IISE Transactions

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

A Flexible System Design Approach for Multi-Facility Capacity Expansion Problems with Risk Aversion

Sixiang Zhao, William B. Haskell & Michel-Alexandre Cardin

A Flexible System Design Approach for Multi-Facility Capacity Expansion Problems with Risk Aversion

Sixiang Zhao*[a], William B. Haskell[b], and Michel-Alexandre Cardin[c]

[a]Sino-US Global Logistics Institute, Shanghai Jiao Tong University, Shanghai, China

[b]Krannert School of Management, Purdue University, West Lafayette, USA

[c]Dyson School of Design Engineering, Imperial College London, UK

*Contact Sixiang Zhao Email: sixiang.zhao@sjtu.edu.cn

Abstract

This paper studies a model for risk aversion when designing a flexible capacity expansion plan for a multi-facility system. In this setting, the decision maker can dynamically expand the capacity of each facility given observations of uncertain demand. We model this situation as a multi-stage stochastic programming problem, and we express risk aversion through the conditional value-at-risk (CVaR) and a mean-CVaR objective. We optimize the multi-stage problem over a tractable family of if–then decision rules using a decomposition algorithm. This algorithm decomposes the stochastic program over scenarios and updates the solutions via the subgradients of the function of cumulative future costs. To illustrate the practical effectiveness of this method, we present a numerical study of a decentralized waste-to-energy system in Singapore. The simulation results show that the risk-averse model can improve the tail risk of investment losses by adjusting the weight factors of the mean-CVaR objective. The simulations also demonstrate that the proposed algorithm can converge to high-performance policies within a reasonable time, and that it is also more scalable than existing flexible design approaches.

Keywords: Capacity expansion problem; System design; Real options; Risk aversion; Multi-stage stochastic programming; Decision rules.

1 Introduction

The capacity expansion problem aims to determine a capacity expansion plan (i.e., the optimal amount and timing of capacity acquisition) to address growing demand. This problem has been studied for a broad variety of systems, such as semiconductor manufacturing (Geng et al., 2009), airport facilities (Sun and Schonfeld, 2015), container terminals (Li et al., 2017), and waste-to-energy systems (Cardin and Hu, 2016). However, this problem is challenging because future demand is uncertain, and also because capital expenditure is very expensive and usually irreversible.

The traditional *inflexible* design determines the capacity expansion plan at the beginning of the planning horizon, before any demand is observed. This plan is then implemented regardless of the realizations of future demand. The inflexible design problem can be formulated as a two-stage stochastic programming problem, which has been widely studied in the literature (Swaminathan, 2002; Geng et al., 2009). However, this method may suffer unexpected costs if the actual demand realizations do not match the forecasts.

Real options analysis copes with this issue by designing *flexible* systems in capacity expansion problems. In real options analysis, the decision to expand capacity is viewed as a series of options that can be exercised over time (Dixit and Pindyck, 1994; de Neufville and Scholtes, 2011). The main advantage of this framework comes from its “wait-and-see” nature: these decisions can be exercised or deferred based on the realizations of uncertain demand. A flexible system has the ability to dynamically adjust its capacity as demand is observed. We can then expand the capacity (i.e., exercise the option) if demand surges, and we may do nothing when demand remains steady. In the multi-facility capacity expansion problem (MCEP), a flexible system has the option not only to adjust the capacity but also to switch service between facilities. For example, if one facility runs out of capacity, then we can either expand the capacity of this facility or allocate the excess demand to another facility. It has been verified in many industrial case studies that flexibility can improve expected performance by 10% to 30% compared with inflexible methods (de Neufville and Scholtes, 2011; Cardin et al., 2017).

In real options theory, it is important to evaluate the economic performance of the flexible system and to quantify the value of flexibility (VoF), which is the difference between the performance of a flexible system and its inflexible counterpart. This is because conferring flexibility on a system usually requires an upfront cost as a premium, compared with an inflexible system with its rigid design (Cardin and Hu, 2016). Another example in capacity expansion problems is that, if we want to expand capacity when observing an increase of demand, then we may buy capacity (e.g., machines) from the spot market. However, this cost may be greater than the cost of contracting additional capacity ahead of time.

Optimizing a flexible MCEP is a dynamic optimization problem with demand uncertainty. We formulate this problem as a multi-stage stochastic programming (MSSP) problem, where the evolution of the uncertain parameters is modeled by a scenario tree. However, this problem is notoriously difficult to solve even for practically-sized problems: the size of the scenario tree grows exponentially with not only the number of stages, but also with the number of uncertain parameters. Even when the capacity decisions are discrete, MSSP may still be inefficient in solving the MCEP.

An MSSP based on decision rules was proposed to address the problem of tractability. In a flexible capacity expansion problem, we need to find the optimal expansion policy, which is a mapping from historical data to capacity decisions. In this method, we optimize over a tractable class of parametrized policies (the decision rules) rather than the space of all policies. Here the focus is on optimizing the parameters of the decision rule rather than optimizing the policy itself: this is more tractable than traditional methods and can scale-up to larger problem instances. Well-known decision rules include linear and piecewise linear rules (Georghiou et al., 2015), although these rules may not be applicable to MCEP because the system is usually modular (i.e., the capacity is discrete).

To solve MCEPs with discrete capacity, if–then decision rules have been proposed to approximate the policy space (Cardin et al., 2017; Zhao et al., 2018; Zhang and Cardin, 2017). An if–then rule states that if the capacity gap of a facility exceeds a threshold, then its capacity is expanded to a certain level, and the capacity is unchanged otherwise. In this framework, we want to find the best if–then decision rule among all such policies. The if–then decision rule mimics the behavior of human decision makers, so they are intuitive and can be interpreted easily by non-experts (Cardin et al., 2017). In contrast, the optimal policies from dynamic programming (DP) and scenario-tree based MSSP may be difficult to interpret and explain in general. In numerical terms, a method based on decision rules can provide high-performance solutions for MCEPs, and its scalability has been verified (Zhao et al., 2018).

Classical MCEP models under the framework of real options theory suppose that the decision makers are risk-neutral, that is, they maximize the expected reward. However, decision makers often have their own attitudes to risk. For example, due to the high expenditure and uncertainty, decision makers may underestimate the value of a flexible system if they are risk averse. To measure risk in our model, we employ the popular conditional value-at-risk (CVaR). The motivation for this choice is three-fold. First, CVaR is a monetary value that represents the expected tail loss, so it is intuitive for decision makers to compare it with the risk-neutral design alternatives in terms of rewards/costs. Second, CVaR is a coherent risk measure that has strong decision-theoretic support (Artzner et al., 1999). Third, CVaR enjoys significant computational advantages compared to other risk-aware objectives (Rockafellar and Uryasev, 2000).

Our present paper investigates a risk-averse MCEP with a mean-CVaR objective. Our specific contributions are as follows:

1. We propose a generic MCEP by considering the risk preferences of the decision maker. The proposed risk-averse model contains the traditional risk-neutral model as a special case. We then solve this risk-averse problem, which is essentially an MSSP, by using an if–then decision rule

to approximate the expansion policy. We are not aware of any studies that have yet used decision rules for risk-averse MCEPs.

2. We propose a customized decomposition algorithm to optimize over decision rules with respect to the risk-averse objective. This algorithm uses subgradient cuts of the cost functions to update the parameters of the decision rule, which is not only more efficient but can also provide insight into the problem when compared with the existing branch-and-cut based decomposition (BACD) algorithm in the literature (Zhao et al., 2018).
3. We verify both theoretically and numerically that the VoF decreases if a risk-averse expansion policy is implemented. We also derive managerial insights by implementing comprehensive numerical studies with data from a case study on a waste-to-energy (WTE) system in Singapore. We find that, in a multi-facility system, a risk-averse expansion policy may not always establish smaller capacity at the beginning when compared with the risk-neutral policy.

The rest of this paper is organized as follows. Section 2 summarizes the relevant literature. Section 3 first introduces the flexible capacity expansion problem and discusses the model assumptions. Then, the risk-averse MCEP model is presented. In Section 4, we approximate the capacity expansion policy of the risk-averse MCEP via if–then decision rules, and we transform the problem to a mixed-integer linear programming problem (MILP). Section 5 then presents a decomposition algorithm to optimize the decision rule. A detailed numerical study of a WTE system is given in Section 6. Finally, the strengths and limitations of the proposed method and opportunities for future research are summarized in Section 7. All proofs are gathered together in the online supplement.

2 Literature Review

2.1 Flexible Capacity Expansion Problems

Capacity expansion problems have been widely studied since the seminal paper by (Manne, 1961). Many variations of this original model have been studied. Comprehensive reviews may be found in (Van Mieghem, 2003; Martínez-Costa et al., 2014). Eberly and Van Mieghem (1997) studied a multi-factor capacity investment problem and characterized the structure of the optimal policy. Kouvelis and Tian (2014) studied a flexible capacity investment problem and investigated the value, when facing uncertain demand, of the option to postpone a commitment to increase capacity. Our work differs from these models because we deal with discrete capacity expansion decisions. Huang and Ahmed (2009) derived, for discrete capacity expansion problems, an analytical bound for the value of the multi-stage problem compared to the two-stage problem, but this result requires linear expansion costs. Cardin and Hu (2016) and Zhao et al. (2018) studied MCEPs with nonlinear expansion costs. The objective functions in these papers, however, are all risk-neutral.

2.2 Risk Measures

To capture the risk preferences of real decision makers, a variety of risk measures have been used in the literature, including utility functions, mean-variance, value-at-risk (VaR), and CVaR. For example, Hugonnier and Morellec (2007) extended the standard real options analysis by introducing a utility function that addresses risk preferences. Birge (2000) incorporated utility functions into a general linear capacity planning model, and formulated the problem as an MSSP problem. Compared to utility functions, CVaR is more intuitive and easier to specify. Decision makers can express their risk preferences by directly adjusting the percentile terms of gains or losses (Krokhmal et al., 2002) rather than choosing a utility function. In addition, CVaR is a coherent risk measure and thus has strong decision-theoretic support (Artzner et al., 1999). Furthermore, CVaR can be formulated as a convex optimization problem and it is thus more tractable that other risk-aware objective functions (Rockafellar and Uryasev, 2000). Maceira et al. (2015) applied CVaR to a multi-stage power generation planning problem and solved it by combining a scenario-tree based method with stochastic dual dynamic programming. Applications of CVaR can also be found in resilient facility location problems (Yu et al., 2017), energy capacity investments (Szolgayová et al., 2011), and transmission network expansion (Delgado and Claro, 2013).

