Application of Markov Decision Processes (MDPs) in Petroleum Industry

Wanying Shi, Jian Guo

Abstract—**Gasoline and diesel fuel is the lifeblood that keeps our daily life moving forward. Inefficient operation of fuel supply leads to unsatisfactory service, time consuming, as well as low economic benefits. Exploring the optimal timing for gas stations to replenish gasoline and diesel is of importance. We propose to apply infinite-horizon Markov Decision Processes (MDPs) to this dynamic problem. Compared with traditional methods for determining the optimal timing of replenishment, such as IB, EOQ, EB, etc., MDPs are better in accurately modeling the situation which needs sequential decision making under uncertainties. For the MDPs modelling gas station replenishment problem, the rewards for any actions taken in the states (the remaining gasoline and diesel inventory status in the oil tank of the gas station) is to keep the duration for stockout and the tanker trucks' waiting time as low as possible. The optimal policy is to maximize the rewards. A real world case study was presented and a revised infinite-horizon MDPs model was constructed to optimize the time for replenishment. Managerial insights guiding the actions gas stations should take to optimize their replenishment strategies are gained.**

Index Terms—**MDPs, optimization, petroleum industry**

I. INTRODUCTION

Gasoline and diesel, the main finished products of the downstream petroleum supply chain, are the lifeblood of people's daily life and businesses. The downstream petroleum supply chain refers to crude oil refining, marketing, and distribution of finished petroleum products. Refining, the remanufacturing process in downstream petroleum supply chain, is the starting point of producing finished petroleum products. There are four transportation ways- pipelines, railway, road, and water transportation to deliver gasoline and diesel to end users. The activities involved in the distribution process can be classified into two stages by end points. Refinery to oil depots (distribution center) is defined as the first (primary) distribution stage, while oil depots to various gas stations are defined as the second distribution stage (See Fig. 1). Gas stations, the main components of the second distribution stage, directly relate to customer needs. On one hand, it is better to keep the oil tanks of gas stations filled with enough gasoline and diesel to supply; however, it may increase the inventory costs as well as the other relevant operation costs of gas stations, such as the transportation costs, depreciation of

tanker trucks, and so forth. On the other hand, if gasoline and diesel in the oil tanks are kept in a low level, the risk of stockout increases. Therefore, it is necessary to explore the optimal timing for gas stations to replenish.
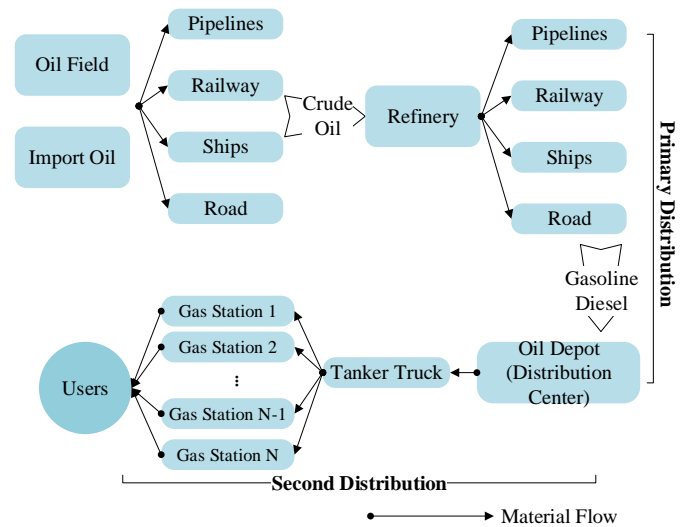

Fig. 1. Downstream petroleum industry

There are a number of literatures exploring the optimal replenishment strategies, which have been applied in various industries. Gallego and Ozer [1] provided the optimal replenishment policies for general multi-echelon inventory problems with incorporating advance demand information into the model. They proved that myopic policies are easy to be determined when demands and involved costs are stationary. Wang and Wu [2] proposed a re-hybrid policy based on hybrid based policy I and hybrid based policy II, which are developed from installation-based (IB) and echelon-based (EB). Re-hybrid policy was effective in reducing costs, especially in the system consisting of one distribution center and many retailers whose demands are independent. Chen, Federgruen, and Zheng [3] developed a set of efficient algorithms to help make optimal pricing and replenishment decisions for a two-echelon distribution system with deterministic demands. Some other articles [4]–[6] also incorporate the time value of money when deciding the optimal replenishment and pricing policy for

deteriorating items.

