Front. Eng. Manag. 2017, 4(3): 338–347
DOI 10.15302/J-FEM-2017032

RESEARCH ARTICLE

Yuan WANG, Ruyan SHOU, Loo Hay LEE, Ek Peng CHEW

A case study on sample average approximation method for stochastic supply chain network design problem

Abstract This study aims to solve a typical long-term strategic decision problem on supply chain network design with consideration to uncertain demands. Existing methods for these problems are either deterministic or limited in scale. We analyze the impact of uncertainty on demand based on actual large data from industrial companies. Deterministic equivalent model with nonanticipativity constraints, branch-and-fix coordination, sample average approximation (SAA) with Bayesian bootstrap, and Latin hypercube sampling were adopted to analyze stochastic demands. A computational study of supply chain network with front-ends in Europe and back-ends in Asia is presented to highlight the importance of stochastic factors in these problems and the efficiency of our proposed solution approach.

Keywords supply chain network, stochastic demand, sampling average approximation, Bayesian bootstrap, Latin hypercube sampling

1 Introduction

A multi-national corporation focused on constructing a detailed and robust supply chain network (SCN). This research area is known as supply chain management (SCM). SCM is categorized into strategic, tactical, and operational phases across a planning horizon (Bender et al., 2002). These phases are differentiated based on strategy detail levels. The strategic phase emphasizes selecting facilities, defining facility purposes, and building a network of suppliers, manufacturers, transporters, retailers, and customers. The tactical phase emphasizes procurement, production schedules and standards, and transportation solutions. The operational phase emphasizes demand forecasting, inventory management, and material flows. Strategic decisions, which often require a tremendous amount of investment, have long-term effects on a supply chain. These decisions cannot be easily changed once they are made (Badri, 1999). Therefore, this study focuses on the SCN design problem, which is one of the long-term decision problems.

A long-term decision is considered robust only when it can last under different types of market influences for a long period of time. Transportation costs, which resulted from facility location selection, already contributed up to 50%–60% of the total distribution costs of a company (Frank, 2002). Given that the products are not mainly transported directly from the manufacturing sites to the customers but through different types of facilities, such as distribution centers (DCs), the time spent on production, transportation, and storage must also be considered in facility location selection. We also observe that an acceptable facility location provides a decrease of 5% to 15% in logistic costs while maintaining or improving customer satisfaction (Ballou, 2001). Consequently, placement of facilities at optimal locations and removal of redundant facilities ensure that the supply chain effectively performs while maintaining low costs, short cycle time, and satisfied requirements (Sule, 2001).

The major factor that complicates long-term SCN decision-making process is uncertainty. Any part of the supply chain may vary over time, and a difference between the current and future situations usually completely alters the decisions. Potential uncertainty sources include demands, supplies, costs, lead times, laws, regulations, and exchange rates (Acar et al., 2010). The models constructed in this paper incorporate uncertain demands to obtain a robust answer to SCN problems. In the existing literature, a large number of studies on deterministic SCN design problem was conducted, and various methods on stochastic demand SCN with limited size were introduced.

Received April 30, 2017; accepted August 14, 2017

Yuan WANG (✉), Ruyan SHOU, Loo Hay LEE, Ek Peng CHEW
Department of Industrial Systems Engineering & Management, National University of Singapore, Singapore 117576, Singapore
E-mail: iseway@nus.edu.sg

A common assumption is that demand follows a certain distribution. While our work is a case study from an actual industrial project, the actual demand data are large but extremely random; thus, providing assumptions is difficult. Therefore, we propose two methods, namely deterministic equivalent method (DEM) and sample average approximation (SAA) method, to solve supply network design problems with large and random uncertain demands. In the DEM method, we present a splitting variable representation of the problem via each instance to jointly optimize the sub-problems. The branch-and-fix coordination (BFC) approach is introduced to coordinate the branch fix phase execution for location selection such that the nonanticipativity constraints are also satisfied. While SAA method is optimizing the average objectives of multiple instances of the model with multiple samples of the random elements. Over the entire year, a few elements usually have no demand; thus, the demand data are sparse and limited. Bayesian bootstrap is one of the approaches in generating samples for random elements. Bayesian bootstrap simulates the posterior distribution of a random element with limited amount of available data (Rubin, 1981). Latin hypercube sampling (LHS) is another method of generating samples for random elements. LHS generates samples by partitioning each distribution into N intervals of equal probability and selecting one sample from each interval (Stein, 1987). Thus, the possibility of samples clustering together in sampling methods, such as Monte Carlo sampling, is avoided.

In summary, the contributions of this paper are as follows.

1. We develop a multi-product, multi-shipment level model with realistic logistic constraints to solve an actual industrial strategy decision-making problem.
2. We present innovative approaches to solve the stochastic model with consideration to stochastic demands.
3. We explore the two efficient sampling techniques in the SAA methods.
4. We present that our method could solve large-scale decision-making problems with large but random data.

The remainder of the paper is organized as follows. In Section 2, we review the existing work on the present SCN design. We propose the deterministic model formulation in Section 3. The DEM method and SAA model are discussed in Section 4. We report the extensive experimental results of the actual case in Section 5. Section 6 concludes the paper.

