Computers and Chemical Engineering 73 (2015) 1–16

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Simulation-based optimization framework for multi-echelon inventory systems under uncertainty

Yunfei Chu[a], Fengqi You[a,*], John M. Wassick[b], Anshul Agarwal[b]

[a] *Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA*
[b] *The Dow Chemical Company, Midland, MI 48674, USA*

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form 13 October 2014
Accepted 20 October 2014
Available online 31 October 2014

Keywords:
Simulation-based optimization
Supply chain management
Agent-based modeling
Cutting plane algorithm
Statistical hypothesis tests

ABSTRACT

Inventory optimization is critical in supply chain management. The complexity of real-world multi-echelon inventory systems under uncertainties results in a challenging optimization problem, too complicated to solve by conventional mathematical programing methods. We propose a novel simulation-based optimization framework for optimizing distribution inventory systems where each facility is operated with the (r, Q) inventory policy. The objective is to minimize the inventory cost while maintaining acceptable service levels quantified by the fill rates. The inventory system is modeled and simulated by an agent-based system, which returns the performance functions. The expectations of these functions are then estimated by the Monte-Carlo method. Then the optimization problem is solved by a cutting plane algorithm. As the black-box functions returned by the Monte-Carlo method contain noises, statistical hypothesis tests are conducted in the iteration. A local optimal solution is obtained if it passes the test on the optimality conditions. The framework is demonstrated by two case studies.

1. Introduction

Supply chain optimization is critical for manufacturing companies to maintain economic viability in the current highly competitive global marketplace (Grossmann, 2005; Neiro and Pinto, 2004; Wassick et al., 2012). A supply chain is a network of interconnected entities, such as suppliers, manufactures, distributors, retailers, and customers, which have different functions ranging from procurement of raw materials, transformation of raw materials into products, and distribution of final products to customers (Chopra and Meindl, 2007; Sarimveis et al., 2008). Supply chain decisions are broadly classified into strategic, tactical, and operational ones (Min and Zhou, 2002). Inventory management and control is a critical decision-making in supply chain optimization (Grossmann, 2005; Papageorgiou, 2009; You and Grossmann, 2008, 2010, 2011a; You et al., 2011). The objective of inventory optimization is to determine optimal inventory at each stocking location in a multi-echelon network for given inventory control policies at each location so as to minimize total cost (such as inventory holding cost, working capital, inventory carrying cost, production cost, etc.) while maintaining acceptable customer service levels.

Though widely studied, multi-echelon inventory optimization is still a challenging problem due to interacting facilities, stochastic parameters, and dynamic behaviors (Zipkin, 2000). Mathematical programming methods, such as multi-stage stochastic programming or stochastic inventory approaches, suffer from computational complexity (You and Grossmann, 2011b; Yue and You, 2013; Zapata et al., 2011). As the operations and networks in a supply chain become more complex, analytical approaches tend to become inadequate and inaccurate to model complicated real-world supply chains (Peidro et al., 2009). In order to solve the complicated problems which are intractable for mathematical programming methods, simulation-based optimized methods offer promising alternatives (Fu, 2002). In a simulation setting, the model is able to capture details that can represent the actual stochastic and dynamic inventory system relatively accurately (Shah, 2005).

Simulation-based optimization methods, though quite versatile for inventory optimization, still face a number of challenges. First, the simulation procedure only provides black-box functions which evaluate outputs for given input values. There are no analytical expressions characterizing the input–output relationship. Compared to a mathematical model where variables and equations inside the model are accessible to the optimization procedure, the black-box function provides rather limited information. In an inventory optimization problem, not only the objective function but also the constraint functions are black-box in nature, which

* Corresponding author. Tel.: +1 847 467 2943; fax: +1 847 491 3728.
E-mail address: you@northwestern.edu (F. You).

http://dx.doi.org/10.1016/j.compchemeng.2014.10.008
0098-1354/

Nomenclature

Dimension

N_D	number of design points
N_F	number of facilities
N_{MC}	number of sampling points (sample size) used in the Monte-Carlo method
$N_Q^{(p)}$	number of cutting plane constraints till iteration p
N_R	number of samples
N_T	number of simulation time points (simulation horizon)
N_X	length of $\mathbf{x}$

Index

d	design point
i, j	facility
p, q	iteration
m	element in $\mathbf{x}$
r	sample
t	time
s	sampling points

Parameter

c_i^H	unit holding cost of facility i
c_i^R	unit reorder cost of facility i
$fr_i^{\min}$	minimum fill rate of facility i
od_{it}	customer order demand received by facility i on day t
ot_i	order processing time for the order sent from facility j to facility i

Statistics

df	length of the half confidence interval of the test on the objective function
dp_i	length of the half confidence interval of the test on constraint function i
$sf^{(r)}$	standard deviation of $\hat{\varphi}^{(r,s)}(\mathbf{x}^{(p)})$
sp_i	significance level of hypothesis test
tf	statistic in the two-sample t-test for check if the objective function value is improved
α	significance level of hypothesis test

Variable in simulation model

AIO_i	average on-hand inventory of facility i
FOD_{it}	immediately fulfilled demand by facility i on day t
FR_i	fill rate of facility i
$INVC$	inventory cost of the inventory system
IOH_{it}	on-hand inventory of facility i at the end of day t
ROD_{it}	aggregate demand received by facility i on day t
ROI_{it}	reorder indicator of facility i on day t
SFD_i	sum of order demands immediately fulfilled by facility i
SRD_i	sum of order demands received by facility i
SRO_i	number of replenishment for facility i
$\Theta^{(r)}$	sample r (a finite set of sampling points)

Variable in optimization algorithm

$f_b^{(p)}$	value of $\varphi(\mathbf{x})$ at iteration p
$h_{b,i}^{(p)}$	value of $\psi_i(\mathbf{x})$ at iteration p
oq_i	order quantity of facility i
rp_i	reorder point of facility i
x_m	element m in $\mathbf{x}$
y_m	auxiliary variable representing the absolute value concerning x_m in $\mathbf{x}$

Vector/Matrix

$\mathbf{e}_{N_X}$	vector of all elements equal to 1
$\mathbf{f}_a^{(p)}$	estimated gradient of $\varphi(\mathbf{x})$ at iteration p
$\mathbf{fd}^{(p)}$	vector of the objective function values over the design points at iteration p
$\mathbf{h}_{a,i}^{(p)}$	estimated gradient of $\psi_i(\mathbf{x})$ at iteration p
$\mathbf{I}_{N_X}$	identity matrix with the dimension of N_X
$\mathbf{x}$	vector of inventory parameters
$\mathbf{x}^{(p)}$	value of x at iteration p
$\mathbf{xd}^{(p,d)}$	value of $\mathbf{x}$ of design point d at iteration p
$\boldsymbol{\theta}$	uncertain parameters

further complicates the optimization procedure. Second, the black-box functions returned by the simulation procedure contain noise due to uncertain parameters. As a result, simulation-based optimization is not merely optimization with black-box functions (Fu, 2002). The outputs of the simulation procedure are often expected values over the uncertain parameter region. In the absence of analytical expressions that link the uncertain input parameters with the uncertainty in the output, the expectations of output variables cannot be calculated exactly. A computational algorithm has to be adopted to estimate the expectations using sampling points. For a finite number of sampling points, there can be discrepancies between the true expectations and their estimates. The discrepancies need to be considered by the optimization procedure as they can have a significant impact on the optimal solution. For example, the discrepancies in the constraint functions can affect the solution feasibility. Such inaccuracies in the estimated black-box functions are often neglected in the literature. As a result, many existing methods fail to validate the optimality of the searched solution.

To address the aforementioned challenges, we propose a novel simulation-based optimization framework which integrates the agent-based modeling and simulation, the Monte-Carlo method, a cutting plane algorithm, an experimental design technique, and statistical hypothesis tests. The framework aims to optimize a multi-echelon inventory system under uncertainties. Its main components are shown in Fig. 1. An agent-based model is built according to the specifications of the multi-echelon system. The agent-based model simulates the inventory system and evaluates the performance measures for given inventory parameters, which define the black-box functions in the objective and the constraints for the optimization procedure. To evaluate the black-box functions, Monte Carlo method is applied which estimates the expectations in terms of the sample averages. The optimization problem with black-box functions is solved by a cutting plane algorithm, which is an iterative search by successively linearizing the black-box functions. The linearized functions are obtained from the first-order response surface models estimated by the fractional factorial experimental design technique. The optimization procedure returns the best parameter value searched in the iteration. The solution optimality is validated by statistical hypothesis tests conducted on the KKT (Karush–Kuhn–Tucker) conditions. When the tests are passed, a local optimal solution is obtained.

The remainder of the paper is organized as follows. A literature review of simulation-based optimization for inventory management and control is presented in Section 2. The problem statement is given in Section 3. The simulation-based optimization framework is proposed in Section 4. In Section 5, the framework is demonstrated by two case studies. The conclusion is given in Section 6.

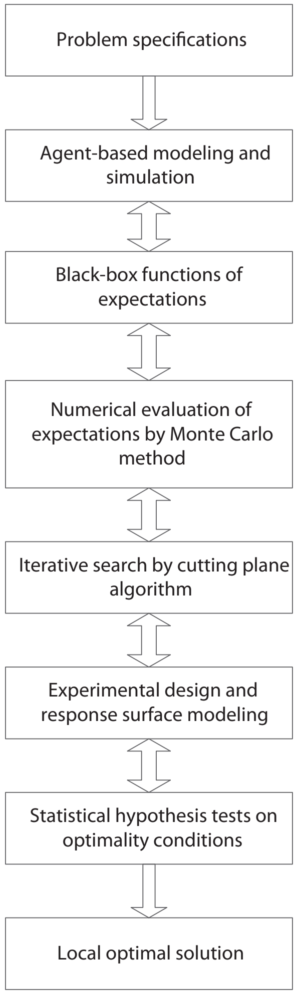

Fig. 1. Outline of the simulation-based optimization framework.

2. Literature review

Stochastic multi-echelon inventory optimization has been widely investigated using mathematical inventory theory. Mathematical programming methods usually solve highly abstracted models that may not capture the complexities of the real business processes. To ensure solvability, these models are often oversimplified with various restrictive assumptions (Kochel and Nielander, 2005). To overcome these restrictions, simulation of more realistic models is often conducted. Simulation models are able to accurately capture actual system behaviors by explicitly considering general specifications in the system organization, system peculiarities, the control policy, the randomness of exogenous and endogenous variables, etc. However, simulation is not a decision tool by itself. It cannot find the best parameter value that will result in an optimal performance. For latter, simulation should be incorporated into an appropriate optimization tool (Fu, 2002).

Simulation can be applied to various kinds of open mathematical models. As an inventory system is a stochastic dynamic system, it can be mathematically described by a set of stochastic differential equations or difference equations. The inventory system can be regarded as a classic coupled tank system in control engineering (Perea-Lopez et al., 2003; Subramanian et al., 2013). Then inventory optimization turns into a stochastic optimal control problem (Ivanov et al., 2012). However, a complicated optimal control problem is usually challenging to solve and some simulation-based optimization methods might be required. For example, Schwartz et al. (2006) proposed a simulation-based optimization framework by applying a simultaneous perturbation stochastic approximation method on inventory systems described by stochastic difference equations. Besides dynamic models, an inventory problem can also be formulated into a mixed integer linear program (MILP), to which a simulation-based optimization method is applied. Jung et al. (2004) presented such an approach by applying a stochastic search algorithm on a simulation model which contains an MILP inside.

