MARITIME EMPTY CONTAINER REPOSITIONING

WITH INVENTORY-BASED CONTROL

LUO YI

NATIONAL UNIVERSITY OF SINGAPORE

2012

MARITIME EMPTY CONTAINER REPOSITIONING WITH INVENTORY-BASED CONTROL

LUO YI

(***B.Eng., Shanghai Jiao Tong University***)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

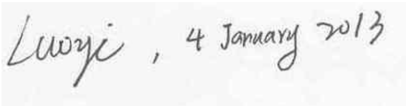
NATIONAL UNIVERSITY OF SINGAPORE

2012

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



LUO YI

4 January 2013

ACKNOWLEDGEMENTS

First and foremost, I thank God for the wisdom and perseverance that he has been bestowed upon me during my doctoral study, and indeed, through my life. “I can do everything through him who gives me strength.” (Philippians 4: 13)

I would like to express my profound gratitude to my supervisors, A/Prof. Lee Loo Hay and A/Prof. Chew Ek Peng for their consistent encouragement and guidance through the whole course of this work. Without their invaluable and illuminating instructions, this thesis would not have reached its present form.

I am grateful for the project collaborators on empty container repositioning, Long Yin and Shao Jijun, for their friendships as well as good advices and collaboration throughout the project. I also would like to express my sincere gratitude to my friend, Li Haobin, for his valuable suggestions and help on coding the simulation model.

I also wish to thank the scholarship support from the Department of Industrial & Systems Engineering in National University of Singapore. Gratitude also goes to all other faculty member and stuff in the Department of Industrial & Systems Engineering, especially the members of System Modeling and Analysis Lab, for their support and advices.

Last, but not the least, I would like to thank my beloved family, especially my boyfriend Zhang Yongfu, for their continuous support and constant love on me.

LUO YI

TABLE OF CONTENTS

SUMMARY

Due to the international trade imbalances between countries, liner operators today often face a challenge to effectively operate empty containers in a dynamic environment. The problem of empty container repositioning therefore is well worth studying and has received considerable attention from both academics as well as industries in recent years. Among a variety of methods proposed for empty container repositioning problem, inventory-based control policies have recently received increasing attention. This thesis focuses on maritime empty container repositioning problem with inventory-based control policies.

Firstly, we address the joint empty container repositioning and container fleet sizing problem in a multi-port system. A threshold-type policy is developed to reposition empty containers periodically. The problem is to optimize the fleet size and the threshold levels of the policy so as to minimize the expected total cost per period. We show that when the fleet size is equal to the sum of the thresholds, the problem can be reduced to a newsvendor problem which can be solved analytically. Meanwhile, we show that it is worth to study the scenario in which the fleet size is not equal to the sum of the thresholds, since this scenario could result in lower operation cost compared to the scenario in which the fleet size is equal to the sum of the thresholds. However, since there is no closed-form formulation of the expected cost function, we build a simulation model and propose a gradient-driven algorithm with infinitesimal perturbation analysis gradient estimator to tackle this problem. The numerical experiments are offered to demonstrate the effectiveness of the proposed policy and to provide some insights for liner operators in managing empty containers.

Then, we extend the previous problem by considering actual service schedule. Previous study simply assumes that the transportation time between each pair of ports is not greater than one unit time, and empty containers can be repositioned between any pair of ports. However, in practical liner shipping operations, empty containers can only be repositioned by vessels, which travel according to the fixed schedules of service routes. And the transportation time between two ports in a service route could be different. Thus, we extend the multi-port system to a liner shipping system with multiple services. This has greatly complicated the empty container repositioning problem. Focusing on empty container, we formulate the problem in a time-driven way and develop two inventory-based control policies to manage empty containers. The numerical studies are provided to examine the relative performances of both policies in some small size problems.

Further, for the problem in the liner shipping system, we optimize the threshold values of a given policy by developing a search algorithm based on the convergent optimization via most-promising-area stochastic search method. A hybrid coordinate and gradient sampling scheme with simultaneous perturbation stochastic approximation gradient estimator is proposed to improve the sampling scheme in the search algorithm in terms of search efficiency. The numerical studies are offered to demonstrate the effectiveness of the proposed sampling scheme and to present the performance of the inventory-based policy in a practical problem.

LIST OF TABLES

LIST OF FIGURES

Chapter 1 INTRODUCTION

Growth in maritime transportation industries has been stimulated by the increase in international merchandise trade as a result of globalization in the last few decades. In particular, the containerization of cargo transportation has been the fastest growing sector of the maritime industries. Containerized cargoes have grown at an annual average rate of about 7.5% over the period 1996 through to 2012 as shown in Figure 1.1.

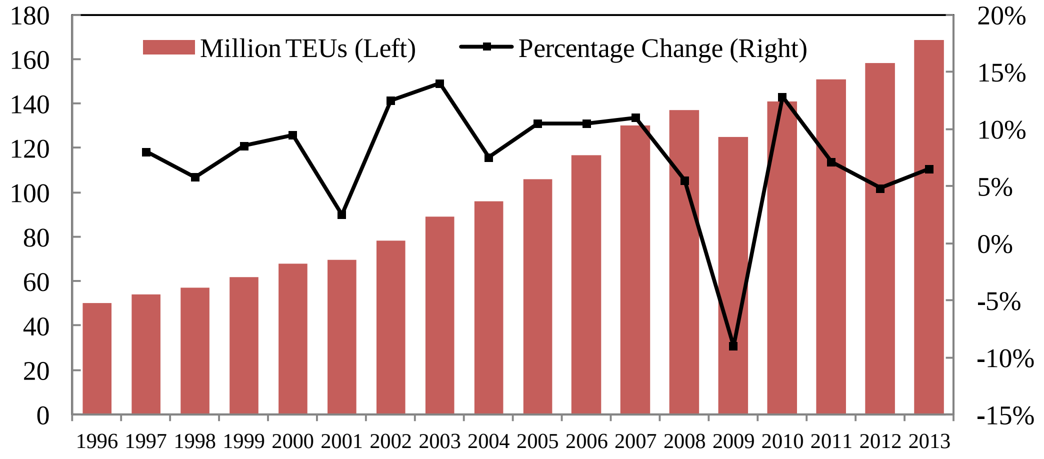


Figure 1.1. Global container trade, 1996-2013 (TEUs and annual percentage change)
Source: United Nations Conference on Trade and Development (2012)

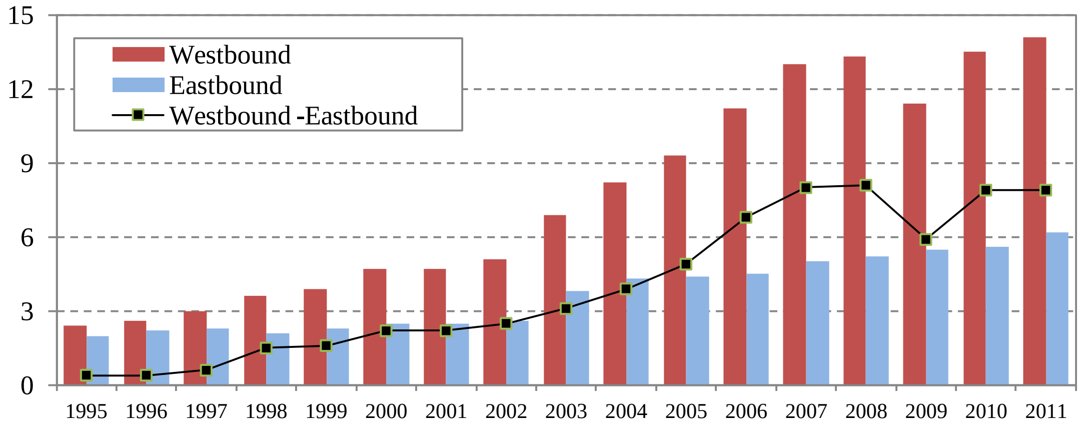
From Figure 1.1, it can be seen that global container trade is continuously increasing, except in 2009. In 2009, global container trade volumes fell for the first time in the history of containerization, which is due to weak global economic conditions. In 2010, however, global container trade experienced robust recovery and volumes bounced back at 12.8% over 2009, among the strongest growth rates in the history of containerization. It is estimated that global container trade is projected to grow by about 6.5% in 2013 to reach 169 million TEUs (Twenty-foot Equivalent Units), equivalent to about 1.5 billion tons. There are tremendous potential for growth in the containerized shipping industry especially with the rapid economic development in economies such like China and India. Meanwhile, the rapid growth of containerized shipping has presented challenges on managing container inevitably, in particular on repositioning empty containers at various geographical levels. The repositioning problem, which is about how to move empty containers from surplus location to demand location in order to satisfy customer shipment demands at the least cost, is known as Empty Container Repositioning (ECR) problem.

In the subsequent sections, we first provide an overview of ECR problem， followed by presenting the objectives and the scope of our study. Then, the contributions of this thesis are presented. Finally, a summary of the contents of this thesis and its structure are presented. A more detailed discussion of previous and on-going research on ECR problem will be presented in Chapter 2.

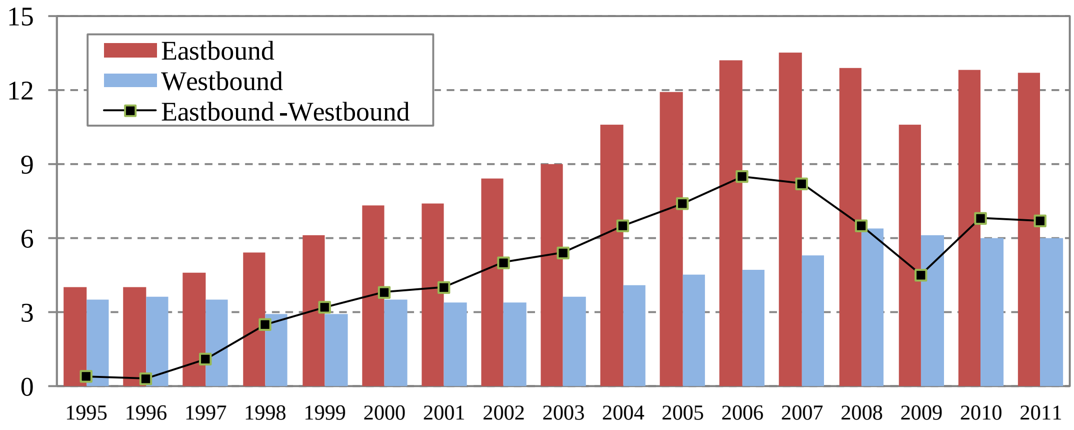
1.1 Overview of empty container repositioning problem

The rapid growth of containerized shipping has effectively removed significant costs and inefficiencies in the movement of materials for shippers. In the container shipping process, it is most efficient that containers are used to hold cargo at all times, possibly by loading new cargo after the previous cargo reaching its destination. However, due to the nature of the equipments used in containerized shipping, this efficient use of containers may not be always possible. Thus, ECR problem persists. This problem is further aggravated by the highly imbalanced nature of global trade.

Figure 1.2 features container trade volumes on the two major East-West container routes between 1995 and 2011. It can be seen that in general, container flows and imbalance between eastbound and westbound trades are continuously increasing. In the



(a) Europe Asia trade route
(Westbound: Asia – Europe; Eastbound: Europe – Asia)



(b) Transpacific trade route
(Eastbound: Asia – North America; Westbound: North America – Asia)

Figure 1.2. Estimated cargo flows along two major container trade routes, 1995-2011 (in Million TEUs)
Source: United National Conference on Trade and Development (2011, 2012)

Europe Asia trade route, in 1995, the flow was 2.4 million TEUs from Asia to Europe and 2.0 million TEUs in the opposite direction. By 2004, the flow had increased to 8.2 million TEUs in the westbound direction and 4.3 million TEUs in the reverse direction. By 2008, the flow reached 13.3 million TEUs in the westbound direction and 5.2 million TEUs in the reverse direction. Although the westbound trade volumes fell in 2009 due to weak economic conditions, it experienced recovery in 2010. Hence, the annual container flow in the westbound direction, i.e., from Asia to Europe, increased by 5.8 million TUEs in the 9 years from 1995 to 2004, and by 5.1 million TEUs in the

4 years from 2004 to 2008. In addition, the imbalance of container flow between the westbound trade and the eastbound trade has also increased. The imbalance was 0.4 million TEUs in 1995, that was 3.9 million TEUs in 2004, and that reached 7.9 million TEUs in 2011. The situation is similar for container flows in the transpacific trade route.

The port in the favorable balance of trade, such as port in mainland China, has a shortage of empty containers. On the other hand, the port in the adverse balance of trade, such as port in Europe, accumulates a large number of surplus empty containers. The unbalanced empty containers between the export-dominated port and the import-dominated port result in a delay of fulfilling customer shipment demands, holding and maintenance costs of unused empty containers, extra purchasing or leasing cost of containers, etc (Liu *et al.*, 2011). Hence, repositioning empty containers from one port to another is an essential mechanism to balance the container flow.

Figure 1.3 shows the ratio of empty containers to total containers handled in ports during 1990 and 2006. It can be seen that until approximately 1996, there was a declining trend in the ratio of empty containers to full containers, since increasingly

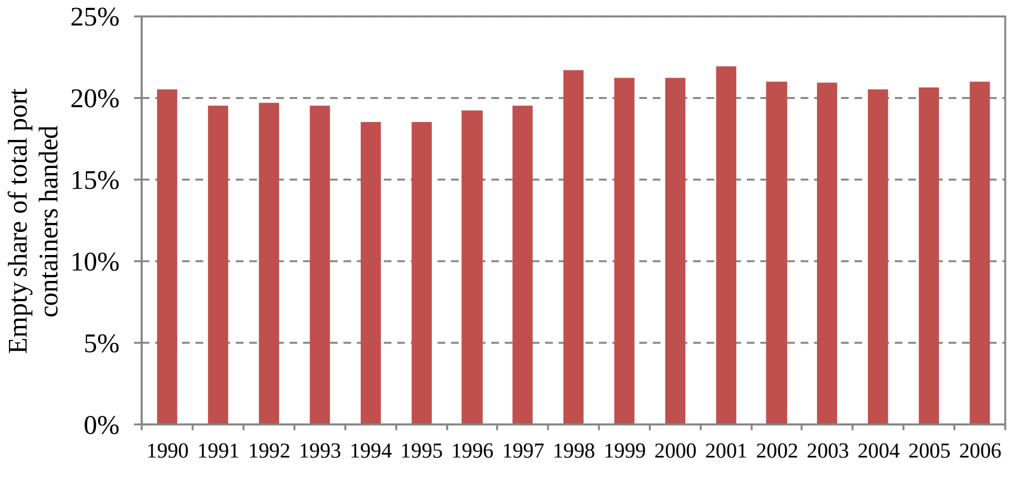


Figure 1.3. Empty share of container movements (1990-2006)
Source: Economic and social commission for Asia and the Pacific (2007)

sophisticated container logistics gradually reduced the number of empty container movements. However, since 1998, the empty container handling has accounted for more than 20% of global port handling activities, mainly due to the very pronounced imbalance in the two main Asian trades with Europe and North America.

Besides, it is estimated that the cost of repositioning empty containers is just under $15 billions, which is 27% of the total world fleet running cost based on the data of 2002 (Song *et al.*, 2005). Drewry also estimated that there were 50 million TEUs of empty container movements in 2009. Assuming a nominal cost of $400 per TEU for each empty container movement, the total cost of repositioning empty containers was estimated about $20 billion in 2009 (United Nations Conference on Trade and Development, 2011). Therefore, efficiently and effectively repositioning empty containers is a crucial issue that shipping companies have to address.

Industrial practitioners and theoretical researchers have come up with various valuable mathematical programming models regarding repositioning empty containers. The decisions about empty containers in those mathematical models are usually presented by a series of values that describe how many empty containers should be moved from which ports to other ports. One major attempt of those models is to find the optimal values of the ECR decisions to balance the container flow or to minimize the related operation cost. However, it is usually difficult to find those optimal values for the ECR decisions and even more difficult to execute the decisions in the situations with high uncertainty (Dong and Song, 2009).

Inventory-based control policies for ECR problem have recently received increasing attention. Such policies utilize the feedback of inventory information to manage empty containers, and are characterized by a set of rules and a set of parameters. Once the rules and parameters are designed, the ECR decisions can be made by following these simple rules. Hence, in practice, inventory-based control policies appear to be more appropriate for making the ECR decisions in systems with uncertainty since they can provide flexibility to cope with dynamic changes and unexpected fluctuation of stochastic factors, such as future customer demands.

Based on the literature review, it has been found that the threshold-type inventory control policy is the optimal repositioning policy for empty containers in some systems, such as one port system and two-depot system (Li *et al.*, 2004; Song, 2005; Song, 2007), and such policies have advantage of being easy-to-operate and easy-to-understand. Hence, our study focuses on ECR problem with inventory-based control policies and tries to provide some managerial insights for shipping companies.

1.2 Research objectives and scope

This thesis studies maritime ECR problem with inventory-based control policies. The specific objectives of this thesis are:

- To study ECR problem in a multi-port system. In particular, a single-level threshold policy with intelligent rule is developed to manage empty containers. Two approaches are proposed to optimize the fleet size and the parameters of the policy by taking advantage of the property of this problem. The performance of the proposed policy is evaluated by some numerical experiments and some insights for managing empty container in the multi-port system are also provided.
- To extend the simulation model of ECR problem in the multi-port system to that in a complex liner shipping system with multiple service routes. In particular, two inventory-based control policies are developed.
- To optimize the parameters of a two-level threshold policy by a proposed search algorithm. In particular, a hybrid sampling scheme with gradient information is proposed to speed up the convergence rate of the search algorithm. The performance of

the proposed sampling scheme is tested on a benchmark function and the proposed algorithm is utilized to tackle an ECR problem in a complex liner shipping system.

The focus of this thesis is to make maritime ECR decisions for liner shipping companies. We only consider one maritime transportation mode, i.e., by vessel. Other transportation modes, such as by barge, by truck or by rail are not considered in this study. Besides, following a traditional assumption (Cheung and Chen, 1998), we only consider one type of container, i.e., TEU. In addition, in the liner shipping system, we assume that the schedule of a service route is given and fixed. The service schedule is the timetable of when each vessel will call at each port in the service route. And a service route is defined as a special sequence of calling ports. Weekly service for each route is commonly provided by most liner shipping companies. That is, in each week, for the port in a service route, it will be called by a vessel deployed in this route at a fixed day according to the service schedule. Finally, note that we do not make decisions on laden container transportation in this study. As laden container transportation problem and ECR problem are usually considered separately in current shipping industry.

1.3 Contributions of the thesis

The contributions of this thesis are listed as follows:

- We develop a single-level threshold policy with a repositioning rule in terms of minimizing repositioning cost to reposition empty containers in a multi-port system. Applying this intelligent policy, the operation cost for repositioning empty container could be significantly reduced.

- We present the proof that for the multi-port system with a single-level threshold policy, the policy with sum of thresholds $\neq$ the fleet size could result in less operation cost than the policy with sum of thresholds $=$ the fleet size. It provides a important insight that keeping more (less) empty containers over the threshold in import-dominated (export-dominated) port in advance when it becomes a deficit (surplus) port could reduce the repositioned in (out) quantity of empty containers.

- We propose a gradient-driven algorithm with infinitesimal perturbation analysis (IPA) gradient estimator to optimize the fleet size and the parameters of the policy for the multi-port system. The value of this gradient can be obtained from only one simulation run in each iteration, which can greatly save the computation time to provide good solution.

- In the procedure to obtain the gradient estimator, we develop a modified stepping stone technique to explore the perturbations on ports. It is innovative and provides a potential methodology contribution in the field of application of the stepping stone method.

- We build a time-driven simulation model with considering multiple service routes and uncertain demand, supply and residual capacity on vessel for empty containers. And inventory-based threshold policies are developed to manage empty containers. The simulation model and the policies can be used by the liner shipping company analysts to explore other operation options in the future.

- We develop a search algorithm based on the Convergent Optimization via Most-Promising-Area Stochastic Search (COMPASS) algorithm to optimize the parameters of an inventory-based policy in a liner shipping system. It provides a potential methodology contribution to the application of COMPASS algorithm in complex systems.

- We propose a Hybrid Coordinate and Gradient Sampling (HCGS) scheme to improve the convergence rate of COMPASS. The HCGS scheme could be easily applied in other random search algorithms to speed up the convergence rate, especially for high-dimensional problems.

- The optimal values of the fleet size and the thresholds of the policy in the general or complex liner shipping system could be used as reference points for shipping companies to make strategic decisions.

1.4 Organization of the thesis

The thesis consists of six chapters. The rest of this thesis is organized as follows. Chapter 2 introduces existing studies about ECR problem. In Chapter 3, the model for ECR problem with a single-level threshold policy in multi-port system is presented. Two approaches, one non-linear programming approach and one gradient-driven algorithm, as well as numerical studies are presented. Chapter 4 studies the complicated ECR problem in a multi-service liner shipping system, and a time-driven simulation model is developed. Numerical studies are provided to analyze the performance of policies for some simple systems. In Chapter 5, for the ECR problem in Chapter 4, a search algorithm based on COMPASS with proposed hybrid sampling scheme is developed to optimize the parameters of a given policy. Numerical studies are conducted and results are presented in this chapter. Finally, in Chapter 6, we conclude this thesis and discuss several issues for future research.

Chapter 2 LITERATURE REVIEW

In this chapter, we present a survey of literature pertinent to ECR problem. There have been a lot of studies considering ECR problem. From the viewpoint of the adopted methodologies, we generally categorize the literature into two groups. The first group tries to seek the solutions about ECR decisions, involving whether to reposition empty containers, to or from which ports, by solving mathematical programming models. The second group focuses on examining inventory-based control policies and their implementation for ECR problem.

2.1 Mathematical programming models for ECR problem

Extensive literature could be found for ECR problem with mathematical programming models. Much of the work has considered the empty equipment allocation and distribution problem and the balancing of demand and supply between depots or terminals to meet future customer demands.

Crainic *et al.* (1989) presented a class of models for the Multicommodity Location/allocation problem with interdepot Balancing requirements (MLB). Their models were initially developed as part of a strategic and tactical planning system for a maritime shipping company. Their general goal was to locate depots for empty containers with the objective of satisfying client demand, while minimizing depot opening and operation costs, bidirectional client-depot transportation costs, as well as the costs of the interdepot movements necessary to counter the unbalancing of demand. These interdepot movements differentiated such problem from classical location/allocation applications. Many studies have attempted to solve the MLB problem. Crainic and Delorme (1993) used a dual-ascent based approach for this problem. Crainic *et al.* (1993b) solved the problem by an adaption of the tabu search metaheurist. The method of Crainic *et al.* (1993b) generally could obtain better solutions than that by Crainic and Delorme (1993), while it requires rather large computing times. Gendron and Crainic (1995) later used a branch-and-bound algorithm, in which bounds were computed by a dual-ascent procedure, to solve the MLB problem. Gendron and Crainic (1997) then presented a parallel branch-and-bound algorithm based on a depth-first branch-and-bound procedure. Bourbeau *et al.* (2000) proposed three strategies that may be used to parallelize most brand-and-bound algorithms specialized to location/network design formulations.

Crainic *et al.* (1993a) described the problem of allocating empty containers in a land distribution and transportation system, which involved allocating empty containers of different types in response to requests by export customers and repositioning other empty containers to depots or ports to meet future demands. They introduced two dynamic deterministic formulations for the single and multi-commodity cases, which offer a general modelling framework for this class of problems. Abrache *et al.* (1999) used the formulation of Crainic *et al.* (1993a) for a multi-commodity empty containers allocation problem in a deterministic and dynamic situation. To solve the model, they proposed a decomposition approach, which was based on the classical restriction framework that takes into account the specificities of the model, particularly the substitution property between the different container types. Belmecheri *et al.* (2009) proposed a mathematical model to optimize empty container movements between regional consignees, shippers, and port terminal, which was solved by the Solver of Excel.

Shen and Khoong (1995) developed a decision support system to solve a large-scale planning problem about the multi-period repositioning of empty containers based on network optimization models. Besides optimizing on the quantities of repositioned empty containers across ports, their system was also able to recommend container leasing-in and off-leasing decisions. Bandeira *et al.* (2009) later proposed a decision support system for integrated distribution of empty and full containers among customers, leasing companies, harbors and warehouses. The problem was modeled as a network, and the underlying mathematical model was operated in stages. First, it prioritized and adjusted demands considering available empty container supplies, and then statically optimized costs. Transportation routes were registered and dynamically controlled, cyclically, for a given time horizon.

Olivo *et al.* (2005) considered the ECR problem with multiple transportation models for logistic companies. They adopted a dynamic network with an hourly time-step for the problem. In the network, arcs represented service routes, inventory links and decisions concerning the time and place to lease containers from external sources. A case study of the Mediterranean basin was conducted to show good computational efficiency of their algorithms. Francesco (2007) provided several mathematical optimization models for ECR problem in terms of minimizing the total operation cost.

Another popular problem that many researchers have looked at is the empty container balancing strategies within the context of a maritime network design problem. Shintani *et al.* (2007) formulated a two-stage model to address the design of container liner shipping service networks by explicitly taking into account empty container repositioning. The first stage model was to construct the calling port sequence, and the second stage model was to estimate the profit of container management with repositioning empty containers. A genetic algorithm-based heuristic was developed to get the optimal port sequence. They had shown that considering ECR in the network design problem could provide a more insightful solution than the one not considering. Subsequently, Chen and Zeng (2010) focused on the optimization of container shipping network under changing cargo demand and freight rates. They formulated this problem as a mixed integer non-linear programming problem with an objective of maximizing the average unit ship-slot profit at three stages. Their first two stages were similar with that in Shintani *et al.* (2007). But they used a series of the matrices of demand to represent fluctuating demand at the second stage rather than only a matrix of average demand. The third stage was to determine and to arrange the optimal configuration of container with the owned container quantity, long-term leasing container quantity and short-term leasing container quantity. A bi-level genetic algorithm was designed to solve the problem.

Imai *et al.* (2009) studied two typical service networks with different ship sizes, i.e., multi-port calling by conventional ship size and hub-and-spoke mega-ship, taking ECR into consideration. In their study, the problem was analyzed in two phases: the service network design and container distribution. In the container distribution phase, the problem was formulated as an integer programming solved by a commercial mathematical programming solver. Their work provided that under some circumstances, the conventional hub-and-spoke operation was an optimal solution without ECR, whereas the multi-port-calling operation dominated when considering the issue of empty containers. It highlights that ECR has significant impacts on the liner shipping network. Meng and Wang (2011) considered a liner shipping service network design problem with combined hub-and-spoke and multi-port-calling operations and ECR. It introduced a concept *segment* defined as a pair of ordered ports served by one shipping line and subsequently developed a mixed-integer linear programming model for the proposed problem. They had shown that it is advantageous to design a network with the consideration of empty containers.

