; each set containing 100 instances randomly generated as described in Section 4.3. When solving the problem instances, the binary search algorithm for the stochastic TSPD is implemented in Matlab R2015a, while the deterministic subproblems (with linear objectives) therein are solved using the general MILP solver in FICO Xpress 7.8. All experiments are run on an Intel 3.4 gigahertz processor with 16 gigabytes memory.

In Section 5.1 we use a small illustrative example to show how the optimal solution changes according to the specific criterion used, e.g., with and without the consideration of uncertainty. We then show the difference between the deterministic and stochastic solutions to a realistically generated instance in Section 5.2. The impacts of different instance characteristics are discussed in Section 5.3, and we evaluate our Monte Carlo approximation method in Section 5.4.

5.1. Optimal route incorporating risk

In the absence of uncertainty, such as in Arnesen et al. (2017) where the problem is to find the best route that visits all cargoes while respecting all routing constraints, the definition of the optimal route can be simple and straightforward (though solving the problem is not): the feasible route with the least completion time. In a stochastic context, however, it is sometimes not clear how to define the optimal route: is it the one that minimizes the expected completion time, its variance or some combination of the two? In real-life planning, it of course depends on the preference of the decision maker.

Fig. 9 shows a small illustrative problem with three terminals and three cargoes as well as two feasible routes for the problem. Three terminals, A, B and C, are represented by grey squares, with their respective waiting time distributions (in hours) and draft limits (abbreviated as DL) shown above, and each terminal contains one cargo with a positive/negative number indicating its pickup/delivery load. For example, there is a delivery cargo weighing 4000 tonnes in Terminal A, in which the waiting time follows distribution $N(10, 5^2)$ and the draft limit is 10,000 tonnes. Note that the waiting times at both Terminals B and C are assumed to be deterministic, hence both follow $N(0, 0)$. Also note that the draft limit at Terminal B is intentionally set to be low (2000 tonnes), so that the ship must unload the delivery cargo at Terminal A prior to visiting Terminal B.

Two feasible routes are then illustrated in Fig. 9 showing their respective visiting sequences: [Anch. \rightarrow A \rightarrow C \rightarrow B \rightarrow Anch.] for Route 1, and [Anch. \rightarrow C \rightarrow A \rightarrow B \rightarrow Anch.] for Route 2. The travel time distribution for each transit is also indicated, e.g., the travel time for the [C \rightarrow A] transit in Route 2 follows $N(12.5, 3.1^2)$

which is obtained through simulation. Also since we assume deterministic waiting times at Terminals B and C, the travel times of all transits destined in these two terminals are also deterministic. Notice that for both routes there is only one transit having an uncertain travel time (with a non-zero variance): the one that is destined in Terminal A, i.e., [Anch. \rightarrow A] for Route 1 and [C \rightarrow A] for Route 2. In this particular case, approaching Terminal A from deep inside the channel (Terminal C), instead of from the anchorage, reduces the travel time variance from 25 to 9.6 (or by 62%).

By comparing the total completion time of the two routes, one can notice that Route 1 has a smaller mean (32), and Route 2 has a smaller variance (9.6). Hence, one that seeks only to minimize the expected completion time would clearly favor Route 1 despite its higher variance which indicates a higher degree of uncertainty. However, if the decision maker takes into account some measure of risk in their objective, the choice of a better route is less straightforward. For example, if we seek to maximize the probability of completing the route within 36 hours (represented by H in our model), Route 2 has a probability of 87.1% in contrast to only 78.8% for Route 1; and if H is set to 40 hours, the numbers are 99.2% for Route 2 and 94.5% for Route 1. Therefore, under some circumstances a shipping company may find Route 2 a better choice.

For this small example problem, Route 1 is in fact the optimal solution if the objective is to minimize the expected completion time; and Route 2 is the optimal solution when the objective is as proposed in our model and when H is at least 33.4 hours (the breakpoint is somewhere between 33.3 and 33.4 where the corresponding probability is around 60%). In the remaining of this paper, we call the solution like Route 1 in the former case, i.e., the smallest-mean route, the *deterministic solution* of a stochastic problem, as finding such a route may be seen as a deterministic problem in which the uncertain travel times are replaced by their means. The *stochastic solution* of a particular problem instance, in contrast, should depend on the value of the threshold H chosen for that instance. Since we generate the test instances on a random basis, including the loading/unloading times and tank cleaning times etc., we do not use a constant H for all instances. Instead, for every instance we set H to a value such that the associated deterministic solution gives a 95% probability of completing within this threshold value. For the example discussed above, this H is 40.2 for the deterministic solution (Route 1), and the stochastic solution (Route 2) in turn gives 99.4% as the probability of keeping the completion time within 40.2 hours. Note that such “95% rule” is arbitrarily chosen in this study as the criterion to determine the optimal stochastic solution, and other values or other criteria may be used in practice.

5.2. Difference between deterministic and stochastic solutions

As mentioned earlier, in the small example illustrated in Fig. 9 the difference in route variance stems from the way the ship approaches Terminal A. By making the [C \rightarrow A] transit, i.e., placing Terminal A in the visiting sequence right after the relatively deep and far Terminal C, the ship sailing Route 2 mitigates the variance associated with the waiting time at Terminal A. This is partly because Terminal A (in the case of Fig. 9) is located very close to the anchorage and therefore the ship, having finished its previous call at Terminal C, is sailing towards the anchorage and hence Terminal A “anyway”. In this way, Route 2 (or the stochastic solution in that case) is able to reduce the total variance by nearly 62% at the expense of a small increase in total mean (less than 2%), compared with its deterministic counterpart Route 1.

