Annals of Operations Research
https://doi.org/10.1007/s10479-018-3042-5

S.I. : BALCOR-2017

Solving the petroleum replenishment and routing problem with variable demands and time windows

Yan Cheng Hsu[1] · **Jose L. Walteros**[1] · **Rajan Batta**[1]

Abstract
In this paper we develop a methodological framework for designing the daily distribution and replenishment operations of petroleum products over a weekly planning horizon by taking into account the perspectives of both the transporter and its customers. The proposed approach considers the possibility of having late deliveries due to the variability of the customers' demands and expected time windows. We first develop an inventory model for the customers to identify the optimal order quantities and time windows. Then, we solve a sequence of mixed-integer optimization models for designing distribution routes based on the order quantities and time windows selected by the inventory models. We design the optimization models so that late deliveries are balanced among the customers in order to mitigate the overall customer dissatisfaction. We test the proposed approach by solving a set of instances adapted from the literature. The empirical results show that the proposed approach can be used for designing the distribution plan for delivering petroleum products in conditions where the operational capabilities of the transporter are limited for generating optimal on-time plans.

Keywords Gasoline distribution · Vehicle routing with time windows · Vehicle routing and scheduling

1 Introduction

The transportation industry plays a critical role in today's global economy fostering the operations of nearly all other industries around the world. Alone in the U.S., according to the U.S. Federal Highway Administration, about 20 billion tons of goods worth more than \$10 trillion were moved across the country just in 2012 (Strocko et al. 2014). Transportation

✉ Jose L. Walteros
josewalt@buffalo.edu

Yan Cheng Hsu
yhsu8@buffalo.edu

Rajan Batta
batta@buffalo.edu

[1] Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, 342 Bell Hall, Buffalo, NY 14260, USA

Published online: 17 September 2018

related goods and services represented approximately 11% of the U.S. gross domestic product in 2000, only being surpassed by the housing, health-care, and food industries (Cohn et al. 2007; Nguyen et al. 2015).

Among the total goods transported in the U.S. in 2012, more than 1.8 billion tons corresponded to gasoline, diesel, and other petroleum-based products, thus becoming the top sixth most transported commodities in the country (Strocko et al. 2014). Petroleum products are still one of the world's most traded commodities, as they continue being the main energy source for the transportation industry. According to the U.S. National Academy of Sciences (Committee on America's Energy Future, National Academy of Sciences 2009), petroleum-based fuels represent about 98% of the energy sources used for mobilizing both people and freight in the U.S. In addition to their use as fuel for both commercial and personal transportation, the need for petroleum products as lubricants, transmission and hydraulic fluids, and as raw materials for many production processes, has made of them ubiquitous and vital for the daily operations of every country.

In order to provide a steady supply of petroleum products that satisfies the growing demand, many suppliers face the highly complex process of planning the routing strategies for distributing these products to their customers, which include gasoline retailers (gas stations), production companies, agricultural companies, and marine centers, among others. This process generally requires the allocation of multiple competing resources (multi-compartmented trucks) while simultaneously satisfying many operational restrictions and regulatory policies [the transportation of gasoline and other petroleum-based products is highly regulated by the Federal Motor Carrier Safety Administration (FMCSA) in the U.S. (Federal Motor Carrier Safety Administration 2015)]. As a result, since publication of the first paper on gasoline distribution (Dantzig and Ramser 1959), there has been a continuous effort to develop quantitative models that support the decision making of petroleum transportation companies.

Over the last few decades numerous solution approaches have been developed to tackle vehicle routing problems that incorporate the distribution requirements of many industries (e.g., Avella and Sforza 2004; Ben Abdelaziz et al. 2002; Campbell and Savelsbergh 2004; Cornillier et al. 2007; Toth and Vigo 2014; van der Bruggen et al. 1995). However, despite the large number of such approaches, there are still several elements of the distribution logistics of some products that require a further analysis. One fundamental assumption that is often considered by many of these technologies is that, given a set of customer requirements in the form of order quantities and delivery time windows, there always exist a distribution plan that is able to satisfy all of such requirements under the transporter operational restrictions (e.g., fleet size, driver schedule, transportation regulations). Yet, scenarios in which the operational capabilities of the transporter may fail to meet the delivery requirements are rarely considered.

In a real-life scenario, the decisions regarding the product orders and the distribution logistics are often made independently and sequentially by the customers and the transporter without any interaction between them during the distribution planning process (Toth and Vigo 2014). In other words, the order quantities and expected delivery time windows—which are selected by each customer solely based on the customer's own interests—are given to the transporter in the form of hard constraints. Then, after collecting all the orders, the transporter makes a routing plan aiming to fulfill all the requirements while minimizing its own operational costs (Andersson et al. 2010; Toth and Vigo 2014). Many mathematical models assume that there are always feasible routes that satisfy such demands and time requirements of all the customers. In contrast, in many competitive markets like the one of petroleum products distribution—in which the transportation decisions must also meet strict regulatory policies—finding feasible solutions that cope with all customer and governmental requisites is often impossible. In other words, due to capacity limitations and further operational requirements, the actual delivery time may deviate from the desired time window for many customers.

For the specific case of petroleum products, there are several reasons why optimally planning the distribution logistics is a complex challenge. In addition to the variable nature of the demands, the limited number of trucks and drivers, the regulatory policies, the difficulties posed by the inherent characteristics of the products (i.e., mostly flammable liquids that must be transported in specialized multi-compartmented trucks and trailers), the heterogeneity of the customers, and the strict time requirements make the problem of identifying optimal delivery routes especially difficult. Moreover, because of the replenishment logistics, most customers request very specific delivery time windows that often overlap among them (e.g., retailers often prefer having replenishments late at night when traffic is low, whereas other customers prefer early morning deliveries before any operation begins). These latter requirements dramatically impact the complexity of the distribution planning, up to the point that even finding feasible delivery schedules is rather challenging.

One approach that has been consistently utilized in the literature to alleviate the aforementioned problems is synchronizing the inventory management with the transportation of commodities (Andersson et al. 2010; Coelho et al. 2013). This allows transporters to manage and control both the transportation and the inventory decisions. These inventory routing models can produce several advantageous strategies for both transporter and customers (i.e., win-win situations), if the customers are willing to delegate the inventory management to the transporter (Andersson et al. 2010). Nevertheless, there are several cases in which these types of customer-transporter inventory agreements are not necessarily implementable or perhaps undesirable for one of the parties.

Establishing inventory management agreements might not necessarily be beneficial for both the transporter and customers in various contexts (Andersson et al. 2010). For instance, if the inventory turnover of a customer is too erratic to be accurately estimated, the transporter may want to refrain from entering into a coordination agreement with such a customer because of the potential risk of provoking major disruptions to the distribution schemes, possibly causing inventory shortages to other customers. Similarly, it is also possible that some customers may want to retain full control of the inventory levels. A typical example corresponds to the case in which the commodity being distributed represents a large monetary investment for the customers, and they require to keep a close track of their expenses. In the context of petroleum products, this happens quite often with family-operated gasoline stations that cannot maintain open accounts with the distributors.

The mathematical models that will be investigated in this paper are thus aimed to tackle scenarios in which the use of these integrated inventory routing models is not viable.

1.1 Relevant literature

The *vehicle routing problem* (VRP) is at the cornerstone of most distribution planning processes. Since the publication of the first paper in the subject in the late 50s (Dantzig and Ramser 1959) (interestingly, a paper about gasoline distribution), a staggering number of studies have been developed to tackle many variants of this problem [for further references see the following comprehensive surveys (Cordeau et al. 2007; Eksioglu et al. 2009; Toth and Vigo 2014)]. This problem and its variations have continuously raised the interest of the academic community because of their practical relevance and inherent difficulty. In fact, several methodological advancements in the field of Operations Research have been discovered by studying routing problems (Cordeau et al. 2007).

Among all the variations that can be found in the literature, the ones that are relevant to this paper are the VRP with time windows (VRPTW), where each customer must be visited during a specific time frame (Baldacci et al. 2012; Bräysy and Gendreau 2005a, b; Gendreau and Tarantilis 2010; Kallehauge et al. 2005); the VRP with multiple compartments (MCVRP), where the vehicles have different capacities and are equipped with multiple compartments that can carry more than one type of product (Al-Khayyal and Hwang 2007; Caramia and Guerriero 2009; Chajakis and Guignard 2003; Christiansen et al. 2011; Derigs et al. 2011; Fallahi et al. 2008; Henke et al. 2015; Jetlund and Karimi 2004; Lahyani et al. 2015; Muyldermans and Pang 2010; Reed et al. 2014); the VRP with stochastic demands (SVRP), where the demands are given by a probability distribution (Bertsimas 1992; Bianchi et al. 2006; Christiansen and Lysgaard 2007; Sungur et al. 2008; Mendoza et al. 2010); the dynamic VRP (DVRP), where the information of some customers becomes available during operation (Pillac et al. 2013); the time window assignment VRP (TWAVRP), which is a variation where time windows have to be assigned by the transporter to the customers before the demand of product given by them is known (Spliet and Desaulniers 2015; Spliet and Gabor 2015).

In addition to the aforementioned problems regarding vehicle routing, *inventory routing problems* (IRPs)—as previously mentioned—represent an alternative approach that is often used to plan the distribution logistics of several products (Andersson et al. 2010; Campbell et al. 1998; Coelho et al. 2013; Cohn et al. 2007; Federgruen and Simchi-Levi 1995; Moin and Salhi 2006). In general, the IRP is a product distribution problem in which one actor—the *manager*—is responsible for both transportation and inventory planning (Coelho et al. 2013). In practice, the manager can either be the producer, the consumer, or the transportation company depending on the type of business. When the managers are the producers or the transportation companies—which is often the case—these integrated policies allow them to select the timing and sizes of the deliveries. For these models to be applicable though, the customers must render complete knowledge of their operational needs and full control over their inventory levels to the manager. In turn, the manager must ensure that the customers will never run out of stock. Nevertheless, in the context of petroleum products that we tackle in this paper, the operational decisions of many customers often require them to maintain full control of the inventory levels, which complicates the use of these latter models. Thus, the main difference between the problem studied in this paper and the IRP is that in the latter suppliers control the inventory management of the customers, whereas in the former, the inventory problem is solved by each customer and the results are given as hard constraints to the transporter.

For the specific context of delivering petroleum products, the first publications that provide specific applications of this kind date back to the 50s (Dantzig and Ramser 1959), 80s (Brown et al. 1987; Brown and Graves 1981) and 90s (Bausch et al. 1995; van der Bruggen et al. 1995). Most of the solution approaches proposed in these papers range between heuristic and exact approaches. For instance, Avella and Sforza (2004) formulated a fuel delivery problem as a set partitioning model and proposed a branch-and-price algorithm to solve the resulting problem—a technique widely used for solving VRPs. In addition to that, Ng et al. (2007) presented a case study on the delivery networks in Hong Kong, which contains tanker assignment and a routing problem with a heterogeneous fleet of compartmented trucks. A decision support system (DSS) approach was developed to solve the vendor managed inventory (VMI) problem. In Cornillier et al. (2007), an exact algorithm was proposed to tackle the single period and single depot case using an unlimited heterogeneous fleet of compartmented tank trucks petroleum replenishment problem. A heuristic for the multi-period and single depot with limited number of trucks was proposed in Cornillier et al. (2008). The same problem with time window constraints was tackled further in Cornillier et al. (2009). More recently, Cornillier et al. (2012) proposed heuristics for the multi-depot station replenishment problem considering time windows, in which the concept of a trip—defined by both a route and the truck used to make deliveries in this route—was firstly introduced to address this problem. Instead of generating possible routes, they introduced a method to generate potential feasible trips. Similarly, as for other VRP variants, the time windows given by the customers are assumed to be fixed and no further considerations are proposed for the cases where the given time windows render the problem infeasible.

1.2 Contributions

This paper aims to develop a methodological framework for designing the daily distribution operation for delivering petroleum products that considers the possibility of having late deliveries due to the variability of the customers' demands and expected time windows. In addition to maximizing the distribution profit, the proposed framework attempts to minimize the dissatisfaction of the customers due to late deliveries by balancing the late deliveries among the customers over the planning horizon. The contributions of this paper can be summarized as follows:

- We propose a methodological framework to solve the daily petroleum distribution problem considering both transporter's and customers' perspectives and run it for a weekly time horizon.
- We develop an inventory problem that models all the scenarios regarding the delivery times for the product orders. This model is used to determine the order quantities and time windows for each customer.
- We propose a sequence of mixed-integer optimization models for designing the distribution routes based on the order quantities and time windows selected by the inventory models.
- We tailor the optimization models so that late deliveries are balanced among the customers in order to mitigate the overall customers' dissatisfaction throughout the time horizon.
- We combine the proposed optimization models with a powerful variable neighborhood search (VNS) to efficiently improve the quality of the distribution routes.
- We test the proposed approach by solving a test bed of instances adapted from the literature. The empirical results show that the proposed approach can be used for designing the distribution plan for delivering petroleum products in conditions where the operational capabilities of the transporter are limited for generating optimal on-time plans.

The remainder of this paper is organized as follows. Section 2 presents a detailed description of the problem at hand; Sect. 3 summarizes the proposed sequential solution approach; Sect. 4 introduces the inventory model that is used to model the gas station decisions; Sect. 5 presents the proposed mathematical formulations and overall solution to generate the distribution plan; Sect. 6 analyzes the results of the empirical study; and finally, Sect. 7 provides the final conclusions and further research directions.

2 Problem description

The petroleum replenishment problem deals with the logistics of delivering petroleum products to a set of customers—gas stations in this case—so that the requirements of such customers in the form of delivery quantities and time windows are fulfilled under the operational capabilities of the transporter, while maximizing the total distribution profits yielded by the distribution operation.