Some researchers also study the ship deployment problem, which is to assign a fleet of ships to a given network with fixed calling sequence, with considering ECR. Liu *et al.* (2011) developed a tactical model which considered jointly container flow management and ship deployment problem. They presented that jointly considering the container flow, including laden container flow as well as empty container flow, and ship deployment problem could improve the utilization of shipping capacity and profitability of the shipping company.

Jula *et al.* (2006) considered the ECR problem, which they referred to as the empty container reuse problem from a different perspective. Their aim was the reduction of the traffic congestion in the Los Angeles and Long Beach area caused heavily by empty container traffic. They considered two empty container reuse methodologies, street-turn and depot-direct. In street-turn, empty containers were directly moved from local consignees to local shippers. In depot-direct, empty containers were moved from depot to local shippers. A network formulation model was constructed in order to optimize empty container movements following the both methodologies. The problem was solved in two phases. In the first phase, they transformed the model into a bipartite transportation network, and in the second phase, they found the best match between supply and demand of empty containers in the transportation network. They showed that when time is critical, empty container reuse is shifted towards depot-direct, since waiting time is minimal in this methodology. On the other hand, when the traveling cost and traffic congestion are the important factors, street-turn methodology provides the best match between supply and demand of empty containers. Based on the empty container reuse problem presented in Jula *et al.* (2006), Chang *et al.* (2006) investigated multi-commodity empty container substituted problem, in which one type of containers can be substituted with another container. They presented that the cost of empty container reuse can be further reduced by allowing the substitution between empty containers. Besides, they also showed that the reuse approaches in Jula *et al.* (2006) can also perform well in a stochastic environment.

Moon *et al.* (2010) studied the ECR problem that considered short-term leasing and purchasing. Mixed integer linear programming and genetic algorithms were used to solve the model and a hybrid genetic algorithm was proposed to reduce the computation time. In the hybrid genetic algorithm, a heuristic method was developed to find feasible solutions for the problem. The main idea of the heuristic method was to try to satisfy customer demands by owned empty containers first and then to reposition surplus empty containers randomly. These feasible solutions obtained from the heuristic method could be further utilized to find the optimal solution by solving a linear mathematical model. They presented that their proposed algorithms were capable of solving problems of larger size. Moon *et al.* (2013) considered the ECR problem with foldable and standard containers. Similar heuristic methods with that in Moon *et al.* (2010) were proposed to find feasible solutions for the problem, in which the decisions about repositioned empty containers were related to the purchasing cost and repositioning cost at each port. Then, a local search algorithm was applied to improve the quality of the obtained feasible solutions. They pointed out that the potential cost saving could be achieved by using foldable containers in maritime transportation.

The effect of the planning horizon length on empty container distribution management has also been looked at by a few researchers. Choong *et al.* (2002)

conducted a computational analysis of the effect of planning horizon length on empty container management for intermodal transportation networks. The analysis is based on an integer programming that aims to minimize total costs related to repositioning empty containers, subject to meeting requirements for moving laden containers. The study concludes that a longer planning horizon can encourage the use of inexpensive, slow transportation modes, such as barge.

Most of above studies reviewed have taken the deterministic approach and the decisions about empty containers are dynamic in time. However, in container shipping, there are many stochastic factors, such as the real transportation time between two ports/depots, future customer demands and returned containers, and the available capacity in vessels for empty containers, etc. Such stochastic factors of the ECR problem have attracted much attention since 1990s.

Cheung and Chen (1998) developed a two-stage stochastic network model to determine the maritime ECR and leasing decisions in terms of minimizing the total operation cost. All information in the first stage was given while some parameters in the second stage were uncertain when decisions in the first stage were made. A stochastic quasi-gradient method and a stochastic hybrid approximation procedure were applied to solve the stochastic model. Lam *et al*. (2007) formulated the ECR problem as a dynamic stochastic programming with the decision policy optimal in the infinite horizon average cost sense. Linear approximation architecture was chosen to approximate the cost function. A simulation based approximate dynamic programming approach was deployed to solve the problem. Erera *et al.* (2009) presented a robust optimization framework based on time space network for dynamic ECR problems. The robust repositioning plan was developed based on the nominal forecast value and could be adjusted under a set of recovery sections. Di Francesco *et al.* (2009) proposed a multi-scenario model to address the ECR problem in a scheduled maritime system. In the scenario-based model, deterministic optimization techniques could be applied to solve the stochastic ECR problem. Long *et al.* (2012) formulated a two-stage stochastic programming model with random demand, supply, ship weigh capacity, and ship space capacity for ECR problem. The sample average approximation method was applied to approximate the expected cost function and they considered the scenario aggregation by combining the approximate solution of the individual scenarios problem. Two heuristic algorithms based on the progressive hedging strategy were applied to solve the problem.

The mathematical programming models often successfully capture the stochastic and dynamic nature of the ECR problem (Song, 2007). However, they also give rise to some concerns.

Firstly, the decisions about empty containers in those mathematical models are usually presented by a series of values that describe how many empty containers should be moved from which ports to other ports. One major attempt of those models is to find the optimal values of ECR decisions to balance the container flow or to minimize the related operation cost. However, it is usually difficult to find those optimal values for ECR decisions due to the computational complexity. And additional mechanisms could be required to execute the decisions in the situations with uncertainties.

Secondly, a pre-specified planning horizon is usually required for the mathematical models. In other words, it is important to select an appropriate planning horizon for the mathematical models. According to Choong *et al.* (2002), the planning horizon had a significant effect on empty container management. They reported that a longer planning horizon could encourage the use of inexpensive slow transportation modes to reposition empty containers.

2.2 Inventory-based control policies for ECR problem

Inventory-based control policies for ECR problem have recently received increasing attentions. Such policies utilize the feedback of inventory information to manage empty containers. Their decision process is more like rule-based rather value-based. That is, the number of repositioned empty containers is not designed in advance; instead, the parameters and the rules of the policy should be designed in advance. For example, (*U*, *D*) policy (Li *et al.*, 2004) is a inventory-based control policy with two parameters, i.e., *U* and *D*. Its rule is repositioning in empty containers up to *U* when the number of empty containers in a port is less than *U*, or repositioning out empty containers down to *D* when the number of empty containers is larger than *D*, doing nothing otherwise. Therefore, once the parameters and the rules of the policy are set up, such policy can be applied to make the ECR decisions involving whether to reposition empty containers, to or from which ports, and in what quantity. Hence, in practice, such policies appear to be more appropriate for making ECR decisions in stochastic systems since they provide flexibility to cope with dynamic changes and unexpected fluctuation of stochastic factors, such as future customer demands.

Such policies, in fact, are quite similar with the policies for traditional inventory systems, such as (s, S) policy, in which numerous literature have been developed, e.g., Iglehart (1963) and Sheopuri *et al.* (2010). However, there are significant differences between the studies about the inventory-based policies for ECR problem with that for the traditional inventory problems. For example, in the maritime transport, container itself has to be carried by vehicle such as vessel. Since most vessels are committed to specific service routes with published schedules that are often fixed for more than six months, the routes for repositioning empty containers are, of course, fixed. Besides, due to the uncertainty of the customer demand and supply for containers, the source points for empty containers repositioned could be varied in different periods.

Focusing on the inventory-based control policies for ECR problem, there are a number of studies. Imai and Rivera (2001) addressed the problem of fleet size planning for refrigerated containers, in which a heuristic strategy was developed to allocate empty containers between each ports. Lai *et al.* (1995) considered the ECR problem from mid-east port to far-east ports. They used a simulation model to evaluate different container allocation policies, which were characterized by a safety stock level, critical allocation point and port priority. Their study is a major milestone in the development of simulation model for container fleet sizing problem with inventory policies. Du and Hall (1997) utilized inventory and queuing theory and proposed a single threshold policy to redistribute empty containers in a hub-and-spoke system with random demands and deterministic travel times. Their model was based on the assumptions that the number of terminals was large and terminal stock-out probabilities were low. A decomposition approach was then developed to determine the fleet size and control parameters for the policy. Feng and Chang (2008) formulated the ECR problem for intra-Asia liner shipping as a two-stage problem. The first stage identified safety stock at each port to determine the number of repositioned empty containers, and the second stage solved the transportation problem by linear programming. The value of safety stock in each port was estimated as an average of difference in known inbound containers and outbound containers during two weeks.

Several researchers turn to explore the inventory-based mechanism in addressing ECR problem in the stochastic systems. Li *et al.* (2004) formulated the one port containerization problem as a non-standard inventory problem with positive and negative demands. They showed that the two-level threshold policy was optimal for the single port system and a value iterative algorithm was proposed to calculate the optimal threshold values. Song and Zhang (2010) also considered ECR problem for a single port. In their study, the flow of containers was treated as continuous fluid and dynamic programming was used to solve the problem. They demonstrated that the optimal policies were given in terms of threshold levels and closed form solutions of these threshold levels could be obtained. Song (2005) and Song and Earl (2008) considered the empty vehicle redistribution problem in a two-depot system with continuous-reviewed. As they had mentioned, a vehicle may be defined as a reusable resource for realization of a given kind of transportation, such as ECR in shipping business. They applied Markov decision process theory to the problem and presented that the optimal stationary policy for ECR was of threshold control type when the long-run average cost and the infinite-horizon expected discounted cost, respectively. The explicit form of the cost function under such threshold controls can be obtained, and then the optimal threshold values can be derived. Song (2007) considered the similar problem with Song (2005) with periodical-reviewed. Song (2007) proved that optimal empty repositioning policy was also of threshold control structure in such system and a value iterative algorithm was applied to find the optimal threshold values. These studies demonstrate that the optimal repositioning policies are of threshold-type, which are characterized by a set of parameters and a set of rules, in some situations such as one-port and two-port systems. Once the parameters and rules are designed in advance, such threshold control type policies are easy to operate.

Above works are extended to focus on the implementation of the threshold-type control policies for ECR problem in more general systems. Li *et al.* (2007) extended the study by Li *et al.* (2004) to a multi-port system. Since multiple ports were considered, the optimal two-level policy for only one port may not be used successfully in the multi-port case. They modified the policy and designed a heuristic algorithm to compute the number of empty containers allocated between each pair of ports. Song and Carter (2008) further extended works by Song (2005) and Song and Earl (2008) to a hub-and-spoke system, in which only the demands between the hub and spokes were considered. It was found that static threshold-type control policy was not usually optimal in such system. Thus, a dynamic decomposition algorithm was proposed. And the analysis showed that the dynamic decomposition procedure can produce a near-optimal policy. Song and Dong (2008) applied the threshold policy for empty container management in a cyclic shipping route problem and demonstrated that the threshold policy significantly outperformed the heuristic policies with simulation results. Heuristic methods based on experiences were used to set up the threshold control parameters.

Dong and Song (2009) considered the joint container fleet sizing and ECR problem in a multi-vessel liner system. A two-level threshold policy was adopted to allocate empty containers and a simulation-based optimization tool based on genetic algorithm was developed to optimize the container fleet size and the parameters of the policy simultaneously. Dong and Song (2009) investigated the impact of inland transportation times and their variability on the optimal container fleet size in liner services. A rule-based policy was applied to manage containers, whose idea was to satisfy important customer demands first and then to reposition empty containers by balancing the container flows in each port-pair. Three simulation-based optimization approaches, golden section, genetic algorithm and simulated annealing were applied to solve the optimal container fleet sizing problem. Song and Dong (2010) studied the

ECR policy with flexible destination ports. Numerical results showed that such policy outperformed the conventional policy significantly in situations where trade demands were imbalanced and container fleet sizes were within reasonable range. Song and Dong (2011) formulated the ECR problem for shipping service routes at the tactical decision level. Two types of flow balancing mechanisms were considered in their study: one is a point-to-point repositioning policy and the other is a coordinated repositioning policy. A heuristic procedure was developed to solve the coordinated balancing problem.

We have reviewed the studies about ECR problem with inventory-based control policies. These prior studies enable us to have a better understanding of the general implementation of such policies for ECR problem. However, from the literature, we find that the current studies are not sufficient in addressing ECR problem with the inventory-based policies in general or complicated liner shipping systems. Table 2.1 summarizes the differences between this thesis and previous studies about implementation of inventory-based policies for ECR problem.

Table 2.1 Comparison between this thesis and previous studies about implementation of inventory-based policies for ECR problem

	General shipping system	Liner shipping system	
		Single service route	Multiple service routes
ECR	Li *et al.* (2007) Song and Carter (2008) Song and Dong (2008) ***Chapter 3***	Dong and Song (2009) Dong and Song (2009) Song and Dong (2011)	Song and Dong (2010) ***Chapters 4 and 5***

Firstly, few of the studies have considered the implementation of such policies in multi-port system with direct empty container flows between each pair of ports. Li *et al.* (2007) considered a two-level threshold policy in a multi-port system and proposed a heuristic algorithm to allocate the empty containers. However, it could be computationally expensive when the number of ports and the fleet size are very large. Besides, the fleet sizing problem is also not fully studied in such system. Moreover, most policies in current works apply simple rule, such as linear rationing rule to reallocate empty containers. The repositioning rule in terms of minimizing the repositioning cost has not been considered yet.

Secondly, the existing studies are not sufficient in addressing ECR problem in the liner shipping system with multiple service routes. There is a series of papers by Song et al. considering the implementation of threshold policies for ECR problem in liner shipping systems. However, to our knowledge, only one study considers the system with multiple service routes. For the problem considering multi-service, it is possible that multiple vessels will arrive at or depart from a same port within one week, or even at the same day. This is one of key differences between studies about multi-service system and that about single-service system. Thus, the rules with considering the cooperation of multiple vessels to reposition empty containers are worth to study. Moreover, to our knowledge, no existing study considers the optimization about the parameters of the inventory-based policy for ECR problem in the multi-service system due to the complexity of this problem.

To fill these gaps, the major aim of our study is to study the implementation of the inventory-based policy for ECR in a general system, and then to extend it to more complicated system with taking multiple service routes into account.

Chapter 3 EMPTY CONTAINER MANAGEMENT IN MULTI-PORT SYSTEM WITH INVENTORY-BASED CONTROL

In this chapter, we address the joint ECR and container fleet sizing problem in a multi-port system with inventory-based control mechanism. A single-level threshold policy with repositioning rule in terms of minimizing the repositioning cost is proposed. The objective is to optimize the fleet size and the parameters of the policy to minimize the expected total cost per period, incurred by repositioning empty containers, holding unused empty containers and leasing empty containers. Taking advantage of an interesting property of the problem, we propose two approaches, a non-linear programming and a gradient-driven algorithm with IPA gradient estimator to tackle this problem.

3.1 Problem formulation

We consider a multi-port system, which is consisting of ports connected to each other. A fleet of owned empty containers (ECs) meets exogenous customer demands, which are defined as the requirements for transforming ECs to laden containers and then transporting these laden containers from original ports to destination ports. A single-level threshold policy with periodical review is adopted to manage ECs. According to the study by Li *et al.* (2004), the optimal repositioning policy minimizing the long-run average cost existed with the two-level structure, but it may degenerate to a single level. Moreover, although the single-level threshold policy cannot be shown to be optimal, it is adopted here because of its simplicity and it is similar to the base stock policy which is widely accepted and applied in many literatures.

At the beginning of a period, the ECR decisions are made for each port, involving whether to reposition ECs, to or from which ports, and in what quantity. Then, when customer demands occur in the period, we can use those ECs that are currently stored at the port and those ECs that are repositioned to the port in this period to satisfy customer demands. If it is not enough, we need to lease additional ECs immediately from vendors.

3.1.1. Modeling assumptions

The following assumptions are made:

- Customer demands must be satisfied in each period; and customer demands for each pair of ports in each period follow independent Normal distributions.
- Short-term leasing is considered and the quantity of the leased ECs is always available in the port at any time.
- The leased ECs are not distinguished from owned ECs, i.e., the shipping company can return owned ECs to the vendors when it has sufficient ECs available.
- The travel time for each pair of ports is less than one period length.
- When the repositioned ECs arrive at the destination ports, they will become available immediately; and when laden containers arrive at the destination ports, they will become empty and be available at the beginning of next period.
- The cost of repositioning an EC from p to port m is the summation of the handling cost of an EC at port p, the handling cost of an EC at port m, and the transportation cost of an EC from p to port m.

3.1.2. Notations

Notations used in this study are presented as follows:

N The fleet size, which is the number of owned ECs

P The set of ports

t The discrete time decision period

P_t^S The surplus port subset in period t

P_t^D The deficit port subset in period t

$x_{p,t}$ The beginning on-hand inventory of port p in period t

$y_{p,t}$ The inventory position of port p in period t after making the ECR decisions

$z_{p,m,t}$ The number of ECs repositioned from port p to port m in period t

$\varepsilon_{p,m,t}$ The random customer demand from port p to port m in period t

$u_{p,t}^S$ The number of estimated EC supply of surplus port p in period t

$u_{p,t}^D$ The number of estimated EC demand of deficit port p in period t

$\mathbf{x}_t$ $= [x_{p,t}]_{p\in P}$ the vector of the beginning on-hand inventory in period t

$\mathbf{y}_t$ $= [y_{p,t}]_{p\in P}$ the vector of the inventory position in period t

$\mathbf{Z}_t$ $= (z_{p,m,t})_{p\in P,m\in P,p\neq m}$ the array of repositioned quantities for all ports

ω_t $= (\varepsilon_{p,m,t})_{p\in P,m\in P,p\neq m}$ the stochastic customer demands in period t

$C_{p,m}^R$ The cost of repositioning an EC from port p to port $m, \forall p \neq m$

C_p^H The cost of holding an EC at port p per period

C_p^L The cost of leasing an EC at port p per period

γ_p The threshold of port p

$\boldsymbol{\gamma}$ $= [\gamma_p]_{p\in P}$ vector of the thresholds

$\eta_{p,t}^O$ $= \sum_{m\in P,m\neq p} \varepsilon_{p,m,t}$ the number of total exported laden containers of port p in period t

$a_{p,t}$ $= \sum_{l\in P,l\neq p} z_{l,p,t} - \sum_{m\in P,m\neq p} z_{p,m,t}$ the net actual imported ECs of port p in period t

$F_p(.)$ The cumulative distribution function for $\eta_{p,t}^O$

$\varphi_{p,t}$ $= \sum_{l\in P,l\neq p} \varepsilon_{l,p,t} - \eta_{p,t}^O$ the net imported laden containers of port p in period t

$E_{p,t}$ The set of ports, whose net actual imported ECs are changed by perturbing the estimated supply or demand of port p in period t

$Q_{p,t}$ The set of ports, whose beginning on-hand inventories in period t are affected by perturbing threshold of port p

$\pi_{p,t}$ The corresponding dual variable for port p constraint in the transportation model in period t

q_t^N The port whose beginning on-hand inventory in period t is affected by perturbing the fleet size

$I\{.\}$ An indicator function, which takes 1 if the condition in the brace is true and otherwise 0

In every period t, the ECR decisions are firstly made at the beginning of this period. Then, the inventory position can be obtain by

$$y_{p,t} = x_{p,t} + a_{p,t}, \forall p \in P \tag{3.1}$$

After customer demands are realized and the laden containers become available, the beginning on-hand inventory in the next period can be updated by

$$x_{p,t+1} = y_{p,t} + \varphi_{p,t}, \forall p \in P \tag{3.2}$$

It should be noted that $x_{p,t}$ can be negative. This is due to the fact that customer demands are random and beyond control. If $x_{p,t}$ is negative, it represents the number of containers that are leased from port p and stored at other ports. In this situation, there are no ECs stored at port p; otherwise, they will be returned to vendors to reduce the number of leased containers according to the assumption. If $x_{p,t}$ is positive, it represents the number of ECs that are available at port p, which implies that there are no container leased out from vendors at this port. Note that there are N owned ECs in the system, we have

$$N = \sum_{p \in P} x_{p,t} \ \ \forall t \tag{3.3}$$

Let $J(\mathbf{x}_t, \boldsymbol{\gamma}, \omega_t)$ be the total cost in period t. It can be defined as:

$$J(\mathbf{x}_t, \boldsymbol{\gamma}, \omega_t) = H(\mathbf{Z}_t) + G(\mathbf{y}_t, \omega_t) \tag{3.4}$$

where the value of $\mathbf{Z}_t$ is determined by the beginning on-hand inventory $\mathbf{x}_t$ and the policy $\boldsymbol{\gamma}$, about which the details will be explained in the next section; $H(\mathbf{Z}_t)$ and $G(\mathbf{y}_t, \omega_t)$ are the EC repositioning cost and the EC holding and leasing cost in period t, respectively. They are defined as follow

$$H(\mathbf{Z}_t) = \sum_{p \in P} \sum_{m \in P, m \neq p} C_{p,m}^R \cdot z_{p,m,t} \tag{3.5}$$

$$G(\mathbf{y}_t, \omega_t) = \sum_{p \in P} \mathrm{g}\left(y_{p,t}, \eta_{p,t}^O\right) = \sum_{p \in P} \left(C_p^H \cdot \left(y_{p,t} - \eta_{p,t}^O\right)^+ + C_p^L \cdot \left(\eta_{p,t}^O - y_{p,t}\right)^+\right) \tag{3.6}$$

where $\mathrm{g}\left(y_{p,t}, \eta_{p,t}^O\right)$ represents the EC holding and leasing cost of port p in period t; $x^+ = \max\,(0, x)$. More specifically, the EC repositioning cost $H(\mathbf{Z}_t)$ refers to the cost of repositioning ECs between multiple ports. The EC holding and leasing cost $G(\mathbf{y}_t, \omega_t)$ is the cost incurred when ECs are stored at some ports and additional ECs are leased from vendors at the other ports.

Next, the single-level threshold policy is developed to make the ECR decisions at the beginning of period t.

3.1.3. A single-level threshold policy

Note that when we make the ECR decisions in a period, customer demands in this period have not been realized yet. Hence, our policy tries to maintain the inventory position at a target threshold value $\boldsymbol{\gamma}$. More specifically, γ_p is the target threshold level of port p. In period t, if the beginning on-hand inventory $x_{p,t}$ is larger than γ_p, i.e., $x_{p,t} > \gamma_p$, then port p is a surplus port and the quantity excess of γ_p should be repositioned out to other ports, which may need ECs, to try to bring the inventory position down to γ_p; if $x_{p,t} < \gamma_p$, then it is a deficit port and ECs should be repositioned in from other ports, which may supply ECs, to try to bring the inventory position up to γ_p; if $x_{p,t} = \gamma_p$, then it is a balanced port and nothing is done.

From the policy, therefore, if there are no surplus or deficit ports in period t, no ECs should be repositioned. Otherwise, ECs should be repositioned from surplus ports to deficit ports in the right quantity at the least movement. Without loss of generality, we consider the ECR decisions in period t. The two subsets, i.e., surplus port subset and deficit port subset can be obtained as $P_t^S = \{i: x_{i,t} > \gamma_i\}$ and $P_t^D = \{j: x_{j,t} < \gamma_j\}$, respectively. For a surplus port, the number of excess ECs, namely the number of estimated EC supply is calculated by Eq. (3.7); and for a deficit port, its number of estimated EC demand by Eq. (3.8).

$$u_{i,t}^S = x_{i,t} - \gamma_i, \forall i \in P_t^S \tag{3.7}$$

$$u_{j,t}^D = \gamma_j - x_{j,t}, \forall j \in P_t^D \tag{3.8}$$

If either P_t^S or P_t^D are empty, we have $\mathbf{Z}_t = 0$. Otherwise, we get that $z_{l,m,t} = 0, \forall l \in (P - P_t^S), m \in (P - P_t^D), l \neq m$ and the value of $z_{i,j,t}, \forall i \in P_t^S, j \in P_t^D$ are determined by solving a transportation model.

Note that when P_t^S and P_t^D are non-empty, the total number of estimated EC supplies $\sum_{i \in P_t^S} u_{i,t}^S$ could be not equal to the total number of estimated EC demands $\sum_{j \in P_t^D} u_{j,t}^D$. Thus, we propose to move as many excess ECs of surplus ports as possible to the deficit ports to satisfy their demands. Hence, when $\sum_{i \in P_t^S} u_{i,t}^S \geq \sum_{j \in P_t^D} u_{j,t}^D$, a transportation model is formulated as follows to determine the repositioned quantities in each period:

$$\min_{z_{i,j,t}} \sum_{i\in P_t^S} \sum_{j\in P_t^D} C_{i,j}^R \cdot z_{i,j,t} \tag{3.9}$$

$$\text{s.t.}\ \sum_{j\in P_t^D} z_{i,j,t} \le u_{i,t}^S,\ \forall i \in P_t^S \tag{3.10}$$

$$\sum_{i\in P_t^S} z_{i,j,t} \ge u_{j,t}^D,\ \forall j \in P_t^D \tag{3.11}$$

$$z_{i,j,t} \ge 0, \forall i \in P_t^S, j \in P_t^D \tag{3.12}$$

Constraints (3.10) ensure that the repositioned out ECs of a surplus port should not exceed its estimated EC supply; constraints (3.11) ensure that the EC demand of a deficit port can be fully satisfied; constraints (3.12) are the non-negative quantity constraints. When $\sum_{i\in P_t^S} u_{i,t}^S < \sum_{j\in P_t^D} u_{j,t}^D$, a similar model is formulated by substituting $\sum_{j\in P_t^D} z_{i,j.t} \ge u_{i,t}^S, \forall i \in P_t^S$ and $\sum_{i\in P_t^S} z_{i,j.t} \le u_{j,t}^D, \forall j \in P_t^D$ for inequations (3.10) and (3.11), respectively.

3.1.4. The optimization problem

Let $J(N, \boldsymbol{\gamma})$ be the expected total cost per period with the fleet size N and policy $\boldsymbol{\gamma}$, when the system is in steady state. The problem, which is to optimize the fleet size and the parameters of the policy in terms of minimizing the expected total cost per period, can be formulated as

$$\min_{N,\boldsymbol{\gamma}} J(N, \boldsymbol{\gamma}) \tag{3.13}$$

subject to the inventory dynamic Eqs. (3.1)~(3.3) and the single-level threshold policy. As in many papers on EC movement (see, e.g., Moon *et al.* (2010) and Long *et al.* (2012)), the variables that relate to the flow of ECs are considered as continuous variables in this study. That is, the fleet size and the parameters of the policy are considered as real numbers. It is fine when the values of these variables are large.

In general, it is difficult to solve problem (3.13), since there is no closed-form formulation for the computation of $J(N, \boldsymbol{\gamma})$ involving the repositioned empty container quantities determined by the transportation models. Next, we analyze the optimization problem and then provide two approaches to solve the problem.