2 Literature review

2.1 Deterministic SCN models

The simplest type of SCN location problem is a single-period single-objective deterministic problem with single or multiple commodities. Geoffrion and Graves (1974) solved a multi-commodity single-period single-objective deterministic capitated problem by using Benders decomposition. However, they did not separate transportation variables for plant-to-DC and DC-to-customer shipments but used quadruply subscripted variables to prevent the loss of commodity origin information. Canel and Khumawala (1997) studied the impact of international trade agreements on a company and proposed a multi-period model that allows companies to select facility locations globally. They also included a large variety of possible costs that might be incurred. They obtained an exact optimal solution with an acceptable computational efficiency by using branch-and-bound (BNB) algorithm. Kouvelis and Rosenblatt (2002) considered a multi-period SCN problem with governmental incentives, tariff, local rules, and taxation. They solved the model by using a mathematical programming software GAMS/OSL2. Klose and Drexl (2005) reviewed different models for solving deterministic distribution location problem. Chakravarty (2005) solved a similar problem with the assistance of an efficient search procedure and Kuhn–Tucker equation. Facility location, production quantities, export and import quantities were also studied. Similarly, Wilhelm et al. (2005) studied a multi-period SCN problem under the North American Free Trade Agreement with consideration to shipment types, material flows, safety labor, and local and international rules. Eskigun et al. (2005) considered several possible shipment strategies in their model. The proposed model in the present study focuses on the different types of conveyance. By contrast, their model focused on the different steps in shipment. Lagrangian heuristics, greedy heuristics, and sub-gradient algorithm were applied to reach the solution. Altiparmak et al. (2006) constructed an SCN model which simultaneously analyzed total cost, customer service level, and capacity utilization. However, these objectives are conflicting. They applied a genetic algorithm to obtain a set of Pareto-optimal solutions. The study was divided into two stages: weight approach and simulated annealing, which were used accordingly. Similarly, Cordeau et al. (2006) integrated the selection of shipment types with deterministic SCN problem. BNB and Benders decomposition were used to reach the solution.

2.2 Stochastic supply chain network models

Erlebacher and Meller (2000) proposed a stochastic multi-objective capitated p-median model to incorporate SCN and inventory. The NP-hard problem was solved by a self-developed heuristic. Similarly, Sabri and Beamon (2000) used the e-constraint and iterative procedure between strategic and operational sub-models to solve their multi-objective stochastic model, which analyzed conflicting objectives, namely total cost and SCN volume flexibility. Alonso-Ayuso et al. (2003) presented a scheme to handle multi-stage decision problems with uncertainties. They discretized the model into different scenarios. The best solution can be reached by analyzing different scenarios using the DEM model, nonanticipativity constraints, and BFC. Guillén et al. (2005) studied the interactions between conflicting objectives, namely net present value, customer satisfaction level, and financial risk, by using a multi-objective stochastic model. They obtained the answer using Pareto-optimal solutions and e-constraint. Shen and Qi (2007) proposed a model to incorporate SCN, safety stock, customer service level, and transportation routing. The stochastic and multi-period model was solved by Lagrangian relaxation-based heuristics and BNB.

The considerable impact of uncertain demands on SCN problems attracted significant attention from researchers. Laval et al. (2005) proposed a stochastic SCN model for Hewlett–Packard. They extrapolated data for non-existent cases, grouped similar scenarios, and used representative scenarios to optimize their model. They highlighted the possibility of introducing bias during data extrapolation and the necessity of data verification by experts. Owen and Daskin (1998) provided a holistic overview of the previous studies to solve stochastic location problems. Probabilistic, queuing, and scenario planning models have been frequently used. Variables are incorporated into probabilistic models with their distributions. Gregg et al. (1988) proposed a solution to allocate public libraries in New York City by using service demand distributions. They also discussed the impact of stochastic demands on location decisions using sensitivity analysis. Queuing models are frequently used to represent stochastic customer demands as stochastic customer arrival rates. Averbakh and Berman (1997) studied the problem of placing a vehicle in a congested network with consideration to customer arrival rates. They defined customer waiting time as the objective and formulated the problem as a demand arrival Poisson process. A self-defined procedure was deployed to obtain the optimal solution as a function of stochastic customer demands. Scenario planning models are used to identify solutions with the most promising outcomes of decision maker-defined uncertainty-incorporated scenarios. Mir-Hassani et al. (2000) proposed a two-stage model for multi-period SCN capacity planning. Benders decomposition was used to solve this stochastic SCN problem with uncertain demand. Tsiakis et al. (2001) proposed a two-stage model for multiproduct, multi-echelon SCN design problems. However, only three demand scenarios were considered in their paper. Kouvelis and Yu (2013) proposed a method to for decision making with uncertainties in the collected data. They discussed the construction of a comprehensively discrete mathematical programming framework to search for the worst scenario. Petridis (2015) provided the optimal design of a multiproduct, multi-echelon supply network under uncertainty of demand, which is normally distributed.

The preceding works mentioned show the various methods that deal with stochastic demand in the SCN design problem; however, none of these methods use actual raw, large, and random industrial data. Thus, solving the problem and providing a distribution fitting assumption is difficult, especially when the raw demand data are sparse. Therefore, the present study aims to solve stochastic SCN design problems by using large but random data.