From the perspective of the customers, during their daily operations, gas stations periodically review their underground tanks and place order requests to the suppliers when the stock levels of their products fall below predefined thresholds. The orders are typically placed a day before the delivery date and consist of a list of product type requests that include the desired quantities and delivery time windows, as shown in Fig. 1. The supplier then collects this information from all of its customers to generate: (1) the truck loading, (2) the delivery routes, and (3) truck distribution schedule, in order to fulfill these orders.

From the perspective of the supplier, the distribution network is defined as follows. Let $G = (N, A)$ be a directed graph where $N = \{0, 1, 2, .., n\}$ is a set of nodes representing the distribution terminal (node 0) and the gas stations (nodes $1, .., n$), and $A = \{(i, j) : i \neq j$ and $i, j \in N\}$ is the set of arcs that represents the road segments connecting the nodes in N. We denote c_{ij} and t_{ij} as the travel costs and travel times associated with the arc (i, j), s_i as the service time of gas station i, and q_i as the order quantity requested by station i. The time window $[a_i, b_i]$ specifies the earliest and latest time limits for performing the petroleum replenishment, i.e., the delivery must occur within the given time window $[a_i, b_i]$. A station requesting more than one type of product can be modeled as multiple stations having the same location and different order quantities and time windows, according to the specific requirements of each product. This permits modeling the inventory problems of all the products independently from each other, as the demand rates may differ significantly (e.g., low octane gasoline has generally a higher demand compared to diesel and high octane gasoline).

Each truck is divided into multiple compartments with known capacities, which are used to upload different types of products. In other words, two distinct grades of petroleum must be placed into two separate compartments to avoid internal contamination. Furthermore, the petroleum stored in each compartment must be fully pumped out when fulfilling the distribution service, as the quality of any remaining petroleum will begin to deteriorate on contact with air. Therefore, if the underground tanks of a gas station fail to accommodate a full-compartmented load of petroleum, the remainder must be sent back to the terminals, resulting in a send-back cost—which is generally high in comparison to other costs. All trucks should begin and end at the terminal and the travel speed of those are considered to be the same. Thus, the petroleum replenishment problem consists of determining:

(1) the quantity and time window of delivery for each gas station;
(2) the loading of the various petroleum products into the truck compartments;
(3) the delivery routes to the gas stations;
(4) the departure time of each truck from the terminal and the arrival time at each of its assigned customers.

In addition, the objective of the distribution problem is twofold: to minimize the expected total costs for gas stations, and to maximize the overall distribution profits for the transporter. In this paper we decouple the petroleum delivery problem into two parts: the gas station inventory problem (i.e., part (1) of the above list) and the transporter distribution problem [i.e., parts (2)–(4)]. First, for the gas station inventory problem, we use a model to determine the order quantity and delivery time window for each gas station and second, for the transporter's distribution problem we propose a sequence of mixed-integer formulations to determine the distribution plan. The description of the proposed framework is given in the following section.

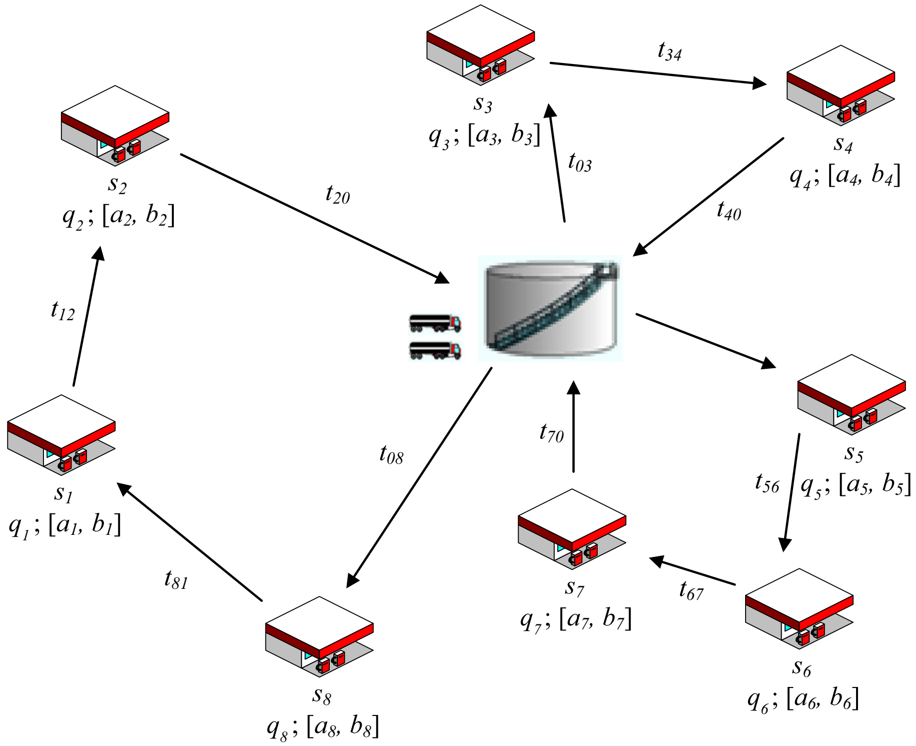


Fig. 1 Distribution network

3 Solution framework

The petroleum replenishment problem deals with both the inventory problem of the gas stations and the distribution problem of the transporter on a daily basis. For each day of the planning horizon, the order quantity and requested time window for each gas station will be identified by an inventory model that is aimed to minimize the expected total costs perceived by the gas stations. Such costs include ordering cost, holding cost, shortage cost, and send-back cost (see Sect. 4). The proposed model takes into consideration that the exact delivery times are not known a priori by the gas stations. Therefore, the order quantities and desired time windows are decided based on an estimation of the delivery times.

Once the order quantity and desired delivery time window are selected by each gas station using the proposed inventory model, we then solve a series of interrelated mixed-integer linear programs that will determine how to load these demands into truck compartments, how to schedule the truck departure and returning times, and how to deliver these demands (see Sect. 5). By solving these models, the petroleum distribution plan for each day can be identified. In order to improve the quality of the distribution routes, these models are coupled with a variable neighborhood search (VNS) procedure.

In the proposed approach, we solve the replenishment problem sequentially day by day. Thus, the resulting distribution plan obtained for each day of the planning horizon is then used to calculate the inputs of the inventory models of the subsequent days. Notice that, after solving the distribution problem of a given day, the actual delivery time for each station can

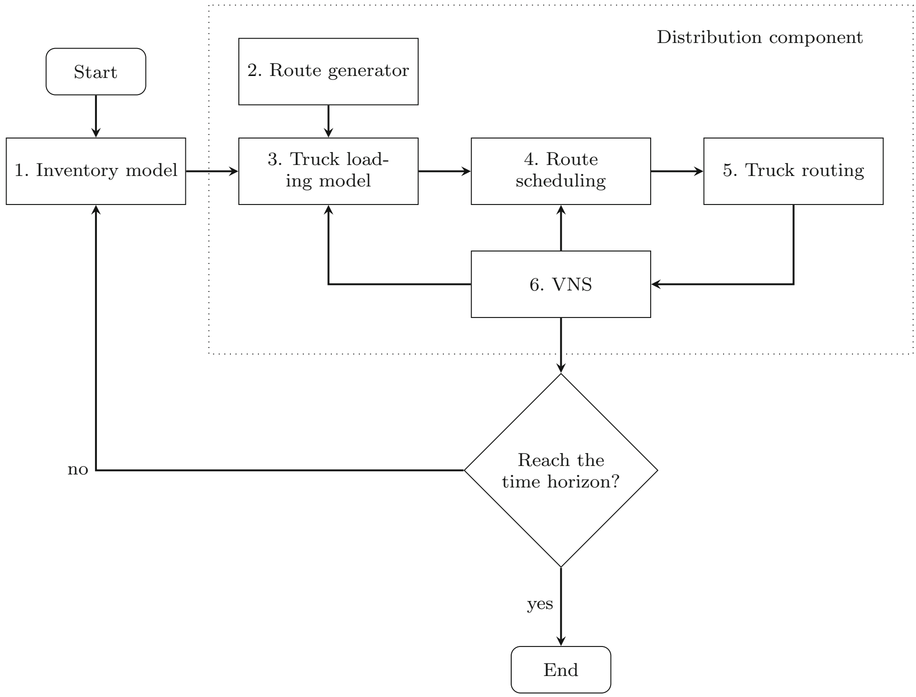


Fig. 2 Solution procedure

be used to compute the initial stock level for each station for the day after and thus, the order decision can be made accordingly. The solution process continues until the considered time horizon is reached. In this section, the steps of the proposed approach, which are presented in Fig. 2 are summarized.

- *Step 1: Inventory model*
  The first step of the solution framework is to generate the product order of each gas station. The information required to generate the orders in this step comprises the initial stock levels, the demand rates, and the tank capacities. In addition to the gas station information, the inventory related costs are also required to calculate the expected total costs. These costs include the unit order cost, fixed order cost, holding cost, shortage cost, and send-back cost. The inventory model is used to determine the order quantities and desired delivery time windows that minimizes the gas stations' expected total costs based on their stock levels and petroleum consumption rates. A full description of this model is given in Sect. 4.
- *Step 2: Route generator*
  The delivery routes are generated for the gas stations that placed orders during the given day according to the inventory model. Typically, a truck contains four to six compartments with different capacities and each gas station requires one or two compartments to satisfy the petroleum requirement. Therefore, having routes serving between one and three stations is common in practice. However, we also consider the situation where the petroleum distribution serves gas stations with lower demand rates. For this case, trucks can visit four to five stations within a route. We describe the process used to generate the candidate routes in Sect. 5.1.

- *Step 3: Truck loading capacity check*
  The total delivery quantities of a route cannot exceed the capacity of the compartments in the truck. The truck loading model is used to determine the assignment of the different petroleum products to each of the truck compartments. The route will be eliminated if the demands of the gas stations of the given route cannot be loaded into the truck. In addition, this model computes the profit of the route, which is calculated by the revenue received for delivering the petroleum minus the travel costs of the route. We describe this model in Sect. 5.2.
- *Step 4: Route scheduling*
  The truck schedule of each feasible route generated is then determined in this step. The objectives are to find the truck departure and returning time, the delivery time for each gas station in the route, and a set of penalties for those routes failing to satisfy the time window constraints. A route that fails to satisfy the time window constraints is not eliminated in this step. Instead, we add a penalty that accounts for the total time the truck following the given route arrives before or after the gas station time windows. Consequently, the generated routes are divided into two sets: feasible (on time) routes and infeasible (late) routes and their costs are updated with the corresponding penalties. We describe this model in Sect. 5.3.
- *Step 5: Truck routing*
  In the truck routing problem, the truck assignment will be decided by maximizing the total profits of delivering the petroleum to all gas stations. The calculation of the objective includes the route profit found in Step 3, as well as the penalties for late deliveries obtained in Step 4. We describe this model in Sect. 5.6.
- *Step 6: Variable neighborhood search*
  The routes produced by the generator are improved by a variable neighborhood search heuristic that comprises several local search procedures. The improvements in the routes are calculated by solving both the resulting scheduling and routing models. We describe this model in Sect. 5.5.
- *Step 7: Termination condition*
  As mentioned before, the petroleum distribution problem is solved sequentially on a daily basis. The results of route scheduling and truck routing for day t are inputs for day $t + 1$. If the considered time horizon is not reached, then return to Step 1 to resolve the problem for the subsequent day. If the time horizon is reached, we terminate the process.

4 The inventory model for the gas stations

The inventory model seeks to determine the desired delivery time window $[a_i, b_i]$ over the time horizon T and the order quantity q_i for each customer $i \in N$ so that the expected customer's total cost is minimized. Let l_i be the actual delivery time for customer $i \in N$. The total cost of each customer given delivery time l_i is named $C_i(l_i)$ and comprises: (1) the ordering cost $P_i(q_i)$, which represents the cost of ordering product to the supplier; (2) the holding cost $H_i(l_i, a_i, b_i, q_i)$, which is the opportunity cost of having inventory; (3) the shortage cost $S_i(l_i, a_i, b_i, q_i)$, which is the cost the customer incurs if at some point it runs out of inventory (e.g., for the case of a retailer, the equity cost associated with losing potential sales); and (4) the send-back cost $B_i(l_i, a_i, b_i, q_i)$, which is a monetary penalty paid to the transporter for ordering product in excess (i.e., more than what the customer can accommodate at delivery time). Notice that the order quantity and delivery time window are selected by each customer before the transporter decides upon the distribution planing. Hence, from the perspective of the customer, the exact delivery time l_i is not known when solving its inventory problem, which implies that the total cost for customer i is a function of the delivery time l_i

$$C_i(l_i) = \min_{[a_i,b_i]\in T, q_i\geq 0} P_i(q_i) + H_i(l_i, a_i, b_i, q_i) + S_i(l_i, a_i, b_i, q_i) + B_i(l_i, a_i, b_i, q_i). \tag{1}$$

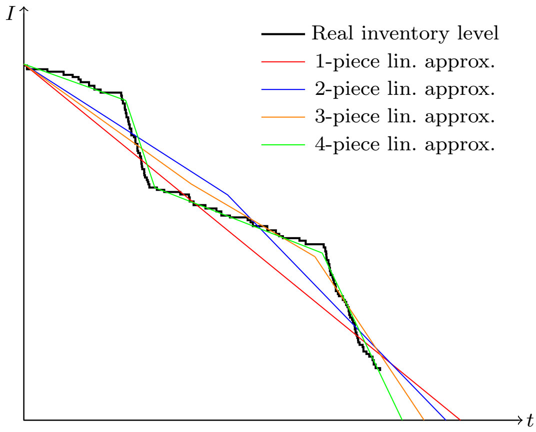
The ordering cost P_i is often given by a function comprising a fixed ordering cost k plus the product between the order quantity q_i and a unitary price per gallon F. The fixed ordering cost is not necessarily a monetary cost, but rather a logistics cost in the sense that both the transporter and gas stations prefer fewer visits to fulfill the demand of product. On the one hand, the gas stations generally prefer ordering product when their inventory is low to avoid mixing large quantities of product that has been sitting in their tanks with the new incoming product. Furthermore, to avoid constantly stirring the sediments in their tanks (as a result of the tank filling process), gas stations typically request large quantities to reduce the number of fulfilling rounds. On the other hand, the transporter prefers to deliver large quantities to the customers because that allows them to have a better use of the trucks storage space, as well as reducing the number of visits to each gas station.