Apart from the open mathematical models, simulation-based optimization methods can be applied to models implemented in different simulation toolkits. Because supply chain management is fundamentally concerned with coherence among multiple decision makers, an agent-based modeling framework is a natural choice for supply chain simulation (Swaminathan et al., 1998). Given that the simulation model provides black-box functions, meta-heuristic methods can be applied which only require the input and output data (Tekin and Sabuncuoglu, 2004). Mele et al. (2006) proposed a simulation-based optimization framework combining agent-based modeling and genetic algorithm. Mansouri (2006) developed a simulated annealing approach to solve a bi-criteria sequencing problem in a two-stage supply chain. Silva et al. (2006) developed a distributed optimization method for a logistic system and its suppliers using ant colonies. The ant colony optimization heuristic was also applied to supply chain design with multi objectives (Moncayo-Martinez and Zhang, 2011). Though meta-heuristic methods are applicable to a general black-box function and have the potential (though often not warranted) to find the global optimal solution, tuning the methods' parameters and validating the searched solution could be a challenge.

Instead of using a simulation model with black-box optimization, using a hybrid model has increasingly attracted attention recently (Blum et al., 2011). The hybrid model is a combination of a simulation model and a mathematical program. Its advantage is that the model structure is partially accessible to the optimization procedure so that it can utilize the information inside the model to employ efficient search algorithms. Consequently, a hybrid method can be more efficient than conventional simulation-based optimization methods which treat the entire model as a black-box function. Castro et al. (2011) developed a three-stage method combining mathematical programming and discrete-event simulation. Nikolopoulou and Ierapetritou (2012) integrated agent-based modeling with an MILP for solving supply chain problems involving production scheduling. A similar iterative method between simulation and optimization was also proposed by Almeder et al. (2009). Chu et al. (2014) proposed a cutting-plane algorithm to solve a hybrid problem combining an agent-based model and an MILP for integrated planning and scheduling. The problems that a hybrid method is applied to often have a hierarchical structure with different decision levels so that they can be segregated into different models: simulation models or mathematical programs. However, an inventory optimization problem may not possess such a feature allowing the hybrid model formulation.

3. Problem statement

There are a variety of inventory systems and no single method can be applied to all systems. In this work, we focus on a distribution inventory system, which is a divergent network (a tree structure) illustrated in Fig. 2. In the distribution inventory system, a facility has only one predecessor but it can have one or multiple successors. The inventory system is driven by customer order demands, which are random variables. We assume customer demand follows a known probability distribution. Besides the end facilities, customer orders can also be placed to the facilities in an intermediate level.

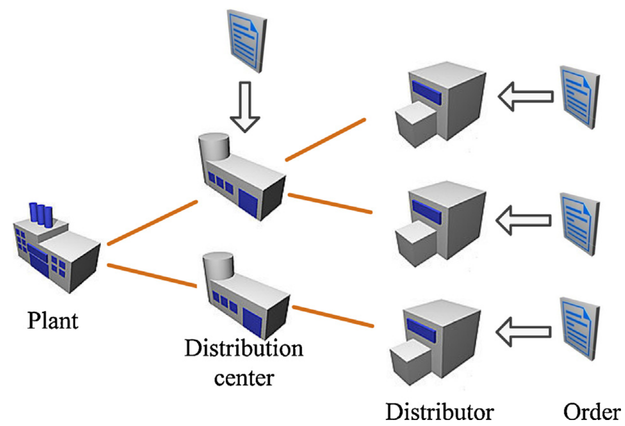

Fig. 2. Illustration of the distribution inventory system.

The inventory system considered in this work consists of a plant, distribution centers, and distributors. The plant is the root of the tree structure. The plant is assumed to always have enough inventory on hand to be able to send a shipment once it receives an order. The inventory control of the plant is not taken into consideration. Each facility, except the plant, is controlled by an inventory policy. In this work, we confine our attention to the (r, Q) policy where r denotes the reorder point and Q is the order quantity. An illustration of the (r, Q) policy is given in Fig. 3. The on-hand inventory (or physical inventory) is the amount of stored materials that are immediately available to satisfy the incoming order demands.

When a replenishment order is placed from a downstream node to an upstream facility in a multi-echelon system, a time period is required for preparing, handling, and delivering goods for that order. This time period is defined as the *order processing time* required for the downstream node, which is the duration from the moment that the order demand is satisfied by the on-hand inventory, to the moment that the shipment is received. The order processing time is a component of the lead time. The lead time is the duration ranging from the moment an order is placed to the moment the shipment is received. In Fig. 2, when a distribution center places an order to the plant, the lead time is equal to the order processing time because the plant is assumed to have unlimited on-hand inventory. However, the equality is not true when a distributor places an order to a distribution center as the lead time of the distributor may or may not include the lead time of the distribution center. If the distribution center has enough on-hand inventory when it receives an order from the distributor, the order is processed immediately and the lead time of the distributor is equal to the order processing time. In another case when the distribution center is out of stock, the order cannot be processed immediately. The distribution center needs to first replenish its own inventory. Afterwards, it can start processing the order from the distributor. In this case, the lead time of the distributor is the sum of the order processing time and the lead time of the distribution center. Obviously, in a multi-echelon inventory system, the lead times depend on the inventory parameters at each echelon. Consequently, we cannot assume that the lead times follow the same given distributions when the inventory parameters are changed. Instead, we specify the probability distributions of the order processing times and let the simulation procedure determine the lead times accordingly.

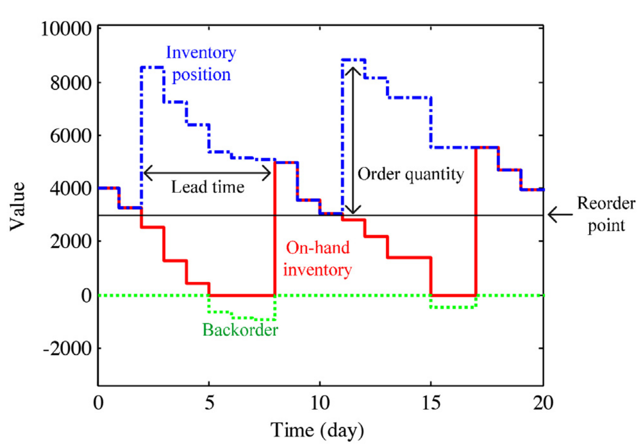

Fig. 3. Diagram of the (r, Q) inventory policy. The time has a discrete value of one day. For clarity, the backorder is shown as the negative value.

We adopt the common settings for inventory management. A facility sends orders to or receives shipments from its intermediate upstream facility. Though it is possible that a facility can be bypassed when it is out of stock, this special case is not considered in the paper.

When the demands are not satisfied by the on-hand inventory, they are backordered. The inventory position is defined as the on-hand inventory minus the backorder plus the amount of materials that has been ordered but has not yet arrived (the on-order inventory). To determine when to replenish inventory for a facility depends on its inventory position instead of its on-hand inventory. When the inventory position is equal to or less than the reorder point, the facility places an order with the order size/demand equal to the order quantity.

The objective of inventory optimization is to minimize the inventory cost while satisfying the order demands with acceptable service levels. In this work we assume the inventory cost is the aggregate of the holding cost (or carrying cost) associated with storing inventory and the reorder cost (or fixed ordering cost/setup cost) incurred each time an order is placed. The total cost can also include the shortage cost associated with the delays and remedial actions when inventory is not available to satisfy order demands. On certain occasions the shortage cost can become difficult to estimate accurately, because it is related to loss of customer goodwill that can be hard to quantify. In such cases a minimum service level is directly added as a constraint to the optimization problem, instead of including the shortage cost in the objective function (Diks et al., 1996).

There are various types of service levels. We focus on the β (or type 2) service level, which in this work is defined as the average fraction of demands delivered immediately from the on-hand inventory. This service level is also called the fill rate. The fill rate considers the probability of a stock out in terms of the size/volume of stockout and thus also includes size of backorders. For more rigorous definition of fill rate, see Chan (2003).

Specifically, the inventory optimization problem is stated as follows.

Assumptions

- Divergent distribution network
- (r, Q) inventory control policy
- Service level defined by fill rate
- Backorder of unsatisfied demand
- Random customer order demands and order processing times

Given

- Network structure
- Time horizon

- Unit holding cost and unit reorder cost
- Minimum desired service level for each facility
- Probability distributions of random parameters

To determine

- Reorder point and order quantity of each facility

Objective

- To minimize the inventory cost (=holding cost + reorder cost)
- To maintain service levels no less than the minimum desired values

4. Simulation-based optimization framework

Optimizing a multi-echelon inventory system encounters several challenges. First, an inventory system in practice is often composed of many facilities in different echelons. The inventory parameters of a facility affect not only the local inventory system but also those of other facilities in the entire network. Such interactions between facilities in a multi-echelon network require a simultaneous optimization approach because a simple sequential optimization approach that optimizes single-stage inventory systems one by one often leads to a poor performance. Another challenge is that a realistic model of an inventory system intrinsically contains various unknown parameters, because inventories are used as buffers against uncertainties of demand fluctuations, lead-time delays, etc. The stochastic nature imposes a major obstacle for applying commonly-used deterministic optimization methods. Next, an inventory system is also an order driven dynamic system (Puigjaner and Lainez, 2008), where the variables such as order demands, inventory levels, and received shipments all change with time. The dynamic behavior can have a critical impact on inventory in each echelon, such as the known "bullwhip effect" (i.e. the variance amplification of order quantities observed in supply chains) (Dejonckheere et al., 2003). In order to capture the dynamic behavior, it is important to track all variables evolving with time, which, as a result, requires a more complex multi-period model. In summary, a real-world inventory system, because of its nonlinear, stochastic, and time-dependent nature, and presence of complex interactions between echelons, can become quite challenging to optimize.

To address some of these challenges, a simulation-based optimization framework is proposed in this section. The framework comprises an agent-based modeling and simulation in Section 4.1, the optimization formulation in Section 4.2, the Monte-Carlo method and statistical hypothesis tests in Section 4.3, and the cutting plane algorithm in Section 4.4.

4.1. Agent-based modeling of multi-echelon system

An order driven inventory network is a distributed decision-making system. Each facility monitors its on-hand inventory and inventory position. It autonomously takes actions of satisfying the incoming orders, backordering the unmet demands, and replenishing the inventory. There is no central controller governing the entire network. Such a nature of the autonomous decision-making makes it suitable to model the inventory system by intelligent agents (Chu et al., 2013, 2015; Julka et al., 2002; Lee and Kimz, 2008). The dynamics of the inventory system can be easily captured by simulating the agent-based model, which is distinct from models formulated as differential equations (Parunak et al., 1998; Rahmandad and Sterman, 2008).

In the agent-based model, agents are created to model important constituents of an inventory system. As facilities are basic components of the system, facility agents are created accordingly to capture their behavior. Besides modeling the actions of how an individual facility manages its inventory, the agent-based model also needs to capture interactions among different facilities. The facilities interact with each other by a forward flow of materials and a backward flow of information. Thus, shipment agents and order agents are created, which are responsible for delivering goods and information among the facility agents. In addition, customer agents are created to generate customer orders which are exogenous inputs driving the inventory system. The interactions between these agents in an inventory system are exhibited in Fig. 4.

The functions of the agents are as follows:

- **Facility agents**: Monitor inventory, handle incoming orders, backorder unsatisfied demands, replenish inventory, and generate order processing times according to the specified distributions.
- **Order agents**: Record order information, including the sender, the receiver, the demand, and the status.
- **Shipment agents**: Record shipment information, including the sender, the receiver, the quantity, and the transportation time.
- **Customer agents**: Generate customer orders according to the given distributions.

