

Received June 1, 2020, accepted June 18, 2020, date of publication June 25, 2020, date of current version July 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004849

# Two-Stage Stochastic Programming for the Refined Oil Secondary Distribution With Uncertain Demand and Limited Inventory Capacity

ZHENGPING LI<sup>1,2</sup>, YUWEI ZHANG<sup>1,2</sup>, AND GUOWEI ZHANG<sup>3,4</sup>

<sup>1</sup>School of Information, Beijing Wuzi University, Beijing 101149, China

<sup>2</sup>School of Management Engineering, Capital University of Economics and Business, Beijing 100070, China

<sup>3</sup>Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

<sup>4</sup>School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding authors: Zhenping Li (lizhenping@bwu.edu.cn) and Yuwei Zhang (morbeny@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71771028, in part by the Beijing Natural Science Foundation under Grant Z180005, in part by the Beijing Municipal University's High-Level Innovation Team Construction Project under Grant IDHT20180510, and in part by the Beijing Science and Technology Innovation Service Capacity Building-High-Grade, Precision and Advanced Subject Construction Project.

**ABSTRACT** In recent years, more and more oil companies adopt initiative delivery mode to make the refined oil secondary distribution scheme. In this work, we focus on the optimization problem of refined oil secondary distribution based on the initiative distribution mode considering stochastic demand and the limited inventory capacity of each petrol station. We present a two-stage stochastic programming model that determines the replenishment quantity of each petrol station based on its existing stock and the available supply quantity of each oil depot, as well as transportation schedule. When the uncertainty in demand can be captured via a finite set of scenarios, the two-stage stochastic programming model is transformed into an equivalent deterministic mixed integer programming model that can be efficiently solved by CPLEX solver. The effectiveness of the two-stage stochastic programming model is verified by simulation on extensive computer-generated instances. To solve practical problems with a large number of scenarios, we propose a method to reduce the problem scale by merging similar scenarios. We demonstrate that compared to the optimal solution obtained from the model with all scenarios, the gap corresponding to the model with merged scenarios is always less than 1%. The results of the sensitivity analysis show that an increase in the inventory capacity leads to a decrease in the total cost within a certain range. The results of this study can help companies making refined oil secondary distribution plan.


**INDEX TERMS** Refined oil secondary distribution, initiative distribution mode, transportation, stochastic demand, limited inventory capacity, two-stage stochastic programming.

## I. INTRODUCTION

Refined oil logistics includes two stages: primary distribution and secondary distribution, which link refineries, transfer depots and petrol stations or clients. Primary distribution refers to the process of transporting refined oil from refineries to oil depots, which is mainly achieved by oil vessels, railway tankers, oil pipeline and other channels in a large-quantity low-frequency mode. Secondary distribution is the process of

transporting the refined oil from oil depots to petrol stations or end-users, which is mainly completed by using petroleum tank vehicles through road transportation in a small-quantity high-frequency mode. The secondary distribution of refined oil accounts for a large proportion of the whole refined oil logistics and results in high logistics costs. Therefore, it is an important means to reduce the logistics cost by optimizing the secondary distribution process of refined oil.

Generally, after arriving at central oil depots through the primary distribution, the refined oil is temporarily stored in the large storage tanks, then transported to each petrol

The associate editor coordinating the review of this manuscript and approving it for publication was Md Asaduzzaman .

station or end-user by secondary distribution to meet their demands. Since the refined oil is dangerous liquid, it needs a tightly closed container or isolated compartment to store and transfer. And the oil loaded in one compartment of a vehicle is not allowed to be unloaded to multiple storage tanks. Making a refined oil secondary distribution plan is to determine the quantity of refined oil to be transported from oil depots to petrol stations, the types of vehicles to be used and their distribution routes, the arriving time of each vehicle at the petrol stations it serves for, and so on, so as to meet the demand of petrol stations or end-users.

There are two modes to be selected for making a refined oil secondary distribution plan: passive delivery mode and initiative delivery mode. In the passive delivery mode, inventory management of each petrol station is controlled by itself rather than oil company/vendor. All petrol stations report their demands to the oil company daily at a certain time. The oil company arranges a refined oil secondary distribution plan for the next day according to the demand data reported by each petrol station.

Adopting passive delivery mode may result in imbalance in resource allocation when the total supply quantity of refined oil from all oil depots is in shortage. On the other hand, when oil depots have adequate stocks, they would not have enough vacant tanks to accept the refined oil from refineries next day if the petrol stations do not report their high demand plans.

Over the last few years, more and more industries began to adopt initiative delivery mode based on vendor managed inventory (VMI) (Marquès *et al.* (2010) [1], Danese(2011) [2]). When adopting the initiative delivery mode, the oil company/supplier determines the replenishment quantity of refined oil for each petrol station based on its existing stock and historical sales volume, and then arrange the distribution plan and vehicle scheduling scheme. Since all managerial decisions, including the supply volume of oil depots, the replenishment quantity of each petrol station, and the transportation vehicles to be used, are controlled by the oil company in initiative delivery mode, the oil company can overall optimize the refined oil secondary distribution plan so as to effectively reduce the inventory and transportation cost of the whole distribution system (Marquès(2010) *et al.* [1]).

In initiative delivery mode, the sales volume of each petrol station at each time period is random, therefore, it is difficult for the oil company to precisely predict the sales volume of each petrol station during next day (or next time period) when making distribution plan. On the other hand, the capacity of each oil storage tank in each petrol station is limited. When implementing the initiative distribution plan, each petrol station might encounter one of the following three scenarios.

(1) The replenishment volume to a petrol station plus its existing stock is sufficient for next day's sales demand, and the volume of unsold oil is less than the capacity of its storage tanks, so they can be stored in the oil storage tanks.

(2) The replenishment volume to a petrol station plus its existing stock is sufficient for next day's sales demand, and the volume of unsold oil exceeds the spare capacity of storage

tanks, so the exceeding part of oil cannot be stored in the storage tanks. This situation will inevitably result in the cost of handling surplus refined oil.

(3) The replenishment volume to a petrol station plus its existing stock is less than next day's sales demand, which causes stockouts in next day. This situation will result in shortage cost.

In the process of implementing the oil secondary distribution plan, when scenario (1) occurs, there are no extra costs incurred, since the sales and the storage capacity of the petrol station are normal. When scenario (2) occurs, the petrol station needs to pay extra costs for handling the surplus oil, which is defined as surplus cost in this work. When scenario (3) occurs, the petrol station needs to pay extra costs for stockouts, which is defined as shortage cost.

When the next day's sales volume of each petrol station is unknown, a key issue for making the refined oil secondary distribution scheme based on the initiative delivery mode is how to determine the replenishment quantity and plan the vehicles distribution routes based on the existing stock of each petrol station and the available supply quantity of each oil depot, such that the sum of transportation costs and the expected recourse costs is minimized.

The remainder of this paper is organized as follows. In section II, we briefly review the related literature and highlight our contributions. In section III, we formulate the refined oil secondary distribution problem as a two-stage stochastic programming, and transform it into an equivalent mixed integer programming model when the number of scenarios is finite. In section IV, we measure the relevance of using a stochastic programming approach. Section V presents the computational experiments and numerical results to verify the effectiveness of the proposed model. Finally, we summarize our work and point out some potential extensions for future research in section VI.

## II. RELATED RESEARCH

The refined oil secondary distribution problem (ROSDP) is a well-known research topic. As pointed out by Popović *et al.* (2012) [4] and Cornillier *et al.* (2011) [5], there were many researches taking into account the inter-relationship between inventory management and transportation. Carotenuto *et al.* (2015) [6] decomposed the final distribution of fuel oil into determining the weekly replenishment plan for each station and arranging petrol stations' visiting sequences (vehicle routes) for each day of the week. Recent studies that focus on multi-period fuel replenishment can also be found in Charusakwong and Lohatepanont (2016) [7], Vidović *et al.* (2014) [8]. The problem can be extended to a vehicle towing a trailer with several compartments and visiting multi-depots. Related research efforts on these topics have been intensified by Zhang *et al.* (2016) [9] and Li *et al.* (2018) [10] considered a heterogeneous fleet of vehicles with multi-compartment used for fuel distribution involving unloading sequence. Several studies from this research area have considered

Neighborhood Search (NS) algorithms (Popović *et al.* [4]; Vidović *et al.* [8]; Li *et al.* [10]), Genetic Algorithms (Carotenuto *et al.* [6]; Zhang *et al.* [9]) and Tabu Search heuristics (Benantar *et al.* [11]). Additionally, a branch-and-price algorithm for exact optimization of the fuel delivery problem was proposed by Avella *et al.* (2004) [12].