3.2 Analysis of the optimization problem

When the transportation model is balanced in period t, i.e., the total number of estimated EC supplies, namely $\sum_{i\in P_t^S} u_{i,t}^S$, is equal to the total number of estimated EC demands, namely $\sum_{j\in P_t^D} u_{j,t}^D$, the excess ECs in the surplus ports can be fully repositioned out to satisfy the demands of the deficit ports, so that the inventory position of each port can be kept at its target threshold level in this period, i.e., $y_{p,t} = \gamma_p$. Further, the repositioning cost in next period will be only related to customer demand in this period. From Eqs. (3.3), (3.7) and (3.8), we obtain that $\sum_{i\in P_t^S} u_{i,t}^S = \sum_{j\in P_t^D} u_{j,t}^D$ $\forall t$ if and only if $N = \sum_{p\in P} \gamma_p$. Hence, we have an important property of the problem as follows:

PROPERTY 3.1: *When the fleet size is equal to the sum of thresholds, the inventory position can be always maintained at a target threshold value and then the empty container repositioning cost in a period is only dependent on the customer demand.*

Let Scenario-I (Scenario-II) represent the scenario in which $N = \sum_{p\in P} \gamma_p$ ($\neq \sum_{p\in P} \gamma_p$). Taking advantage of this property, we propose two approaches to tackle problem (3.13) under both scenarios, respectively.

3.2.1. Scenario-I

Consider the problem under Scenario-I. From Property 3.1, it implies that $y_{p,t} \equiv \gamma_p \ \forall p \in P, t$, and only the EC holding and leasing cost in a period is related to values of N and $\boldsymbol{\gamma}$. Hence, the optimal solution which minimizes $J(N, \boldsymbol{\gamma})$ should be equivalent to the optimal solution which minimizes the expected EC holding and leasing cost per period. From (3.6), since that holding and leasing cost function for each port is independent, the problem (3.13) under Scenario-I can be reduced to a newsvendor problem, which is formulated as a Non-Linear Programming (NLP):

$$\min_{\boldsymbol{N},\boldsymbol{\gamma}} \sum_{p\in P} E\left(C_p^H \cdot \left(\gamma_p - \eta_p^O\right)^+ + C_p^L \cdot \left(\eta_p^O - \gamma_p\right)^+\right)$$

$$\text{s.t. } N = \sum_{p\in P} \gamma_p$$

where the subscript t in the notation of $\eta_{p,t}^O$ is dropped since customer demands in each period are independent and identically distributed. This newsvendor problem can be further decomposed into P separate sub-newsvendor problems with the optimal solutions as follows:

$$\gamma_p^* = F_p^{-1}\left(C_p^L / \left(C_p^L + C_p^H\right)\right) \tag{3.14}$$

$$N^* = \sum_{p\in P} \gamma_p^* \tag{3.15}$$

where $F_p^{-1}(.)$ is the inverse function of $F_p(.)$.

3.2.2. Scenario-II

The problem under Scenario-II is more complex than that under Scenario-I, since the EC repositioning cost is also affected by the values of N and $\boldsymbol{\gamma}$. Consider that the minimum expected holding and leasing cost of problem (3.13) can only be achieved under Scenario-I, since only under Scenario-I, the inventory position of each port can be kept at its optimal threshold level. The problem under Scenario-II is worth to study if and only if the minimum expected EC repositioning cost under this scenario could be less than that under Scenario-I.

There is no closed-form formulation for the expected EC repositioning cost. Therefore, without loss of generality, we mathematically compare the EC repositioning costs in a period under both scenarios with the same customer demands. More specifically, we have Scenario-I with parameters $(N, \boldsymbol{\gamma}^{\mathrm{I}})$ and Scenario-II with parameters $(N, \boldsymbol{\gamma}^{\mathrm{II}})$ where $N = \sum_{p\in P} \gamma_p^{\mathrm{I}} \neq \sum_{p\in P} \gamma_p^{\mathrm{II}}$. Let other various quantities in both scenarios be distinguished by displaying the arguments $.^{\mathrm{II}}$ and $.^{\mathrm{I}}$, respectively.

From property 3.1, we can get that under Scenario-I, we have $\mathbf{y}_t^{\mathrm{I}} = \boldsymbol{\gamma}^{\mathrm{I}}, \forall t$. Similarly, under Scenario-II, we can have $\mathbf{y}_t^{\mathrm{II}} \geq \boldsymbol{\gamma}^{\mathrm{II}}$ $(\leq \boldsymbol{\gamma}^{\mathrm{II}})$ if N is $> \sum_{p\in P} \gamma_p^{\mathrm{II}}$ $(< \sum_{p\in P} \gamma_p^{\mathrm{II}})$. We know that when N is $> \sum_{p\in P} \gamma_p^{\mathrm{II}}$ $(< \sum_{p\in P} \gamma_p^{\mathrm{II}})$, the transportation model in each period should be unbalanced, i.e., the total number of estimated EC supplies, namely $\sum_{i\in P_t^S} u_{i,t}^S$, is greater (less) than the total number of estimated EC demands, namely $\sum_{j\in P_t^D} u_{j,t}^D$. This unbalanced situation, such as $\sum_{i\in P_t^S} u_{i,t}^S > \sum_{j\in P_t^D} u_{j,t}^D$, means that all EC demands at the deficit ports can be satisfied and some excess ECs still stay at some surplus ports. Hence, the inventory position of each port in each period could be not less (greater) than its target threshold level when N is $> \sum_{p\in P} \gamma_p^{\mathrm{II}}$ $(< \sum_{p\in P} \gamma_p^{\mathrm{II}})$.

Note that when there are EC movements in a period, the EC repositioning cost will be the minimum objective value of the transportation model. Hence, we first investigate the constraints of ports under both scenarios and have following lemmas.

LEMMA 3.1: *For all period* t, *we have* $P_{t+1}^{S,\mathrm{I}} \subseteq P_{t+1}^{S,\mathrm{II}}$ *and* $P_{t+1}^{D,\mathrm{II}} \subseteq P_{t+1}^{D,\mathrm{I}}$ *if* $N > \sum_{p\in P} \gamma_p^{\mathrm{II}}$; $P_{t+1}^{S,\mathrm{II}} \subseteq P_{t+1}^{S,\mathrm{I}}$ *and* $P_{t+1}^{D,\mathrm{I}} \subseteq P_{t+1}^{D,\mathrm{II}}$, *if* $N < \sum_{p\in P} \gamma_p^{\mathrm{II}}$ *with same customer demand* ω_t.

PROOF: With the help of Eq. (3.2), we have $x_{p,t+1}^{\mathrm{I}} = \gamma_p^{\mathrm{I}} + \varphi_{p,t}, \forall p \in P$. It implies that $\varphi_{i,t} > 0, \forall i \in P_{t+1}^{S,\mathrm{I}}$ and $\varphi_{j,t} < 0, \forall j \in P_{t+1}^{D,\mathrm{I}}$.

Note that we investigate both scenarios with same ω_t. We have $x_{i,t+1}^{\mathrm{II}} > y_{i,t}^{\mathrm{II}} \geq \gamma_i^{\mathrm{II}}, \forall i \in P_{t+1}^{S,\mathrm{I}}$ if $N > \sum_{p\in P} \gamma_p^{\mathrm{II}}$, and $x_{j,t+1}^{\mathrm{II}} < y_{j,t}^{\mathrm{II}} \leq \gamma_j^{\mathrm{II}}, \forall j \in P_{t+1}^{D,\mathrm{I}}$ if $N < \sum_{p\in P} \gamma_p^{\mathrm{II}}$. Hence, we can obtain that $P_{t+1}^{S,\mathrm{I}} \subseteq P_{t+1}^{S,\mathrm{II}}$ if $N > \sum_{p\in P} \gamma_p^{\mathrm{II}}$, and $P_{t+1}^{D,\mathrm{I}} \subseteq P_{t+1}^{D,\mathrm{II}}$ if $N < \sum_{p\in P} \gamma_p^{\mathrm{II}}$.

Considering that $P_t^S \cup P_t^D = P$ for any scenario, we have $P_{t+1}^{D,\mathrm{II}} \subseteq P_{t+1}^{D,\mathrm{I}}$ if $N > \sum_{p\in P} \gamma_p^{\mathrm{II}}$ and $P_{t+1}^{S,\mathrm{II}} \subseteq P_{t+1}^{S,\mathrm{I}}$ when $N < \sum_{p\in P} \gamma_p^{\mathrm{II}}$. ■

From lemma 3.1, there are only two possible cases about the port subsets under both scenarios. One is that there are same port subsets. The other is that there are greater (smaller) surplus port subset and smaller (greater) deficit port subset under Scenario-II than that under Scenario-I when the fleet size is greater (less) than the sum of thresholds. Furthermore, for the port that is common to surplus or deficit port subsets under both scenarios, we investigate its quantity of excess or deficit ECs.

LEMMA 3.2: *For all period* t, *we have*

1) *If* $N > \sum_{p\in P} \gamma_p^{\mathrm{II}}$, *for* $i \in P_{t+1}^{S,\mathrm{I}}$ *and* $j \in P_{t+1}^{D,\mathrm{II}}$ ($P_{t+1}^{D,\mathrm{II}} \neq \emptyset$), *we have* $u_{i,t+1}^{S,\mathrm{II}} \geq u_{i,t+1}^{S,\mathrm{I}}$ *and* $u_{j,t+1}^{D,\mathrm{II}} \leq u_{j,t+1}^{D,\mathrm{I}}$.

2) *If* $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$, *for* $i \in P_{t+1}^{S,\mathrm{II}}$ ($P_{t+1}^{S,\mathrm{II}} \neq \emptyset$) *and* $j \in P_{t+1}^{D,\mathrm{I}}$, *we have* $u_{i,t+1}^{S,\mathrm{II}} \leq u_{i,t+1}^{S,\mathrm{I}}$ *and* $u_{j,t+1}^{D,\mathrm{II}} \geq u_{j,t+1}^{D,\mathrm{I}}$.

with same customer demand ω_t.

PROOF: From Eqs. (3.2), (3.7) and (3.8), we get that $u_{i,t+1}^{S} = y_{i,t} + \varphi_{i,t} - \gamma_i$ and $u_{j,t+1}^{D} = \gamma_j - y_{j,t} - \varphi_{j,t}, \forall i \in P_{t+1}^{S}, j \in P_{t+1}^{D}$. Hence, with the help of lemma 3.1, when $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$, we have $u_{i,t+1}^{S,\mathrm{II}} \geq \varphi_{i,t} = u_{i,t+1}^{S,\mathrm{I}}, \forall i \in P_{t+1}^{S,\mathrm{I}}$, and if $P_{t+1}^{D,\mathrm{II}}$ is non-empty, we have $u_{j,t+1}^{D,\mathrm{II}} \leq -\varphi_{j,t} = u_{j,t+1}^{D,\mathrm{I}}, \forall j \in P_{t+1}^{D,\mathrm{II}}$. Similarly, when $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$, we have $u_{i,t+1}^{S,\mathrm{II}} \leq u_{i,t+1}^{S,\mathrm{I}}, \forall i \in P_{t+1}^{S,\mathrm{II}}$ if $P_{t+1}^{S,\mathrm{II}} \neq \emptyset$, and $u_{j,t+1}^{D,\mathrm{II}} \geq u_{j,t+1}^{D,\mathrm{I}}, \forall j \in P_{t+1}^{D,\mathrm{I}}$. ■

From lemma 3.2, it implies that for example, when the fleet size is greater than the sum of thresholds, for the common surplus port under both scenarios, its number of excess empty containers under Scenario-II is not less than that under scenario-I. We next present the proposition about the repositioning costs under both scenarios.

PROPOSITION 3.1: *For all* t, *with same customer demand* ω_t, *we have*

$$H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})\big|_{\omega_t} \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})\big|_{\omega_t}$$

PROOF: Since $P_t^{S,\mathrm{I}}$ and $P_t^{D,\mathrm{I}}$ are both nonempty, from the transportation model, we have $H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t} > 0$. Hence, when either $P_{t+1}^{D,\mathrm{II}}$ or $P_{t+1}^{S,\mathrm{II}}$ is empty, we can get that $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} = 0 < H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$.

When $P_{t+1}^{D,\mathrm{II}}$ is nonempty in the case of $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$, the constraints of the transportation models in period $t+1$ of both scenarios should be inequations (3.10) and (3.11), which are as follows:

$$\sum_{j \in P_{t+1}^{D}} z_{i,j,t+1} \leq u_{i,t+1}^{S}, \forall i \in P_{t+1}^{S}$$

$$-\sum_{i \in P_{t+1}^{S}} z_{i,j,t+1} \leq -u_{j,t+1}^{D}, \forall j \in P_{t+1}^{D}$$

Based on the sensitivity analysis about right-hand side (RHS) of linear programming (LP), we get that increasing the RHS coefficient value of the above constraint could reduce the minimum objective value. Depending on the port subsets under both scenarios, from lemmas 3.1 and 3.2, there are only two possible cases:

- Case 1: $P_{t+1}^{S,\mathrm{II}} = P_{t+1}^{S,\mathrm{I}}$ and $P_{t+1}^{D,\mathrm{II}} = P_{t+1}^{D,\mathrm{I}}$. It implies that the transportation models under both scenarios have similar constraints for all ports $i \in P_{t+1}^{S,\mathrm{I}}, j \in P_{t+1}^{D,\mathrm{I}}$. Hence, based on the model under Scenario-I, we increase the RHS values from $u_{i,t+1}^{S,\mathrm{I}}$ and $-u_{j,t+1}^{D,\mathrm{I}}$ up to $u_{i,t+1}^{S,\mathrm{II}}$ and $-u_{j,t+1}^{D,\mathrm{II}}$, respectively. Then, the new model will be equivalent to the model under Scenario-II, and we can have $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$.

- Case 2: $P_{t+1}^{S,\mathrm{II}} \supset P_{t+1}^{S,\mathrm{I}}$ and $P_{t+1}^{D,\mathrm{II}} \subset P_{t+1}^{D,\mathrm{I}}$. It implies that the transportation models under both scenarios have similar constraints for some ports $i \in P_{t+1}^{S,\mathrm{I}}, j \in P_{t+1}^{D,\mathrm{II}}$. Similarly, based on the model under the Scenario-I (Scenario-II), we increase (decrease) the RHS values of ports that are $\in \left(P_{t+1}^{D,\mathrm{I}} - P_{t+1}^{D,\mathrm{II}}\right)$ ($\left(P_{t+1}^{S,\mathrm{II}} - P_{t+1}^{S,\mathrm{I}}\right)$) up (down) to zero. Then, the minimum objective cost of the new model-I (model-II), denoted by C' (C''), should be not greater (less) than that of the original model, i.e., we have $C' \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$ ($C'' \geq H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t}$). For the new model-I and model-II, they have similar constraints for ports $i \in P_{t+1}^{S,\mathrm{I}}, j \in P_{t+1}^{D,\mathrm{II}}$. Hence, as that in case 1, we have $C'' \leq C'$. Finally, we get that $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} \leq C'' \leq C' \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$.

Similarly, when $P_{t+1}^{S,\mathrm{II}} \neq \emptyset$ in the case of $\sum_{p \in P} \gamma_p^{\mathrm{II}} > N$, we can get that $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$. ■

From this proposition, it implies that the expected EC repositioning cost per period under Scenario-II could be less than that under Scenario-I. Therefore, it is worth to study the problem under Scenario-II. Note that tackling problem (3.13) under

Scenario-II is a non-trivial task. Simulation optimization technique could be a useful tool to solve our problem (Tekin and Sabuncuoglu, 2004). Hence, we adopt the simulation technique to estimate $J(N, \boldsymbol{\gamma})$ with given values of N and $\boldsymbol{\gamma}$ as follows:

$$J(N, \boldsymbol{\gamma}) \approx \frac{1}{T}\sum_{t=1}^{T} J(\mathbf{x}_t, \boldsymbol{\gamma}, \omega_t) = \frac{1}{T}\sum_{t=1}^{T}\left(H(\mathbf{x}_t, \boldsymbol{\gamma}) + G(\mathbf{y}_t, \omega_t)\right) \tag{3.16}$$

where T is the number of the simulation periods. Then, our problem becomes the optimization via simulation problem. The solution method is presented in the next section.

3.3 IPA-based gradient algorithm

Due to the rapid growth of computing power and the burst of the applications of optimization via simulation, many search algorithms have been designed in the past ten years to solve the optimization via simulation problems (see Fu (2002) and Fu *et al.* (2005) for a comprehensive review). One category of well-known search algorithm is driven by gradient information, such as the steepest descent algorithm (Debye, 1909) and stochastic approximation algorithm (Robbins and Monro, 1951). Such algorithms seek the next solution along the steepest descent (negative gradient) direction at the current solution resulting in fast convergence speed. Under certain conditions, a gradient-driven algorithm converges to a local optimal point almost surely (Spall, 2003).

The success of a gradient-driven algorithm relies heavily on the quality of gradient estimate, especially in the stochastic problems. If the inside structure of the simulation model is known, direct gradients could be obtained to enhance the effectiveness of the gradient algorithm. Among many approaches for estimating the gradient (see, e.g., Fu and Hu (1997) , Ho and Cao (1991) and Fu (2008)) , IPA technique is one of the best-known gradient estimation techniques, which is designed to obtain derivatives of continuous parameters and to estimate all partial derivatives from a single run by keeping track of related statistics of certain events during a run. It is able to estimate the gradient of the objective function from one single simulation run, thus to reduce the computational time (Ho and Cao, 1991). Moreover, it has been shown that variance of IPA estimator is lower, compared with many other gradient estimators (Suri, 1989). Hence, IPA technique has been successfully applied to many problems, such as a two-echelon assembled-to-stock problem (Chew *et al.*, 2009), stochastic flow models (Chen *et al.*, 2010; Wardi *et al.*, 2010), persistent monitoring problems (Cassandras *et al.*, 2011), budget allocation for effective data collection (Wong *et al.*, 2011), multi-location transshipment problem with capacitated production (Özdemir *et al.*, In press), etc.

In this study, taking advantage of the inside structure of our simulation model, we propose an IPA-based gradient algorithm to tackle problem (3.13) under Scenario-II. The overall gradient technique is briefly described in Figure 3.1. For a given $(N, \boldsymbol{\gamma})$ at iteration n, we run simulation to estimate the expected cost $J(N, \boldsymbol{\gamma})$ in (3.16). At the same time, we can also estimate the gradient vector of the expected cost $\nabla J(N, \boldsymbol{\gamma})_{(N,\boldsymbol{\gamma})=(N,\boldsymbol{\gamma})_n}$. Briefly, the simulation is run by firstly making the ECR decision in a given period by solving transportation models and then customer demands are realized. At the same time, the cost and the gradient in this period can also be computed. After all periods are run, the overall cost and gradient can be computed from the individual values obtained in each period. Based on the gradient information, we update parameters by using the steepest descent algorithm, i.e., $(N, \boldsymbol{\gamma})_{n+1} = (N, \boldsymbol{\gamma})_n - \alpha_n\left(\nabla J(N, \boldsymbol{\gamma})_{(N,\boldsymbol{\gamma})=(N,\boldsymbol{\gamma})_n}\right)$. α_n is the step size at the n-th iteration.

After $(N, \boldsymbol{\gamma})$ is updated, we continue the same process for new iteration until the stopping criterion is met.

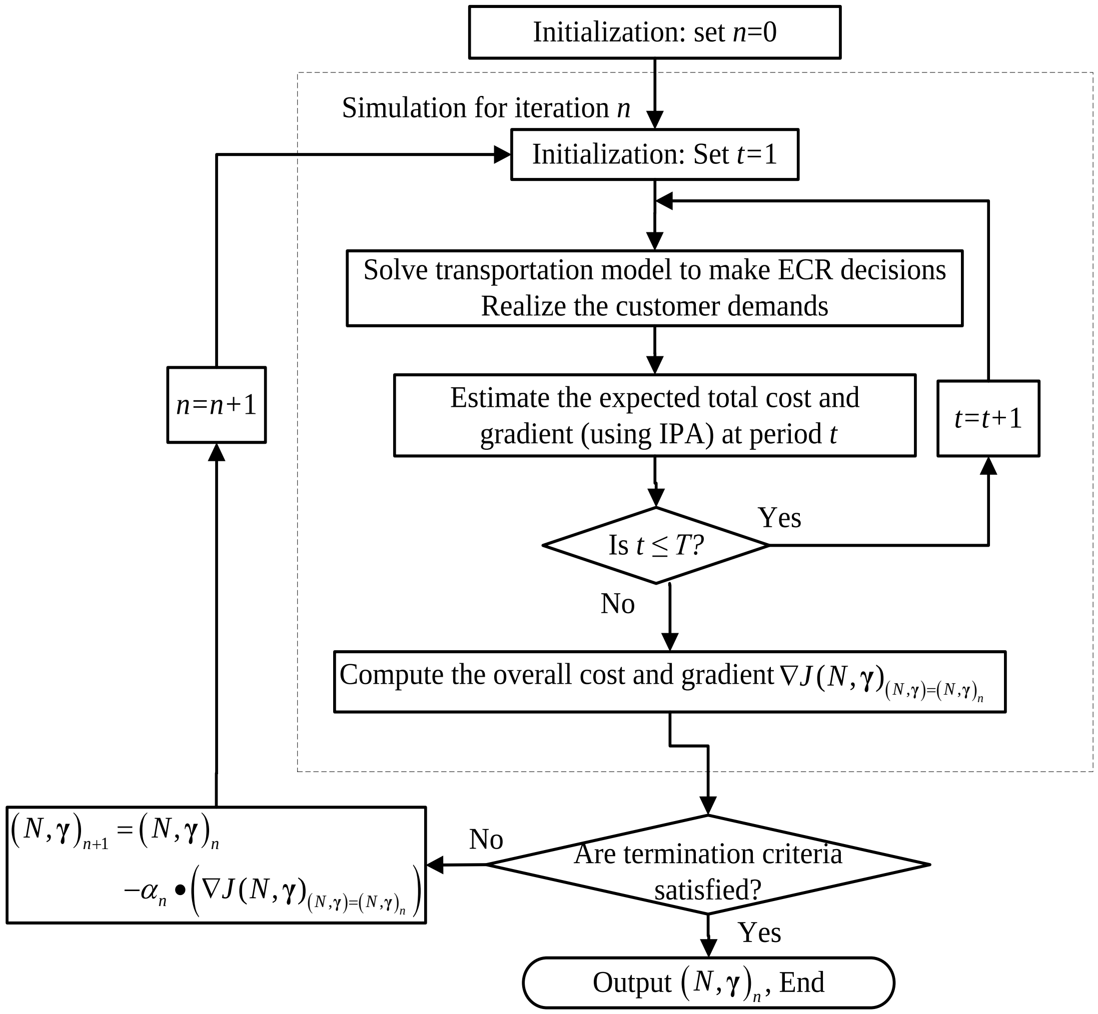


Figure 3.1. The flow of the IPA-based gradient technique

To estimate the gradient of expected cost, we take a partial derivation of (3.16) with respect to the fleet size and the threshold of port i, respectively and have

$$\frac{\partial J(N, \boldsymbol{\gamma})}{\partial N} \approx \frac{1}{T} \cdot \sum\nolimits_{t=1}^{T} \frac{\partial J(\mathbf{x}_t, \boldsymbol{\gamma}, \omega_t)}{\partial N}$$

$$\approx \frac{1}{T} \cdot \sum\nolimits_{t=1}^{T} \left(\frac{\partial H(\mathbf{Z}_t)}{\partial N} + \sum\nolimits_{p \in P} \frac{\partial E\left(\mathrm{g}\left(y_{p,t}, \eta_{p,t}^{O}\right)\right)}{\partial y_{p,t}} \cdot \frac{\partial y_{p,t}}{\partial N} \right) \tag{3.17}$$

$$\frac{\partial J(N,\boldsymbol{\gamma})}{\partial \gamma_i} \approx \frac{1}{T}\cdot\sum_{t=1}^{T}\frac{\partial J(\mathbf{x}_t,\boldsymbol{\gamma},\omega_t)}{\partial \gamma_i}$$
$$\approx \frac{1}{T}\cdot\sum_{t=1}^{T}\left(\frac{\partial H(\mathbf{Z}_t)}{\partial \gamma_i}+\sum_{p\in P}\frac{\partial E\left(\mathrm{g}\left(y_{p,t},\eta_{p,t}^{O}\right)\right)}{\partial y_{p,t}}\cdot\frac{\partial y_{p,t}}{\partial \gamma_i}\right) \tag{3.18}$$

Here, for the EC holding and leasing cost function, we use the expected cost function to estimate the gradient instead of sample cost function since we are able to get the explicit function to evaluate the average gradient. $\partial H(\mathbf{Z}_t)/\partial N$or $\partial H(\mathbf{Z}_t)/\partial \gamma_i$, which measures the impact of the transportation cost in period t when the fleet size or threshold of port i is changed, can be found using the dual model information from the transportation model. $\partial \mathrm{g}(y_{p,t},\eta_{p,t}^{O})/\partial y_{p,t}$, which measures the impact of the holding and leasing cost function of port p in period t when the inventory position is changed, can be easily found taking the derivation of Eq. (3.6) with respect to the inventory position level. $\partial y_{p,t}/\partial N$ or $\partial y_{p,t}/\partial \gamma_i$, which measures the impact of the inventory position level of port p in period t when the fleet size or threshold of port i change, can be estimated using the IPA technique.

Next, we analyze the gradient of expected total cost with respect to the threshold, followed by that to the fleet size.

3.3.1. Gradient with respect to the threshold

In this section, we study the gradient of expected total cost with respect to the threshold with given fleet size N.

The nominal path is defined as the sample path generated by the simulation model with parameter $\boldsymbol{\gamma}$ and the perturbed path as the sample path generated using the same model and same random seeds, but with parameter $(\boldsymbol{\gamma})'$, where $(\boldsymbol{\gamma})' = \boldsymbol{\gamma} + \Delta\boldsymbol{\gamma}$. Without loss of generality, we only perturb the threshold of port i and keep the thresholds of the other ports unchanged, i.e., $(\gamma_i)' = \gamma_i + \Delta\gamma_i$ and $(\gamma_p)' = \gamma_p$ for other $p \in (P - \{i\})$, where the value of $\Delta\gamma_i$ is sufficiently small. By "sufficiently" small, we mean such that the surplus port and deficit port subsets are same in the both paths in every period. That is, if port p is a surplus port in period t in the nominal path, it will still be a surplus port in the perturbed path. Oftentimes, we will present the changes in various quantities by displaying with argument Δ. For example, $\Delta\mathbf{x}_t$ represents the change in the beginning on-hand inventory in period t.