The objective of the agent-based modeling is to simulate the inventory system and return the performance measures for the given inventory parameters. The performance measures are black-box functions to be used by the optimization procedure to search

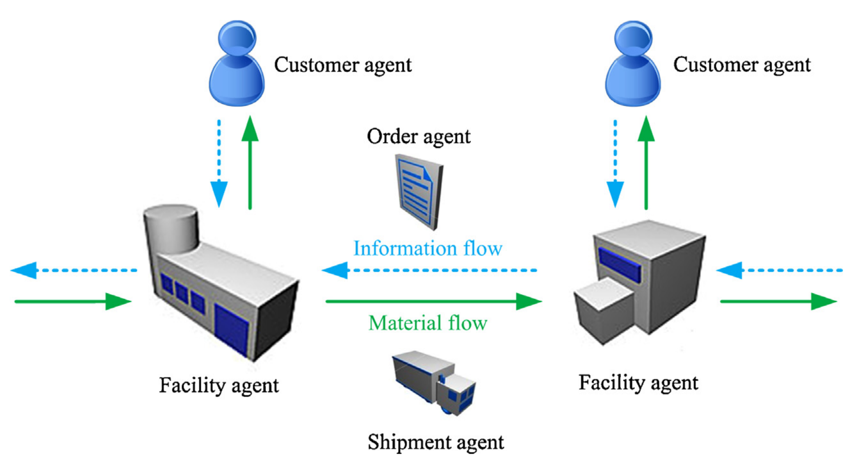

Fig. 4. The agent-based model for a multi-echelon inventory system. For clarity, only two facilities are shown.

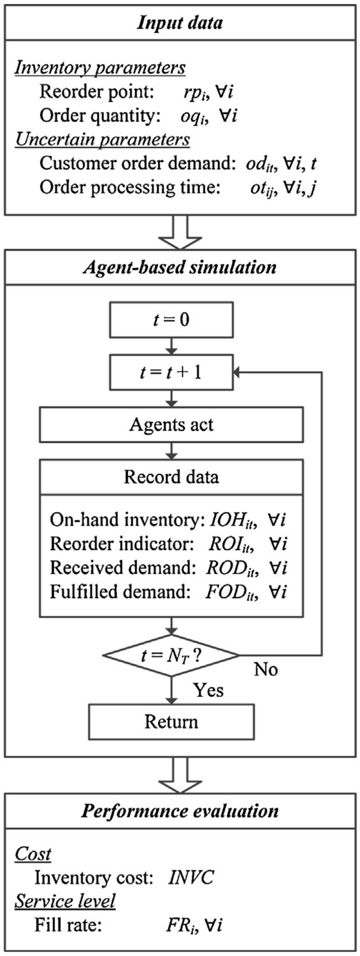

Fig. 5. Agent-based simulation that evaluates the black-box functions for the optimization procedure.

for the optimal inventory parameters. The diagram of the agent-based simulation that evaluates the black-box functions are shown in Fig. 5. In the simulation model, a facility is index by i. The simulation time is discrete with the unit of one day. The simulation time point is indexed by t and the simulation horizon is indicated by N_T.

Each facility manages its inventory according to the (r, Q) policy. The reorder point of facility i is denoted by rp_i and its order quantity is denoted by oq_i. The customer order demand received by facility i on day t is denoted by od_{it}. As the order processing times are random variables in the inventory system, they are also regarded as inputs of the agent-based simulation as to investigate their effects on the inventory performances. The order processing time for the order sent from facility j to facility i is denoted by ot_{ij}. Note that the order processing time, unlike customer demand, is not a time-dependent quantity. We should also note that the inventory parameters and the uncertain parameters are all variables in the inventory system. However, they are fixed parameters for the agent-based model and their values are changed by the optimization procedure outside the simulation procedure.

The simulation starts from $t=0$ and terminates at $t=N_T$. The review period is one day. Each day, the facility agent interacts with each other to manage their inventories according to the (r, Q) policy. At the end of day t, facility agent i records the on-hand inventory, denoted by IOH_{it}. It also records the indicator variable ROI_{it} for inventory replenishment. When an order is placed to replenish inventory of facility i at day t, $ROI_{it}=1$. The on-hand inventory and the reorder indicator are used to calculate the inventory cost. To calculate the fill rate, facility agent i records the aggregate demand of incoming orders on day t, denoted by ROD_{it}, which is the sum of the exogenous customer order demand and the endogenous demand from downstream facilities. If the on-hand inventory is insufficient to satisfy all incoming orders, the unsatisfied demand is backordered. The backorders will be satisfied when the inventory is sufficient. As the backorders affect the service level, the total amount of backorders is constrained by the threshold of the service level. The portion of the order demand that is satisfied immediately by the on-hand inventory of facility i on day t is denoted by FOD_{it}.

Using the daily records, the performance measures of the inventory system are evaluated after the simulation. The average on-hand inventory of facility i is

$$AIO_i = \frac{1}{N_T}\sum_{t=1}^{N_T} IOH_{it}, \quad \forall i \tag{1}$$

The number of replenishments for facility i in the time horizon is

$$SRO_i = \sum_{t=1}^{N_T} ROI_{it}, \quad \forall i \tag{2}$$

The inventory cost of the inventory system is

$$INVC = \sum_{i=1}^{N_F} c_i^H AIO_i + \sum_{i=1}^{N_F} c_i^R SRO_i \tag{3}$$

where c_i^H is the unit holding cost of facility i and c_i^R is the unit reorder cost. The number of facilities is N_F.

The fill rate of a facility is also calculated from the daily data. The sum of order demands received by facility i in the time horizon is

$$SRD_i = \sum_{t=1}^{N_T} ROD_{it} \tag{4}$$

and the sum of demands fulfilled without delay is

$$SFD_i = \sum_{t=1}^{N_T} FOD_{it} \tag{5}$$

The fill rate of the facility is calculated as

$$FR_i = \frac{SFD_i}{SRD_i}, \quad \forall i \tag{6}$$

To facilitate the discussion on the optimization procedure, the black-box functions evaluated by the agent-based simulation are expressed explicitly as

$$INVC = f(\mathbf{x}, \theta) \tag{7}$$

$$FR_i = g_i(\mathbf{x}, \theta), \quad \forall i \tag{8}$$

where the input data are stacked into the vector of inventory parameters as

$$\mathbf{x} = [rp_1, \ldots, rp_{N_F}, oq_1, \ldots, oq_{N_F}] \tag{9}$$

and the vector of uncertain parameters as

$$\theta = [od_{11}, \ldots, od_{N_F N_T}, ot_{11}, \ldots, ot_{N_F N_F}] \tag{10}$$

4.2. Formulation of inventory optimization problem

Inventory optimization aims to minimize the inventory cost while maintaining acceptable service levels. The optimization problem is formulated using the performance functions in Eqs. (7) and (8) evaluated by the agent-based simulation. Note that, besides the inventory parameters, the performance functions in Eqs. (7) and (8) also depend on the uncertain parameters. As probability distributions of the uncertain parameters are given, the expected performance measures are calculated in the optimization problem.

Using the vector expression in Eq. (10), the probability density function of the uncertain parameters is assumed to be

$$\theta \sim p(\theta) \tag{11}$$

The expected inventory cost is defined as

$$\varphi(\mathbf{x}) = E_\theta[f(\mathbf{x}, \theta)] \tag{12}$$

where the expectation is calculated by

$$E_\theta[f(\mathbf{x}, \theta)] = \iint_\theta f(\mathbf{x}, \theta) p(\theta) d\theta \tag{13}$$

The expected fill rate of a facility is

$$\psi_i(\mathbf{x}) = \frac{E_\theta[SFD_i(\mathbf{x}, \theta)]}{E_\theta[SRD_i(\mathbf{x}, \theta)]}, \quad \forall i \tag{14}$$

where SFD_i and SRD_i are defined in Eqs. (4) and (5), respectively. An alternative definition of the expected fill rate is the expected value of $g_i(\mathbf{x}, \theta)$ in Eq. (8). However, the definition in Eq. (14) appears more commonly in literature (Chen et al., 2003). It will be seen in Eq. (22), when the expectations are evaluated by the Monte-Carlo method, the definition in Eq. (14) is the fraction of immediately satisfied demands over all samples. This is a nature extension of the fill rate (in Eq. (6)) for a single sample of uncertain parameters.

Using the expected performance functions, the inventory optimization problem is formulated as

$$(\text{INV\_OPT}) \quad \min_{\mathbf{x}} \varphi(\mathbf{x}) \tag{15}$$

$$\text{s.t.}$$

$$\psi_i(\mathbf{x}) \geq fr_i^{\min}, \quad \forall i \tag{16}$$

where $fr_i^{\min}$ is the minimum desired service level of facility i. The functions $\varphi(\mathbf{x})$ and $\psi_i(\mathbf{x})$ are defined in Eqs. (12) and (14), respectively. The problem (INV_OPT) is a nonlinear program (NLP) with continuous decision variables stacked in the vector **x**. However, as the objective function and the constraint functions are all black-box functions, the problem cannot be solved by a gradient-based optimization solver. In addition, the black-box functions are expectations which cannot be calculated analytically, incurring extra difficulties in solving the problem.

4.3. Expectation evaluation by Monte-Carlo method

Expectation evaluation requires calculating a high-dimensional integral as that in Eq. (13). Because the performance functions evaluated by the agent-based simulation have no analytical expressions and there are many uncertain parameters, exact calculation of the expectations is not feasible. A general alternative is to numerically calculate the expectations by the Monte-Carlo method that estimates the expectations by averages over a finite number of sampling points (Asmussen and Glynn, 2007).

In the Monte-Carlo method, a sample of uncertain parameters is generated, which is a finite set of sampling points defined as

$$\Theta^{(r)} = \left\{ \theta^{(r,1)}, \theta^{(r,2)}, \ldots, \theta^{(r,s)}, \ldots, \theta^{(r,N_{MC})} \right\} \tag{17}$$

The sample is denoted by $\boldsymbol{\Theta}$ and indexed by r. A sampling point in the sample is indexed by s. The sample size is indicated by N_{MC}. The parameters are sampled from the probability density function $p(\boldsymbol{\theta})$.

At each sampling point $\theta^{(r,s)}$, the performance functions are evaluated as

$$f^{(r,s)}(\mathbf{x}) = f(\mathbf{x}, \theta^{(r,s)}) \tag{18}$$

$$SFD_i^{(r,s)}(\mathbf{x}) = SFD_i(\mathbf{x}, \theta^{(r,s)}), \quad \forall i \tag{19}$$

$$SRD_i^{(r,s)}(\mathbf{x}) = SRD_i(\mathbf{x}, \theta^{(r,s)}), \quad \forall i \tag{20}$$

Then the estimated expectations are

$$\hat{\varphi}^{(r)}(\mathbf{x}) = \frac{1}{N_{MC}} \sum_{s=1}^{N_{MC}} f^{(r,s)}(\mathbf{x}) \tag{21}$$

$$\hat{\psi}_i^{(r)}(\mathbf{x}) = \frac{\sum_{s=1}^{N_{MC}} SFD_i^{(r,s)}(\mathbf{x})}{\sum_{s=1}^{N_{MC}} SFR_i^{(r,s)}(\mathbf{x})}, \quad \forall i \tag{22}$$

To distinguish the estimated value from true value, the estimated value is expressed with a caret.

Though applicable to a general black-box function, the Monte-Carlo method is essentially an estimation method. The estimated value approaches to the true value as the number of sampling points tends to infinity according to the law of large numbers (Shapiro et al., 2009). Unfortunately, a finite set of sampling points has to be employed in practice and the discrepancy between the estimated value and the true value is unavoidable. As the discrepancy can have an impact on the optimization procedure, e.g. affecting the solution feasibility, it needs to be investigated carefully. In this work, statistical hypothesis tests are conducted to investigate the results returned by the Monte-Carlo method. Specifically, the hypothesis tests are conducted on the objective function and the constraint functions in the optimization problem.

