Sample Average Approximation for Stochastic Empty Container Repositioning

Ek Peng CHEW[a], Loo Hay LEE[b], Yin LONG[c]
Department of Industrial & Systems Engineering
National University of Singapore
Singapore 119260
e-mail: [a]isecep@nus.edu.sg; [b] iseleelh@nus.edu.sg; [c]iselongy@nus.edu.sg

***Abstract*—To incorporate uncertainties in empty container repositioning problem, we formulate a two-stage stochastic programming model with random demand, supply, ship weight capacity and ship space capacity. The Sample Average Approximation (SAA) method is applied to approximate the expected value function. Several non-independent and identically distributed sampling schemes are considered to enhance the performance of the SAA method. Numerical experiments show that near-optimal solutions could be provided by the SAA method with these sampling schemes.**

Keywords-empty container repositioning; simulation optimization; sample average approximation; supersaturated design

I. INTRODUCTION

One main issue in the containerized transportation is the imbalance of container flow, which is the result of global trade imbalance between different regions. Thus, maintaining higher operational cost efficiencies in repositioning empty containers becomes a crucial issue in shipping industry.

There are increasing studies on empty container flows in recent years [1][2]. In the maritime transportation, container operators have to deal with some uncertain factors like the real transportation time between two ports/deports, future demand and supply, the in-transit time of returning empty container from customers, and the available capacity in vessels for empty containers transportation, etc. The uncertain nature of parameters is taken into account in several studies [3][4]. A multi-scenario model was proposed in [5] to address the Empty Container Repositioning (ECR) problem in a scheduled maritime system. In their study, opinions of shipping companies were considered to generate scenarios when the distributions of uncertain parameters cannot be estimated through historical data. Our study focuses on developing scenario-based model when the distribution of uncertain parameters can be estimated through historical data. Random scenarios could be generated based on these distributions. However, it is difficult to solve the stochastic ECR problem with a large number of scenarios. In this case, we apply the Sample Average Approximation (SAA) method to solve the stochastic ECR problem with multiple scenarios.

The SAA problem with multiple scenarios is usually difficult to solve due to its large scale. In this study, we try to enhance the SAA method with well-planned samplings which are more representative. The motivation of this idea is to get acceptable solutions by solving SAA problems with a small number of scenarios. Independent and identically distributed (i.i.d) sampling is well studied for construction approximations [6][7]. On the other hand, SAA with non-i.i.d samplings is also studied in recent year [8][9]. Empirically, it was shown that Latin Hypercube (LH) and Antithetic Variates (AV) methods outperform those under i.i.d sampling, with LH outperforming AV [10]. In [11], U design was used to further enhance the accuracy of the SAA and their theoretical results showed that the SAA with U designs can significantly outperform those with LH designs. In this study, we try to do as a whole very few experiments (even less than the number of degrees of freedom of the system when that is possible) and still get a satisfying approximation. To our knowledge, no existing study applies the SAA method with non-i.i.d samplings to the stochastic ECR problem where the distributions of uncertain parameters are known. This study is to fill in this gap.

The paper is structured in the following way. Section I concerns introduction; In Section II, we provide the description of a basic deterministic model and our two-stage stochastic model for ECR. Section III shows the solving methodologies to solve our proposed model. The SAA method and the sampling schemes to enhance SAA are explained. Section IV presents the results of computational studies. Finally, we give conclusions and outline directions for future research in Section V.

II. PROBLEM FORMULATION

The focus of this study is to make operational level maritime ECR decisions for shipping companies. As the global container transportation network is large and complex, the ocean liners usually divide the global network into several regions and appoint regional operators to manage the container flows for each region. Because of the long lead time of the across-region empty containers, ocean liners usually make ordering decisions depending on forecasting to make decisions on ordering. Due to the booking system used in the maritime transportation, demand, supply and the available ship capacity in the near future could be forecasted accurately. However, it is difficult to obtain accurate forecasting for more than one or two weeks. Currently, container operators make decisions based on the nominal forecast value. Because of the differences between the expected value and the realized value, inefficient solutions

 ISBN: 978-1-61208-161-8

may be produced. These decisions have to be recovered at real-time operations.