2.3 Solution Methods for Capacity Expansion Problems

The early work on MCEP modeled the problem as a Markov decision process (MDP) and solved it with exact DP (Wu and Chuang, 2010) or approximate dynamic programming (ADP) (Zhao et al., 2017). However, these methods are subject to the curse of dimensionality. More specifically, the size of the action space of the MDPs grows exponentially in the number of facilities. Alternative methods of solving a risk-averse MDP can be found in (Ruszczyński, 2010; Haskell and Jain, 2015), but these methods can be inefficient when the actions are discrete and high-dimensional.

Alternatively, scenario tree-based MSSP has been widely applied to MCEP (Huang and Ahmed, 2009; Taghavi and Huang, 2016). In this method, the evolution of uncertain parameters is modeled as a scenario tree, and the model is solved with decomposing by fixing the allocation plan (Huang and Ahmed, 2009), by a Benders decomposition-based heuristic (Taghavi and Huang, 2016), or by Lagrangian relaxation (Taghavi and Huang, 2018). Nevertheless, the size of the scenario tree grows exponentially when the number of stages or the number of uncertain parameters increases.

To solve MCEPs with discrete capacity, many have investigated if–then decision rules and proposed customized decomposition algorithms to optimize the parameters of the decision rules, such as the Lagrangian decomposition method (Cardin et al., 2017) and the BACD algorithm (Zhao et al., 2018). The solution technique in the present paper is similar to the BACD algorithm proposed in (Zhao et al., 2018). However, our study differs from the previous literature in that we construct cuts for the master problem with the subgradients of the cumulative future cost function, which is much less time consuming and can provide a better interpretability for the resulting policy.

3 Model Formulation

3.1 Problem Descriptions and Assumptions

We consider a multi-facility capacity expansion problem, where multiple customers are served over a finite planning horizon. We first set the initial capacity before any realizations of the uncertain demand are observed. Then, in each period, customer demand is observed and allocated to the facilities subject to available capacity. In our model, profit is earned by satisfying customer demand and costs are incurred by capacity expansion. If the capacity is insufficient then an additional penalty is incurred. In practice, the penalty can be interpreted as lost sales, or the overtime costs for workers and machines that are incurred to meet the demand. We make the following assumptions:

A1. The expansion lead time is negligible, so the capacity expanded at the end of the previous time period will be available at the beginning of the next period.
A2. Contraction of capacity is not allowed.
A3. The demand distribution is known and can be simulated.
A4. The expansion cost function is piecewise linear (but possibly nonconvex).

Assumption A1 is common in strategic capacity planning problems (Huang and Ahmed, 2009; Sun and Schonfeld, 2015; Cardin et al., 2017). Assumption A2 holds in many industries where the capacity investment is irreversible. For example, the capacity of airport facilities, such as highway links and ports, is difficult to reduce once established (Sun and Schonfeld, 2015). Furthermore, many unpredictable shocks are industry-specific: a steel manufacturer intends to sell a steel plant when the market is depressed but it is likely that the plant has little value under these circumstances (Dixit and Pindyck, 1998).

Assumption A4 is general. If expansion costs benefit from economies of scale, the cost function will be concave with respect to the expanded capacity. A piecewise linear function can represent/approximate a variety of concave cost functions; for example, the fixed-charge function and the power function (Van Mieghem, 2003). Essentially, because capacity expansion decisions are discrete, we can use piecewise linear functions to represent any appropriate cost functions.

In the inflexible capacity expansion problem, the capacity plan does not respond to changes in demand (see Figure 1a). In a flexible system with options for adjusting the capacity, the decision maker first observes customer demand and then decides, at the end of each period, whether to expand the capacity (see Figure 1b & 1c). In addition, when there are multiple facilities, we can switch services between facilities if one runs out of capacity. Therefore, we need to determine both when and by how much to expand the capacity, and also which facility to expand. These decisions are characterized by a *capacity expansion policy*, which is a function that maps historical demand to capacity decisions (i.e., when, how much, and which facility to expand).

The objective is to maximize the economic performance of the system by optimizing the capacity expansion policy. If the decision maker is risk-neutral, the economic performance can be characterized by the expected net present value (ENPV) of the cumulative profits. However, if the decision maker is risk-averse, a risk metric may be incorporated into the objective function to capture the risk preferences.

3.2 Formulation of Risk-Averse MCEP

Now we introduce the flexible MCEP with risk aversion. The notation for this model is summarized in Table 1.

For our flexible capacity expansion problem, we introduce a set of facilities $\mathcal{N} \triangleq \{1, \ldots, N\}$, a set of customers $\mathcal{I} \triangleq \{1, \ldots, I\}$, and a finite discrete planning horizon $\mathcal{T} \triangleq \{1,\dots,T\}$. We define $\mathcal{T}_0 \triangleq \mathcal{T} \cup \{0\}$ when the planning horizon includes 0. Let $\mathbb{K} \triangleq \{K \in \mathbb{Z}_+^N \mid K \le K^{\max}\}$ be the finite set of possible capacity levels, where $K^{\max} \triangleq (K_1^{\max},\dots,K_N^{\max})$ is the vector of maximum possible capacity levels. Denote by $K_t \triangleq (K_{1t},\dots,K_{Nt}) \in \mathbb{K}$ the vector of the capacity installed at the end of time period $t \in \mathcal{T}_0$, where K_0 denotes the vector of initial capacities. Write $K_{[t]} \triangleq (K_0, K_1,\dots,K_t)$ for the history of installed capacity levels up to time t. Let $\Delta K_{nt} \triangleq K_{nt} - K_{n(t-1)}$ and $\Delta K_t \triangleq (\Delta K_{1t},\dots,\Delta K_{Nt})$ be the changes in capacity at time $t \in \mathcal{T}$. Without loss of generality, we assume that no capacity is installed at the *beginning* of $t = 0$ so that $\Delta K_0 = K_0$ (note that K_0 is the amount of capacity installed at the *end* of period $t = 0$).

Let $\Xi_t \subset \mathbb{R}^I$ denote the sample space for the demand in time $t \in \mathcal{T}_0$, and $\xi_t \triangleq (\xi_{1t},\dots,\xi_{It})$ as the realized demand, such that $\xi_t \in \Xi_t$. Without loss of generality, we assume that the demand in period $t = 0$ (i.e., $\xi_0 \in \Xi_0$) is known. Further, let $\xi_{[t]} \triangleq (\xi_0, \xi_1,\dots,\xi_t)$ denote the history of demand up to time t, let $\Xi \triangleq \times_{t=0}^{T} \Xi_t$ be the set of all possible demand sequence realizations, and let $\xi \triangleq \xi_{[T]}$ denote a demand realization over the entire planning horizon.

3.2.1 *Capacity Expansion Policies for the Flexible MCEP*

In the flexible MCEP, capacity decisions are made sequentially based on realized demand. Specifically, the capacity decisions K_t are made according to a capacity expansion policy, which is a mapping from the historical demand $\xi_{[t]}$ to the expansion decisions, for all $t \in \mathcal{T}$. We require the expansion policies to be *non-anticipative*: the capacity decision in time t may only depend on $\xi_{[t]}$ (it cannot use future information). We then define the non-anticipative policies for the flexible MCEP as mappings

$$\mathcal{K}_t : \Xi_0 \times \cdots \times \Xi_t \mapsto \mathbb{K}, \;\; \forall t \in \mathcal{T}_0,$$

where $K_t = \mathcal{K}(\xi_{[t]})$ for any $\xi_{[t]} \in \Xi_0 \times \cdots \times \Xi_t$. Further, we let $\mathcal{K}_{[t]} \triangleq (\mathcal{K}_0, \ldots, \mathcal{K}_t)$ denote the truncation of the policy up to time *t*. Under Assumption A2, the policy space (i.e., the set of non-anticipative and feasible policies) for the flexible MCEP is

$$\overline{\mathcal{K}} \triangleq \left\{ (\mathcal{K}_0, \ldots, \mathcal{K}_T) \middle| \mathcal{K}_{t-1}\left(\xi_{[t-1]}\right) \le \mathcal{K}_t\left(\xi_{[t]}\right), \forall t \in \mathcal{T}, \xi \in \Xi \right\}.$$

3.2.2 *Profits and Costs*

In the MCEP, profit is earned by satisfying the customer demand, and is determined by an allocation problem. Let $\Pi_t(K_{t-1}, \xi_t)$ denote the profit given the realized demand ξ_t and the installed capacity K_{t-1}. Let z_{int} be the demand allocated from customer $i \in \mathcal{I}$ to facility $n \in \mathcal{N}$ in time $t \in \mathcal{T}$. Then, $\Pi_t(K_{t-1}, \xi_t)$ is determined by the value of the following linear program problem:

$$\begin{aligned}
&\Pi_t\left(K_{t-1}, \xi_t\right) \triangleq \max_{z_{int}} \sum_{i \in \mathcal{I}} \sum_{n \in \mathcal{N}} r_{int} z_{int} - \sum_{i \in \mathcal{I}} b_{it} \left(\xi_{it} - \sum_{n \in \mathcal{N}} z_{int} \right) \\
&\text{s.t.} \sum_{n \in \mathcal{N}} z_{int} \le \xi_{it}, \quad \forall i \in \mathcal{I}, \\
&\sum_{i \in \mathcal{I}} z_{int} \le K_{n(t-1)}, \quad \forall n \in \mathcal{N}, \\
&z_{int} \ge 0, \quad \forall i \in \mathcal{I}, n \in \mathcal{N},
\end{aligned} \qquad (1)$$

where r_{int} is the unit revenue for satisfying customer *i*'s demand with facility *n*, and b_{it} is the unit penalty for the unsatisfied demand of customer *i*.

The expansion cost is given by a piecewise linear function. For all $t \in \mathcal{T}_0$, let $c_t(\Delta K_t)$ be the capacity expansion cost given ΔK_t. Denote by $\mathcal{L} \triangleq \{1, \ldots, L\}$ the set of indices for *L* line segments, and let $\left(a_{n1}, \ldots, a_{n(L+1)}\right)$ be the set of breakpoints for the expansion costs for facility $n \in \mathcal{N}$ such that $a_{n1} = 0$ and $K_n^{\max} < a_{n(L+1)}$. Let p_{nlt} and q_{nlt} be the slope and intercept of the l^{th} line segment of the expansion costs for facility *n* in time *t*. For all $t \in \mathcal{T}_0$, the cost function is then:

$$c_t\left(\Delta K_t\right) \triangleq \left\{ \sum_{n\in\mathcal{N}} c_{nt}\left(\Delta K_{nt}\right) \middle| c_{nt}\left(\Delta K_{nt}\right) = p_{nlt}\Delta K_{nt} + q_{nlt}, \text{ if } \Delta K_{nt} \in \left[a_{nl}, a_{n(l+1)}\right), \forall l \in \mathcal{L} \right\}. \quad (2)$$

Eq. (2) can represent arbitrary finite cost functions as ΔK_{nt} is finite. We can set $L = K_n^{\max} + 1$ so that each line segment corresponds to a specific expansion cost at point ΔK_{nt}.