However, our preliminary tests have shown that in a more realistic scenario which normally has more cargoes and terminals involved, and waiting time uncertainty in almost every terminal, the difference between the deterministic and stochastic solutions can

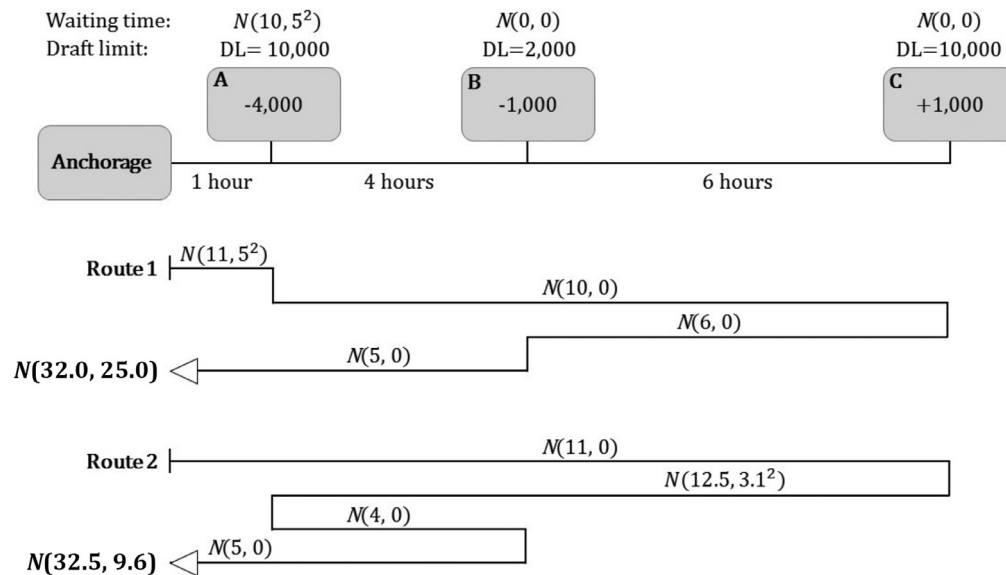


Fig. 9. Comparing two different routes for an illustrative problem: Route 1 is the smallest-mean route, whereas Route 2 has slightly higher mean but much lower variance.

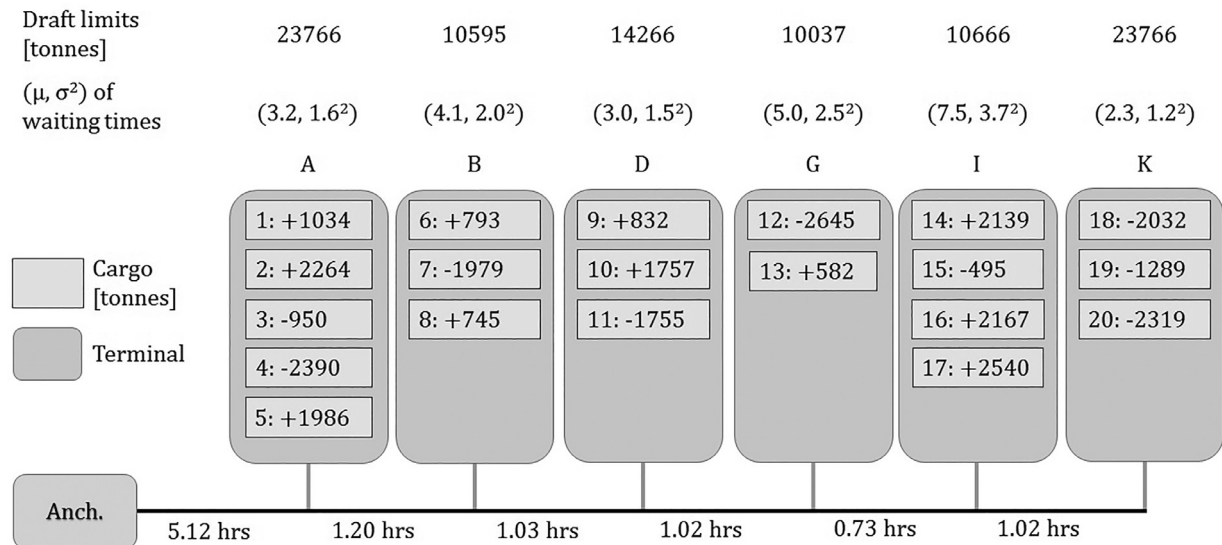


Fig. 10. An example instance Even-46 from the Even set.

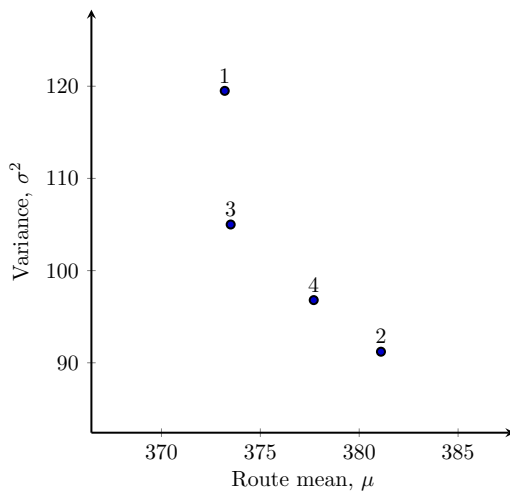
be less significant. To confirm this we start by solving all 100 instances in the *Even* set. The results show that out of 100 instances, only 18 of them have a stochastic solution (using the “95% rule”) that differs from its deterministic counterpart, i.e., for the rest 82 instances the deterministic solution is also the optimal one in the stochastic setting.