In the petroleum product distribution context then, the fixed ordering cost can then be modeled as high monetary cost when compared to the holding costs, which translates into gas stations requesting enough gasoline to fill their tanks as much as possible. We will assume that the order quantity q_i is made so that upon arrival, the transporter delivers enough gasoline to fill the underground tanks of the gas station. Therefore, such an order quantity depends on the initial stock r_i, the demand rate $d_i(t)$, and the expected delivery time. Ideally, the gas station would expect the product to be delivered at the specific time within the time window for which the possible shortages or send-backs, and holding costs are minimized. Nevertheless, since the gas station does not know in advance the actual delivery time l_i—as this is decided by the transporter after all the orders are collected—the gas station must then base its decisions on the idea that the delivery truck will arrive at any time within the time window $[a_i, b_i]$. For this reason, from the perspective of the gas stations, since the transporter is allowed to arrive at any time within the time window, the delivery time $l_i \sim U[a_i, b_i]$ is modeled as a random variable uniformly distributed over the interval given by a_i and b_i. Furthermore, the order quantity q_i must be then chosen so that the expected total costs are minimized. In general, shortage costs and send-back costs are often more expensive than the holding costs. Thus, the customers aim to select time windows so that the stock-out time T_i^e, also referred to as the time to empty (i.e., the expected time in which the customer runs out of inventory), occurs after the window upper bound b_i.

4.1 Demand rate and inventory levels

The demand observed by the gas stations for each of the petroleum products is the result of a random process in which a set of customers arrive randomly to the gas station to fill their tanks. The demand process is not particularly homogeneous throughout the day because the number of customers is correlated with the automobile traffic near the station—which, if the gas station is located within a typical urban area, is known to follow a bimodal distribution following to the early inbound and late outbound traffic flows during a typical work day (Hallenbeck et al. 1997). The inventory level of the products in between tank fills thus evolves naturally in a non-increasing staircase fashion, as depicted in Fig. 3, where each of

Fig. 3 Demand distribution



the small down steps correspond to a client purchasing gasoline and the slope changes are given by an increase or decrease on the customer arrival rate depending on the time of the day. The quantities demanded by the customers do not vary much from the average, but the number of customers may drastically change during the day, producing different demand rate patterns.

The quality of the inventory model clearly depends on the approach used to model the demand rates. There are several models that have been developed over the last few decades to tackle different variations of inventory problems with variable demands (Simchi-Levi et al. 2005). In this paper we resort to a widely used approach in which the demand rate $d_i(t)$ is estimated from historical data using average demand quantities observed during different time segments of the day. Depending on the desired precision, each day can be partitioned into a set of sequentially disjoint time segments and then a specific demand rate can be assigned to each of them. The real inventory level is then approximated as a non-decreasing piecewise linear function—as depicted in Fig. 3—where the slope of each segment corresponds to the estimated demand rate at that particular time. Clearly, the more time segments are used, the closer the static representation would be from the demand of a typical day.

There are several advantages and disadvantages of using this approach. On the one hand, calculating the holding, send-back, shortage, and ordering costs is straightforward, as one only requires to calculate fixed values obtained from the resulting estimated function. Also, for gas stations with low demand volatility (i.e., stations with steady demand patterns during the week), piecewise-linear estimations often provide fair approximations, despite their simplicity. On the other hand, static estimations of this kind tend to over or underestimate the costs, particularly the holding costs. Moreover, for instances in which the demand has a high variance during the day, the static estimation may produce inaccurate results. Even with a relatively accurate demand rate estimation, a single static representation of the demand rates will not fully capture the stochastic nature of the demand process.

To overcome this issue and take advantage of the flexibility of the static models, it is possible to estimate the slopes of the piecewise linear approximation of the inventory function as random variables from historical data, and then randomly generate several demand scenarios producing low, medium, and high types of demand rates and inventory levels, as in a Monte Carlo sampling method. Then, each of the demand scenarios can be solved individually as a static model and the resulting routes obtained for each of them can be compared across the full set selecting the ones that better performed in average. More specifically, for each customer and each day of the weekly horizon we use a sampling based approach in which we generate κ scenarios selecting the demand rate for each station randomly according to the estimated distribution. Each scenario is then solved individually on a daily basis.

4.2 Mathematical model

In this section we describe the inventory model of any given station $i \in N$. We focus our attention on the case in which the inventory level is estimated using a linear approximation (i.e., a 1-piece approximation) because the number of cases that must be considered is small enough to be described in detail in this section. Nevertheless, the inventory model for more detailed piecewise linear approximations can be generated easily in similar fashion, keeping track of the slope changes for each linear piece.

From the modeling point of view, it is advantageous to define the order quantity q_i in terms of a time value $c_i \in [a_i, b_i]$ as follows. Given the demand rate d_i and the initial inventory r_i, for any given $c_i \in [a_i, b_i]$, the amount of gasoline left in the tank at time c_i is $\max\{0, r_i - c_i d_i\}$. Since gas stations are assumed to request enough gasoline q_i to fill their tanks, we have that $Q_i - q_i = \max\{0, r_i - c_i d_i\}$, which in turn implies that $q_i = Q_i - \max\{0, r_i - c_i d_i\}$.

The following assumptions are made for the gas station inventory model:

- For each demanded product, the gas station solves an inventory model to determine $[a_i, b_i]$ and q_i.
- The delivery time l_i is assumed uniformly distributed over the interval $[a_i, b_i]$.
- The time horizon for the petroleum distribution of each operational day is discretized into 12 hours.
- No backlogging of demand is allowed (i.e., sales not met due to shortages are lost).
- The expected demand rate d_i is assumed to be constant and estimated from historical data given a particular sampled scenario (see Sect. 4.1).
- Depending on the desired granularity of the time horizon and demand settings (i.e., the randomly generated demand rates), different demand profiles can be generated for each gas station.
- Inventory is continuously reviewed and all replenishment decisions are made at the beginning of each time period and once the orders are placed, those cannot be changed.
- Each gas station has a known starting level r_i and that information is not disclosed to the supplier.
- The order quantity q_i is equal to the capacity of the underground tank minus the amount left at a time $c_i \in [a_i, b_i]$. That is, $q_i = Q_i - \max\{0, r_i - c_i d_i\}$.
- The expected stock-out time of each station T_i^e is given by the demand rate profile and the initial stock of the stations.
- The possible time window choices are given by a discrete finite set.

The mathematical notation for the inventory model is provided in Table 1.

The objective of the model is to identify optimal values for a_i, b_i, and c_i (thus, q_i), so as to minimize the overall inventory costs. To properly model the above inventory problem, there are several cases that must be considered depending on the values of a_i, b_i, q_i, l_i, and T_i^e. Particularly, if for the given values of $[a_i, b_i]$, the stock-out time T_i^e falls either: before a_i, between a_i and b_i or, after b_i. When gas stations receive the orders, four scenarios may occur:

- Scenario 1: the stock-out time T_i^e of the station occurs prior to a_i.
- Scenario 2: the stock-out time T_i^e of the station occurs between a_i and c_i. This scenario comprises two cases that depend on whether the delivery lead time l_i is shorter or longer

Table 1 Mathematical notation for the inventory model

Notation	Definition
Q_i	The underground tank capacity of station i
T_i^e	Stock out time of station i
d_i	Demand rate of station i
l_i	Delivery lead time, which is assumed to follow a uniform distribution with limits a_i and b_i
r_i	Initial stock level of underground tank of station i
F	The order cost per gallon
k	Fixed cost per delivery
h	Holding cost (per unit per unit time)
s	Shortage cost (per unit per unit time)
p	Send-back cost (per unit per unit time)
a_i	Earliest time window of station i
b_i	Latest time window of station i
q_i	Order quantity of gas station i

than the stock-out time T_i^e. i.e., the following two possibilities: (1) $l_i \leq T_i^e$ and (2) $T_i^e \leq l_i$.

- Scenario 3: the stock-out time T_i^e of the station occurs between c_i and b_i. This scenario comprises three cases that depend on the position of the delivery lead time l_i within the time window. i.e., the following three cases: (1) $l_i \leq c_i \leq T_i^e$, (2) $c_i \leq l_i \leq T_i^e$, and (3) $c_i \leq T_i^e \leq l_i$.
- Scenario 4: the stock-out time T_i^e of the station occurs after b_i. This scenario comprises two cases as well, which depend on if the delivery lead time l_i is shorter or longer than c_i. In other words, we have the following two possibilities: (1) $l_i \leq c_i$ and (2) $c_i \leq l_i$.

The four scenarios described above are summarized in Table 2 and the corresponding expected total cost calculations are presented afterwards.

- Scenario 1: $T_i^e \leq a_i$
  In Scenario 1, the stock-out time T_i^e of the station occurs prior to a_i. This implies that there will be a cost associated with a petroleum shortage. Consequently, the send-back

Table 2 Summary of scenarios

	Stock-out time	Case in scenario
Scenario 1	$T_i^e \leq a_i$	
Scenario 2	$a_i \leq T_i^e \leq c_i$	Case 1: $l_i \leq T_i^e$
		Case 2: $T_i^e \leq l_i$
Scenario 3	$c_i \leq T_i^e \leq b_i$	Case 1: $l_i \leq c_i \leq T_i^e$
		Case 2: $c_i \leq l_i \leq T_i^e$
		Case 3: $c_i \leq T_i^e \leq l_i$
Scenario 4	$b_i \leq T_i^e$	Case 1: $l_i \leq c_i$
		Case 2: $l_i \geq c_i$

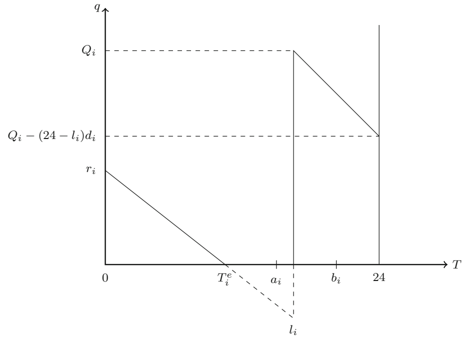


Fig. 4 The inventory level of Scenario 1

cost is zero and the order quantity q_i is equal to the underground tank capacity Q_i. Figure 4 provides a graphical representation of the inventory of Scenario 1.
Since the order quantity is equal to the tank capacity, the ordering cost is given by a fixed ordering cost plus the order variable cost as follows

$$k + FQ_i,$$

where k is the fixed cost and F is the unit order cost per gallon. The expected daily holding volume can be calculated from the inventory function depicted in Fig. 4. The expected holding volume before the underground tank becomes empty is $r_i T_i^e/2$ and the expected daily holding volume after order arrival is $Q_i(24-l_i)-(24-l_i)^2 d_i/2$. Therefore, we obtain the following expected daily holding cost:

$$\left(\frac{r_i T_i^e}{2} + Q_i(24-l_i) - \frac{(24-l_i)^2 d_i}{2}\right)h.$$

The total shortage cost is given by

$$\left(\frac{(l_i - T_i^e)^2 d_i}{2}\right)s,$$

which is given by the area above the inventory function during the shortage period times the shortage cost s.
The send-back cost is zero because the underground tank will become empty prior to receiving the ordered petroleum and as such it can accommodate the full order. Hence, the total cost for Scenario 1, denoted by $C_i^{S1}(l_i)$ is then:

$$\begin{aligned} C_i^{S1}(l_i) &= \text{ordering cost + holding cost + shortage cost + send-back cost} \\ &= k + FQ_i + \left(\frac{r_i T_i^e}{2} + Q_i(24-l_i) - \frac{(24-l_i)^2 d_i}{2}\right)h + \left(\frac{(l_i - T_i^e)^2 d_i}{2}\right)s \end{aligned} \tag{2}$$

Given that l_i is assumed to be uniformly distributed within the time window, to calculate the expected cost, we take the following integral over a_i and b_i for l_i. Thus, the expected total cost of Scenario 1, $E[C_i^{S1}]$, is given by:

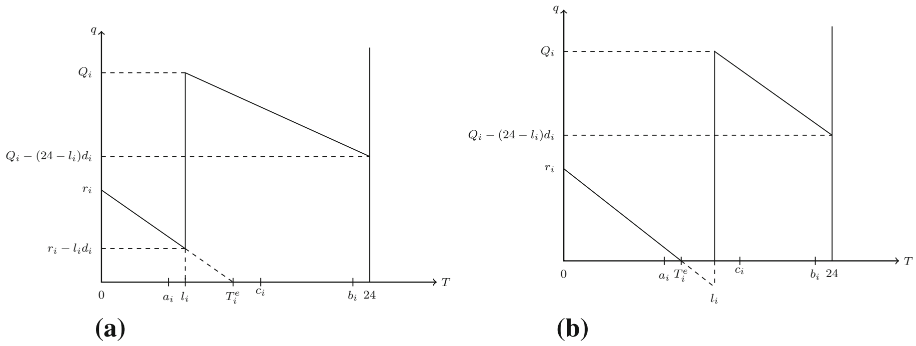


Fig. 5 The inventory model of Scenario 2. **a** Scenario 2, Case 1. **b** Scenario 2, Case 2

$$E\left[C_i^{S1}\right]=\int_{a_i}^{b_i} C_i^{S1}(l_i)\frac{1}{b_i-a_i}\,dl_i \tag{3}$$

- Scenario 2: $a_i \le T_i^e \le c_i$.
  In this scenario, since the stock-out time occurs before c_i, the order quantity is equal to the tank capacity Q_i; thus, as for Scenario 1, the ordering cost is $k+FQ_i$. To evaluate the other costs, it should be noted that the difference between l_i, and T_i^e within $[a_i, b_i]$ play a significant role affecting the final outcome of the expected total cost. There will be a send-back cost and no shortage cost if the station receives the order prior to running out of petroleum ($l_i < T_i^e$). Conversely, there will be a shortage cost and no send-back cost if the station runs out of product before receiving petroleum ($T_i^e < l_i$). We consider both these cases in Scenario 2 as follows:

  - **Case 1**: the petroleum is received before the stock-out time ($l_i \le T_i^e$).