A two-sample t-test is conducted to check if there is an improvement in the objective function value. The absence of an improved objective value provides a criterion to terminate the search algorithm in the optimization procedure. Suppose the objective function value at an iteration is $\hat{\varphi}^{(r)}(\mathbf{x}^{(p)})$ where p indicates the iteration. The objective function value at the next iteration is assumed to be $\hat{\varphi}^{(r+1)}(\mathbf{x}^{(p+1)})$. Because the objective function values are estimated with different samples, the improvement cannot be checked by simply comparing the two estimated values. Instead, a hypothesis test is formulated as

$$H_0: \quad \varphi(\mathbf{x}^{(p+1)}) \leq \varphi(\mathbf{x}^{(p)}) \tag{23}$$

$$H_1: \quad \varphi(\mathbf{x}^{(p+1)}) > \varphi(\mathbf{x}^{(p)}) \tag{24}$$

If the null hypothesis is rejected, there is no improvement in iteration $p+1$ so that the search is terminated. The hypothesis test is conducted based on the test statistic as

$$tf = \frac{\hat{\varphi}^{(r+1)}(\mathbf{x}^{(p+1)}) - \hat{\varphi}^{(r)}(\mathbf{x}^{(p)})}{\sqrt{\left((sf^{(r+1)})^2 + (sf^{(r)})^2 \right) / N_{MC}}} \tag{25}$$

where $sf^{(r+1)}$ and $sf^{(r)}$ are standard derivations of $f^{(r+1,s)}(\mathbf{x}^{(p+1)})$ and $f^{(r,s)}(\mathbf{x}^{(p)})$, respectively. The null hypothesis (23) is rejected if (Toutenburg, 2009)

$$tf > t_{1-\alpha, 2(N_{MC}-1)} \tag{26}$$

where $t_{1-\alpha,2(N_{MC}-1)}$ is the value of the inversed t distribution at the quantile $1-\alpha$ with the degree of freedom $2(N_{MC}-1)$. The significance level of the test is designated by α. Inequality (26) is further expressed as

$$\hat{\varphi}^{(r+1)}(\mathbf{x}^{(p+1)}) > \hat{\varphi}^{(r)}(\mathbf{x}^{(p)}) + df \tag{27}$$

where

$$df = t_{1-\alpha,2(N_{MC}-1)} \sqrt{\left((sf^{(r+1)})^2 + (sf^{(r)})^2 \right) / N_{MC}} \tag{28}$$

Another hypothesis test is conducted to validate the solution feasibility. As $\hat{\psi}_i^{(r)}(\mathbf{x})$ is an estimate according to Eq. (22), the inequality $\hat{\psi}_i^{(r)}(\mathbf{x}) \geq fr_i^{\min}$ does not imply $\psi_i(\mathbf{x}) \geq fr_i^{\min}$. To ensure that the latter inequality is satisfied with a given confidence level of $1-\alpha$, the hypothesis test is

$$H_0: \quad \psi_i(\mathbf{x}) \geq fr_i^{\min}, \quad \forall i \tag{29}$$

$$H_1: \quad \psi_i(\mathbf{x}) < fr_i^{\min}, \quad \forall i \tag{30}$$

If the null hypothesis is not rejected, the solution of $\mathbf{x}$ is feasible.

The test is conducted based on the distribution of $\hat{\psi}_i^{(r)}(\mathbf{x})$. Unlike the estimate of the objective function, $\hat{\psi}_i^{(r)}(\mathbf{x})$ calculated in Eq. (22) is not an average. Instead, it is a ratio of two averages. Thus, the distribution of $\hat{\psi}_i^{(r)}(\mathbf{x})$ cannot be obtained simply from those of $SFD_i^{(r,s)}(\mathbf{x})$ and $SRD_i^{(r,s)}(\mathbf{x})$. A sampling approach is employed to acquire the distribution of $\hat{\psi}_i^{(r)}(\mathbf{x})$. From N_R samples of $\hat{\psi}_i^{(r)}(\mathbf{x})$, where $r = 1, \ldots, N_R$, the standard derivation is calculated as

$$sp_i = \sqrt{\frac{1}{N_R-1}\sum_{r=1}^{N_R}\left(\hat{\psi}_i^{(r)}(\mathbf{x}) - \frac{\sum_{r'=1}^{N_R}\hat{\psi}_i^{(r')}(\mathbf{x})}{N_R}\right)^2}, \quad \forall i \tag{31}$$

The left end of the $1-\alpha$ confidence interval is $\hat{\psi}_i^{(r)}(\mathbf{x}) - sp_i \cdot t_{1-\alpha,N_R-1}$. Therefore, the null hypothesis is not rejected when $\hat{\psi}_i^{(r)}(\mathbf{x}) - sp_i \cdot t_{1-\alpha,N_R-1} \geq fr_i^{\min}$, or equivalently

$$\hat{\psi}_i^{(r)}(\mathbf{x}) \geq fr_i^{\min} + dp_i, \quad \forall i \tag{32}$$

where dp_i is the length of the half confidence interval, calculated as

$$dp_i = sp_i \cdot t_{1-\alpha,N_R-1}, \quad \forall i \tag{33}$$

The value of dp_i can be regarded as a safety distance. It guarantees $\psi_i(\mathbf{x}) \geq fr_i^{\min}$ is satisfied with the confidence level of $1-\alpha$ when $\psi_i(\mathbf{x})$ is estimated by the Monte-Carlo method. After the safety distance is added, the fill rate constraint (16) becomes

$$\psi_i(\mathbf{x}) \geq fr_i^{\min} + dp_i, \quad \forall i \tag{34}$$

Finally, the last hypothesis test is conducted to test if the inequality (34) is active, which is

$$H_0: \quad \psi_i(\mathbf{x}) = fr_i^{\min} + dp_i, \quad \forall i \tag{35}$$

$$H_1: \quad \psi_i(\mathbf{x}) > fr_i^{\min} + dp_i, \quad \forall i \tag{36}$$

When the feasibility test in Eq. (29) is passed, $\psi_i(\mathbf{x})$ is no less than $fr_i^{\min}$. Therefore, the alternative hypothesis in Eq. (36) is only one lateral. The inequality (34) is active if the null hypothesis (35) is not rejected, equivalently when

$$\begin{aligned}\hat{\psi}_i^{(r)}(\mathbf{x}) &< (fr_i^{\min} + dp_i) + dp_i \\ &= fr_i^{\min} + 2dp_i\end{aligned}, \quad \forall i \tag{37}$$

This test is used to validate the KKT conditions.

The two statistical tests on a fill rate function can be illustrated in Fig. 6. The estimate $\hat{\psi}_i^{(r)}(\mathbf{x})$ is a random variable dependent upon sample r. If inequality (34) is satisfied, i.e. $fr_i^{\min} + dp_i \leq \hat{\psi}_i^{(r)}(\mathbf{x})$, then the minimum service level $fr_i^{\min}$ is outside the confidence interval. Thus, the service level constraint $\psi_i(\mathbf{x}) \geq fr_i^{\min}$ is satisfied with the probability confidence. Also, if inequality (37) is satisfied, then the safety service level $fr_i^{\min} + dp_i$ falls inside the confidence interval, which cannot be distinguished from $\hat{\psi}_i^{(r)}(\mathbf{x})$ in a statistical sense. In this case, the test on their equality (35) is not rejected.

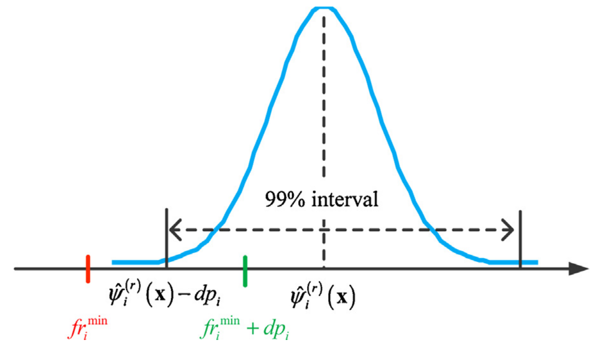

Fig. 6. Illustration of statistical tests on the fill rate.

4.4. Iterative search by cutting plane method

After the expectations are evaluated by the Monte-Carlo method, the optimization problem with the estimated functions becomes

$$\text{(INV\_OPT\_EST)} \qquad \min_{\mathbf{x}} \ \hat{\varphi}(\mathbf{x}) \qquad \text{s.t.} \tag{38}$$

$$\hat{\psi}_i(\mathbf{x}) \geq fr_i^{\min} + dp_i, \quad \forall i \tag{39}$$

The safety distances, dp_i, $\forall i$, are added to the service level constraints to compensate for the possible discrepancies between the estimated fill rates and the true values. The optimization problem is an NLP with black-box functions and it is solved by the cutting plane algorithm presented in this subsection.

The cutting plane algorithm is an iterative search. It starts from a feasible solution, denoted by $\mathbf{x}^{(p)}$. Explicitly, the vector $\mathbf{x}^{(p)}$ consists of elements as

$$\mathbf{x}^{(p)} = \left[x_1^{(p)}, \ldots, x_m^{(p)}, \ldots, x_{N_X}^{(p)}\right]^T \tag{40}$$

where an element is indexed by m and the dimension of $\mathbf{x}^{(p)}$ is N_X. Then the algorithm linearizes the objective function and the constraint functions at $\mathbf{x}^{(p)}$, which become

$$\hat{\varphi}^{(r)}(\mathbf{x}) \approx (\mathbf{x} - \mathbf{x}^{(p)})^T f_a^{(p)} + f_b^{(p)} \tag{41}$$

$$\hat{\psi}_i^{(r)}(\mathbf{x}) \approx (\mathbf{x} - \mathbf{x}^{(p)})^T \mathbf{h}_{a,i}^{(p)} + h_{b,i}^{(p)}, \quad \forall i \tag{42}$$

The coefficients of $\mathbf{f}_a^{(p)}$, $\mathbf{h}_{a,i}^{(p)}$, $f_b^{(p)}$, and $h_{b,i}^{(p)}$ are obtained from first-order response surface models, which are estimated by an experimental design technique.

We use the fractional factorial design (Box et al., 2005) to estimate the response surface models. The design technique generates a set of design points around $\mathbf{x}^{(p)}$ as

$$\mathbf{xd}^{(p,d)} = \mathbf{x}^{(p)} + \rho \cdot \mathbf{d}_d^T, \quad \forall d \tag{43}$$

where $\mathbf{xd}^{(p,d)}$ denotes the design point indexed by d, ρ is the scaling factor, and $\mathbf{d}_d$ is a design vector from the design matrix of

$$\mathbf{DM} = \begin{bmatrix} \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_d \\ \vdots \\ \mathbf{d}_{N_D} \end{bmatrix}_{N_D \times N_X} \tag{44}$$

The number of design points is indicated by N_D. The design matrix has elements with value of −1 or 1 and its columns are orthogonal with each other, having

$$\mathbf{DM}^T\mathbf{DM} = N_D\mathbf{I}_{N_X} \tag{45}$$

where $\mathbf{I}_{N_X}$ is the identity matrix with the dimension N_X.

At a design point $\mathbf{xd}^{(p,d)}$, the objective function is evaluated as $\hat{\varphi}^{(r)}(\mathbf{xd}^{(p,d)})$. All function values over the design points form a vector given by

$$\mathbf{fd}^{(p)} = \left[\hat{\varphi}^{(r)}(\mathbf{xd}^{(p,1)}), \ldots, \hat{\varphi}^{(r)}(\mathbf{xd}^{(p,d)}), \ldots, \hat{\varphi}^{(r)}(\mathbf{xd}^{(p,N_D)})\right]^T \tag{46}$$

The interception value of $f_b^{(p)}$ in Eq. (41) is calculated as

$$f_b^{(p)} = \hat{\varphi}^{(r)}(\mathbf{x}^{(p)}) \tag{47}$$

and the coefficient vector is estimated from the least squares solution

$$\begin{aligned}\mathbf{f}_a^{(p)} &= \frac{1}{\rho}(\mathbf{DM}^T\mathbf{DM})^{-1}\mathbf{DM}^T(\mathbf{fd}^{(p)} - f_b^{(p)}\mathbf{e}_{N_X}) \\ &= \frac{1}{\rho N_D}\mathbf{DM}^T(\mathbf{fd}^{(p)} - f_b^{(p)}\mathbf{e}_{N_X})\end{aligned} \tag{48}$$

The column vector $\mathbf{e}_{N_X}$ has all elements equal to 1. Similarly, the coefficient vectors $\mathbf{h}_{a,i}^{(p)}$ and the interception $h_{b,i}^{(p)}$ are obtained by the experimental design technique.