We have a discount factor $0 < \gamma \leq 1$. Given policies $\mathcal{K}_{[T]} \in \overline{\mathcal{K}}$ and the profit/cost structure described above, the cumulative future costs from time period $t = 0$ to $t = T$ for a particular $\xi \in \Xi$ are

$$Q\left(\mathcal{K}_{[T]}, \xi\right) \triangleq c_0\left(\mathcal{K}_0\left(\xi_{[0]}\right)\right) + \sum_{t=1}^{T} \gamma^t \left(c_t\left(\mathcal{K}_t\left(\xi_{[t]}\right) - \mathcal{K}_{t-1}\left(\xi_{[t-1]}\right)\right) - \Pi_t\left(\mathcal{K}_{t-1}\left(\xi_{[t-1]}\right), \xi_t\right)\right).$$

If the decision maker is risk neutral, then the objective of the flexible MCEP is to find the capacity expansion policy, i.e., $\mathcal{K}_{[T]}$, that maximizes the ENPV:

$$\text{ENPV}_{\text{flex}} \triangleq \max_{\mathcal{K}_{[T]} \in \overline{\mathcal{K}}} \mathbb{E}\left[-Q\left(\mathcal{K}_{[T]}, \xi\right)\right]. \quad (3)$$

3.2.3 *A Risk-Averse Flexible MCEP*

When the variance of demand is high, the profits in the risk-neutral Problem (3) can be low for some particular realizations of ξ. Therefore, we need consider the risk of $Q\left(\mathcal{K}_{[T]}, \xi\right)$. Specifically, we incorporate CVaR into the objective function to address this issue. One of the major motivations for using CVaR is that it is readily comparable to the risk-neutral objective function using the mean value, because they are in the same units. In addition, CVaR is a coherent risk measure and also enjoys computational advantages. Alternatively, if one incorporates e.g. mean-variance in the objective function, it is not comparable to the risk-neutral design alternative because of the inconsistency of the units, and the VoF is hard to quantify.

We first recall the definition of CVaR. For a continuous random variable X and a confidence level $\alpha \in (0,1)$, $\mathrm{CVaR}_\alpha(X)$ is the expectation conditional on $X \geq \mathrm{VaR}_\alpha(X)$, where $\mathrm{VaR}_\alpha(X) \triangleq \inf\{y \in \mathbb{R} \mid P\{X \leq y\} \geq \alpha\}$ (Sarykalin et al., 2008). This coincides with the definition of "expected shortfall" (Acerbi, 2002). If $\alpha \to 1$, then $\mathrm{CVaR}_\alpha(X)$ approaches the worst-case cost; whereas if $\alpha \to 0$, it approaches the expected value of X.

We introduce a weight factor $0 \leq \beta \leq 1$ to obtain the mean-CVaR objective function for the MCEP (Shapiro, 2011):

$$\mathrm{ENPV}_\alpha(\beta) \triangleq \max_{\mathcal{K}_{[T]} \in \overline{\mathcal{K}}} \beta\, \mathbb{E}\left[-Q\left(\mathcal{K}_{[T]}, \xi\right)\right] + (1-\beta)\left[-\mathrm{CVaR}_\alpha\left(Q\left(\mathcal{K}_{[T]}, \xi\right)\right)\right]. \quad (4)$$

In this formulation, a decision maker can compromise between risk-neutral and risk-averse policies by adjusting the weight β. If we choose $\beta = 1$, we recover the original risk-neutral model; but if $\beta = 0$, we minimize $\mathrm{CVaR}_\alpha\left(Q\left(\mathcal{K}_{[T]}, \xi\right)\right)$.

We have the following result, since $\mathrm{CVaR}_\alpha(X) \geq \mathbb{E}[X]$ for all $\alpha \in (0,1)$ for any continuous random variable X.

Proposition 1. Given Problems (3) and (4), we have: (i) $\mathrm{ENPV}_\alpha(\beta)$ is non-decreasing in β given any $\alpha \in (0,1)$, and (ii) $\mathrm{ENPV}_{\mathrm{flex}} \geq \mathrm{ENPV}_\alpha(\beta)$ for any $\beta \in [0,1]$ and $\alpha \in (0,1)$.

Proposition 1 states that $\mathrm{ENPV}_\alpha(\beta)$ will not exceed $\mathrm{ENPV}_{\mathrm{flex}}$ given appropriately chosen α and β. Furthermore, as β decreases, decision makers place more emphasis on minimizing CVaR rather than minimizing the expected cost, and so $\mathrm{ENPV}_\alpha(\beta)$ decreases. That is, decision makers will tend to undervalue the system performance when they become more risk averse.

From the perspective of real options theory, we often want to evaluate the VoF compared to an inflexible benchmark problem, because implementing a flexible design for a system usually requires upfront costs. The risk-neutral problem (i.e., Problem (3)) and the risk-averse problem (i.e., Problem (4)) have the same level of flexibility (i.e., both have capacity adjustment options and facility switching options), and so the willingness of the decision maker to enable flexibility may decrease as risk aversion increases.

4 Decision Rule Approximation of the Capacity Expansion Policy

Problem (4) is an MSSP problem with a non-convex objective function. It is widely believed that MSSP is "computationally intractable already when medium-accuracy solutions are sought" (Shapiro and Nemirovski, 2005).

To develop a computationally tractable solution strategy, we approximate $\overline{\mathcal{K}}$ with decision rules. In other words, we restrict the policy space to a class of parametrized functions $\tilde{\mathcal{K}}(\Theta) \subset \overline{\mathcal{K}}$, where $\Theta \subset \mathbb{R}^{\dim(\Theta)}$ is some admissible set of parameters. Then, we can optimize the parameters $\theta \in \Theta$ which determine the decision rule $\tilde{\mathcal{K}}(\Theta)$, instead of optimizing over all non-anticipative policies in $\overline{\mathcal{K}}$. Of course, a particular decision rule does not guarantee global optimality over the problem space. They offer, however, significant computational advantages and managerial insights that help operators leverage the benefits of flexibility.

4.1 If–Then Decision Rules

We focus on if–then decision rules to approximate the policy space $\overline{\mathcal{K}}$. The motivations for our choice of if–then decision rules are threefold. First, expansion decisions are binary by their very nature—the capacity is either expanded or it is not. In addition, the output of the decision rule should be integral because capacity is discrete, so a nonlinear decision rule is required. Second, if–then decision rules mimic the decision-making behavior of human beings and are more intuitive and interpretable in practical implementation (Cardin et al., 2017). Third, some optimal if–then policies for capacity expansion problems have been reported in the literature (Eberly and Van Mieghem, 1997; Angelus et al., 2000), and some numerical results have shown that if–then decision rules can yield high-performance solutions for discrete MCEPs (Cardin et al., 2017; Zhao et al., 2018).

An if–then decision rule in a single facility setting is stated as: if the capacity gap (i.e., demand minus capacity) of the facility exceeds a threshold, then we expand the capacity up to a certain level. However, capacity gaps for individual facilities are hard to quantify in multi-facility problems, as a facility can serve more than one customer. To address this feature of the MCEP, we introduce a weight matrix in order to calculate the weighted capacity gap for each facility. We take $W \in \mathbb{R}^{I \times N}$ as a preset weight matrix such that $W^\top \xi_t \in \mathbb{R}^N$. Then, we compute the weighted capacity gaps $\lfloor W^\top \xi_t \rceil - K_{t-1}$, where $\lfloor \cdot \rceil$ denotes the operation of rounding to the nearest integer. These weighted capacity gaps are the trigger conditions for our if–then rules. In the remainder of this subsection, we will first present the general form of the if–then decision rule and then introduce the choices of *W*.

Let $\theta_{1,n}$ denote the parameter that controls capacity adjustments when the if–then rule is triggered and let $\theta_{2,nt}$ be the threshold for the trigger condition. Define the parameter vectors $\theta_1 \triangleq (\theta_{1,1}, \dots, \theta_{1,N})$ and $\theta_2 \triangleq (\theta_{2,11}, \dots, \theta_{2,NT})$. The admissible sets for the parameters θ_1 and θ_2 are $\Theta_1 \triangleq \{\theta_1 \in \mathbb{Z}^N \mid 0 \le \theta_1 \le \theta_1^{\max}\}$ and $\Theta_2 \triangleq \{\theta_2 \in \mathbb{R}^{N \times T} \mid 0 \le \theta_2 \le \theta_2^{\max}\}$, respectively. We further define $\theta \triangleq (\theta_1, \theta_2)$ and $\Theta \triangleq \Theta_1 \times \Theta_2$.

Our if–then decision rule is as follows: for all $n \in \mathcal{N}, t \in \mathcal{T}, \xi \in \Xi, K_{t-1} \in \mathbb{K}$, we define

$$\mathcal{K}_{nt}\left(K_{n(t-1)}, \xi_t; \theta\right) \triangleq \begin{cases} \lfloor \sum_{i \in \mathcal{I}} W_{in} \xi_{it} \rceil + \theta_{1,n}, & \text{if } \lfloor \sum_{i \in \mathcal{I}} W_{in} \xi_{it} \rceil - K_{n(t-1)} \ge \theta_{2,nt} \text{ and } \\ & \quad \lfloor \sum_{i \in \mathcal{I}} W_{in} \xi_{it} \rceil + \theta_{1,n} \le K_n^{\max}, \\ K_{n(t-1)}, & \text{otherwise.} \end{cases} \tag{5}$$

For given $\theta \in \Theta$, we let $\tilde{\mathcal{K}}_{[T]}(\theta) \triangleq \left(\mathcal{K}_0, \mathcal{K}_1(\cdot;\theta), \ldots, \mathcal{K}_T(\cdot;\theta)\right)$ denote the vector of parametrized if–then decision rules encoded by the above policy. Note that we assume Ξ_0 is known, so the policy for the initial capacity decision (i.e., $\mathcal{K}_0 : \Xi_0 \mapsto \mathbb{K}$) is independent of *θ*.

Policy (5) states that if the weighted capacity gap of facility *n* (i.e., $\lfloor \sum_{i \in \mathcal{I}} W_{in} \xi_{it} \rceil - K_{n(t-1)}$) exceeds the threshold $\theta_{2,nt}$ *and* the expanded capacity does not exceed the maximum possible capacity level, then we expand the capacity of facility *n* up to level $\lfloor \sum_{i \in \mathcal{I}} W_{in} \xi_{it} \rceil + \theta_{1,n}$. Otherwise, the capacity is unchanged. Note that $\theta_{1,n}$ is integral, so the decision rule automatically yields integral expansion decisions.