Though the research articles on exploring general optimal replenishment strategies as well as their applications are plentiful, there have been very few literatures exploring the optimal replenishment strategies for gas stations with uncertainties so far. Tang and Li [7] discussed the optimal replenishment policy of refined oil through demand forecasting and effective usage of tanker trucks under VMI mode. Wang and Cao [8] applied EOQ to determine the optimal replenishment quantity for gas stations so as to reduce inventory cost. Other relevant articles [9]–[11] regard gas stations as nodes in the downstream petroleum supply chain, which need to strengthen safety and information construction.

The objective of this research is to explore the optimal timing for gas stations to replenish considering the changing customer demands. A revised MDPs model with infinite time horizon was built to describe the dynamic problem and explore the decision under uncertainties. The remainder of this research is organized as follows. A detailed literature review was made on the theory basis and wide applications of MDPs. Then we applied MDPs to one real world case study about China's petroleum industry. Combined with the practical situation and characteristics, we improved the MDPs model to make it applicable in petroleum industry. The research was concluded with a discussion on the managerial implications of the model application, and future work was also pointed out.

II. GENERAL MARKOV DECISION PROCESSES (MDPs)

MDPs are mostly applied in a situation or system which consists of a series of discrete event stochastic processes and can be controlled by sequential decisions [12]–[14]. Markov decision processes turns out pretty useful in optimizing problems and has been widely used in various research areas, such as computer science, robotics, economics, and automated control [15], but not common in petroleum industry. Sandikci [16] pointed out that MDPs include several necessary components, which are described below (see Table 1).

TABLE 1
NECESSARY COMPONENTS OF MARKOV DECISION PROCESSES (MDPs)

Components	Description
State space (S)	Set of mutually exclusive that can describe all the possible conditions collectively
Action space (A)	Set of different choices (decisions) at each state
Transition probability (p)	Probability of transferring one state (i) to another state (j), i,j ∈ S
Reward (r)	Value resulting from the actions taken at any state
Decision rule (d)	The decision made for a specific state
Decision epoch (t)	The time points when the decisions are made
Policy (π)	The sequential decisions made for a time range

At any given time point t, there is always one state $i \in S$ corresponding to one decision process. The decision maker may choose any action a from a set of $A(i)$ that is available at state i, which leads the current state i to move to a new state j randomly and the accompanying reward $r_t(i, a)$ come along. Transition probability $p(j|i, a)$ represents the probabilities of moving from state i to j if action a was taken at state i. At each

step of transition, a certain rewards will be received. The objective of MDPs is to explore an optimal policy which determines a sequence of actions so as to maximize the rewards obtained. However, future rewards are discounted along with the time at a constant discounting rate [17]. We use $r_t^*(i_t)$ to represent the maximized expected rewards when the current state is i at time t. Then the future time horizon is $(T - t)$; the expected rewards and the optimal policy will be determined by solving the following recursive Bellman equations [18].

$$
r_t^*(i_t)
$$
$$
= \begin{cases} r_t(i_t), \ i_t \in \mathrm{S}; t = T \\ \max \left\{ r_t(i_t, a) + (1 - \theta) \sum_{j \in S} p_t(j|i_t, a) \, r_{t+1}^*(j) \right\} \\ \qquad t = 1, 2, \dots, T - 1 \end{cases} \quad (1)
$$

Notation:
$r_t^*(i_t)$: Maximized rewards when the state at time t is i
$r_t(i_t)$: Rewards when the state at time t is i
T: Total time horizon
$r_t(i_t, a)$: The immediate reward for taking action a when the state is i at time t
θ: Discounting rate used to measure the future rewards, and the value range for θ is: $0 \le \theta \le 1$
$p_t(j|i_t, a)$: Probability transition from state i to j by taking action a at time t
$r_{t+1}^*(j)$: Maximized rewards when the state at time $t + 1$ is j

The equation to maximize total expected rewards at time t is divided into two situations. For any time points along the time horizon except the terminal point, the maximized expected reward function equals to the summation of the rewards at current time and the discounted future rewards until the time point second to the terminal point ($t = 1, 2, \dots, T - 1$). When it comes to the reward at terminal time point (T), the maximized reward of the system is comparatively easy to calculate, simply the normal reward to be received at that time.