A. The Deterministic Model

If all parameters in the planning horizon are known, the deterministic time space model for ECR could be formulated as follows, where the actual service schedule and most port requirements are considered. Details of this model could be found in [12].

$$\text{Min} \sum_{t=1,2,...,T} TC^t$$

$$\begin{aligned} TC^t = & \sum_{k\in K}\sum_{i\in P} c^u_{t,i,k} \sum_{(s,v)\in\{(s,v)|v\in V, s\in S_v, p_{v,s}=i, t\in A_{v,s}\}} u_{t,s,v,k} \\ & + \sum_{k\in K}\sum_{i\in P} c^w_{t,i,k} \sum_{(s,v)\in\{(s,v)|v\in V, s\in S_v, p_{v,s}=i, t\in D_{v,s}\}} w_{t,s,v,k} \\ & + \sum_{i\in P}\sum_{k\in K} (c^y_{t,i,k} y_{t,i,k} + c^z_{t,i,k} z_{t,i,k}) \\ & + \sum_{k\in K} \sum_{(s,v)\in\{(s,v)|v\in V, s\in S_v, t\in D_{v,s}\}} c^x_{t,s,v,k} x_{t,s,v,k} \end{aligned}$$

Subject to

$$\sum_{k\in K}(g_k \times x_{t,s,v,k}) \le \gamma_{t,s,v} \quad \forall\ (v,s,t)\in\{(v,s,t)|v\in V, s\in S_v, t\in D_{v,s}\} \quad (1)$$

$$\sum_{k\in K}(h_k \times x_{t,s,v,k}) \le \sigma_{t,s,v} \quad \forall\ (v,s,t)\in\{(v,s,t)|v\in V, s\in S_v, t\in D_{v,s}\} \quad (2)$$

$$x_{t-b_{v,s},s,v,k} - u_{t,s+1,v,k} + w_{t+d_{v,s},s+1,v,k} = x_{t+d_{v,s},s+1,v,k} \quad \forall\ (v,s,t)\in\{(v,s,t)|v\in V, s\in S_v, t\in D_{v,s}\} \quad (3)$$

$$x_{t-b_{v,s},s,v,k} \ge u_{t,s+1,v,k} \quad \forall k\in K, \forall (v,s,t)\in\{(v,s,t)|v\in V, s\in S_v, t\in A_{v,s+1}\} \quad (4)$$

$$y_{t-1,i,k} + \sum_{(s,v)\in\{(s,v)|v\in V, s\in S_v, p_{v,s}=i, t\in A_{v,s}\}} u_{t,s,v,k} - \psi_{t,i,k} + z_{t,i,k} + \theta_{t,i,k} - \sum_{(s,v)\in\{(s,v)|v\in V, s\in S_v, p_{v,s}=i, t\in D_{v,s}\}} w_{t,s,v,k} = y_{t,i,k} \quad \forall\ k\in K, \forall i\in P, \forall t=1,2,...,T \quad (5)$$