Due to the complex characteristics of ROSDP, Li *et al.* (2016) [10] and Hanczar (2012) [13] proposed two-stage heuristic algorithms - first grouping the petrol stations, then assigning vehicles and planning the distribution paths. Popović *et al.* (2012) [4], Vidović *et al.* (2014) [8] and Hanczar (2012) [13] considered both the stock level of petrol stations and the vehicles' distribution routes simultaneously, and investigated the joint optimization problem of inventory control and vehicle routing. Zhang *et al.* (2019) [14] used a multi-scenario mixed integer programming model to describe the problem of the transportation process of oil products, considering the stochastic hub disruption and uncertain demands. Most studies mentioned above are based on the situation where the demands of petrol stations are deterministic. Although the sales volume of each station is stochastic, most researchers use average sales volume to replace random demands, so that the stochastic problem is simplified to a deterministic counterpart.

The refined oil secondary distribution problem based on stochastic demands can be considered as a variant of the Stochastic Transportation Problem (STP). The stochastic transportation problem was proposed by Williams (1963) [15] in 1963, and it was solved by a decomposition algorithm. Tsai *et al.* (2011) [16] described the transportation problem with uncertainty delivery capacities and costs. Thapalia *et al.* (2012) [17] studied the single-commodity network flow problem with stochastic edge capacities, compared optimal stochastic solution with the deterministic counterparts, and then provided a heuristic algorithm for the stochastic problem. Hinojosa *et al.* (2014) [18] presented the transportation problem with stochastic demands. They constructed a two-stage stochastic transportation model formulation of single commodity with stochastic demand and incorporated the existence of multiple commodities and capacities at sources. Singh *et al.* (2019) [19] formulated an uncertain three-dimensional transportation problem with multi-objective decision-making and proposed a solution methodology based on chance-constraint programming techniques. Huang *et al.* (2018) [20] developed a stochastic programming model to handle milk collection and delivery process with maximum route duration limitation, external cooling facility options, and travel time uncertainty.

However, there seems to be a lack of research considering the transportation problem of the refined oil secondary distribution with inventory capacity. Most of existing studies assume that demand rates or sales volume of petrol stations are deterministic. Moreover, existing researches on stochastic transportation problems rarely consider both limited supply quantity at sources and the limited storage capacity at destinations. According to the actual situation of ROSDP,

in the case that the sales volume (or sales rate) of petrol stations is stochastic, all constraints must be considered in planning the refined oil secondary distribution scheme based on the initiative distribution mode, including the supply limit of oil depots, the storage capacity limit of petrol stations, the compartment capacity of transport vehicles, and so on. Although these constraints make the problem more complicated, they can provide greater flexibility for final decision makers. To the best of our knowledge, so far there are no models or algorithms to tackle this problem.

In this paper, we focus on the refined oil secondary delivery problem based on the initiative distribution mode. We proposed a two-stage stochastic programming model for refined oil secondary distribution problem with stochastic demand and limited inventory capacity at each petrol station, limited supply quantity at each depot, and considering the capacity of vehicle's compartments. When the number of scenarios is limited, the stochastic programming model is reduced to an equivalent deterministic mixed integer programming model and the algorithm for solving the model is designed. Numerical experiments demonstrate the validity of the proposed model and algorithm. The computational effectiveness of the stochastic programming model is evaluated by using the value of the stochastic solution (VSS) and the expected value of perfect information (EVPI).

### III. PROBLEM FORMULATION

#### A. PROBLEM DESCRIPTION

The refined oil secondary distribution problem (ROSDP) can be described as follows:

A set of oil depots  $I$  supply a single type of refined oil to a set of petrol stations  $J$ . The supply quantity of each oil depot is limited. A set of vehicles  $K$  can be used to transport refined oil from oil depots to petrol stations. Each oil depot is equipped with a sufficient number of all types of vehicles, and each type of vehicle has a specific limited loading capacity and a fixed cost. The transportation cost for delivering refined oil is proportional to the travel distance and loading volume. In order to ensure safety and conveniently calculate the replenishment volume, the oil loaded in each compartment of one vehicle can only be unloaded to one petrol station.

Under the initiative distribution mode or VMI, by the end of each day, the petroleum company needs to make the next day's refined oil secondary distribution plan according to the supply quantity of oil depots, the existing stock of each petrol station and the number of available transportation vehicles. The refined oil secondary distribution plan includes the quantity of refined oil transported from each oil depot to each petrol station, the type of transport vehicles and drivers to perform each distribution task, the time of each vehicle arriving at petrol stations, and other information.

Because the daily sales volume of each petrol station is uncertain, if a petrol station is out of stock next day, it will result in penalty costs, assuming that the unit shortage cost

at each petrol station is known. If the replenishment volume transported to a petrol station exceeds its demand and the storage tank of the petrol station is full, then some special measures should be taken to properly handle the surplus refined oil, (for example, the surplus refined oil has to be transported back to the depot). Extra costs need to pay for dealing with the surplus refined oil, assuming that the unit cost for dealing with the surplus oil at each petrol station is known.

When the sales volume of each petrol station in the next day is uncertain, an important problem is how to determine the replenishment quantity of refined oil to be transported from each oil depot to each petrol station, and the type of vehicles to be used, so as to minimize the sum of the distribution costs and the expected recourse costs?

This problem can be formulated as a two-stage stochastic programming model. Since next day's sales volume cannot be predicted accurately when making the distribution plan, the problem has obviously two-stage characteristics. In the first stage, according to the limited supply quantity of oil depots and available transport vehicles, the quantity of the replenishment oil to be delivered from each depot to each petrol station and the type of vehicles to be used can be determined, so the distribution costs can be calculated. In the second stage, when the sales volume of each petrol station is known, the shortage quantity or surplus volume at each petrol station can be determined, so the total recourse costs can be calculated. Accordingly, the overall costs can be determined as well.

## B. ASSUMPTIONS

To simplify the problem, several assumptions are made as follows:

(1) Each transportation vehicle only serves for one petrol station at a time. In other words, all the refined oil transported by one vehicle can only be unloaded to one petrol station. It is not allowed to be separated unloaded to two or more petrol stations.

(2) The maximum capacity of each type of vehicle is less than the volume of the storage tank of any petrol station. In other words, if the storage tank is empty, the refined oil delivered by any vehicle can be completely unloaded to the tank without overflow.

(3) After the refined oil secondary distribution plan is made, all the transportation tasks are performed according to the plan, and the modification of the distribution plan is not considered.

(4) We assume that each vehicle only performs one delivery task every day and it will arrive at its petrol station in the given time window. So, we do not take into account either the overflow caused by a vehicle's earlier arrival at a petrol station, or the stockout caused by a vehicle's later arrival at a petrol station.

(5) Only the overflow or stockout caused by mismatching between the replenishment volume and next day's demand are considered.

(6) The sales volume or demand of each petrol station is a stochastic variable. There are three possible levels for the sales volume of each petrol station: high, medium and low.

(7) The sales volumes of all petrol stations comprise a stochastic vector. Since the sales volume of each petrol station is related to each other, we assume that the number of scenarios that affect the sales volume vector of all petrol stations is finite, and the occurrence probability of each scenario is known.

## C. NOTATIONS AND VARIABLES

The relevant notations and variables of the model for the problem are described as follows:

### 1) SETS

$I$ : set of oil depots

$J$ : set of petrol stations

$K$ : set of types of transportation vehicles

$S$ : set of scenarios affecting petrol stations' sales volume

### 2) INDICES

$i$ : index of oil depots

$j$ : index of petrol stations

$k$ : index of types of transportation vehicles

$s$ : index of scenarios

### 3) SYMBOLS

$a_i$ : supply quantity of oil depot  $i$

$R_j$ : maximum storage tank capacity of petrol station  $j$

$r_j$ : existing stock of petrol station  $j$

$c_j$ : unit shortage cost of petrol station  $j$

$d_j$ : unit surplus cost of petrol station  $j$

$h_k$ : fixed cost of vehicle type  $k$

$q_k$ : capacity of vehicle type  $k$

$g_{ij}$ : unit transportation cost from oil depot  $i$  to petrol station  $j$

### 4) STOCHASTIC PARAMETERS

$\xi$ : stochastic scenario which affects the sales volume of each petrol station

$S = \{\omega_1, \omega_2, \dots, \omega_{|S|}\}$ : set of all stochastic scenario realization, which contains a finite number of elements

$b_j(\xi)$ : actual sales volume of petrol station  $j$  when scenario  $\xi$  occurs

### 5) DECISION VARIABLES

$x_{ij}$ : quantity of refined oil to be transported from oil depot  $i$  to petrol station  $j$

$z_{ijk}$ : number of vehicles of type  $k$  used to deliver refined oil from depot  $i$  to petrol station  $j$

$y_j^+(\xi)$ : surplus quantity of refined oil in petrol station  $j$  under scenario  $\xi$

$y_j^-(\xi)$ : shortage quantity of refined oil in petrol station  $j$  under scenario  $\xi$