A. MDPs with Finite Time Horizon

MDPs with finite time horizon was quite useful in modeling the problem or system that can be terminated at a specific time point. To maximize the rewards, we can solve the finite-horizon MDPs with Equation (1) directly to decide which state to visit and with what probability. There are extensive applications, such as a finite-year service planning in a service system. Bauerle and Rieder [20] mentioned the stochastic linear-quadratic control problems application, in which the optimal decision rule is a linear function of the states, and thus easy to compute. They also suggest some application areas, such as logistics, healthcare, energy systems, and so forth.

B. MDPs with Infinite Time Horizon

MDPs with infinite time horizon was appropriate in modeling the problem or system that never has a terminal point along the time horizon. The condition that a finite-horizon MDPs model with various states and long enough time horizon can be approximated to an infinite-horizon MDPs model.

Compared with finite-horizon MDPs, addition convergence assumptions have to be satisfied in infinite-horizon MDPs model [18]. For MDPs with infinite time horizon, we usually assume it is stationary and no not need the time parameter (t), which makes the value function be the greatest r-sub harmonic function or the unique value for optimal rewards. According to Alagoz, et al [19], the equation used to find the maximized rewards (optimal policy) for state i is as follows:

$$R(i) = \max \{r(i, a) + (1 - \theta) \sum_{j \in S} p(j|i, a) R(j)\}$$
(2)

$$A = \bigcup_{i=1}^{N} A(i), \quad i \in S$$
(3)

$$\sum_{j \in S} p(j|i, a) = 1, \quad a \in A(i), \ i \in S$$
(4)

In the above equations, R(i) and R(j) represent the maximized rewards at state i and j separately. $A(i)$ represent the available action set for state $i \in$ S. $p(j|i, a)$ represent the transition probability from state i to j by taking action $a \in A(i)$. Equation (2) explains that the maximized rewards for MDPs with infinite time horizon. Equation (3) signifies that the action space A is the union of the available action sets $A(i)$ for all of the states within the system. Since infinite-horizon MDPs can be assumed to be stationary, Equation (4) explains that all of the involved stationary transition probabilities are summed to be one. No time concept t is necessary in infinite-horizon MDPs.

III. REVISED MARKOV DECISION PROCESSES (MDPs) IN PETROLEUM INDUSTRY

The petroleum supply chain sometimes operates with high costs, low efficiency, or even stops due to the frequent replenishment or supply rupture of petroleum products. Most of the time, the problem occurs at the downstream petroleum supply chain, where gas stations need to get gasoline and diesel from distribution center to supply customers. In this part, we apply a revised MDPs model to solve the existing petroleum problem, namely providing the optimal timing policy for gas stations to reorder gasoline and diesel to guarantee enough supply while keeping the inventory as low as possible.

A. Revised MDPs Model in Petroleum Industry

We construct a revised MDPs model in which there is a distribution center which is responsible for supplying gasoline and diesel to a bunch of gas stations. Since gasoline is consumed much faster than diesel, we focus on the optimal timing determination of gasoline replenishment in this research. Uncertain customer demands lead to a fluctuant consuming rate of the gasoline stored at gas stations for any time period. So this is a dynamic problem with stochastic processes. One noticed characteristic is that the time and quantity of gasoline needed at time t only has something to do with the status of the remaining gasoline in the oil tank of gas stations at time $(t - 1)$ but nothing to do with the states prior to time $(t - 1)$. So the

stochastic process is also a typical Markov chain. We use an advised MDPs model, which is built on the properties of Markov chain, to determine the optimal timing of gas station replenishments when the gasoline from distribution center is always available to the decision maker. We seek a policy describing the remaining gasoline status in which immediate replenishment is the optimal strategy and those where waiting until next time point is the optimal strategy. The details of the MDPs components applied in petroleum industry are represented below (see Table 2).

TABLE 2
COMPONENTS OF MARKOV DECISION PROCESSES (MDPs) IN PETROLEUM

Components	Description
State space (S)	The remaining gasoline status in the oil tank of gas station
Action space (A)	Two actions: reorder gasoline immediately or wait until next review period
Transition probability (p)	Probability of transferring between two states after taking any action
Reward (r)	-(duration of stockout + the waiting time of tanker trucks returned from distribution center)
Decision rule (d)	For any state, minimize the summation of the duration of stockout and waiting time of tanker trucks
Policy (π)	The sequential decisions to be made