$$u_{t,s,v,k}, w_{t,s,v,k}, x_{t,s,v,k}, y_{t,i,k}, z_{t,i,k} \ge 0 \quad (6)$$

The objective function is to minimize the total operation cost in the planning horizon. TC^t is the total operational cost at time *t*. The operational costs include the handling cost (unloading cost and loading cost), the holding cost, the penalty cost and the transportation cost. $u_{t,s,v,k}$, $w_{t,s,v,k}$, $x_{t,s,v,k}$, $y_{t,i,k}$, and $z_{t,i,k}$ are decision variables, where $u_{t,s,v,k}$ is the number of empty containers of size k unloaded at stop *s* from service *v* at time *t*, $w_{t,s,v,k}$ is the number of empty containers of size k loaded from stop *s* onto service *v* at time *t*, $x_{t,s,v,k}$ is the number of empty containers of size *k* transported from stop *s* to next stop on service *v* leaving stop *s* at time *t*, $y_{t,i,k}$ is the number of empty containers of size *k* stored at port *i* at time *t*, and $z_{t,i,k}$ is the number of empty containers of size *k* that cannot be satisfied by the empty containers stored at port *i* at time *t*. $c^u_{t,i,k}$, $c^w_{t,i,k}$, $c^x_{t,s,v,k}$, $c^y_{t,i,k}$, and $c^z_{t,i,k}$ are the corresponding cost parameters of unloading, loading, transportation, storing and penalty respectively. *V* is the set of services. *P* is the set of ports. *Q* is the set of regions. *K* is the set of container sizes. And S_v is the set of stops on service *v*. $D_{v,s}$ is the set of periods in which service *v* departs from its stop *s*. $A_{v,s}$ is the set of periods in which service *v* arrives at its stop *s*.

Constraint (1) and constraint (2) are ship capacity constraints. $\gamma_{t,s,v}$ is residual space capacity on service *v* when it leaves stop *s* at time *t*, and $\sigma_{t,s,v}$ is residual weight capacity on service *v* when it leaves stop *s* at time *t*. g_k and h_k are volume and weight of one container of size *k*. These two constraints should be considered when there is a service leaving the port. Constraint (3) guarantees the balance of the container flows at each service. $b_{v,s}$ is the transportation time from stop *s* to next stop on the service *v*. $d_{v,s}$ is the number of days that the service *v* stays at stop *s*. Constraint (4) ensures that the number of empty containers unloaded from a vessel should not exceed the total number of empty containers in the vessel. These two constraints should be considered when there is a service arriving at a port. Constraint (5) considers the balance of the container flows at each port at each time. $\theta_{t,i,k}$ is the supply of empty containers of size *k* in port *i* at time *t*, and $\psi_{t,i,k}$ is the demand of empty containers of size *k* in port *i* at time *t*. $p_{v,s}$ is the port corresponding to the stop *s* on service *v*. Constraint (6) ensures that all variables are non-negative.

B. The Stochastic Model

In this study, we develop a stochastic programming model which takes account into four uncertain parameters, i.e. the demand (the empty containers that picked up by the customers to load cargos), the supply (the empty containers that returned by the customers), the available ship space capacity and available ship weight capacity for empty containers. Other uncertain factors like the transportation time between two ports are not considered. We also do not consider container substitution in this study. We assume that service schedule is given and fixed in the planning horizon. This assumption is valid as the planning horizon of our operation model is short (several weeks), and the service schedule is not changed frequently. In order to incorporate the deterministic information and the uncertain information, a two-stage stochastic programming is developed. This model is run in a rolling horizon manner. ECR decisions are made at the beginning of stage 1 and will be made again when new information is collected.

Let $\omega \in \Omega$ denotes a scenario that is unknown when decisions at stage 1 are made, but that is known when the decisions at stage 2 are made, where Ω is the set of all scenarios. A two-stage stochastic model for ECR is formulated as follows.

Stage 1

$$\min \quad g(x_1) = c_1 x_1 + E_p[Q(x_1, \xi(\omega))] \tag{7}$$

$$\text{subject to} \quad A_1 x_1 = a_1 \tag{8}$$

$$B_1 x_1 = v \tag{9}$$

$$x_1 \geq 0 \tag{10}$$

Stage 2: For a realized scenario ω, we have

$$Q[x_1, \xi(\omega)] = \min \quad c_2\, x_2(\omega) \tag{11}$$

$$\text{subject to} \quad A_2 x_2(\omega) = a_2(\omega) \tag{12}$$

$$B_2 x_2(\omega) = v(x_1) \tag{13}$$

$$x_2(\omega) \geq 0 \tag{14}$$

x_1: Decisions at stage 1

$x_2(\omega)$: Decisions for scenario ω at stage 2 given x_1

c_1, c_2 : The cost vector at stage 1 and stage 2 respectively

A_1, B_1, A_2, B_2: The coefficient matrices of x_1 or x_2

a_1, $a_2(\omega)$: The RHS of constraint (8) and constraint (12) respectively

v: The vector of end container states of stage 1. It is the empty container inventory at each port and at each vessel at the end of stage 1. $v = \{v_1, v_2, ..., v_K\}$