**D. A TWO-STAGE STOCHASTIC PROGRAMMING MODEL**

ROSDP can be formulated as a two-stage stochastic programming model as follows:

The first-stage model:

$$\min SP = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} h_k z_{ijk} + \sum_{i \in I} \sum_{j \in J} g_{ij} x_{ij} + E_{\xi}(Q(x, z, \xi)) \tag{1}$$

$$s.t. \sum_{j \in N} x_{ij} \leq a_i \quad i \in I \tag{2}$$

$$\sum_{k \in K} z_{ijk} q_k \geq x_{ij} \quad i \in I, j \in J \tag{3}$$

$$x_{ij} \geq 0 \quad i \in I, j \in J \tag{4}$$

$$z_{ijk} \geq 0 \quad \text{integer } i \in I, j \in J, k \in K \tag{5}$$

The second-stage recourse function:

$$Q(x, z, \xi) = \min \left( \sum_{j \in J} c_j y_j^-(\xi) + \sum_{j \in J} d_j y_j^+(\xi) \right) \tag{6}$$

$$s.t. y_j^-(\xi) \geq b_j(\xi) - r_j - \sum_{i \in I} x_{ij} \quad j \in J \tag{7}$$

$$y_j^+(\xi) \geq \sum_{i \in I} x_{ij} + r_j - b_j(\xi) - R_j \quad j \in J \tag{8}$$

$$y_j^-(\xi) \geq 0 \quad j \in J \tag{9}$$

$$y_j^+(\xi) \geq 0 \quad j \in J \tag{10}$$

In the first stage, with unknown values of scenarios, the model determines the quantity of refined oil to be transported from each depot to each petrol station and the number of vehicles to perform the distribution task.

The objective function (1) minimizes the sum of distribution costs and the expected recourse costs of the second stage, where the distribution costs include the fixed costs and the transportation costs. Constraints (2) represent the supply quantity constraints of each oil depot. The total quantity of refined oil transported from depot  $i$  to each petrol station cannot exceed its limited supply volume. Constraints (3) ensure that the total capacity of vehicles to perform the distribution task from depot  $i$  to petrol station  $j$  can hold the quantity of oil that should be delivered from depot  $i$  to station  $j$ . Constraints (4) and (5) define variable constraints. In the second stage, for a given scenario value of  $\xi$ , the sales volume of each petrol station is known. Based on the replenishment quantity obtained in the first stage, both the shortage quantity and the surplus quantity of each petrol station can be determined, so the second-stage total costs can be formulated by summing up the shortage costs and surplus costs.

In the second stage model, for a given scenario value  $\xi$ , the recourse function (6) is to minimize the total costs based on the first stage solution  $x$  and  $z$ . Constraints (7) represent the inequality that the shortage quantity of each petrol station  $j$  should satisfy. Constraints (8) denote the inequality that the

surplus quantity of each petrol station  $j$  should satisfy. Constraints (9) and (10) ensure that both the shortage quantity and the surplus quantity of each petrol station  $j$  are nonnegative variables.

**E. EQUIVALENT DETERMINISTIC MIXED INTEGER PROGRAMMING MODEL**

When the realizations of stochastic scenario vector  $\xi$  are finite, the two-stage stochastic programming model can be formulated as a deterministic mixed integer programming model.

Assume that  $S = \{\omega_1, \omega_2, \dots, \omega_{|S|}\}$  is the set of all realizations of stochastic scenario  $\xi$  and the probability of scenario realization  $\omega_s$  is  $p_s = P\{\xi = \omega_s\}$ , where

$$\sum_{s \in S} p_s = 1.$$

For simplification, let  $y_{js}^+, y_{js}^-$  respectively represent the surplus quantity and the shortage quantity of petrol station  $j$  in scenario realization  $\omega_s$ , and  $b_{js}$  represents the actual sales volume of petrol station  $j$  in scenario realization  $\omega_s$ . Then the two-stage stochastic programming model can be transformed into a deterministic mixed integer programming model as follows.

$$\begin{aligned} \min SP = & \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} h_k z_{ijk} + \sum_{i \in I} \sum_{j \in J} g_{ij} x_{ij} \\ & + \sum_{s \in S} p_s \left( \sum_{j \in J} c_j y_{js}^- + \sum_{j \in J} d_j y_{js}^+ \right) \end{aligned} \tag{11}$$

$$s.t. \sum_{j \in J} x_{ij} \leq a_i \quad i \in I \tag{12}$$

$$\sum_{k \in K} z_{ijk} q_k \geq x_{ij} \quad i \in I, j \in J \tag{13}$$

$$y_{js}^- \geq b_{js} - r_j - \sum_{i \in I} x_{ij} \quad j \in J, s \in S \tag{14}$$

$$y_{js}^+ \geq \sum_{i \in I} x_{ij} + r_j - b_{js} - R_j \quad j \in J, s \in S \tag{15}$$

$$x_{ij} \geq 0, \quad i \in I, j \in J \tag{16}$$

$$z_{ijk} \geq 0 \quad \text{integer } i \in I, j \in J, k \in K \tag{17}$$

$$y_{js}^- \geq 0 \quad j \in J, s \in S \tag{18}$$

$$y_{js}^+ \geq 0 \quad j \in J, s \in S \tag{19}$$

The objective function of the two-stage stochastic programming is to minimize the sum of distribution cost of the first stage and the expected storage cost and surplus cost of the second stage. It is in fact a joint optimization problem of two stages. The global optimal solution of the two-stage stochastic programming can be obtained by solving its equivalent deterministic mixed integer programming model. Its optimal value is denoted by SP.

*Example 1:* Consider an instance of ROSDP with  $|I| = 2$  depots and  $|J| = 4$  petrol stations,  $|K| = 2$  types of



TABLE 1. Information associated with oil depots and petrol stations.

$j$ $g_{ij}$	1	2	3	4	$a_i$
$i$					
1	1	2	2	4	60
2	4	2	3	1	90
$R_j$	20	20	30	30	-
$r_j$	5	5	10	10	-
$c_j$	100	100	100	100	-
$d_j$	20	20	20	20	-

TABLE 2. Fixed costs and capacities of each type of vehicle.

Vehicle type	Fixed cost	Capacity
1	200	10
2	300	20

TABLE 3. Demand of each petrol station in each scenario and the probability of each scenario.

$s$	$s_1$	$s_2$	$s_3$
$j$ $b_{js}$			
1	10	20	30
2	30	40	60
3	40	30	20
4	60	40	20
$p_s$	0.3	0.4	0.3

TABLE 4. Results associated with each depot and each petrol station  $j$ .

$j$	1	2	3	4
$i$ $x_{ij}$				
1	20(20*1)	30(20*1+10*1)	10(10*1)	0
2	0	20(20*1)	20(20*1)	50(20*2+10*1)
Total replenishment quantity	20	50	30	50

vehicles, and  $|S| = 3$  scenarios. Other parameters are given in Table 1–Table 3:

We solve the mixed integer programming model with CPLEX solver, and the result can be obtained in less than 1 second. The minimal cost is 3020 yuan. The optimal solutions are listed in Table 4.

The solution of the first stage is presented in Table 4. The numbers outside of the parentheses indicate the quantity of refined oil transported from oil depot  $i$  to petrol station  $j$ , and the numbers inside of the parentheses show the specific transportation plan, which consists of the types and number of vehicles to be used. For example, the total replenishment quantity delivered from depot 2 to petrol station 4 is 50 tons. The specific plan  $20*2 + 10*1$  means that two vehicles of capacity 20 tons and one vehicle of capacity 10 tons are used to perform the transportation task. According to the solution in Table 4, all 150 tons of refined oil supplied from two depots

are delivered to four petrol stations. The last row of Table 4 lists the total replenishment quantity to each petrol station. The total distribution cost of the first stage is 2630 yuan.

Based on the first stage transportation plan, the shortage quantity and the surplus quantity of each petrol station in the second stage with given scenarios are presented in Table 5 and Table 6 respectively.

TABLE 5. Shortage quantity of each petrol station in each scenario.

station shortage scenario	1	2	3	4
$s_1$	0	0	0	0
$s_2$	0	0	0	0
$s_3$	5	5	0	0

TABLE 6. Surplus quantity of each station in each scenario.

station shortage scenario	1	2	3	4
$s_1$	0	5	0	0
$s_2$	0	0	0	0
$s_3$	0	0	0	10

Table 5 and Table 6 illustrate that when scenario 1 occurs, there will be 5 tons of surplus oil in petrol station 2 and it results in 100 yuan surplus cost. When scenario 2 occurs, the supply and demand of four petrol stations are balanced, without causing surplus or shortage costs. When scenario 3 occurs, there will be 5 tons of shortage oil in both petrol station 1 and petrol station 2, and 10 tons of surplus oil in petrol station 4. So, the recourse cost corresponding to scenario 3 is 1200 yuan, which consists of 1000 yuan shortage cost and 200 yuan surplus cost. Based on the probabilities of three scenarios, the expected recourse cost of the second-stage is 390 yuan.