We use one instance (for which deterministic and stochastic solutions are not the same) as an example, described in Fig. 10. We call this instance *Even-46* as it is the 46th instance in the *Even* set. This problem includes six terminals and 20 cargoes to pick up or deliver, and unlike the previous small illustrative example in Fig. 9, the anchorage point is quite far from all of the terminals as is the case in the Port of Houston.

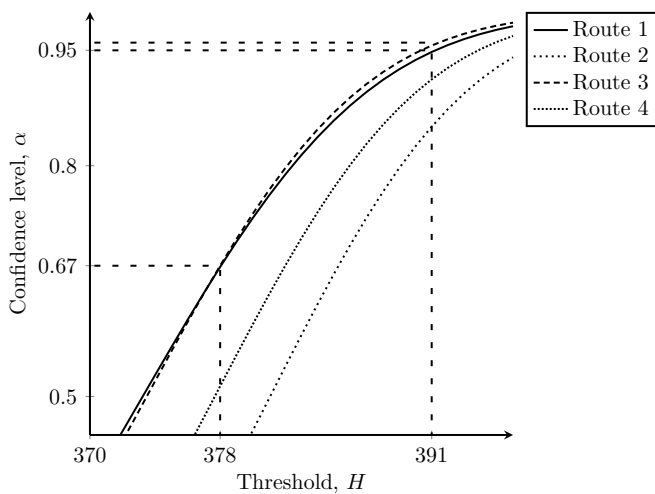
Using the binary search method introduced in Section 3.2 we find four routes, shown in Fig. 11a as numbered points on a (μ, σ^2) -plane indicating their route means and variances. Route 1 is essentially the deterministic solution to the problem having the smallest route mean, while the other three can be seen as its alternatives with a higher mean but lower variance. In Fig. 11b we

show the cumulative distribution function (CDF) for the completion time of the four routes. The stochastic solution is determined by setting the threshold H to 391, such that Route 1 has a confidence level of 95% (that it can be finished within 391 hours), and by comparing the resulting confidence levels given by all four routes. Route 3 is, therefore, chosen as the stochastic solution to the problem as it gives the highest confidence level of around 96% when $H = 391$. Also note that $H = 378$ is the breakpoint above which Route 3 replaces Route 1 to be the stochastic solution, which is also shown in Fig. 11b.

In this example, the quality of Route 1 (the deterministic solution) is fundamentally good. For the given $H = 391$ the confidence level of Route 1 is only about 1% lower than that of Route 3 (the stochastic solution) which indicates that the deterministic solution is as good as the stochastic one in this case. However, compared with its deterministic counterpart, Route 3 manages to reduce the route variance by around 12% (from 119.5 to 105.0) at the expense of only 0.07% increase in route mean (from 373.24 to 373.49). To compare the deterministic and stochastic solutions in



(a) Means and variances of the four feasible routes, represented by numbered points on a (μ, σ^2) -plane.



(b) The cumulative distribution function (CDF) for the completion time of the four routes. Given some threshold H one can compare the probability (confidence level) of each route finishing within this threshold.

Fig. 11. Four feasible routes for instance *Even-46* found using the binary search method.

Table 4

Comparison of the transits which differ between Routes 1 and 3 for instance *Even-46*.

Route 1			Route 3		
Transit	μ	σ^2	Transit	μ	σ^2
[A \rightarrow K]	9.7	$2.3^2 = 5.4$	[A \rightarrow D]	8.2	$2.9^2 = 8.4$
[K \rightarrow D]	7.6	$4.4^2 = 19.0$	[D \rightarrow K]	7.5	$2.3^2 = 5.5$
[D \rightarrow G]	10.8	$4.6^2 = 20.8$	[K \rightarrow G]	12.6	$4.1^2 = 16.9$
Sum	28.1	45.2	Sum	28.3	30.8

more detail, we show the terminal and cargo visiting sequence of the two routes in Fig. 12. The travel time distribution is shown above each transit, and the cargoes being serviced at a terminal visit are indicated by their numbers (see Fig. 10 for the numbering of all cargoes). We also highlight the parts of the two routes that differ from each other in Fig. 12, i.e., [A \rightarrow K \rightarrow D \rightarrow G] for Route 1 in contrast to [A \rightarrow D \rightarrow K \rightarrow G] for Route 3, and further compare these different parts in Table 4.

From Table 4 we can see that three transits are different between Route 1 and Route 3, and that the [D \rightarrow K] transit in Route 3 is the one that contributes the most to the reduced route variance of the stochastic solution. On the other hand, the opposite [K \rightarrow D] transit in Route 1, which is an “originating from deep inside the channel and heading towards the anchorage” transit, has a high variance (19.0). Remember that for the illustrative example in Fig. 9, a similar type of transit ([C \rightarrow A]) that originates from deep inside the channel is able to reduce the relatively high waiting time variance at the destination terminal (from 25 to 9.6). This is because in that illustrative example the anchorage is very close to the destination terminal, and thus the ship is more likely to be parking at the anchorage when the waiting time is longer than expected. This is not the case in Port of Houston where the anchorage is located further into the sea. Therefore, despite that the [K \rightarrow D] transit is started by “heading towards the anchorage and the destination terminal anyway”, its travel time variance is still high as the ship may find itself having to turn around at different points between terminal A and the anchorage.