$v(x_1)$: The vector of initial container states of stage 2 given x_1

The objective function of stage 1 is to minimize the total operational cost in the planning horizon. $c_1 x_1$ is the operation cost at stage 1. $E_p[Q(x_1, \xi(\omega))]$ is the expected cost at stage 2, where p is the probability distribution of uncertain parameters. We assume that the probability distribution p on Ω is known in the stage 1. $Q(x_1, \xi(\omega))$ is the objective function of stage 2, which is the operational cost at stage 2 given x_1 and scenario ω. Constraint (8) and constraint (12) includes the typical constraints of ECR problem in the deterministic model, i.e. ship capacity constraint, service flow constraint, and port flow constraint. Constraint (9) is to set the end container states of stage 1, which are also the initial container states of stage 2. Constraint (13) is to set the initial container states of stage 2.

III. SOLVING METHODOLOGY

Our stochastic problem is hard to solve as it is difficult to evaluate the expected cost of stage 2 for a given x_1, i.e. $E_p[Q(x_1, \xi(\omega))]$. It requires the solutions of a large number of stage 2 optimization problems. In this study, we consider applying the SAA method to solve the stochastic ECR problem. The basic idea of SAA method is that the expected objective function of the stochastic problem is approximated by a sample average estimate derived from a random sample and the resulting SAA problem could then be solved by deterministic optimization techniques [6].

A. The SAA Method

A sample with N scenarios { ω^1, ω^2 ,…, ω^N } is generated according to the probability distribution p. The SAA problem is formulated as follows.

$$\hat{g}_N = \min \quad c_1 x_1 + \frac{1}{N}\sum_{n=1}^{N}[c_2 x_2(\omega^n)] \tag{15}$$

Subject to (8), (9); and (12), (13) *for n=1,2,…,N*

$$v \geq 0,\ x_1 \geq 0,\ x_2(\omega^n) \geq 0 \quad \textit{for } n=1,2,\dots,N \tag{16}$$

The SAA problem can then be solved by deterministic optimization methods. The optimal solution $\hat{x}$ and the optimal value $\hat{g}_N$ of the SAA problem could be obtained. $\hat{x}$ is a candidate solution of the true problem. Note that when this sample is an i.i.d random sample of the random vector, $\hat{g}_N(x)$ is called a (standard) Monte Carlo estimator of $g(x)$. To evaluate the objective value given the candidate solution $\hat{x}$, we consider generating an independent sample with N' scenarios, where N' is much larger than N. Let

$$\hat{g}_{N'}(\hat{x}) = \min \quad c_1\, \hat{x} + \frac{1}{N'}\sum_{n=1}^{N'} Q[\hat{x}, \xi(\omega^n)] \tag{17}$$

$\hat{g}_{N'}(\hat{x})$ is defined to estimate the objective value $g(\hat{x})$ of an feasible solution $\hat{x}$.

In order to get a better solution, we can generate M independent samples equally with N scenarios. By solving the corresponding M independent SAA problems, we can get M candidate solutions $\hat{x}_N^1$, $\hat{x}_N^2$, …, $\hat{x}_N^M$ and the objective values $\hat{g}_N^1$, $\hat{g}_N^2$, …, $\hat{g}_N^M$. It is natural to take $\hat{x}^*$ as one of the optimal solutions of these SAA problems which provides the smallest estimated objective value,

$$\hat{x}^* \in \arg\min\{\hat{g}_{N'}(\hat{x}_N) : \hat{x}_N \in \{\hat{x}_N^1, \hat{x}_N^2, ..., \hat{x}_N^M\}\} \tag{18}$$