Therefore, the optimal value of the two-stage stochastic programming is 3020 yuan.

#### IV. MEASURING THE RELEVANCE OF USING A STOCHASTIC APPROACH

In this work, we use two measures to evaluate the two-stage stochastic programming: the value of stochastic solution (VSS) and the expected value of perfect information (EVPI).

##### A. THE VALUE OF STOCHASTIC SOLUTION (VSS)

When the random demand  $b_j(\xi)$  of each petrol station  $j \in J$  is replaced by its expected value  $\bar{b}_j$ , the two-stage stochastic programming model is transformed into a deterministic mixed integer programming model as follows.

$$\min EV = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} h_k z_{ijk} + \sum_{i \in I} \sum_{j \in J} g_{ij} x_{ij} + \sum_{j \in J} c_j y_j^- + \sum_{j \in J} d_j y_j^+ \quad (20)$$

$$s.t. \sum_{j \in J} x_{ij} \leq a_i \quad i \in I \quad (21)$$

$$\sum_{k \in K} z_{ijk} q_k \geq x_{ij} \quad i \in I, j \in J \quad (22)$$

$$y_j^- \geq \bar{b}_j - r_j - \sum_{i \in I} x_{ij} \quad j \in J \quad (23)$$

$$y_j^+ \geq \sum_{i \in I} x_{ij} + r_j - \bar{b}_j - R_j \quad j \in J \quad (24)$$

$$x_{ij} \geq 0, \quad i \in I, j \in J \quad (25)$$

$$z_{ijk} \geq 0 \quad \text{integer } i \in I, j \in J, k \in K \quad (26)$$

$$y_j^- \geq 0 \quad j \in J \quad (27)$$

$$y_j^+ \geq 0 \quad j \in J \quad (28)$$

The optimal solution  $\hat{x}$ ,  $\hat{z}$ ,  $\hat{y}^-$ ,  $\hat{y}^+$  of this deterministic mixed integer programming model gives a first-stage feasible solution  $\hat{x}$ ,  $\hat{z}$  of the original two-stage stochastic programming problem. Fixing the first-stage solution  $\hat{x}$ ,  $\hat{z}$ , and solving the recourse model equation (29-33) of the second stage for each scenario  $s \in S$ , we can find the optimal recourse solution  $y_{js}^{-*}$ ,  $y_{js}^{+*}$  ( $j \in J$ ) of the second stage in each scenario  $s \in S$ .

$$\min RC_s = \sum_{j \in J} c_j y_{js}^- + \sum_{j \in J} d_j y_{js}^+ \quad (29)$$

$$s.t. y_{js}^- \geq \bar{b}_j - r_j - \sum_{i \in I} \hat{x}_{ij} \quad j \in J \quad (30)$$

$$y_{js}^+ \geq \sum_{i \in I} \hat{x}_{ij} + r_j - \bar{b}_j - R_j \quad j \in J \quad (31)$$

$$y_{js}^- \geq 0 \quad j \in J \quad (32)$$

$$y_{js}^+ \geq 0 \quad j \in J \quad (33)$$

So, the minimal expected cost  $EEV$  of the expected value model can be found by equation (34).

$$EEV = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} h_k \hat{z}_{ijk} + \sum_{i \in I} \sum_{j \in J} g_{ij} \hat{x}_{ij} + \sum_{s \in S} p_s \left( \sum_{j \in J} c_j y_{js}^{-*} + \sum_{j \in J} d_j y_{js}^{+*} \right) \quad (34)$$

In model (20)-(28), the random demand  $b_j(\xi)$  of each petrol station  $j \in J$  is replaced by its expected value  $\bar{b}_j$ , so it has fewer number of variables and constraints than the equivalent mixed integer linear programming model (11)-(19), and it is much easier to be solved. But we can only find the first stage solution  $\hat{x}$ ,  $\hat{z}$  of the original two-stage stochastic programming problem by solving model (20)-(28). To further calculate the recourse cost of each scenario  $s \in S$  in the second stage, we must solve a recourse model (29-33) and find the solution  $y_{js}^{-*}$ ,  $y_{js}^{+*}$  ( $j \in J$ ) for each scenario  $s \in S$ .

Since the first stage solution  $\hat{x}$ ,  $\hat{z}$ , and the second stage solution  $y_{js}^{-*}$ ,  $y_{js}^{+*}$  ( $j \in J, s \in S$ ) are obtained by solving different models, the solution obtained by combining  $\hat{x}$ ,  $\hat{z}$  and  $y_{js}^{-*}$ ,  $y_{js}^{+*}$  ( $j \in J, s \in S$ ) is a local optimal solution of the

two-stage stochastic programming model. And the objective value  $EEV$  must be larger than that of SP.

The difference between  $EEV$  and the optimal value of the stochastic problem  $SP$  is defined as the value of the stochastic solution *i.e.*  $VSS = EEV - SP$ .

In order to eliminate the effect of value magnitude and obtain a better perception of  $VSS$ , we adopt the related percent of  $VSS$  defined by Hinojosa *et al.* (2014) [18].

$$VSS_R = 100 \times \frac{1}{SP} (EEV - SP) \quad (35)$$

Both  $VSS$  and  $VSS_R$  can be used to verify the efficient of two-stage stochastic programming, the larger  $VSS$  or  $VSS_R$  is, the more superior the stochastic programming method is.

## B. THE EXPECTED VALUE OF PERFECT INFORMATION (EVPI)

$EVPI$  represents the value of perfect information, which is the value of stochastic solution  $SP$  minus the value of wait-and-see solution  $WS$ . To obtain  $EVPI$ , we should calculate a set of optimal values  $WS_s$  of the problem with each scenario  $s \in S$  by solving the mixed integer programming as follows:

$$\min WS_s = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} h_k z_{ijks} + \sum_{i \in I} \sum_{j \in J} g_{ij} x_{ijs} + \sum_{j \in J} c_j y_{js}^- + \sum_{j \in J} d_j y_{js}^+ \quad (36)$$

$$s.t. \sum_{j \in J} x_{ijs} \leq a_i \quad i \in I \quad (37)$$

$$\sum_{k \in K} z_{ijks} q_k \geq x_{ijs} \quad i \in I, j \in J \quad (38)$$

$$y_{js}^- \geq b_{js} - r_j - \sum_{i \in I} x_{ijs} \quad j \in J \quad (39)$$

$$y_{js}^+ \geq \sum_{i \in I} x_{ijs} + r_j - b_{js} - R_j \quad j \in J \quad (40)$$

$$x_{ijs} \geq 0, \quad i \in I, j \in J \quad (41)$$

$$z_{ijks} \geq 0 \quad \text{integer } i \in I, j \in J, k \in K \quad (42)$$

$$y_{js}^- \geq 0 \quad j \in J \quad (43)$$

$$y_{js}^+ \geq 0 \quad j \in J \quad (44)$$

The model (36)-(44) can be regarded as a special case of the deterministic programming model (11)-(19), where there is only one scenario  $s$  in the set of all scenarios  $S$ . The optimal solution of model (36)-(44) means that if we can accurately predict the actual demand of scenario  $s$  to be occurred in future, we would find the optimal solutions  $x_{ijs}$ ,  $z_{ijks}$ ,  $y_{js}^-$ ,  $y_{js}^+$  based on the determined information of scenario  $s$ , and obtain the minimal cost  $WS_s$  corresponding to this optimal solution.

After solving each model of (36)-(44) under every scenario, we gather them into the wait-and-see solution  $WS$ :

$$WS = \sum_{s \in S} p_s \cdot WS_s$$

Obviously,  $WS$  is smaller than  $SP$ . The difference between  $SP$  and  $WS$  is defined by  $EVPI$ , which can be used to measure the value of information.

$EVPI$  is defined as follows:

$$EVPI = SP - WS$$

Similarly, we define a related percent of  $EVPI_R$  as follows:

$$EVPI_R = 100 \times \frac{1}{SP}(SP - WS)$$

In this work, we use  $VSS_R$  and  $EVPI_R$  to measure the relevance of the stochastic programming approach.

### C. CALCULATE $VSS_R$ AND $EVPI_R$ OF EXAMPLE 1

For the problem in Example 1, if we replace the random demand  $b_j(\xi)$  of each petrol station by its expected demand value  $\bar{b}_j$ , the first stage solution can be found by solving the mixed integer programming model based on the expected value of demands. The detailed information of the solution is provided in Table 7, and the total cost of the first stage is 1861 yuan. In Table 7, the value outside of the parentheses describes the replenishment quantity from oil depot  $i$  to petrol station  $j$ , whereas the value inside of the parentheses describes the capacity of vehicle type times the number of vehicles used to deliver the replenishment quantity. For example, the information in the second row and the forth column is  $30(20*1 + 10*1)$ , which means that 30 tons of refined oil should be transported from depot 2 to station 4, and the delivery task is performed by one vehicle with 20-ton capacity and one vehicle with 10-ton capacity.