Based on the analysis of the example instance *Even-46* and its solutions we emphasize the following insights. First, when planning a route inside a port it is important to properly address the special sailing patterns (if any) instead of making certain assumptions to simplify the problem. In this study, for example, we see that the actual travel time for an in-port transit is simultaneously determined by the distance between the origin and destination terminals, the waiting time uncertainty in the destination terminal and the relative location of the anchorage. This combined effect cannot be captured, in any way, by assuming that the ship can wait “right outside” the terminal even if the uncertainty of such waiting time is taken into account. Second, one should always try to solve the stochastic problem in order to obtain some alternatives in addition to smallest-mean route (the deterministic solution). A company with risk-averse requirements may favor routes with a low variance in order to facilitate their planning for the voyage afterwards. In this case, the criteria with regards to smallest-mean or even our probabilistic objective based on some given threshold may be less important; the company should instead be presented with a set of alternative routes each representing some trade-off between low expected completion time and low risk. Third, the fact that the observation regarding the transits heading towards the anchorage (derived from the illustrative example in Fig. 9) being not a consistent conclusion (when the anchorage is further away) suggests that this is not an appropriate rule for determining the mean-variance trade-off. It also points to the fact that the relative distance from the anchorage to the terminals may be an important factor, which we discuss in more detail in Section 5.3.

5.3. Comparing the impacts of different instance characteristics

The results for instance *Even-46* in fact represent the general situation of most test instances that we realistically generated based on Port of Houston, i.e., the deterministic (smallest-mean) solution, which considers the special sailing pattern but only utilizes the means (without variances) of the travel times, is as good as the stochastic solution under most circumstances. However, it is worth noting that one should not always assume good performance of the deterministic solution before actually solving the stochastic problem. As shown by the illustrative example in Fig. 9, under certain circumstances the stochastic solution may still be considerably better when considering some measure of risk.

We test on some other sets of instances in order to investigate the impact of different instance settings. This is to answer the question: Under what circumstances is the stochastic solution more likely to differ from its deterministic counterpart? Apart from the *Even*, *Far* and *Split* sets that represent different types of

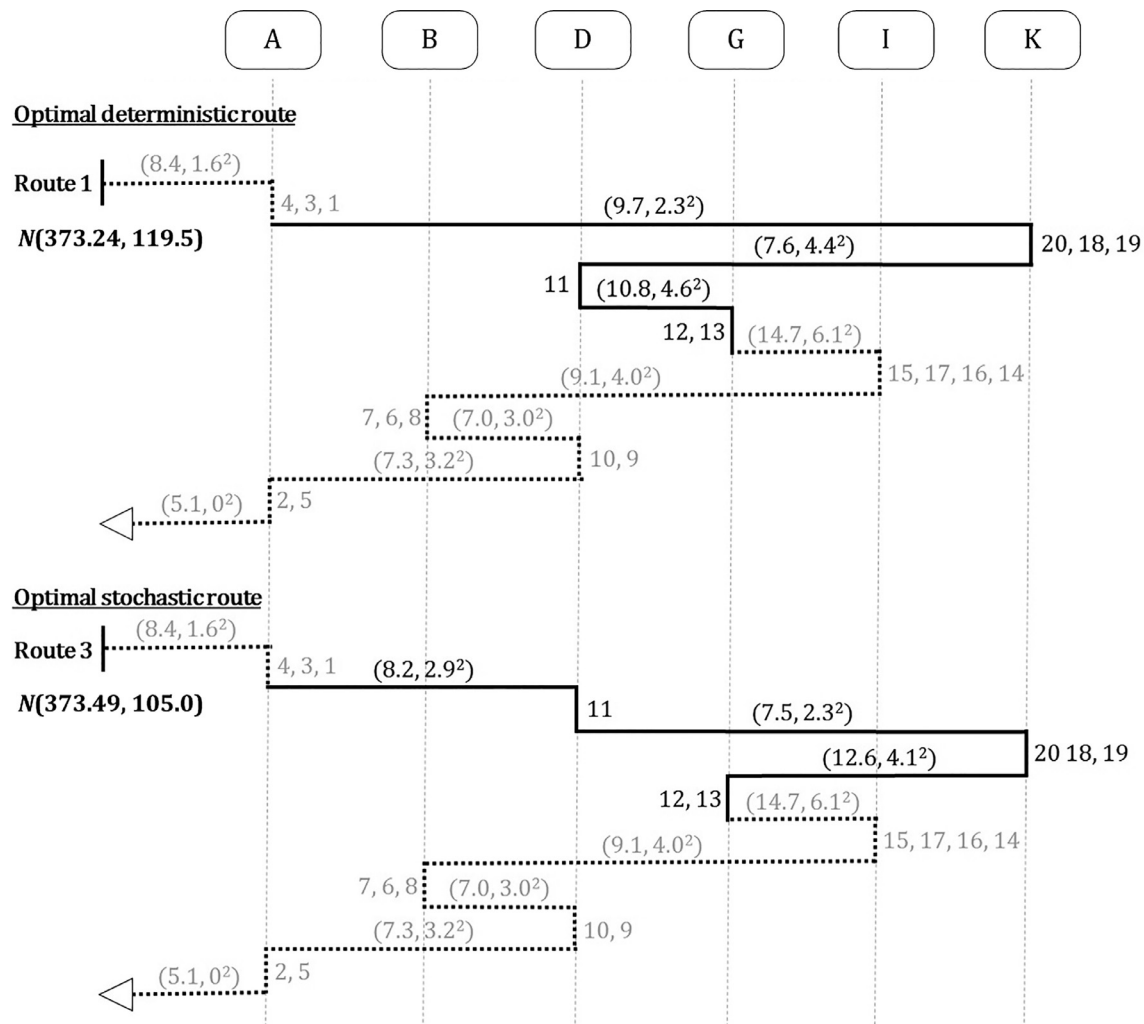


Fig. 12. The terminal and cargo visiting sequence of the deterministic and stochastic solutions to instance Even-46.