3 Deterministic model formulations

3.1 Model background

The company SCN comprises a series of operations, including fabrication and sorting at front = ends (FEs), assembly and final test at back-ends (BEs), and several inventory points and distribution centers, which are called die banks (DBs). The Company's FEs are located in Europe whereas its BEs are located in Asia Pacific (APAC). A supply network design is similar to the "storage at distributor and carrier delivery" design. This design functions under an integrated system of push and pull processes. Figures 1 and 2 provide an overview of the SCN of the company.

Push process is used in anticipation of a customer order. Materials, semi-finished products, or final products are prepared in advance by forecasting customer demands. This procedure reduces the amount of time spent between receiving customer order and fulfilling the order. However, ensuring that products are available when a customer order arrives requires a high inventory level on hand. Thus, the inventory cost increases, diminishing the overall profit. By contrast, the pull process is used in response to a customer order. Products are only manufactured or assembled after receiving the order. A considerable amount of time is spent between receiving the order and fulfilling the order. However, the inventory level and cost are reduced. The company adopts a mixture of both processes to achieve a balance between responsiveness to customer order and cost efficiency.

DBs are the most important part of SCN because they are disposition points for semi-finished products (Christian and Thomas, 2014). DBs behave as a boundary that separates the push from the pull process. Products are fabricated at FEs based on make-to-forecast strategy and then shipped to DBs. Subsequently, when customers place their orders, the assembling facilities (AssyLocations) at BEs use the product inventories at DBs for the assembly, which is based on the assemble-to-order strategy (Dinesh, 2010). Moreover, DBs serve as decentralization points of product batches because different products are packed and shipped together in batches when they have the same designated FE and sales code. Thus, economies of scale are achieved, and transportation cost is reduced.

Therefore, considering the supply chain strategies that the company adopted and the considerable importance of

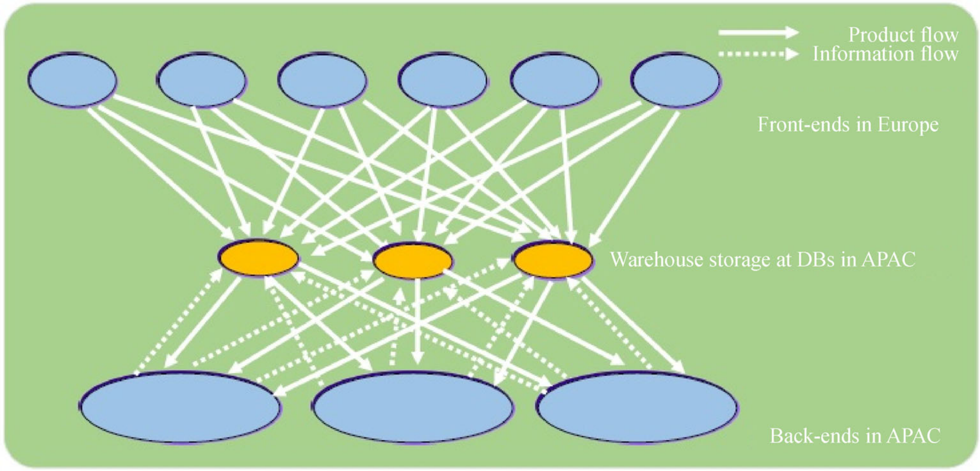

Fig. 1 Company SCN design

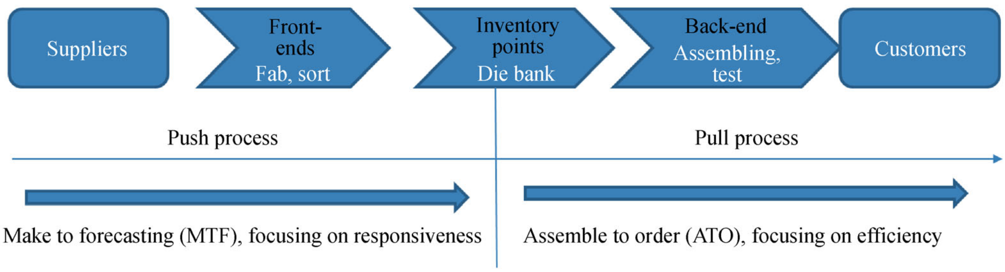

Fig. 2 Supply chain strategies of the company

DBs, the present study assists the company in making decisions regarding the suitable places in APAC to construct DBs. In this section, we develop the deterministic model with the following assumptions.

- For confidential agreement reasons, we simplify the network into three layers: FE, BE, and DB. The four candidate DB locations include Batam, Singapore, Philippines, and Shanghai. A few sites in BE can also be potential DBs. FE and BE locations cannot be published for confidential reasons and thus are represented in capital letters.
- Transportation and storage costs are included in the total cost, whereas other costs, such as inventory, order, and shortage costs, are excluded.
- The model describes the reality as much as possible. For example, in a few constraints, different components that share the same project code must ship to the same hub.