In Policy (5), the design of the weight matrix *W* is case-specific. For example, it can be a preset allocation matrix of the demands of the customers conditional on there being sufficient capacity. From a managerial point of view, the entry W_{in} of the weight matrix can also be interpreted as the profit coefficient of customer *i* with respect to facility *n*. In other words, if the per unit demand from customer *i* is more profitable for facility *n*, then we should tend to allocate more demand from customer *i* to facility *n* and so W_{in} should be larger. For example, if there are the same number of facilities as customers (i.e., *I* = *N*), and there is a bijective map from each facility/customer to its most profitable customer/facility counterpart, then the weight matrix can be the *N* × *N* identity matrix. In this case, the weighted capacity gap of a facility is calculated by subtracting the current capacity from the most profitable demand served by this facility. We refer interested readers to (Zhao et al., 2018) for further discussion of the choice of *W*.

Now we want to optimize the scenario-independent policy parameters $\theta \in \Theta$ of Policy (5), rather than the policy $\mathcal{K}_{[T]} \in \overline{\mathcal{K}}$, with respect to the mean-CVaR objective function. Problem (4) can then be approximated by

$$\max_{\theta\in\Theta} \; \beta\, \mathbb{E}\left[-Q\left(\tilde{\mathcal{K}}_{[T]}(\theta),\xi\right)\right] + (1-\beta)\left[-\mathrm{CVaR}_{\alpha}\left(Q\left(\tilde{\mathcal{K}}_{[T]}(\theta),\xi\right)\right)\right], \quad (6)$$

which optimizes over policies of the form (5). Based on (Rockafellar and Uryasev, 2000, Theorem 2), we introduce an auxiliary variable $u \in \mathbb{R}$ and recall the variational form of $\mathrm{CVaR}_{\alpha}(X) = \inf_{u\in\mathbb{R}}\left\{u + \frac{1}{1-\alpha}\mathbb{E}[X-u]_{+}\right\}$, where $[\cdot]_{+}$ denotes $\max\{\cdot,0\}$. Then, Problem (6) is equivalent to

$$\max_{\theta\in\Theta, u\in\mathbb{R}} \; -(1-\beta)u - \mathbb{E}\left[\beta Q\left(\tilde{\mathcal{K}}_{[T]}(\theta),\xi\right) + \frac{1-\beta}{1-\alpha}\left[Q\left(\tilde{\mathcal{K}}_{[T]}(\theta),\xi\right) - u\right]_{+}\right]. \quad (7)$$

4.2 MILP Transformations of the Decision Rule-based Model

In this subsection, we transform Problem (7) into an MILP. First, we use sample average approximation to transform the expectation in the objective function into a finite sum. Let $\mathcal{S} \triangleq \{1,\ldots,S\}$ be the set of indices of the scenario, and $\{\xi^1,\ldots,\xi^S\}$ be a set of sample paths generated via Monte Carlo simulation of customer demand, where $\xi^s \triangleq (\xi_1^s,\ldots,\xi_T^s)$ for all $s\in\mathcal{S}$. We assume equal probabilities for the generated sample paths in this paper, but the proposed method also works if the samples have non-uniform probabilities. The SAA of Problem (7) is then

$$\max_{\theta\in\Theta, u\in\mathbb{R}} \; -(1-\beta)u - \frac{1}{S}\sum_{s\in\mathcal{S}}\left[\beta Q\left(\tilde{\mathcal{K}}_{[T]}(\theta),\xi^s\right) + \frac{1-\beta}{1-\alpha}\left[Q\left(\tilde{\mathcal{K}}_{[T]}(\theta),\xi^s\right) - u\right]_{+}\right]. \quad (8)$$

To incorporate the specifics of $\tilde{\mathcal{K}}_{[T]}(\theta)$ into this optimization, we develop the exact MILP formulation of Problem (8) by using the Big-M method. We introduce a set of auxiliary binary variables $\delta_t^s \triangleq (\delta_{1t}^s,\ldots,\delta_{Nt}^s)$ for each $t\in\mathcal{T}$, $s\in\mathcal{S}$, where δ_{nt}^s is such that the capacity of facility n is expanded to $\lfloor \sum_{i\in\mathcal{I}} W_{in}\xi_{it}^s \rceil + \theta_{1,n}$ if $\delta_{nt}^s = 1$; otherwise, the capacity is unchanged. To enforce this constraint, we introduce a large constant $M > 0$, the $N \times N$ identity matrix $\mathbb{I}_{N\times N}$, and an N-dimensional vector of ones $\mathbf{1}_N$.

To linearize the nonlinear term $\left[Q\left(\tilde{\mathcal{K}}_{[T]}(\theta), \xi\right) - u\right]_+$, we can introduce an auxiliary variable $\eta^s \in \mathbb{R}$ for each $s \in \mathcal{S}$, such that $\eta^s \geq Q\left(\tilde{\mathcal{K}}_{[T]}(\theta), \xi^s\right) - u$ and $\eta^s \geq 0$. We ultimately obtain the following reformulation of Problem (8)

$$\min_{K_0, u, \theta, \eta^s, \delta_t^s, K_t^s} \beta c_0(K_0) + (1-\beta) u + \frac{1}{S}\sum_{s\in\mathcal{S}}\left[\beta \sum_{t\in\mathcal{T}} \gamma^t \left(c_t\left(K_t^s - K_{t-1}^s\right) - \Pi_t\left(K_{t-1}^s, \xi_t^s\right)\right) + \frac{1-\beta}{1-\alpha}\eta^s\right] \quad (9a)$$

$$\text{s.t.} \lfloor W^\top \xi_t^s \rceil - K_{t-1}^s - \theta_{2,t} \geq \left(M \mathbb{I}_{N\times N}\right)^\top \left(\delta_t^s - \mathbf{1}_N\right), \ \forall t \in \mathcal{T}, s \in \mathcal{S}, \quad (9b)$$

$$\lfloor W^\top \xi_t^s \rceil - K_{t-1}^s - \theta_{2,t} < \left(M \mathbb{I}_{N\times N}\right)^\top \delta_t^s, \ \forall t \in \mathcal{T}, s \in \mathcal{S}, \quad (9c)$$

$$K_t^s \leq \lfloor W^\top \xi_t^s \rceil + \theta_1 + \left(M \mathbb{I}_{N\times N}\right)^\top \left(\mathbf{1}_N - \delta_t^s\right), \ \forall t \in \mathcal{T}, s \in \mathcal{S}, \quad (9d)$$

$$K_t^s \geq \lfloor W^\top \xi_t^s \rceil + \theta_1 + \left(M \mathbb{I}_{N\times N}\right)^\top \left(\delta_t^s - \mathbf{1}_N\right), \ \forall t \in \mathcal{T}, s \in \mathcal{S}, \quad (9e)$$

$$K_t^s - K_{t-1}^s \leq \left(M \mathbb{I}_{N\times N}\right)^\top \delta_t^s, \ \forall t \in \mathcal{T}, s \in \mathcal{S}, \quad (9f)$$

$$K_t^s \leq K^{\max}, \ \forall t \in \mathcal{T}, s \in \mathcal{S}, \quad (9g)$$

$$\eta^s \geq c_0(K_0) + \sum_{t\in\mathcal{T}} \gamma^t \left(c_t\left(K_t^s - K_{t-1}^s\right) - \Pi_t\left(K_{t-1}^s, \xi_t^s\right)\right) - u, \quad \forall s \in \mathcal{S}, \quad (9h)$$

$$\eta^s \in \mathbb{R}_+, \ K_t^s \in \mathbb{R}^N, \ \delta_t^s \in \{0,1\}^N, \quad \forall t \in \mathcal{T}, s \in \mathcal{S}, \quad (9i)$$

$$K_0^s = K_0, \ \forall s \in \mathcal{S}, \quad (9j)$$

$$K_0 \in \mathbb{K}, \ u \in \mathbb{R}, \ \theta \in \Theta. \quad (9k)$$

Problem (9) is an MILP (note that the piecewise cost functions $c_t(\cdot)$ can be expressed via a set of linear constraints with binary and continuous auxiliary variables). Constraints (9b)–(9g) are the Big-M formulation for the if–then decision rule. These constraints are non-anticipative because the scenario-dependent capacity decisions K_t^s are determined by θ and the historical information (i.e., $\xi^s_{[t]}$), having no access to future information. In addition, the right-hand sides of Policy (5) are integral, so Constraints (9b)–(9g) map from continuous capacity levels to discrete capacity expansion decisions. Constraint (9j) is introduced so that K^s_0 is well defined in Eq. (9h).

5 Solution Method: A Decomposition Algorithm

In this section, we propose a decomposition algorithm for Problem (9). We first reformulate Problem (9) as a two-stage stochastic programming problem. The first stage determines the initial capacity, the parameters of the decision rule, and the auxiliary variable *u*. The second stage contains all of the subsequent dynamics of the system, including the capacity expansion decisions and the allocation decisions, in response to customer demand, and returns the future costs. We can further decompose the second stage over scenarios, compute the subgradients of the recourse function, and then update the initial capacity and the parameters of the decision rule.

5.1 Two-Stage Decomposition

The first stage corresponding to Problem (9) determines the parameters θ, the initial capacity K_0, and the auxiliary variable u from the variational form of CVaR, which are all scenario independent (i.e., here-and-now decisions). The second-stage decisions determine the capacity plan $\left(K^s_1, \dots, K^s_T\right)$ and the auxiliary variables η^s and δ^s_t, which depend on the scenario (i.e., wait-and-see decisions). The recourse function defined by the second-stage problem for a specific scenario $s \in \mathcal{S}$ is:

$$R^s\left(K_0, u, \theta\right) \triangleq \min_{\eta^s, \delta^s_t, K^s_t} \left\{ \beta \sum_{t=1}^{T} \gamma^t \left(c_t \left(K^s_t - K^s_{t-1} \right) - \Pi_t \left(K^s_{t-1}, \xi^s_t \right) \right) + \frac{1-\beta}{1-\alpha} \eta^s, \text{ s.t. } (9\text{b}) - (9\text{j}) \right\}. \quad (10)$$

The recourse function here returns the multi-period revenue over the time periods $t = 1, \dots, T$. Then, the first stage problem is

$$\min_{K_0,u,\theta}\left\{\beta c_0\left(K_0\right)+\left(1-\beta\right)u+\frac{1}{S}\sum_{s=1}^{S}R^s\left(K_0,u,\theta\right),\ \text{s.t.}(9\text{k})\right\}. \quad (11)$$

However, this two-stage problem is difficult to solve because the recourse function $R^s\left(K_0,u,\theta\right)$ is highly nonconvex due to the integer wait-and-see variables (i.e., δ_t^s) and the nonconvex costs $c_t(\cdot)$. Therefore, traditional methods like the Benders decomposition are not applicable.