After linearization, the cutting plane algorithm solves the following linearized problem

$$(\text{INV\_LINp}) \quad \mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}}\left(\mathbf{x} - \mathbf{x}^{(p)}\right)^T\mathbf{f}_a^{(p)} + f_b^{(p)} \quad \text{s.t.} \tag{49}$$

$$(\mathbf{x} - \mathbf{x}^{(q)})^T\mathbf{h}_{a,i}^{(q)} + h_{b,i}^{(q)} \geq fr_i^{\min} + dp_i, \quad \forall i, \quad q \leq N_Q^{(p)} \tag{50}$$

$$\sum_{m=1}^{N_X}\left|x_m - x_m^{(p)}\right| \leq \delta \tag{51}$$

The objective in Eq. (49) is the linearized $\hat{\varphi}^{(r)}(\mathbf{x})$ at $\mathbf{x}^{(p)}$ and the functions on the left-hand side of constraint (50) are linearized $\hat{\psi}_i^{(r)}(\mathbf{x})$. The constraint includes all cutting planes generated in the previous iteration indexed by q. $N_Q^{(p)}$ denotes the number of cutting plane constraints generated till the current iteration p. Because at each iteration multiple cutting plane constraints can be added, the number of constraints $N_Q^{(p)}$ is typically not equal to the iteration index p. The first-order response surface models are only local approximations of the original nonlinear functions. Thus, inequality (51) is added to confine the search scope. The value of δ is updated adaptively during the iterative search. The absolute values in the constraint can be expressed by linear inequalities as

$$\sum_{m=1}^{N_X} y_m \leq \delta \tag{52}$$

$$y_m \geq x_m - x_m^{(p)}, \quad \forall m \tag{53}$$

$$y_m \geq x_m^{(p)} - x_m, \quad \forall m \tag{54}$$

where y_m is a non-negative variable introduced to express the absolute value. The optimal solution of the linearized problem (INV_LIN$_p$) provides the value for the next iteration.

The iterative procedure of the cutting plane algorithm is exhibited in Fig. 7. First, the search algorithm is initialized by setting the number of sampling points, N_{MC}, for the Monte-Carlo method. Then the iteration index p is set to one and the search starts from a feasible solution $\mathbf{x}^{(1)}$ that satisfies the service level constraints. A feasible solution is obtained by setting the inventory parameters to large values. It guarantees a feasible solution because the fill rates are usually higher for large values of reorder points and the order quantities. The black-box functions are evaluated by the Monte-Carlo method using a sample, indexed by r. At each iteration, a new sample is generated, different from the previous ones.

At iteration p, the current solution is $\mathbf{x}^{(p)}$. The objective function and the constraint functions are linearized at $\mathbf{x}^{(p)}$ by the first-order response surface models. Then the linearized problem (INV_LIN$_p$) is solved to return the solution $\mathbf{x}^*$. Before $\mathbf{x}^*$ is used at the next iteration, its feasibility is validated by the hypothesis test in Eqs. (29)–(32). If the test fails, a new cutting plane constraint is generated and appended to the linearized problem. If the feasibility test is passed, another hypothesis test is conducted to check if the objective function value has improved in this iteration. The two-sample t-test in Eqs. (23)–(28) is conducted. If the test is passed, the next iteration begins with $\mathbf{x}^{(p+1)} = \mathbf{x}^*$. Otherwise, the solution $\mathbf{x}^*$ fails to achieve an improved objective function value. One reason for this can be that the search range denoted by δ in constraint (51) is so large that the linearized objective function cannot well approximate the nonlinear one. The remedy is to shrink δ and solve the linearized problem again. The iteration terminates when the search range is smaller than the specified value $\delta^{\min}$.

Next, the optimality of the returned solution $\mathbf{x}^{(p)}$ is validated by a hypothesis test on the KKT conditions of the problem (INV_OPT), which will be presented in Eqs. (55)–(67). If the test is passed, a local optimal solution of $\mathbf{x}^{(p)}$ is returned. Otherwise, a two-sample t-test is conducted again to test if the reduction in the objective function value from $\varphi(\mathbf{x}^{(1)})$ to $\varphi(\mathbf{x}^{(p)})$ is statistically larger than a threshold value γ. If there is no improvement, the algorithm terminates with a sub-optimal solution. Otherwise, when the final solution $\mathbf{x}^{(p)}$ is better than the initial solution $\mathbf{x}^{(1)}$, the algorithm restarts from $\mathbf{x}^{(p)}$. To refine the results, the number of sampling points in the Monte-Carlo method is increased within a bounded range. With a better starting guess and bigger sample size, the algorithm restarts and retraces all the aforementioned steps.

Restarting the algorithm at a better solution addresses some inherent difficulties. The first one arises from the Monte-Carlo method. A critical parameter in the method is the number of sampling points, N_{MC}. As an asymptotic method, the Monte-Carlo method returns an accurate estimate when N_{MC} is sufficiently large. However, a large N_{MC} incurs extra computational efforts. When the Monte-Carlo method is applied only once for a common simulation purpose, the increased computational burden is often affordable. In contrast, in a simulation-based optimization framework, the Monte-Carlo method has to be applied many times at every iteration. As a result, increasing N_{MC} has a more significant impact on the computational performance. Therefore, selecting a proper N_{MC} is of great importance. Unfortunately, the selection is based on the information of the distribution that the Monte-Carlo method aims to estimate. If the variations in the estimated value are small, a small number of sampling points is adequate. However, if the variations are large, a large number of sampling points should be used to achieve the same accuracy. Determining the number of sampling points is further complicated by the optimization procedure because the decision variables, which change during the iterative procedure, may affect the accuracy of the Monte-Carlo method. As the distribution to be estimated is unknown, the Monte-Carlo method usually adopts a trial-and-error approach to determine the number of sampling points (Asmussen and Glynn, 2007). Restarting the cutting plane algorithm accommodates the trial-and-error approach. Initially, when the starting solution is far from the optimal one, there is no need to have a very accurate estimation and the number of sampling points can be small. When the solution is close to the optimal one, a large number of sampling points is required to ensure an accurate estimation. The adaptive approach

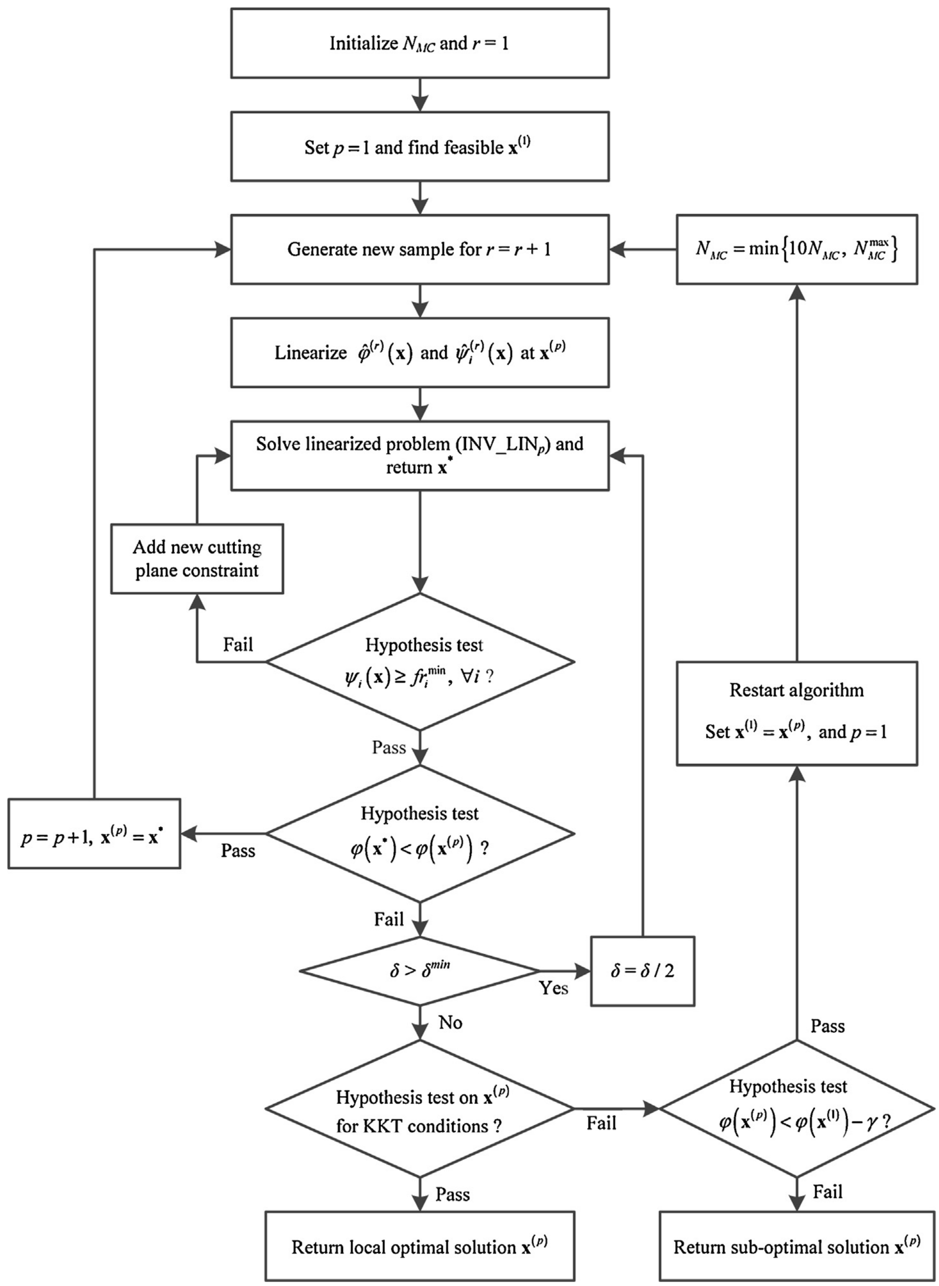

Fig. 7. Iterative procedure of the cutting plane algorithm along with hypothesis tests.

for determining N_{MC} reduces the computational efforts while ensuring the estimation accuracy.

The second benefit from restarting the cutting plane algorithm at a new better solution is to alleviate the non-convexity effects. As a service level measure, the fill rate is often a convex function when the service level is close to one (Atlason et al., 2004; Jung et al., 2008). However, the convexity is not guaranteed in general. The cutting plane algorithm is a heuristic as it may inadvertently exclude some feasible region, possibly resulting in a sub-optimal solution. The impact of non-convexity is alleviated by removing all previous cutting plane constraints when the algorithm restarts at a new solution.

The final step in the iterative procedure in Fig. 7 is the optimality test on the final solution $\mathbf{x}^{(p)}$. Specifically, the KKT conditions of the problem (INV_OPT) are

$$\nabla\varphi(\mathbf{x}^{(p)}) = \sum_i \lambda_i \nabla\psi_i(\mathbf{x}^{(p)}) \tag{55}$$

$$\lambda_i(\psi_i(\mathbf{x}^{(p)}) - fr_i^{\min} - dp_i) = 0, \quad \forall i \tag{56}$$

$$\psi_i(\mathbf{x}^{(p)}) - fr_i^{\min} \geq 0, \quad \forall i \tag{57}$$

$$\lambda_i \geq 0, \quad \forall i \tag{58}$$

where λ_i is a Lagrange multiplier. Eq. (55) is the stationary equality where the gradient of the objective function is a linear combination of the gradients of the constraint functions. According to inequality (58), all coefficients in the linear combination should be non-negative.