3.2 Notations

The following notations are used in the model.

$$\text{I: Set of FEs. i} = \begin{cases} 1 & D \\ 2 & K \\ 3 & R \\ 4 & V \end{cases}$$

$$\text{J: Set of DBs. j} = \begin{cases} 1 & Batam \\ 2 & Singapore \\ 3 & Philippines \\ 4 & Shanghai \end{cases}$$

$$\text{K: Set of BEs. k} = \begin{cases} 1 & P \\ 2 & S \\ 3 & B \\ 4 & U \end{cases}$$

C: Set of Product Types. $c \in \{1...270\}$
S: Set of Shipment Levels.

$$s = \begin{cases} 1 & (\ 0,\ 25\ )products \\ 2 & [25,50)products \\ 3 & [50,100)products \\ 4 & [100,150)products \\ 5 & [150,750)products \\ 6 & [750,1500)products \\ 7 & [1500,3750)products \end{cases}$$

The decision variables and parameters are illustrated in Tables 1 and 2, respectively.

Our objective function minimizes the total annual cost, including transportation costs of product shipment from FEs to DBs, transportation costs of product shipment from DBs to BEs, storage costs of DBs, and fixed transportation cost between Batam and site U. The function is defined as follows:

$$Min\ Z = r_{SGD}^{EURO}\left(\sum_i\sum_j\sum_s\sum_c \alpha_{\mathrm{ijcs}}\cdot c_{ijs}^{\alpha}\right.$$
$$\left.\sum_j\sum_k\sum_s\sum_c \beta_{\mathrm{jkcs}}\cdot c_{jks}^{\beta} + \sum_j f_j r_j \cdot \mathrm{M}\right)$$
$$+\sum_t \delta\cdot C^{r}\cdot T\cdot W.$$

Subject to:

$$\sum\nolimits_j\sum\nolimits_s X_{ijcs} \geqslant \sum\nolimits_k d_{ikc} \forall i,c, \tag{1}$$

$$\sum\nolimits_j\sum\nolimits_s Y_{jkcs} \geqslant \sum\nolimits_k d_{ikc}\ \forall i,c, \tag{2}$$

$$\sum\nolimits_i\sum\nolimits_s X_{ijcs} = \sum\nolimits_k\sum\nolimits_s Y_{jkcs}\ \forall j,c, \tag{3}$$

$$d_{ikc} - L_s < K\cdot\ \vartheta_{ikc}^{L_s}\ \ U_s - d_{ikc} \leqslant K\cdot\ \vartheta_{ikc}^{U_s}$$

$$\vartheta_{ikc}^{L_s} + \vartheta_{ikc}^{U_s} - 1 \leqslant K\cdot\ \eta_{ikcs}\ \forall i,k,c,s, \tag{4)–(6}$$

$$\sum\nolimits_j \alpha_{ijcs} = \sum\nolimits_k \eta_{ikcs}\ \forall i,c,s, \tag{7}$$

$$\sum\nolimits_j \beta_{jkcs} = \sum\nolimits_i \eta_{ikcs}\ \forall k,c,s, \tag{8}$$

Table 1 Decision variables

Decision variables	Description
X_{ijcs}	No. of type c products shipped from FE i to DB j by shipment level s
Y_{jkcs}	No. of type c products shipped from DB j to BE k by shipment level s
α_{ijcs}	1 if type c products are shipped from FE i to DB j by shipment level s, 0 otherwise
β_{jkcs}	1 if type c products are shipped from DB j to BE k by shipment level s, 0 otherwise
γ_j	1 if DB j is selected, 0 otherwise
δ	1 if the route between Batam and U is selected (as it is a monthly fixed cost), 0 otherwise
$\zeta_{ikc}^{L_s}$	1 if demand d_{ikc} is larger than or equal to the lower bound of shipment level s, 0 otherwise
$\zeta_{ikc}^{U_s}$	1 if demand d_{ikc} is smaller than the upper bound of shipment level s, 0 otherwise
η_{ikcs}	1 if shipment level s is used for demandd$_{ikc}$, 0 otherwise
θ_{c}	1 if demand of type c products is nonzero, 0 otherwise

Table 2 Parameters used in the model

Parameters	Description
d_{ikc}	Demand of type c products from BE k, which is supposed to be produced at designated FE i
L_s	Lower bound of shipment level s
U_s	Upper bound of shipment level s
c_{ijs}^{α}	Transportation cost from FE i to DB j per carton by shipment level s in EURO
c_{jks}^{β}	Transportation cost from DB j to BE k per carton by shipment level s in EURO
c^{γ}	Transportation cost between B and U per trip in SGD
f_j	Storage cost of DB j per month in EURO
r_{Euro}^{SGD}	Exchange rate from EURO to SGD (based on the exchange rate on 27/10/16)
(r_{Euro}^{SGD} = 1.52)	
M	No. of months per year (M = 12)
W	No. of weeks per year (W = 52)
T	No. of trips between B and U per week (t = 7)
K	A large integer (K = 100000)

$$\sum_i \sum_k d_{ikc} \leqslant k \cdot \theta_c \ \forall c, \tag{9}$$

$$\sum_i \sum_j \sum_s \alpha_{ijcs} = \theta_c \ \forall c, \tag{10}$$

$$\sum_j \sum_k \sum_s \beta_{jkcs} = \theta_c \ \forall c, \tag{11}$$

$$X_{ijcs} \leqslant K \cdot \alpha_{ijcs} \ \forall i,j,c,s, \tag{12}$$

$$Y_{jkcs} \leqslant K \cdot \beta_{ikcs} \ \forall j,k,c,s, \tag{13}$$

$$\sum_i \sum_s \sum_c \alpha_{ijcs} \leqslant K \cdot r_j \ \forall j, \tag{14}$$

$$\sum_k \sum_s \sum_c \beta_{jkcs} \leqslant K \cdot r_j \ \forall j, \tag{15}$$

$$\sum_c \sum_s \beta_{14cs} \leqslant K \cdot \delta \ \forall j , \tag{16}$$

$$X_{ijcs}, Y_{jkcs} \geqslant 0, \tag{17}$$

$$\alpha_{ijcs}, \beta_{jkcs}, r_j, \delta, \vartheta_{ikc}^{L_s}, \vartheta_{ikc}^{U_s}, \eta_{ikcs}, \theta_c \in \{0,1\} \ \forall i,j,k,c,s, \tag{18}$$