    As depicted in Fig. 5a, gas station i will not run out of petroleum, but will incur a send-back cost. The expected daily holding volume before receiving the petroleum is $r_il_i - l_i^2d_i/2$ and the expected daily holding volume after the order arrival is $Q_i(24-l_i)-(24-l_i)^2d_i/2$. Hence, the daily holding cost is:

    $$\left(r_il_i-\frac{l_i^2d_i}{2}+Q_i(24-l_i)-\frac{(24-l_i)^2d_i}{2}\right)h.$$

    The quantity that exceeds the tank capacity when receiving the petroleum is $r_i-l_id_i$, so the send-back cost is $(r_i-l_id_i)p$. The total cost for this case, denoted by $C_i^{S2C1}(l_i)$ is then:

    $$C_i^{S2C1}(l_i)=k+FQ_i+\left(r_il_i-\frac{l_i^2d_i}{2}+Q_i(24-l_i)-\frac{(24-l_i)^2d_i}{2}\right)h + (r_i-l_id_i)p. \tag{4}$$

  - **Case 2**: the petroleum is received after the stock-out time ($T_i^e \le l_i$).

    As depicted in Fig. 5b, gas station i will run out of petroleum, but will not incur a send-back cost. The expected daily holding volume before receiving petroleum is

$T_i^e r_i/2$ and the expected daily holding volume after the order arrival is $Q_i(24-l_i) - (24-l_i)^2 d_i/2$. Hence, the daily holding cost is:

$$\left(\frac{T_i^e r_i}{2} + Q_i(24-l_i) - \frac{(24-l_i)^2 d_i}{2}\right)h.$$

The expected volume of lost sales due to the shortage of petroleum is $(l_i - T_i^e)^2 d_i/2$, so the shortage cost is $((l_i - T_i^e)^2 d_i/2)s$. The total cost for this case, denoted by $C_i^{S2C2}(l_i)$ is then:

$$C_i^{S2C2}(l_i) = k + FQ_i + \left(\frac{T_i^e r_i}{2} + Q_i(24-l_i) - \frac{(24-l_i)^2 d_i}{2}\right)h + \left(\frac{(l_i - T_i^e)^2 d_i}{2}\right)s. \tag{5}$$

Similarly as for Scenario 1, to calculate the expected cost, we take the integral over the interval from a_i to T_i^e for (4) and from T_i^e to b_i for (5). The expected total cost of Scenario 2, $E[C_i^{S2}]$, is given by:

$$E[C_i^{S2}] = \int_{a_i}^{T_i^e} C_i^{S2C1}(l_i)\frac{1}{b_i - a_i}\,dl_i + \int_{T_i^e}^{b_i} C_i^{S2C2}(l_i)\frac{1}{b_i - a_i}\,dl_i \tag{6}$$

– Scenario 3: $c_i \le T_i^e \le b_i$.
As for Scenario 2, the differences between l_i, T_i^e, and c_i within $[a_i, b_i]$ are also fundamental for evaluating the expected total cost. In this case, however, since the stock-out time occurs after c_i, the order quantity is no longer Q_i, but q_i, which is the customer's tank capacity minus the expected available inventory at c_i. The send-back cost, shortage cost, holding cost, and order cost depend on whether the station receives the order prior to c_i ($l_i \le c_i \le T_i^e$), between c_i and the stock-out time ($c_i \le l_i \le T_i^e$) , or after the stock-out time ($T_i^e \le l_i$). We consider these three cases in Scenario 3.
In all three cases, the order quantity of station i is set to the amount required to fill the underground tank at the time c_i. Since the underground tank will not become empty before the order expected arrival time, the order quantity will not be equal to the full capacity of the tank, as depicted in Fig. 6. The order quantity for all the three cases in this scenario is given by:

$$Q_i - (r_i - c_i d_i)\,, \tag{7}$$

where $r_i - c_i d_i$ represents the quantity left in the tank at the expected arrival time.

  – **Case 1**: the petroleum is received before c_i (i.e., $l_i \le c_i \le T_i^e$).

  The expected daily holding volumes before and after the order arrival for this case are $r_i l_i - l_i^2 d_i/2$ and $(24-l_i)Q_i - (24-l_i)^2 d_i/2$, respectively, resulting in the following expected daily holding costs:

$$\left(r_i l_i - \frac{l_i^2 d_i}{2} + (24-l_i)Q_i - \frac{(24-l_i)^2 d_i}{2}\right)h.$$

  The shortage cost is zero since the station receives the petroleum before they run out; on the other hand, the amount of petroleum that is sent back is $c_i d_i - l_i d_i$, resulting in a cost of

$$(c_i d_i - l_i d_i)p.$$

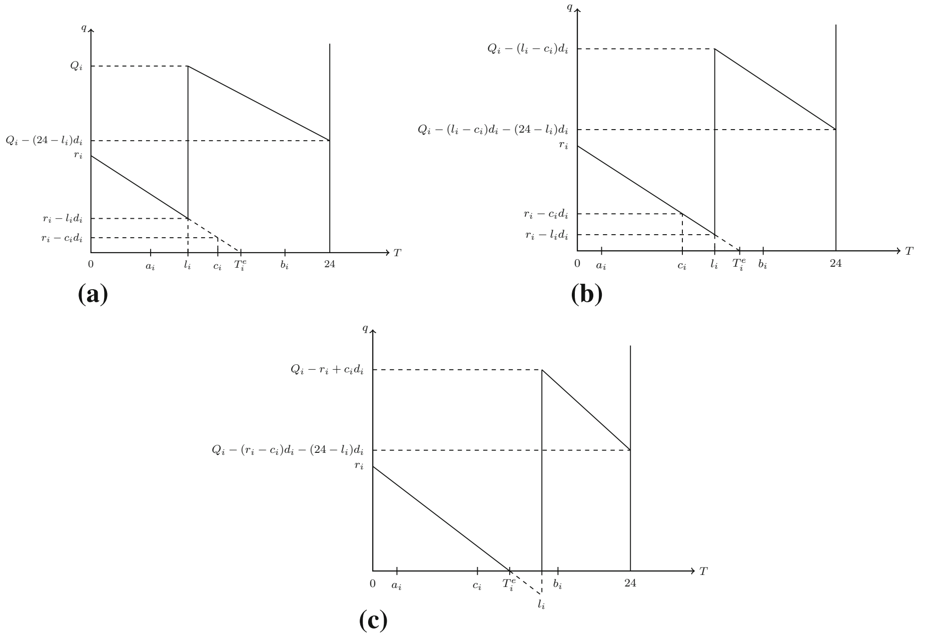


Fig. 6 The inventory model of Scenario 3. **a** Scenario 3, Case 1. **b** Scenario 3, Case 2. **c** Scenario 3, Case 3

Let $C_i^{S3C1}(l_i)$ denote the total cost for Scenario 3, Case 1. That is,

$$
\begin{aligned}
C_i^{S3C1}(l_i) = k &+ F(Q_i - r_i + c_i d_i) \\
&+ \left(r_i l_i - \frac{l_i^2 d_i}{2} + (24 - l_i) Q_i - \frac{(24 - l_i)^2 d_i}{2} \right) h \\
&+ (c_i d_i - l_i d_i) p.
\end{aligned} \tag{8}
$$

- **Case 2**: the petroleum is received between c_i and the stock-out time ($c_i \le l_i \le T_i^e$).

  For this particular case, the expected daily holding volumes before and after the order arrival are $r_i l_i - l_i^2 d_i/2$ and $(24-l_i)(Q_i-(l_i-c_i)d_i)-(24-l_i)^2 d_i)/2$, respectively, resulting in the following expected daily holding costs:

$$
\left(r_i l_i - \frac{l_i^2 d_i}{2} + (24 - l_i)\left(Q_i - (l_i - c_i)\, d_i\right) - \frac{(24 - l_i)^2 d_i}{2} \right) h.
$$

  Both the shortage cost and send-back cost are zero since the station receives the petroleum before they run out and the ordered quantity will not exceed the tank capacity because the actual order arrival time occurs after the expected arrival time

c_i. Let $C_i^{S3C2}(l_i)$ denote the total cost for Scenario 3, Case 2. That is,

$$C_i^{S3C2}(l_i) = k + F\Big(Q_i - r_i + c_i d_i\Big) + \left(r_i l_i - \frac{l_i^2 d_i}{2} + (24 - l_i)\,(Q_i - (l_i - c_i)\,d_i) - \frac{(24 - l_i)^2 d_i}{2}\right) h. \tag{9}$$

– **Case 3**: the petroleum is received after the stock-out time ($T_i^e \leq l_i$).

  The holding cost is given by:

$$\left(\frac{r_i T_i^e}{2} + \Big(Q_i - r_i + c_i d_i\Big)(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}\right) h,$$

  where $(r_i T_i^e)/2$ represents the expected daily holding volume before running out of petroleum and

$$\Big(Q_i - r_i + c_i d_i\Big)(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}$$

  is the volume after receiving the petroleum.
  The expected volume of lost sales is $(l_i - T_i^e)^2 d_i/2$, which results in a shortage cost given by

$$\frac{(l_i - T_i^e)^2 d_i}{2} s.$$

  Also, the send-back cost is zero since the tanks are always able to accept the full order in this case. The total cost for Scenario 3, Case 3 is then:

$$C_i^{S3C3}(l_i) = k + F\Big(Q_i - r_i + c_i d_i\Big) + \left(\frac{r_i T_i^e}{2} + \Big(Q_i - r_i + c_i d_i\Big)(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}\right) h + \frac{(l_i - T_i^e)^2 d_i}{2} s. \tag{10}$$

  Furthermore, taking integral over the range a_i to c_i for Case 1, over c_i to T_i^e for Case 2, and over T_i^e to b_i for Case 3—i.e., combining expressions (8), (9) and (10) and integrating for their respective limits—the expected total cost of Scenario 3, $E[C_i^{S3}]$, is:

$$E[C_i^{S3}] = \int_{a_i}^{c_i} C_i^{S3C1}(l_i) \frac{1}{b_i - a_i}\, dl_i + \int_{c_i}^{T_i^e} C_i^{S3C2}(l_i) \frac{1}{b_i - a_i}\, dl_i + \int_{T_i^e}^{b_i} C_i^{S3C3}(l_i) \frac{1}{b_i - a_i}\, dl_i, \tag{11}$$

  for $c_i \in [a_i, T_i^e]$.

– Scenario 4: $b_i < T_i^e$.
  As depicted in Fig. 7, the order cost in Scenario 4 is also given by expression (7) because T_i^e occurs after c_i. Furthermore, since the stock-out time occurs after b_i, the shortage cost

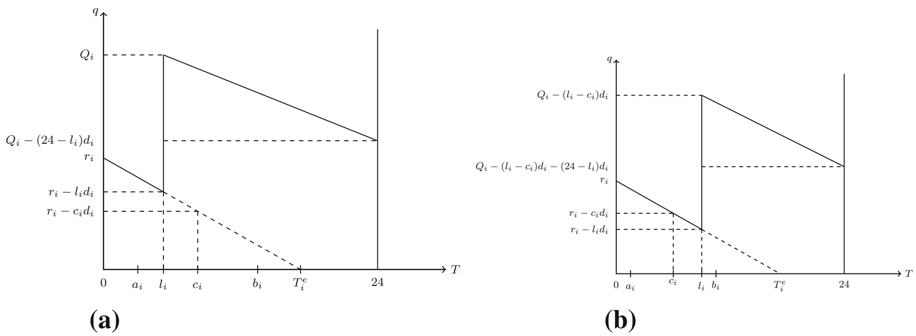


Fig. 7 The inventory model of Scenario 4. **a** Scenario 4, Case 1. **b** Scenario 4, Case 2

is zero. The send-back cost and the holding cost depend on whether the station receives the order before or after c_i. We consider these two cases in Scenario 4.