Fortunately, this problem has exploitable structure. Once the initial capacity K_0, the control parameters θ, and u are all fixed, we can determine $\left(K_1^s,\dots,K_T^s\right)$ and evaluate the expected future costs via Policy (5) given the demand vector ξ^s in scenario $s\in\mathcal{S}$. Then, we can compute the subgradients of the recourse functions $R^s\left(K_0,u,\theta\right)$ and update the scenario-independent decisions.

We remark that the subgradients of $R^s\left(K_0,u,\theta\right)$ with respect to θ_2 are always zero because θ_2 only appears in the trigger conditions of Policy (5). In this case, θ_2 cannot be updated via the subgradients of the recourse function. To address this difficulty, we further decompose Problem (11) and update θ_2 separately via a stochastic approximation scheme (see Step 1 of the decomposition algorithm).

5.2 Algorithm Procedure

In this subsection, we briefly summarize the procedure and the innovation of the algorithm (see the online supplement for the detailed implementation). Our algorithm updates the solutions iteratively via a cut generation method. Specifically, we solve the epigraph formulation of Problem (11) in the m^{th} iteration:

$$\min_{K_0,u,\theta_1,y}\ \beta c_0\left(K_0\right)+\left(1-\beta\right)u+y \quad (12a)$$

$$\text{s.t.}\ y\geq\left(\phi^{m'}\right)^{\top}\left(K_0,u,\theta_1\right)+\phi_0^{m'},\ \forall m'=1,\dots,m-1, \quad (12b)$$

$$K_0\in\mathbb{K},\ u\in\mathbb{R},\ \theta_1\in\Theta_1,\ y\in\mathbb{R}, \quad (12c)$$

where Eq. (12b) contains the cuts generated up to iteration *m*, and $\phi^{m'}$ and $\phi_0^{m'}$ are the coefficients corresponding to these cuts. Once the above problem is solved, we denote its optimal solution as $\left(K_0^m, u^m, \theta_1^m\right)$. The algorithm then proceeds iteratively via three major steps:

- Step 1: Fix $\left(K_0^m, u^m, \theta_1^m\right)$ and update θ_2^m by averaging the solutions via a stochastic approximation scheme.
- Step 2: Once $\left(K_0^m, u^m, \theta_1^m, \theta_2^m\right)$ are fixed, we compute subgradients of the recourse functions $R^s(\cdot)$ for all $s \in \mathcal{S}$ at $\left(K_0^m, u^m, \theta_1^m\right)$, and construct subgradient cuts.
- Step 3: We add the subgradient cuts to Eq. (12b), and compute $\left(K_0^{m+1}, u^{m+1}, \theta_1^{m+1}\right)$ by solving Problem (12). We then go back Step 1 and proceeds the algorithm iteratively until our termination conditions are met.

We now elaborate on these three steps in more detail.

Step 1: Update θ_2 via Stochastic Approximation

Suppose that we have fixed the first-stage decisions $\left(K_0^m, u^m, \theta_1^m\right)$ in the *m*th iteration. One can update θ_2 by solving

$$\min_{\theta_2 \in \Theta_2} \frac{1}{S} \sum_{s=1}^{S} R^s\left(K_0^m, u^m, \theta_1^m, \theta_2\right),$$

where the objective is the expected future costs provided by Problem (10). However, this may result in a large-scale MILP because the number of scenarios *S* can be large.

Instead, we select one scenario at a time and construct a corresponding single-scenario problem. Let $k \geq 0$ be a counter for the inner iterations in Step 1. We assign equal probability to each scenario in $\mathcal{S}$, and randomly select one scenario

$\left(\xi_1^{s_k}, \ldots, \xi_T^{s_k}\right)$ for $s_k \in \mathcal{S}$ without replacement. Based on Problem (10), we construct the following single-scenario problem to optimize θ_2 given $\left(K_0^m, u^m, \theta_1^m\right)$ and ξ^{s_k}:

$$\min_{\theta_2 \in \Theta_2} R^{s_k}\left(K_0^m, u^m, \left(\theta_1^m, \theta_2\right)\right) + c_\theta \sum_{t \in \mathcal{T}} \sum_{n \in \mathcal{N}} \theta_{2,nt}, \qquad (13)$$

where $c_\theta > 0$ is a small constant that is intended to regularize θ_2. The objective of this problem is to find the optimal θ_2 such that the cumulative costs given sample path $\left(\xi_1^{s_k}, \ldots, \xi_T^{s_k}\right)$ are minimized.

We then average the optimal θ_2 of the sampled single-scenario problems via an update rule (see Subsection B.1 of the online supplement for details). By this method, we can compute an approximate θ_2 by evaluating a portion of, rather than all of, the scenarios in set $\mathcal{S}$. In particular, for practically-sized problem instances, a single-scenario problem only has a few hundred binary variables, which can be directly solved by commercial solvers in seconds.

Step 2: Calculating the Subgradients of the Recourse Function

Once θ_2^m is computed, we can fix $\left(K_0^m, u^m, \theta_1^m, \theta_2^m\right)$ and calculate the subgradients of the recourse functions with respect to $\left(K_0^m, u^m, \theta_1^m\right)$. Given the subgradients, we then generate subgradient cuts and update $\left(K_0, u, \theta_1\right)$ by solving Problem (12).

Denote $\phi^{m,s} \in \partial R^s$ as a subgradient of the recourse function with respect to $\left(K_0^m, u^m, \theta_1^m\right)$ and $\phi_0^{m,s}$ as the corresponding intercept, such that

$$R^s\left(K_0^m, u^m, \theta_1^m\right) = \left(\phi^{m,s}\right)^\top \left(K_0^m, u^m, \theta_1^m\right) + \phi_0^{m,s}, \quad \forall s \in \mathcal{S}, m \geq 1.$$

According to Eq. (10), we need to compute the subgradients of $c_t(\cdot)$ and $\Pi_t(\cdot)$ in order to compute $\phi^{m,s}$ and $\phi_0^{m,s}$. We put the detailed computation of $\phi^{m,s}$ and $\phi_0^{m,s}$ in the online supplement (Subsection B.2) and briefly summarize the procedures below:

- We first derive a closed form for the capacity decisions $K_t^{m,s}$ with respect to (K_0^m, u^m, θ_1^m) given scenario $s \in \mathcal{S}$ (see Lemma B.1).
- We compute the subgradient of $c_t(\cdot)$ according to the definition of Eq. (2) (see Lemma B.2).
- We compute the subgradient of $\Pi_t(\cdot)$ by solving the dual of Problem (1) (see Lemma B.3).
- Finally, we compute $\phi^{m,s}$ and $\phi_0^{m,s}$, and construct the subgradient cut (see Proposition B.2).

Given $\phi^{m,s}$ and $\phi_0^{m,s}$ for all $s \in \mathcal{S}$ and $m \geq 1$, we can compute $\phi^m \triangleq \frac{1}{S}\sum_{s \in \mathcal{S}} \phi^{m,s}$ and $\phi_0^m \triangleq \frac{1}{S}\sum_{s \in \mathcal{S}} \phi_0^{m,s}$. Then, a new subgradient cut is given by

$$y \geq (\phi^m)^\top (K_0, u, \theta_1) + \phi_0^m. \qquad (14)$$

We see that if $(K_0, u, \theta_1) = (K_0^m, u^m, \theta_1^m)$, then the cut recovers $\frac{1}{S}\sum_{s \in \mathcal{S}} R^s(K_0^m, u^m, \theta^m)$. Otherwise, it returns the recourse along the computed subgradient.

Step 3: Update (K_0, u, θ_1) by Solving the First-Stage Problem

Suppose we have a set of subgradient cuts computed from Step 2 up to iteration *m*. We add the subgradient cuts to Problem (12), and solve the problem in iteration *m* + 1. Problem (12) is an MILP with $N + N \times L$ integer variables. For practically-sized problems (e.g., $N \leq 10$ and $L \leq 20$), it can be directly solved with commercial solvers.

The decomposition algorithm terminates when the value of the objective function of Problem (11), computed for the best-found solution (K_0, u, θ), is close enough to that of Problem (12) *or* when a preset number of iterations is reached. The details of the entire procedure are provided in the online supplement (Subsection B.3).

6 Numerical Study: Capacity Planning for a WTE System

The numerical study in this section is adapted and extended from a real case study of a multi-facility WTE system in Singapore (Cardin and Hu, 2016; Zhao et al., 2018). The decision maker aims to deploy WTE facilities to dispose of food waste. There are five candidate sites in different sectors of Singapore for establishing the WTE facilities. The WTE facility disposes of food waste collected from each sector using an innovative anaerobic digestion technique. This technique transforms the food waste into biogas, which can then be used to generate electricity. If the collected food waste exceeds the disposal capacity of the WTE system, undisposed waste will be sent to a landfill, incurring greater disposal costs (e.g., penalties). The revenues of the WTE system come from selling the electricity and from the fees collected for disposing of the food waste. The costs consist of unit disposal costs, transportation costs, penalty costs for undisposed food waste, and disposal capacity expansion costs. We omit annual fixed costs, which makes this setting slightly different from the problem in (Cardin and Hu, 2016; Zhao et al., 2018). However, the proposed method can still be used with some modifications.

This problem can be formulated as an MCEP with five WTE facilities and five customers (i.e., sectors). The waste generated in each sector can be viewed as stochastic demand for disposal. The decision maker wants to maximize the economic performance by optimizing the capacity expansion policy of the WTE facilities over a 15-year planning horizon.

Given that the inspection of historical waste generation patterns show a clear combination of mean drift and random fluctuations, the generation of food waste in each sector (i.e., the demand of each customer) is assumed to be standard geometric Brownian motion (GBM) (Cardin and Hu, 2016). Demand evolves according to:

$$\xi_{it} = \left(\bar{\mu} + \bar{\sigma}\omega_t\right)\xi_{i(t-1)}, \quad \forall\, i \in \{1, \dots, 5\}, t \in \{1, \dots, 15\},$$

where ξ_{it} is the amount of waste generated in sector i at time t, $\bar{\mu}$ is the percentage drift, $\bar{\sigma}$ is the percentage volatility, and ω_t are standard i.i.d. normal random variables for all $t \in \mathcal{T}$. In the numerical study, we assume that $\bar{\mu}$ is 4% and $\bar{\sigma}$ is 16%, and the initial waste is the vector $\xi_0 = [498, 518, 293, 460, 382]$ (unit: tonnes per day).