The revised model is a stationary infinite-horizon MDPs model and discounted with total expected future rewards (see Fig.2). The infinite time horizon signifies the service life of gas stations is long enough under the system control. In the MDPs model, the decision epochs are measured in hours for the reason that gasoline consumption cycle is comparatively short. The states represent the status of how much gasoline left at the oil tank of the gas station and we use the percentage of the full capacity to describe the state. For example, at the very first beginning, namely state 1, we assume the oil tank is full; so state 1 is 100%. Along with the gasoline consumption, there comes a group of different states with different inventory percentages. Until the last state N, assuming no gasoline left at all, so state N is zero. For the remaining gasoline inventory status at each time point, the decision maker has two available decisions to choose: reorder the new gasoline from the distribution center immediately or wait until another review period. However, one point needs to be made clear. As we indicated before, road transportation is the main transportation mode and tanker trucks are the proper vehicles to use for transporting gasoline from the distribution center to the right gas stations. Usually the tanker trucks are categorized according to their full capacities, such as 5 tons, 10 tons, 15 tons, etc., and they will be filled up before driving to gas stations. Then, the gasoline carried in tanker trucks will be unloaded into gas stations completely. It is very dangerous for tanker trucks driving on the road with part of gasoline (less than full capacity) in the tank because the chemical reaction would possibly result in explosion. Therefore, if the gas station reorders too early or the consumption rate is unexpectedly considerably small during the lead time (the time interval between placing order and receiving gasoline), which makes the available capacity of the oil tank (gas station) too small to contain the gasoline carried in

the tanker truck returning from the distribution center, it is no wonder that the tanker truck needs to wait until the enough available capacity of the oil tank is made at the gas station. Thus waiting time occurs. It is not economic to keep the tanker truck waiting. In order to reduce transportation costs and other costs concerned with tanker trucks (i.e., purchasing cost, maintenance costs, depreciation costs, tax, etc.), the number of tanker trucks kept in use is limited. Therefore, waiting causes a waste of resources. The decision maker should reorder gasoline wisely to prevent the tanker trucks' waiting happening. However, if the gas station reorders too late or the consuming rate is unexpectedly large during the lead time, that being said, there is great probability that no gasoline left in the gas station before the tanker truck filled with gasoline comes back. Under this condition, the gas station would lose part of its customers and might have the image damaged. Therefore, the decision maker should also reorder at an optimal timing point to avoid stockout. In conclusion, if the decision maker chooses the "Reorder" option in the current decision epoch, a post-reorder reward is obtained, which can be represented as [- (duration of stockout + the waiting time of tanker trucks returning from the distribution center)]. Or, if the decision maker chooses to postpone the reorder until the next review period by taking the action "Wait", a pre-reorder reward is received as [- (duration of stockout)]. If there is no stockout before reorder at the beginning of the next review period, then the pre-reorder reward will equal to zero, which is the perfect condition. The fact that post-reorder rewards depend on the remaining gasoline inventory status makes the reward not assigned to the reordered state but to the action of reorder from each specific remaining gasoline percentage. Again, we use the general time measurement unit- hour to measure rewards. Another necessary component for MDPs model is transition probability, which determines the progression of the remaining gasoline status, or we can say the probability of transferring between any two of the states randomly. If the current state is i and the decision maker chooses to "Wait" for one more review period, then the gas station can stay at its current remaining gasoline inventory percentage state with probability p($i|i$, W), or move to another state j randomly with probability p($j|i$, W). However, the action "Wait" means no more gasoline from the distribution center coming in while the remaining gasoline at the gas station stays stable or is being consumed; thus state j ranks equal with or behind state i. The remaining gasoline inventory level at state j is lower than or at most equal with the remaining gasoline inventory level at state i. On the other hand, "RO" means the oil tank of the gas station will be filled. So the gas station cannot stay at its current state i but to move ahead to some other states randomly in which the remaining gasoline inventory percentage is comparatively higher than the gasoline inventory percentage at state i.
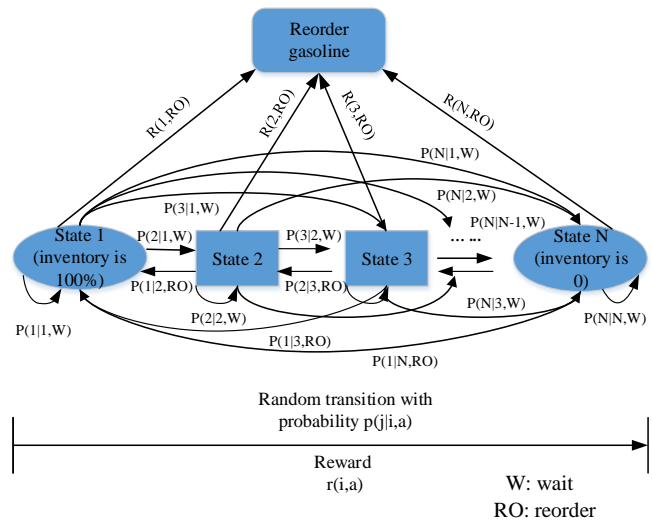