To estimate the performance of SAA method, we need to calculate the optimality gap, i.e. the difference between the lower bound and the upper bound. This gap can be used to evaluate the quality of the solution. As $\hat{x}^*$ is a feasible solution of the stochastic ECR problem, $\hat{g}_{N'}(\hat{x}^*)$ gives an estimate of the upper bound of the true optimal objective value of the true problem. On the other hand, as N realized scenarios are considered in the SAA problems, the objective value of the SAA problem $\hat{g}_N$ has a negative bias. Let $\bar{g}_N^M$ denotes the average objective value of the M SAA problems,

$$\bar{g}_N^M = \frac{1}{M}\sum_{m=1}^{M}\hat{g}_N^m \tag{19}$$

$\bar{g}_N^M$ provides a statistical estimate for a lower bound of the true optimal value of the true problem. The optimality gap could be estimated as

 ISBN: 978-1-61208-161-8

$$\hat{g}_{N'}(\hat{x}^*) - \bar{g}_N^M \qquad (20)$$

B. Non-i.i.d Samplings

Due to its large scale, the SAA problem (15)-(16) for the real scale ECR is difficult to solve. In this case, the sampling should be well-planned. We try to generate samplings with a small number of scenarios (the number of scenarios is even less than the random variables of the stochastic ECR problem) and still get acceptable solutions. In this study, three sampling schemes are considered to enhance the performance of the SAA method for the stochastic ECR problem.

1) Latin Hypercube (LH) Sampling: In computer experiments, it is well know that LH design achieves maximum stratification in one-dimensional projections. The idea is to partition the sample space, and the number of sample points on each region should be proportional to the probability of that region. This way we ensure that the number of sampled points on each region will be approximately equal to the expected number of points to fall in that region.

2) AG Design: AG deign is a supersaturated design which is introduced in [13]. One good property of the AG design is that the saturation increase rather fast with the number of scenarios. Besides, for a two-level design with m scenarios and n factors ($m \times n$), each column has the same number of -1's and 1's in an even case. This property is necessary for a stable regression analysis as each variable has to be evaluated fairly from its smallest values to its highest values.

3) AGLH Design: We also propose a superstaturated design which combines the AG deisn and LH sampling, e.g., a two-level case, we can generate the AGLH design as follows,

a) Generate a AG design, B

b) Randomly permute the rows, columns and symbols of B ($m \times n$)

c) In each columns of B, replace the m/2 0s by a uniform random permutation of 1,…,m/2. The m/2 1s by a uniform random permutation of m/2+1,…,m.

d) Coupling B with U[0,1] random variables and we can get our desire design, C.

IV. NUMERICAL STUDY

To evaluate the performance of the SAA method, we first generate an ECR transportation network as shown in Fig. 1. Five ports, three services, and one type of container (twenty-foot standard container) are considered. The planning horizon is three weeks, and we define the first week as stage 1 and the second and the third weeks are stage 2. All information in the first week is known when decisions in stage 1 are made, while some parameters in stage 2 are unknown when decisions in stage 1 are made. These parameters are known when decisions in stage 2 are made. The lead time of across-region empty containers is one week. The service schedules are given in Fig. 1.