– **Case 1**: the petroleum is received before c_i

Figure 7a depicts the inventory level for this case, for which the send-back cost is given by:

$$(c_i d_i - l_i d_i)\, p,$$

where $c_i d_i - l_i d_i$ is the difference between order quantity and the actual received quantity. The expected holding volumes before and after order arrival are

$$r_i l_i - \frac{l_i^2 d_i}{2}$$

and

$$Q_i(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}.$$

Thus, the holding cost is

$$\left(r_i l_i - \frac{l_i^2 d_i}{2} + Q_i(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}\right)h.$$

We denote the total cost for the Scenario 4, Case 1 as $C_i^{S4C1}(l_i)$, which is thus given by

$$\begin{aligned} C_i^{S4C1}(l_i) &= k + F(Q_i - r_i + c_i d_i) \\ &\quad + \left(r_i l_i - \frac{l_i^2 d_i}{2} + Q_i(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}\right)h \\ &\quad + (c_i d_i - l_i d_i)\, p. \end{aligned} \tag{12}$$

– **Case 2**: the petroleum is received after c_i (i.e., $l_i \geq c_i$)

Figure 7b depicts the inventory level for this case, there will be no shortage nor send-back costs and the holding cost is quite similar to the holding cost in Case 1, except the expected holding volume after order arrival is

$$\Big(Q_i - (l_i - c_i)\, d_i\Big)(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}$$

Thus, the total cost for Scenario 4, Case 2, denoted as $C_i^{S4C2}(l_i)$, is given by

$$C_i^{S4C2}(l_i) = k + F\left(Q_i - r_i + c_i d_i\right)$$
$$\left(r_i l_i - \frac{l_i^2 d_i}{2} + \left(Q_i - (l_i - c_i)d_i\right)(24 - l_i) - \frac{(24 - l_i)^2 d_i}{2}\right)h. \qquad (13)$$

We then take the expectation over the limits for l_i, from a_i to c_i for Case 1 and from c_i to b_i for Case 2. The expected total cost of Scenario 4, referred to as, $E[C_i^{S4}]$, is then given by

$$E[C_i^{S4}] = \int_{a_i}^{c_i} C_i^{S4C1}(l_i)\frac{1}{b_i - a_i}\,dl_i + \int_{c_i}^{b_i} C_i^{S4C2}(l_i)\frac{1}{b_i - a_i}\,dl_i. \qquad (14)$$
$$\text{for } c_i \in [a_i, b_i].$$

The objective of the inventory model is to select the optimal time window $[a_i, b_i]$, and order quantity q_i (by finding the optimal c_i) for each gas station among all possible discrete delivery times for a_i and b_i specified by transporters, so as to minimize the total inventory costs. Based on the value of T_i^e given by $d_i(t)$, we can identify which scenario will be applied and the expected total costs can be further calculated. The optimal time window and order quantity for each gas station can be found then by solving (15).

$$(c_i, [a_i, b_i]) \in \arg\min\{E[C_i(l_i)]\},\ i \in N; \qquad (15)$$

where

$$E[C_i(l_i)] = \begin{cases} E[C_i^{S1}(l_i)], & T_i^e \le a_i, \\ \min\{E[C_i^{S2}(l_i)], E[C_i^{S3}(l_i)]\}, & a_i \le T_i^e \le b_i, \quad i \in N. \\ E[C_i^{S4}(l_i)], & b_i \le T_i^e, \end{cases} \qquad (16)$$

Furthermore, notice that for a given realization of l_i (i.e., after the transporter has decided upon the distribution routes), it is possible to calculate the actual total cost incurred by the gas station by selecting the scenario that corresponds to the value T_i^e and use a similar procedure to evaluate the resulting value for $C_i(l_i)$, for all $i \in N$. As will be described in the following section, we use the actual costs obtained after the distribution operation of each day to smooth the possible tardiness imbalances among the customers.

4.3 Send-back buffer and lower bounds on the replenishment process

To void incurring unnecessary send-back costs for requesting more product than what the tanks are capable to hold when the transporter arrives, gas stations typically set up a buffer $\epsilon_i \ge 0$ so that they order enough product to reach $Q_i - \epsilon_i$ instead of Q_i. This is particularly useful for cases in which the time window is relatively narrow, mainly because even small values of ϵ_i can protect the gas station of having to reject a portion of the order without compromising the operation of the gas station. Nevertheless, for wider time windows, the buffer may not be sufficient protection.

Setting up the send-back buffer in the above inventory model is straightforward, as a send-back cost exists if the excess in the product order is larger than ϵ_i. We now proceed to discuss the distribution part of the problem.

In addition to the send-back buffer, customer $i \in N$ can also impose a limitation on the expected amount of product to be held in the tank at the end of the replenishment time window b_i. That is, the inventory level at time b_i should be at least a certain percentage ϵ of the tank size Q. This typically prevents scenarios in which time windows that are too wide are selected. Extremely wide time windows could allow the transporter to arrive too early, thus having to potentially return large quantities of product; or arriving too late, which may result in unwanted shortages. The limitation then requires that at time b_i, the product left in the tank satisfies $Q - d(b_i - c_i) \geq \epsilon Q$.

5 The distribution problem of the transporter

The proposed integrated routing model seeks to determine distribution plan and the strategic decisions of the transporter by taking into account the customer decisions given by the inventory model presented in the previous section.

5.1 Route generator

As mentioned before, we attempt to generate all the possible delivery routes for the gas stations that placed orders during each given day, based upon the results of the inventory models. For the cases where few gas stations are served per route, we generate all possible route permutations visiting one, two and three stations for a total of $O(|N|^3)$ routes. As for the cases where the route visits more gas stations, since we solve both the truck loading model and the route scheduling model for all of those routes every day of the planning horizon, generating all possibilities is simply impractical and would take longer than the desired computational time limit. Notice that for each day of the planning horizon all routes must be tested. In other words, a route that was considered infeasible for the first day, may be feasible for the demands and time windows of the second day. Therefore, instead of generating all the possible permutations of four and five gas stations, we start by randomly generating a smaller subset of such routes to reduce the total computational load.

Since, some good routes may not be generated in the random route generation, we test different sets of randomly generated routes to see the impact of the random generation. We then proceed to generate improved routes by applying a *Variable Neighborhood Search* (VNS) (Mladenovic and Hansen 1997) over the generated routes. VNS is a powerful metaheuristic that has been commonly used to tackle routing problems directly or, like in this case, to complement other optimization methodologies (Hemmelmayr et al. 2009; Dayarian et al. 2016; Kytöjoki et al. 2007). Given its simplicity and documented efficacy, we rely on a VNS to generate improved routes. Nonetheless, other effective metaheuristic search mechanisms in combinatorial optimization (e.g, GRASP or tabu search) can also take advantage of the same decomposition approach presented here. We describe the VNS used in our approach in Sect. 5.5.

Traditional approaches often rely on column generation when searching for alternative routes that could potentially benefit the distribution process (Christiansen and Lysgaard 2007). In the context of column generation for solving VRPs, a so-called pricing subproblem is used to generate new, unexplored, feasible routes that have the potential of reducing the distribution costs. Instead of enumerating all the possible routes, the pricing subproblem only generates new routes if there is a potential reduction of the transporter objective function by including such potential routes. In fact, when coupled with a branching scheme (i.e., when using branch and price), the obtained solutions are guaranteed to be optimal.

The idea of the pricing subproblem is thus to generate a sequence of the customers that will be visited by the route being generated, guaranteeing that the capacity of the compartments in the truck, as well as the time windows and other requirements are satisfied (Feillet et al. 2010;

Feillet 2010). VRP pricing subproblems are commonly modeled as elementary shortest path problems, which depending on the route length and required conditions can be quite difficult (Lozano et al. 2016).

In the petroleum distribution context described in this paper, there are several challenges that render the task of combining all the route requirements within a monolithic pricing subproblem to be rather difficult. One specific hurdle comes from the fact that routes that do not necessarily satisfy the time windows may still be considered as part of the distribution plan. Given the fact that some distribution instances can be infeasible due to the strict nature of the time window requirements, incorporating in the subproblem the customer sequence selection, the multi-compartment loading component, the goal of producing route schedules that minimize late deliveries, the penalties for the late deliveries in the previous day, and the pricing component may be unproductive. In this paper, despite not guaranteeing and optimality certificate, we resort to the random generation option described before because it allows us to solve each of the aforementioned components sequentially. Furthermore, the random generation is easily be embedded within a improving local-search based heuristic to increase the quality of the routes generated—as described in Sect. 5.5.

Let R be a set of candidate routes generated. For each of the routes in R, we solve the truck loading model that checks if the order quantities of the given route can be loaded into truck compartments. Furthermore, if the route satisfies the truck loading constraints, we then generate the schedule for the route based on the given time windows of the customers of each route.

5.2 Truck loading model

In addition to satisfying the constraints given by the time windows, a feasible route must also satisfy the truck compartment capacity constraints. As the trucks used to deliver have multiple compartments with different capacities, the demand feasibility of the routes is checked in the truck loading model. If the demand of a route cannot be loaded into any truck, this route will be eliminated from the set of candidate routes. Therefore, for every possible route $r \in R$, we solve the following tuck loading model to test the loading feasibility of the route. The notation and the decision variables are described in Table 3.

Then the truck loading model is given by:

$$q_i \leq \sum_{c \in C} Q_c y_{ic} \quad \forall i \in N_r, \tag{17}$$

$$\sum_{i \in N_r} y_{ic} \leq 1 \quad \forall c \in C, \tag{18}$$

$$y_{ic} \in \{0, 1\} \quad \forall i \in N_r,\ c \in C. \tag{19}$$

Table 3 Mathematical notation for the truck loading model

Notation	Description
N_r	The set of gas stations served by the route r being tested
C	The set of compartments for the given truck
q_i	The delivery quantity to station i
Q_c	The capacity of compartment c
y_{ic}	A binary variable equal to 1 if demand of gas station i is assigned to compartment c, and 0 otherwise

Constraints (17) enforce that the delivery quantity of station i cannot exceed the sum of compartment capacities of the loaded truck assigned to such a gas station. Constraints (18) ensure only one demand can be loaded into each compartment. Constraints (19) define the decision variables. Notice that the loading model is in fact a satisfiability problem, as any feasible solution can be used for loading the truck compartments. If route r does not satisfy the truck loading model, it is removed from set R.

Additionally, in this stage we compute the profit of the route, which is calculated by the revenue perceived by delivering the petroleum minus the travel costs of the route. Since the gas station visit sequence is given by the route, and the order quantities of each customer are calculated a priori by the inventory models, if there is a feasible loading distribution of route r—given by the solution of model (17)–(19)—we then calculate the corresponding profit of r and the transportation cost of the route and use it as an input for the truck routing model.

5.3 Route scheduling model

The route scheduling model aims to find the truck schedule of each candidate route $r \in R$, as well as a set of penalties for having late deliveries. Given the time windows obtained by the inventory models and the starting hour of the delivery shift, we check the feasibility of the time window constraints for each gas station of candidate route $r \in R$. Since the number of trucks is limited and the time window requirements for the gas stations of route r could be too close or potentially overlap, the delivery time windows are not always satisfied for all customers. For that reason, in the route scheduling model we assign penalties to the routes for which the stations receive petroleum at undesired times. The penalties consist of additional costs for truck arrival times either before a_i or after b_i, and are proportional to the truck arrival time at stations outside the specified time windows, as shown in (20). Notice that we also add additional penalties for deviating from the optimal value of c_i identified in the inventory model. This is intended to ensure that, if possible, the tucks should try to arrive at time c_i to reduce the possibility for the customers to incur in shortage or send-back costs. Although, when considering the final value for the penalties of the route, only the penalties for deviating from the time window $[a_i, b_i]$ are considered. We weight the time deviations of the objective with the order quantity q_i, so as to give preference to larger orders. The routes with no deviations (positive or negative) are routes whose schedule guarantees that the truck arrives at every station at the desired time c_i. The objective of this model is to find the optimal truck arrival time at each station in the route so that the penalties are minimized.

The definition of the notation used for this model is given in Table 4. The route scheduling model is formulated as follows:

$$\min \quad \sum_{i \in N_r} (u_i + v_i + \gamma m_i + \delta n_i) q_i \tag{20}$$

$$\text{s.t.} \quad x_i - c_i = v_i - u_i \quad \forall i \in N_r, \tag{21}$$

$$x_1 \geq h + t_0, \tag{22}$$

$$x_i \geq x_{i-1} + s_{i-1} + t_{i-1} \quad \forall i \in N_r \setminus \{1\}, \tag{23}$$

$$x_i + m_i \geq a_i \quad \forall i \in N_r, \tag{24}$$

$$x_i - n_i \leq b_i \quad \forall i \in N_r, \tag{25}$$

$$a_i \leq x_i \leq m_l \quad \forall i \in N_r, \tag{26}$$

$$x_i \in R^+, \quad u_i \in R^+, \quad v_i \in R^+, \quad m_i \in R^+, \quad n_i \in R^+ \quad \forall i \in N_r. \tag{27}$$

Table 4 Mathematical notation and decision variables for the route scheduling model

Notation	Description
N_r	Subset of stations in given route r
h	Starting hour of the shift
m_l	Last hour of the shift
s_i	Service time for station i
t_i	Travel time from station i to station $i+1, \quad \forall i \in N_r$
a_i	Lower bound of the time window of station i
b_i	Upper bound of the time window of station i
γ	The penalty incurred if the truck arrives before a_i
δ	The penalty incurred if the truck arrives after b_i
x_i	Truck arrival time at station i
u_i	Negative time deviation from c_i of the truck arrival for station i
v_i	Positive time deviation from c_i of the truck arrival for station i
m_i	Amount of time the truck arrives before a_i for station i
n_i	Amount of time the truck arrives after b_i for station i

The objective function (20) minimizes penalties associated with the time deviations from the expected arrival times of route r. Constraints (21) define the positive and negative deviations of the delivery time from c_i for each gas station i. Constraints (22) ensure that the truck arrival time at the first station of each route occurs after the starting hour of the shift plus the travel time from terminal to the first station. Constraints (23) enforce that the truck arrival time at all the gas stations in the route (except for the first station) is greater than the arrival time plus the service time of the preceding station, and travel time between the stations. Constraints (24), (25) define the truck arrival time between the limits of the time window $[a_i, b_i]$. Constraints (26) require that the truck arrival time at station i lies within the earliest time window a_i and the maximum delivery time in one day m. Finally, constraints (27) define the decision variables.