We perturb γ_i with the same quantities in all periods and the representative perturbation flow in period t is shown in Figure 3.2. The perturbation of $\Delta\mathbf{x}_t$ and the perturbation of $\Delta\gamma_i$ will affect the estimated EC supplies or demands, i.e., $u_{p,t}^S$ or $u_{p,t}^D$ of some ports, which are the RHS of the constraints in the transportation model. Hence, the perturbation of $\Delta u_{p,t}^S$ or $\Delta u_{p,t}^D$ could change the repositioning cost and the net actual imported ECs, i.e., $a_{p,t}$ of some ports. The perturbations of $\Delta a_{p,t}$ and $\Delta\mathbf{x}_t$ will affect the perturbation on the EC inventory positions $y_{p,t}$ of some ports. Further, perturbation of $\Delta y_{p,t}$ will affect the holding and leasing cost in this period and the beginning on-hand inventory of next period.

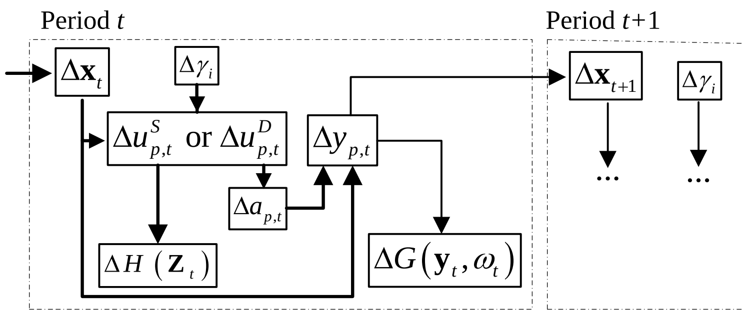


Figure 3.2 The perturbation flow

From the definition of $Q_{i,t}$, we have $Q_{i,t} = \{p: \Delta x_{p,t} \neq 0\}$. Thus, depending on the status of $Q_{i,t}$, only two scenarios are possible – one with $Q_{i,t} = \emptyset$, the other with $Q_{i,t} \neq \emptyset$. Since the value of $\mathbf{x}_1$ is given, it implies that $\Delta \mathbf{x}_1 = 0$ and thus $Q_{i,1} = \emptyset$. If the perturbation in the first period is propagated to the second period, it could lead $Q_{i,2} \neq \emptyset$. Hence, analysis on the perturbations in the first two periods could be representative of the general two scenarios. Consequently, we investigate the perturbations in the first two periods to conclude the general formulations for the perturbation terms in (3.18).

3.3.1.1. Perturbation with respect to the threshold in the first period

We trace the perturbations in the period $t = 1$ following the flow in Figure 3.2. Since $\Delta \mathbf{x}_t = 0$, we have $\Delta u_{i,t}^S = -\Delta \gamma_i$ or $\Delta u_{i,t}^D = \Delta \gamma_i$, and $\Delta u_{p,t}^S$ or $\Delta u_{p,t}^D = 0$ for other $p \in P$ from Eqs. (3.7) and (3.8). It implies that only the RHS of port i constraint is changed.

From the transportation model, the perturbation of $\Delta u_{i,t}^S$ or $\Delta u_{i,t}^D$ will affect the perturbation of $\Delta a_{p,t}$, depending on the status of port i constraint at the optimal solution. Only two scenarios are possible:

- Scenario 1: The port i constraint is non-binding. The perturbation of $\Delta u_{i,t}^S$ or $\Delta u_{i,t}^D$ will not affect the optimal repositioned quantities. Hence, it implies that $\Delta a_{p,t} = 0$ for all $p \in P$.
- Scenario 2: The port i constraint is binding. The perturbation could affect the optimal repositioned quantities of some ports. However, only for a pair of ports, i.e., i and k, the net actual imported empty containers will be changed (explained in section 3.3.2); while for the other ports, no changes. It implies that $\Delta a_{p,t} \neq 0$ for any $p = i, k$, and $\Delta a_{p,t} = 0$ for other $p \in (P - \{i, k\})$.

Note that the ECR decisions are made by solving a transportation model (in Section 3.1.3). If an inequality constraint holds with equality at the optimal solution, the constraint is said to be *binding*, as the solution cannot be varied in the direction of the constraint. Otherwise, if an inequality constraint holds as a strict inequality at the optimal solution (that is, does not hold with equality), the constraint is said to be *non-binding*, as the point could be varied in the direction of the constraint.

From the definition of $E_{i,t}$, we get that $E_{i,t} = \{p: \Delta a_{p,t} \neq 0\}$. Thus, if the port i constraint is non-binding, we have $E_{i,t} = \phi$. Otherwise, we have $E_{i,t} = \{i, e_{i,t}\}$ with $e_{i,t} = k$. It implies that $e_{i,t}$ is unique in period t. To find set $E_{i,t}$, a modified stepping stone approach (elaborated in section 3.3.2) is proposed.

Without investigating the details on the perturbation of $\Delta a_{p,t}$, we next consider the perturbation on $y_{p,t}$. From Eq. (3.1), we have $\Delta y_{p,t} = \Delta a_{p,t}, \forall p \in P$ since $\Delta x_{p,t} = 0$, $\forall p \in P$. Hence, we can get that:

- If $E_{i,t} = \phi$, $\Delta y_{p,t} = 0$ for all $p \in P$.
- If $E_{i,t} \neq \phi$, $\Delta y_{p,t} \neq 0$ for any $p \in E_{i,t}$, and $\Delta y_{p,t} = 0$ for other $p \in (P - E_{i,t})$.

From the transportation model, for the port with binding constraint, its inventory position will be equal to the threshold. Hence, we have $\Delta y_{i,t} = \Delta \gamma_i$. Further, we can get that $\Delta y_{e_{i,t},t} = -\Delta \gamma_i$ since $\sum_{p \in P} \Delta y_{p,t} = \Delta N = 0$ from Eqs. (3.1) and (3.3).

We continue to study how the perturbations are carried forward to the next period. From Eq. (3.2), we have $\Delta x_{p,t+1} = \Delta y_{p,t}, \forall p \in P$. It implies that the perturbation on $x_{p,t+1}$ will be fully from the perturbation on $y_{p,t}$, and then we have $Q_{i,t+1} = \{p: \Delta y_{p,t} \neq 0\}$. From the analysis on the perturbation of $\Delta y_{p,t}$ in the case of $Q_{i,t} = \emptyset$, we have

$$Q_{i,t+1} = \begin{cases} \emptyset, & \text{if } E_{i,t} = \emptyset \\ \{i, q_{i,t+1}\}, & \text{otherwise} \end{cases} \tag{3.19}$$

where $q_{i,t+1} = e_{i,t}$ when $E_{i,t} \neq \emptyset$. And when $Q_{i,t+1} \neq \emptyset$, we have $\Delta x_{i,t+1} = \Delta\gamma_i$, $\Delta x_{q_{i,t+1},t+1} = -\Delta\gamma_i$ and $\Delta x_{p,t+1} = 0$ for other $p \in (P - Q_{i,t+1})$. It indicates that if there are perturbations propagated to the next period, only for a pair of ports, i.e., i and $q_{i,t+1}$, their beginning on-hand inventories will be changed by $\Delta\gamma_i$ and $-\Delta\gamma_i$, respectively; while for the other ports, no changes.

From the analysis on the perturbations of the relevant variables, we have next lemma. Note that the value of the EC repositioning cost $H(\mathbf{Z}_t)$ is equivalent to the value of the objective in the transportation model presented in Eq. (3.9).

LEMMA 3.3: *In the period* t=1, *we have*

1) $\partial H(\mathbf{Z}_t)/\partial\gamma_i = (-1)^{I(i \in P_t^S)} \cdot \pi_{i,t}$

2) $U_{p,t} = \begin{cases} -C_p^L, & \text{if } y_{p,t} < 0 \\ (C_p^H + C_p^L) \cdot F_p(y_{p,t}) - C_p^L, & \text{if } y_{p,t} \geq 0 \end{cases}, \forall p \in P$

3) $\frac{\partial y_{p,t}}{\partial\gamma_i} = \begin{cases} 1, & \text{if } I(Q_{i,t+1} \neq \emptyset, p = i) = 1 \\ -1, & \text{if } I(Q_{i,t+1} \neq \emptyset, p = q_{i,t+1}) = 1 \\ 0, & \text{otherwise} \end{cases}, \forall p \in P$

where $U_{p,t} = \partial E\left(\mathrm{g}(y_{p,t}, \eta_{p,t}^O)\right)/\partial y_{p,t}$.

PROOF: Since only the RHS of port i 's constraint is changed in the first period with $\Delta u_{i,t}^S = -\Delta\gamma_i$ or $\Delta u_{i,t}^D = \Delta\gamma_i$, the perturbation on the optimal repositioning cost value in the first period is estimated by $\Delta H(\mathbf{Z}_t) = (-1)^{I(i \in P_t^S)} \cdot \Delta\gamma_i \cdot \pi_{i,t}$ from the sensitivity analysis about RHS of LP. By hypothesis the value of $\Delta\gamma_i$ is infinitesimally small. The change in the RHS value is within its allowable range and the status of all constraints remains unchanged in the perturbed path. In our problem with real variables, the probability of having degenerate optimal solutions or multiple optimal solutions is close to 0. Hence, $\frac{\partial H(\mathbf{Z}_t)}{\partial \gamma_i} = \lim_{\Delta\gamma_i \to 0} \frac{\Delta H(\mathbf{Z}_t)}{\Delta \gamma_i} = (-1)^{I(i \in P_t^S)} \cdot \pi_{i,t}$.

For assertion 2), by taking the first derivation of the holding and leasing cost function of port p in Eq. (3.6) with respect to the inventory position level, it is easy to have the equations.

For assertion 3), recall that the perturbation on $y_{p,t}$ is equal to the perturbation on $x_{p,t+1}$ for all $p \in P$, i.e., $\Delta y_{p,t} = \Delta x_{p,t+1}$. We have $\frac{\partial y_{p,t}}{\partial \gamma_i} = \lim\limits_{\Delta\gamma_i \to 0} \frac{\Delta y_{p,t}}{\Delta \gamma_i} = \lim\limits_{\Delta\gamma_i \to 0} \frac{\Delta x_{p,t+1}}{\Delta \gamma_i}, \forall p \in P$. Hence, if $Q_{i,t+1} = \emptyset$, we have $\partial y_{p,t}/\partial \gamma_i = 0$ for all $p \in P$. Otherwise, we have $\partial y_{i,t}/\partial \gamma_i = 1$, $\partial y_{q_{i,t+1},t}/\partial \gamma_i = -1$, and $\partial y_{p,t}/\partial \gamma_i = 0$ for other $p \in (P - Q_{i,t+1})$. ■

3.3.1.2. Perturbation with respect to the threshold in the second period

If $Q_{i,2} = \emptyset$, the analysis about the perturbation in the second period will be similar with that in the first period. Hence, we study the other case with $Q_{i,t} = \{i, q_{i,t}\}$ for t=2.

Similarly, we analyze the perturbation on relevant variables. We have $\Delta u_{q_{i,t},t}^S = -\Delta\gamma_i$ or $\Delta u_{q_{i,t},t}^D = \Delta \gamma_i$ and $\Delta u_{p,t}^S$ or $\Delta u_{p,t}^D = 0$ for other $p \in (P - \{q_{i,t}\})$. It implies that only the RHS of port $q_{i,t}$ constraint is changed. Hence, if the port $q_{i,t}$ constraint is non-binding, we have $E_{q_{i,t},t} = \emptyset$ with $\Delta a_{p,t} = 0$ for all $p \in P$. Otherwise, we have $E_{q_{i,t},t} = \{q_{i,t}, e_{q_{i,t},t}\}$ with $\Delta a_{p,t} \neq 0$ for any $p \in E_{q_{i,t},t}$, and $\Delta a_{p,t} = 0$ for other $p \in (P - E_{q_{i,t},t})$.

Differently, the perturbation on $y_{p,t}$ could be from the both perturbations of $\Delta\mathbf{x}_t$ and $\Delta a_{p,t}$ in this period since $\Delta\mathbf{x}_t \neq 0$. Hence, from Eq. (3.1), i.e., $y_{p,t} = x_{p,t} + a_{p,t}$, we obtain:

- If $E_{q_{i,t},t} = \phi$, we have $\Delta y_{i,t} = \Delta\gamma_i$, $\Delta y_{q_{i,t},t} = -\Delta\gamma_i$, and $\Delta y_{p,t} = 0$ for other $p \in \left(P - \{i, q_{i,t}\}\right)$.
- If $E_{q_{i,t},t} \neq \phi$, we have $\Delta y_{p,t}$ could be $\neq 0$ for any $p = i, e_{q_{i,t},t}$, and $\Delta y_{p,t} = 0$ for other $p \neq i, e_{q_{i,t},t}$. Depending on whether $e_{q_{i,t},t} = i$, we get that:
    - If $e_{q_{i,t},t} \neq i$, then it implies that $\Delta y_{i,t} = \Delta x_{i,t} = \Delta\gamma_i$, and $\Delta y_{e_{q_{i,t},t},t} = -\Delta\gamma_i$.
    - If $e_{q_{i,t},t} = i$, then it implies that $\Delta y_{i,t} = 0$.

From the analysis about the perturbation of $\Delta y_{p,t}$ in the case of $Q_{i,t} \neq \emptyset$, we have:

$$Q_{i,t+1} = \begin{cases} \emptyset, & \text{if } E_{i,t} \neq \emptyset \text{ and } e_{q_{i,t},t} = i \\ \{i, q_{i,t+1}\}, & \text{otherwise} \end{cases} \tag{3.20}$$

where $q_{i,t+1}$ is $= q_{i,t}$ when $E_{q_{i,t},t} = \emptyset$, or $= e_{q_{i,t},t}$ when $E_{q_{i,t},t} \neq \emptyset$ and $e_{q_{i,t},t} \neq i$. When $Q_{i,t+1} \neq \emptyset$, the values of the perturbations of $\Delta x_{i,t+1}$ and $\Delta x_{q_{i,t+1},t+1}$ are same with that in the first period.

3.3.1.3. Total perturbation with respect to the threshold

According to the analysis about $Q_{i,t}$ for $t = 1, 2$, we can conclude that in any period t, $Q_{i,t}$ will be either empty or comprised of a pair of ports, i.e., $\{i, q_{i,t}\}$. It implies that the analysis about the perturbations in period $t > 2$ by the perturbation of $\Delta\gamma_i$ will be similar with that in either of the first two periods.

From lemma 3.3 and (3.18), in a general form, we have:

$$\frac{\partial J(N,\boldsymbol{\gamma})}{\partial \gamma_i} \approx \frac{1}{T}\cdot\sum_{t=1}^{T}\begin{pmatrix} I(Q_{i,t}=\emptyset)\cdot(-1)^{I(i\in P_t^S)}\cdot\pi_{i,t} \\ +I(Q_{i,t}\neq\emptyset)\cdot(-1)^{I(q_{i,t}\in P_t^S)}\cdot\pi_{q_{i,t},t} \\ +I(Q_{i,t+1}\neq\emptyset)\cdot(U_{i,t}-U_{q_{i,t+1},t}) \end{pmatrix} \tag{3.21}$$

where the first two terms of the RHS of (3.21) present the perturbations on the EC repositioning cost in two conditions, i.e., $Q_{i,t}=\emptyset$ and $Q_{i,t}\neq\emptyset$, respectively; the third term presents the perturbation on the holding and leasing cost when $Q_{i,t+1}\neq\emptyset$; for $t=1$, $Q_{i,t}=\emptyset$, and for $t>1$, $Q_{i,t}$ can be obtained from either (3.19) or (3.20), depending on the status of $Q_{i,t-1}$; the values of $U_{i,t}$ and $U_{q_{i,t+1},t}$ can be calculated from assertion 2) of lemma 3.3.

Next, we present the method to find $E_{i,t}$, followed by analyzing the gradient of expected total cost with respect to the fleet size with given policy.

3.3.2. Modified stepping stone method

In this section, we first use a numerical example to show that how the perturbation in the supply or demand of one port with binding constraint, such as i, will change the net actual imported ECs of a pair of ports. Then, a modified stepping stone (MSS) method is proposed to find $E_{i,t}$, which is defined as the set of ports whose net actual imported empty containers is changed by perturbing the estimated supply or demand of port i in period t.

Consider a numerical example with three surplus ports and two deficit ports. ECs are repositioned from the surplus ports to the deficit ports. Assume the total number of EC supplies is greater than the total number of EC demands. And the transportation tableau is presented Figure 3.3. Note that one dummy node is created in the transportation tableau to ensure that the total number of demands is equal to the total

Surplus port	Deficit port			
	1	2	3(dummy)	Change of supply
1	●			
2	●		●	
3		●	●	
Change of demand				

Figure 3.3 The transportation tableau
● Basic cell

number of supplies. For illustration, the value of cell (m, l) represents the number of ECs repositioned from the m^{th} surplus port to the l^{th} deficit port. Two sub-examples are investigated.

<u>S-E1: Perturb the empty container supply of the first surplus port by $\Delta\gamma$</u>

Consider the balance of the transportation tableau. Perturbing the number of EC supply of the first surplus port by $\Delta\gamma$ implies that the number of EC demand of the dummy deficit port will be perturbed by $\Delta\gamma$. However, since cell $(1, 3)$ is a non-basic cell (see Figure 3.3), its value should be kept as zero and cannot be changed by the perturbation. To track the changes on the basic variables, a loop is developed, which begins at cell $(1, 3)$ and is back to this cell (see Figure 3.4).

The loop in Figure 3.4 consists of successive horizontal and vertical segments whose end nodes must be basic variables, except for the two segments starting and

Surplus port	Deficit port			
	1	2	3(dummy)	Change of supply
1	$\Delta\gamma$ ●		◎	$\Delta\gamma$
2	$-\Delta\gamma$ ●		● $\Delta\gamma$	
3		●	●	
Change of demand			$\Delta\gamma$	

Figure 3.4 Perturb the number of EC supply of the first surplus port by $\Delta\gamma$
◎ The beginning cell

ending at the non-basic variable. More specifically, the changes on the basic variables can be obtained as follows: first increase the value of cell $(1, 1)$ by $\Delta\gamma$; then go around the loop, alternately decrease and then increase basic variables in the loop by $\Delta\gamma$, i.e. decrease the value of cell $(2, 1)$ and increase the value of cell $(2, 3)$ by $\Delta\gamma$. It is observed that the total numbers of repositioning out ECs at the first and the second surplus ports will be increased and decreased by $\Delta\gamma$, respectively.

S-E2: Perturb the number of EC demand of the first deficit port by $\Delta\gamma$

Perturbing the number of EC demand of a deficit port (not the dummy deficit port) by $\Delta\gamma$ will result in the total number of demands from deficit ports (excluding the dummy deficit port) increasing by $\Delta\gamma$. To maintain the balance of the transportation tableau, a hypothetic surplus port is introduced and its supply is assumed to increase by $\Delta\gamma$. Besides, to make the transportation tableau non-degeneracy, the cell (the hypothetic surplus port, the dummy deficit port), i.e., cell (4, 3) in Figure 3.5, is set as a basic cell.

To track the changes on the basic variables by the perturbation of $\Delta\gamma$ on the number of EC demand of the first deficit port, similar loop is developed in Figure 3.5. It is observed that the total number of repositioning in ECs at the first deficit port and the total number of repositioning out ECs at the second surplus port will be both

Surplus port	Deficit port			
	1	2	3(dummy)	Change of supply
1	•			
2	$\Delta\gamma$ •		• $-\Delta\gamma$	
3		•	•	
4 (hypothetic)	◎		• $\Delta\gamma$	$\Delta\gamma$
Change of demand	$\Delta\gamma$			

Figure 3.5 Perturb the number of EC demand of the first deficit port by $\Delta\gamma$

increased by $\Delta\gamma$.

Just as the loop using in the typical stepping stone method (Charnes and Cooper, 1954), the loop in either Figure 3.4 or Figure 3.5 is unique. The uniqueness of the loop guarantees that there will be only a pair of ports, whose net actual imported ECs will be affected by perturbing the supply or demand of one port with binding constraint. Similar results can be obtained under the other case in which estimated total number of EC supplies is less than the estimated total number of EC demands.

We can conclude that the perturbation in the estimated supply or demand of one port with binding constraint will only change the net actual imported ECs of a pair of ports. Next, we develop a Modified Stepping Stone (MSS) method to find the set of ports whose net actual imported ECs could be changed by perturbing the estimated supply or demand of one port in period t. Some additional notations are defined in Table 3.1.

Table 3.1 List of notations for MSS method

	Description
M_t	the number of surplus ports in period t
L_t	the number of deficit ports in period t
$I_{i,t}^{S}$	the index of the surplus port i in the transportation tableau of period t
$I_{j,t}^{D}$	the index of the deficit port j in the transportation tableau of period t

When the number of EC supply or demand of port i in period t is perturbed, a MSS method is proposed to find $E_{i,t}$ with the rules as follows.

Procedure of MSS

Step 1. Build the transportation tableau based on the transportation model solutions in period t. Create a dummy deficit node, $L_t + 1$ and a dummy surplus node, $M_t + 1$. And arbitrarily insert the value ε into cell $(M_t + 1,\ L_t + 1)$.

Step 2. If port i is a surplus port in period t, i.e. $i \in P_t^S$, select cell $(I_{i,t}^S,\ L_t + 1)$; if port i is a deficit port in period t, i.e. $i \in P_t^D$, select cell $(\ M_t + 1, I_{i,t}^D,)$.

Step 3. If the selected cell is a basic cell, no port's total repositioned empties quantity will be affected, set $E_{i,t} = \emptyset$ and stop. Otherwise, go to Step 4.

Step 4. Beginning at the selected cell, trace a loop back to the cell, turning corners only on basic cells. Only successive horizontal & vertical moves are allowed.

Step 5. If the total empty container supply is more than the total empty container demand, record the basic cell $(v,\ L_t + 1)$, $v \neq (\ M_t + 1)$ in this loop and find port k with $I_{k,t}^S = v$. Otherwise, record the basic cell $(\ M_t + 1, v)$, $v \neq (\ L_t + 1)$ in this loop and find port k with $I_{k,t}^D = v$. Then, $E_{i,t} = \{i, e_{i,t}\}$ with $e_{i,t} = k$.

3.3.3. Gradient with respect to the fleet size

Similarly with analysis in section 3.3.1, we study the gradient of expected total cost with respect to the fleet size with a given policy in this section.

We define the nominal path as the sample path generated by the simulation model with parameter N and the perturbed path as the sample path generated using the same model and same random seeds, but with parameter $(N)'$, where $(N)' = N + \Delta N$ and ΔN is sufficiently small. Since $\sum_{p \in P} \Delta x_{p,t} = \Delta N$ from Eq. (3.3), we can perturb $x_{i,1}$ by ΔN to investigate the perturbations. That is, we set $q_t^N = i$ with $\Delta x_{q_t^N,t} = \Delta N$ and $\Delta x_{p,t} = 0$ for other $p \in (P - \{q_t^N\})$ in the period $t = 1$.

The perturbation flow for this problem is similar with that shown in Figure 3.2, but the unique difference is that there is no perturbation from the threshold. In other words, the perturbations on various variables are only from the perturbation on the beginning on-hand inventory.

3.3.3.1. Perturbation with respect to the fleet size in the first period

In the period $t = 1$, we have $\Delta u^S_{q_t^N,t} = \Delta N$ or $\Delta u^D_{q_t^N,t} = -\Delta N$ and $\Delta u^S_{p,t}$ or $\Delta u^D_{p,t} = 0$ for other $p \in (P - \{q_t^N\})$. It implies that only the RHS of port q_t^N constraint is changed. Hence, we have

- If the port q_t^N constraint is not binding, we have $E_{q_t^N,t} = \emptyset$. And then $\Delta y_{q_t^N,t} = \Delta N$ and $\Delta y_{p,t} = 0$ for other $p \in (P - \{q_t^N\})$.
- Otherwise, we have $E_{q_t^N,t} = \left\{ e_{q_t^N,t} \right\}$. And then $\Delta y_{e_{q_t^N,t},t} = \Delta N$ and $\Delta y_{p,t} = 0$ for other $p \in \left(P - E_{q_t^N,t}\right)$.

Similarly, we have

$$q_{t+1}^N = \begin{cases} q_t^N, & \text{if } E_{q_t^N,t} = \emptyset \\ e_{q_t^N,t}, & \text{otherwise} \end{cases} \tag{3.22}$$

where $\Delta x_{q_{t+1}^N,t+1} = \Delta N$ and $\Delta x_{p,t+1} = 0$ for other $p \in \left(P - E_{q_t^N,t}\right)$. It implies that in each period, there is a unique port whose beginning on-hand inventory will be affected by ΔN. And we can have next lemma.

LEMMA 3.4: *In any period t, we have*

$$\frac{\partial y_{p,t}}{\partial N} = \begin{cases} 1, \text{if } p = q_{t+1}^{N} \\ 0, \text{otehwise} \end{cases}, \forall p \in P$$

PROOF: The proof is similar with that in the assertion 3) of the lemma 3.1. ■

3.3.3.2. Total perturbation with respect to the fleet size

From lemma 3.4 and (3.17), in a general from, we have:

$$\frac{\partial J(N, \boldsymbol{\gamma})}{\partial N} \approx \frac{1}{T} \cdot \sum_{t=1}^{T} \left((-1)^{I(q_t^N \in P_t^S)+1} \cdot \pi_{q_t^N,t} + U_{q_{t+1}^N,t} \right) \tag{3.23}$$

where the first term of the RHS of (3.23) presents the perturbation on the EC repositioning cost; the second term presents the perturbation on the holding and leasing cost; for $t = 1$, $q_t^N = i$, and for $t > 1$, q_t^N can be obtained from (3.22); the value of $U_{q_t^N,t}$ can be calculated from the assertion 2) of lemma 3.3.

3.4 Numerical experiments

In this section we aim to evaluate the performance of the proposed single-level threshold policy (STP) and provide some insights for shipping companies.

While our policy can be applied in a multi-port system with an arbitrary number of ports, we use three problems that differ in the number of ports: problem 1 has 6 ports, problem 2 has 9 ports and problem 3 has 12 ports, which represent small, moderate and large systems, respectively. Besides, considering that the trade imbalance could be the most important factor affecting the performance of the proposed policy, we design three kinds of trade imbalance patterns, including balanced trade pattern, moderately imbalanced trade pattern and severely imbalanced trade pattern. Therefore, a total of 9 cases are studied.