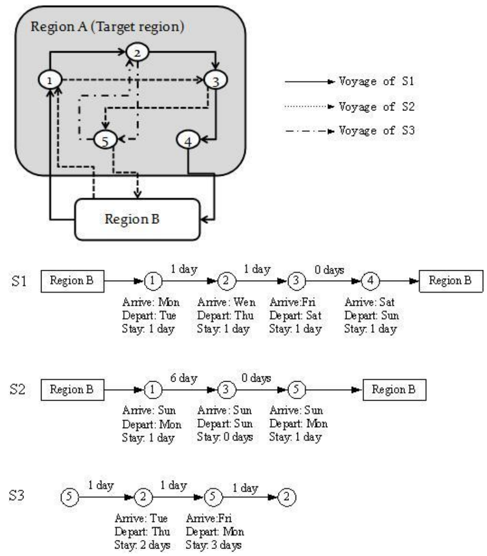

Figure 1. A network with three services and five ports

We apply the SAA method to solve the two-stage stochastic problem (with i.i.d sampling). We can solve the SAA problem directly by using CPLEX11.2 when the sample size N is not too large (N < 1000). We set the sample size N as 100. The number of scenarios to evaluate the solution N' is set to be 1000. Replication number is set to be 20. The performance of the SAA method (N = 100) for the small scale case is examined with the key results shown in Table I. As shown in Table I, the estimated objective value of the true problem $\hat{g}_{N'}(\hat{x}^*)$ is 3359.97, and a statistic lower bound for the objective value $\bar{g}_N^M$ is 3335.45. The optimality gap $\hat{g}_{N'}(\hat{x}^*) - \bar{g}_N^M$ is 24.52 (0.73% of $\hat{g}_{N'}(\hat{x}^*)$) which is quite small. The small optimality gap implies that the SAA method can provide solutions with good quality.

In the case study above, samples are independently and identically distributed. We also consider applying the supersaturated design to generate samples. The results of the SAA method with AG design are shown in Table II. The optimal estimated objective value we can obtain with N=10 (note that the number of sample scenarios is smaller that the number of random variables of the ECR problem, i.e. 56) is 3361.81, which is quite close to the optimal estimated objective value in Table I, i.e. 3359.97, with N=100. It indicates that SAA method based on supersaturated design

TABLE I. RESULTS OF THE SAA METHOD (N=100)

Estimate	Value
$\bar{g}_N^M$	3335.45
$\hat{g}_{N'}(\hat{x}^*)$	3359.97
$\hat{g}_{N'}(\hat{x}^*) - \bar{g}_N^M$	24.52(0.73%)

 ISBN: 978-1-61208-161-8

can provide good solutions with a small number of sample scenarios.

The performances of the SAA method with i.i.d sampling, LH sampling, AG design, and AGLH design are compared in Fig. 2 and Fig. 3 (all with N=10). The replication number is 1000 and we can obtain 1000 feasible solutions for each sampling method. In Fig. 2, the mean and confidence interval of the estimated objective value $\hat{g}_{N'}(\hat{x})$ of each sampling method are analyzed. We find that all the three non-i.i.d samplings can reduce the average estimated objective value and the variance of the estimated objective value. Fig. 3 is the probability plot. Based on Fig. 3, we find that the non-i.i.d samplings are less likely to provide bad solutions compared with the i.i.d sampling. We also find that the SAA method with AGLH design has the smallest probability to provide bad solutions.

TABLE II. RESULTS OF THE SAA METHOD (N=10, AG DESIGN)

Estimate	Value
$\bar{g}_N^M$	3275.80
$\hat{g}_{N'}(\hat{x}^*)$	3361.81
$\hat{g}_{N'}(\hat{x}^*)-\bar{g}_N^M$	86.01(2.56%)

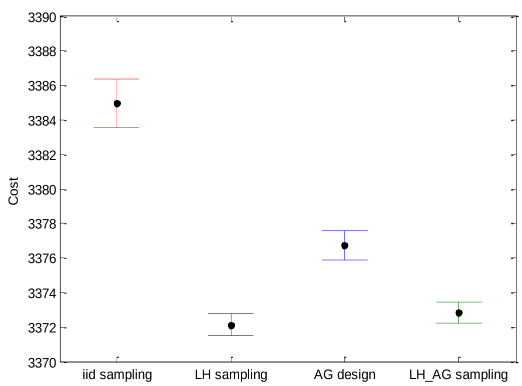

Figure 2. Expect cost estimates (α=0.05)