Fig. 2. Markov decision processes (MDPs) in petroleum industry

This infinite-horizon MDPs model for determining the optimal timing for gas stations to replenish includes the following three assumptions:
1) The gasoline and diesel reserved at the distribution center is always enough and available to supply to gas stations.
2) Gasoline is delivered normally, and no accidents, bad weather, or other factors that would prevent gasoline from being delivered normally.
3) The discounting factor used to measure future rewards stays constant, no changing along with the time.

There are seven parameters and one decision variable (see Table 3) in this revised infinite-horizon MDPs model. Let $R(i)$ be the reward of the gas station's gasoline reorder applying the optimal policy when the remaining gasoline inventory percentage (state) is i, i =100% … 0, where 0 represents stockout. We use $R(i, RO)$ to represent the post-reorder expected discounted rewards when the gas station state is i at the time when reorder happens. We also use $p(j|i, W)$ to represent the stationary probability that the gas station state will be j at the review period $t + 1$ when it is i at the review period t given the taken action "Wait".

TABLE 3
PARAMETERS AND VARIABLES IN MARKOV DECISION PROCESSES (MDPs)

	Parameters	Decision Variable
t	Review period (time interval of checking the remaining gasoline status)	
N	Total number of the remaining gasoline inventory states	
i, j	Gas station states at different review periods (i, j ≤ N)	The reward of the gas station using the optimal policy
p(j|i)	stationary transition probability from state i to another state j randomly	R(i)
R(i, RO)	Post-reorder expected discounted rewards	
T(s)	Stockout time period for pre-reorder	
θ	Discounting rate	

With applying the above parameters, the optimal policy to maximize the total rewards resulting from the actions taken at gas stations is found by solving the following recursive equations (5) ~ (6):

$$R(i) = \max \begin{cases} R(i, RO) \\ \{-T(s) + (1 - \theta) \sum_{j=1}^{N} p(j|i) \, R(j)\} \end{cases}$$

(5)

$$\sum_{j=1}^{N} p(j|i) = 1, \qquad a \in A(i), \ i,j \in S$$

(6)

From the above revised infinite-horizon MDPs model, we got known $R(i)$ equals either $R(i, RO)$ corresponding to taking action "reorder" or negative stockout time period for pre-reorder plus future discounted expected rewards written as (-duration of stockout - waiting time of tanker trucks returning from distribution center), which corresponds to action "Wait" until next review period. The optimal policy is the set of actions at the corresponding states that lead to the maximized reward value. The revised MDPs model does not include the transitions to the "reordered" states for two reasons. On one hand, we rely on too much information about pre-reorder states rather than the post-reorder states when deciding the optimal timing to reorder; on the other hand, the rewards of transition between these two types of states (one type belongs to pre-reorder while the other one belongs to post-reorder) equals zero.

B. Case Study

1) Background

A study was conducted within one of the largest Chinese petroleum companies. The company's businesses involve exploiting crude oil, refining (gasoline), and transporting gasoline and diesel to gas stations. The capacities for the vehicles used are separately 5 tons, 10 tons, and 15 tons. There are totally eight distribution centers in the name of this petroleum company, and each distribution center takes charge of supplying gasoline to a group of gas stations. Recently, the operation of gas stations experiencing a low efficiency has been noticed. The gas stations sometimes experience stockout which leads to a serious consequence (i.e., losing customers). However, it also happens that the tanker trucks getting back from the distribution center are waiting to fill the oil tanks of the gas stations. The situation of low efficiency necessitates a study on exploring the optimal timing for gas stations to reorder and replenish their oil tanks. The gas stations check the remaining gasoline inventory status every other one hour. So the review period is one hour.