Constraint sets (1) and (2) guarantee that the demand for type c products is satisfied by the type c products shipped from FE i to DB j and from DB j to BE k. Equation (3) imposes a balance between the number of type c products shipped from FEs to a specific DB j and the number of type c products shipped from that DB to BEs. Constraint sets (4), (5), (6), (7), and (8) ensure an appropriate shipment level to be selected to transport a specific product type based on its demand. Constraint sets (9), (10), and (11) state that a specific product type only has at most one designated FE, DB, and BE, as well as one shipment level. Constraint sets (12) and (13) ensure that shipment routes with a specific shipment level from FEs to DBs and from DBs to BEs can only be activated when a few products are shipped along these routes. Constraint sets (14) and (15) are connecting constraints to guarantee products passing through if the site is chosen as DB. Constraint (16) checks if shipment route between Batam and U is activated. Finally, constraint sets (17) and (18) are the non-negativity and binary restrictions on decision variables, respectively.

4 Stochastic programming method

4.1 Deterministic equivalent model

The variance of demand values of a specific product type can be extremely large, and a few demand values are equal to zero. Therefore, the demands for the product types in one year are rather random and uncertain. We divided the analysis into two methods to obtain a systematic approach of analyzing uncertain demands. In the first method, we apply the idea of DEM with nonanticipativity constraints and BFC scheme to obtain the final decision by inputting the exact one year demand data into the deterministic model. Alonso-Ayuso et al. (2003) introduced this method to solve mixed 0–1 programs under uncertainty in the objective function coefficients or parameters, which are the storage costs in our base model. This method also considers multiple values of a parameter as multiple instances of the deterministic model and simultaneously solves these instances. Similarly, we split our one year demand into 12 instances and used these instances to represent the DEM model.

A DEM model defines our base model as follows, where W is the set that contains all the instances, w is an instance in set W with the probabilityp^w, y^w denotes the set of binary DB location decision variables of instance w, and x^w denotes the set of other decision variables of instances. A and B are coefficient sets of x^w and y^w in the objective function, respectively.

Minimize $Z = \sum_{w \in W} P^w (\alpha^w x^w + b^w y^w)$,

Subject to: $\mathrm{A}\ \mathrm{x}^\mathrm{w} + \mathrm{B}\ \mathrm{y}^\mathrm{w} = \mathrm{h}^\mathrm{w}\ \forall \mathrm{w} \in \mathrm{W}$

$\mathrm{y}^\mathrm{w} \in \{0,1\} \forall \mathrm{w} \in \mathrm{W}, \mathrm{x} \geqslant 0$.

In our problem, we present a splitting variable representation of the problem for each month; thus, each month could be regarded as a sub-problem with equal probability. The DEM model can be easily solved by jointly optimizing the sub-problems. Unfortunately, the product demands in each instance are different from the others, and the instances are separately optimized; thus, two or more different final decisions can be observed. As such, explicitly adding the nonanticipativity constraints to the DEM model is essential. The nonanticipativity constraints force the final decisions of all the instances to be the same and should not be arbitrarily changed. The nonanticipativity constraints are as follows:

$$\mathrm{x}^\mathrm{w} - \mathrm{x}^{\mathrm{w}+1} = 0\ \forall \mathrm{w} \in \mathrm{W}.$$

We adopt the idea of BFC scheme and fixed DB location decision variables, which are exactly the same for all the instances and branched on the other DB location decision variables, to efficiently satisfy the nonanticipativity constraints in the DEM model. Therefore, the problem is broken into sub-problems, which can be jointly optimized with less effort. A simple example of the BFC scheme is shown in Fig. 3 as follows.

In our DEM model, we split the problem into sub-problems by month; thus, the probability p^w is equal to 1.

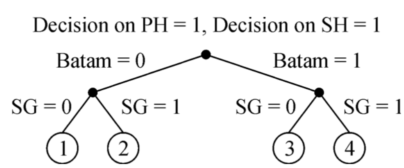

Fig. 3 BFC on DB decision variables

In other words, our DEM model is deterministic because it utilized the previous fixed demand data. Hence, we introduce another method to analyze the stochastic demand.

4.2 Sample average approximation method

For the second method, investigating the final decision based on the exact one-year demand data are insufficient to make a conclusion because the demand values are highly uncertain. The demand values in the future might alter the current decision. Therefore, the underlying demand distribution of each product type must be identified, and the analysis must be extended to more demand values. We adopted the idea of SAA from Kim et al. (2015) to solve OR problems by optimizing the average objective value of multiple model instances based on multiple samples generated from stochastic distributions. We applied two sampling methods, namely Bayesian bootstrap and LHS, due to limited data points to draw the demand distributions; both methods are useful in this situation.