Eq. (56) is the complementarity equality. Instead of the common expression, $\lambda_i(\psi_i(\mathbf{x}^{(p)}) - fr_i^{\min}) = 0$, in the deterministic case, the safety distance dp_i is added. Recalling Fig. 6, to ensure the feasibility of constraint (57) is satisfied with the significance level of α, $fr_i^{\min}$ should be outside the confidence interval. Consequently, the test of $\psi_i(\mathbf{x}^{(p)}) = fr_i^{\min}$ will get automatically rejected, implying that the constraint (57) will not be active. This difficulty arises from the fact that $\psi_i(\mathbf{x}^{(p)})$ cannot be calculated exactly. As its estimate inevitably involves noise, the constraint functions cannot be pushed to the bounds, and safety distances are always required. In other words, ensuring constraint feasibility automatically excludes the situation when constraints can become active in the stochastic scenario, which is illustrated in Fig. 6. Therefore, a safety distance is added to the complimentary equality (56). The equality $\psi_i(\mathbf{x}^{(p)}) = fr_i^{\min} + dp_i$ indicates that the constraint function can be pushed to the bounds.

Considering that the black-box functions $\varphi(\mathbf{x}^{(p)})$ and $\psi_i(\mathbf{x}^{(p)})$ are not known, their estimates $\hat{\varphi}^{(r)}(\mathbf{x}^{(p)})$ and $\hat{\psi}_i^{(r)}(\mathbf{x}^{(p)})$ are used in the test. As the primal feasibility constraints (57) have been taken into account in the cutting plane algorithm, they get satisfied. The focus is placed on the complementary constraint (56) and the stationary equality (55).

Conducting the hypothesis test in Eqs. (35)–(37), the active constraints $\psi_i(\mathbf{x}^{(p)}) = fr_i^{\min} + dp_i$ can be identified. The index set for the active constraints are generated as

$$I_a = \{i | \psi_i(\mathbf{x}^{(p)}) = fr_i^{\min} + dp_i\} \tag{59}$$

According to the complementary constraint (56), the Lagrange multipliers should be zero for the non-active constraints, specifically

$$\lambda_i = 0, \quad \forall i \notin I_a \tag{60}$$

Finally, the stationary equality (55) is verified, which becomes

$$\nabla\varphi(\mathbf{x}^{(p)}) = \sum_{i \in I_a} \lambda_i \nabla\psi_i(\mathbf{x}^{(p)}) \tag{61}$$

Again, the exact values of $\nabla\varphi(\mathbf{x}^{(p)})$ and $\nabla\psi_i(\mathbf{x}^{(p)})$ cannot be known. Using their estimates, the stationary equality becomes

$$\nabla\hat{\varphi}(\mathbf{x}^{(p)}) = \sum_{i \in I_a} \lambda_i \nabla\hat{\psi}_i(\mathbf{x}^{(p)}) \tag{62}$$

where the gradients are estimated from Eqs. (41) and (42) as

$$\nabla\hat{\varphi}(\mathbf{x}^{(p)}) = \mathbf{d}_a^{(p)} \tag{63}$$

$$\nabla\hat{\psi}_i(\mathbf{x}^{(p)}) = \mathbf{x}_{a,i}^{(p)}, \quad \forall i \tag{64}$$

Because the estimated gradients contain noise, equality (62) cannot be satisfied exactly as in the deterministic case. In the stochastic scenario, the equality is regarded as a linear regression model. Thus, we test if the linear regression model has a good fit. The goodness of fit can be validated by solving the following constrained least squares problem

$$(\text{CON\_LS}) \quad \varepsilon = \min_{\lambda_i \in I_a} \left(\nabla\hat{\varphi}(\mathbf{x}^{(p)}) - \sum_{i \in I_a} \lambda_i \nabla\hat{\psi}_i(\mathbf{x}^{(p)}) \right)^T \left(\nabla\hat{\varphi}(\mathbf{x}^{(p)}) - \sum_{i \in I_a} \lambda_i \nabla\hat{\psi}_i(\mathbf{x}^{(p)}) \right) \tag{65}$$

s.t.

$$\lambda_i \geq 0, \quad \forall i \in I_a \tag{66}$$

The squared fitting error ε indicates how well $\nabla\hat{\varphi}(\mathbf{x}^{(p)})$ is expressed as a linear combination of $\nabla\hat{\psi}_i(\mathbf{x}^{(p)})$ with non-negative coefficients λ_i. To quantify the goodness of fit, the coefficient of determination (Rao et al., 2008) is calculated as

$$R^2 = 1 - \frac{\varepsilon}{\nabla\hat{\varphi}(\mathbf{x}^{(p)})^T \nabla\hat{\varphi}(\mathbf{x}^{(p)})} \tag{67}$$

If R^2 is close to one, the linear model (62) has a good fit and the equality (61) is satisfied in a statistical sense. In this case, the KKT conditions are validated and a local optimal solution is obtained.

We should note that the inventory optimization problem contains black-box functions in the objective and in the constraints. These black-box functions cannot be evaluated without errors by the Monte-Carlo method. Because of these challenges, it is difficult to guarantee that the iterative algorithm converges to the optimal solution for all inventory systems with various network structures and distributions of uncertain parameters. However, once we obtain a solution, we can validate its optimality by applying the statistical tests on the KKT conditions.

5. Case study

To demonstrate the simulation-based optimization framework proposed in Section 4, we apply it to two case studies. The agent-based model is coded in Java using Eclipse. The linearized problem (INV_LIN$_p$) and the constrained least squares problem (CON_LS) are solved by MATLAB. All algorithms are implemented in a desktop with a 3.10 GHz CPU running Windows 7 Professional (64 bit).

5.1. Two-echelon system

The first case study is a two-echelon inventory system shown in Fig. 8. The system consists of five facilities: a plant, a distribution center and three distributors. The plant is assumed to have the unlimited inventory and its inventory management is not considered. The remaining four facilities are indexed by a number shown in the figure. They manage their inventories according to the (r, Q) policy. Each facility, including the distribution center can receive customer orders. The daily demand of a customer order is assumed to be a Gaussian random variable with the mean and the standard deviation given in Fig. 8. The order processing times are assumed to be uniform random variables with the lower bound and the upper bound given in Fig. 8.

The objective of the inventory optimization problem is to minimize the inventory cost while maintaining the acceptable service

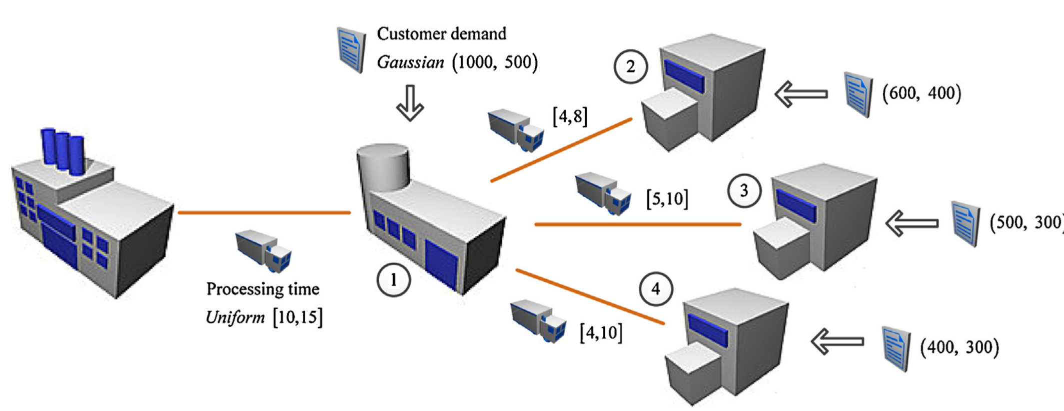

Fig. 8. Two-echelon inventory system. The daily customer demands follow the Gaussian distribution with (mean, standard deviation) labeled beside the order icons. The order processing times follow the uniform distribution with [lower bound, upper bound] labeled beside the shipment icons.

levels. The unit holding cost of the facilities is 1 m.u./unit/day where m.u. refers to money unit. The unit reorder cost of the distribution center is 500 m.u./order while those for the distributors are 100 m.u./order. The minimum service levels for all facilities are set as 95%. The time horizon is 365 days.

The inventory optimization problem is solved by the simulation-based optimization method proposed in Section 4. The iterative cutting plane algorithm is applied according to the procedure shown in Fig. 7. The number of sampling points in the Monte-Carlo method is 1000. The number is sufficiently large to ensure an accurate estimate. In the Monte-Carlo method, the mean value is calculated as the sample average approximation (Kleywegt et al., 2002). The sampling points are generated by the Latin Hypercube method (Helton and Davis, 2003). The number of sampling points is an important parameter for a Monte-Carlo method. The number should be large enough to ensure an accurate approximation of the mean value by the sample average. Meanwhile it is desirable to be as small as possible to reduce the computational efforts. However, finding the best number is not the focus of this work. One can refer to the literature for various approaches used to determine the number of sampling points (Liu, 2008).

Other sampling methods can also be adopted. It is shown in Table 2 that the lengths of the 99% confidence intervals using the sample size are all less than 1%. The iterative process is displayed in Fig. 9. The gap between the objective function values in two adjacent iterations is set to 1%. The search algorithm terminates when the difference in the objective function values is less than the gap. With each iteration, the inventory cost reduces while the fill rates are pushed to the bounds. The optimization algorithm stops after 15 iterations and the computational time (the wall-clock time) is 270.6 s.

The initial and optimal solutions are listed in Table 1. The indices of the reorder points and the order quantities correspond to those of the facilities in Fig. 8. The optimal function values are listed in Table 2 along with the range of the 99% confidence intervals. The range of a confidence interval is small (less than 0.5%), reflecting a small variation in the estimate returned by the Monte-Carlo method.

To validate the solution feasibility and optimality, Table 2 also lists the minimum service levels and the safety service levels. Every minimum service level is less than the obtained service level minus the range of the 99% confidence interval. Therefore, the service level constraints are satisfied with 99% probability. To find the set of active constraints defined in Eq. (59), the hypothesis test in Eqs. (35)–(37) is conducted. The test shows that all the service level constraints are active. Consequently, all Lagrange multipliers in equality (61) can be non-zero. Their values, calculated by solving the constrained least squares problem (CON_LS), are $\lambda_1 = 8.10$, $\lambda_2 = 2.09$, $\lambda_3 = 1.81$, and $\lambda_4 = 1.51$. The coefficient of determination calculated from Eq. (67) is $R^2 = 0.968$. It is close to one, implying the linear regression model (62) satisfies the goodness of fit condition. Therefore, the stationary equality (61) holds in a statistical sense. The optimal solution thus passes the test on KKT conditions and, therefore, its local optimality is validated.

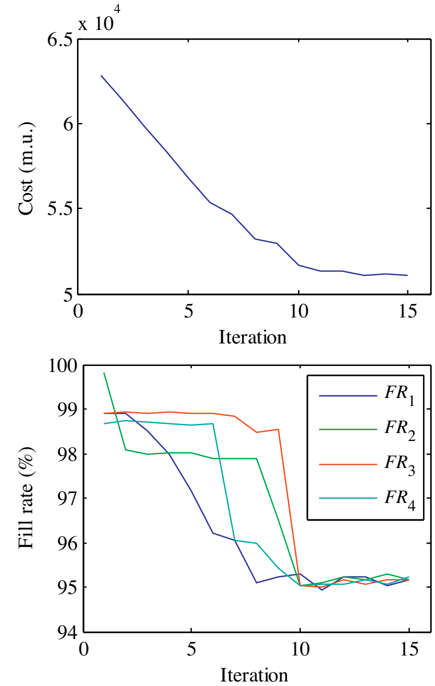

Fig. 9. Iterative results of the two-echelon inventory system.