For each problem, the values of cost parameters and the average customer demand from port p to port m in a period, denoted by $\mu_{p,m}$, are randomly generated. For example, for problem 1, the value of the holding cost parameter, i.e., the value of $C_p^H\ \forall p \in P$, is uniformly generated from the interval (0, 5); the value of the repositioning cost parameter, i.e., the value of $C_{p,m}^R\ \forall p, m \in P, p \neq m$, is uniformly taken from the interval (5, 10) ; the value of the leasing cost parameter, i.e., the value of $C_p^L\ \forall p \in P$, is uniformly generated from the interval (10, 30). This reflects the general view that the cost of repositioning an EC is greater than the cost of holding an EC, while much less than the cost of leasing an EC in a period. In the balanced trade pattern, the value of $\mu_{p,m}$ is uniformly generated from the interval (0, 200), and we set $\mu_{p,m} = \mu_{m,p}\ \forall p \neq m$ to balance customer demands between any pair of ports. In the moderately (severely) imbalanced trade pattern, we double (treble) the values of one port's or several ports' exported laden containers and keep other values remain the same as those in the balanced trade pattern. The customer demands, i.e., $\varepsilon_{p,m,t} \forall p \neq m$, are assumed to follow normal distribution with mean $\mu_{p,m}$ and standard variance $0.2 \cdot \mu_{p,m}$, and be left-truncated at zero. In Appendix A, the values of the cost parameters and the average customer demands in different trade imbalance patterns of problem 1 are presented.

For a multi-port system with STP, we can obtain its optimal fleet size and thresholds through sequentially solving the problem (3.13) under Scenario-I and Scenario-II. Under Scenario-I, the NLP is solved by Matlab (version 7.0.1). Under Scenario-II, the IPA-gradient based algorithm is coded in Visual C++ 5.0. All the numerical studies are tested on a PC with 2.67GHz Processor under the Microsoft Vista Operation System. Based on preliminary experiments, we set the simulation period $T = 10100$ with 100 warm-up periods. The termination criteria for searching are that the maximum iteration, namely n_{max}, is achieve or the expected total cost in current iteration is significantly greater than that in last iteration. We set $n_{max} = 1000$.

3.4.1. Policy performance evaluation

To evaluate the performance the proposed STP, a match back policy (MBP) is introduced for comparison. Such policy is widely accepted and applied in practice and its principle is to balance the container flow in each pair of ports. In other words, ECs to be repositioned from port p to port m in period $t+1$ should try to match the difference between the total number of laden containers exported from port m to port p and the total number of laden containers exported from port p to port m in period t. Mathematically, we have

$$\left(z_{p,m,t+1}\right)_{MBP} = \left(\varepsilon_{m,p,t} - \varepsilon_{p,m,t}\right)^{+}, \forall t \tag{3.24}$$

When MBP is adopted, the repositioning cost is independent from the fleet size. Hence, the fleet size which minimizes the expected holding and leasing cost will be the optimal fleet size minimizing the expected total cost. We have $(N^*)_{MBP} = \sum_{p \in P} F_p^{-1}\left(C_p^L / \left(C_p^L + C_p^H\right)\right)$.

To facilitate the comparison of STP and MBP, the percentage of expected total cost reduction achieved by STP from MBP is given in Figure 3.6. We use a doublet, i.e., (number of ports, trade pattern), to present a particular case. For example, (6, B), (6, M-IB), and (6, S-IB) mean the cases for problem 1 with 6 ports in the balanced, moderately imbalanced and severely imbalanced trade patterns, respectively.

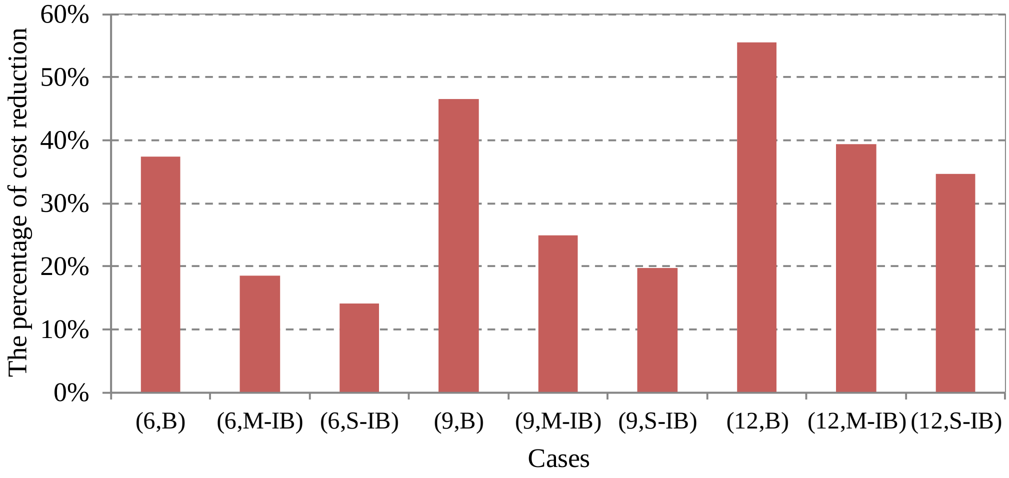


Figure 3.6 Percentages of total cost reduction achieved by STP from MBP

There are three main observations from the results. First, STP outperforms MBP in all cases. As expected, the reduction of total cost by STP is major from the reduction of the repositioning cost. Second, as the system becomes larger, STP can reduce more cost. Hence, it is important for shipping companies to use better method in repositioning ECs, instead of resorting to simple way such as the MBP, especially in the complex system. Another observation is that for a problem, it seems that in the imbalanced trade patterns, the advantage of using STP over MBP seems not as great as that in the balanced trade pattern. One possible explanation is that MBP performs well in the imbalanced trade, in which a number of ECs should be repositioned to balance the trade flow. However, STP still can reduce cost from MBP in the imbalanced trade pattern. For example, under the case of 12 ports problem, the cost reduction can be as high as 30%.

3.4.2. Policy performance sensitivity to the fleet size

Considering the fact that container fleet is often fixed by operators in practice, we further investigate the policy performance under both policies with the given fleet size.

If the fleet size is given, the optimal thresholds under STP can be found by the two proposed approaches with little adjustments. Hence, let N^* be the optimal fleet size under STP. We vary the fleet size from $0.7N^*$ to $1.3N^*$ in all 7 cases to investigate the effect of the fleet size on the expected total cost.

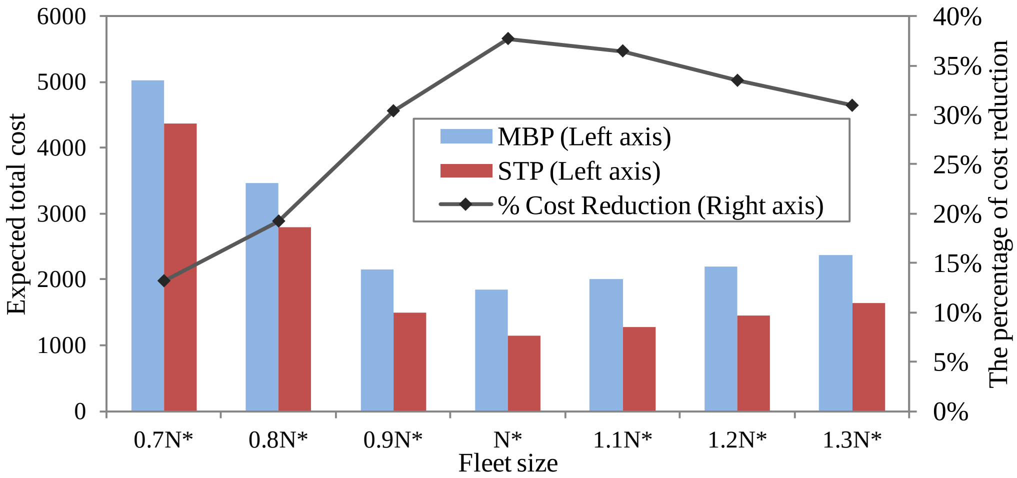


Figure 3.7 Comparison of STP and MBP in case (6, B)

Figure 3.7 shows the expected total cost under both policies in case (6, B) with different fleet sizes. First, it is also observed that STP outperforms MBP in all different fleet size cases. It reveals that the expected total cost per period savings achieved by STP over MBP are of the order of 13.18%~37.72%. One possible explanation is that the proposed policy makes the ECR decisions in terms of minimizing the repositioning cost. Besides, the trend of the diamond line shows that, the cost saving achieved by STP from MBP increases gradually when the fleet size increases, and as the fleet size increases further, the saving decreases. It is also possible to have a small cost saving percentage when the fleet size is too little. The reason is that when the fleet size is severely insufficient to satisfy customer demands, a large number of ECs are leased and few requirements for repositioning ECs. Third, the results also show that the minimum expected total cost per period appears to be convex with respect to the fleet size under both policies. It reflects the intuition that the optimal fleet size is the trade-off between the repositioning cost and the holding and leasing cost.

Similarly observations can be also found in other cases.

3.4.3. Sensitivity analysis of the thresholds

Since the single-level threshold policy is applied to manage ECs in the multi-port system, the thresholds of the policy will significantly affect the performance of the proposed policy. Many factors may impact the thresholds of the policy. The most significant factors are the fleet size and the parameters of leasing cost and holding cost. Next, based on the case (6, B), we explore the sensitivity of the thresholds.

As shown in Figure 3.8, the optimal threshold values generally increase as the fleet size increases. More interesting, as the fleet size increases from $1.1N^*$ to $1.3N^*$, ports 1~5 have very small increments on their thresholds, while port 6 has large increment on its threshold. This reflects the fact that as the fleet size is much greater than its optimal value, the low holding cost at port 6 works in favor of keeping a large number of ECs at this port. While as the fleet size decrease from $0.9N^*$ to $0.7N^*$, port

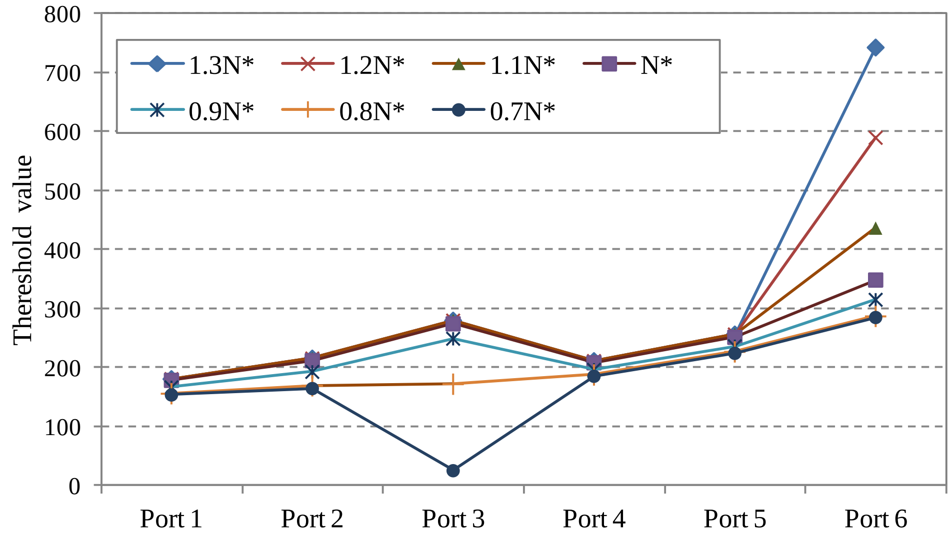


Figure 3.8 The optimal threshold value for case (6, B) under STP

3 has the largest decrement on its threshold. A possible explanation is that as owned ECs are insufficient at all ports, many leasing containers have to be used to satisfy customer demands. And the low leasing cost at port 3 supports it to keep low inventory level of ECs and lease a large number of ECs.

Next, the sensitivity of the thresholds to the holding and leasing cost parameters is considered. From (3.18), we obtain that $\partial\left(\frac{\partial(J(N,\boldsymbol{\gamma}))}{\partial\gamma_i}\right)/\partial C_i^L \leq 0$ and $\partial\left(\frac{\partial(J(N,\boldsymbol{\gamma}))}{\partial\gamma_i}\right)/\partial C_{q_{i,t}}^L \geq 0$. It implies that $\frac{\partial(J(N,\boldsymbol{\gamma}))}{\partial\gamma_i}$ will decrease (increase) as C_i^L ($C_{q_{i,t}}^L$) increases. Thus, when the leasing cost of port i increases, the threshold for this port will increase while the thresholds for some other ports decrease. The similar property for the holding cost parameter can be derived. That is, as the holding cost of port i increases, the threshold of this port will decrease while the thresholds for some other ports increase.

Focusing on the case (6, B) with optimal fleet size under STP, we consider two more cases, i.e., cases A and B, in which the holding and leasing cost parameters of port 6 are increased by 2 times, respectively. Figure 3.9 shows the results about thresholds changes by cases A and B from the original case.

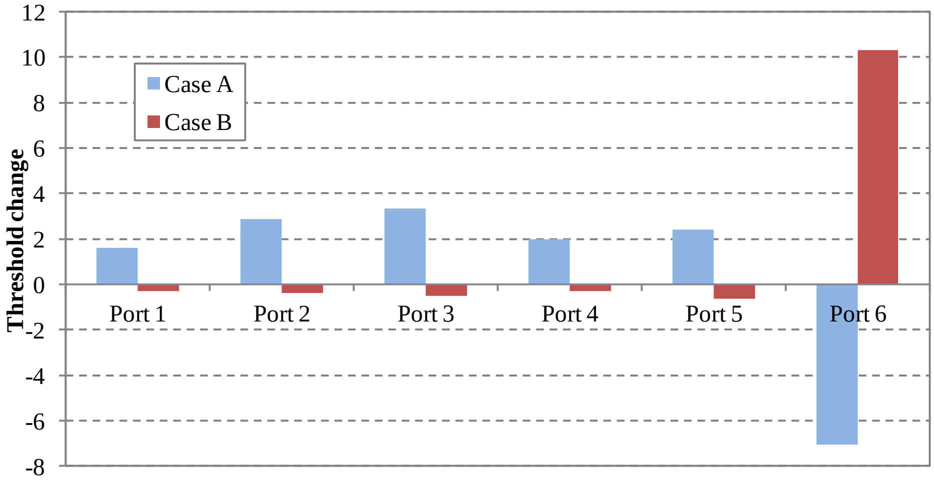


Figure 3.9 Optimal thresholds changes in cases A and B from the original case

It can be observed that when the holding (leasing) cost parameter of port 6 increases by 2 times, the threshold of port 6 decreases (increases), but the thresholds of other ports increase (decrease). This testifies above phenomenon. Hence, the results reflect the fact that a higher leasing cost at a port works in favor of keeping a large number of ECs while a higher holding cost encourages repositioning out more ECs in a surplus port or repositioning in less EC in a deficit port.

3.5 Summary

In this chapter, we address the joint ECR and container fleet sizing problem in a multi-port system. A single-level threshold policy is developed to reposition ECs periodically. Two approaches, a non-linear programming and IPA-based gradient technique are developed to solve the problem, which optimizes the fleet size and the parameters of the policy under Scenario-I and Scenario-II, respectively. The numerical results provide insights that by intelligently repositioning ECs, we can significantly reduce the total operation cost.

As we assumed that the ECs can be dispatched between any pair of ports within one period, an extension of this work is needed to relax the one-period assumption and to consider more complex liner shipping system.

Chapter 4 INVENTORY-BASED EMPTY CONTAINER REPOSITIONING IN LINER SHIPPING SYSTEM

In this chapter, we consider ECR problem in liner shipping system with multiple service routes. Two threshold control policies are proposed. A time-driven simulation model is formulated. Numerical studies are provided to show the application of the simulation model and the performance of the proposed policies.

4.1 Problem description

In this study, we extend the previous study to consider ECR problem with the inventory-based threshold policy in the liner shipping system with multiple service routes. As we have presented former, a service route is defined as a special sequence of calling ports, which construct a closed loop. Vessel(s) is deployed in the loop to provide weekly service. For example, if the cycle travelling time of a service route is four weeks, four vessels are assigned in this service route.

We consider the inventory-based threshold policy with flexible destination. That is, the ports of destinations to unload ECs are not determined in advance when ECs are loaded to a vessel from an original port, and when the vessel arrives at a port, ECs can be unloaded as need. Such policy has been widely applied in current shipping industry and its good performance under some conditions has also been demonstrated by Song and Dong (2010).

We aim to develop a simulation model with considering the dynamic and uncertain customer demand, supply and residual capacity.

4.2 Problem formulation

We formulate the problem in a time-driven way. One period is one day. The state of the system can be updated at the end of each period.

4.2.1. Modeling assumptions

Several assumptions are used in this study.

- Once a vessel is assigned to a service route, it will stay at this service route for the whole planning horizon. This is usually the case in practice, as liner companies do not change vessel deployment schedule regularly.
- Laden containers have priority to occupy the vessel capacity and ECs only take up the residual space; the residual vessel space available for EC is randomly generated.
- When a vessel arrives at a port, we only unload ECs.
- When a vessel departs from a port, if there is insufficient vessel capacity for loading laden containers, some ECs will be unloaded from the vessel to release the capacity for laden containers. Otherwise, we can load ECs.
- All customer demands must be satisfied and cannot be delayed to the next period.
- Short-term leasing is considered and the quantity of the leased ECs is always available in the port at any time.
- The leased containers are not distinguished from owned containers. More specifically, we can return owned ECs instead of the leased ones to vendors in order to reduce the leasing time.
- We return the number of the leased ECs to the vendor in the original port after a specific period.

4.2.2. Notations

To describe the system, we introduce the following notation.

Index sets

P The set of ports

G The set of all service routes

S_g The set of stops on service route g

V_g The set of vessels on service route g

$A_{v,s}$ The set of periods in which vessel v arrives at stop s

$D_{v,s}$ The set of periods in which vessel v departs from stop s

$S_{t,v}$ The set of stops which vessel v arrives at or departs from within period t

$V_{t,i}^{a}$ The set of vessels which arrive at port i within period t

$V_{t,i}^{d}$ The set of vessels which depart from port i within period t

Deterministic parameters

$p_{g,s}$ The port corresponding to the stop s on service route g

R_v The carrying capacity in TEUs of vessel v

$c_{t,i}^{u}$ Cost of unloading an EC from a vessel at port i at period t

$c_{t,i}^{w}$ Cost of loading an EC to a vessel at port i at period t

$c_{t,s,v}^{x}$ The transportation cost of an EC going from stop s to next stop on vessel v leaving at period t

$c_{t,i}^{y}$ Per period cost for storing an EC at port i at period t

$c_{t,i}^{z}$ Cost of leasing an EC at port i at period t

T^r The specific period, after which the leased EC should be returned

Random data

$\xi_{t,i}$ Customer supply at port i at period t

$\zeta_{t,i}$ Customer demand at port i at period t

$\gamma_{t,s,v}$ The residual capacity on vessel v for ECs when it leaves stop s at period t

State variables

$x_{t,v}$ The number of ECs on vessel v at the end of period t ($v \in V_g, g \in G$)

$y_{t,i}$ The number of ECs stored at port i at the end of period t ($i \in P$)

Decision variables

$u^a_{t,s,v}$ The number of ECs unloaded from vessel v when it arrives at stop s at period t $(v \in V_g, g \in G, s \in S_g, t \in A_{v,s})$

$u^d_{t,s,v}$ The number of ECs unloaded from vessel v when it departs from stop s at period t $(v \in V_g, g \in G, s \in S_g, t \in D_{v,s})$

$w_{t,s,v}$ The number of ECs loaded at stop s to vessel v at period t $(v \in V_g, g \in G, s \in S_g, t \in D_{v,s})$

$l_{t,i}$ The number of ECs leased at port i at period t ($i \in P$)

In this study, we consider three major sources of randomness, namely customer demand, customer supply and residual capacity on vessel for ECs. Customer demand refers to ECs picked up by customers from the inventory depot to load their cargoes, whose actual value at a particular time is highly uncertain. On other hand, customer supply refers to ECs returned from consignees (customers) to the inventory depot, which is also out of shipping company's control. Finally, the actual number of laded containers on a vessel is also not known with certainty. For example, the weight of individual laden containers can vary substantially. A large number of heavy containers can reduce the effective capacity on vessel for ECs. We assume that the customer demands, customer supplies and vessel capacities for ECs are independent random variables.

4.2.3. State transition

The states of the system are updated at the end of a period. More specifically, at the end of period t, for each port and each vessel, we have

$$y_{t,i} = y_{t-1,i} - l_{t-T^r,i} + \xi_{t,i} - \zeta_{t,i} + l_{t,i} + \sum_{(v,s)\in\{(v,s)|v\in V_g, g\in G, s\in S_g, p_{g,s}=i, t\in A_{v,s}\}} u^a_{t,s,v} + \sum_{(v,s)\in\{(v,s)|v\in V_g, g\in G, s\in S_g, p_{g,s}=i, t\in D_{v,s}\}} \left(u^d_{t,s,v} - w_{t,s,v}\right) \quad \forall i \in P \tag{4.1}$$

$$x_{t,v} = x_{t-1,v} - \sum_{(s)\in\{(s)|s\in S_g, t\in A_{v,s}\}} u^a_{t,s,v} - \sum_{(s)\in\{(s)|s\in S_g, t\in D_{v,s}\}} \left(u^d_{t,s,v} - w_{t,s,v}\right) \quad \forall (v) \in \left\{\left((v)\middle| v \in V_g, g \in G\right)\right\} \tag{4.2}$$

The end inventory at period t is equal to the original inventory plus total container flows at period t for each port and vessel, respectively.

Within period t, for vessel v, if set $S_{t,v}$ is empty, it implies that this vessel is at sea or staying at a port in this period. Otherwise, it implies vessel v will arrive at or depart from those stops in the set $S_{t,v}$. When $S_{t,v}$ is nonempty, we sequentially rank the stops in $S_{t,v}$ according to the increasing called time, and have $\mathscr{S}_{t,v} = \left\{s_1, \dots, s_e, \dots s_{|S_{t,v}|}\right\}$, where s_e is the e-th called stop and $|S_{t,v}|$ is the number of stops in this set. Then, for vessel v with $S_{t,v} \neq \emptyset$ in period t, we need to define other variables:

$k^a_{t,s,v}$ The number of ECs on vessel v just before this vessel arrives at stop s at period t $\left(v \in V_g, g \in G, s \in S_{t,v}, t \in A_{v,s}\right)$

$k^b_{t,s,v}$ The number of ECs currently on the vessel v at time t and staying on this vessel at stop s $\left(v \in V_g, g \in G, s \in S_{t,v}\right)$

$k^d_{t,s,v}$ The number of ECs to be repositioned from stop s to next stop on vessel v departing at period t $\left(v \in V_g, g \in G, s \in S_{t,v}, t \in D_{v,s}\right)$

Then, we can have

$$k^a_{t,s_e,v} = \begin{cases} x_{t-1,v}, & \text{if } e = 1, t \in A_{v,s_e} \\ k^d_{t,s_{e-1},v}, & \text{if } 1 < e \leq |S_{t,v}| \end{cases} \tag{4.3}$$

$$k^b_{t,s_e,v} = \begin{cases} x_{t-1,v}, & \text{if } e = 1, t \notin A_{v,s_e} \\ k^a_{t,s_e,v} - u^a_{t,s_e,v}, & \text{if } 1 < e \leq |S_{t,v}|, t \in A_{v,s_e} \end{cases} \tag{4.4}$$

$$k^d_{t,s_e,v} = k^b_{t,s_{e-1},v} - u^d_{t,s_e,v} + w_{t,s_e,v} \tag{4.5}$$

where Eqs. (4.3)~(4.5) guarantee the balance of container flows on vessel v.

4.2.4. Inventory-based threshold policy

Inventory-based threshold policy $\boldsymbol{\theta}$ is applied to make the ECR decisions at the beginning of period t for all ports.

We consider two inventory-based policies with different levels of threshold number, i.e., a single-level threshold policy and a two-level threshold policy. Denote $\boldsymbol{\theta}^{\mathrm{SL}}$ and $\boldsymbol{\theta}^{\mathrm{TL}}$ as the single-level and the two-level threshold policies, respectively. We have $\boldsymbol{\theta} \in \{\boldsymbol{\theta}^{\mathrm{SL}}, \boldsymbol{\theta}^{\mathrm{TL}}\}$ with $\boldsymbol{\theta}^{\mathrm{SL}} = [\theta_i]_{i \in P}$ and $\boldsymbol{\theta}^{\mathrm{TL}} = \left[\left(\theta_i^D, \theta_i^U\right)\right]_{i \in P}$, where θ_i and $\left(\theta_i^D, \theta_i^U\right)$ are the thresholds for port i in both policies, respectively. For the two-level threshold policy, the lower level threshold should be not greater than the upper level threshold for each port, i.e., $\theta_i^D \leq \theta_i^U, \forall i \in P$. Note that the number of ECs stored at port i at the end of period t, i.e., $y_{t,i}$, is equivalent to the on-hand EC inventory at the beginning of period $t+1$. The basic rules for both policies are as follows:

- for the single-level threshold policy: If $y_{t,i} > \theta_i$, then the quantity excess of θ_i should be repositioned out to other ports, which may need ECs, to try to bring the inventory position down to θ_i; if $y_{t,i} < \theta_i$, then ECs should be repositioned in from other ports, which may supply ECs, to try to bring the inventory position up to θ_i; if $y_{t,i} = \theta_i$, then nothing is done.

- for the two-level threshold policy: if $y_{t,i} > \theta_i^U$, then port i is a surplus port and the quantity excess of θ_i^U should be repositioned out to other ports to try to bring the inventory position down to θ_i^U; if $y_{t,i} < \theta_i^D$, then it is a deficit port and ECs should be repositioned in from other ports to try to bring the inventory position up to θ_i^D; if $\theta_i^D \leq \ y_{t,i} \leq \theta_i^U$, then nothing is done.