The WTE facilities are modular so the capacity decisions are discrete. One unit of capacity can dispose of 100 tonnes of food waste per day, and the maximum capacities for the five candidate sites are given by $K^{\max} = [16, 10, 10, 10, 10]$. The WTE system enjoys economies of scale, and capacity expansion costs are given by a power function. We linearize the power function to derive the cost function presented in Eq. (2). We set the discount factor $\gamma = 0.926$ according to (Cardin and Hu, 2016). The detailed data, the evaluation of the unit revenue (i.e., r_{int}) and the expansion costs, and supplemental numerical studies can be found in the online supplement (see Section C).

6.1 Comparison of Different Design Approaches

We compare two policies for the capacity planning of the multi-facility WTE system:

- An *inflexible policy* that sets up a static capacity expansion plan at the beginning of the planning horizon and does not revise the plan regardless of the realizations of stochastic demand for waste disposal.
- A *flexible if–then policy* that dynamically adjusts the capacity of the WTE facilities based on realizations of stochastic demand.

The inflexible policy is derived from an inflexible MCEP model, which can be solved to optimality by Benders decomposition (Benders, 1962). The inflexible policy can be viewed as an attainable lower bound for the multi-stage MCEP considered in this paper. The detailed model and algorithm to solve the inflexible model can be found in the online supplement (see Subsection C.3).

The flexible if–then policy is determined by our proposed method. For our simulations, we design the weight matrix W in the decision rule (5) using the technique of (Zhao et al., 2018). We consider the allocation of waste of sector $i \in \mathcal{I}$ to its three closest facilities after analyzing the distances between the candidate sites and tuning the parameters. As an illustration, the weight factors for facility n = 1 are given by $(w_{i1})_{i \in \mathcal{I}} = (0.63, 0.44, 0, 0, 0)^{\top}$ (see the online supplement for a detailed calculation). Then, the decision rule for facility n = 1 in time $t \in \mathcal{T}$ is

$$K_{1t} = \begin{cases} \left\lfloor \frac{0.63\xi_{1t} + 0.44\xi_{2t}}{100} \right\rceil + \theta_{1,1}, & \text{if } \left\lfloor \frac{0.63\xi_{1t} + 0.44\xi_{2t}}{100} \right\rceil - K_{1(t-1)} \geq \theta_{2,1t} \text{ and } \left\lfloor \frac{0.63\xi_{1t} + 0.44\xi_{2t}}{100} \right\rceil + \theta_{1,1} \leq 16, \\ K_{1(t-1)}, & \text{otherwise.} \end{cases}$$

The flexible and inflexible models are solved separately by generating 4,000 demand scenarios via Monte Carlo simulation. To compare them, we evaluate the economic performance of each design by using an identical evaluation sample set. The evaluation sample set consists of 12,000 scenarios that are generated via Monte Carlo simulation. The rationale underlying this evaluation test is to eliminate the bias introduced by using different sample sets to optimize different models.

In our numerical studies, we analyze the economic performance of the system given different risk preferences of the decision maker reflected by the parameter β. Given the inflexible baseline design, we can calculate the difference between the economic performance of the flexible policy and its inflexible counterpart (i.e., the VoF).

6.2 Simulation Results and Discussion

6.2.1 *The proposed method captures the decision-maker's risk preferences*

In this subsection, we test the performance of the proposed method under different risk preferences by tuning α and β. In practice, we suggest fixing α and tuning β because the parameter α is more intuitive and more easily interpretable.

Common choices of α are 0.99, 0.95, and 0.90, which means that we try to minimize the mean values of the worst 1%, 5%, and 10% losses, respectively (Krokhmal et al., 2002).

In the remainder of this subsection, we set $\alpha = 0.95$ unless otherwise specified. When $\alpha = 0.95$, we minimize the expected costs above the 95^{th}-percentile of all costs. Equivalently, the objective of Problem (4) is to maximize the expected profits below the 5^{th}-percentile. A detailed two-way sensitivity analysis on α and β can also be found in the online supplement.

The simulation results for different policies are presented in Table 2, where the five metrics are the statistical results of the net present values (NPVs) computed using 12,000 scenarios in the evaluation test. The expected net present values (ENPVs) are the mean values of the NPVs of the test scenarios, and “Min” (“Max ”) indicates the result of the worst (best) scenario. Each case is run three times and then the values are averaged.

If we use an inflexible capacity expansion policy, the ENPV is 238.0 million S$, the 5^{th}-percentile of the NPVs is 145.0 million, and the worst-case NPV is –246.9 million. If we use a flexible policy (with $\alpha = 0.95$ and $\beta = 0.99$), then the ENPV is 295.7 million, the 5^{th}-percentile is 221.9 million, and the worst-case NPV is 83.6 million. We see that the flexible policy dominates the inflexible policy in terms of the five metrics presented in Table 2. In particular, the worst-case NPV is significantly improved. The worst-case NPV for the inflexible policy is negative, while the worst-case NPV of the flexible policies ranges from 78.2 million to 87.8 million.

These results demonstrate that the 5^{th}-percentile of the NPVs decreases as β increases, while the ENPV increases as β increases (see Figure 2a). If we fix $\alpha = 0.95$ and change β from 0.99 to 0.01, the 5^{th}-percentile rises from 221.9 million to 230.7 million and the ENPV declines from 295.7 million to 286.3 million. In addition, the VoF decreases from 57.7 million to 48.3 million, which means that the expected value gained from flexibility decreases as the weight factor β decreases. This phenomenon may occur because our particular flexible capacity-expansion policy focuses on improving the upside potential. This means that from the perspective of a risk-averse decision-maker, the system gains less value by having the option of adjusting its capacity. The result from Figure 2a also verifies the conclusion of Proposition 1 numerically.

6.2.2 *The risk-averse policy is more conservative in expansion*

We examine the optimal solutions of Problem (4) for different risk preferences. The optimal initial capacity given $\beta = 0.99$ is $K_0^* = [6,6,0,8,4]$ and the optimal parameter of the decision rule is $\theta_1^* = [1,4,0,2,2]$. In contrast, the optimal solutions given $\beta = 0.01$ is $K_0^* = [5,5,4,5,4]$ and $\theta_1^* = [0,4,0,1,1]$.

Intuitively, one may expect that a more conservative decision maker will choose smaller initial capacities. However, it is interesting to see from the simulation results that the risk-averse expansion policy may not always establish smaller initial capacities. The initial capacity of facility n = 3 under the risk-neutral policy ($\beta = 0.99$) is zero. This happens because the initial amount of waste from sector i = 3 is small ($\xi_{30} = 293$), so this policy tends to take the risk of transporting the demand from this sector to others. However, the risk-averse policy ($\beta = 0.01$) establishes four units of capacity in this sector. For other sectors that have more initial waste (i.e., facility 1, 2, 4, 5), the policy with $\beta = 0.01$ installs less initial capacity, and its expansion in these sectors is also more conservative when the decision rule is triggered.

The cumulative density functions of the evaluation samples of these two policies are plotted in Figure 2b. We see that the policy with $\beta = 0.01$ does not fully exploit the upside expansion opportunity, whereas the one with $\beta = 0.99$ does, but it does reduce the downside risk because it is more conservative.

6.2.3 *Flexible policies are more robust than inflexible policies against variations in the amount of waste*

We analyze the sensitivity of the VoF to the percentage volatility of the amount of waste. As can be seen from Table 3, the VoF of the risk-neutral flexible policy (i.e., $\beta = 0.99$) increases as the volatility increases but decreases when the volatility declines. However, the VoF of the risk-averse flexible policy (i.e., $\beta = 0.01$) slightly decreases when the volatility increases from 0.16 to 0.24. This is because the VoF is calculated in terms of the ENPV. Even though the VoF may increase when the amount of waste has higher volatility, the downside risk of the system is also higher (more specifically, the 5^{th}-percentile of the NPV decreases from 230.7 to 166.4 million S$). In this situation, a risk-averse decision maker may underestimate the VoF.

We also test the robustness of the policies when the demand model generating evaluation samples differs from the one generating the training samples. These results imply that the flexible policy is more robust than the inflexible policy to inaccuracy in the demand model (see Subsection C.2 of the online supplement).

6.2.4 *The proposed method yields high performance policies and is more scalable*

We compare the proposed method with: (i) an ADP with neural networks being approximators (Zhao et al., 2017); and (ii) the decision rule-based method solved by BACD (Zhao et al., 2018). As ADP and BACD are designed for risk-neutral MCEPs, we set β = 1 in our proposed method. According to Proposition 1, any policies with $\beta < 1$ yield smaller ENPVs. Among these methods, the result from ADP can be viewed as an upper bound, as it can derive ϵ-optimal solutions with high probability when the approximator is rich enough (Munos and Szepesvári, 2008). The numerical studies were performed on a workstation with an Intel Xeon Gold 5218 CPU and 32 GB RAM via a Matlab R2018a environment. All simulations were accelerated by parallel computing with 16 cores.

The simulation results are presented in Table 4. The ENPV calculated by the proposed method is 295.8 million, which is close to the performance of ADP and slightly outperforms BACD. The CPU time of the proposed method is only 391 s, but it takes 4107 s for BACD and more than 100 hours for ADP. Though the CPU time of the inflexible design is merely 351 s, it is dominated by all of the flexible designs in terms of the metrics listed in Table 4.

The proposed method is much faster than BACD. In BACD, cuts are generated by solving large-scale LPs. However, the proposed method generates the subgradient cuts via some analytical results, and by solving small-scale LPs, which is more time-efficient than BACD. We test the scalability of these two methods via problem instances with $N = 5$ and $S = 4000$. We then increase the number of customers I and the number of time periods T. As can be seen in Table 5, the CPU times of the proposed method are less than 1000 seconds when the problem size increases, but the time of BACD increases significantly. These results indicate that the proposed method is more scalable than BACD in solving large-scale problems.

7 Conclusion

In this paper, we have established a flexible model for multi-facility capacity expansion that captures the risk preferences of decision makers using a mean-CVaR objective. To solve our risk-averse model, we approximated the capacity expansion policy of the multi-stage problem with an if–then decision rule, and then optimized over if–then decision rules with a customized decomposition algorithm.

Our simulation results show that the decision maker is able to choose a policy with a higher ENPV or with a higher 5^{th} -percentile of the NPVs by simply adjusting a weight factor in the objective function. In addition, the ENPV decreases as the decision maker becomes more risk averse, indicating that the decision maker may prefer to pay less for flexibility. Further, our simulation results show that a risk-averse expansion policy may not always establish smaller initial capacity—a risk-neutral policy may expand less than a risk-averse policy in certain facilities if the corresponding demand is low.