4.2.1 Two-stage stochastic model

We present the two-stage stochastic model, which assumes that all parameters in stage one are known while a few parameters in stage two are unknown, to extend the previous deterministic model to a stochastic setting. The first stage comprises the location decisions y, and the second stage comprises all costs related in an optimal fashion under the realized uncertain demands. The general formulation of the two-stage stochastic model is expressed as follows:

Minimize $f(y) = c^T y + E[Q(y,\varepsilon)]$,

Subject to: $\forall\ y \in Y$ in$\{0,1\}$,

where $Q(y, \varepsilon)$ is the optimal objective function of the second-stage problem.

Minimize $q^T x$,

Subject to: $A\ x^s + B\ y^s = h^s \quad \forall s \in S$

$x \geqslant 0$,

where $s \in S$ denotes an unknown scenario in stage one but is known when the decisions at stage two are made, and S is the set of all scenarios.

4.2.2 Sample average approximations

The two-stage stochastic models are extremely difficult to solve because the cost expectation $E[Q(y,\varepsilon)]$ in the objective is complicated. Thus, we deal with the problem by using the SAA method. The basic idea of the SAA method is using random samples to approximate the expected objective function of a stochastic problem, as shown in the following equation. With random samplings, the problem becomes deterministic and solvable by optimization techniques. The proposed SAA procedures are described as follows.

$$\min_{y \in Y} \left\{ \tilde{f}_N(Y) := c^T y + \frac{1}{N}\sum_{n=1}^{N} Q(\mathrm{y}, \in^{\mathrm{n}}) \right\}.$$

Step 1: Generate M independent samples with the sample size N. Then, solve the corresponding M independent SAA problems

$$\min_{y \in Y} \left\{ c^T y + \frac{1}{N}\sum_{n=1}^{N} Q\left(\mathrm{y}, \xi_j^n\right) \right\}.$$

Step 2: Compute the sample average:

$$\overline{Z}_{N,M} := \frac{1}{(M)}\sum_{j=1}^{M}(Z_N^j).$$

Let z^* denote the actual optimal value given that $\overline{z}_{N,M}$ is an unbiased estimation of Ez_N. Thus, $Ez_N \leqslant z^*$ and the estimator $\overline{z}_{N,M}$ provide a statistical lower bound of the optimal objective value.

Step 3: Estimate the actual objective function value as follows:

$$\tilde{f}_{N'}(\overline{y}) := c^T \overline{y} + \frac{1}{N'}\sum_{n=1}^{N'} \mathrm{Q}\left(\overline{y}, \xi_j^n\right).$$

We evaluate the M candidate solutions in a large size N', where $\tilde{Z}_{N'}(\overline{y})$ is an unbiased estimator of $f(\overline{y})$, and $\tilde{Z}_{N'}(\overline{y})$ is an upper bound of Z^*.

Step 4: Compute the optimality gap as follows: $\tilde{Z}_{N'}(\overline{y}) - \overline{Z}_{N,M}$.

4.2.3 Sampling methods

Each sample d_i which is used in SAA, is obtained by using Bayesian bootstrap or LHS method. On the one hand, Bayesian bootstrap simulates the posterior parameter distributions instead of simulating distributions of a statistical estimator of that parameter, similar to the study of Rubin (1981). On the other hand, LHS generates random samples of parameters by partitioning the distributions and selecting one sample from each partition to prevent samples from clustering together Stein (1987).

5 Experiments

In this section, we first separately introduce the results from DEM model and SAA method. Then, the comparative results are presented to evaluate the performance of our proposed methods. IBM ILOG Cplex Optimization Studio Version 12.5 is used to run the model under an environment, which comprises an Intel (R) Core (TM) i5-3570 CPU, 3.40 GHz processors, and an 8 GB RAM. The demand samples are generated based on Bayesian bootstrap and LHS methods in Rstudio. Library packages, Bayesian bootstrap and LHS were used.

5.1 DEM results

Different from the DEM model which comprises multi-stage decision variables and proposed by Alonso-Ayuso et al. (2003), DB location decision variables in our deterministic-based model are simultaneously provided only once. Thus, the model has only one stage.

We obtain the final decision result matrix by separately running the base model for 12 months, as shown in Table 3. The result shows that the final decisions across 12 months are nearly the same. DB Philippines and DB Shanghai are selected all the time. The only difference occurred in November 2016 and January 2017, in which decision differs between DB Batam and DB Singapore.

Table 3 DB decisions obtained based on 12-month demands

Month	Batam	Singapore	Philippines	Shanghai	Objective (S$)
Jul-16	1	0	1	1	110,689.6
Aug-16	1	0	1	1	121,579.5
Sep-16	1	0	1	1	124,029.2
Oct-16	1	0	1	1	121,665.8
Nov-16	0	1	1	1	128,591.7
Dec-16	1	0	1	1	121,233.4
Jan-17	0	1	1	1	124,160.3
Feb-17	1	0	1	1	111,106.9
Mar-17	1	0	1	1	102,843
Apr-17	1	0	1	1	91,406.94
May–17	1	0	1	1	78,181.76
Jun-17	1	0	1	1	64,648.94

Next, we fix the decisions on DBs Philippines and Shanghai in one case and branched the decisions on DBs Batam and Singapore into four cases, as shown in Fig. 3. We run the base model in each case with 12-month product demand data and obtain the minimum total cost as shown in the objective function of the following modified DEM model.

$$Z \min = \min\{Z_i\} \ \forall i,$$

where $Z_i = Minmize \sum_{w \in W} P^w(\alpha^w x^w + b^w y^w)$, and i denotes the set of cases, $i \in \{1,2,3,4\}$.