Based on the first stage solution in Table 7, both the shortage and the surplus quantities of each station in each scenario can be obtained by solving the second-stage recourse model. The shortage quantity of each station in each scenario is listed in Table 8.

When the stochastic demand of each petrol station is replaced by its expected demand, the total replenishment quantity of refined oil transported from each depot to each petrol station will decrease, so the total distribution costs

TABLE 7. The first stage solution based on expected value of demands.

$x_{ij}$	1	2	3	4
1	15(20*1)	18(20*1)	20(20*1)	0
2	0	20(20*1)	0	30(20*1+10*1)
Total replenishment quantity	15	38	20	30

TABLE 8. Shortage quantity of each station in each scenario based on the expected solutions.

station shortage scenario	1	2	3	4
$s_1$	0	0	10	20
$s_2$	0	0	0	0
$s_3$	10	17	0	0

would decrease accordingly. However, when the actual sales volume increases, the shortage cannot be avoided. When scenario 1 occurs, the shortage quantities at petrol station 3 and petrol station 4 are 10 tons and 20 tons respectively, which result in 3000 yuan shortage cost. When scenario 3 occurs, the shortage quantities of petrol station 1 and petrol station 2 will be 10 tons and 17 tons respectively, which result in 2700 yuan shortage cost. There is no shortage cost under scenario 2.

Furthermore, since the total replenishment quantity transported to each petrol station is relatively small, no surplus occurs under each scenario.

Summing up the shortage cost and the surplus cost, the total expected cost based on the expected value model can be obtained.

$$EEV = 1861 + 3000 \times 0.3 + 0 \times 0.4 + 2700 \times 0.3 = 3571$$

$$VSS_R = 100 \times \frac{EEV - SP}{SP} = 100 \times \frac{3571 - 3020}{3020} = 18.25$$

For the problem in Example 1, the perfect information solutions of three scenarios can be found by solving three mixed integer programming models, and the optimal values of three scenarios are 2165, 1855 and 1965 yuan respectively. Therefore, the wait-and-see solution can be calculated as follows:

$$WS = 2165 \times 0.3 + 1855 \times 0.4 + 1965 \times 0.3 = 1981$$

$$EVPI_R = 100 \times \frac{SP - WS}{SP} = 100 \times \frac{3020 - 1981}{3020} = 34.4$$

As analyzed above, for the refined oil secondary distribution problem with random demand, the expected total cost corresponding to the two-stage stochastic programming solution is lower than the expected total cost corresponding to the expected value solution. The larger fluctuation the demand has, the more obviously superior the stochastic solution is. On the other hand, the expected cost of wait-and-see solution with perfect information is lower than that of the stochastic solution, so the value of perfect information is great.

### V. COMPUTATIONAL EXPERIMENTS

In this section, we conduct numerical simulations on small, medium, and large size examples respectively, and report the computational results to evaluate the performance of the proposed two-stage stochastic programming model. Section V-A describes the generation of the experimental instances. Section V-B reports and analyzes the computational results. Section V-C reduces the problem scale by merging similar scenarios and assesses the performance of the merged model. Section V-D reports the sensitivity analysis of inventory capacity and the demand.

All computational experiments were performed on a PC with an Intel(R) Core(TM) i7 processor with 2.80 GHz and 8 GB of RAM. Programs with default parameters were implemented using ILOG CPLEX Studio Academic 12.8.0.



### A. INSTANCE GENERATION

To generate the instances, we consider four factors associated with the dimension of the test instances: the numbers of (i) depots, (ii) petrol stations, (iii) vehicle types and (iv) scenarios. For each factor, we consider one or several values which correspond to the different dimensions of test instances.

In particular, the number of depots  $|I|$  is chosen from the set  $\{2, 4, 6\}$ , the number of petrol stations  $|J|$  is chosen from  $\{20, 50, 100\}$ , the number of vehicle types  $|K|$  is 3, and the number of scenarios  $|S|$  is chosen from  $\{4, 8, 12, 20\}$ .

Once the dimensions are set, one instance was generated as follows:

For each depot  $i \in I$ , the supply  $a_i$  was drawn from a uniform distribution  $U[u_0 - 40, u_0 + 40]$ , where

$$u_0 = 40 \times |J| / |I|.$$

For each  $j \in J$ , the capacity  $R_j$  was randomly drawn from  $\{20, 30, 40\}$ . The initial inventory  $r_j$  was randomly drawn from  $\{5, 10, 15\}$ . The unit shortage penalty cost  $c_j$  was drawn from  $\{90, 100, 110\}$ . The unit surplus handling cost  $d_j$  was drawn from  $\{10, 20, 30\}$ . Demand  $b_j$  was generated from three possible uniform distributions, which corresponds to the low demand level, medium demand level, and high demand level respectively:

- (i)  $U[10, 30]$ , (ii)  $U[30, 40]$  and (iii)  $U[40, 60]$ .

For each  $i \in I, j \in J$ , the unit transportation cost  $g_{ij}$  was drawn from a uniform distribution  $U[1, 4]$ .

For each vehicle type  $k \in K$ , the capacity  $q_k$  was set as follows:  $q_1 = 10, q_2 = 15, q_3 = 20$ ; the fixed cost  $h_k$  was set as follows:  $h_1 = 200, h_2 = 250, h_3 = 300$ .

Our data sets consist of four different types of scenarios:

- (S1) all the demand levels are drawn from interval (i) representing low demand level;
- (S2) all the demand levels are drawn from interval (ii) representing medium demand level;
- (S3) all the demand levels are drawn from interval (iii) denoting high demand level;
- (S4) For each petrol station, a demand level is chosen randomly from three different intervals (i) (ii) (iii) with the same probability.

The number of each type of scenarios is summarized in Table 9.

$$p_s = P\{\xi = s\} = \frac{1}{|S|}$$

For each combination of parameters  $|I|, |J|, |K|$  and  $|S|$ , five instances were generated. In total, 180 instances were generated.

**TABLE 9. Number of each type of scenarios.**

$ S $	$ S_1 $	$ S_2 $	$ S_3 $	$ S_4 $
4	1	1	1	1
8	1	1	1	5
12	2	2	2	6
20	4	4	4	8

### B. COMPUTATIONAL RESULTS AND ANALYSIS

For each instance, we have used the solver ILOG CPLEX Studio Academic 12.8.0 to solve the equivalent deterministic mixed integer programming model. The time limit is set to be 2000s. If an optimal solution was found before the time limit was reached, we recorded the CPU time (in seconds) to obtain the optimal solution. When the time limit was reached, we saved the relative gap between the objective value of the best integer solution and the best-known lower bound of the optimal solution value.

Table 10 and Table 11 report the results of the instances associated with each combination of parameters  $|I|, |J|, |S|$ . In Table 10 and Table 11, the rows were grouped into several blocks, one for each parameter value of  $|I|$ . Within each block, the rows were further grouped into sub-blocks by the parameter value of  $|J|$ . Each subblock corresponds to a parameter value of  $|J|$ . Each row in a subblock corresponds to an instance with a particular combination of  $|I|, |J|$  and  $|S|$ .

Table 10 summarizes the results for the stochastic programming of refined oil secondary distribution with inventory capacity constraints. The average of percent gaps listed in

**TABLE 10. Average gap and CPU time (seconds) for the stochastic programming of refined oil secondary distribution with inventory capacity constraints.**

$ I $	$ J $	$ S $	Gap(%)	CPU	Opt
2	20	4	0	0.2552	5
		8	0	0.2569	5
		12	0	0.2614	5
		20	0	0.2782	5
	50	4	0	0.2702	5
		8	0	0.4607	5
		12	0	0.4825	5
		20	0	0.6652	5
	100	4	0	0.5543	5
		8	0	0.6302	5
		12	0	1.0003	5
		20	0	1.3124	5
4	20	4	0	0.3540	5
		8	0	0.3739	5
		12	0	0.4201	5
		20	0	0.4936	5
	50	4	0	0.4418	5
		8	0	0.6212	5
		12	0	1.0883	5
		20	0	1.2676	5
	100	4	0	1.2456	5
		8	0	1.4337	5
		12	0	1.7249	5
		20	0	2.1428	5
6	20	4	0	0.3859	5
		8	0	1.3304	5
		12	0	1.4493	5
		20	0	1.6492	5
	50	4	0	0.6153	5
		8	0	1.9647	5
		12	0	3.1148	5
		20	0	3.7210	5
	100	4	0	3.3789	5
		8	0	4.6599	5
		12	0	4.8406	5
		20	0	6.7231	5