It can be seen that the single-level threshold policy is a special instance of the two-level threshold policy when the lower-level threshold is equal to the upper-level threshold. In general, at the beginning of period t, we make the ECR decisions following three steps:

Step 1- Estimate the number of repositioned ECs for all ports

Denote $\mathcal{O}_{t,i}$ and $\mathcal{I}_{t,i}$ as the estimated total number of ECs repositioned out from and repositioned into port i within period t, respectively. From the basic rule of single-level policy, we have

$$\begin{cases} \mathcal{O}_{t,i} = y_{t-1,i} - \theta_i,\ \mathcal{I}_{t,i} = 0, & \text{if } y_{t-1,i} > \theta_i \\ \mathcal{O}_{t,i} = 0, \mathcal{I}_{t,i} = \theta_i - y_{t-1,i}\,, & \text{if } y_{t-1,i} < \theta_i \\ \mathcal{O}_{t,i} = 0, \mathcal{I}_{t,i} = 0, & \text{if } y_{t-1,i} = \theta_i \end{cases} \quad \forall i \in P \tag{4.6}$$

Similarly, with the two-level policy, we have

$$\begin{cases} \mathcal{O}_{t,i} = y_{t-1,i} - \theta_i^U,\ \mathcal{I}_{t,i} = 0, & \text{if } y_{t-1,i} > \theta_i^U \\ \mathcal{O}_{t,i} = 0, \mathcal{I}_{t,i} = \theta_i^D - y_{t-1,i}, & \text{if } y_{t-1,i} < \theta_i^D \\ \mathcal{O}_{t,i} = 0, \mathcal{I}_{t,i} = 0, & \text{if } \theta_i^D \leq \ y_{t-1,i} \leq \theta_i^U \end{cases} \quad \forall i \in P \tag{4.7}$$

Step 2- Estimate the number of repositioned ECs unloaded from or loaded to vessels for all ports

For port i, it is possible that multiple vessels visit it at period t. Hence, ranking the vessels which will depart from port i in period t according the increasing departure time, we define $\mathcal{V}_{t,i}^{d} = \left\{v_1^d, \dots, v_e^d, \dots v_{\left|V_{t,i}^d\right|}^d\right\}$, where $\left|V_{t,i}^d\right|$ is the number of vessels in $V_{t,i}^d$ and v_e^d is the e-th departure vessel with $v_e^d \in V_\mathrm{g}, \mathrm{g} \in G, i = p_{\mathrm{g},s}, t \in D_{v_e^d,s}$. Similarly, ranking the vessels which will arrive at port i in period t according the increasing arrival time, we define $\mathcal{V}_{t,i}^{a} = \left\{v_1^a, \dots, v_e^a, \dots v_{\left|V_{t,i}^a\right|}^a\right\}$, where $\left|V_{t,i}^a\right|$ is the number of vessels in $V_{t,i}^a$ and v_e^a is the e-th arrival vessel with $v_e^a \in V_\mathrm{g}, \mathrm{g} \in G, i = p_{\mathrm{g},s}, t \in A_{v_e^a,s}$. In general, we set the allocation rules as follows:

- If $\mathcal{I}_{t,i} > 0$, for the vessels which will arrive at port i within stage t, based on their arrival sequences, i.e., $\mathcal{V}_{t,i}^{a}$, we try to unload as many ECs as possible from the vessels until the total number of unloaded ECs is equal to $\mathcal{I}_{t,i}$
- If $\mathcal{O}_{t,i} > 0$, for the vessels which will depart from port i within stage t, based on their departure sequences, i.e., $\mathcal{V}_{t,i}^{d}$, we try to load as many ECs as possible onto the vessels until the total number of loaded ECs is equal to $\mathcal{O}_{t,i}$

Step 3- Implement the repositioning plan subject to constraints

To simplify the narrative, we first define $\mathscr{O}_{t,s,v}$ ($\mathscr{I}_{t,s,v}$) as the estimated remaining number of ECs, which are needed to be repositioned out from (repositioned into) stop s, just before the vessel v departs from (arrives at) this stop in period t. We have

$$\mathscr{O}_{t,s,v} = \mathcal{O}_{t,i} - \sum_{\left(v_j^d,s'\right)\in\left\{\left(v_j^d,s'\right)\middle|v_j^d\in\mathcal{V}_{t,i}^d,1\le j\le e-1,v_j^d\in V_\mathrm{g},\mathrm{g}\in G,p_{\mathrm{g},s'}=i\right\}} u_{t,s',v_j^d}^{d} \quad (4.8)$$

$$\forall(v,s,t) \in \left\{(v,s,t)\middle|v \in \mathcal{V}_{t,i}^{d}, v = v_e^d, v \in V_\mathrm{g}, \mathrm{g} \in G, p_{\mathrm{g},s} = i\right\}$$

$$\mathscr{I}_{t,s,v} = \mathscr{q}_{t,i} - \sum_{\left(v_j^a,s'\right)\in\left\{\left(v_j^a,s'\right)\middle|v_j^a\in\mathscr{V}_{t,i}^a,1\le j\le e-1,v_j^a\in V_{\mathrm{g}},\mathrm{g}\in G,p_{\mathrm{g},s'}=i\right\}} u_{t,s',v_j^a}^a$$
$$\forall(v,s,t)\in\left\{(v,s,t)\middle|v\in\mathscr{V}_{t,i}^a, v=v_e^a, v\in V_{\mathrm{g}},\mathrm{g}\in G,p_{\mathrm{g},s}=i\right\} \tag{4.9}$$

When vessel v arrives at stop s at period t, if there is no EC repositioning in requirement in this stop, we do not unload ECs. Otherwise, we unload ECs from the vessel to satisfy this requirement, while the number of unloaded ECs should not exceed the total number of ECs on this vessel. That is,

$$u_{t,s,v}^a = min\left\{k_{t,s,v}^a,\mathscr{I}_{t,s,v}\right\}$$
$$\forall(v,s,t)\in\left\{(v,s,t)\middle|v\in V_{\mathrm{g}},\mathrm{g}\in G,s\in S_{\mathrm{g}},t\in A_{v,s}\right\} \tag{4.10}$$

Note that laden containers have priority to occupy the vessel capacity. Hence, when vessel v departs from stop s at period t, we first examine whether there is sufficient vessel capacity for loading laden containers. If yes, no ECs on vessel will be unloaded. Otherwise, some ECs will be unloaded from the vessel to release the capacity for the laden containers, and we have

$$u_{t,s,v}^d = max\left\{k_{t,s,v}^b-\gamma_{t,s,v},0\right\}$$
$$\forall(v,s,t)\in\left\{(v,s,t)\middle|v\in V_{\mathrm{g}},\mathrm{g}\in G,s\in S_{\mathrm{g}},t\in D_{v,s}\right\} \tag{4.11}$$

Then, if there is EC repositioning out requirement in this stop, we load ECs to this vessel. The number of ECs loaded should not exceed the residual capacity on vessel.

$$w_{t,s,v} = min\left\{\gamma_{t,s,v}-k_{t,s,v}^b+u_{t,s,v}^d,\ \mathscr{O}_{t,s,v}\right\}$$
$$\forall(v,s,t)\in\left\{(v,s,t)\middle|v\in V_{\mathrm{g}},\mathrm{g}\in G,s\in S_{\mathrm{g}},t\in D_{v,s}\right\} \tag{4.12}$$

4.2.5. Cost function

Let J_t denote the daily cost in our problem, and we have

$$\begin{aligned} J_t = &\sum_{i\in P} c_{t,i}^{u} \sum_{(v,s)\in\{(v,s)|v\in V_g, g\in G, s\in S_g, p_{g,s}=i, t\in A_{v,s}\}} u_{t,s,v}^{a} \\ &+ \sum_{i\in P} c_{t,i}^{u} \sum_{(v,s)\in\{(v,s)|v\in V_g, g\in G, s\in S_g, p_{g,s}=i, t\in D_{v,s}\}} u_{t,s,v}^{d} \\ &+ \sum_{i\in P} c_{t,i}^{w} \sum_{(v,s)\in\{(v,s)|v\in V_g, g\in G, s\in S_g, p_{g,s}=i, t\in D_{v,s}\}} w_{t,s,v} \\ &+ \sum_{i\in P} \left(c_{t,i}^{y}\cdot y_{t,i}\right) + \sum_{i\in P}\left(c_{t,i}^{z}\cdot l_{t,i}\right) \\ &+ \sum_{(v,s)\in\{(v,s)|v\in V_g, g\in G, s\in S_g, t\in D_{v,s}\}} c_{t,s,v}^{x}\cdot k_{t,s,v}^{d} \end{aligned} \tag{4.13}$$

where the first two items on the RHS of (4.13) are the unloading handling costs, which refer to the cost of unloading ECs from vessels to ports when vessels arrive at and depart from ports, respectively; the third item is the loading handling cost referring to the cost of loading ECs from ports to vessels; the fourth item is the holding cost, which presents the storage cost when ECs are stored at ports; the fifth item is the leasing cost incurred when ECs are leased from vendors; the sixth item is the transportation cost capturing the cost of transporting ECs on the vessels.

Given a policy $\boldsymbol{\theta}$, we measure its performance by the expected total cost per period. Let $J(\boldsymbol{\theta})$ be the expected total cost per period with policy $\boldsymbol{\theta}$. We have

$$J(\boldsymbol{\theta}) \approx \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} Y(\boldsymbol{\theta},\omega_n) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{T}\sum_t J_t\right) \tag{4.14}$$

where $Y(\boldsymbol{\theta},\omega_n)$ is a sample average cost per period, and ω_n is a realization of the stochastic process of the random customer supplies, customer demands and residual capacities on vessels for ECs. The system is subject to the policy rules, the state transition Eqs. (4.1) and (4.2), and the non-negative integer quantity constraint, i.e., all the decisions variables and state variables should be non-negative integer

$$u^a_{t,s,v}, u^d_{t,s,v}, w_{t,s,v}, k^a_{t,s,v}, k^b_{t,s,v}, k^d_{t,s,v}, x_{t,i}, y_{t,i} \geq 0 \text{ and integer} \tag{4.15}$$

Note that in Chapter 3, we simplify ECR problem by considering variables that related to the flow of ECs as continuous variables. Now, we keep the integer constraint on these variables related to ECs in order to study the performance of the inventory-based policy in this case.

4.3 Numerical study

In this section, we present the application of simulation model and compare the performances of the proposed policies. We conduct our experiments using a set of problems with actual service routes of one shipping company. The data of the mean values for uncertain customer supplies, customer demands and residual capacities is generated by using a problem generator that is described below.

4.3.1. Data generator

From the idea of Cheung and Chen (1998), we develop a data generator to produce the mean values for uncertain customer supplies, customer demands and residual capacities for our problem. Some additional notations for data generator are defined in Table 4.1.

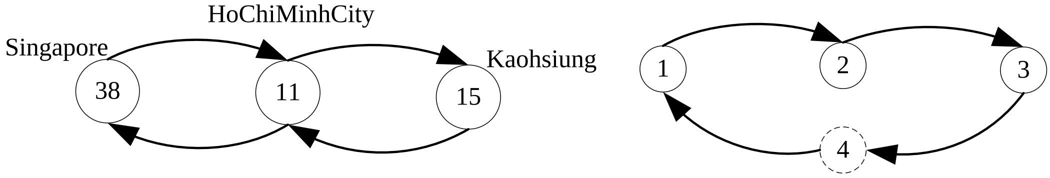
Table 4.1 List of notations for data generator

	Description
$\bar{d}_{i,j,v,t}$	The average number of laden containers to be transported from stop i to stop j on vessel v departing at period t $\left(v \in V_{\mathrm{g}}, \mathrm{g} \in G, i, j \in S_{t,v}, i \neq j, t \in D_{v,i}\right)$
$\bar{R}_{s,r,v,t}$	The average residual capacity on vessel v for ECs when it leaves stop s at period t to its next stop r
$\tau_{k,i,v}$	The time between vessel v departs from stop k and arrive at stop i in one voyage
$\bar{D}_{i,t}$	The average customer demand at port i at time t
$\bar{S}_{i,t}$	The average customer supply at port i at time t

Consider that customer demands, customer supplies and the residual capacities of vessels are all high related to laden containers. We first generate the value of $\bar{d}_{i,j,v,t}$ by

$$\bar{d}_{i,j,v,t} = I\left\{(i,j)_{\mathrm{g}}\right\} \cdot P\left(\phi_i^{Out} \cdot \phi_j^{In} \cdot \bar{D}\right) \; \forall t \in D_{v,i}, v \in V_{\mathrm{g}}, \mathrm{g} \in G, i, j \in S_{\mathrm{g}}, i \neq j \quad (4.16)$$

where $I\left\{(i,j)_{\mathrm{g}}\right\}$ is denoted as an indicator function, which takes 1 if voyage(i,j) in service route g is open and otherwise 0; ϕ_i^{In} and ϕ_i^{Out} are the in-bound and out-bound demand potential for laden container at port i, respectively; $\bar{D}$ represents the overall mean demand of the laden container; $Poi(m)$ represents a random variable of Poisson distribution with mean m. The reason to define $I\left\{(i,j)_{\mathrm{g}}\right\}$ is that some stops in a service route may correspond to a same port since ports in a service route may be called twice or more in a round-trip. Figure 4.1 shows the route of service SVX.



a) Original calling route of service SVX   b) Corresponding stop calling route

Figure 4.1 Liner service SVX

For service SVX, stop 2 and stop 4 in the corresponding stop calling route refer to the port of HoChiMinhCity. Therefore, there should be no customer cargos from stop 2 to 4, i.e., $I\{(2{,}4)_{\mathrm{SVX}}\} = 0$. Besides, physically, customer cargos from original port to destination port will be transported by the cheapest stop index-pair in the trip. Therefore, for example, customer cargos from port Singapore to port HoChiMinhCity should be transported from stop 1 to stop 2, while not from stop 1 to stop 4, i.e., $I\{(1{,}2)_{\mathrm{SVX}}\} = 1$ and $I\{(1,4)_{\mathrm{SVX}}\} = 0$.

Several days before the scheduled departure of the vessel, ECs must be available to customers for loading cargos. Similarly, after laden containers arrive at a stop (i.e., a port) and are shipped to the customers, the customers will keep the containers for several days before they return the empty ones to the stop. Hence, we denote $\alpha_{i,l}^{D}$ as the probability that an EC is needed l days before the scheduled departure in port i, and $\alpha_{i,l}^{S}$ as the probability that an EC is returned l days after the scheduled arrival in port i. Then, the mean customer demand and mean customer supply for port i at period t can be computed by

$$\bar{D}_{i,t} = \sum_{l}\left(\alpha_{i,l}^{D} \cdot \sum_{(s,v)\in\{(s,v)|v\in V_{\mathrm{g}},\mathrm{g}\in G,s\in S_{\mathrm{g}},p_{\mathrm{g},s}=i,(t+l)\in D_{v,s}\}} \sum_{j\in S_{\mathrm{g}}} \bar{d}_{s,j,v,t+l}\right) \tag{4.17}$$

$$\bar{S}_{i,t} = \sum_{l}\left(\alpha_{i,l}^{S} \sum_{(s,v)\in\{(s,v)|v\in V_{\mathrm{g}},\mathrm{g}\in G,s\in S_{\mathrm{g}},p_{\mathrm{g},s}=i,(t-l)\in A_{v,s}\}} \sum_{k\in S_{\mathrm{g}}} \bar{d}_{k,s,v,t-l-\tau_{k,i,v}}\right) \tag{4.18}$$

Recall that liner shipping company provides weekly service for each port. For the system, the average total number of weekly customer demands should be equal to the average total number of weekly customer supplies. Besides, residual capacity can be calculated as

$$\bar{R}_{s,r,v,t} = Q_v - \sum_{(i,j,t')\in\{(i,j,t')|(i,j)\in U_{s,r,v},t'\in D_{v,s},t'\le t\}} \bar{d}_{i,j,v,t'} \tag{4.19}$$

$$\forall t \in D_{v,s}, v \in V_{\mathrm{g}}, \mathrm{g} \in G, s, r \in S_{\mathrm{g}}$$

where we denote

$U_{s,r,v} = \{(i,j) | \text{vessel } v \text{ departing from stop } i \text{ to stop } j \text{ will call stop } s \text{ and stop } r\}$

4.3.2. Performance of inventory-based policies

In this section we present the application of our simulation model with utilizing the two policies to manage ECs, i.e., the single-level threshed policy and the two-level threshed policy, for two small size problems.

Both small size problems have three ports, but differ in the number of service routes: problem 1 has 1 service route and problem 2 has 2 service routes, which represent single- service route and multi-service routes systems, respectively. Besides, the uncertainty of customer demands can be represented by probability distributions. So we consider two types of distributions, i.e., exponential and normal distribution for the daily customer demand, daily customer supply and residual capacity on vessel for ECs. And the degree of uncertainty, i.e., the standard deviation, is decreasing from exponential to normal distribution. Therefore, a total of 4 cases for both policies are studied. The input data for both problems are presented in Appendix B, and we set the transportation cost $c_{t,s,v}^{x} = 10$. The networks for both problems are presented in Figure 4.2 and Figure 4.3, respectively.

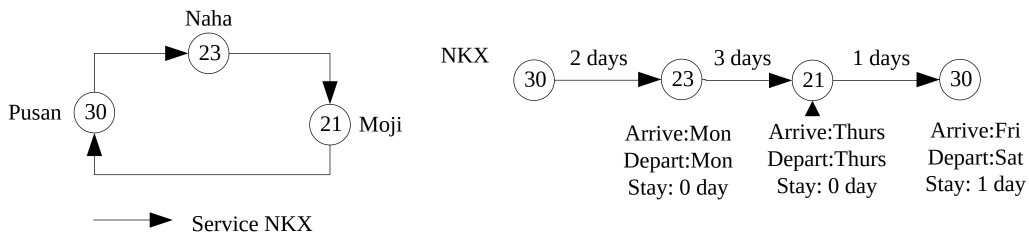


Figure 4.2 the network of problem 1: one service route and three ports

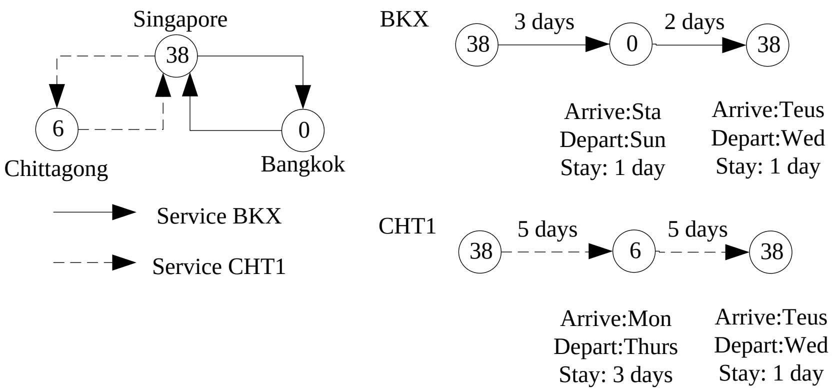


Figure 4.3 the network of problem 2: two service routes and three ports

Given a case, we try to numerically find the optimal values of a given policy by brute-force enumerating all the possible designs, in which the best design of thresholds could be included. For all designs, we use optimal computing budget allocation algorithm (Chen *et al.*, 2000) to allocate the budget for each design in order to save the computation time. The results are presented in Table 4.2.

Table 4.2 Minimum costs for both problems

Distribution	Problem 1			Problem 2		
	SL	TL	Improve %	SL	TL	Improve %
Normal	18215.48	17837.51	2.08%	6346.866	6211.838	2.13%
Exponential	48301.9	44187.07	8.52%	13092.32	11998.18	8.36%

(SL: single-level threshold policy; TL: two-level threshold policy)

To facilitate the comparison of the two-level policy and the single-level policy, the percentage of total cost reduction achieved by the two-level threshold policy from the single-level threshold policy is presented in the column “Improve %”. From Table 4.2, it can be seen that two-level threshold policy outperforms over the single-level threshold policy in all cases. The cost reduction percentages achieved by the two-level threshold policy under exponential distribution is significant more than those under normal distribution. It implies that two-level threshold policy could be more effective to manage ECs, especially in the problem with high uncertainty.

4.4 Summary

In this chapter, we consider ECR problem in the liner shipping system with multi-service. A simulation model is formulated in time-driven way with considering uncertain customer demand, customer supply and residual capacity on vessel for ECs. Two threshold policies, i.e., a single-level threshold policy and a two-level threshold policy, are developed to manage ECs. We investigate the performance of both policies in several small size problems by numerical experiments. The results show that the two-level threshold policy outperforms over the single-level threshold policy, especially in the system with high uncertainty. Hence, focusing on the two-level threshold policy, we extend the study to optimize the parameters of the policy in next chapter.

Chapter 5 COMPASS WITH HYBRID SAMPLING FOR EMPTY CONTAINER REPOSITIONING IN LINER SHIPPING SYSTEM

We continue previous work to optimize the parameters of the two-level threshold policy. A search algorithm based on convergent optimization via most-promising-area stochastic search method is utilized to solve our optimization problem. A hybrid coordinate and gradient sampling scheme with simultaneous perturbation stochastic approximation gradient estimator is proposed to improve the sampling scheme in search method in terms of search efficiency.

5.1 Introduction

From the literature, few studies consider the optimization of the parameters of inventory-based policies in the practical problem with multiple service routes. Dong and Song (2009) considered the joint fleet sizing and ECR problem in a multi-vessel, multi-port and multi-voyage system. A simulation-based optimization tool using Genetic Algorithm (GA) was developed to optimize the fleet size and the parameters of the thresholds of policy. However, only single service route is considered in their study. Song and Dong (2010) studied an ECR policy with flexible destination ports in a multi-service system. The threshold values of the policy were determined heuristically based on demand statistic and service frequency. Their numerical results presented that the performance of the policy was quite sensitive to the threshold values. Hence, it is interesting to find the optimal parameters of the inventory-based policy and to investigate the performance of such policy in practical problem with multiple service routes. From (4.14), the problem, which is to optimize parameters of a given policy in terms of minimizing the expected total cost per period, can be formulated as

$$min\{J(\boldsymbol{\theta})\}, \boldsymbol{\theta} \in \Theta = \Phi \cap \mathcal{I}^d \tag{5.1}$$

where $\boldsymbol{\theta}$ is the vector of the decision variables (also called a solution), i.e., the vector of the parameters of the inventory-based threshold policy; Θ is the feasible region; Φ is a convex and compact set in $\Re^d$, which is the of d-dimensional vectors with real elements, and Φ is defined by a set of linear constraints; $\mathcal{I}^d$ is the set of d-dimensional vectors with integer elements; d is the number of parameters of the given policy. In this study, we consider that Θ is finite.

In this study, we only consider the two-level threshold policy to manage ECs. Similar with the problem in Chapter 3, this problem is also an optimization via simulation problem. However, the gradient-based IPA algorithm is not suitable for this optimization problem. The major reason is that the decision variables are integer, which does not satisfy the basic assumption for IPA technique. Besides, even that the integer assumption could be relaxed; the IPA gradient is not intuitive to be obtained in the complex liner shipping system. Hence, we turn to consider other search algorithms.

Note that the thresholds of the policy in problem (5.1) should be integer. Our problem is further classified to the discrete optimization via simulation (DOvS) problem. There have been a number of approaches to DOvS problems proposed in the research literature (see Nelson (2010) for a review). When the solution space is large, adaptive random search dominates the research literature (Nelson, 2010). Most of adaptive random search algorithms converge to the global optimum, such as the stochastic ruler method of Yan and Mukai (1992) and nested partitions algorithm of

Shi and Ólafsson (2000). However, such algorithms do not scale up efficiently for practical problems with a large number of feasible solutions, which implies that global convergence has little practical meaning (Nelson, 2010). Besides, Dong and Song (2009) point that even though the standard deviation of simulation outputs over multiple samples is small, their proposed GA-based algorithm may still produce a local optimal solution. Hence, for problem (5.1), it is more practical to find the local optimum than the global optimum.

In this study, we develop a search algorithm based on a locally adaptive random search method, i.e., COMPASS method, to solve problem (5.1). A hybrid sampling scheme, i.e., HCGS scheme, with Simultaneous Perturbation Stochastic Approximation (SPSA) gradient estimator is proposed to speed up the convergence rate of the search algorithm.

Next, we present the literature review about COMPASS, followed by the details of the COMPASS algorithm with SPSA-based HCGS scheme.

5.2 Literature review

The COMPASS method is a randomized optimization method that has been recently developed for local optimization of DOvS problems (Hong and Nelson, 2006). The main idea of the COMPASS method is to construct the most promising area (MPA) that most likely contains the best solution around the current sample-best solution, then uniformly sample new solutions from the MPA and evaluate all of the chosen solutions. Hong and Nelson (2006) have showed that COMPASS can solve both constrained and partially constrained or unconstrained DOvS problems with integer-ordered decision variables. And with probability 1, the algorithm converges to the set of locally optimal solutions under some mild conditions. Since its convention, COMPASS has received increasing attention and was applied to solve various problems, such as the project management problem (Kuhl and Tolentino-Peña, 2008), the multi-objective problem (Lee *et al.*, 2011).

Hong and Nelson (2007) later designed a framework for locally convergent random search algorithms, where they showed that the local convergence can be easily achieved. They also proposed a revised COMPASS to speed up the convergence of COMPASS by modifying the estimate scheme of original COMPASS, i.e., simulating only a subset of all chosen solutions instead of all chosen solutions.

The revised COMPASS was embedded in software called Industrial Strength COMPASS (ISC) to solve practical problems (Xu *et al.*, 2010). ISC has three phases: global search, local search and final clean up. In the first global search phase, a niching genetic algorithm is utilized to identify several regions possible with competitive locally optimal solutions. In the second local search phase, COMPASS with constraint pruning technique is adapted to find the locally optimal solution for each of the regions identified in the global phase. In the third cleanup phase, a two-stage ranking & selecting procedure is applied to select the best solution among all identified locally solutions and estimate the true value of the selected solution. Numerical experiments showed that the ISC algorithm was competitive with OptQuest (which is the most popular commercial optimization via simulation software products) for problems up to dimension 10 in terms of finite-time performance. However, ISC dramatically slowed down when dimensionality increased beyond 10.

The low convergence rate of COMPASS has hindered its application in practical problems. Hence, Hong *et al*. (2010) investigated why COMPASS slows down for high-dimensional problem. They reported that the uniform sampling scheme is a primary reason for the deterioration of COMPASS for high-dimensional problems, since it is not efficient to find better solutions from the MPA. Then, they proposed a coordinate sampling (CS) scheme to alleviate this problem so as to speed up the convergence rate of COMPASS. Both empirical and analytical evidence showed that compared with uniform sampling scheme, CS scheme is a much more efficient sampling scheme.

The main idea of CS scheme is to sample a solution that differs in only one coordinate from the current sample best solution. The following is a high-level description of CS scheme:

Step 1. Randomly choose a coordinate axis.

Step 2. Draw a line passing through the current best sample solution and parallel to the chosen coordinate axis.

Step 3. Uniformly sample a solution with integer elements on the line from step 2 within the MPA.

It can be seen that CS scheme just randomly chooses the coordinate axis to sample new solution without utilizing any information of the structure of the cost function, such as gradient information. Therefore, there is some room to improve the efficiency of the sampling scheme by making use of gradient information.

5.3 COMPASS algorithm with SPSA-based HCGS scheme

In this section, we develop COMPASS algorithm with SPSA-based HCGS scheme to solve problem (5.1). We introduce the procedure of COMPASS algorithm, followed by the proposed HCGS scheme.