Fig. 10 shows the inventory trajectories for the initial inventory parameters and the optimal ones listed in Table 1. Inventory optimization reduces the reorder points to reduce the on-hand inventory levels. The reduction raises the backorders and decreases the service levels. However, because the minimum service levels are taken into account in the constraints, the decreased service levels are still maintained above the acceptable threshold values by the algorithm.

5.2. Three-echelon system

The second case study is a three-echelon inventory system shown in Fig. 11, which consists of eight facilities. The plant is assumed to have unlimited inventory and its inventory

Table 1
Initial and optimal solutions for the two-echelon inventory system.

Facility	Reorder point				Order quantity			
	1	2	3	4	1	2	3	4
Initial solution	40,000	6000	5000	8000	40,000	8000	7000	6000
Optimal solution	32,606.1	3556.6	3956.7	9289.2	39,548.3	9289.2	7025.2	5654.1

Table 2
Optimal function values for the two-echelon inventory system.

	Inventory cost	Fill rate			
		Facility 1	Facility 2	Facility 3	Facility 4
Value	51,043.6	95.16	95.15	95.16	95.21
99% interval	±108.9	±0.14	±0.13	±0.13	±0.13
Minimum service level	–	95.00	95.00	95.00	95.00
Safety service level	–	95.14	95.13	95.13	95.13

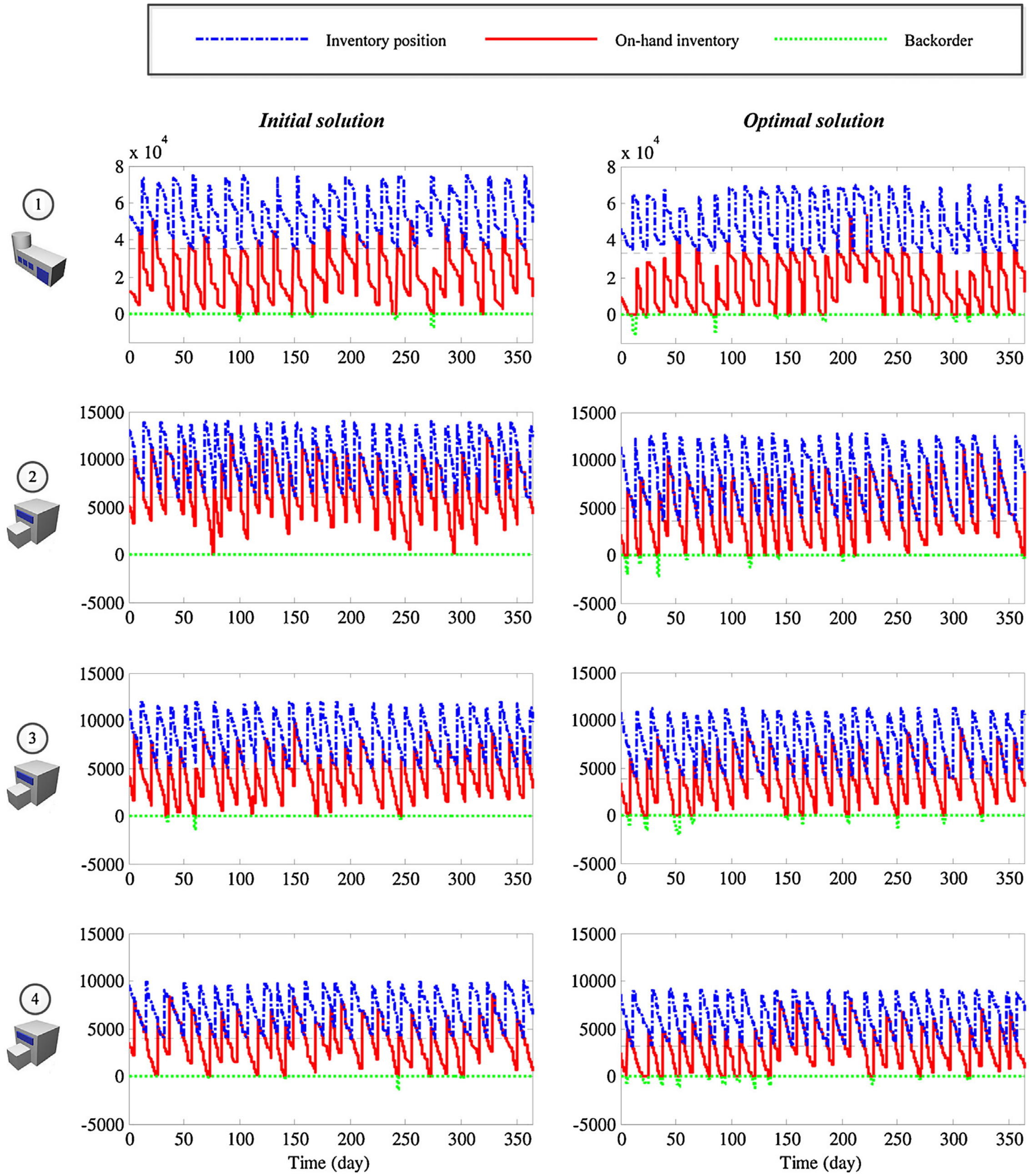

Fig. 10. Inventory trajectories of the two-echelon system. The left panel shows the results for the initial inventory parameters and the right panel shows the results for the optimal inventory parameters. The reorder points are shown as the horizontal black dash lines. For clarity, the backorders are shown as negative values.

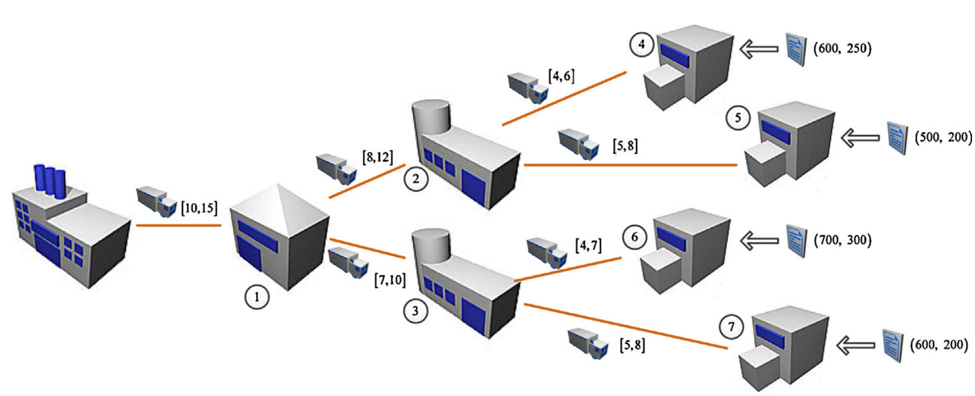

Fig. 11. Three-echelon inventory system. The daily customer demands follow the Gaussian distribution with (mean, standard deviation) labeled beside the order icons. The order processing times follow the uniform distribution with [lower bound, upper bound] labeled beside the shipment icons.

management is not taken into account. The inventory parameters are optimized for the warehouse (No. 1), two distribution centers (No. 2, 3), and four distributors (No. 4–7). The daily customer order demand follows a Gaussian distribution and the order processing times follow a uniform distribution. The parameters for the distributions are shown in Fig. 11.

The objective of the inventory optimization problem is to minimize the inventory cost while maintaining acceptable service levels. The minimum time unit is set as one day, which is small enough to accurately describe the event occurrence in practice. The time unit can be easily reduced to increase the time resolution. A smaller time unit incurs more computational efforts for the simulation program while it has few effects on the optimization algorithm. The unit holding cost of the facilities is 1 m.u./(unit day). The unit reorder cost of the warehouse, the distributions centers, and the distributors is 300 m.u./order, 200 m.u./order, and 100 m.u./order, respectively. The minimum service levels for all facilities are set as 95%. The time horizon is 365 days.

We should note that the inventory network is a dynamic system driven by customer order demands. The optimization problem is different from a supply chain design problem. We do not simply determine the static flow rates among the facilities. Instead, we need to investigate the actions of each facility every day. In the context of multi-echelon optimization, such a three-echelon problem is not trivial to solve.

The inventory optimization problem is solved by the simulation-based optimization method proposed in Section 4. The termination gap between the objective function values in two adjacent iterations is set as 1%. The number of sampling points in the Monte-Carlo method is initially set as 100, in case the initial solution is far from the optimal one. While following the iterative procedure in Fig. 7, the algorithm restarts at the 23rd iteration and the number of sampling points in the Monte-Carlo method gets increased to 1000 as the solution comes close to the optimal one. The sampling size is sufficiently large to ensure small 99% confidence intervals which is shown in Table 4. It took 36 iterations to get to the optimum, with a computational time (the wall-clock time) of 751.3 s.

The iterative procedure is shown in Fig. 12. Similar to the two-echelon system, the optimization algorithm decreases the objective function while pushing the fill rates to bounds. However, larger oscillations occur in the fill rate trajectories than those in Fig. 9. The reason is that the three-echelon system incorporates more stages than the two-echelon system, resulting in stronger interactions among the facilities. When the fill rate of one facility increases, for the other facility it may decrease owing to the strong interactions among the facilities. Therefore, more iterations are required to push the fill rates to the bounds simultaneously, resulting in oscillations.The initial solution and the optimal solution are listed in Table 3. The optimal function values along with the confidence interval range are listed in Table 4. Every minimum service level is less than the obtained service level minus the range of the 99% confidence interval. Therefore, the service level constraints are satisfied with 99% probability. To find the set of active constraints defined in Eq. (59), the hypothesis test in Eqs. (35)–(37) is conducted. The test shows that all service level constraints are active. Consequently, all Lagrange multipliers in equality (61) can be non-zero. Their values, calculated by solving the constrained least squares problem (CON_LS), are $\lambda_1 = 11.53$, $\lambda_2 = 2.69$, $\lambda_3 = 5.13$, $\lambda_4 = 4.02$, $\lambda_5 = 2.10$, $\lambda_6 = 2.15$, and $\lambda_7 = 1.46$. The coefficient of determination calculated from Eq. (67) is $R^2 = 0.953$. It is close to one, implying the linear regression model (62) satisfies the goodness of fit condition. Therefore, stationary equality (61) holds in a statistical sense. The optimal solution thus passes the test on the KKT conditions and, therefore, its local optimality is validated.

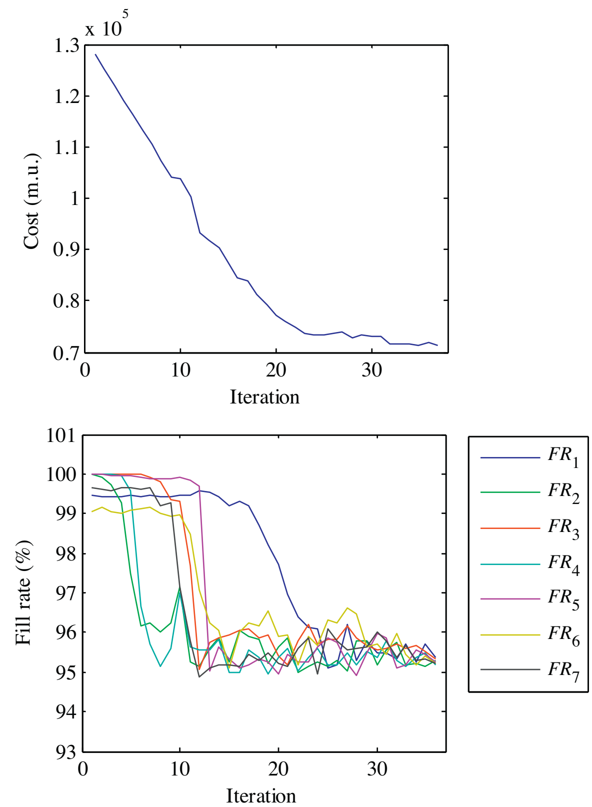

Fig. 12. Iterative results of the three-echelon inventory system.