Table 4 shows that selecting DB Batam provides the lowest total cost of S$108400.32. The highest total cost of S$117467.97 was incurred when all DBs are selected due to a large increase in the storage cost. Therefore, the optimal decision from the DEM method is selecting DBs Batam, Philippines, and Shanghai with the lowest total cost of S$108400.32.

5.2 SAA results

We obtain 20 groups with a demand sample size of 50 by each sampling method after running the Bayesian bootstrap and LHS packages on Rstudio. Noticeably, by solving the SAA model with the generated demand samples, both sampling methods indicate that the first DB decision (selecting Batam, Philippines, and Shanghai) provides the lowest average objectives compared to the other DB decision sets. However, directly concluding that the first DB decision (selecting Batam, Philippines, and Shanghai) is the optimal decision does not consider the occasionality when the low average objective is caused by generated low demands. Thus, we evaluate each potential location(s) option in a large demand sample size of 500, as generated by the Bayesian bootstrap and LHS methods. We assume that a sample size of 500 is most likely to avoid such occasionality.

Table 4 Objectives obtained based on 12-month demands for four cases defined by BFC

Month	Batam = 0 Singapore = 0	Batam = 0 Singapore = 1	Batam = 1 Singapore = 0	Batam = 1 Singapore = 1
Jul-16	9,539.40	9,559.51	9,224.13	10,009.09
Aug-16	10,589.77	10,277.75	10,131.62	10,722.50
Sep-16	10,702.54	10,544.21	10,335.76	10,961.02
Oct-16	10,660.80	10,363.03	10,138.82	10,717.72
Nov-16	11,509.58	10,715.97	10,764.60	11,107.57
Dec-16	10,579.95	10,374.62	10,102.78	10,779.61
Jan-17	11,026.95	10,346.69	10,353.65	10,713.01
Feb-17	9,590.83	9,463.51	9,258.91	9,884.18
Mar-17	8,972.40	9,073.50	8,570.25	9,480.6
Apr-17	7,941.56	8,246.82	7,671.25	8,718.36
May–17	6,755.48	7,171.92	6,515.15	7,646.97
Jun-17	5,549.94	6,238.71	5,387.41	6,726.34
Annual cost (S$)	113,419.2	112,376.24	108,400.32	117,467.97

We observe that both methods highlight that the DB decision of selecting DBs Batam, Philippines, and Shanghai is the optimal decision that provides the lowest total cost. This result coincides with the result obtained from the DEM model with BFC in Subsection 4. We also provide the top three solution results with their corresponding gap in each solution in Table 5. The table shows that both sample techniques can generate robust near optimal solutions.

5.3 Comparison results

In this subsection, we compare our proposed results with the deterministic model in Section 3 using simple annual average demand input. Most of the literature in deterministic environment simply assumed that the demand is the average number. The purpose of this comparison is to show the difference between large data input and one simple average input. The comparative results are shown in

Table 5 Top 3 SAA Solution values and Gaps

DB Decision Values				Large sample results and gaps (S$)			
Batam	Singapore	Philippines	Shanghai	Bayesian Bootstrap	Bayesian Bootstrap Gap	LHS	LHS Gap
1	0	1	1	101,726.62	65.68	115,654.37	6.44
1	0	1	0	104,274.91	17.33	120,050.98	29.62
1	0	0	1	107,798.59	68.84	122,940.80	18.74

Table 6 Comparative results between stochastic and deterministic demand

Deterministic result	DEM result (S$)	SAA result (S$)
Shanghai + Philippines	Batam + Shanghai + Philippines	Batam + Shanghai + Philippines
198,846.04	108,400.32	101,726.62

Table 6. A significant difference exists in solutions between the deterministic model with assumption input and the stochastic model with large data. Thus, the importance of data in the SCN design problem is enhanced.

6 Conclusions and future work

We studied a large-scale SCN design problem under uncertainties. A mathematical model that minimizes total cost comprising storage and transportation costs was developed to solve an actual industrial problem. We analyzed the impact of stochastic demand and the importance of data on the final decision.

One of our main contributions is investigating the impact of stochastic demand on the decision-making process of SCN design problems. Another contribution of this study is providing a solution scheme for stochastic SCN design problem by using the large data of the company. An example of how company historical data are utilized to guide decision making is also provided.

The scope of our research includes storage and transportation costs. However, other quantitative factors, such as inventory costs and fill rates, as well as qualitative factors, such as tariffs, laws, and regulations in different countries, can be considered for future studies. In addition, rather than using three layers, this problem can be extended to the entire global SCN, including procurement, fabrication, sorting, testing, and delivery. Another fertile area for future studies is a concurrent analysis of the impact on the final decision when multiple uncertainties simultaneously vary. Additional effects on the final decision due to interactions between different uncertainties might be observed by simultaneously considering multiple uncertainties.