terminal layout, we generate one additional set of instances where the anchorage is closer to the terminals. This corresponds to the fact that currently new anchorages, layberths and other lay-by facilities that are closer to the channel (e.g., a mid-bay anchorage inside Galveston Bay) are being actively developed to address the scheduling issues in Houston and to provide better services for chemical ship operators (see Kruse (2015) for more details). This may also represent a situation faced in some other port having a similarly narrow ship channel but with the possibility for the ships to park closer to the terminals therein. To achieve this we simply modify every instance in the *Even* set by shortening the distance between Terminal A and the anchorage from 5.12 hours to 1 hour, see Fig. 13 for an approximated illustration. We call the resulting set of 100 new instances the *Even-1h* set.

To evaluate the impact of using a stochastic model rather than its deterministic counterpart, it is common to compute the Value of Stochastic Solution (VSS) for a stochastic problem, see Birge (1982) and Maggioni and Wallace (2012) for more details. However, since our deterministic and stochastic models have different types of objectives, the computation of VSS might not be as meaningful as for typical stochastic programming problems. We therefore propose the following alternative measures. For each instance set, we solve all 100 instances and record the number of instances for which the deterministic and stochastic solutions are different. Furthermore, for every pair of deterministic and stochastic solutions that are not the same, we introduce a SD/Mean ratio to

Table 5

Results of solving four sets of instances *Even*, *Far*, *Split* and *Even-1h* showing two measures: percentage of instances that have different deterministic and stochastic solutions, and average SD/Mean ratio for these particular instances.

Instance set	<i>Even</i>	<i>Far</i>	<i>Split</i>	<i>Even-1h</i>
No. of instances	100	100	100	100
% diff determ & stoch	18%	20%	22%	56%
Average SD/Mean ratio	1.24	1.51	1.99	2.29

indicate how much the stochastic solution is able to, compared with its deterministic counterpart, reduce the standard deviation of the route relative to the resulting increase in route mean. For example, the deterministic and stochastic solutions for the instance *Even-46* (see Fig. 12) correspond to $N(373.24, 119.5)$ and $N(373.49, 105.0)$, respectively, and the SD/Mean ratio in this case is calculated as $(10.93-10.25)/(373.49-373.24) = 2.72$, where 10.93 and 10.25 are the square roots of 119.5 and 105.0, respectively. The results are presented in Table 5 showing two measures, “percentage of instances that have different deterministic and stochastic solutions” and “average SD/Mean ratio for these particular instances”, for four sets of instances *Even*, *Far*, *Split* and *Even-1h*.

From Table 5 we see marginal increases in the numbers of instances having different deterministic and stochastic solutions for the *Far* and *Split* settings, compared with the *Even* setting. The

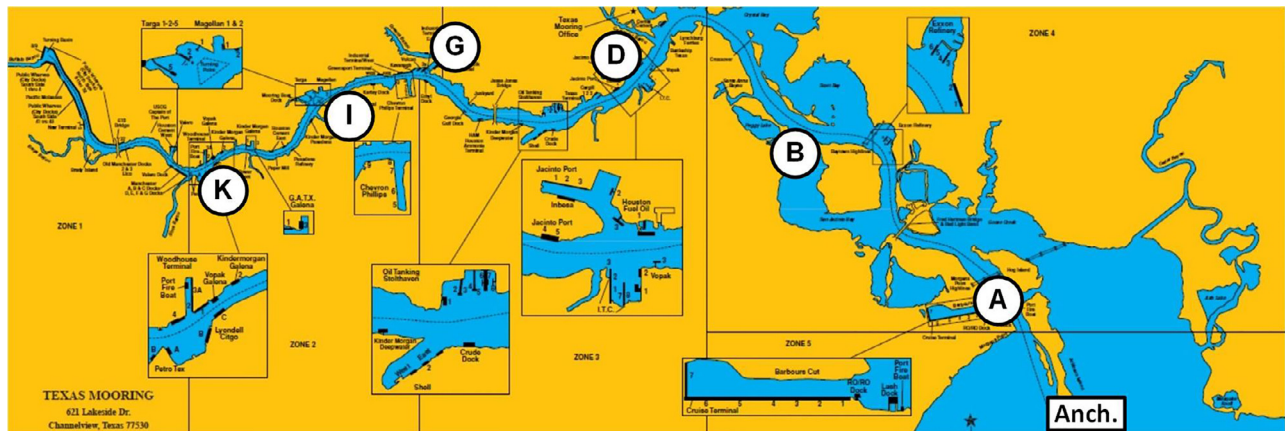


Fig. 13. The *Even-1h* layout, where the distance between Terminal A and the anchorage is shortened to 1 hour.

increases in average SD/Mean ratio are larger, e.g., 1.99 for the *Split* set in contrast to 1.24 for the *Even* set. In general, we may conclude that the layout of the terminals has certain but limited impact on the two measures; and when the terminals both close to the outlet and deep inside the channel are to be serviced (hence the *Split* layout), taking route variance into account is more likely to be beneficial.