**TABLE 11. Relative percentage  $EVPI_R$ ,  $VSS_R$  and the corresponding CPU time (in seconds).**

$ I $	$ J $	$ S $	$EVPI_R$ (%)	CPU-1	$VSS_R$ (%)	CPU-2
2	20	4	32.15	0.5954	37.69	0.1027
		8	39.17	0.4067	33.14	0.1477
		12	38.84	0.6667	30.26	0.2153
		20	38.44	1.0508	32.07	0.3245
	50	4	35.67	0.6250	29.11	0.2822
		8	39.13	1.5808	32.88	0.4548
		12	40.69	1.1682	33.69	0.3369
		20	37.91	2.9764	30.49	0.5375
	100	4	36.18	1.0432	30.40	0.2960
		8	39.78	1.8644	31.77	0.4112
		12	39.58	2.3702	30.61	0.5841
		20	40.38	4.9297	29.97	0.7738
4	20	4	31.35	0.3032	34.97	0.1506
		8	38.38	1.2905	29.93	0.3172
		12	43.13	1.5788	33.65	0.2216
		20	40.13	1.3108	30.87	0.3860
	50	4	37.18	0.6057	34.22	0.2913
		8	37.82	1.1284	35.19	0.3471
		12	39.89	2.0595	34.82	0.3825
		20	39.96	5.6106	31.32	0.5653
	100	4	38.05	1.2401	34.59	0.4879
		8	38.98	2.8925	33.13	0.5865
		12	40.33	4.2657	31.71	0.6703
		20	41.44	6.9802	29.86	0.8995
6	20	4	35.89	0.3331	35.43	0.1387
		8	40.19	0.6661	38.31	0.2247
		12	40.13	7.9282	34.67	0.2863
		20	42.88	3.2842	31.85	0.3465
	50	4	38.68	1.0223	35.09	0.2926
		8	39.60	1.6367	39.28	0.4518
		12	39.23	2.5457	33.31	0.4932
		20	40.63	4.4795	34.46	0.6371
	100	4	38.71	1.9441	31.93	0.5638
		8	38.47	6.9933	34.63	0.6991
		12	37.72	13.8856	32.07	0.8713
		20	39.56	14.8160	31.66	1.1337

the columns headed by ‘Gap’ is defined by  $100 \times (\text{best-lowbound})/\text{lowbound}$ , where ‘best’ is the objective value of the best integer solution and ‘lowbound’ is the best lower bound found by the solver.

The column headed by ‘CPU’ lists the average CPU times (seconds) required by the solver to obtain the optimal/best solution. The column headed by ‘Opt’ lists the number of instances whose optimal solutions were found within time limit.

From the results in Table 10, we found that the optimal solutions of all instances were obtained within 6.7231 seconds. When the parameter value of  $|I|$ ,  $|J|$  or  $|S|$  increases, the CPU time required for obtaining the optimal solution increases too. For the largest instance with 6 depots, 100 petrol stations and 20 scenarios, none of experimental CPU time for finding the optimal solution exceeds 7 seconds. So the solver can be used to find the optimal solutions of real problems.

Table 11 provides the relative percentage  $EVPI_R$  and  $VSS_R$ , and the CPU time (in seconds) required to obtain each value. An entry in the column headed by ‘ $EVPI_R$ ’ (or ‘ $VSS_R$ ’) gives the average relative value of the information (or the stochastic solution) as introduced in Section IV. An entry in the column

headed by ‘CPU-1’ (or ‘CPU-2’) corresponds to the average CPU time (seconds) required by the solver to obtain the value of ‘ $EVPI_R$ ’ (or ‘ $VSS_R$ ’). In Table 11, the minimal value of  $EVPI_R$  is larger than 31%, and the minimal value of  $VSS_R$  is larger than 29%. In other words, both  $EVPI_R$  and  $VSS_R$  are quite significant across all the instances. Table 11 verifies the necessity and effectiveness of using a stochastic approach for the problem we aim to address.

Since the instances in Table 10 and Table 11 are randomly generated, it is difficult to find the correlation between ‘ $EVPI_R$ ’ (or ‘ $VSS_R$ ’) and the parameter values in  $|I|$ ,  $|J|$  and  $|S|$ .

From each subblock with same parameter values  $|I|$  and  $|J|$  in Table 10, we found that the CPU time required for obtaining the optimal solution increases with the number of scenarios  $|S|$ .

In practice, since the sales volume of each petrol station might take a lot of values (it even can be regarded as a continuous random variable), there is a large number of scenarios related to the sales volumes of all petrol stations. In this case, it will take much longer CPU time to solve the equivalent mixed integer programming model of the two-stage stochastic programming problem. How to find approximately optimal solutions in short time is the key issue for solving large scale practical problems.

### C. REDUCING THE PROBLEM SCALE BY MERGING SIMILAR SCENARIOS

For a given practical problem, the parameters  $|I|$ ,  $|J|$  and  $|K|$  are usually determined in prior. The CPU time for solving the stochastic programming model increases with the number of scenarios  $|S|$  increases.

When the number of oil depots and petrol stations is large, the size of SP model in one scenario is already large. With the number of scenarios increases, the size of SP model will increase rapidly. Since the actual demand of each petrol station might be a discrete random variable with a large number of possible values, or a continuous random variable, the number of scenarios in the SP model might be astronomical. If we solve the SP model with all scenarios by solver, the running time will be out of acceptable range. For examples with 6 depots, 100 petrol stations, and 100 scenarios, the average CPU time is more than 10 minutes. By merging a large number of similar scenarios into fewer ones, we can effectively reduce the size of SP model and rapid the running speed.

On the other hand, there will be errors when replacing the original problem by the one after merging scenarios. Next, we will study the gap of solutions obtained from the model with merged scenarios. By fixing the number of oil depots  $|I|$ , the number of petrol stations  $|J|$ , and the number of vehicle types  $|K|$ , we take a practical problem with 20 scenarios as an example and analyze the gap of solutions obtained from the model with merged scenarios.

Assume that 20 scenarios are divided into 4 types according to the classification rules in Section V-A, and each type

TABLE 12. Comparative analysis of the results obtained by two models before and after merging scenarios.

I	J	before scenarios merging			after scenarios merging			Gap
		S (1)	CPU(1)	Cost(1)	S (2)	CPU(2)	Cost(2)	
2	20	20	0.3152	17020	8	0.1964	17113	0.55%
	30	20	0.4254	24064	8	0.2483	24214	0.63%
	40	20	0.5426	33277	8	0.3171	33539	0.79%
	50	20	0.6582	40373	8	0.3645	40715	0.85%
	60	20	0.7804	49457	8	0.4827	49827	0.75%
	70	20	0.8916	57436	8	0.4487	57691	0.44%
	80	20	1.0014	63828	8	0.5076	64112	0.44%
	90	20	1.1849	73376	8	0.5734	73965	0.80%
	100	20	1.6666	81119	8	0.6442	81533	0.51%
4	20	20	0.6323	15591	8	0.3700	15629	0.24%
	30	20	0.7665	23682	8	0.4358	23808	0.53%
	40	20	0.9395	30581	8	0.4388	30731	0.49%
	50	20	1.1220	39615	8	0.4866	39743	0.32%
	60	20	2.4714	48022	8	0.6422	48288	0.55%
	70	20	1.4234	54604	8	0.6482	54893	0.53%
	80	20	1.8161	63640	8	0.9015	63854	0.34%
	90	20	2.3464	71610	8	1.0172	71981	0.52%
	100	20	6.4099	79121	8	1.1608	79537	0.53%
6	20	20	0.7709	15048	8	0.4557	15091	0.29%
	30	20	0.8457	23661	8	0.4627	23719	0.25%
	40	20	1.4002	30840	8	0.5944	30972	0.43%
	50	20	3.5485	39012	8	0.7150	39274	0.67%
	60	20	4.3294	47883	8	2.5910	48208	0.68%
	70	20	2.1312	54600	8	1.1289	54944	0.63%
	80	20	5.0823	63593	8	2.9142	64151	0.88%
	90	20	5.8444	72145	8	2.2160	72747	0.83%
	100	20	8.2379	77816	8	4.5368	78263	0.57%

contains 5 scenarios. As shown in Table 9, the value range of each petrol station’s sale volume is identical in the first three types of scenarios. For example, in the first type of scenarios S1, the sales volume of each petrol station is drawn from interval [10, 30], which corresponds to the low demand level. Therefore, five scenarios of the same type can be merged into one scenario. After merging the five scenarios of the same type, the sales volume of each petrol station is replaced by the average of sales volumes under five scenarios. Obviously, the value of average sales volume is taken from interval of [10, 30], which is the same interval as that of sales volumes taken from in each scenario of S1 type. The occurrence probability of the merged scenario is taken as the sum of the occurrence probability of each scenario before merging.