Although we emphasized CVaR in this paper, our proposed method can be extended to other risk measures. For example, if we capture risk with a utility function, then we can still use if–then decision rules to approximate the policy space, and then calculate subgradients of the cost/profit functions to construct cuts to update the here-and-now decisions.

Our algorithm is an improvement on the BACD algorithm in (Zhao et al., 2018). BACD is proposed to solve risk-neutral MCEPs with if–then rules approximation, but our algorithm can handle risk-averse problems. In addition, we improve the algorithm with respect to the cut-generation method (i.e., Step 2 of the algorithm). In BACD, the cuts are generated via branch-and-cut, where large-scale LPs need to be solved in each iteration. In this paper, subgradient cuts are generated via some analytical results, and by solving small-scale LPs. Our numerical studies show that the proposed algorithm is more scalable than BACD, especially as the number of time periods or the number of customers increases.

Though our algorithm may not obtain an optimal policy, our numerical studies show that its performance is close to ADP—which can find a near-optimal policy but is extremely slow. In addition, the improvement of the system performance of the proposed decision-based method over the baseline design (i.e., the inflexible MCEP) is more than 20%. This improvement is even higher when the demand model is inaccurate.

There are many possibilities for future work. In this paper, we use if–then decision rules to solve strategic capacity expansion problem. Further research can incorporate expansion lead time, or explore other families of decision rules. We may also consider robustness against an unknown demand model. Our present numerical experiments suggest that our approach has some intrinsic robustness against demand uncertainty, so we can build on this initial proof of concept.

Acknowledgments

This research was supported by the National Natural Science Foundation of China [grant numbers 72001141, 71931007, 72031007], and by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

Notes on contributors

Dr. Sixiang Zhao is an Assistant Professor in the Sino-US Global Logistics Institute, Shanghai Jiao Tong University. He received his B.Eng. degree in Industrial Engineering from Hunan University, his M.Eng. degree in Industrial Engineering from Shanghai Jiao Tong University, and his Ph.D. degree in Industrial Systems Engineering & Management from the National University of Singapore. Prior to joining Shanghai Jiao Tong University, Dr. Zhao worked as a Senior Data Scientist at GrabTaxi Holdings Pte. Ltd. His research interests include optimization under uncertainty, real options analysis, and dynamic decision-making problems.

Dr. William B. Haskell received his B.S. Mathematics and M.S. Econometrics degrees from the University of Massachusetts Amherst in 2005 and 2006, respectively. He then obtained his M.S. Operations Research, M.A. Mathematics, and Ph.D. Operations Research degrees from the University of California Berkeley in 2007, 2010, and 2011, respectively. He is currently an Assistant Professor in the Supply Chain and Operations Management Area in the Krannert School of Management at Purdue University. Dr. Haskell's research focus is on algorithms for convex optimization and dynamic programming, with an emphasis on risk-aware decision-making.

Dr. Michel-Alexandre Cardin is a Senior Lecturer (eq. Associate Professor) in Computational Aided Engineering at the Dyson School of Design Engineering, Imperial College London, where he leads the Strategic Engineering Laboratory. His work focuses on the development and evaluation of new computational aided methodologies, digital processes, and algorithms to support the design of engineering systems, with applications in infrastructure and financial systems. Before joining Imperial College, Dr. Cardin worked as a Quantitative Researcher in the hedge fund industry, developing strategies for derivatives trading using machine learning. He also worked as an Assistant Professor at the National University of Singapore, where he served as Principal Investigator on international collaborations like Future Resilient Systems with ETH Zürich, and Singapore-MIT Alliance for Research and Technology. Dr. Cardin holds a PhD in Engineering Systems and a Master of Science in Technology and Policy from MIT, a Master of Applied Science in Aerospace Engineering from the University of Toronto, Honors BSc in Physics from McGill University in Canada, and graduate of the Space Science Program at the International Space University. He is currently serving as Associate Editor for the *ASME Journal of Mechanical Design*, and served on the Editorial Boards of the *INCOSE Journal Systems Engineering* and *IEEE Transactions on Engineering Management*.

References

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.

Angelus, A., Porteus, E., and Wood, S. C. (2000). Optimal sizing and timing of modular capacity expansions. *Graduate School of Business Research Paper*, (1479R2).

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252.

Birge, J. R. (2000). Option methods for incorporating risk into linear capacity planning models. *Manufacturing & Service Operations Management*, 2(1):19–31.

Cardin, M.-A. and Hu, J. (2016). Analyzing the tradeoffs between economies of scale, time-value of money, and flexibility in design under uncertainty: Study of centralized versus decentralized waste-to-energy systems. *ASME Journal of Mechanical Design*, 138(1):011401–011401.

Cardin, M.-A., Xie, Q., Ng, T. S., Wang, S., and Hu, J. (2017). An approach for analyzing and managing flexibility in engineering systems design based on decision rules and multistage stochastic programming. *IISE Transactions*, 49(1):1–12.

de Neufville, R. and Scholtes, S. (2011). *Flexibility in Engineering Design*. MIT Press.

Delgado, D. and Claro, J. (2013). Transmission network expansion planning under demand uncertainty and risk aversion. *International Journal of Electrical Power & Energy Systems*, 44(1):696–702.

Dixit, A. K. and Pindyck, R. S. (1994). *Investment Under Uncertainty*. Princeton University Press.

Dixit, A. K. and Pindyck, R. S. (1998). Expandability, Reversibility, and Optimal Capacity Choice. *National Bureau of Economic Research*.

Eberly, J. C. and Van Mieghem, J. A. (1997). Multi-factor dynamic investment under uncertainty. *Journal of Economic Theory*, 75(2):345–387.

Geng, N., Jiang, Z., and Chen, F. (2009). Stochastic programming based capacity planning for semiconductor wafer fab with uncertain demand and capacity. *European Journal of Operational Research*, 198(3):899–908.

Georghiou, A., Wiesemann, W., and Kuhn, D. (2015). Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming*, 152(1-2):301–338.

Haskell, W. and Jain, R. (2015). A convex analytic approach to risk-aware Markov decision processes. *SIAM Journal on Control and Optimization*, 53(3):1569–1598.

Huang, K. and Ahmed, S. (2009). The value of multistage stochastic programming in capacity planning under uncertainty. *Operations Research*, 57(4):893–904.

Hugonnier, J. and Morellec, E. (2007). Real options and risk aversion. SSRN Scholarly Paper ID 422600, Social Science Research Network, Rochester, NY.

Kouvelis, P. and Tian, Z. (2014). Flexible capacity investments and product mix: optimal decisions and value of postponement options. *Production and Operations Management*, 23(5):861–876.

Krokhmal, P., Palmquist, J., and Uryasev, S. (2002). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of risk*, 4:43–68.

Li, H., Zhou, C., Lee, B. K., Lee, L. H., Chew, E. P., and Goh, R. S. M. (2017). Capacity planning for mega container terminals with multi-objective and multi-fidelity simulation optimization. *IISE Transactions*, 49(9):849–862.

Maceira, M., Marzano, L., Penna, D., Diniz, A., and Justino, T. (2015). Application of CVaR risk aversion approach in the expansion and operation planning and for setting the spot price in the Brazilian hydrothermal interconnected system. *International Journal of Electrical Power & Energy Systems*, 72:126–135.

Manne, A. S. (1961). Capacity expansion and probabilistic growth. *Econometrica*, 29(4):632–649.

Martínez-Costa, C., Mas-Machuca, M., Benedito, E., and Corominas, A. (2014). A review of mathematical programming models for strategic capacity planning in manufacturing. *International Journal of Production Economics*, 153:66–85.

Munos, R. and Szepesvári, C. (2008). Finite-time bounds for fitted value iteration. *Journal of Machine Learning Research*, 9(May):815–857.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.

Ruszczyński, A. (2010). Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming*, 125(2):235–261.

Sarykalin, S., Serraino, G., and Uryasev, S. (2008). Value-at-risk vs. conditional value-at-risk in risk management and optimization. In Chen, Z.-L., Raghavan, S., Gray, P., and Greenberg, H. J., editors, *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*, pages 270–294. INFORMS.

Shapiro, A. (2011). Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63–72.

Shapiro, A. and Nemirovski, A. (2005). On complexity of stochastic programming problems. In *Continuous Optimization*, pages 111–146. Springer.

Sun, Y. and Schonfeld, P. (2015). Stochastic capacity expansion models for airport facilities. *Transportation Research Part B: Methodological*, 80:1–18.

Swaminathan, J. M. (2002). Tool procurement planning for wafer fabrication facilities: a scenario-based approach. *IIE Transactions*, 34(2):145–155.

Szolgayová, J., Fuss, S., Khabarov, N., and Obersteiner, M. (2011). A dynamic CVaR-portfolio approach using real options: an application to energy investments. *European Transactions on Electrical Power*, 21(6):1825–1841.

Taghavi, M. and Huang, K. (2016). A multi-stage stochastic programming approach for network capacity expansion with multiple sources of capacity. *Naval Research Logistics*, 63(8):600–614.

Taghavi, M. and Huang, K. (2018). A Lagrangian relaxation approach for stochastic network capacity expansion with budget constraints. *Annals of Operations Research*, 284:605–621.

Van Mieghem, J. A. (2003). Commissioned paper: Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management*, 5:269–302.

Wu, C.-H. and Chuang, Y.-T. (2010). An innovative approach for strategic capacity portfolio planning under uncertainties. *European Journal of Operational Research*, 207(2):1002–1013.

Yu, G., Haskell, W. B., and Liu, Y. (2017). Resilient facility location against the risk of disruptions. *Transportation Research Part B: Methodological*, 104:82–105.

Zhang, S. and Cardin, M.-A. (2017). Flexibility and real options analysis in emergency medical services systems using decision rules and multi-stage stochastic programming. *Transportation Research Part E: Logistics and Transportation Review*, 107:120–140.

Zhao, S., Haskell, W. B., and Cardin, M.-A. (2017). An approximate dynamic programming approach for multi-facility capacity expansion problem with flexibility design. In *IIE Annual Conference Proceedings*, pages 440–445.

Zhao, S., Haskell, W. B., and Cardin, M.-A. (2018). Decision rule-based method for flexible multi-facility capacity expansion problem. *IISE Transactions*, 50(7):553–569.