When the anchorage is closer to the terminals (the *Even-1h* setting), we see significant increases in both measures. Benchmarked with the original *Even* set, the number of instances having different deterministic and stochastic solutions has more than tripled from 18 to 56, and the average SD/Mean ratio also almost doubled from 1.24 to 2.29. This confirms our observation that the relative distance from the anchorage to the terminals is an important factor in the in-port routing problem discussed in this study. When a shipping company is facing a similar routing problem where the anchorage is close to the terminals, e.g., at Port of Houston with a new anchorage location or at some other port, it is important to examine the alternatives to the smallest-mean route as there may be some route with a slightly increased mean but much lower variance.

5.4. Evaluation of the Monte Carlo approximation

Finally, we examine the validness of the normality assumption for a complete route and of using Monte Carlo estimates for the means and variances of the travel times. This is to evaluate what and how much is lost by using standard normal distributions to approximate the skewed distributions with “leftmost bars” as shown in Fig. 8.

We use as examples Routes 1 and 3 of instance *Even-46* as illustrated in Fig. 12. We have shown that by summing up the mean and variance (computed through Monte Carlo approximation) of every individual transit along the route, we can obtain the approximated normal distributions for the two routes, $N(373.24, 119.5)$ and $N(373.49, 105.0)$ for Route 1 and Route 3, respectively. However, to get a feel for their “real” completion time distributions, we now perform whole-route simulations for these two routes.

Take Route 1 for instance. This route, i.e., [Anch. → A → K → D → G → I → B → D → A → Anch.], consists of nine transits. At each run of the whole route simulation, we follow the route, sequentially draw the random waiting time and calculate the corresponding travel time for each transit, and record the final completion time of the route. We run the simulation in this way 10,000 times, and obtain the same number of outcomes for the final completion time. In Fig. 14a these 10,000 outcomes are shown as a histogram where the x- and y-axes are completion

Table 6

Measuring the mean, variance and skewness of the whole-route simulation results, compared with approximations.

	Simulation		Approximation	
	Route 1	Route 3	Route 1	Route 3
Mean	373.37	373.58	373.24	373.49
Variance	150.3	130.2	119.5	105.0
Skewness	−0.14	−0.21	0	0

time and frequency, respectively. By comparing the histogram to the approximated standard normal distribution (shown as the bell curve in Fig. 14a), we see that the approximation is quite good and has described the “true” distribution well. The whole-route simulation is also done for Route 3 and a similar conclusion may be drawn from the resulting diagram shown in Fig. 14b.

We then show in Table 6 the results of measuring the mean, variance and skewness of the simulation outcomes for the two routes (in the Simulation column), and compare them with their “theoretical” counterparts derived from approximation (in the Approximation column). We see that the difference in the variance for the same route between simulation and approximation is still quite significant, e.g., for Route 1 the sample variance of the 10,000 outcomes is 150.3 while the approximation only gives an estimation of 119.5. Such underestimation of route variance comes from the fact that our approximation procedure disregards the skewness and those “leftmost bars” when generating the Monte Carlo estimates for individual transits. It is, however, important to emphasize that using the approximated distributions still gives us the fair comparison across different routes, e.g., Route 3 having significantly lower variance but slightly increased mean in comparison with Route 1. This observation also applies to other instances in our experiments. Also, based on the whole-route simulation results in Table 6 the SD/Mean ratio (decrease in standard deviation relative to increase in mean) given by choosing Route 3 over Route 1 is 4.05, in contrast to 2.74 based on the approximation results. This is also a strong indication that the benefit of sailing the stochastically obtained route may be larger in practice than what we have shown in this study in theory using approximation.

Therefore, despite its underestimation on the variance of an individual route, the Monte Carlo approximation provides fair and reliable indications with respect to the mean-variance trade-off across different routes. This correct prediction of the mean-variance trade-off may also reflect the robustness of the problem with respect to solution methods, i.e., other approximation methods might also indicate such mean-variance trade-off even the

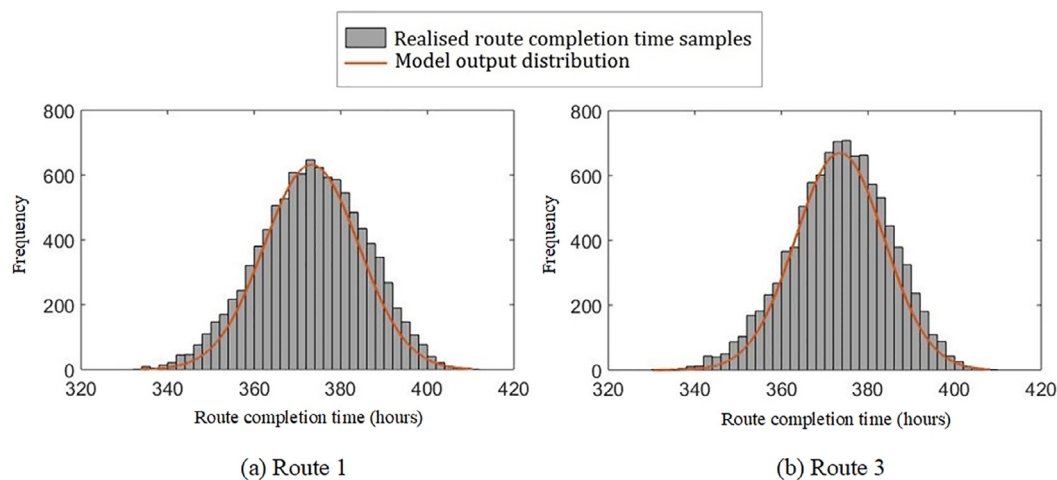


Fig. 14. Results of the whole-route simulation for two routes showing the goodness of fit of the completion time realizations to normal distribution.