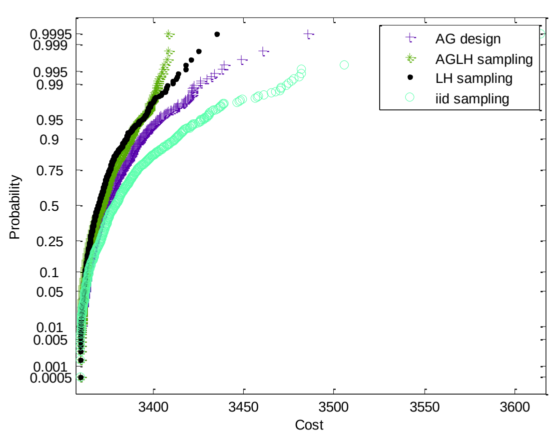

Figure 3. Probability plot of the objective estimates

V. CONCLUSION AND FUTURE STUDY

In this study, we developed an operational model to solve the ECR problem. In order to incorporate uncertainties, we built a two-stage stochastic model with uncertain demand, supply, residual ship space capacity, and residual ship weight capacity. The distributions of these parameters can be estimated based on historical data. We applied the SAA method to solve this stochastic problem. In the future, we will consider applying the SAA method to real-scale problems. Based on the results in numerical study section, we found that using LH design, AG design, and the combination of AG design and LH design to enhance the performance of SAA is promising. A direct extension of this work is to explore other sampling schemes to control scenario generation and thus improve the quality of solutions. Another possible direction for future research is to study the convergence rate of these samplings for stochastic programming.

REFERENCES

[1] A. Olivo, P. Zuddas, M. D. Francesco, and A. Manca, "An Operational Model for Empty Container Management," Maritime Economics and Logistics, vol. 7, pp. 199-222, 2005.

[2] C. M. Feng and C. H. Chang, "Empty Container Reposition Planning for intra-Asia Liner Shipping," Maritime Policy and Management, vol. 35, pp. 469-489, 2008.

[3] R. K. Cheung and C. Y. Chen, "A Two-Stage Stochastic Network Model and Solution Method for the Dynamic Empty Container Allocation Problem," Transportation Science, vol. 32, pp. 142-162, 1998.

[4] A. L. Erera, J. C. Morales, and M. Savelsbergh, "Robust optimization for empty repositioning problems," Operations Research, vol. 57, pp. 468-483, 2009.

[5] M. D. Francesco, T. G. Crainic, and P. Zuddas, "The effect of multi-scenario policies on empty container repositioning," Transportation Research Part E: Logistics and Transportation Review, vol. 45, pp. 758-770, 2009.

[6] A. J. Kleywegt, A. Shapiro, and T. Homem-De-Mello, "The Sample Average Approximation Method for Stochastic Discrete Optimization," SIAM Journal on Optimization, vol. 12, pp. 479-502, 2002.

[7] J. Wei and M. J. Realff, "Sample average approximation methods for stochastic MINLPs," Computers and Chemical Engineering, vol. 28, pp. 333-346, 2004.

[8] T. Homem-De-Mello, "On Rates of Convergence for Stochastic Optimization Programs Non-Independent and Identically distributed Sampling," SIAM J. OPTIM, vol. 19, pp. 524-551, 2008.

[9] H. Xu, "Uniform Exponential Convergence of Sample average random functions under general sampling with allocation in stochastic programming," Journal of Mathematical Analysis and Applications, vol. 368, pp. 692-710, 2010.

[10] M. B. Freimer, J. T. Linderoth, and D. Thomas, "The Impact of sampling methods on bias and variance in stochastic linear programs." Computational Optimization and Applications. 2010, DOI: 10.1007/s10589-010-9322-x

 ISBN: 978-1-61208-161-8

[11] B. Q. Tang and P. Z. G. Qian, “Enhancing the sample average approximation method with U design,” Biometrika 2010, pp. 1-14.

[12] L. H. Lee, E. P. Chew, Y. Long, Y. Luo, and J. J. Shao, “A Dynamic Approach to Empty Container Flow Management,” International Association of Matitime Economists 2010 Annual Conference, Lisbon, Portugal, July, pp. 7-9.

[13] S. Ahlinder and I. Gustafsson, “On Super Saturated Experiment Design,” unpublished.