The last type of scenarios S4 is different from the first three scenarios. In S4 type of scenarios, the sales volume of the same petrol station in different scenarios may take value from different demand intervals. If the scenarios are merged, the sales volume of the same petrol station in different scenarios might be offset each other, so the average sales volume under the merged scenarios might be significantly different from the sale volume of the same petrol station in any scenarios before merging.

Therefore, the scenarios in type S4 can be merged into multiple scenarios according to the demand interval (high, medium and low) that each petrol station’s sale volume is drawn from. Given two or more scenarios, if the sales volume of every petrol station in all scenarios falls in the same interval, then the scenarios can be merged into one scenario. Otherwise, they cannot be merged.

The *K*-means clustering method can be directly used for merging scenarios. If multiple scenarios are merged, the average sales volume of each petrol station in all scenarios is taken as the sales volume of this petrol station in the merged scenario, and the probability of the merged scenario is the sum of the probabilities of each scenario before merging.

To analyze the efficiency and effectiveness of merging scenarios, we first solve the two-stage stochastic programming model corresponding to merged scenarios by CPLEX solver, and obtain the transportation scheme of the first stage; then, based on the transportation scheme in the first stage, the recourse cost of the second stage under each scenario before merging is calculated, and the total expected cost corresponding to the problem after merging scenarios is further calculated.

For each instance, the CPU time for solving the stochastic programming models before and after merging scenarios are recorded and compared. The expected cost of the model with merged scenarios is compared with the optimal value of the model before merging scenarios. We also analyze the change of error rate and CPU time after merging scenarios. The detailed results are listed in Table 12.

In Table 12, the columns headed by “|S| (1)”, “CPU (1)”, “cost (1)” represent the number of scenarios, CPU time and the total expected costs respectively related to the model before merging scenarios; the columns headed by “|S| (2)”, “CPU (2)”, “cost (2)” represent the number of scenarios, CPU time and the total expected costs respectively related to the model after merging scenarios. The last column headed by “Gap” represents the total costs increasing rate

corresponding to the solution obtained by the model after merging scenarios, the Gap is calculated by the equation as follows:

$$Gap = (cost (2) - cost (1)) / cost (1)$$

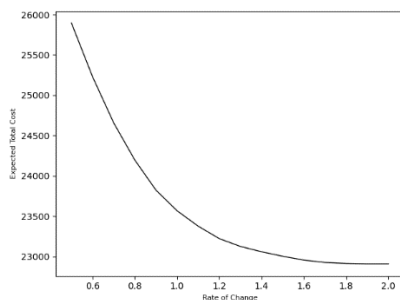
It can be seen from Table 12 that after merging similar scenarios, the scale of the problem becomes smaller, and the CPU time for solving the problem is reduced accordingly. After merging 20 scenarios into 8 scenarios, the CPU time is reduced to about 50% of the original. Compared with the optimal solution obtained by solving the model with 20 scenarios, the error rate of the model with 8 merged scenarios is less than 1%. So, whenever the practical problem has a large number of scenarios, we can first merge some similar scenarios using clustering methods, and then solve the smaller scale problem with merged scenarios to obtain an approximately optimal solution of the original problem in much shorter CPU time.

**D. INFLUENCE OF INVENTORY CAPACITY AND DEMAND CHANGES**

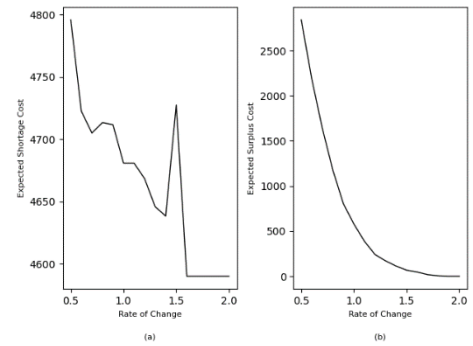
We perform additional testing on the sensitivity of inventory capacity and the demand to the optimal solution. We generate an instance with 2 depots, 30 petrol stations, 3 types of vehicles and 12 scenarios by the method in V-A.

Firstly, we vary the inventory capacity of each petrol station by the same ratios  $\lambda$ , and calculate the optimal solution under each parameter  $\lambda \in [0.5, 2]$ . Figure 1 depicts the relationship between the total cost of optimal solution and the value of parameter  $\lambda$ . The results clearly show that increasing the inventory capacity results in the total cost reduction. Decreasing the inventory capacity by 50% leads to a total cost increase of 9.8%. Vice versa, increasing inventory capacity by 90% results in the minimal total cost. If the inventory capacity increases further, the total cost can't decrease any more. We also conduct further investigation to study the influence of parameter  $\lambda$  on the expected shortage and surplus cost. The results are depicted in Figure 2(a) and 2 (b) respectively. It is easily seen that a larger inventory capacity result in lower shortage cost and surplus cost because it enables more refined oil to be stored and the more uncertainty demands will be satisfied.

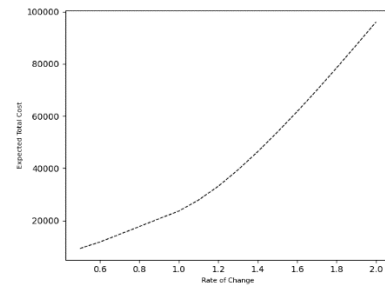
Furthermore, we study the influence of increasing or decreasing the demand of each petrol station in each scenario



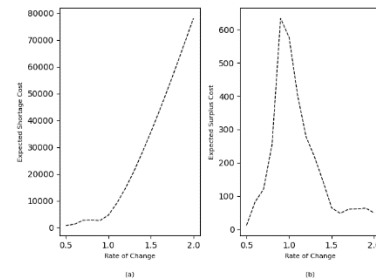
**FIGURE 1.** The effect of the total cost under different parameter  $\lambda$ .



**FIGURE 2.** The effect of (a) the expected shortage cost and (b) the expected surplus cost under different parameter  $\lambda$ .



**FIGURE 3.** The effect of the total cost under different parameter  $\mu$ .



**FIGURE 4.** The effect of (a) the expected shortage cost and (b) the expected surplus cost under different parameter  $\mu$ .

by the same ratios  $\mu$ , and calculate the optimal solution under each parameter  $\mu \in [0.5, 2]$ . Figure 3 depicts the relationship between the minimal cost of optimal solution and the value of parameter  $\mu$ . The results show that increasing demands lead to the total cost increases. Decreasing demands by 50% results in a total cost reduction of 61%, whereas 50% higher demands increase the total cost by more than 128%. We also analyze how different values of parameter  $\mu$  affect the expected shortage and surplus. The results are shown in Figure 4(a) and 4(b) respectively. We observe that the higher demands results in the higher shortage cost due to the limited supply. Notice that the surplus cost increases sharply to a peak of 634 and decreases rapidly to 50, since the higher demands narrow the gap between the total supply and demand.

Hence, the tests clearly show that a larger inventory capacity results in lower total cost because it reduces both the shortage cost and the surplus cost. In practice, the inventory



capacity of each petrol station should be determined based on its demand and delivery amounts. If the conditions permit, a greater inventory capacity of each petrol station should be taken into consideration.

## VI. CONCLUSION

In this paper, we investigate the refined oil secondary distribution problem faced by oil companies which adopt initiative delivery mode. The uncertain demand of petrol stations and limited inventory capacity of petrol stations are considered. We formulate the refined oil secondary distribution problem with stochastic demand and limited capacity as a two-stage stochastic programming model. In the first stage, a decision is made about the replenishment quantity of refined oil transported from each depot to each petrol station, and the number of each type of vehicles used to perform each transportation task so as to minimize the sum of distribution costs and the expected recourse costs. In the second stage, a decision is made to minimize the recourse costs including the shortage penalty costs and the surplus handling costs at each petrol station in a given scenario. By analyzing the simulation results of instances with different scales, we found that if the sales volume at each petrol station is stochastic, the expected total cost of the inventory transportation scheme obtained by the two-stage stochastic programming model is about 30%-40% lower than that of the deterministic programming model using the average sales volume. On the other hand, the number of oil depots, the number of petrol stations and the number of vehicle types in practical instances will not exceed the maximum scales of the simulation examples in this work. So when the number of scenarios is no more than 20, we can solve the equivalent deterministic mixed integer problem corresponding to the two-stage stochastic programming by CPLEX solver and obtain global optimal solutions in a short time.

When the two-stage stochastic programming model is transformed into the deterministic mixed integer programming model, the number of scenarios is the main factor affecting the scale of the problem. Since it needs longer CPU time for solving the larger scale equivalent mixed integer programming model, for practical problems with larger number of scenarios, we can merge some similar scenarios so as to obtain approximately optimal solutions by the solver in a shorter CPU time.

The results of sensitivity analysis indicate that the total cost decreases with corresponding increase in inventory capacity within a certain range. Therefore, the petrol stations should use the relatively larger storage tank if it is possible.