methods are not accurate by themselves. In practice, this is a good problem attribute (though not obvious at all); it means that the decision maker can safely put aside the skewness and “leftmost bars” (as in Fig. 8) of the travel time distributions to find a set of good routes, and then take the skewness and leftmost bars back into consideration by means of a final re-evaluation among the set of routes to make a final decision. More specifically, we suggest that one starts with the proposed approximation method in order to find the theoretically smallest-mean route (deterministic solution) as well as its alternatives. Then, we suggest that one re-evaluates these routes by running the whole-route simulation to obtain a more accurate comparison, in order to choose the one optimal route for the shipping company based on its chosen threshold.

6. Conclusion

In this paper we have introduced a ship routing problem with uncertain travel times that arises in the chemical shipping industry. Based on the study of Arnesen et al. (2017), we have proposed a stochastic TSPDP model that takes in consideration the uncertain travel times for in-port transits and the special sailing pattern in the Port of Houston, in which the objective is finding the route which maximizes the probability that its total length does not exceed a given threshold. We solve the problem by adapting the binary search enumeration method proposed in Nikolova et al. (2006).

Several sets of test instances are generated on a random basis representing different cases faced in Port of Houston, including three types of terminal layout (*Even*, *Far* and *Split*) and also one additional situation where the anchorage is closer to the terminals (*Even-1h*) which may also arise in reality. Assuming independent normally distributed travel times, Monte Carlo simulation is used to estimate the means and variances of the distributions for the individual transits when generating the instances. We show that it is important to properly address the special sailing pattern, and that the actual travel time for an in-port transit is simultaneously determined by the distance between the origin and destination terminals, the waiting time uncertainty in the destination terminal and the relative location of the anchorage. This combined effect cannot be captured, in any way, by assuming that the ship can wait “right outside” the terminal even if the uncertainty of such waiting time is taken into account.

Our computational study also shows that in the case of Odfjell operating in the Port of Houston, the deterministic solution,

obtained by utilizing only the means (without variances) of the travel times, is as good as the stochastic solution under most circumstances. However, we show that in cases when the anchorage is closer to the terminals, it is important to examine the alternatives to the smallest-mean (deterministic) route since there may be some route with a slightly increased mean but much lower variance, which may be particularly valuable for some companies, as it benefits the planning of the voyage to the next port if the company knows with higher certainty how long the current port call will take.

By evaluating the effects of the approximation which encompasses the usage of Monte Carlo estimates for the means and variances of travel times, we show that although the normality assumption for a complete route holds, the approximation procedure underestimates the route variance. However, the comparison across different routes based on the approximated distributions still gives the correct indication. Therefore, we suggest that one should first use the approximation method and the model proposed in this study to find the smallest-mean route and its alternatives, and then re-evaluate them by running whole-route simulations.

Many perspectives remain open with respect to this study. First, in this paper we assume that the tanker only send the NOR (Notice of Readiness) to one terminal after servicing a cargo, whereas in practice the tanker sometimes send out NOR to up to three different terminals and only sail to the one that responds first. Therefore a possible future study is to find the optimal strategy regarding the selection of terminals to send the NOR after servicing a cargo, in the light of different uncertain waiting time distributions for the terminals. Second, apart from the exact method proposed by Nikolova et al. (2006) used in this paper, there are other approaches in the literature for solving similar problems, e.g., Cheng and Lisser (2015) show that the maximum probability shortest path problem can be reformulated and solved as a chance constrained problem. A comparison of the two methodologies can then be part of another possible future study.

Acknowledgments

The authors acknowledge financial support from the project “Green shipping under uncertainty (GREENSHIPRISK)” partly funded by the Research Council of Norway under grant number 233985. We are also grateful to Odfjell that provided real data for the test instances used in the computational study.