We focus on studying one distribution center which locates in the northeast part of China, and is responsible for providing gasoline for total 126 gas stations. We collected the relevant data about this distribution center and one of its gas stations from the year 2010 to 2012 through observing the reorder frequencies of this gas station, especially the duration of stockout and the tanker trucks' waiting time. The full capacity of the gas station is 10 tons. The remaining gasoline inventory percentage status was divided into 101 states, such as 100%,

99% … 1%, 0%. Also, the average lead time for the gas station is 2 hours assuming there are no unexpected accidents, bad weather, and any other factors influencing the normal delivery.

2) Results Analysis

We find the policy determining the optimal timing for the gas station to reorder using the revised infinite-horizon MDPs model. Given the market interest rate (general inflation rate included) during 2010 ~2012, a discounted rate to measure future rewards is determined to be 4% on average. Therefore, the applied model changes to be as follows.

$$R(i) = \max \begin{cases} R(i, RO) \\ \{-T(s) + (1 - 4\%) \sum_{j=1}^{101} p(j|i) \, R(j)\} \end{cases}$$

(7)

$$\sum_{j=1}^{101} p(j|i) = 1, \qquad i,j \in [0,100]$$

(8)

According to Denardo [21], the policy iteration algorithm is recommended to solve the above MDPs model. It starts with an arbitrary value until the algorithm finds the policy that returns the maximized rewards though iterative operation. Then the algorithm will stop at the maximized rewards and yield the optimal decision rule. The backward induction was applied in policy iteration algorithm to compare the reward functions one by one. If the reward functions for any two successive steps stay the same, then the optimal policy is generated. We implemented this procedure by running a well-structured C++ program. The developed optimal policy is that the gas station should wait until the remaining gasoline inventory percentage reaches 22% and then to reorder gasoline from distribution center.

IV. CONCLUSION AND FUTURE WORK

This research applied a revised infinite-horizon MDPs model to petroleum industry, determining the optimal timing for gas station to reorder and replenish its oil tank. Unlike the traditional mathematical methods applied in optimizing petroleum supply chain, revised infinite-horizon MDPs model constructed in this paper is more proper to solve practical gasoline reorder problems with uncertain demands, which makes this research meaningful. The result of the case study shows that no matter how the factors (i.e., customer demands) change outside, once the remaining gasoline inventory drops to 22% of the full capacity, the decision maker should place the reorder and make sure that the duration for both the stockout and tanker truck's waiting should be kept as low as possible, and zero is the perfect condition.

This research work brings managerial insights to general gas station replenishment problems on how to determine the optimal timing of reorder under uncertainties. The revised MDPs model in this research could be applied to any similar situations. The future work can try to explore the optimal timing for gas stations to replenish if the gasoline stored in the distribution center is limited.

R<small>EFERENCES</small>

[1] S. C. Graves, S. P. Willems, "Optimizing strategic safety stock placement in supply chains," *Manufacturing & Service Oper.Management*, vol. 2, no. 1, pp. 68-83, Jan. 2000.

[2] Y. Wang, J.-Q. Wu, "Re-hybrid replenishment strategies in coordinated inventory distribution systems," *Systems Engineering – Theory & Practice*, vol. 31, no. 2, pp. 252-261, Feb. 2011.

[3] F.-R. Chen, A. Federgruen, and Y.-S. Zheng, "Near-optimal pricing and replenishment strategies for a retail/distribution system," *Operations Research*, vol. 49, no. 6, pp. 839-853, Dec. 2001.

[4] H. M. Wee, A. T. Law, "Replenishment and pricing policy for deteriorating items taking into account the time-value of money," *Int. J. Production Economics*, vol. 71, no. 1, pp. 213-220, May 2001.

[5] T. Roy, K. S. Chaudhuri, "Optimal pricing for a perishable item under time-price dependent demand and time-value of money," *Int. J. of Operational Research*, vol. 7, pp. 133-151, Jan. 2010.

[6] T. P. Hsieh, C. Y. Dye, "Pricing and lot-sizing policies for deteriorating items with partial backlogging under inflation," *Expert Systems with Applications*, vol. 37, no. 10, pp. 7234-7242, Oct. 2010.

[7] L. Tang, S.-J. Li, "Under the VMI pattern the refined oil makes up the goods question to study," *China Water Transport*, vol. 6, no. 11, pp. 1-3, Dec. 2006.

[8] L. Wang, D. Cao, "Optimized strategy for inventory management of oil products," Oil & Gas Storage and Transportation, vol. 28, no. 12, pp. 6-9, Dec. 2009.