Table 3
Initial and optimal solutions for the three-echelon inventory system.

	Facility	1	2	3	4	5	6	7
Initial solution	Reorder point	40,000	20,000	20,000	5000	5000	5000	5000
	Order quantity	50,000	20,000	20,000	5000	5000	5000	5000
Optimal solution	Reorder point	22,218.4	8780.6	8607.3	4079.8	3923.9	4394.2	4126.0
	Order quantity	44,132.0	18,146.9	19,169.7	7297.6	5914.5	6429.2	5653.3

Table 4
Optimal function values for the three-echelon inventory system.

	Inventory cost	Fill rate						
		FR_1	FR_2	FR_3	FR_4	FR_5	FR_6	FR_7
Value	71,331.5	95.36	95.29	95.32	95.20	95.25	95.23	95.21
99% interval	±131.4	±0.25	±0.17	±0.18	±0.13	±0.14	±0.14	±0.13
Minimum service level	–	95.00	95.00	95.00	95.00	95.00	95.00	95.00
Safety service level	–	95.25	95.17	95.18	95.13	95.14	95.14	95.13

6. Conclusion

We proposed a simulation-based optimization framework for solving multi-echelon inventory problems. The inventory systems have a divergent structure where each facility is installed with the (r, Q) control policy. The inventory parameters were optimized to minimize the inventory cost while maintaining acceptable service levels quantified by the fill rates. The objective functions and the constraint functions were evaluated by the agent-based modeling and simulation. Their expectations were then estimated by the Monte Carlo method using sampling points of the uncertain parameters according to their probability distributions. The estimated expectation functions were used in the optimization problem as black-box functions. The problem was solved by a cutting plane algorithm that iteratively linearized the black-box functions using the first-order response surface models estimated via the fractional factorial experiment design. Because the black-box functions were estimates by the Monte Carlo method using a finite number of sampling points, they inevitably contained noise. Thus, hypothesis tests were conducted to validate the improvements in the objective function value and the feasibility of the service level constraints. If the final solution passed the statistical tests on the optimal conditions, a local optimal solution was obtained.

The simulation-based optimization framework was demonstrated for two case studies. The first case study was a two-echelon system with four facilities while the second was a three-echelon system with seven facilities. In the iterative optimization procedure, the cost objective function decreased while the constraints functions were pushed to the bounds specified by the minimum service levels. In both case studies, the local optimal solutions were found in 270.6 and 751.3 s, respectively. Their optimality were validated by the statistical tests on the KKT conditions.

Acknowledgement

The authors gratefully acknowledge financial support from The Dow Chemical Company.

References

Almeder C, Preusser M, Hartl RF. Simulation and optimization of supply chains: alternative or complementary approaches? OR Spectr 2009;31:95–119.
Asmussen S, Glynn PW. Stochastic simulation: algorithms and analysis, vol. 57. Springer; 2007.
Atlason J, Epelman MA, Henderson SG. Call center staffing with simulation and cutting plane methods. Ann Oper Res 2004;127:333–58.
Blum C, Puchinger J, Raidl GR, Roli A. Hybrid metaheuristics in combinatorial optimization: a survey. Appl Soft Comput 2011;11:4135–51.
Box GE, Hunter JS, Hunter WG. Statistics for experimenters. New York: Wiley; 2005.
Castro PM, Aguirre AM, Zeballos LJ, Mendez CA. Hybrid mathematical programming discrete-event simulation approach for large-scale scheduling problems. Ind Eng Chem Res 2011;50:10665–80.
Chan FTS. Performance measurement in a supply chain. Int J Adv Manuf Technol 2003;21:534–48.
Chen JH, Lin DKJ, Thomas DJ. On the single item fill rate for a finite horizon. Oper Res Lett 2003;31:119–23.
Chopra S, Meindl P. Supply chain management. Strategy, planning & operation. Springer; 2007.
Chu Y, You F, Wassick JM. Hybrid method integrating agent-based modeling and heuristic tree search for scheduling of complex batch processes. Comput Chem Eng 2014;60:277–96.
Chu Y, You F, Wassick JM, Agarwal A. Integrated planning and scheduling under production uncertainties: bi-level model formulation and hybrid solution method. Comput Chem Eng 2015;72:255–72.
Chu YF, Wassick JM, You FQ. Efficient scheduling method of complex batch processes with general network structure via agent-based modeling. AIChE J 2013;59:2884–906.
Dejonckheere J, Disney SM, Lambrecht MR, Towill DR. Measuring and avoiding the bullwhip effect: a control theoretic approach. Eur J Oper Res 2003;147:567–90.
Diks EB, deKok AG, Lagodimos AG. Multi-echelon systems: a service measure perspective. Eur J Oper Res 1996;95:241–63.
Fu MC. Optimization for simulation: theory vs. practice. Inf J Comput 2002;14:192–215.
Grossmann I. Enterprise-wide optimization: a new frontier in process systems engineering. AIChE J 2005;51:1846–57.
Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81:23–69.
Ivanov D, Dolgui A, Sokolov B. Applicability of optimal control theory to adaptive supply chain planning and scheduling. Ann Rev Cont 2012;36:73–84.
Julka N, Srinivasan R, Karimi I. Agent-based supply chain management-1: framework. Comput Chem Eng 2002;26:1755–69.
Jung JY, Blau G, Pekny JF, Reklaitis G, Eversdyk D. Integrated safety stock management for multi-stage supply chains under production capacity constraints. Comput Chem Eng 2008;32:2570–81.
Jung JY, Blau G, Pekny JF, Reklaitis GV, Eversdyk D. A simulation based optimization approach to supply chain management under demand uncertainty. Comput Chem Eng 2004;28:2087–106.
Kleywegt AJ, Shapiro A, Homem-de-Mello T. The sample average approximation method for stochastic discrete optimization. SIAM J Optim 2002;12:479–502.
Kochel P, Nielander U. Simulation-based optimisation of multi-echelon inventory systems. Int J Prod Econ 2005;93–94:505–13.
Lee JH, Kimz CO. Multi-agent systems applications in manufacturing systems and supply chain management: a review paper. Int J Prod Res 2008;46:233–65.
Liu JS. Monte Carlo strategies in scientific computing. Springer; 2008.
Mansouri SA. A simulated annealing approach to a bi-criteria sequencing problem in a two-stage supply chain. Comput Ind Eng 2006;50:105–19.
Mele FD, Guillen G, Espuna A, Puigjaner L. A simulation-based optimization framework for parameter optimization of supply-chain networks. Ind Eng Chem Res 2006;45:3133–48.
Min H, Zhou GG. Supply chain modeling: past, present and future. Comput Ind Eng 2002;43:231–49.
Moncayo-Martinez LA, Zhang DZ. Multi-objective ant colony optimisation: a meta-heuristic approach to supply chain design. Int J Prod Econ 2011;131:407–20.
Neiro SMS, Pinto JM. A general modeling framework for the operational planning of petroleum supply chains. Comput Chem Eng 2004;28:871–96.
Nikolopoulou A, Ierapetritou MG. Hybrid simulation based optimization approach for supply chain management. Comput Chem Eng 2012;47:183–93.
Papageorgiou LG. Supply chain optimisation for the process industries: advances and opportunities. Comput Chem Eng 2009;33:1931–8.
Parunak HVD, Savit R, Riolo RL. Agent-based modeling vs. equation-based modeling: a case study and users' guide. In: Multi-agent systems and agent-based simulation. Springer; 1998. p. 10–25.
Peidro D, Mula J, Poler R, Lario FC. Quantitative models for supply chain planning under uncertainty: a review. Int J Adv Manuf Technol 2009;43:400–20.
Perea-Lopez E, Ydstie BE, Grossmann IE. A model predictive control strategy for supply chain optimization. Comput Chem Eng 2003;27:1201–18.
Puigjaner L, Lainez JM. Capturing dynamics in integrated supply chain management. Comput Chem Eng 2008;32:2582–605.
Rahmandad H, Sterman J. Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models. Manage Sci 2008;54:998–1014.
Rao C, Toutenburg H, Shalabh HC, Schomaker M. Linear models and generalizations. Least squares and alternatives. 3rd ed. Berlin, Heidelberg/New York: Springer; 2008.
Sarimveis H, Patrinos P, Tarantilis CD, Kiranoudis CT. Dynamic modeling and control of supply chain systems: a review. Comput Oper Res 2008;35:3530–61.
Schwartz JD, Wang WL, Rivera DE. Simulation-based optimization of process control policies for inventory management in supply chains. Automatica 2006;42:1311–20.
Shah N. Process industry supply chains: advances and challenges. Comput Chem Eng 2005;29:1225–35.
Shapiro A, Dentcheva D, Ruszczynski A. Lectures on stochastic programming. Lectures on stochastic programming: modeling and theory, vol. 9. Philadelphia: Siam; 2009. p. 1–436.
Silva CA, Sousa JMC, Runkler TA, Da Costa J. Distributed optimisation of a logistic system and its suppliers using ant colonies. Int J Syst Sci 2006;37:503–12.
Subramanian K, Rawlings JB, Maravelias CT, Flores-Cerrillo J, Megan L. Integration of control theory and scheduling methods for supply chain management. Comput Chem Eng 2013;51:4–20.
Swaminathan JM, Smith SF, Sadeh NM. Modeling supply chain dynamics: a multiagent approach. Decis Sci 1998;29:607–32.
Tekin E, Sabuncuoglu I. Simulation optimization: a comprehensive review on theory and applications. Iie Trans 2004;36:1067–81.
Toutenburg H. Statistical analysis of designed experiments. 3rd ed. New York, USA: Springer; 2009.
Wassick JM, Agarwal A, Akiya N, Ferrio J, Bury S, You FQ. Addressing the operational challenges in the development, manufacture, and supply of advanced materials and performance products. Comput Chem Eng 2012;47:157–69.
You FQ, Grossmann IE. Mixed-integer nonlinear programming models and algorithms for large-scale supply chain design with stochastic inventory management. Ind Eng Chem Res 2008;47:7802–17.
You FQ, Grossmann IE. Integrated multi-echelon supply chain design with inventories under uncertainty: MINLP models, computational strategies. AIChE J 2010;56:419–40.
You FQ, Grossmann IE. Balancing responsiveness and economics in process supply chain design with multi-echelon stochastic inventory. AIChE J 2011a;57:178–92.
You FQ, Grossmann IE. Stochastic inventory management for tactical process planning under uncertainties: MINLP models and algorithms. AIChE J 2011b;57:1250–77.

You FQ, Pinto JM, Grossmann IE, Megan L. Optimal distribution-inventory planning of industrial gases. II. MINLP models and algorithms for stochastic cases. Ind Eng Chem Res 2011;50:2928–45.

Yue DJ, You FQ. Planning and scheduling of flexible process networks under uncertainty with stochastic inventory: MINLP models and algorithm. AIChE J 2013;59:1511–32.

Zapata JC, Pekny J, Reklaitis GV. Simulation-optimization in support of tactical and strategic enterprise decisions. In: Kempf KG, Keskinocak P, Uzsoy R, editors. Planning production and inventories in the extended enterprise: a state of the art handbook, vol. 151. New York: Springer; 2011. p. 593–627.

Zipkin PH. Foundations of inventory management, vol. 2. New York: McGraw-Hill; 2000.