References

- Ando, N., & Taniguchi, E. (2006). Travel time reliability in vehicle routing and scheduling with time windows. *Networks and Spatial Economics*, 6(3–4), 293–311.
- Arnesen, M. J., Gjestvang, M., Wang, X., Fagerholt, K., Thun, K., & Rakke, J. G. (2017). A traveling salesman problem with pickups and deliveries, time windows and draft limits: Case study from chemical shipping. *Computers & Operations Research*, 77, 20–31.
- Arnsdorf, I., & Murtaugh, D. (2014). Big ships play texas chicken in congested houston channel. <https://www.bloomberg.com/news/articles/2014-02-27/houston-ship-channel-congested-by-u-dot-s-dot-oil-and-gas-boom>. Accessed 1 Sep 2017.
- Berbeglia, G., Cordeau, J. F., Gribkovskaia, I., & Laporte, G. (2007). Static pickup and delivery problems: a classification scheme and survey. *TOP*, 15(1), 1–31.
- Bertazzi, L., & Maggioni, F. (2015). Solution approaches for the stochastic capacitated traveling salesman location problem with recourse. *Journal of Optimization Theory and Applications*, 166(1), 321–342.
- Birge, J. R. (1982). The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming*, 24(1), 314–325.
- Carraway, R. L., Morin, T. L., & Moskowitz, H. (1989). Generalized dynamic programming for stochastic combinatorial optimization. *Operations Research*, 37(5), 819–829.
- Chang, T. S., Wan, Y. W., & OOI, W. T. (2009). A stochastic dynamic traveling salesman problem with hard time windows. *European Journal of Operational Research*, 198(3), 748–759.
- Cheng, J., & Lisser, A. (2015). Maximum probability shortest path problem. *Discrete Applied Mathematics*, 192, 40–48.
- Christiansen, M., Fagerholt, K., Nygreen, B., & Ronen, D. (2007). Chapter 4 maritime transportation. In C. Barnhart, & G. Laporte (Eds.), *Handbooks in operations research and management science*: Vol. 14 (pp. 189–284). Elsevier.
- Gendreau, M., Laporte, G., & Séguin, R. (1996). Stochastic vehicle routing. *European Journal of Operational Research*, 88(1), 3–12.
- Gómez, A., Mariño, R., Akhavan-Tabatabaei, R., Medaglia, A. L., & Mendoza, J. E. (2016). On modeling stochastic travel and service times in vehicle routing. *Transportation Science*, 50(2), 627–641.
- Hammer, H. (2013). The chemical tanker market. Master's thesis Norwegian School of Economics.
- Jula, H., Dessouky, M., & Ioannou, P. (2006). Truck route planning in nonstationary stochastic networks with time windows at customer locations. *IEEE Transactions on Intelligent Transportation Systems*, 7(1), 51–62.
- Kao, E. P. C. (1978). A preference order dynamic program for a stochastic traveling salesman problem. *Operations Research*, 26(6), 1033–1045.
- Kruse, C. J. (2015). Analysis of chemical tanker transits on the houston ship channel. Technical report Texas A&M Transportation Institute.
- Lei, H., Laporte, G., & Guo, B. (2012). A generalized variable neighborhood search heuristic for the capacitated vehicle routing problem with stochastic service times. *TOP*, 20(1), 99–118.
- Li, X., Tian, P., & Leung, S. C. (2010). Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. *International Journal of Production Economics*, 125(1), 137–145.
- Maggioni, F., & Wallace, S. W. (2012). Analyzing the quality of the expected value solution in stochastic programming. *Annals of Operations Research*, 200(1), 37–54.
- Malaguti, E., Martello, S., & Santini, A. (2018). The traveling salesman problem with pickups, deliveries, and draft limits. *Omega*, 74, 50–58.
- Nikolova, E. (2010). High-performance heuristics for optimization in stochastic traffic engineering problems. In I. Lirkov, S. Margenov, & J. Wasniewski (Eds.), *Proceedings of the large-scale scientific computing (LSSC)*. In *Lecture notes in computer science*: Vol. 5910 (pp. 352–360). Berlin Heidelberg: Springer.
- Nikolova, E., Kelner, J. A., Brand, M., & Mitzenmacher, M. (2006). Stochastic shortest paths via quasi-convex maximization. In Y. Azar, & T. Erlebach (Eds.), *Proceedings of the algorithms - ESA*. In *Lecture notes in computer science*: Vol. 4168 (pp. 552–563). Berlin Heidelberg: Springer.
- Oyola, J., Arntzen, H., & Woodruff, D. L. (2016). The stochastic vehicle routing problem, a literature review, part I: models. *EURO Journal on Transportation and Logistics*. doi:10.1007/s13676-016-0100-5.
- Perboli, G., Gobbato, L., & Maggioni, F. (2017). A progressive hedging method for the multi-path travelling salesman problem with stochastic travel times. *IMA Journal of Management Mathematics*, 28(1), 65–86.
- Rakke, J. G., Christiansen, M., Fagerholt, K., & Laporte, G. (2012). The traveling salesman problem with draft limits. *Computers & Operations Research*, 39(9), 2161–2167.
- Russell, R. A., & Urban, T. L. (2008). Vehicle routing with soft time windows and erlang travel times. *Journal of the Operational Research Society*, 59(9), 1220–1228.
- Sniedovich, M. (1981). Analysis of a preference order traveling salesman problem. *Operations Research*, 29(6), 1234–1237.
- Ta, D., Dellaert, N., van Woensel, T., & de Kok, T. (2013). Vehicle routing problem with stochastic travel times including soft time windows and service costs. *Computers & Operations Research*, 40(1), 214–224.
- UNCTAD (2016). Review of maritime transport 2016. In *Proceedings of the United Nations conference on trade and development*. Geneva, Switzerland.
- Wang, X., Arnesen, M. J., Fagerholt, K., Gjestvang, M., & Thun, K. (2018). A two-phase heuristic for an in-port ship routing problem with tank allocation. *Computers & Operations Research*, 91, 37–47.
- Woensel, T. V., Kerbache, L., Peremans, H., & Vandaele, N. (2007). A queueing framework for routing problems with time-dependent travel times. *Journal of Mathematical Modelling and Algorithms*, 6(1), 151–173.
- Yan, S., Wang, S. S., & Chang, Y. H. (2014). Cash transportation vehicle routing and scheduling under stochastic travel times. *Engineering Optimization*, 46(3), 289–307.
- Zhang, J., Lam, W. H. K., & Chen, B. Y. (2013). A stochastic vehicle routing problem with travel time uncertainty: Trade-off between cost and customer service. *Networks and Spatial Economics*, 13(4), 471–496.
- Zhang, T., Chaovalitwongse, W., & Zhang, Y. (2012). Scatter search for the stochastic travel-time vehicle routing problem with simultaneous pick-ups and deliveries. *Computers & Operations Research*, 39(10), 2277–2290.