[9] Y.-L. Guo, Q.-F. Hong, and X. Lai, "Exploration and application of KVM in each system of gas stations," *Journal of BPMTI*, vol. 20, no. 3, pp. 46-50, Sep. 2013.

[10] J.-Q. Chen, J.-H. Wang, and L. Z, "Hydrocarbon VOCs pollution in refueling service station and its control techniques," Chinese *Journal of Environmental Engineering,* vol. 1, no. 3, pp. 84-91, Mar. 2007.

[11] Z.-H. Wang, "Optimization of secondary distribution of oil products in SINOPEC corporation," *Modern Chemical Industry*, vol.24, no.2, pp.30-36, Aug.2004.

[12] M. Baykal-Gursoy, K. Gursoy, "Semi-Markov Decision Process," in *Probability in the Engineering and Informational Sciences*, vol. 21, New York: Cambridge University Press, 2007, pp. 635-657.

[13] R. Bellman, *Dynamic Programming.* New Jersey: Princeton University Press, 1957, ch.4.

[14] R. A. Howard, *Dynamic Programming and Markov Processes.* Boston: The Technology Press of MIT, 1960, ch.3.

[15] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* New York: Wiley-Interscience, 1994, ch.3.

[16] B. Sandikci, L. M. Maillart, A. J. Schaefer, O. Alagoz, M. S. Roberts. "Estimating the patient's price of privacy in liver transplantation," *Operations Research*, vol.56, no.6, pp.1393-1410, Dec.2008.

[17] V. S. Borkar, V. Ejov, J. A. Filar, G. T. Nguyen,, "Hamiltonian Cycle Problem and Markov Chains," in *International Series in Operations Research & Management Science*, vol. 171, F. S. Hillier, C. C. Price, Ed. New York: Springer, 2012,pp. 49-51.

[18] N. Bauerle, U. Rieder, *Markov Decision Processes with Applications to Finance.* New York: Springer, 2011, ch.7.

[19] O. Alagoz, H. Hsu, A. J. Schaefer, and M. S. Roberts, "Markov decision processes: A tool for sequential decision making under uncertainty," *Medical Decision Making*, vol. 30, no. 4, pp. 474-483, Aug. 2010.

[20] N. Bauerle, U. Rieder, "Markov decision processes," *Jahresber Dtsch Math-Ver*, vol.212, pp.217-243, Sep.2010.

[21] E. V. Denardo, *Dynamic Programming: Models and Applications.* New Jersey: Prentice-Hall, 2003, ch.8.

Wanying Shi (China, July 1988) is currently a member of IEEE, INFORMS, and IIE. She received her bachelor degree in logistics management, department of management science and engineering, Zhengzhou Institute of Aeronautical Industry Management (Zhengzhou, Henan Province, China 450046) in 2010. She got the master degree in logistics engineering, school of economics and management, Beijing Jiaotong University (Beijing, China 100044) in 2012. She is currently a PhD fellow in engineering management, department of industrial engineering and engineering management, Western New England University (Springfield, MA, USA 01119).
Her publications are listed below:
The Study on Site Selection of Coal Distribution Center based on Logistics Green Grade, International Conference on Logistics, Informatics and Service Science, June 2011, ISBN: 978-989-8425-63-8.
Measuring College Student Satisfaction: Analyzing Interactions among Student Attributes, The Industrial and Systems Engineering Research Conference, May 2014.
Her current research interests are supply chain risk analysis and service value measurement and optimization.

Jian Guo (China, September 1986) is currently the member of IEEE, INFORMS, and IIE. He received his bachelor degree in logistics management, college of transportation management, Dalian Maritime University (Dalian, China 116026) in 2009. He got the master degree in logistics management and engineering, school of economics and management, Beijing Jiaotong University (Beijing, China 100044) in 2011. He is currently a graduate student in business management, college of business, Western New England University (Springfield, MA, USA 01119).
He used to work in PetroChina Planning and Engineering Institute (2011-2013) as an Engineer. Parts of his publications are listed below:
Modeling Li-Ion Battery Capacity Fade Using Designed Experiments. The Industrial and Systems Engineering Research Conference, May 2014
Measuring College Student Satisfaction: Analyzing Interactions among Student Attributes, The Industrial and Systems Engineering Research Conference, May 2014.
His current research interests are modelling battery capacity modelling and data science.