Once the truck arrival time at each station of the route is known, one can compute the penalty for delivering petroleum to the gas stations in route r as:

$$\sum_{i \in N_r} (\gamma m_i + \delta n_i) q_i. \tag{28}$$

In addition to the penalties, the truck schedule can be recovered from the value of the $\mathbf{x}$'s variables. Notice that expression (28) can be used as the objective function of the route scheduling model instead of (20). We added the extra penalties $u_i + v_i$ for each $i \in N_r$ as it is beneficial for the customers to receive the product at a time close to the optimal point c_i.

5.4 ϵ-constraint method for the route scheduling problem

Typically, in the presence of time windows that overlap, or when some time windows are too narrow the time window requirements are difficult to satisfy. One of the possible situations that can occur is that a small set of the gas stations gets continuously penalized with late deliveries for several days of the time horizon, which may result in potential loss of such customers. To avoid this, we attempt to balance the late deliveries over the time horizon by applying the ϵ-constraint method (Miettinen 1999) for the multiple time periods. For time period t, we first minimize the maximum time window violation of any gas station by transforming the model as follows:

$$\min \quad z \tag{29}$$

$$\text{s.t.} \quad z \geq (u_i + v_i + \gamma m_i + \delta n_i) q_i + p^t_{cum} \quad \forall i \in N_r, \tag{30}$$

$$(\mathbf{x}, \mathbf{u}, \mathbf{v}, \mathbf{m}, \mathbf{n}) \in \Omega. \tag{31}$$

Here p^t_{cum} represents the cumulative penalty of station i in route r for day t which is given by

$$p^t_{cum} = \sum_{\tau=0}^{t-1} (\gamma m^\tau_i + \delta n^\tau_i) q_i \tag{32}$$

and Ω represents the constraint set given by (21)–(27). Also, variables $\mathbf{m}^\tau$ and $\mathbf{n}^\tau$ are assumed to be the optimal deviations of the scheduling problems of day τ. The solution found in the formulation above is denoted as z^*. We then minimize the summation of violations by replacing z by the optimal solution z^* in the formulation, which results in the following model.

$$\min \sum_{i \in N_r} (u_i + v_i + \gamma m_i + \delta n_i) q_i \tag{33}$$

$$\text{s.t.} \quad z^* \geq (u_i + v_i + \gamma m_i + \delta n_i) q_i + p^t_{cum} \quad \forall i \in N_r, \tag{34}$$

$$(\mathbf{x}, \mathbf{u}, \mathbf{v}, \mathbf{m}, \mathbf{n}) \in \Omega. \tag{35}$$

5.5 Variable neighborhood search

Variable Neighborhood Search (VNS) is a well-known metaheuristic that exploits systematically the idea of *neighborhood* change, both to identify local minima and to the escape from the valleys that contain them (Mladenovic and Hansen 1997). VNS typically combines several basic neighborhoods with a shaking component (i.e., a randomized shuffling procedures) within a simple, yet efficient search framework. Given its simplicity and adaptability, VNS has been commonly used over the last few decades to solve many variations of VRPs (Hemmelmayr et al. 2009; Dayarian et al. 2016; Kytöjoki et al. 2007), as well as other combinatorial optimization problems. There are other alternative specialized metaheuristics (e.g., genetic algorithms, GRASP, tabu search) that could be used in the context of this paper too. However, considering that the route improvement corresponds to a single component of the proposed approach, we resort to VNS for the main improvement component.

The main purpose of the VNS is then to improve the pool of candidate routes R from which the truck routing model—described in Sect. 5.6—selects the routes that will be used. As mentioned before, it is possible that the randomly generated set of routes R does not contain a sufficient amount of good candidates to yield an efficient routing plan; thus, the VNS is used in an attempt to purge set R. For this improving component, we use the traditional general VNS scheme, as proposed by Mladenovic and Hansen (1997).

The potential source of inefficiency in the candidate routes in R is twofold: either (1) the sequence in which the vehicle visits the customers (gas stations) in the route or (2) the group of customers comprising the route are not good enough to yield an efficient route schedule.

To alleviate these inefficiencies, there are several traditional *neighborhoods* (e.g., Prins 2004) that have been developed over the last couple of decades, which can be embedded within the VNS to improve the routes. The VNS is only applied over routes that serve more than four gas stations. We now describe the list of neighborhoods we used, which are divided into two categories: neighborhoods that are used to find improvements in the route sequences and neighborhoods that are used to find alternative routes with different members.

Sequence improvement: Given a route $r \in R$, the neighborhoods of this type correspond to alternative routes comprised by the same list of gas stations in r, but following a different sequence.

- *Position swaps*: This type of neighborhood contains the alternative routes to a route $r \in R$ in which any pair of gas stations i and j in r have their positions swapped.
- *Intra-route 2-opt*: This type of neighborhood contains the alternative routes to a route $r \in R$ in which the successors k and l of any pair of non-consecutive gas stations i and j are inverted.

Member improvement: Given routes r and p in R, the neighborhoods of this type correspond to alternative routes comprised by a list of different gas stations.

- *Member replacement*: This type of neighborhood contains the alternative routes to a route $r \in R$ in which any gas station i in r replaced by a gas station j not in R.
- *Member removal*: This type of neighborhood contains the alternative routes to a route $r \in R$ in which any gas station i in r is absent.
- *Member addition*: This type of neighborhood contains the alternative routes to a route $r \in R$ in which any gas station i not r is present.
- *Inter-route 2-opt*: This type of neighborhood contains the alternative routes to routes r and p in R in which the successors k and l of any pair of gas stations $i \in r$ and $j \in p$ are inverted.

If an exploration over a sequence improvement neighborhood is performed, since we do not change the gas stations that are visited, the loading of the compartments remains unchanged; therefore, only the route scheduling model is solved, to check the quality of the new route produced. On the other hand, if an exploration is performed over neighborhoods of the member improvement type, the truck loading model must be resolved to verify feasibility.

The VNS then replaces bad routes in R with better alternatives and augments R with newly generated routes.

5.6 Truck routing model

Finally, the objective of the truck routing model is to assign the optimal routes to the truck that will be used to fulfill the distribution plan. The trucks are allowed to make multiple deliveries provided that the delivery schedules of the routes do not overlap and the trucks return at the terminal by the allowable time limit. The notation and decision variable are defined in Table 5.

Table 5 Mathematical notation and decision variable for the truck routing model

Notation	Description
K	The truck set
R_k	The possible route set of that can be served by truck k, $\forall k \in K$
N_r	The set of stations served by route r, $\forall r \in R_k$
a_{ir}	A binary parameter equal to 1 if station i is served by route r, $\forall r \in R_k, \forall k \in K$
ρ_r	The profit of route r, $\forall r \in R$
q_r	The penalty of route r if route r is a "late" route, $\forall r \in R$
x_{rk}	A binary variable that takes the value of 1 if route r is operated by truck k, $\forall k \in K, \forall r \in R_k$

Then the routing model is given by:

$$\max \quad \sum_{r\in R_k}\sum_{k\in K}(\rho_r - q_r)x_{rk} \tag{36}$$

$$\text{s.t.} \quad \sum_{r\in R_k}\sum_{k\in K} a_{ir}x_{rk} = 1 \quad \forall i \in N, \tag{37}$$

$$\sum_{r\in R_k(t)} x_{rk} \leq 1 \quad \forall t \in T,\ k \in K, \tag{38}$$

$$x_{rk} \in \{0, 1\} \quad \forall r \in R_k,\ k \in K. \tag{39}$$

For this formulation, the objective function (36) maximizes the total profit, which is difference between the profit and penalty of the routes. Constraints (37) state that each station is visited exactly once. Constraints (38) ensure that the delivery times of the selected routes cannot overlap for each discrete time period t. We consider T as 24 hours here. For example, the time slot between 2 and 3 can only be occupied by one route for each truck assuring that the selected routes will not overlap in the time slot from 2 to 3. Constraints (39) define the decision variables.

6 Computational results

The proposed solution framework was coded in Java 8 with the API of CPLEX 12.6. All the experiments were performed on computing cluster equipped with twelve nodes, each having a 12-core Intel Xeon E5-2620 v3 2.4GHz processor, 128 GB of RAM, running Linux x86 64, CentOS 7.2. We first present the procedure for generating the test instances in Sect. 6.1; the performance measures of the solution framework are presented in Sect. 6.2; and then, the impact of several parameters on the generated distribution plans is studied thereafter, and the stochastic demand case is described towards the end of this section.

6.1 Test problems

In order to evaluate the proposed petroleum distribution framework, we use a test bed of 15 randomly generated instances with 50 customers adapted from Cornillier et al. (2009), after

Table 6 Parameters list

Parameter	Value used
The unit order cost (F)	$2.0
Fixed cost per delivery (k)	$100
Holding cost (h)	$0.002
Shortage cost (s)	$3.0
Send-back cost (p)	$0.8
Delivery time window limit hour	12
Average travel speed (mile/h)	40
Variable travel cost per mile	$1.05
Service time (min)	45
Penalty of truck arrival time before a_i	2.0
Penalty of truck arrival time after b_i	2.0

Table 7 Truck configurations

Type	No. of compartments	Capacities (gallons)	No. of trucks
1	6	4490, 1585, 2641, 2641, 1849, 2641	10
2	5	4226, 1585, 1585, 2641, 4226	5
3	4	4226, 2113, 3170, 3698	5

including the additional components introduced in this study. Among all the possible tank configurations provided in Cornillier et al. (2009), we use as a base model a medium size tank configuration that we latter vary to analyze its impact on the solution quality (see Sect. 6.3). We converted the petroleum quantity units from liters to gallons and the distance units from kilometers to miles as per use in the typical context of U.S. scenarios. Furthermore, we introduced additional information required for the instances to be used in the context of our approach. The parameters used regarding the inventory costs and penalties can be found in Table 6. We selected those figures to match some of the values used by Cornillier et al. (2007). We consider a fleet of 20 trucks whose compartment compositions are given in Table 7. As for the demands, we generate demand rates for the test instances so that each station orders every one or two days. All this to see the patterns that emerge regarding the time windows and the impact of balancing late deliveries.

We further study the impact that the tank capacity of the gas stations and the number of randomly generated routes serving more than three gas stations have on the solutions; see Sects. 6.3 and 6.4, respectively. We introduced a correlation between the demand rates and the tank capacities of the gas station to reflect the fact that gas stations with larger tanks are expected to have higher demand rates. The tank capacities we used are given in Table 8, labeled C1 to C5. Furthermore, we also vary the total number of randomly generated routes with more than three customers. We solve the replenishment problem for five scenarios labeled R1 to R5 in which we generate 100, 1000, 10,000, 20,000, and 50,000 of such routes, respectively.

Table 8 Tank capacities

Scenarios	Tank 1	Tank 2	Tank 3
C1	5547	8395	9246
C2	5151	7795	8585
C3	4755	7196	7925
C4	4358	6596	7264
C5	3962	5996	6604

6.2 Performance of the solution framework

We tested the performance of the proposed approach over the 15 instances described in Sect. 6.1. We performed 50 different runs for each instance, varying the tank capacities, the demand rates, and the number of routes that are randomly generated. The data presented in Table 9 corresponds to the averages and standard deviations of the 50 runs and includes: for the gas stations, (1) the expected cost estimated by the inventory models, (2) the real cost obtained by calculating the costs based on the routes given by the distribution model and (3) the average deviation in hours of the delivery times with respect to the optimal ordering point per day. For the transporter, Table 10 presents the average profit and number of late deliveries per day. The real cost for the gas stations refers to the actual cost incurred by a station including all ordering cost, holding cost, shortage cost, and send-back cost after knowing when the station actually receives the petroleum by solving the truck routing problem.

We observe that the real costs are typically higher than the expected costs in all instances, despite the fact that the average tardiness for a gas station is less than one hour (the minimum expected length of the time windows). This can be explained by noting that it is more common for the transporter to arrive after the ideal delivery time c_i. Furthermore, we can see that even though 20 trucks are used to deliver all the demands, the number of late deliveries still ranges from 13.57 to 18.47 times everyday on average. The main reason is that most stations request similar and tight time windows, so it is difficult for the transporter to deliver all demands within the time windows even after using a large fleet of trucks. Nevertheless, as mentioned before, the average late deliveries are not that far from the selected time windows.

6.3 Impact of the tank capacity and demand rates

The tank capacities of the stations have a strong effect on the total order quantities because of the correlation that was introduced between the tank size and the demand rates. Consequently, there is also an effect on the total petroleum quantity distributed by the transporter. We ran our algorithms for solving the 15 instances using different levels of the tank capacities (C1-C5). Figure 8 depicts the average customer costs and the transporter profits for the different scenarios. Unsurprising, the higher the expected and real cost for the stations occur when the tank capacities are increased. From the station perspective, this is because the demands are larger which potentially increases the inventory costs.