5.3.1. COMPASS algorithm

Some notations used in COMPASS algorithm are presented in Table 5.1.

Table 5.1 List of notations for COMPASS algorithm

	Description
$\mathcal{V}(k)$	The set of all solutions visited through iteration k
$\boldsymbol{\theta}_k^*$	The solution with the smallest aggregated sample mean among all $\boldsymbol{\theta} \in \mathcal{V}(k)$
$\mathcal{C}_k$	$= \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \Theta$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}_k^*\| \leq \|\boldsymbol{\theta} - \varphi\| \ \forall \boldsymbol{\varphi} \in \mathcal{V}(k)$and $\boldsymbol{\varphi} \neq \boldsymbol{\theta}_k^*\}$, the MPA at iteration k
m	The number of solutions to sample in each iteration
$a_k(\boldsymbol{\theta})$	The additional observations allocated to $\boldsymbol{\theta}$ at iteration k
$N_k(\boldsymbol{\theta})$	$= \sum_{i=0}^{k} a_k(\boldsymbol{\theta})$, the total number of observations on solution $\boldsymbol{\theta}$ at iteration k for every $\theta \in \mathcal{V}(k)$
$\bar{J}_k(\boldsymbol{\theta})$	The sample mean of all $N_k(\boldsymbol{\theta})$ observations of $Y(\boldsymbol{\theta}, \omega)$ at iteration k

The MPA $\mathcal{C}_k$ consists of all feasible solutions that are at least as close to the current sample best solution as to other visited solution. It is the part of feasible region Θ, which is assumed to be finite in this study. The procedure of COMPASS algorithm (Hong and Nelson, 2006) used in this study is as follows:

The COMPASS algorithm

Step 1. Set iteration count $k = 0$. Find $\boldsymbol{\theta}_0 \in \Theta$, set $\mathcal{V}(0) = \{\boldsymbol{\theta}_0\}$ and $\boldsymbol{\theta}_k^* = \boldsymbol{\theta}_0$. Determine $a_0(\boldsymbol{\theta}_0)$ according to a simulation-allocation rule (SAR). Take $a_0(\boldsymbol{\theta}_0)$ observation from $\boldsymbol{\theta}_0$, set $N_0(\boldsymbol{\theta}_0) = a_0(\boldsymbol{\theta}_0)$, and calculate $\bar{J}_0(\boldsymbol{\theta}_0)$. Let $\mathcal{C}_0 = \Theta$.

Step 2. Let $k = k + 1$. Independently sample $\boldsymbol{\theta}_{k1}, \boldsymbol{\theta}_{k2}, \ldots, \boldsymbol{\theta}_{km}$ using the SPSA-based HCGS sampling scheme from $\mathcal{C}_{k-1}$. Remove any duplicates from $\boldsymbol{\theta}_{k1}, \boldsymbol{\theta}_{k2}, \ldots, \boldsymbol{\theta}_{km}$, and let $\mathcal{V}_k$ be the remaining set. Then, we have $\mathcal{V}(k) = \mathcal{V}(k-1) \cup \mathcal{V}_k$. Determine $a_k(\boldsymbol{\theta})$ according to SAR for every $\boldsymbol{\theta}$ in $\mathcal{V}(k)$. For all $\boldsymbol{\theta} \in \mathcal{V}(k)$, take $a_k(\boldsymbol{\theta})$ observation, and then update $N_k(\boldsymbol{\theta})$ and $\bar{J}_k(\boldsymbol{\theta})$.

Step 3. Let $\boldsymbol{\theta}_k^* = \arg min_{\boldsymbol{\theta} \in \mathcal{V}(k)} \bar{J}_k(\boldsymbol{\theta})$. Construct $\mathcal{C}_k$ and go to step 1.

At the end of iteration, if there is more than one solution having the smallest aggregated sample mean, we select $\boldsymbol{\theta}_k^*$ randomly from the set of solutions having the smallest aggregated sample mean.

5.3.2. SPSA-based HCGS scheme

In this section, we first introduce the HCGS scheme, followed by presenting the SPSA gradient estimator.

5.3.2.1. HCGS scheme

The HCGS scheme is based on the CS scheme (Hong *et al.*, 2010) but the gradient information is added by generating a special solution with a gradient-based scheme. Hence, two kinds of solutions constitute the set of new solutions in the HCGS scheme: the *gradient solution* and the *CS solutions*. The former is a single solution generated by a gradient-based scheme while the latter are generated by the CS scheme. For the gradient-based scheme, it uniformly samples a solution with integer elements following the direction of the negative gradient from the current best sample solution within the MPA.

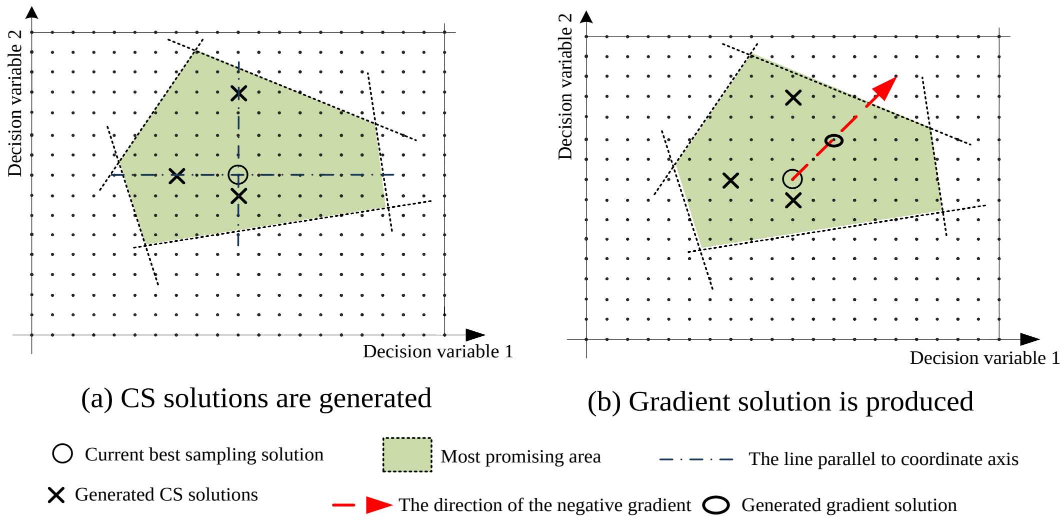


Figure 5.1 New solutions generation process in HCGS scheme

Let $\mathbf{g}(\boldsymbol{\theta})$ be the gradient at solution $\boldsymbol{\theta}$. In general, the procedure of solution generation at iteration k by HCGS scheme is summarized in Figure 5.1 with following three steps:

HCGS scheme at iteration k

Step 1. (Initialization): set the set of sampled solutions $L = \emptyset$, and the number of sampled solution $j = 0$.

Step 2. (Sample CS solutions): repeat the following sub-steps until $j = m - 1$

(a) Uniformly sample an integer i from 1 to d

(b) Determine the values of r_{min} and r_{max} that are the minimum and maximum integer values of r that ensure $\boldsymbol{\theta}^*_{k-1} + r \cdot \mathbf{e}_i \in \mathcal{C}_{k-1}$, respectively. $\mathbf{e}_i$ is the ith column of a $d \times d$ identity matrix.

(c) Uniformly sample an integer r from$[r_{min}, r_{max}]$. Let $j = j + 1$, and we have one solution $\boldsymbol{\theta}_{kj} = \boldsymbol{\theta}^*_{k-1} + r\mathbf{e}_i$. Let $L = L \cup \{\boldsymbol{\theta}_{kj}\}$.

Step 3. (Sample a gradient solution): Let $\boldsymbol{\theta}'$ be the point going from current best sample solution $\boldsymbol{\theta}^*_k$ in the direction of the negative gradient $-\mathbf{g}(\boldsymbol{\theta}^*_{k-1})$, i.e., $\boldsymbol{\theta}' = \boldsymbol{\theta}^*_{k-1} - r \cdot \mathbf{g}(\boldsymbol{\theta}^*_{k-1})$. We denote r^G_{max} as the maximum value of r that ensures $\boldsymbol{\theta}'$ is within the MPA $\mathcal{C}_{k-1}$ Then, uniformly sample a real number r from $(0, r^G_{max})$. Let $j = j + 1$, and we have a solution

$$\boldsymbol{\theta}_{kj} = \Pi_{\Theta}\left(\Pi_{\text{Int}}\left(\boldsymbol{\theta}^*_{k-1} - r \cdot \mathbf{g}(\boldsymbol{\theta}^*_{k-1})\right)\right) \tag{5.2}$$

And we return $L = L \cup \{\boldsymbol{\theta}_{kj}\}$.

Note that the generated solution is constrained to be with integer components. Hence, in Eq. (5.2), Π_{Int} represents an operation to adjust $\boldsymbol{\theta}'$ to be a solution with integer elements. If either ith component of $\boldsymbol{\theta}'$, i.e., θ'_i is not an integer, we always round it down to the next integer, i.e., $\lfloor \theta'_i \rfloor$ if $\mathrm{g}_i(\boldsymbol{\theta}^*_{k-1}) < 0$, and round it up to the next integer, i.e., $\lceil \theta'_i \rceil$ if $\mathrm{g}_i(\boldsymbol{\theta}^*_{k-1}) > 0$. $\mathrm{g}_i(\boldsymbol{\theta}^*_{k-1})$ is the ith component of the gradient $\mathbf{g}(\boldsymbol{\theta}_{k-1}^*)$. Then, if $\Pi_{\text{Int}}(\boldsymbol{\theta}') \notin \Theta$, we project it into the nearest point in the feasible region by operation $\Pi_\Theta(.)$.

As the new scheme samples some solutions by the CS scheme, we can show that the proposed scheme still satisfies the required condition for ensuring the local convergence (Hong and Nelson, 2007). The proposition is as follows.

PROPOSITION 5.1*: The HCGS scheme ensures that* $\Pr\{\boldsymbol{\theta} \in \mathcal{V}_k\} > 0$ *for all* $\boldsymbol{\theta} \in \mathcal{LN}(\boldsymbol{\theta}_{k-1}^*) \cap \mathcal{C}_k$*, where* $\mathcal{LN}(\boldsymbol{\theta})$ *denotes the local neighborhood of a solution* $\boldsymbol{\theta}$ *and includes all feasible solutions that have Euclidean distance* 1 *from* $\boldsymbol{\theta}$, i.e., $\mathcal{LN}(\boldsymbol{\theta}) = \{\boldsymbol{\upsilon} \in \Theta: \|\boldsymbol{\theta} - \boldsymbol{\upsilon}\| = 1\}$.

PROOF: Denote the n unique solutions in $\mathcal{V}_k$ as $\boldsymbol{\theta}_{k1}, \boldsymbol{\theta}_{k2}, \dots, \boldsymbol{\theta}_{kn}$, which are independently sampling within $\mathcal{C}_{k-1}$. The first $n-1$ solutions are CS solutions and the last one is a gradient solution. For all $\boldsymbol{\theta} \in \mathcal{LN}(\boldsymbol{\theta}_{k-1}^*)$, we have $\Pr\{\boldsymbol{\theta} \in \mathcal{V}_k\} = 1 - \Pr\{\boldsymbol{\theta} \notin \mathcal{V}_k\} = 1 - \Pr\{\boldsymbol{\theta}_{k1} \neq \boldsymbol{\theta}, \boldsymbol{\theta}_{k2} \neq \boldsymbol{\theta}, \dots, \boldsymbol{\theta}_{kn} \neq \boldsymbol{\theta}\} = 1 - \Pr\{\boldsymbol{\theta}_{k1} \neq \boldsymbol{\theta}\}^{n-1} \cdot \Pr\{\boldsymbol{\theta}_{kn} \neq \boldsymbol{\theta}\}$.

When we sample a CS solution, every coordinate axis has a probability of $1/d$ being chosen, and every coordinate axis has at most $|\mathcal{C}_{k-1}|$ feasible solutions, where $|\mathcal{C}_{k-1}|$ dentoes the number of solutions within the MPA in the $k-1$ iteration. Note that $\mathcal{C}_{k-1}$ is the part of feasible region Θ, which is finite in our study. Thus, we can have $\Pr\{\boldsymbol{\theta}_{k1} = \boldsymbol{\theta}\} \geq \frac{1}{d|\mathcal{C}_{k-1}|} \geq \frac{1}{d|\Theta|}$. Further, we obtain that $\Pr\{\boldsymbol{\theta} \in \mathcal{V}_k\} \geq 1 - \Pr\{\boldsymbol{\theta}_{k1} \neq \boldsymbol{\theta}\}^{n-1} \geq 1 - \left(1 - \frac{1}{d|\Theta|}\right)^{n-1} > 0$. ■

The HCGS scheme exploits the advantage of both the CS scheme and the gradient-based scheme. It maintains the local convergence property of COMPASS and speeds up the convergence rate by utilizing the gradient information. Next, we present the method to obtain the gradient, i.e., $\mathbf{g}(\boldsymbol{\theta}_{k-1}^{*})$ for the HCGS scheme.

5.3.2.2. SPSA-based gradient

As a general rule, if direct gradient information is conveniently and readily available, it is generally to one's advantage to use this information, such as our gradient method with IPA estimator in Chapter 3. However, for many problems, direct gradient estimates may not be readily available. In this case, gradient estimates based on (noisy) measurements of performance measure itself are the only recourse. Hence, in this study, the SPSA algorithm is applied to estimate the gradient for the HCGS scheme.

SPSA algorithm is originally developed by Spall (1992) for the problems in the multivariate Kiefer-Wolfowitz setting. Since its convention, it has received considerable attention and been used in a wide variety of setting such as statistical parameter estimation, adaptive control and many other applications (Spall, 2003). For the details literature about SPSA, see the webpage http://www.jhuapl.edu/spsa/. The basis of SPSA method is an efficient and intuitive "simultaneous perturbation" estimation of the gradient from only two noisy function measurements. Compared with other gradient-base method, SPSA has two advantages: it requires only two simulations per gradient estimate, regardless of the number of input variables, and it can treat the simulation model as a black box, i.e., no knowledge of the working of the system is required (Fu *et al.*, 2005).

We can obtain one gradient at the best solution $\boldsymbol{\theta}_{k}^{*}$ by SPSA method with the following procedure:

Basic SPSA procedure

Step 1. Randomly generate the perturbation vector $\boldsymbol{\Delta} = [\Delta_1, \ldots, \Delta_d]^T$, where Δ_i is independent Bernoulli random variables taking the values ± 1 with probability $1/2$.

Step 2. Evaluate two measurements of the cost function $Y(.)$ in Eq. (4.14) based on the simultaneous perturbation around the current $\boldsymbol{\theta}_k^*$ with one replication, and compute the approximation to the gradient by

$$\mathbf{g}(\boldsymbol{\theta}_k^*) = \frac{Y(\boldsymbol{\theta}_k^* + c\boldsymbol{\Delta}) - Y(\boldsymbol{\theta}_k^* - c\boldsymbol{\Delta})}{2c} \boldsymbol{\Delta}^{-1} \tag{5.3}$$

Note that we only consider solution with integer components. Thus, the above procedure requires several modifications:

- In the original procedure of SPSA method (Spall, 1992), the parameter c is decreasing sequence with respect to the iteration of simulation, such as k in COMPASS algorithm. Although it may change according to the problem, it is suitable to define it as a fixed value (Gerencsér *et al.*, 1999; Gerencsér *et al.*, 2001; Hill *et al.*, 2004) for the discrete problem and we use $c = 1$ in the study. This, together with the choice of $\boldsymbol{\Delta}$ makes sure that all evaluates are on the integer lattice on which we optimize.

- Evaluations have to stay within the feasible set area. If both $\boldsymbol{\theta}_k^* \pm c\boldsymbol{\Delta}$ lie outside the feasible region, we try to generate a new perturbation vector, with which at least one of the evaluations lies inside the feasible region. And the number of maximum time to try is given. If only one evaluation lies inside, one-sided SPSA (Chen *et al.*, 1999) is applied to calculate magnitude of the gradient. That is, we have $\mathbf{g}(\boldsymbol{\theta}_k^*) = \frac{Y(\boldsymbol{\theta}_k^*+c\boldsymbol{\Delta})-\bar{J}_k(\boldsymbol{\theta}_k^*)}{c}\boldsymbol{\Delta}^{-1}$ or $\frac{Y(\boldsymbol{\theta}_k^*-c\boldsymbol{\Delta})-\bar{J}_k(\boldsymbol{\theta}_k^*)}{-c}\boldsymbol{\Delta}^{-1}$, when there is only point $\boldsymbol{\theta}_k^* + c\boldsymbol{\Delta}$ or $\boldsymbol{\theta}_k^* - c\boldsymbol{\Delta}$ inside the feasible region.

Besides, from Spall (1992), we make another modification on the basic SPSA procedure. That is, instead of computing only one search direction $\boldsymbol{\Delta}$ and evaluating the cost function in forward and backward direction, we generates several, say S, search directions, resulting in $2S$ function evaluations. By incorporating all these computations, we can obtain an average gradient as

$$\bar{\mathbf{g}}(\theta_k^*) = \frac{1}{S}\sum_{s=1}^{S} \mathbf{g}^{(s)}(\boldsymbol{\theta}_k^*) \tag{5.4}$$

where $\mathbf{g}^{(s)}(\theta_k^*)$ is a gradient calculated by Eq. (5.3) with a random perturbation vector $\boldsymbol{\Delta}$. We then use the average gradient $\bar{\mathbf{g}}(\theta_k^*)$ as gradient in Eq. (5.2) to generate the gradient solution. Despite the expense of the additional function evaluations, it could allow a better approximation of the true gradient (Spall, 1998) and will thus provide a better gradient solution.

5.4 Numerical experiments

In this section, we first evaluate the performance of the COMPASS algorithm with SPSA-based HCGS scheme using a test function. Then, the proposed method is used to solve one ECR problem. The proposed algorithm is coded in Microsoft Visual studio C# 2010. All the numerical studies are tested on a PC with 2.67GHz Processor under the Microsoft Vista Operation System.

5.4.1. Performance of COMPASS algorithm with SPSA-based HCGS scheme

The sphere function is a common test function as defined in (5.5).

$$J_{\text{sphere}}(\boldsymbol{\theta}) = \sum_{i=1}^{d} \theta_i^2 \tag{5.5}$$

It has a unique global minimum occurs at $\theta_i^* = 0\ \forall i = 1,2,\dots,d$ with $J_{\text{sphere}}(\boldsymbol{\theta}^*) = 0$. Set $-100 \le \theta_i \le 100\ \forall i = 1,2,\dots,d$, we use this function to compare the performances of COMPASS algorithm with different sampling schemes, i.e., CS scheme and HCGS scheme.

Since there is no stochastic noise in the cost function, we only need to evaluate a newly chosen solution once to know its objective value. This allows us to eliminate the effect of the estimation scheme and to focus solely on the sampling scheme. We examine two performance measures. One is the number of visited solutions that the algorithm evaluates until it first hits the global optimum $\boldsymbol{\theta}^*$, and the other is the CPU time that the algorithm spends until it first hit $\boldsymbol{\theta}^*$. The first measure quantifies the amount of simulation effort required by COMPASS, and the second measure quantifies the amount of computational overhead in running COMPASS (because additional function evaluations are needed to estimate the gradient for the gradient-based sampling scheme).

The search algorithms are started with the same initial solution $\theta_i = 80\ \forall i = 1,2,\ldots,d$, and stopped when the MPA is singleton. Repeat the process for 10 replications; we then report the average performance. In all numerical experiments, we set $m = 5$, i.e., we sample five solutions from the MPA in each iteration.

First, to evaluate the effectiveness of the HCGS scheme, we compare the performance of the COMPASS algorithm with CS scheme with that with HCGS scheme by varying the dimension from 2 to 50. For HCGS scheme, the true gradient information of sphere function, i.e., $\partial J_{\text{sphere}}(\boldsymbol{\theta}^*)/\partial\theta_{ki}^* = 2\theta_{ki}^*\ \forall i = 1,2,\ldots,d$ is used to generate the gradient solution in Eq. (5.2).

The results are presented in Table 5.2, where “Number of evaluated solutions” and “CPU time” mean the average number of solutions evaluated by COMPASS and the average CPU time, respectively, until the algorithm first finds the optimal solution. The CPU time is measured in milliseconds.

Table 5.2 CS vs. HCGS with true gradient

Dimension	Number of evaluated solutions		CPU time	
	CS	HCGS-true	CS	HCGS-true
2	42.1	28.4	5.9	5.3
5	124.8	41.6	21.3	4.8
10	319.8	50.7	170.3	9.2
15	513.8	50.4	508.7	10.5
20	789.1	60.4	1442.8	20.3
30	1337.4	92.6	4398.3	56.9
40	2015.6	122.2	11423.1	132.2
50	2580.4	160.5	22407.9	290.7

From Table 5.2, it is clear that the COMPASS algorithm with HCGS scheme is much more efficient in solving the problem than that with CS scheme, especially when the dimension of the problem is high. When the dimension is 2, the difference between two algorithms is negligible. When the dimension is 15, compared to the COMPASS algorithm with CS scheme, the COMPASS algorithm with HCGS scheme evaluates fewer than 10% of the solutions and spends less than 3% of the computation overhead. The gradient solution in the gradient-based sampling scheme contributes to the drastic differences. The results demonstrate the effectiveness of the HCGS scheme and indicate that when the gradient has enough accuracy, the gradient solution could greatly improve the convergence rate of the COMPASS algorithm.

Next, we compare the performances of the COMPASS algorithm with CS scheme and that with SPSA-based HCGS scheme. For the SPSA-based HCGS scheme, we vary the number of random directions S from 1 to 6. Table 5.3 and Table 5.4 show the results.

From Table 5.3 and Table 5.4, it can be seen that the COMPASS algorithm with HCGS scheme outperforms over that with CS scheme when $S \geq 3$. When $S = 2$, COMPASS algorithm with HCGS scheme evaluates fewer solutions than that with CS

Table 5.3 CS vs. HCGS with SPSA gradient in terms of number of evaluated solutions

Dimension	CS	SPSA-based HCGS					
		$S = 1$	$S = 2$	$S = 3$	$S = 4$	$S = 5$	$S = 6$
2	42.1	37.8	29.1	31.8	28.5	31	30.6
5	124.8	138.9	115.3	104.8	98.7	80.8	81.8
10	319.8	355	301.8	263.5	221	196	189.7
15	513.8	572	516.1	452.8	390	364.7	318.7
20	789.1	858.7	751.6	644.9	583.5	457.8	468.9
30	1337.4	1412.3	1307.1	1175.3	1040.2	871.1	789.9
40	2015.6	2046.2	1989.5	1656	1533.3	1346.9	1227.3
50	2580.4	2835.4	2602.9	2382.6	2068	1850.9	1698.6

Table 5.4 CS vs. HCGS with SPSA gradient in terms of CPU time

Dimension	CS	SPSA-based HCGS					
		$S = 1$	$S = 2$	$S = 3$	$S = 4$	$S = 5$	$S = 6$
2	5.9	8.7	4.5	2.2	2.2	1.3	1.5
5	21.3	28.9	18.8	16.6	16.4	9.8	8.9
10	170.3	232.3	179.6	136	89.4	75.1	71.6
15	508.7	683.1	563.1	460.8	401.3	284.1	221.1
20	1442.8	1788.4	1405.4	1037.5	857.5	541.4	556.6
30	4398.3	5587	4844	4013.3	3136	2310	1927.1
40	11423.1	13690.7	13382.2	9474.6	7924.2	6221.3	5169
50	22407.9	29496.4	25207.4	20670.1	15440.8	12974.5	11104.8

scheme for some cases; however, it needs a little more computational overhead to find the optimal solution. One possible explanation for such issue is that HCGS scheme needs additional computational overhead to evaluate the SPSA gradient in each iteration. When we set the appropriable value for S, the COMPASS algorithm with HCGS scheme is much more efficient in solving the problem than that with CS scheme, especially when the dimension of the problem is high. Hence, the results demonstrate the effectiveness of the SPSA gradient approximation. Besides, it shows that the more random directions we use to average, the better HCGS scheme with SPSA gradient could perform. However, since the number of evaluations for SPSA is increased with S, there is a tradeoff between the accuracy of gradient approximation and the computation time cost.

5.4.2. Application for ECR problem

In this section we show the performance of the COMPASS algorithm with SPSA-based HCGS scheme for a large size ECR problem.

The problem has 12 ports with 4 service routes. The network for this ECR problem is presented in Figure 5.2. The unit costs of holding, leasing, unloading and loading are given in Table 5.5, which are from real data. We set the transportation cost $c^x_{t,s,v} = 10$.