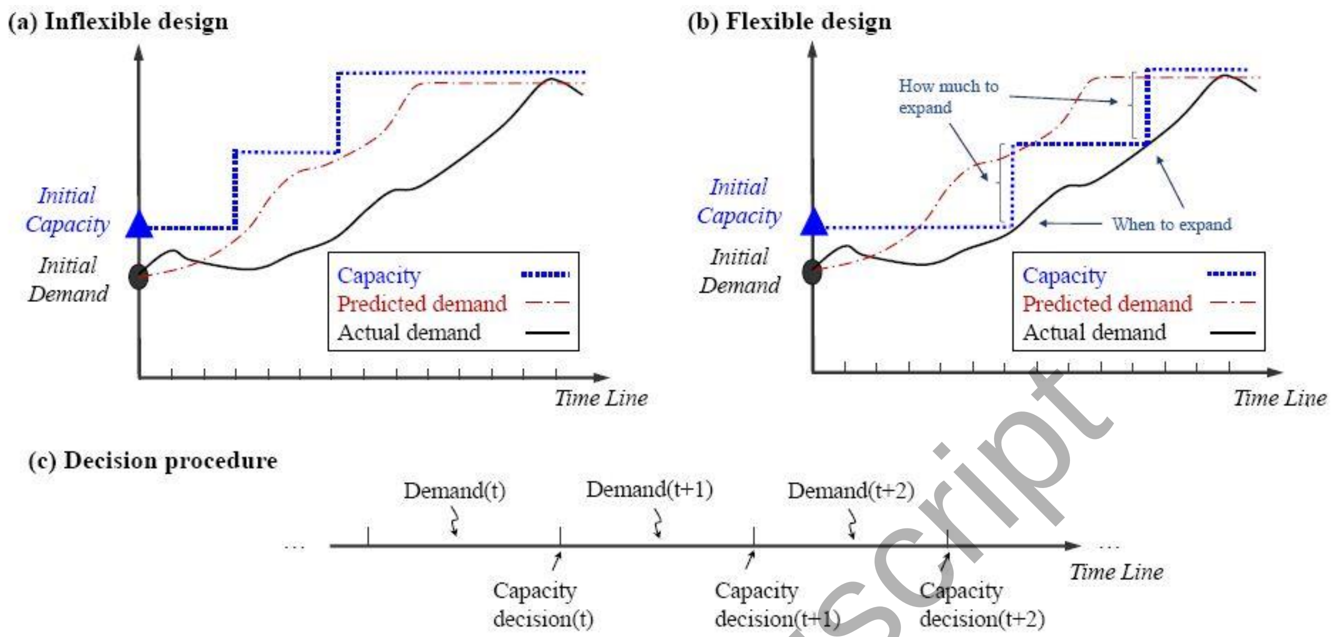

Fig. 1 Illustration of (a) inflexible design approach and (b) flexible design approach, and (c) decision procedure of a flexible capacity expansion policy.

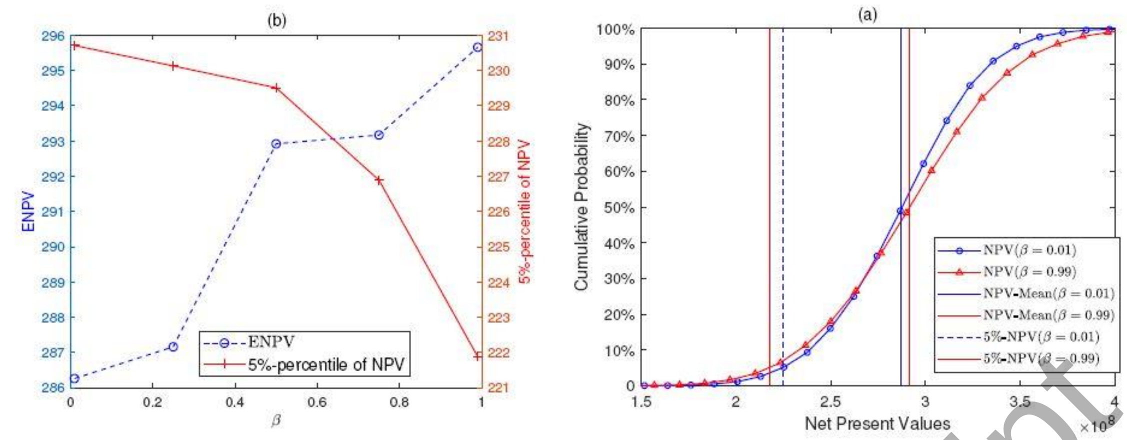

Fig. 2 (a) Simulation results given different *β*, and (b) cumulative distribution functions of different policies.

Table 1 Notations for the flexible MCEP

SETS	
$\mathcal{I}$	Set of customers, $i \in \mathcal{I}$ and $\lvert\mathcal{I}\rvert = I$
$\mathcal{N}$	Set of facilities, $n \in \mathcal{N}$, and $\lvert\mathcal{N}\rvert = N$
$\mathcal{T}$	Set of time periods, $t \in \mathcal{T}$, $\lvert\mathcal{T}\rvert = T$, and $\mathcal{T}_0 \triangleq \mathcal{T} \cup \{0\}$
$\mathcal{L}$	Set of line segments of the piecewise expansion cost, $l \in \mathcal{L}$ and $\lvert\mathcal{L}\rvert = L$
Ξ_t	Sample space of the uncertain demand in time $t \in \mathcal{T}$
$\mathbb{K}$	The feasible set of capacity $\mathbb{K} \triangleq \left\{ K \in \mathbb{Z}_+^N \,\middle\vert\, K \le K^{\max} \right\}$
$\overline{\mathcal{K}}$	Set of capacity expansion policies that are feasible to MCEP
Θ	Admissible set for parameters of the if–then decision rule, $(\theta_1, \theta_2) \in \Theta$
PARAMETERS AND FUNCTIONS	
ξ_{it}	Amount of demand generated from customer $i \in \mathcal{I}$ in time $t \in \mathcal{T}$; its vector form is $\xi_t = (\xi_{1t}, \ldots, \xi_{It})$
$K_n^{\max}$	The maximum capacity of facility $n \in \mathcal{N}$; its vector form is $K^{\max} = \left(K_1^{\max}, \ldots, K_N^{\max}\right)$
γ	Discount factor, $0 < \gamma \le 1$
r_{int}	Unit revenue from satisfying customer $i \in \mathcal{I}$ with facility $n \in \mathcal{N}$ in time $t \in \mathcal{T}$
b_{it}	Unit penalty cost for unsatisfied customer $i \in \mathcal{I}$ in time $t \in \mathcal{T}$

SETS	
p_{nlt}, q_{nlt}	Slope/intercept of the l^{th} line segment of the expansion costs corresponding to facility n in time t
α	Confidence level of CVaR, $0 < \alpha < 1$
β	Weight factor of the objective function, $0 \le \beta \le 1$
$\Pi_t(\cdot)$	Profit function given the installed capacity and realized demand in time $t \in \mathcal{T}$
$c_t(\cdot)$	Expansion cost function in time $t \in \mathcal{T}$
$Q(\cdot)$	Cumulative future costs given a policy $K_{[T]} \in \overline{\mathcal{K}}$ and a particular $\xi \in \Xi$
VARIABLES	
K_{nt}	Capacity of facility $n \in \mathcal{N}$ at the end of time $t \in \mathcal{T}_0$; its vector form is $K_t = (K_{1t}, \dots, K_{Nt})$
$\mathcal{K}_t(\cdot)$	Expansion policy mapping from the historical demand $\xi_{[t]}$ to the capacity decisions in time $t \in \mathcal{T}$
z_{int}	Amount of demand allocated from customer $i \in \mathcal{I}$ to facility $n \in \mathcal{N}$ in time $t \in \mathcal{T}$
u	Auxiliary variable for CVaR
$\theta_{1,n}$	Parameter to adjust capacity when the if–then rule for facility $n \in \mathcal{N}$ is triggered; its vector form is
	$\theta_1 = (\theta_{1,1}, \dots, \theta_{1,N})$
$\theta_{2,nt}$	Threshold for the trigger condition of the if–then rule for facility $n \in \mathcal{N}$ in time $t \in \mathcal{T}$; its vector form
	is $\theta_2 = (\theta_{2,11}, \dots, \theta_{2,NT})$

Table 2 Simulation results with varying α and β (unit: million S$).

Method	α	β	**Metric (unit: million S$)**					**VoF**
			Min	5^{th} - percentile	ENPV	95^{th} - percentile	Max	
Flexible policy	0.95	0.01	78.2	230.7	286.3	343.0	431.2	48.3
	0.95	0.25	79.4	230.1	287.2	344.3	431.6	49.2
	0.95	0.5	80.4	229.5	292.9	356.6	435.6	55.0
	0.95	0.75	87.8	226.9	293.2	359.8	444.6	55.2
	0.95	0.99	83.6	221.9	295.7	365.8	461.5	57.7
	0.50	0.01	93.2	228.4	291.5	351.8	420.7	53.5
	0.05	0.01	119.5	221.7	295.2	364.6	451.7	57.2
Inflexible policy	-	-	–246.9	145.0	238.0	307.3	346.5	-

Table 3 Sensitivity analysis on the percentage volatility (unit: million S$).

Case	Policy	5^{th} -percentile	Mean	VoF
GBM(0.04, 0.08)	Flexible policy ($\beta = 0.01$)	280.4	314.5	37.0
	Flexible policy ($\beta = 0.99$)	277.4	315.9	38.6
	Inflexible policy	228.9	277.3	-
GBM(0.04, 0.16)	Flexible policy ($\beta = 0.01$)	230.7	286.3	48.3
	Flexible policy ($\beta = 0.99$)	221.9	295.7	57.7
	Inflexible policy	145.0	238.0	-
GBM(0.04, 0.24)	Flexible policy ($\beta = 0.01$)	166.4	238.6	48.0
	Flexible policy ($\beta = 0.99$)	159.2	261.1	70.5
	Inflexible policy	60.1	190.6	-

Table 4 Comparing the proposed method with the benchmarks.

	Method	Metric (unit: million S$)					CPU time
		Min	5^{th} -per.	Mean	95^{th} -per.	Max	
Flexible policy	ADP	123.0	224.2	296.2	372.0	471.1	≥ 100 h
	Proposed method	117.6	219.5	295.8	366.1	448.5	391 s
	BACD	82.8	220.1	293.4	361.3	425.7	4107 s
Inflexible policy	-	–246.5	144.9	237.8	306.5	343.2	351 s

Table 5 Scalability test of the proposed method given $N = 5$ and $S = 4000$.

Method	Indicator	$(I = 8, T = 15)$	$(I = 10, T = 15)$	$(I = 10, T = 20)$
Proposed method	CPU time (seconds)	473	508	754
	ENPV (million S$)	237.7	306.3	400.9
BACD	CPU time (seconds)	5499	6749	20437
	ENPV (million S$)	237.0	307.5	399.0