When analyzing the late deliveries and the length of the tardiness periods, it is interesting to see a higher number of late deliveries and longer periods of tardiness when the tank capacities are reduced, see Fig. 9a, b. With smaller tanks, the gas stations tend to place smaller orders more frequently. This in turn results in having trucks visiting more stations per route to cope with the increased number of order requests. In consequence, when more stations are visited by the same route, it is more likely that a larger number of late deliveries and longer periods

Table 9 Performance of the solution framework from the customer's point of view

Instance	Expected cost		Real cost		Tardiness	
	Avg.	SD	Avg.	SD	Avg.	SD
L1	10,146.98	1614.99	10,212.40	1638.37	0.24	0.07
L2	9548.61	1147.57	9617.71	1137.19	0.23	0.02
L3	9243.38	1062.17	9305.14	1060.09	0.23	0.05
L4	10,096.47	1575.94	10,171.07	1587.71	0.23	0.04
L5	8986.65	1051.71	9049.72	1058.33	0.27	0.05
L6	9407.33	1079.35	9461.47	1082.39	0.20	0.02
L7	9616.34	1111.11	9663.46	1103.09	0.24	0.05
L8	9657.46	1165.27	9737.38	1166.11	0.26	0.03
L9	9560.12	1152.84	9606.41	1150.65	0.24	0.02
L10	9202.03	1079.84	9255.07	1086.33	0.20	0.03
L11	9394.28	1123.26	9461.38	1109.23	0.21	0.05
L12	9446.51	1138.26	9483.03	1130.43	0.18	0.02
L13	10,258.40	1759.71	10,315.08	1766.91	0.21	0.05
L14	9299.88	958.32	9360.98	969.61	0.27	0.06
L15	9694.92	1160.46	9744.85	1154.08	0.20	0.04

Table 10 Performance of the solution framework from the transporter's point of view

Instance	Profit		No. late deliveries	
	Avg.	SD	Avg.	SD
L1	2630.64	452.25	14.90	2.41
L2	3062.76	560.54	16.52	1.60
L3	2938.40	543.32	16.97	0.93
L4	2919.24	591.06	15.37	1.75
L5	2617.84	517.48	18.47	1.00
L6	2807.49	581.52	13.57	1.39
L7	2750.44	464.26	17.23	1.29
L8	3154.71	567.90	17.55	2.05
L9	2770.68	489.22	17.42	1.70
L10	2608.05	515.57	16.74	1.13
L11	2765.56	580.91	15.59	2.32
L12	2461.84	596.44	14.20	0.51
L13	2777.91	461.55	15.85	2.66
L14	2753.28	504.21	16.17	2.43
L15	2923.77	547.89	16.90	0.55

of tardiness occur because the trucks need to finish several deliveries before visiting the last few stations in the route. This leads to longer waiting times for the rest of stations in such a route.

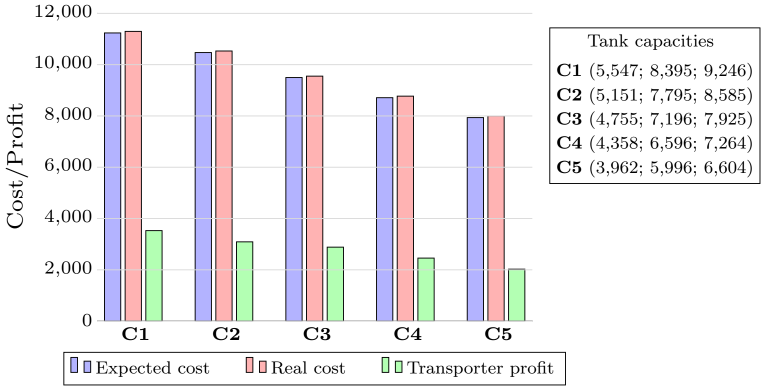


Fig. 8 The economic impact of varying the tank capacity level

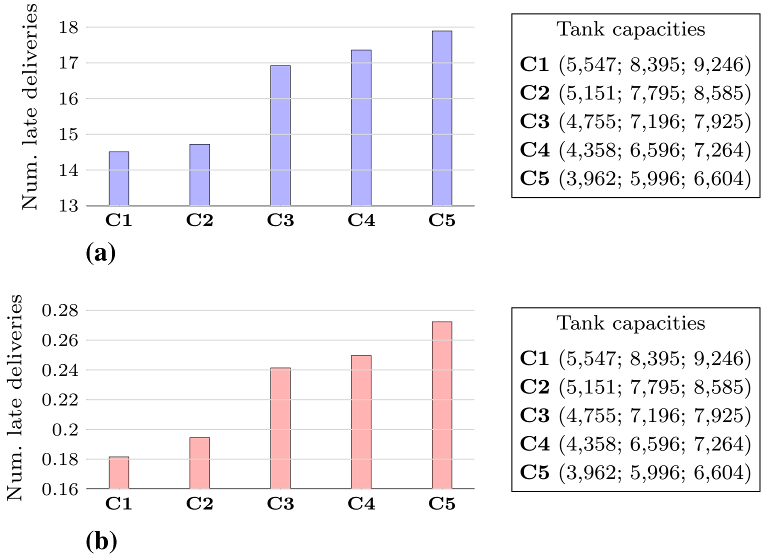


Fig. 9 The impact on the delivery schedules. **a** The impact on the number of late deliveries. **b** The impact on the tardiness

6.4 Impact of the number of randomly generated routes

In order to see the impact that using different numbers of randomly generated routes serving more than three gas stations has on the quality of the solution, we tested the five scenarios R1 to R5 in which we generate 100, 1000, 10,000, 20,000, and 50,000 of such routes. We also vary the instance for the different tank capacity levels, using for testing cases C1 and C5. Table 11 shows that the percentage of profit improvement and the length of the tardiness periods do not improve significantly in higher level of tank capacity (C1). This is because the order quantities of the gas stations are relatively high. Thus, there is a low probability that trucks are able to load the demands of four stations or more in the same route. However, in the lower level of tank capacity (C5), the total order quantity may be low, so trucks have higher chance to visit more stations in the same route. The profit clearly increases with the

Table 11 The impact of the number of randomly generated routes

Instance	C1				C5			
	Profit	%Im	#LD	Tardiness	Profit	%Im	#LD	Tardiness
R1	3528.97	0.0000	14.52	0.18	2014.86	0.0000	17.62	0.26
R2	3528.64	– 0.0093	14.50	0.18	2014.81	– 0.0025	17.63	0.26
R3	3528.71	0.0020	14.54	0.18	2024.00	0.4564	17.77	0.27
R4	3528.87	0.0047	14.48	0.18	2022.14	– 0.0920	17.82	0.27
R5	3528.85	– 0.0055	14.50	0.18	2026.55	0.2181	17.86	0.27

%Im: percentage of improvement; #LD: number of late deliveries

Table 12 Tardiness when the late deliveries are balanced among the gas stations

Instance	Avg.	Max	Min	Max–min	No. of late deliveries
R5C1	0.48	3.38	0.00	3.38	12.00
R5C2	0.90	5.07	0.00	5.07	14.28
R5C3	0.92	5.80	0.00	5.80	15.28
R5C4	0.53	3.97	0.00	3.97	15.57
R5C5	0.62	4.43	0.00	4.43	14.71

use of more of these long routes. Note that the station tardiness increases by increasing the number of route generation in C5, but not in C1 due to the higher chance that trucks will visit up to four stations causing the late arrival time to the later stations in order within a route.

6.5 The importance of balancing late deliveries among the customers

To evaluate the effect of balancing the late deliveries among the customers has on the total tardiness of the distribution plan, we use Instance L12—which is the one having the longest periods of tardiness, as presented in Table 9. We select the number of randomly generated routes to be 50,000 (R5), to avoid the performance from affecting by choosing poor routes. Furthermore, we run the experiments for all the tank capacities C1 to C5. The results found when using our approach to attempt balancing late delivery are presented in Table 12. Additionally, we modify the optimization framework removing the ϵ-constraints approach introduced in Sect. 5.3, in an attempt to solve the problem without balancing the late deliveries. The results of this approach are presented in Table 13. It can be seen from these results that using the ϵ-constraints approach lowers the average tardiness and the difference between the maximum and minimum tardiness, which indicates the proposed policy tends to produce distribution schemes for which the overall tardiness is reduced. In addition, the average number of late delivery is increased using this policy since it attempts to deliver to customers, which have no late deliveries previously, late in order to balance the late deliveries.

6.6 Demand rate sampling scenarios

To capture the stochastic effect of the demand rates (see Sect. 4.1), we randomly generated 50 demand scenarios for one of the 50-node instances described above. The demand rates for

Table 13 Tardiness when the late deliveries are not balanced among the gas stations

Instance	Avg.	Max	Min	Max–min	No. of late delivery
R5C1	0.51	3.88	0.00	3.88	11.42
R5C2	0.89	5.07	0.00	5.07	14.42
R5C3	1.05	5.81	0.00	5.81	15.14
R5C4	0.55	4.00	0.00	4.00	14.14
R5C5	0.62	4.43	0.00	4.43	14.42

Table 14 Profit obtained for each scenario

Scenario	Profit	Scenario	Profit
1	4595.56	26	4466.73
2	4659.69	27	4424.73
3	4763.96	28	4747.18
4	4749.05	29	4475.51
5	4612.01	30	4643.50
6	4723.65	31	4639.29
7	4620.06	32	4589.68
8	4580.21	33	4688.13
9	4610.80	34	4606.05
10	4608.12	35	4716.73
11	4755.54	36	4689.91
12	4668.05	37	4580.93
13	4665.33	38	4596.07
14	4649.21	39	4625.47
15	4497.26	40	4652.97
16	4502.11	41	4581.80
17	4548.07	42	4616.13
18	4667.25	43	4533.49
19	4601.22	44	4687.70
20	4714.53	45	4583.50
21	4420.65	46	4541.64
22	4726.11	47	4585.23
23	4712.79	48	4655.09
24	4613.48	49	4548.45
25	4581.62	50	4599.66

all scenarios were randomly generated for each customer from independent normal distributions. The mean of the distribution was set to be equal to the demand used in the previous experiments. We generated the random demand rates varying the standard deviation of the distribution, using values ranging from 5 to 10% from the mean. Each of such scenarios was then solved individually using the proposed static model. Table 14 presents the best solution obtained for each of the randomly generated scenarios. The routes and schedules produced for the 50 instances were quite different due to the variations on the demand quantities and time windows, but the profit obtained was in general consistent across all the instances.

Finally, to identify if there exists a set of routes that performs generally well for most of the generated demand scenarios, we take the routes produced by the static execution of each scenario and tested them under the demands and time windows of all the other scenarios. It is important to notice that performing this comparison is not necessarily an easy task. For instance, one can notice that the routes in one scenario may not be feasible for others, since the stations that require a replenishment may differ in between scenarios. Thus, the profit calculation must be adjusted accordingly to produce a fair comparison. More specifically, if certain particular stations in a scenario do not require replenishment in another scenario, those stations would be simply skipped in the routes that cover them. For instance, if a route for scenario A stops and replenishes stations $\{a, b, c\}$, as in Fig. 10a, but station c is not required to be served in scenario B. We update the route used in scenario A to evaluate the profit in scenario B after removing c, as shown in Fig. 10b. On the other hand, if a station that is not served under a given scenario, say station c in scenario C requires to be served in other scenario, say D, we add the additional cost to this single trip delivery as seen in Fig. 10c, d, respectively.

Table 15 lists the mean and standard deviation of all scenarios and the routes with the highest profit in the scenario will be adopted. The average performance for the route set of each scenario is significantly inferior to the best solution found (cf. Table 14), which is not surprising given the changes on the demands.

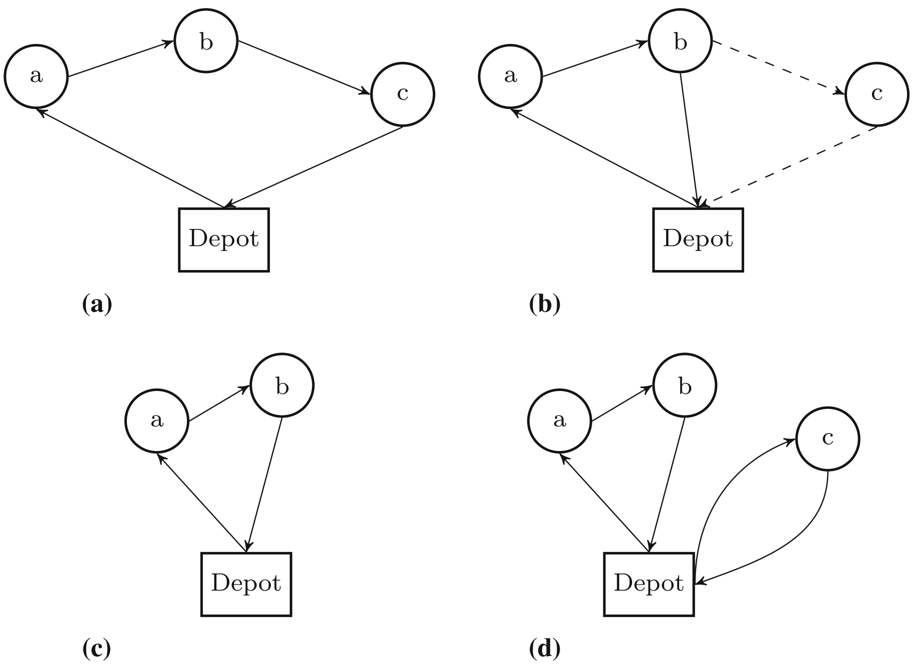


Fig. 10 Example of scenarios with different routes. **a** Route in scenario A. **b** Resulting route if used in scenario B. **c** Route in scenario C. **d** Resulting route if used in scenario D

Table 15 Average profit and standard deviation of each scenario when tested with the demands of the other scenarios

Scenario	Mean	SD	Scenario	Mean	SD
1	3741.63	200.59	26	3815.67	197.27
2	3706.83	289.46	27	3429.47	226.21
3	3816.63	249.82	28	3744.73	235.39
4	3719.72	328.00	29	3830.81	158.64
5	3591.53	286.64	30	3671.13	278.11
6	3792.46	255.49	31	3801.36	255.18
7	3768.83	251.85	32	3827.85	197.73
8	3802.96	194.98	33	3498.52	245.66
9	3623.12	210.92	34	3851.11	238.98
10	3821.16	223.48	35	3851.09	209.93
11	3775.28	239.76	36	3669.84	277.04
12	3615.46	262.10	37	3572.30	282.95
13	3720.29	275.46	38	3853.64	239.02
14	3764.26	226.97	39	3706.32	241.00
16	3607.95	236.79	41	3755.07	235.44
17	3773.09	226.71	42	3689.00	257.45
18	3653.66	248.74	43	3255.67	346.17
19	3634.38	269.36	44	3602.90	246.24
20	3640.23	270.08	45	3667.49	296.24
21	3582.56	217.62	46	3742.34	233.66
22	3797.86	236.23	47	3721.58	273.68
23	3838.95	231.63	48	3673.21	278.65
24	3782.45	265.87	49	3768.53	242.61
15	3645.39	203.65	40	3487.10	269.77
25	3924.43	224.73	50	3706.99	228.34

7 Concluding remarks

We have developed a methodological framework for designing the daily distribution and replenishment operations of petroleum products over a weekly horizon. The proposed models consider the option of allowing late deliveries in the cases where the expected time windows selected by the gas stations are too close to each other and potentially overlapping. The proposed approach consists of solving a series of optimization models for identifying the gas station petroleum demands and time windows, as well as for designing the distribution logistics. One of the main features of the optimization models is that the late deliveries are balanced among the customers in order to mitigate the overall customer dissatisfaction. The proposed approach was tested on a set of randomly generated problems adapted from the literature. The empirical results showed that the proposed approach is a viable option for designing distribution plans in contexts where the variability and complexity of the customer orders results in having unavoidable late deliveries. Further directions would be to embed the truck loading and the route scheduling models within a column generation approach that, instead of testing a random sample of the possible routes, generates the optimal candidate routes sequentially, as needed.