Table 5.5 The cost parameters for the ECR problem

Port number	7	12	14	16	22	24	29	33	34	38	41	48
Holding	1.5	1.5	1.5	5	5	1.5	1.5	1.5	1.5	1.5	5	5
Leasing	438	540	369	246	246	303	191	303	438	261	246	246
Unloading	82	107	73	50	50	61	63	61	82	52	50	50
Loading	82	107	73	50	50	61	63	61	82	52	50	50

The daily customer demands, supplies and residual capacities of vessels are assumed to follow normal distribution. For example, we assume that the daily demand of port i, i.e., $\zeta_{t,i}$ follows normal distribution $N\left(\overline{D}_{i,t}, \left(0.2\overline{D}_{i,t}\right)^2\right)$ and is left-truncated at zero. The mean values of daily supply and demand for each port are represented in Table 5.6 and Table 5.7, respectively

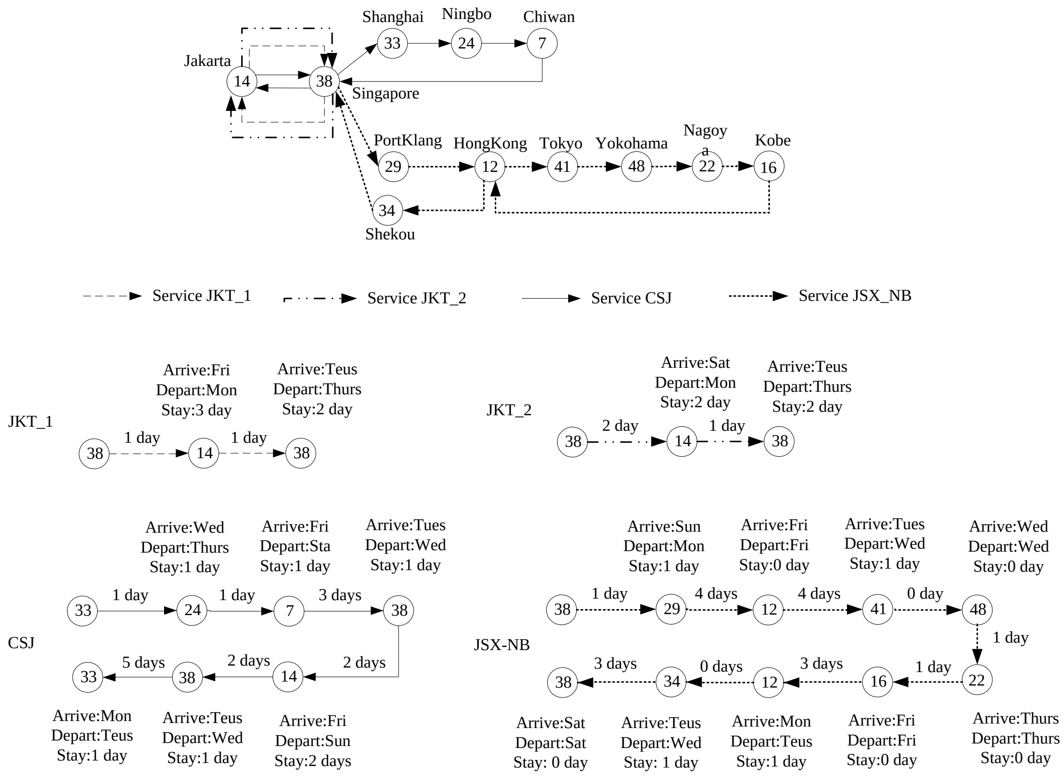


Figure 5.2 the network of ECR problem: 4 service routes and 12 ports

Table 5.6 Average daily customer supply for the ECR problem

Day	Port number											
	7	12	14	16	22	24	29	33	34	38	41	48
Mon	82	128	119	183	70	105	228	0	0	263	94	0
Tues	41	128	119	183	70	0	228	71	0	263	94	0
Wed	41	128	119	0	70	0	228	71	270	229	187	0
Thurs	41	128	119	0	70	105	228	71	271	240	94	219
Fri	41	128	20	0	140	105	0	71	271	240	94	219
Sat	82	128	119	183	140	105	0	71	0	178	94	219
Sun	82	128	99	183	140	105	0	71	0	263	94	0
Total	410	896	714	732	700	525	912	426	812	1676	751	657

Table 5.7 Average daily customer demand for the ECR problem

Day	Port number											
	7	12	14	16	22	24	29	33	34	38	41	48
Mon	60	174	34	93	104	118	104	95	153	212	135	87
Tues	60	110	0	187	208	118	104	0	153	212	135	175
Wed	60	110	0	187	208	118	104	0	0	212	68	87
Thurs	60	155	166	187	104	0	104	95	0	212	68	87
Fri	60	110	214	93	104	0	104	95	153	212	68	87
Sat	60	110	248	93	104	0	104	95	153	212	68	87
Sun	60	110	83	93	104	118	104	95	153	212	135	87
Total	420	879	745	933	936	472	728	475	765	1484	677	697

For the simulation, we set a warm-up period of 40 weeks and then average total cost over the next 120 weeks. For the COMPASS algorithm, we generate 8 new solutions at each iteration and set the number of random directions $S = 8$. To allocate the observation for each solution, we use an equal SAR with $N_k(\boldsymbol{\theta}) = N_k$ for all $\theta \in \mathcal{V}(k)$, where $N_k = max\{15, \lceil 15 \cdot (\log k)^{1.01} \rceil\}$. The search algorithm can be stopped when the current best sampling solution, i.e., $\boldsymbol{\theta}_k^*$ does not change for concessive 50 iterations.

The line in Figure 5.3 is the sample path of the COMPASS algorithm with SPSA-based HCGS scheme averaged over 30 macroreplications for the ECR problem (only the results of the first 100 iterations are shown). It shows the convergence of the proposed algorithm.

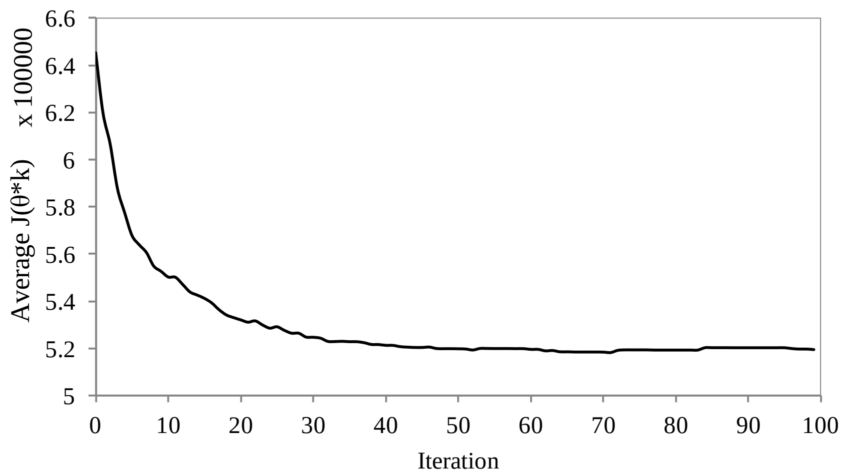


Figure 5.3 Average $J(\theta_k^*)$ by COMPASS with SPSA-based HCGS scheme for the ECR problem

The best solution, i.e., the best thresholds of all ports, is given in Table 5.8 (based on a single replication). It can be seen that those import-dominated ports, such as port 24, port 29 and port 41 usually have lower threshold values, and those export-dominated ports, such as port 14, port 16 and port 22 and port 48 have higher threshold values. It is in agreement with the intuition that setting high thresholds in export-

Table 5.8 The best solution for the ECR problem

	Port number											
	7	12	14	16	22	24	29	33	34	38	41	48
Lower-level	17	89	53	221	303	25	19	50	64	159	72	214
Upper-level	96	235	808	247	335	138	94	188	287	206	113	272

dominated ports and low threshold in import-dominated ports could encourage repositioning EC from import-dominated ports to export-dominated ports. Besides, it is observed that the thresholds of port 12, port 34 and port 38, which are import-dominated ports, are also high. This can be explained by the high variability of demand and supply in those ports. For port 16 and port 22, which have similar weekly demands, port 22 has higher thresholds than port 16 because it has less weekly supplies. And for port 12 and port 29, which have similar weekly supplies, port 12 has higher thresholds because it has higher weekly demands. All the results show the reasonability of the best solution.

5.5 Summary

In this section, we develop a search algorithm based on COMPASS to optimize the parameters of the two-level threshold policy in liner shipping system with multiple service routes. To improve the convergence rate of COMPASS, we propose a SPSA-based HCGS scheme to sample the new solutions in each iteration. The result demonstrates that the proposed sampling scheme can significantly improve the convergence rate of COMPASS. Then, we solve a practical ECR problem by the COMPASS algorithm with SPSA-based HCGS scheme. The result shows the convergence of the proposed algorithm and the reasonability of the best solution.

Chapter 6 CONCLUSION AND FUTURE RESEARCH

6.1 Conclusion

This thesis studied ECR problem with inventory-based control policies. It contributes to the implementation of inventory-based control policies in general and complex liner shipping systems and some methodological issues to optimize the parameters of such policies.

Chapter 3 studied ECR problem in a multi-port system consisting of ports connected to each other. A single-level threshold policy with intelligent rule in terms of minimizing repositioning cost was proposed to manage ECs. Then, we aimed to optimize the fleet size and the parameters of the given policy. When analyzing the optimization problem, we found one interesting property of the problem, i.e., when the fleet size is equal to the sum of thresholds, the optimal values of the thresholds only depend on the holding and leasing cost function. It implies that the optimization problem can be reduced to a newsvendor problem, when the fleet size is equal to the sum of the thresholds. Meanwhile, we also proved mathematically that it is worth to study the scenario in which the fleet size is not equal to the sum of the thresholds, since this scenario could produce lower operation cost compared to the scenario in which the fleet size is equal to the sum of the thresholds. This provides an important insight that keeping more (less) ECs over the threshold in import-dominated (export-dominated) port in advance when it becomes a deficit (surplus) port could reduce the repositioned in (out) quantity of ECs.

Since there is no closed-form formulation of the expected cost function when the fleet size is not equal to the sum of the thresholds, we built a simulation model and proposed a gradient-driven search algorithm to tackle the problem. For the gradient-driven search algorithm, utilizing the knowledge of inside structure of the simulation model, we designed an efficient gradient estimator by IPA technique. In the procedure to obtain the IPA gradient estimator, we developed a modified stepping stone technique to explore the perturbations on ports. It is innovative and provides a potential methodology contribution in the field of application of the stepping stone method. In the numerical runs, we demonstrated the effectiveness of the proposed policy and provided some insights for liner operators in managing ECs.

In Chapter 4, we built a simulation model for a practical ECR problem with multi-service, and developed two threshold-type policies to manage ECs. The simulation model and the policies can be used by the shipping company analysts to explore other operation options in the future. The experiment results showed that the two-level threshold policy outperforms over the single-level threshold policy, especially in the systems with high uncertainty.

Chapter 5 further discussed ECR problem with multiple service routes and focused on optimizing the parameters of the two-level threshold policy. A search algorithm based on COMPASS was developed to solve the optimization problem, which provides a potential methodology contribution to the application of COMPASS in complex systems. To improve the convergence of COMPASS, we proposed a hybrid sampling scheme, i.e., the HCGS scheme by taking use of the gradient information. Considering that the simulation model was a black-box in this problem, SPSA technique was applied to estimate the gradient for the HCGS scheme. The results showed the effectiveness of the proposed algorithm. And the HCGS scheme could be easily applied in other random search algorithms to speed up the convergence rate. Besides, a numerical example was offered to demonstrate the convergence of the proposed algorithm and to show the reasonability of the best solution.

6.2 Future research topics

Despite the contributions described above, the research presented in this thesis has some inevitable limitations. Future research related to the topics reported in this thesis may be carried out in the areas listed below.

In the liner shipping system with multiple service routes, it is possible that multiple vessels will visit one port within one week. Hence, future research should attempt to study the policies with considering the priority for each service route based on each port. Besides, we assumed that all of the leased containers will be returned after T^r periods. This assumption is quite limited because when we need to return is an important decision for ECR. Therefore, we can relax this assumption in the future work.

Although we have demonstrated the effectiveness of the HCGS scheme, other sampling schemes which make use of the gradient information are worth to study further. In addition, the use of both standard and foldable for ECR is also worth to study.

BIBLIOGRAPHY

Abrache, J., T. Crainic and M. Gendreau, 1999. A new decomposition algorithm for the deterministic dynamic allocation of empty containers. Transportation Research, 11(3): 49-99.

Bandeira, D.L., J.L. Becker and D. Borenstein, 2009. A dss for integrated distribution of empty and full containers. Decision Support Systems, 47(4): 383-397.

Belmecheri, F., T. Cagniart, L. Amodeo, F. Yalaoui and C. Prins, 2009. Modelling and optimization of empty container reuse: A real case study. In: Computers & Industrial Engineering, 2009. CIE 2009. International Conference on. IEEE: pp: 1106-1109.

Bourbeau, B., T.G. Crainic and B. Gendron, 2000. Branch-and-bound parallelization strategies applied to a depot location and container fleet management problem. Parallel Computing, 26(1): 27-46.

Cassandras, C.G., X.C. Ding and X. Lin, 2011. An optimal control approach for the persistent monitoring problem. In: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on. IEEE: pp: 2907-2912.

Chang, H., H. Jula, A. Chassiakos and P. Ioannou, 2006. Empty container reuse in the los angeles/long beach port area. In: Proceedings of the National Urban Freight Conference.

Charnes, A. and W.W. Cooper, 1954. The stepping stone method of explaining linear programming calculations in transportation problems. Management Science, 1(1): 49-69.

Chen, C. and Q. Zeng, 2010. Designing container shipping network under changing demand and freight rates. Transport, 25(1): 46-57.

Chen, C.H., J. Lin, E. Yücesan and S.E. Chick, 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. Discrete Event Dynamic Systems, 10(3): 251-270.

Chen, H.F., T.E. Duncan and B. Pasik-Duncan, 1999. A kiefer-wolfowitz algorithm with randomized differences. Automatic Control, IEEE Transactions on, 44(3): 442-453.

Chen, M., J.Q. Hu and M.C. Fu, 2010. Perturbation analysis of a dynamic priority call center. Automatic Control, IEEE Transactions on, 55(5): 1191-1196.

Cheung, R.K. and C.Y. Chen, 1998. A two-stage stochastic network model and solution methods for the dynamic empty container allocation problem. Transportation Science, 32(2): 142-162.

Chew, E.P., L.H. Lee and Y.L. Lau, 2009. Component allocation and ordering policy in a multi-component, multi-product assembled to stock system. OR Spectrum, 32(2): 293-317.

Choong, S.T., M.H. Cole and E. Kutanoglu, 2002. Empty container management for intermodal transportation networks. Transportation Research Part E, 38(6): 423-438.

Crainic, T., M. Gendreau and P. Dejax, 1993a. Dynamic and stochastic models for the allocation of empty containers. Operations Research, 41(1): 102-126.

Crainic, T., M. Gendreau, P. Soriano and M. Toulouse, 1993b. A tabu search procedure for multicommodity location/allocation with balancing requirements. Annals of Operations Research, 41(4): 359-383.

Crainic, T.G., P. Dejax and L. Delorme, 1989. Models for multimode multicommodity location problems with interdepot balancing requirements. Annals of Operations Research, 18(1): 277-302.

Crainic, T.G. and L. Delorme, 1993. Dual-ascent procedures for multicommodity location-allocation problems with balancing requirements. Transportation Science, 27(2): 90-101.

Debye, P., 1909. Näherungsformeln für die zylinderfunktionen für grosse werte des arguments und unbeschränkt veränderliche werte des index. Mathematische Annalen, 67(4): 535-558.

Di Francesco, M., T.G. Crainic and P. Zuddas, 2009. The effect of multi-scenario policies on empty container repositioning. Transportation Research Part E, 45(5): 758-770.

Dong, J.X. and D.P. Song, 2009. Container fleet sizing and empty repositioning in liner shipping systems. Transportation Research Part E, 45(6): 860-877.

Dong, J.X. and D.P. Song, 2009. Quantifying the impact of inland transport times on container fleet sizing in liner shipping services with uncertainties. OR Spectrum, 34(1): 155-180.

Du, Y. and R. Hall, 1997. Fleet sizing and empty equipment redistribution for center-terminal transportation networks. Management Science, 43(2): 145-157.

Economic and social commission for Asia and the Pacific, 2007. Regional shipping and port development: Container traffic forecast 2007 update. Available from http://www.unescap.org/ttdw/publications/tis_pubs/pub_2484/pub_2484_fulltext.pdf.

Erera, A.L., J.C. Morales and M. Savelsbergh, 2009. Robust optimization for empty repositioning problems. Operations Research, 57(2): 468-483.

Feng, C.M. and C.H. Chang, 2008. Empty container reposition planning for intra-asia liner shipping. Maritime Policy & Management, 35(5): 469-489.

Francesco, D.i.M.D., 2007. New optimization models for empty container management.

Fu, M. and J.Q. Hu, 1997. Conditional monte carlo: Gradient estimation and optimization applications. Kluwer Academic Publishers Norwell, MA.

Fu, M.C., 2002. Optimization for simulation: Theory vs. Practice. INFORMS Journal on Computing, 14(3): 192-215.

Fu, M.C., 2008. What you should know ab``out simulation and derivatives. Naval Research Logistics (NRL), 55(8): 723-736.

Fu, M.C., F.W. Glover and J. April, 2005. Simulation optimization: A review, new developments, and applications. In: Simulation Conference, 2005 Proceedings of the Winter. IEEE: pp: 83-95.

Gendron, B. and T.G. Crainic, 1995. A branch-and-bound algorithm for depot location and container fleet management. Location Science, 3(1): 39-53.

Gendron, B. and T.G. Crainic, 1997. A parallel branch-and-bound algorithm for multicommodity location with balancing requirements. Computers & Operations Research, 24(9): 829-847.

Gerencsér, L., S.D. Hill and Z. Vágó, 1999. Optimization over discrete sets via spsa. In: Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future. ACM, New York, NY, USA: pp: 466-470.

Gerencsér, L., S.D. Hill and Z. Vágó, 2001. Discrete optimization via spsa. In: American Control Conference, 2001. Proceedings of the 2001. IEEE: pp: 1503-1504.

Hill, S.D., L. Gerencsér and Z. Vágó, 2004. Stochastic approximation on discrete sets using simultaneous difference approximations. In: American Control Conference, 2004. Proceedings of the 2004. IEEE: pp: 2795-2798

Ho, Y.C. and X.R. Cao, 1991. Discrete event dynamic systems and perturbation analysis. Boston, Massachusetts: Kluwer Academic Publishers.

Hong, L.J. and B.L. Nelson, 2006. Discrete optimization via simulation using compass. Operations Research, 54(1): 115-129.

Hong, L.J. and B.L. Nelson, 2007. A framework for locally convergent random-search algorithms for discrete optimization via simulation. ACM Transactions on Modeling and Computer Simulation (TOMACS), 17(4): 19.

Hong, L.J., B.L. Nelson and J. Xu, 2010. Speeding up compass for high-dimensional discrete optimization via simulation. Operations Research Letters, 38(6): 550-555.

Iglehart, D.L., 1963. Optimality of (s, s) policies in the infinite horizon dynamic inventory problem. Management Science, 9(2): 259-267.

Imai, A. and F. Rivera, 2001. Strategic fleet size planning for maritime refrigerated containers. Maritime Policy & Management, 28(4): 361-374.

Imai, A., K. Shintani and S. Papadimitriou, 2009. Multi-port vs. Hub-and-spoke port calls by containerships. Transportation Research Part E: Logistics and Transportation Review, 45(5): 740-757.

Jula, H., A. Chassiakos and P. Ioannou, 2006. Port dynamic empty container reuse. Transportation Research Part E, 42(1): 43-60.

Kuhl, M.E. and R.A. Tolentino-Peña, 2008. A dynamic crashing method for project management using simulation-based optimization. In: Simulation Conference, 2008. WSC 2008. Winter. pp: 2370-2376.

Lai, K.K., K. Lam and W.K. Chan, 1995. Shipping container logistics and allocation. The Journal of the Operational Research Society, 46(6): 687-697.

Lam, S.W., L.H. Lee and L.C. Tang, 2007. An approximate dynamic programming approach for the empty container allocation problem. Transportation Research Part C, 15(4): 265-277.

Lee, L.H., E.P. Chew and H.B. Li, 2011. Multi-objective compass for discrete optimization via simulation. In: Simulation Conference (WSC), Proceedings of the 2011 Winter. pp: 4065-4074.

Li, J.A., S.C.H. Leung, Y. Wu and K. Liu, 2007. Allocation of empty containers between multi-ports. European Journal of Operational Research, 182(1): 400-412.

Li, J.A., K. Liu, S.C.H. Leung and K.K. Lai, 2004. Empty container management in a port with long-run average criterion. Mathematical and Computer Modelling, 40(1-2): 85-100.

Liu, X., H.Q. Ye and X.M. Yuan, 2011. Tactical planning models for managing container flow and ship deployment. Maritime Policy & Management, 38(5): 487-508.

Long, Y., L.H. Lee and E.P. Chew, 2012. The sample average approximation method for empty container repositioning with uncertainties. European Journal of Operational Research, 222(1): 65-75.

Meng, Q. and S. Wang, 2011. Liner shipping service network design with empty container repositioning. Transportation Research Part E: Logistics and Transportation Review, 47(5): 695-708.

Moon, I., A.-D. Do Ngoc and R. Konings, 2013. Foldable and standard containers in empty container repositioning. Transportation Research Part E: Logistics and Transportation Review, 49(1): 107-124.

Moon, I.K., A.D.D. Ngoc and Y.S. Hur, 2010. Positioning empty containers among multiple ports with leasing and purchasing considerations. OR Spectrum, 32(3): 765-786.

Nelson, B.L., 2010. Optimization via simulation over discrete decision variables. Tutorials in operations research, @ 2010 INFORMS: 193-207.

Olivo, A., P. Zuddas, M. Di Francesco and A. Manca, 2005. An operational model for empty container management. Maritime Economics & Logistics, 7(3): 199-222.

Özdemir, D., E. Yücesan and Y.T. Herer, In press. Production, manufacturing and logistics multi-location transshipment problem with capacitated production. European Journal of Operational Research.

Robbins, H. and S. Monro, 1951. A stochastic approximation method. The Annals of Mathematical Statistics, 22(3): 400-407.

Shen, W.S. and C.M. Khoong, 1995. A dss for empty container distribution planning. Decision Support Systems, 15(1): 75-82.

Sheopuri, A., G. Janakiraman and S. Seshadri, 2010. New policies for the stochastic inventory control problem with two supply sources. Operations Research, 58(3): 734-745.

Shi, L. and S. Ólafsson, 2000. Nested partitions method for global optimization. Operations Research, 48(3): 390-407.

Shintani, K., A. Imai, E. Nishimura and S. Papadimitriou, 2007. The container shipping network design problem with empty container repositioning. Transportation Research Part E, 43(1): 39-59.

Song, D. and J. Dong, 2011. Flow balancing-based empty container repositioning in typical shipping service routes. Maritime Economics & Logistics, 13(1): 61-77.

Song, D.P., 2005. Optimal threshold control of empty vehicle redistribution in two depot service systems. IEEE Transactions on Automatic Control, 50(1): 87-90.

Song, D.P., 2007. Characterizing optimal empty container reposition policy in periodic-review shuttle service systems. Journal of the Operational Research Society, 58(1): 122-133.

Song, D.P. and J. Carter, 2008. Optimal empty vehicle redistribution for hub-and-spoke transportation systems. Naval Research Logistics, 55(2): 156 - 171.

Song, D.P. and J.X. Dong, 2008. Empty container management in cyclic shipping routes. Maritime Economics & Logistics, 10(4): 335-361.

Song, D.P. and J.X. Dong, 2010. Effectiveness of an empty container repositioning policy with flexible destination ports. Transport Policy, 18(1): 92-101.

Song, D.P. and C.F. Earl, 2008. Optimal empty vehicle repositioning and fleet-sizing for two-depot service systems. European Journal of Operational Research, 185(2): 760-777.

Song, D.P., J. Zhang, J. Carter, T. Field, J. Marshall, J. Polak, K. Schumacher, P. Sinha-Ray and J. Woods, 2005. On cost-efficiency of the global container shipping network. Maritime Policy & Management, 32(1): 15-30.

Song, D.P. and Q. Zhang, 2010. A fluid flow model for empty container repositioning policy with a single port and stochastic demand. SIAM Journal on Control and Optimization, 48(5): 3623-3642.

Spall, J.C., 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. Automatic Control, IEEE Transactions on, 37(3): 332-341.

Spall, J.C., 1998. Implementation of the simultaneous perturbation algorithm for stochastic optimization. Aerospace and Electronic Systems, IEEE Transactions on, 34(3): 817-823.

Spall, J.C., 2003. Introduction to stochastic search and optimization: Estimation, simulation, and control. John Wiley and Sons.

Suri, R., 1989. Perturbation analysis: The state of the art and research issuesexplained via the gi/g/1 queue. Proceedings of the IEEE, 77(1): 114-137.

Tekin, E. and I. Sabuncuoglu, 2004. Simulation optimization: A comprehensive review on theory and applications. IIE Transactions, 36(11): 1067-1081.

United Nations Conference on Trade and Development, 2011. Review of maritime transport. Available from http://unctad.org/en/Docs/rmt2011_en.pdf.

United Nations Conference on Trade and Development, 2012. Review of maritime transport. Available from http://unctad.org/en/PublicationsLibrary/rmt2012_en.pdf.

Wardi, Y., R. Adams and B. Melamed, 2010. A unified approach to infinitesimal perturbation analysis in stochastic flow models: The single-stage case. Automatic Control, IEEE Transactions on, 55(1): 89-103.

Wong, W.P., W. Jaruphongsa and L.H. Lee, 2011. Budget allocation for effective data collection in predicting an accurate dea efficiency score. Automatic Control, IEEE Transactions on, 56(6): 1235-1246.

Xu, J., B.L. Nelson and J. Hong, 2010. Industrial strength compass: A comprehensive algorithm and software for optimization via simulation. ACM Transactions on Modeling and Computer Simulation (TOMACS), 20(1): 3.

Yan, D. and H. Mukai, 1992. Stochastic discrete optimization. SIAM Journal on Control and Optimization, 30(3): 594-612.

APPENDICES

Appendix A. Input data for problem 1 in chapter 3

Table A.1 The values of the cost parameters for problem 1 in Chapter 3

Original port p	Destination port m	$C^R_{p,m}$	C^H_p	C^L_p
1	2	8.421	4.396	16.537
	3	9.917		
2	1	6.761	2.374	12.395
	3	5.848		
3	1	9.027	1.900	24.599
	2	7.053		

Table A.2 Average customer demands in different trade imbalance patterns for problem 1 in Chapter 3

Original port p	Destination port m	$\mu_{p,m}$		
		Balanced	Moderately imbalanced	Severely imbalanced
1	2	95.950	191.900	287.850
	3	126.385	252.771	379.155
2	1	95.950	95.950	95.950
	3	176.833	176.833	176.833
3	1	126.385	126.385	126.385
	2	176.833	176.833	176.833

Appendix B. Input data for problems 1 and 2 in Chapter 4

Table B.1 Average daily customer supply and demand for problem 1 in Chapter 4

Day	The daily supply in the port			The daily demand in the port		
	P-21	P-23	P-30	P-21	P-23	P-30
Mon	8	15	19	21	9	0
Tues	8	15	19	21	9	0
Wed	8	15	19	21	9	25
Thurs	8	15	19	21	9	29
Fri	17	15	0	21	9	29
Sat	17	15	19	21	9	0
Sun	17	15	19	21	18	0

Table B.2 The values of the cost parameters for problem 1 in Chapter 4

Port name	Port number	Holding	Leasing	Unloading	Loading
Moji	21	5	246	50	50
Naha	23	5	246	50	50
Pusan	30	1.5	246	57	57

Table B.3 Average daily customer supply and demand for problem 2 in Chapter 4

Day	The daily supply in the port			The daily demand in the port		
	P-0	P-6	P-38	P-0	P-6	P-38
Mon	13	7	18	8	8	38
Tues	13	7	0	8	8	38
Wed	7	7	20	8	16	0
Thurs	7	7	20	8	8	0
Fri	7	7	20	8	8	0
Sat	7	7	20	8	8	0
Sun	13	7	20	8	8	38

Table B.4 The values of the cost parameters for problem 2 in chapter 4

Port name	Port number	Holding	Leasing	Unloading	Loading
Bangkok	0	1.5	126	25	25
Chittagong	6	1.5	270	54	54
Singapore	38	1.5	261	52	52