In order to calculate the replenishment quantity conveniently, the oil delivered by one vehicle is only allowed to be unloaded to one petrol station. It is not allowed to be separately unloaded to multiple petrol stations. This assumption is applicable to the case that a petrol station only sells a single type of refined oil product and the storage tank capacity is larger than the vehicle's compartment capacity. In practice, when a petrol station sells several types of oil products,

and the storage tank capacity of each type of refined oil is smaller than the total capacity of a vehicle, the oil company can use the vehicles with several independent compartments for distribution. At this time, different compartments of one vehicle can be loaded with different types of oil products, and the oil loaded in different compartments can be unloaded to one or more oil storage tanks in one or more petrol stations.

The refined oil secondary transportation problem with multiple compartments vehicles is the generalization of the problem in this work. In the future research, we will study this more general problem.

The CPLEX solver can be used to solve the equivalent mixed integer programming model for the refined oil secondary transportation problem with single type of oil in this work. But when the number of scenarios increases, the CPU time consumed for solving the model will increase, and the solver cannot find exact optimal solutions in a short time. Therefore, it is an important research direction to design exact algorithms or heuristic algorithms based on the characteristics of practical problems.

## ACKNOWLEDGMENT

The authors sincerely thank Dr. Ruisheng Wang in Harvard University for editing and proofreading the manuscript. The authors sincerely thank Dr. Francisco Saldanha-da-Gama and Dr. Jvliang Zhang for useful comments and suggestions.

## REFERENCES

- [1] G. Marquès, C. Thierry, J. Lamothe, and D. Gourc, "A review of vendor managed inventory (VMI): From concept to processes," *Prod. Plan. Control*, vol. 21, no. 6, pp. 547–561, Sep. 2010, doi: [10.1080/09537287.2010.488937](https://doi.org/10.1080/09537287.2010.488937).
- [2] P. Danese, "Towards a contingency theory of collaborative planning initiatives in supply networks," *Int. J. Prod. Res.*, vol. 49, no. 4, pp. 1081–1103, Feb. 2011, doi: [10.1080/00207540903555510](https://doi.org/10.1080/00207540903555510).
- [3] C. Archetti and M. G. Speranza, "The inventory routing problem: The value of integration," *Int. Trans. Oper. Res.*, vol. 23, no. 3, pp. 393–407, 2016, doi: [10.1111/itor.12226](https://doi.org/10.1111/itor.12226).
- [4] D. Popović, M. Vidović, and G. Radivojević, "Variable neighborhood search heuristic for the inventory routing problem in fuel delivery," *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13390–13398, Dec. 2012, doi: [10.1016/j.eswa.2012.05.064](https://doi.org/10.1016/j.eswa.2012.05.064).
- [5] F. Cornillier, F. Boctor, and J. Renaud, "Heuristics for the multi-depot petrol station replenishment problem with time windows," *Eur. J. Oper. Res.*, vol. 220, no. 2, pp. 361–369, Jul. 2012, doi: [10.1016/j.ejor.2012.02.007](https://doi.org/10.1016/j.ejor.2012.02.007).
- [6] P. Carotenuto, S. Giordani, S. Massari, and F. Vagaggini, "Periodic capacitated vehicle routing for retail distribution of fuel oils," *Transp. Res. Procedia*, vol. 10, pp. 735–744, Jan. 2015, doi: [10.1016/j.trpro.2015.09.027](https://doi.org/10.1016/j.trpro.2015.09.027).
- [7] N. Charusakwong and M. Lohatepanont, "Optimization approach for multi-period fuel replenishment," *Eng. J.*, vol. 20, no. 5, pp. 239–261, Nov. 2016.
- [8] M. Vidović, D. Popović, and B. Ratković, "Mixed integer and heuristics model for the inventory routing problem in fuel delivery," *Int. J. Prod. Econ.*, vol. 147, pp. 593–604, Jan. 2014, doi: [10.1016/j.ijpe.2013.04.034](https://doi.org/10.1016/j.ijpe.2013.04.034).
- [9] L.-F. Zhang, W.-L. Yi, and X.-L. Liu, "Two-stage optimization algorithm for large scale secondary petroleum product delivery planning," *Syst. Eng.-Theory Pract.*, vol. 36, no. 11, pp. 2951–2963, Nov. 2016.
- [10] Z.-P. Li, W.-F. Zhou, Y.-W. Zhang, G. Yang, and R. Liu, "Research on vehicle routing problem of refined oil secondary distribution considering unloading sequence constraints," *Decis. Control*, to be published, doi: [10.13195/j.kzyjc.2018.1756](https://doi.org/10.13195/j.kzyjc.2018.1756).
- [11] A. Benantar, R. Ouafi, and J. Boukachour, "A petrol station replenishment problem: New variant and formulation," *Logistics Res.*, vol. 9, no. 1, p. 6, Apr. 2016, doi: [10.1007/s12159-016-0133-z](https://doi.org/10.1007/s12159-016-0133-z).



- [12] P. Avella, M. Boccia, and A. Sforza, "Solving a fuel delivery problem by heuristic and exact approaches," *Eur. J. Oper. Res.*, vol. 152, no. 1, pp. 170–179, Jan. 2004, doi: [10.1016/S0377-2217\(02\)00676-8](https://doi.org/10.1016/S0377-2217(02)00676-8).
- [13] P. Hanczar, "A fuel distribution problem—application of new multi-item inventory routing formulation," *Procedia-Social Behav. Sci.*, vol. 54, pp. 726–735, Oct. 2012, doi: [10.1016/j.sbspro.2012.09.790](https://doi.org/10.1016/j.sbspro.2012.09.790).
- [14] W. Zhang, Z. Li, Q. Liao, H. Zhang, B. Wang, S. Huang, N. Xu, and Y. Liang, "A stochastic linear programming method for the reliable oil products supply chain system with hub disruption," *IEEE Access*, vol. 7, pp. 124329–124340, 2019, doi: [10.1109/ACCESS.2019.2938326](https://doi.org/10.1109/ACCESS.2019.2938326).
- [15] A. C. Williams, "A stochastic transportation problem," *Oper. Res.*, vol. 11, no. 5, pp. 759–770, Oct. 1963, doi: [10.1287/opre.11.5.759](https://doi.org/10.1287/opre.11.5.759).
- [16] M.-T. Tsai, J.-D. Saphores, and A. Regan, "Valuation of freight transportation contracts under uncertainty," *Transp. Res. Part E, Logistics Transp. Rev.*, vol. 47, no. 6, pp. 920–932, Nov. 2011, doi: [10.1016/j.tre.2011.03.005](https://doi.org/10.1016/j.tre.2011.03.005).
- [17] B. K. Thapalia, T. G. Crainic, M. Kaut, and S. W. Wallace, "Single-commodity network design with random edge capacities," *Eur. J. Oper. Res.*, vol. 220, no. 2, pp. 394–403, Jul. 2012, doi: [10.1016/j.ejor.2012.01.026](https://doi.org/10.1016/j.ejor.2012.01.026).
- [18] Y. Hinojosa, J. Puerto, and F. Saldanha-da-Gama, "A two-stage stochastic transportation problem with fixed handling costs and *a priori* selection of the distribution channels," *TOP*, vol. 22, no. 3, pp. 1123–1147, Oct. 2014, doi: [10.1007/s11750-014-0321-4](https://doi.org/10.1007/s11750-014-0321-4).
- [19] S. Singh, A. Pradhan, and M. P. Biswal, "Multi-objective solid transportation problem under stochastic environment," *Sādhanā*, vol. 44, no. 5, p. 105, Apr. 2019, doi: [10.1007/s12046-019-1094-0](https://doi.org/10.1007/s12046-019-1094-0).
- [20] K. Huang, K.-F. Wu, and M. N. Ardiansyah, "A stochastic dairy transportation problem considering collection and delivery phases," *Transp. Res. Part E, Logistics Transp. Rev.*, vol. 129, pp. 325–338, Sep. 2019, doi: [10.1016/j.tre.2018.01.018](https://doi.org/10.1016/j.tre.2018.01.018).



**ZHENPING LI** was born in Shandong, China, in 1966. She received the Ph.D. degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, in 2004. She is currently a Professor with the School of Information, Beijing Wuzi University. Her research interests include operations research and management science, logistics optimization, and intelligent algorithm.



**YUWEI ZHANG** received the B.S. degree in logistics management from Lanzhou Jiaotong University, China, in 2015, and the M.S. degree in logistics engineering from Beijing Wuzi University, China, in 2019. She is currently pursuing the Ph.D. degree in management science and engineering with the Capital University of Economics and Business, China. Her research interests include integer programming and stochastic programming.



**GUOWEI ZHANG** received the B.S. and M.S. degrees from North China Electric Power University, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in operations research with the Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China. His research interests include combinatorial optimization, logistics, approximation algorithm, and machine learning.

...