References

Acar Y, Kadipasaoglu S, Schipperijn P (2010). A decision support framework for global supply chain modelling: An assessment of the impact of demand, supply and lead-time uncertainties on performance. International Journal of Production Research, 48(11): 3245–3268

Alonso-Ayuso A, Escudero L F, Teresa Ortuño M (2003). BFC, a branchand-fix coordination algorithmic framework for solving some types of stochastic pure and mixed 0–1 programs. European Journal of Operational Research, 151(3): 503–519

Altiparmak F, Gen M, Lin L, Paksoy T (2006). A genetic algorithm approach for multi-objective optimization of supply chain networks. Computers & Industrial Engineering, 51(1): 196–215

Averbakh I, Berman O (1997). Minimax regret p-center location on a network with demand uncertainty. Location Science, 5(4): 247–254

Badri B A (1999). Combining the analytic hierarchy process and goal programming for global facility location-allocation problem. International Journal of Production Economics, 62(3): 237–248

Ballou R H (2001). Unresolved issues in supply chain network design. Information Systems Frontiers, 3(4): 417–426

Bender T, Hennes H, Kalcsics J, Melo T, Nickel S (2002). Location software and interface with gis and supply chain management. Applications and Theory, 233–274

Canel C, Khumawala B M (1997). Multi-period international facilities location: An algorithm and application. International Journal of Production Research, 35(7): 1891–1910

Chakravarty A K (2005). Global plant capacity and product allocation with pricing decisions. European Journal of Operational Research, 165(1): 157–181

Christian S, Thomas P (2014). Supply chain integration and practical problems in semiconductor manufacturing. Research Gate

Cordeau J F, Pasin F, Solomon M M (2006). An integrated model for logistics network design. Annals of Operations Research, 144(1): 59–82

Dinesh (2010). Semiconductor industry planning process simplified. http://savysapconsultant.wordpress.com/2010/01/16/semiconductor-industry-%E2%80%93-planning-process-simplified

Erlebacher S J, Meller R D (2000). The interaction of location and inventory in designing distribution systems. IIE Transactions, 32(2): 155–166

Eskigun E, Uzsoy R, Preckel P V, Beaujon G, Krishnan S, Tew J D

(2005). Outbound supply chain network design with mode selection, lead times and capacitated vehicle distribution centers. European Journal of Operational Research, 165(1): 182–206

Frank R (2002). Natural site selection for distribution center site selection. Operations and Fulfillment

Geoffrion A M, Graves G W (1974). Multicommodity distribution system design by benders decomposition. Management Science, 20 (5): 822–844

Gregg S, Mulvey J, Wolpert J (1988). A stochastic planning system for siting and closing public service facilities. Environment & Planning A, 20(1): 83–98

Guillén G, Mele F, Bagajewicz M, Espuña A, Puigjaner L (2005). Multiobjective supply chain design under uncertainty. Chemical Engineering Science, 60(6): 1535–1553

Kim S, Pasupathy R, Henderson S G (2015). A guide to sample average approximation. In: Handbook of Simulation Optimization. Springer, 207–243

Klose A, Drexl A (2005). Facility location models for distribution system design. European Journal of Operational Research, 162(1): 4–29

Kouvelis P, Rosenblatt M J (2002). A mathematical programming model for global supply chain management: Conceptual approach and managerial insights. In: Supply Chain Management: Models, Applications, and Research Directions. Springer, 245–277

Kouvelis P, Yu G (2013). Robust discrete optimization and its applications. Springer Science & Business Media, 14

Laval C, Feyhl M, Kakouros S (2005). Hewlett-packard combined or and expert knowledge to design its supply chains. Interfaces, 35(3): 238–247

MirHassani S, Lucas C, Mitra G, Messina E, Poojari C A (2000). Computational solution of capacity planning models under uncertainty. Parallel Computing, 26(5): 511–538

Owen S H, Daskin M S (1998). Strategic facility location: A review. European Journal of Operational Research, 111(3): 423–447

Petridis K (2015). Optimal design of multi-echelon supply chain networks under normally distributed demand. Annals of Operations Research, 227(1): 63–91

Rubin D B (1981). The bayesian bootstrap. Annals of Statistics, 9(1): 130–134

Sabri E H, Beamon B M (2000). A multi-objective approach to simultaneous strategic and operational planning in supply chain design. Omega, 28(5): 581–598

Shen Z J M, Qi L (2007). Incorporating inventory and routing costs in strategic location models. European Journal of Operational Research, 179(2): 372–389

Stein M (1987). Large sample properties of simulations using latin hypercube sampling. Technometrics, 29(2): 143–151

Sule D R (2001). Logistics of Facility Location and Allocation.America: CRC Press

Tsiakis P, Shah N, Pantelides C C (2001). Design of multi-echelon supply chain networks under demand uncertainty. Industrial & Engineering Chemistry Research, 40(16): 3585–3604

Wilhelm W, Liang D, Rao B, Warrier D, Zhu X, Bulusu S (2005). Design of international assembly systems and their supply chains under nafta. Transportation Research Part E, Logistics and Transportation Review, 41(6): 467–493