References

Al-Khayyal, F., & Hwang, S. J. (2007). Inventory constrained maritime routing and scheduling for multi-commodity liquid bulk, part i: Applications and model. *European Journal of Operational Research*, *176*(1), 106–130.

Andersson, H., Hoff, A., Christiansen, M., Hasle, G., & Løkketangen, A. (2010). Industrial aspects and literature survey: Combined inventory management and routing. *Computers & Operations Research*, *37*(9), 1515–1536.

Avella, P., Boccia, M., & Sforza, A. (2004). Solving a fuel delivery problem by heuristic and exact approaches. *European Journal of Operational Research*, *152*(1), 170–179.

Baldacci, R., Mingozzi, A., & Roberti, R. (2012). Recent exact algorithms for solving the vehicle routing problem under capacity and time window constraints. *European Journal of Operational Research*, *218*(1), 1–6.

Bausch, D. O., Brown, G. G., & Ronen, D. (1995). Consolidating and dispatching truck shipments of mobil heavy petroleum products. *Interfaces*, *25*(2), 1–17.

Ben Abdelaziz, F., Roucairol, C., & Bacha, C. (2002). Deliveries of liquid fuels to SNDP gas stations using vehicles with multiple compartments. In *2002 IEEE international conference on systems, man and cybernetics* (Vol. 1, pp. 478–483).

Bertsimas, D. J. (1992). A vehicle routing problem with stochastic demand. *Operations Research*, *40*(3), 574–585.

Bianchi, L., Birattari, M., Chiarandini, M., Manfrin, M., Mastrolilli, M., Paquete, L., et al. (2006). Hybrid metaheuristics for the vehicle routing problem with stochastic demands. *Journal of Mathematical Modelling and Algorithms*, *5*(1), 91–110.

Bräysy, O., & Gendreau, M. (2005a). Vehicle routing problem with time windows, part I: Route construction and local search algorithms. *Transportation Science*, *39*(1), 104–118.

Bräysy, O., & Gendreau, M. (2005b). Vehicle routing problem with time windows, part II: Metaheuristics. *Transportation Science*, *39*(1), 119–139.

Brown, G. G., Ellis, C. J., Graves, G. W., & Ronen, D. (1987). Real-time, wide area dispatch of mobil tank trucks. *Interfaces*, *17*(1), 107–120.

Brown, G. G., & Graves, G. W. (1981). Real-time dispatch of petroleum tank trucks. *Management Science*, *27*(1), 19–32.

Campbell, A., Clarke, L., Kleywegt, A., & Savelsbergh, M. (1998). The inventory routing problem. In T. G. Crainic & G. Laporte (Eds.), *Fleet management and logistics, centre for research on transportation* (pp. 95–113). New York: Springer US. chap 4,

Campbell, A., & Savelsbergh, M. W. P. (2004). A decomposition approach for the inventory-routing problem. *Transportation Science*, *38*(4), 488–502.

Caramia, M., & Guerriero, F. (2009). A milk collection problem with incompatibility constraints. *Interfaces*, *40*(2), 130–143.

Chajakis, E. D., & Guignard, M. (2003). Scheduling deliveries in vehicles with multiple compartments. *Journal of Global Optimization*, *26*(1), 43–78.

Christiansen, C. H., & Lysgaard, J. (2007). A branch-and-price algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research Letters*, *35*(6), 773–781.

Christiansen, M., Fagerholt, K., Flatberg, T., Haugen, Ø., Kloster, O., & Lund, E. H. (2011). Maritime inventory routing with multiple products: A case study from the cement industry. *European Journal of Operational Research*, *208*(1), 86–94.

Coelho, L. C., Cordeau, J. F., & Laporte, G. (2013). Thirty years of inventory routing. *Transportation Science*, *48*(1), 1–19.

Cohn, A., Root, S., Wang, A., & Mohr, D. (2007). Integration of the load-matching and routing problem with equipment balancing for small package carriers. *Transportation Science*, *41*(2), 238–252.

Committee on America's Energy Future, National Academy of Sciences. (2009). *America's energy future: Technology and transformation*. Washington: National Academies Press.

Cordeau, J. F., Laporte, G., Savelsbergh, M. W. P., & Vigo, D. (2007). Vehicle routing. In C. Barnhart & G. Laporte (Eds.), *Transportation* (Vol. 14, pp. 367–428). Amsterdam: Elsevier.

Cornillier, F., Boctor, F. F., Laporte, G., & Renaud, J. (2007). An exact algorithm for the petrol station replenishment problem. *Journal of the Operational Research Society*, *59*(5), 607–615.

Cornillier, F., Boctor, F. F., Laporte, G., & Renaud, J. (2008). A heuristic for the multi-period petrol station replenishment problem. *European Journal of Operational Research*, *191*(2), 295–305.

Cornillier, F., Boctor, F. F., & Renaud, J. (2012). Heuristics for the multi-depot petrol station replenishment problem with time windows. *European Journal of Operational Research*, *220*(2), 361–369.

Cornillier, F., Laporte, G., Boctor, F. F., & Renaud, J. (2009). The petrol station replenishment problem with time windows. *Computers & Operations Research*, *36*(3), 919–935.
Dantzig, G. B., & Ramser, J. H. (1959). The truck dispatching problem. *Management Science*, *6*(1), 80–91.
Dayarian, I., Crainic, T. G., Gendreau, M., & Rei, W. (2016). An adaptive large-neighborhood search heuristic for a multi-period vehicle routing problem. *Transportation Research Part E: Logistics and Transportation Review*, *95*(Supplement C), 95–123.
Derigs, U., Gottlieb, J., Kalkoff, J., Piesche, M., Rothlauf, F., & Vogel, U. (2011). Vehicle routing with compartments: Applications, modelling and heuristics. *OR Spectrum*, *33*(4), 885–914.
Eksioglu, B., Vural, A. V., & Reisman, A. (2009). The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering*, *57*(4), 1472–1483.
Fallahi, A. E., Prins, C., & Wolfler Calvo, R. (2008). A memetic algorithm and a tabu search for the multi-compartment vehicle routing problem. *Computers & Operations Research*, *35*(5), 1725–1741.
Federal Motor Carrier Safety Administration. (2015). *Rules of practice for motor carriers, intermodal equipment provider, broker, freight forwarder, and hazardous materials proceedings*. Technical report, U.S. Department of Transportation.
Federgruen, A., & Simchi-Levi, D. (1995). Analysis of vehicle routing and inventory-routing problems. In M. O. Ball, T. L. Magnanti, C. Monma, & G. L. Nemhauser (Eds.), *Handbooks in operations research and management science* (Vol. 8, pp. 297–373). Amsterdam: Elsevier.
Feillet, D. (2010). A tutorial on column generation and branch-and-price for vehicle routing problems. *4OR*, *8*(4), 407–424.
Feillet, D., Gendreau, M., Medaglia, A. L., & Walteros, J. L. (2010). A note on branch-and-cut-and-price. *Operations Research Letters*, *38*(5), 346–353.
Gendreau, M., & Tarantilis, C. D. (2010). *Solving large-scale vehicle routing problems with time windows: The state-of-the-art*. Montrea: CIRRELT.
Hallenbeck, M., Rice, M., Smith, B., Cornell-Martinez, C., & Wilkinson, J. (1997). *Vehicle volume distributions by classification*. Technical report.
Hemmelmayr, V. C., Doerner, K. F., & Hartl, R. F. (2009). A variable neighborhood search heuristic for periodic routing problems. *European Journal of Operational Research*, *195*(3), 791–802.
Henke, T., Speranza, M. G., & Wäscher, G. (2015). The multi-compartment vehicle routing problem with flexible compartment sizes. *European Journal of Operational Research*, *246*(3), 730–743.
Jetlund, A. S., & Karimi, I. A. (2004). Improving the logistics of multi-compartment chemical tankers. *Computers & Chemical Engineering*, *28*(8), 1267–1283.
Kallehauge, B., Larsen, J., Madsen, O. B. G., & Solomon, M. M. (2005). Vehicle routing problem with time windows. In D. J. Gand Desrosiers & M. M. Solomon (Eds.), *Column generation* (pp. 67–98). New York: Springer US (book section 3).
Kytöjoki, J., Nuortio, T., Bräysy, O., & Gendreau, M. (2007). An efficient variable neighborhood search heuristic for very large scale vehicle routing problems. *Computers & Operations Research*, *34*(9), 2743–2757.
Lahyani, R., Coelho, L. C., Khemakhem, M., Laporte, G., & Semet, F. (2015). A multi-compartment vehicle routing problem arising in the collection of olive oil in Tunisia. *Omega*, *51*, 1–10.
Lozano, L., Duque, D., & Medaglia, A. L. (2016). An exact algorithm for the elementary shortest path problem with resource constraints. *Transportation Science*, *50*(1), 348–357.
Mendoza, J. E., Castanier, B., Guéret, C., Medaglia, A. L., & Velasco, N. (2010). A memetic algorithm for the multi-compartment vehicle routing problem with stochastic demands. *Computers & Operations Research*, *37*(11), 1886–1898 (metaheuristics for Logistics and Vehicle Routing).
Miettinen, K. (1999). Nonlinear multiobjective optimization. Kluwer Academic Publishers.
Mladenovic, N., & Hansen, P. (1997). Variable neighborhood search. *Computers & Operations Research*, *24*(11), 1097–1100.
Moin, N. H., & Salhi, S. (2006). Inventory routing problems: A logistical overview. *Journal of the Operational Research Society*, *58*(9), 1185–1194.
Muyldermans, L., & Pang, G. (2010). On the benefits of co-collection: Experiments with a multi-compartment vehicle routing algorithm. *European Journal of Operational Research*, *206*(1), 93–103.
Ng, W. L., Leung, S. C. H., Lam, J. K. P., & Pan, S. W. (2007). Petrol delivery tanker assignment and routing: A case study in hong kong. *Journal of the Operational Research Society*, *59*(9), 1191–1200.
Nguyen, L. X., Zhang, J., Zhou, S., Beningo, S., Chambers, M., Ford, C., & Notisand, K. (2015). *National transportation statistics*. Technical report, U.S. Department of Transportation's Bureau of Transportation Statistics.
Pillac, V., Gendreau, M., Guéret, C., & Medaglia, A. L. (2013). A review of dynamic vehicle routing problems. *European Journal of Operational Research*, *225*(1), 1–11.

Prins, C. (2004). A simple and effective evolutionary algorithm for the vehicle routing problem. *Computers & Operations Research*, *31*(12), 1985–2002.
Reed, M., Yiannakou, A., & Evering, R. (2014). An ant colony algorithm for the multi-compartment vehicle routing problem. *Applied Soft Computing*, *15*, 169–176.
Simchi-Levi, D., Chen, X., & Bramel, J. (2005). *The logic of logistics*. New York: Springer.
Spliet, R., & Desaulniers, G. (2015). The discrete time window assignment vehicle routing problem. *European Journal of Operational Research*, *244*(2), 379–391.
Spliet, R., & Gabor, A. F. (2015). The time window assignment vehicle routing problem. *Transportation Science*, *49*(4), 721–731.
Strocko, E., Sprung, M., Nguyen, L., Rick, C., & Sedor, J. (2014). *Freight facts and figures 2013*. Technical report. Federal Highway Administration
Sungur, I., Ordoñez, F., & Dessouky, M. (2008). A robust optimization approach for the capacitated vehicle routing problem with demand uncertainty. *IIE Transactions*, *40*(5), 509–523.
Toth, P., & Vigo, D. (2014). *Vehicle routing: Problems, methods, and applications* (2nd ed.). Philadelphia: SIAM.
van der Bruggen, L., Gruson, R., & Salomon, M. (1995). Reconsidering the distribution structure of gasoline products for a large oil company. *European Journal of Operational Research*, *81*(3), 460–473.